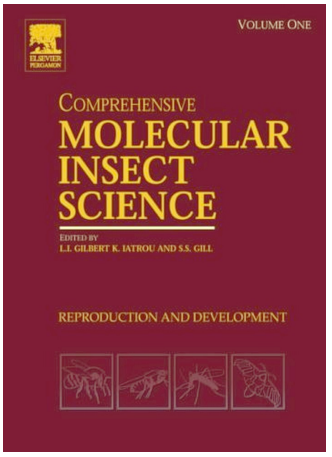
VOLUME FOUR

COMPREHENSIVE MOLECULAR INSECT SCIENCE

EDITED BY
L.I. GILBERT K. IATROU AND S.S. GILL

BIOCHEMISTRY AND MOLECULAR BIOLOGY



Comprehensive Molecular Insect Science

Seven-Volume Set, Volume 1-7

by Kostas Iatrou, Sarjeet S. Gill

Product Details

- **Hardcover:** 3300 pages
- **Publisher:** Pergamon (December 20, 2004)
- **Language:** English
- **ISBN:** 044451516X
- **Product Dimensions:** 12.0 x 9.5 x 9.9 inches
- **Shipping Weight:** 29.3 pounds.

List Price: $2,650.00

Book Description

The important developments and achievements in modern insect science since the first edition have led to this new revised, expanded and retitled reference work. New chapters in this edition include the ecdysone receptor, lipocalins, bacterial toxins, etc. Each article may also be read independently, as a review of that particular subject.

Comprehensive Molecular Insect Science also covers new technologies that have been developed since 1985. The Index Volume will help researchers to track specific subjects and particular species of insects cited in several volumes.

Available Online in 2005 via ScienceDirect!
' Extensive browse and search across subject, thematic, alphabetical, author and cited author indexes - as applicable to the wor k
' Basic and advanced search functionality within volumes, parts of volumes, or across the whole work
' Build, save and re-run searches as well as combine saved searches
' Internal cross-referencing between articles in the work, plus dynamic linking to journal articles and abstract databases, maki ng navigation flexible and easy
' All articles are available as full text HTML files, and as PDF files that can be viewed, downloaded or printed out in their or iginal print format

The online edition of this title is kept under regular editorial review, and enhanced as warranted to reflect the state of current knowledge. Enhancements may take place as frequently as annually and may consist of new versions of existing articles, new articles, or additional material.

Key Features:

' Revised edition of critically acclaimed **Comprehensive Insect Physiology, Biochemistry and Pharmacology** (1985, Pergamon Press, 13 Volume Set)
' Written by a distinguished international group of contributors
' Well-organized format provides for concise, readable entries, easy searches, and thorough cross-references
' Complete up-to-date coverage of many important topics - essential information for scientists, students, and professionals alik e

About the Author

Kostas Iatrou is the Director of the "Insect Molecular Genetics and Biotechnology" Research Group at the Institute of Biology, NCSR "Demokritos" and an adjunct Professor of the University of Calgary. He is the former Professor of Biochemistry and Molecular Biology at the Faculty of Medicine, University of Calgary, Canada and Director of the Institute of Biology, National Institute of Scientific Research "Demokritos", Athens Greece. His major research interests are in the areas of insect cell differentiation, development and reproduction. His parallel research activities include the development of insect, insect cell and insect virus-based functional genomics tools and technologies that can be used for a wide variety of biotechnological applications impacting on the fields of insect pest control human disease treatment.

Sarjeet S. Gill is Professor of Cell Biology in the Department of Cell Biology and Neuroscience at the University of California, Riverside, and Entomologist in the Agricultural Experimental Station on the same campus. His research interests are in biochemistry, molecular biology and insect toxicology, with a focus on mechanisms of bacterial and viral pathogenesis, and the role of the midgut and Malpighian tubules in insect homeostasis. Dr. Gill's research also employs the use of whole genome tools to analyze gene function. His research has been continually funded through the NIH and USDA, he continues to serve on national and international committees related to his research interests, and is an elected fellow of the American Association for the Advancement of Science.

Comprehensive Molecular Insect Science
Copyright © 2005 Elsevier B.V. All rights reserved.

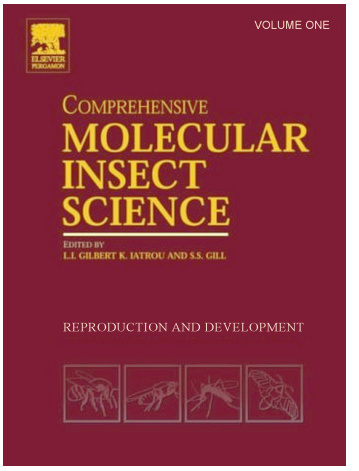
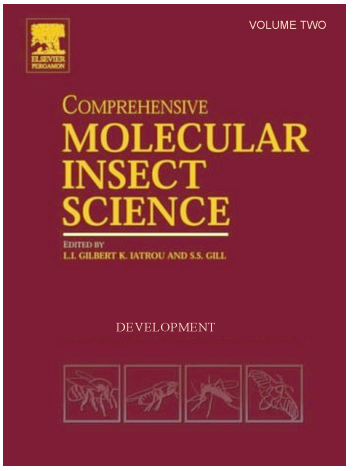


Table of Contents

Volume 1: Reproduction and Development

Volume 2: Development



2.1 Myogenesis and Muscle Development, Pages 1-43, S. M. Abmayr, L. Balagopalan, B. J. Galletta and S. -J. Hong

2.2 The Development of the Flight and Leg Muscle, Pages 45-84, J. O. Vigoreaux and D. M. Swank

2.3 Functional Development of the Neuromusculature, Pages 85-134, D. E. Featherstone and K. S. Broadie

2.4 Hormonal Control of the Form and Function of the Nervous System, Pages 135-163, J. W. Truman

2.5 Programmed Cell Death in Insect Neuromuscular Systems during Metamorphosis, Pages 165-198, S. E. Fahrbach, J. R. Nambu and L. M. Schwartz

2.6 Heart Development and Function, Pages 199-250, R. Bodmer, R. J. Wessells, E. C. Johnson and H. Dowse

2.7 Tracheal System Development and Morphogenesis, Pages 251-289, A. E. Uv and C. Samakovlis

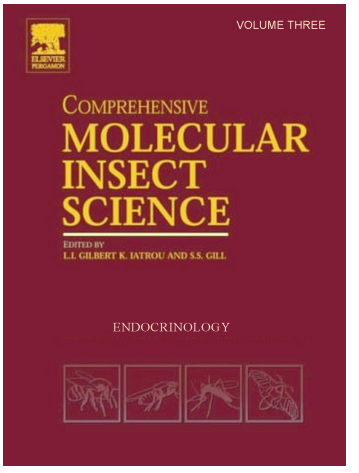
2.8 Development of the Malpighian Tubules in Insects, Pages 291-314, B. Denholm and H. Skaer

2.9 Fat-Cell Development, Pages 315-345, D. K. Hoshizaki

2.10 Salivary Gland Development and Programmed Cell Death, Pages 347-368, D. J. Andrew and M. M. Myat

2.11 Silk Gland Development and Regulation of Silk Protein Genes, Pages 369-384, E. Julien, M. Coulon-Bublex, A. Garel, C. Royer, G. Chavancy, J. -C. Prudhomme and P. Couble

Volume 3: Endocrinology



3.1 Neuroendocrine Regulation of Insect Ecdysis, Pages 1-60, D. Zitnan and M. E. Adams

3.2 Prothoracicotropic Hormone, Pages 61-123, R. Rybczynski

3.3 Ecdysteroid Chemistry and Biochemistry, Pages 125-195, R. Lafont, C. Dauphin–Villemant, J. T. Warren and H. Rees

3.4 Ecdysteroid Agonists and Antagonists, Pages 197-242, L. Dinan and R. E. Hormann

3.5 The Ecdysteroid Receptor, Pages 243-285, V. C. Henrich

3.6 Evolution of Nuclear Hormone Receptors in Insects, Pages 287-318, V. Laudet and F. Bonneton

3.7 The Juvenile Hormones, Pages 319-408, W. G. Goodman and N. A. Granger

3.8 Feedback Regulation of Prothoracic Gland Activity, Pages 409-431, S. Sakurai

3.9 Hormonal Control of Reproductive Processes, Pages 433-491, A. S. Raikhel, M. R. Brown and X. Belles

3.10 Hormones Controlling Homeostasis in Insects, Pages 493-550, D. A. Schooley, F. M. Horodyski and G. M. Coast

3.11 Circadian Organization of the Endocrine System, Pages 551-614, X. Vafopoulou and C. G. H. Steel

3.12 Hormonal Control of Diapause, Pages 615-650, D. L. Denlinger, G. D. Yocum and J. P. Rinehart

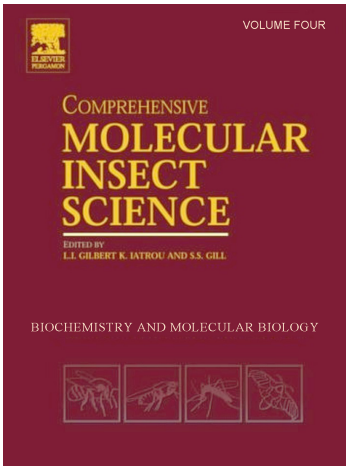
3.13 Endocrine Control of Insect Polyphenism, Pages 651-703, K. Hartfelder and D. J. Emlen

3.14 Biochemistry and Molecular Biology of Pheromone Production, Pages 705-751, G. J. Blomquist, R. Jurenka, C. Schal and C. Tittiger

3.15 Molecular Basis of Pheromone Detection in Insects, Pages 753-803, R. G. Vogt

3.16 Endocrinology of Crustacea and Chelicerata, Pages 805-842, E. S. Chang and W. R. Kaufman

Volume 4: Biochemistry and Molecular Biology



4.1 Insect Cytochrome P450, Pages 1-77, R. Feyereisen

4.2 Cuticular Proteins, Pages 79-109, J. H. Willis, V. A. Iconomidou, R. F. Smith and S. J. Hamodrakas

4.3 Chitin Metabolism in Insects, Pages 111-144, K. J. Kramer and S. Muthukrishnan

4.4 Cuticular Sclerotization and Tanning, Pages 145-170, S. O. Andersen

4.5 Biochemistry of Digestion, Pages 171-224, W. R. Terra and C. Ferreira

4.6 Lipid Transport, Pages 225-246, D. J. Van der Horst and R. O. Ryan

4.7 Proteases, Pages 247-265, M. R. Kanost and T. E. Clarke

4.8 Lipocalins and Structurally Related Ligand-Binding Proteins, Pages 267-306, H. Kayser

4.9 Eicosanoids, Pages 307-339, D. W. Stanley

4.10 Ferritin, Pages 341-356, J. J. Winzerling and D. Q. -D. Pham

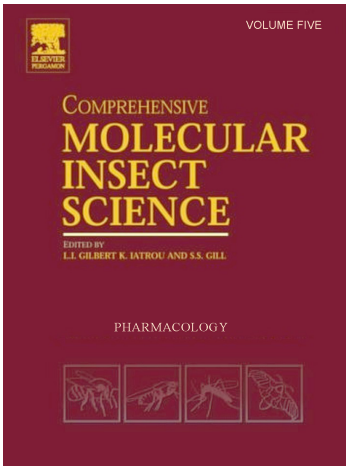
4.11 Tick-Talk, the Cellular and Molecular Biology of *Drosophila* Circadian Rhythms, Pages 357-394, P. H. Taghert and Y. Lin

4.12 Insect Transposable Elements, Pages 395-436, Z. Tu

4.13 Transposable Elements for Insect Transformation, Pages 437-474, A. M. Handler and D. A. O'Brochta

4.14 Insect Cell Culture and Recombinant Protein Expression Systems, Pages 475-507, P. J. Farrell, L. Swevers and K. Iatrou

Volume 5: Pharmacology



5.1 Sodium Channels, Pages 1-24, D. M. Soderlund

5.2 The Insecticidal Macrocyclic Lactones, Pages 25-52, D.Rugg, S. D. Buckingham, D. B. Sattelle and R. K. Jansson

5.3 Neonicotinoid Insecticides, Pages 53-105, P. Jeschke and R. Nauen

5.4 GABA Receptors of Insects, Pages 107-142, S. D. Buckingham and D. B. Sattelle

5.5 Insect G Protein-Coupled Receptors: Recent Discoveries and Implications, Pages 143-171, Y. Park and M. E. Adams

5.6 Scorpion Venoms, Pages 173-220, E. Zlotkin

5.7 Spider Toxins and their Potential for Insect Control, Pages 221-238, F. Maggio, B. L. Sollod, H. W. Tedford and G. F. King

5.8 Insecticidal Toxins from *Photorhabdus* and *Xenorhabdus*, Pages 239-253, R. H. ffrench-Constant, N. Waterfield and P. Daborn

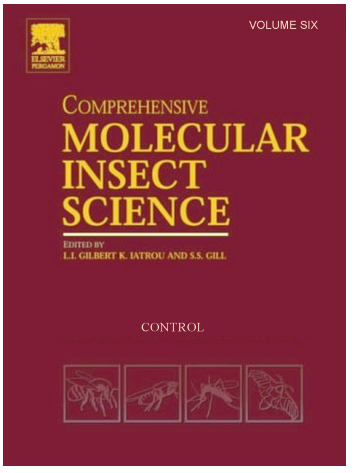
5.9 Amino Acid and Neurotransmitter Transporters, Pages 255-307, D. Y. Boudko, B. C. Donly, B. R. Stevens and W. R. Harvey

5.10 Biochemical Genetics and Genomics of Insect Esterases, Pages 309-381, J. G. Oakeshott, C. Claudianos, P. M. Campbell, R. D. Newcomb and R. J. Russell

5.11 Glutathione Transferases, Pages 383-402, H. Ranson and J. Hemingway

5.12 Insect Transformation for Use in Control, Pags 403-410, P. W. Atkinson, D. A. O'Brochta and A. S. Robinson

Volume 6: Control



6.1 Pyrethroids, Pages 1-29, B. P. S. Khambay and P. J. Jewess

6.2 Indoxacarb and the Sodium Channel Blocker Insecticides: Chemistry, Physiology, and Biology in Insects, Pages 31-53, K. D. Wing, J. T. Andaloro, S. F. McCann and V. L. Salgado

6.3 Insect Growth- and Development-Disrupting Insecticides, Pages 55-115, T. S. Dhadialla, A. Retnakaran and G. Smagghe

6.4 Azadirachtin, a Natural Product in Insect Control, Pages 117-135, A. J. Mordue (Luntz), E. D. Morgan and A. J. Nisbet

6.5 The Spinosyns: Chemistry, Biochemistry, Mode of Action, and Resistance, Pages 137-173, V. L. Salgado and T. C. Sparks

6.6 *Bacillus thuringiensis*: Mechanisms and Use, Pages 175-205, A. Bravo, M. Soberón and S. S. Gill

6.7 Mosquitocidal *Bacillus sphaericus*: Toxins, Genetics, Mode of Action, Use, and Resistance Mechanisms, Pages 207-231, J. -F. Charles, I. Darboux, D. Pauron and C. Nielsen-Leroux

6.8 Baculoviruses: Biology, Biochemistry, and Molecular Biology, Pages 233-270, B. C. Bonning

6.9 Genetically Modified Baculoviruses for Pest Insect Control, Pages 271-322, S. G. Kamita, K. -D. Kang, B. D. Hammock and A. B. Inceoglu

6.10 The Biology and Genomics of Polydnaviruses, Pages 323-360, B. A. Webb and M. R. Strand

6.11 Entomopathogenic Fungi and their Role in Regulation of Insect Populations, Pages 361-406, M. S. Goettel, J. Eilenberg and T. Glare

6.12 Pheromones: Function and Use in Insect Control, Pages 407-459, T. C. Baker and J. J. Heath

Volume 7: Indexes

EDITOR-IN-CHIEF

LAWRENCE I. GILBERT
University of North Carolina
Chapel Hill, NC
USA

Editors

KOSTAS IATROU
National Center for Scientific Research "Demokritos"
Athens, Greece

SARJEEET S. GILL
University of California
Riverside, CA
USA

INTRODUCTION

The first edition of this treatise, entitled *Comprehensive Insect Physiology, Biochemistry, and Pharmacology* and edited by Gerald A. Kerkut (see p. ix for the Dedication to Prof. Kerkut) and Lawrence I. Gilbert, was published in 1985. The preliminary meetings regarding that 13-volume series began in 1980 and it was agreed that there would be 12 volumes "that would provide an up-to-date summary and orientation on the physiology, biochemistry, pharmacology, behavior, and control of insects that would be of value to research workers, teachers, and students." It was felt that the volumes would provide a classical background to the literature, include all the critical basic material, and then emphasize the literature from 1950 to the present day (1984). By mid-1981 most of the chapters were assigned to authors and the project was underway. It was agreed upon that there would be a final Volume 13, an Index volume, that combined the subject, species, and author indexes for all 12 volumes so that all material in those volumes could be located easily. This was a monumental undertaking and the 13 volumes finally appeared in 1985. It received plaudits from reviewers and from researchers around the world. One of us (LG.) was approached by Carrol Williams, an internationally renowned insect physiologist, at an International Congress of Entomology following the publication of the first edition, and Professor Williams noted that he had a choice between purchasing the 13-volume set for his laboratory and buying a new centrifuge. He then stated that he chose the 13-volume series and had never regretted it for a moment. There were more than 50 000 references to the literature in the first edition and more than 10 000 species of insects were referred to, and all of this information was readily available through the thirteenth volume (index). It is of interest that volumes of that original series are still being purchased some 19 years after this series was published.

Over the past 5–6 years, one of us (LG.) was approached many times by colleagues around the world asking when a second edition of this series would be available. There was an obvious need because of the dramatic explosion of knowledge in insect science due to the utilization of molecular biological paradigms and techniques culminating in the elucidation of the *Drosophila* genome (fly database), the *Anopholes* genome, and of course this explosion continues. The use of this technology was really only beginning when the first series was developed and in 2000, 15 years after the publication of the first edition, we and Elsevier agreed that a new edition was necessary. This would be a series in which each chapter would begin by devoting a few pages summarizing the work in a particular research area up to the publication of the first series (1984–1985), and the remainder of the chapter would be a review of the literature from 1985 to 2003. It was agreed upon that each chapter would make full use of data arising from the utilization of molecular technologies if such data were applicable to that particular chapter subject. We chose to fulfill that premise in six volumes plus an index volume and that both a printed version and an electronic version would be available. As is evident, the title of the series has been changed, both to reflect the emphasis on molecular approaches and to alleviate the use of the obviously cumbersome title *Comprehensive Insect Physiology, Biochemistry, Pharmacology, and Molecular Biology,* which would have to be typed every time a researcher refers to a chapter in the series. We do not pretend, however, that this is a truly comprehensive treatise, since critical areas of research, e.g., ecology, evolutionary biology, and behavior, have not been included, but we have chosen to retain the use of the word "Comprehensive" for continuity with the first series.

The reader will note that the number of chapters in the six volumes is certainly well below the number of chapters in the original 12 volumes. The editors made very subjective decisions based on their own experience and knowledge regarding those areas where the most progress had been made over the past 18 years and omitted areas that have not grown significantly. Surely, there are areas of omission that many readers will feel were due to arbitrary decisions or lack of knowledge of the editors. That is bound to be the case although we spent a great deal of time deciding which chapters should be included, but there is little doubt that personal research bias played a role in those decisions. We are certainly confident that these volumes will be of great value to the research community, including postdoctoral investigators and graduate students. The new publishing technology will allow upgrades by the publisher every several years for those areas of research in which significant breakthroughs occur. This, of course, will only occur in the electronic format of the series and it is hoped that at least for the first few upgrades the original authors will take part.

Lawrence I. Gilbert
Kostas Iatrou
Sarjeet Gill

DEDICATION

Professor Gerald Allan Kerkut: 19 August 1927–6 March 2004

Gerald Kerkut began his scientific career in Cambridge, graduating with a first class degree in Natural Sciences (1945–1948). He then elected to remain at Cambridge for his Ph.D. (1948–1951), studying locomotion in starfish under the supervision of Professor Eric Smith in the Department of Zoology. Gerald continued at Cambridge for a further three years (1951–1954) as a junior fellow of his college, Pembroke. During this period, he studied the electrical activity of the gastropod central nervous system, initially selecting slugs but, due to problems of identification, changed to the garden snail, *Helix aspersa*. A species he was to use for nearly 30 years.

In 1954 Gerald moved to the University of Southampton to take up a lectureship in Animal Physiology within the Department of Zoology. Together with Professor Kenneth A. Munday, Gerald established the Department of Physiology and Biochemistry in 1959 and in 1966 was appointed to the second chair of Physiology and Biochemistry. He remained at Southampton throughout his academic life, retiring in 1992, when he was appointed Emeritus Professor. Gerald continued an active association with the University up until his death. He was invited by a number of universities to become chairman of their zoology or physiology departments but he always declined, preferring to remain at Southampton.

Gerald Kerkut had a first class analytical mind, a prodigious capacity for hard work and an extensive knowledge of the scientific literature. He was always ready to challenge accepted dogma and this was very well illustrated in his inaugural lecture entitled "The Missing Pieces," delivered in December 1968 (Kerkut, 1969). In this he reviewed his research on a number of topics, including the variable ionic composition of neurons, the role of the sodium–potassium pump in the maintenance of the resting potential, fast orthodromic and slow antidromic axon transport and amino acids as transmitters. Although much of his earlier research used the snail, Gerald also worked on insects where he was interested in the effect of sudden temperature changes on the nervous system. Using the cockroach leg preparation he observed that a sudden fall in temperature resulted in a transient increase in activity. Intracellular recordings showed that a decrease in temperature resulted in a depolarization of the membrane potential, which was responsible for the transient increase in activity. Q-10 values for changes in membrane potential of insect muscle were greater than those predicted from the Nernst equation. This led to his work on electrogenic metabolic pumps, summarized in his book with Barbara York, "*The Electrogenic Sodium Pump*" (Kerkut and York, 1971). Gerald also employed insect preparations for his research on glutamic acid as a transmitter and that certain insect neuron cell bodies possessed overshooting action potentials, for example, the octopamine-containing dorsal unpaired median (DUM) cells.

Having been a great advocate of the use of isolated invertebrate central nervous system preparations, Gerald turned his attention to isolated mammalian preparations in the late 1970s. In particular, he developed the use of the isolated spinal cord in conjunction with Jeff Bagust. Gerald was particularly interested in sensory integration in the dorsal horn of the spinal cord. Although a great protagonist of

isolated preparations, Gerald stressed that observations using isolated preparations must always be related to the living animal.

Gerald Kerkut was also very involved in the publication of scientific data both in terms of books and journals. In the late 1950s he met Robert Maxwell, the founder of Pergamon Press. They immediately formed an excellent rapport, and Maxwell encouraged Gerald to explore his idea of starting a journal in comparative physiology and biochemistry. To cut a long story short, this resulted in the publication of *Comparative Biochemistry and Physiology* with its first number in 1960 and Gerald as co-editor with Bradley T. Scheer. The journal proved a great success and eventually expanded into three sections, namely, Physiology, Biochemistry, and Pharmacology. Gerald continued to edit the journal until 1994. Another of his editing successes was *Progress in Neurobiology*, which he co-edited with John Phillis. Gerald also edited a very successful series of monographs in Zoology, which were published by Pergamon Press. Gerald's interest in publishing work on insects is illustrated by his co-editing with Lawrence Gilbert in 1985 of the first edition of *Comprehensive Insect Physiology, Biochemistry, and Pharmacology*. He was particularly proud of the high standard of scholarship in these volumes.

Gerald was always interested in evolution and with Maxwell's support published "*The Implications of Evolution*" in 1960 (Kerkut, 1960). In this he examined the problem of the evolution and interrelationships of the animal phyla. This was a topic of lasting interest to him and one to which he returned in a mini-review entitled "*Possible Evolutionary Futures for Mankind*" (Kerkut, 1988). In 1988 a Festschrift was organized in Gerald's honor and the proceedings published as a special edition of *Comp. Biochem. Physiol.* Gerald provided the opening chapter, in which he reviewed his research career up to that time (Kerkut, 1989).

Gerald Kerkut enjoyed music, art, and travel. He was an accomplished pianist and had an extensive collection of art books. Up until 6–7 years ago, Gerald traveled widely in the Americas and the Far East. Gerald enjoyed teaching and interacting with undergraduates and postgraduates. Many will retain enduring memories of his humor and concerned interest in their welfare. During his active research period, Gerald trained over 80 postgraduates and their success will provide a lasting legacy to his memory.

References

Kerkut, G.A., **1960**. The Implications of Evolution. Oxford, Pergamon, p. 174.
Kerkut, G.A., **1969**. The Missing Pieces. University of Southampton, Southampton, UK, p. 15.
Kerkut, G.A., **1988**. Possible evolutionary futures for mankind. *Comp. Biochem. Physiol. 90A*, 5–10.
Kerkut, G.A., **1989**. Studying the isolated central nervous system; a report on 35 years: more inquisitive than aquisitive. *Comp. Biochem. Physiol. 93A*, 9–24.
Kerkut, G.A., York, B., **1971**. The Electrogenic Sodium Pump. Scientechnica, Bristol, UK, p. 182.

Robert Walker
Head of Honours School
School of Biological Sciences
University of Southampton

PERMISSION ACKNOWLEDGMENTS

The following material is reproduced with kind permission of Nature Publishing Group

Figure 11 of Chapter 1.8 Insect Homeotic Complex Genes and Development, Lessons from *Drosophila* and Beyond

Figures 4 and 11 of Chapter 1.12 The Development of the Olfactory System

Figures 3, 6, 7, 8 and 9 of Chapter 3.5 The Ecdysteroid Receptor

Figure 5 of Chapter 3.11 Circadian Organization of the Endocrine System

Figure 9 of Chapter 5.4 GABA Receptors of Insects

http://www.nature.com/nature

The following material is reproduced with kind permission of American Association for the Advancement

Figure 8 of Chapter 5.4 GABA Receptors of Insects

http://www.sciencemag.org

Notes on the Subject Index

To save space in the index the following abbreviations have been used:

ETH – ecdysis triggering hormone
GPCRs – G protein-coupled receptors
PBAN – pheromone biosynthesis activating neuropeptide
PDV – polydnaviruses
PTTH – prothoracicotropic hormone
QSAR – qualitative structure-activity relation
RDL – resistance to dieldrin

4.1 Insect Cytochrome P450

R Feyereisen, INRA, Centre de Recherches de Sophia Antipolis, Sophia Antipolis, France

© 2005, Elsevier BV. All Rights Reserved.

4.1.1. Introduction

4.1.1.1. Overview

Cytochromes P450 or *CYP* genes constitute one of the largest family of genes with representatives in virtually all living organisms, from bacteria to protists, plants, fungi, and animals (Werck-Reichhart and Feyereisen, 2000). An ever-growing number of P450 sequences are available, and the role of P450 enzymes is being documented in an ever-growing number of physiological processes. The human genome carries about 57 P450 genes and insect genomes carry about a hundred. Each P450 protein is the product of a distinct *CYP* gene and P450 diversity is the result of successive gene (or genome) duplications followed by sequence divergence. The typically 45–55 kDa P450 proteins are heme-thiolate enzymes that are known collectively to catalyze at least 60 chemically distinct reactions. Their essential common feature is the absorbance peak near 450 nm of their Fe^{II}–CO complex for which they are named. P450 enzymes are best known for their monooxygenase role, catalyzing the transfer of one atom of molecular oxygen to a substrate and reducing the other to water. The simple stoichiometry (eqn [1]) commonly describes the monooxygenase or mixed-function oxidase reaction of P450:

$$RH + O_2 + NADPH + H^+ \rightarrow ROH + H_2O + NADP^+ \quad [1]$$

However, oxygen atom transfer is not the only catalytic function of P450 enzymes. They also show activity as oxidases, reductases, desaturases, isomerases, etc. (Mansuy, 1998). There are soluble forms of P450 (in bacteria) and membrane-bound forms (in microsomes and mitochondria of eukaryotes). P450s are dependent on redox partners for their supply of reducing equivalents from NADH or NADPH (NADPH cyt P450 reductase and cyt b_5 in

microsomes; a ferredoxin and a ferredoxin reductase in mitochondrial and bacterial systems). Some bacterial P450s are fusion proteins with their redox partners, either P450 reductase (see below) or a phthalate family oxygenase reductase (PFOR) (De Mot and Parret, 2002). Because of their complex catalytic mechanism, P450 enzymes often, perhaps always, generate superoxide or hydrogen peroxide from "unsuccessful," or uncoupled reactions, leading to oxidative stress in cells.

Many P450 proteins are specialized in the metabolism of endogenous substrates (steroid hormones, lipids, etc.) but much of their notoriety has been associated with the metabolism or detoxification of xenobiotics (natural products, drugs, pesticides, etc.). In insects, they are involved in many cases of insecticide resistance. In well-studied cases, P450 enzymes generate more toxic compounds, e.g., carcinogenic metabolites of aryl hydrocarbons. Many foreign compounds, as well as endogenous metabolites, can induce the transcription of P450 genes through complex interactions with nuclear receptors and bHLH-PAS proteins, such as the *Ah* (aryl hydrocarbon) receptor of vertebrates. There are many excellent reviews and books clearly presenting the state of knowledge on P450s, generally with an emphasis on mammalian and bacterial enzymes (Omura *et al.*, 1993; Schenkman and Greim, 1993; Ortiz de Montellano, 1995b). Reliance on the advances made in noninsect systems is both necessary and risky. Necessary because the structural and functional homology of P450 enzymes allows general principles to emerge. Risky because such principles may not always apply to peculiar aspects of insect P450 evolution. The chapter will certainly fail to be comprehensive in view of the enormous literature pertaining to P450 but it will try to convey the state of our current understanding as a basis for future research. Many aspects of P450 biology are intertwined and complex (structure, activity, induction) so the linear and compartmentalized treatment of the subject as done here will probably fail to pass muster with the P450 novice and specialist alike. Indeed, the interpretations and choices of examples cited are mine, and the reader is invited to browse through text and references repeatedly. To help fill the enormous gaps in our knowledge on insect P450, the reader is also invited to join the field of P450 research.

4.1.1.2. Historical Framework

This chapter humbly tries to update the exhaustive reviews of Hodgson on microsomal monooxygenases (Hodgson, 1985) and of Agosin on the role of microsomal oxidations in insecticide degradation (Agosin, 1985) published in the first edition of this series. The present chapter will focus on advances since 1985, and the earlier literature will be cited when needed for clarity or historical import. Both 1985 chapters had insect cytochromes P450 as their main focus and both revealed the enticing complexity of P450 catalysis and regulation. These two chapters, as well as insightful reviews that preceded them (Wilkinson and Brattsten, 1972; Brattsten, 1979b; Brooks, 1979), give a good account of the spirit of the times.

The great variety of reactions catalyzed by P450s was well recognized, but it was not really known how the multiplicity of enzymes and the suspected loose substrate specificity were blended to achieve this variety. The existence of multiple P450 forms was accepted but the true measure of this multiplicity was not foreseen. This multiplicity, once understood at the molecular genetic level, led to the early adoption of a common nomenclature in 1987 – the *CYP* genes. The substrate specificity of individual P450 enzymes was a matter of conjecture, because enzyme purification from insect sources and reconstitution of significant activity *in vitro* proved to be extremely difficult. Many cases of insecticide resistance were clearly attributable to alterations in the P450 system – a quantitative increase and a qualitative change. The molecular basis for this change was unknown, only a few resistance genes had been formally recognized at the genetic level, and the nature of the mutations leading to resistance was barely discussed (Terriere and Yu, 1974).

The role of insect P450 enzymes in the biosynthesis and degradation of endogenous compounds was becoming increasingly clear, but their relation to P450s involved in xenobiotic metabolism was unclear – were they truly the same class of enzymes? In contrast, an evolutionary link between the capacity to metabolize, and resist, synthetic insecticides and the ability of insects to thrive on chemically well-defended plants was recognized early on (Gordon, 1961). The metabolism of plant secondary compounds by P450s and the role these compounds played in regulating P450 activities brought insects forward as exquisite models to study the ecological role of P450 enzymes. Insects have now become equally exquisite models to study the evolution of P450s.

The first cloning and sequencing of a P450 cDNA (rat CYP2B1) was in 1982 (Fuji-Kuriyama *et al.*, 1982) and that of an insect P450 (housefly CYP6A1) followed in 1989 (Feyereisen *et al.*, 1989). The first crystal structure of a soluble bacterial P450, P450cam, was published in 1985 (Poulos

et al., 1985). The first structure of a mammalian microsomal P450 engineered for solubility followed in 2000 (Williams *et al*., 2000a). After several heroic attempts at wresting activities from purified insect enzymes reconstituted *in vitro*, it was the heterologous expression of cloned insect P450 cDNAs in 1994 (Andersen *et al*., 1994; Ma *et al*., 1994) that allowed the rational study of individual P450 enzymology. Genome sequencing projects finally revealed the cast of P450 characters: insects have twice as many *CYP* genes as mammals, but only a third that of plants (David Nelson's cytochrome P450 homepage (web link)).

4.1.1.3. Vocabulary

Although P450 enzymes and *CYP* genes are displacing mixed-function oxidases (MFOs), microsomal oxidases, polysubstrate monooxygenases (PSMO) or simply monooxygenases in the vocabulary, words of caution are needed at the outset. The word "cytochrome" was initially given to the liver P450 pigment (Omura, 1993; Estabrook, 1996) and it is still associated with P450 enzymes. However, P450s are formally not cytochromes but rather heme-thiolate proteins (Mansuy, 1998). Not all P450-dependent reactions are monooxygenations, and not all monooxygenases are P450 enzymes (Mansuy and Renaud, 1995). In particular, flavin monooxygenases or FMOs are NADPH-dependent enzymes that catalyze some reactions similar or identical to those catalyzed by P450 enzymes. In mammalian species, FMOs are microsomal enzymes best known for their *N*- and *S*-oxidation activities (Ziegler, 2002). Until recently, FMOs had not been identified in insects. In the cinnabar moth, *Tyria jacobaea*, a soluble FMO specifically N-oxidizes pyrolizidine alkaloids such as senecionine, seneciphylline, monocrotaline, and axillarine (Lindigkeit *et al*., 1997; Naumann *et al*., 2002). Only two FMO genes of as yet unknown function have been recognized in the *Drosophila* genome, but there may well be many more in other species. Whether insect FMOs are involved in xenobiotic *N*- or *S*-oxidations is still unknown.

A word of caution, too, about the "microsomal" oxidases. This chapter will show that insect P450s can be found in microsomal membranes as well as in mitochondria, and that subcellular localization does not portend a particular physiological role or catalytic competence.

4.1.1.4. Nomenclature

A nomenclature of P450 genes and proteins was introduced when only 65 sequences were known (Nebert *et al*., 1987). Gene families were initially designated by Roman numerals, but the proliferation of diverse sequences rapidly became discouraging, even to those versed in classics. The rules of nomenclature were then revised to their current form (Nebert *et al*., 1991; Nelson *et al*., 1993, 1996) where a CYP prefix, followed by an arabic numeral designates the family (all members nominally >40% identical), a capital letter designates the subfamily (all members nominally >55% identical) and an arabic numeral designates the individual gene (all italics) or message and protein (no italics) (**Figure 1**). Different P450 enzymes are generally products of different genes; they are not isozymes or isoforms. The identity (%) rules for family and subfamily designations are not strictly adhered to, but names once adopted are rarely changed. Initially, many insect P450s were arbitrarily lumped into the CYP6 and the CYP4 families even though they had less than 40% amino acid identity with CYP6A1 or with vertebrate CYP4 proteins. Naming genes in the lumper mode made the CYP6 and CYP4 families the largest ones in insects by a cascade effect. CYP6B1 is only 32% identical to CYP6A1 (Cohen *et al*., 1992), so placing it in the CYP6 family "forced" many subsequent sequences into that family even if they did not meet the 40% criterion. The splitter mode prevailed at the completion of genome projects, which led to a new proliferation of CYP families in insects, the CYP300 series. A termite P450 claimed the welcoming designation of CYP4U2 (GenBank AF046011). Gotoh (1993) has introduced a useful nomenclature of higher order than CYP families: the E (for eukaryotic type) and B (for bacterial type) "classes" and subclasses (I, II, III, etc.) that regroup CYP families on the basis of phylogeny. Nelson (1998) has similarly introduced the notion of "clans," but the precise criteria for naming Gotoh's "classes" and Nelson's "clans" have not been defined.

Alleles of a gene are named as subscripts v1, v2 (e.g., *CYP6B1v2*, Cohen *et al*., 1992). The human P450 polymorphisms are named according to a clear nomenclature. Pseudogenes are noted by the

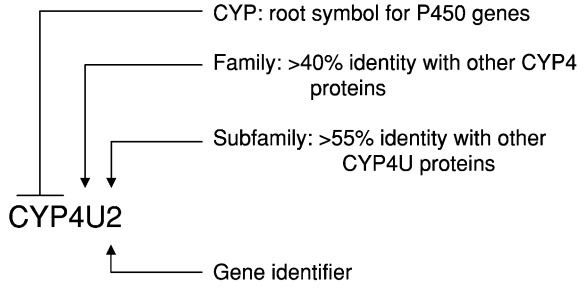


Figure 1 Scheme of the P450 nomenclature.

suffix P. This suffix ought to be added to the closest paralog that is an active gene, e.g., *CYP9E2* and *CYP9E2P1* in *Blattella germanica* (Wen *et al*., 2001). However this is not always done, as the closest paralog is sometimes not easily recognized. In following the tradition that predates the CYP nomenclature, P450 enzymes can be named with a small suffix, such as P450cam, the camphor hydroxylase of *Pseudomonas putida* later named CYP101; P450BM3 the fatty acid hydroxylase of *Bacillus megaterium* (CYP102); or P450scc, the cholesterol side-chain cleavage enzyme (CYP11A1). In insects, few P450 enzymes have been named in this way. P450Lpr is the predominant P450 in the pyrethroid-resistant strain Learn-Pyr of the housefly, later identified as CYP6D1 (Tomita and Scott, 1995). P450hyd (Reed *et al*., 1994) is the P450 forming hydrocarbons in the housefly. P450MA is a P450 purified from the Munsyana strain of the German cockroach (Scharf *et al*., 1998). In the *Drosophila* gene nomenclature (Lindsley and Zimm, 1992), only the initial letter is capitalized, hence *CYP6A1* in the housefly and *Cyp6a2* in *Drosophila*.

The CYP nomenclature of Nebert *et al*. was clearly designed to reflect the evolutionary relationships between the genes as evidenced by the degree of sequence identity of the proteins they encoded. As such it is a Darwinian nomenclature. P450 proteins have been categorized into classes (Ravichandran *et al*., 1993) that reflect the types of electron delivery to the active site. Class I proteins require both an FAD-flavoprotein reductase and a ferredoxin-type protein, class II P450s require only an FAD and FMN diflavin reductase, class III enzymes are self-sufficient, and class IV P450s receive electrons directly from NADPH. The utility of this nonevolutionary classification is debatable.

4.1.2. Diversity and Evolution of Insect P450 Genes

4.1.2.1. Sequence Diversity

4.1.2.1.1. P450 sequences from classical cloning techniques The first insect P450s cloned and sequenced were CYP6A1 from *Musca domestica* in 1989 (Feyereisen *et al*., 1989), CYP4C1 from *Blaberus discoidalis* in 1991 (Bradfield *et al*., 1991), CYP6A2 and CYP4D1 from *Drosophila* (Gandhi *et al*., 1992; Waters *et al*., 1992) as well as CYP6B1 from *Papilio polyxenes* in 1992 (Cohen *et al*., 1992), and CYP4D2 from *Drosophila* in 1994 (Frolov and Alatortsev, 1994). The methods used in these early studies were screening of cDNA expression libraries with polyclonal (CYP6A1) or monoclonal (CYP6A2) antibodies to (partially) purified P450 proteins of insecticide-resistant flies (CYP6A1, CYP6A2). For CYP6B1, microsequencing of N-terminal and internal sequences of a P450-sized band on a SDS-PAGE gel of *P. polyxenes* larval midgut proteins led to the design of degenerate oligonucleotide probes for RT-PCR on midgut poly(A)$^+$ RNA. The cloned PCR product was used in turn as a probe to screen a xanthotoxin-induced midgut cDNA library. Significantly, two cDNAs representing alleles of the *CYP6B1* gene were thus isolated. Classical molecular approaches of this type have continued to yield new P450 sequences (Tomita and Scott, 1995; Wang and Hobbs, 1995; Winter *et al*., 1999) and the initial P450 sequences have served as probes to isolate related sequences in the same species (Cohen and Feyereisen, 1995; Hung *et al*., 1995a, 1996; Maitra *et al*., 1996), or in phylogenetically close species (Li *et al*., 2000a, 2001). Interestingly, CYP9A1 of *Heliothis virescens* was cloned by screening an expression library with monoclonal antibodies that also served to clone CYP6A2 of *Drosophila*, despite the fact that the two sequences are only 32.4% identical (Rose *et al*., 1997).

4.1.2.1.2. Serendipity and P450 discovery In contrast to the targeted approaches, CYP4C1 was obtained in 1991 by differential screening of a cockroach fat body cDNA library (Bradfield *et al*., 1991). The probes consisted of cDNA obtained from fat bodies of hypertrehalosemic hormone-treated roaches, or decapitated controls. The cloning and identification of *Drosophila Cyp4d1* and *Cyp4d2* were equally serendipitous, as investigators were interested in transcripts in the *prune* region at the tip of the X chromosome (Gandhi *et al*., 1992; Frolov and Alatortsev, 1994). *Drosophila Cyp18* was initially cloned in a screen for ecdysone-inducible genes (Hurban and Thummel, 1993) and subsequently obtained as full-length cDNA (Bassett *et al*., 1997). *Cyp4e1* deserves a special historical mention. Its sequence was discovered independently in a 9 kb genomic sequence of the 44D region of *Drosophila* encompassing a cluster of cuticle protein genes (GenBank Acc. K00045) by T. Holton and D. Nero in 1991, each searching databases with P450 sequences (personal communications). This partial sequence of insect *Cyp4e1* is thus the first (1983), though not annotated, record of an insect P450 in GenBank.

4.1.2.1.3. Exponential amplification of new P450 sequences by PCR By 1994, sufficient information about vertebrate and insect P450s allowed the

isolation of fragments of P450 cDNAs and genes by PCR methods with degenerate oligonucleotides corresponding to consensus sequences. A variety of approaches were taken, two of which being particularly successful: in the first, sequences in the I helix and surrounding the conserved cysteine (see Section 4.1.3.3.1) were used to isolate PCR products of about 450–500 bp that coded for about 130 amino acids or almost a third of the full-length P450. In the second, the sequence around the conserved cysteine was used to design a first primer, with oligo(dT) serving as anchor on the poly(A)$^+$ message. Fragments of varying length were obtained by this 3′-RACE strategy, the C-terminal 30–50 amino acids sequence of the P450 and a variable 3′ UTR sequence (review: Snyder *et al.*, 1996). The PCR approaches led to the description of 17 new *CYP4* genes in *A. albimanus* (Scott *et al.*, 1994); 5 *CYP4* genes from *Manduca sexta* (Snyder *et al.*, 1995); 8 *CYP* genes from *H. armigera* (Pittendrigh *et al.*, 1997); 4 genes of the new *CYP28* family from *Drosophila* species (Danielson *et al.*, 1997); 14 P450 fragments from *Ceratitis capitata* (Danielson *et al.*, 1999), 95 new sequences from 16 drosophilid species (Fogleman *et al.*, 1998), etc. (Amichot *et al.*, 1994; Stevens *et al.*, 2000; Tares *et al.*, 2000; Scharf *et al.*, 2001; Wen *et al.*, 2001; Ranson *et al.*, 2002b). The PCR method has often been used as a first step in the isolation of full-length P450 clones in insects (Snyder *et al.*, 1995; Danielson and Fogleman, 1997; Hung *et al.*, 1997; Danielson *et al.*, 1998, 1999; Guzov *et al.*, 1998; Kasai *et al.*, 1998a, 2000; Ranasinghe and Hobbs, 1998; Sutherland *et al.*, 1998; Stevens *et al.*, 2000; Wen *et al.*, 2001; Wen and Scott, 2001b; Liu and Zhang, 2002) and ticks (Crampton *et al.*, 1999; He *et al.*, 2002).

The large number of partial P450 sequences obtained by cloning and sequencing an even larger number of PCR products has led to two related problems. The first is that small sequence differences can be found between very closely related PCR products. Are these artifactual, or do they represent allelic variation? In the study on *A. albimanus* for instance, 64 clones were sequenced of which 47 encoded P450 fragments, describing 17 genes (Scott *et al.*, 1994). In some clones, up to seven nucleotide differences from the closest sequence were seen. A few nucleotides may not just differentiate two alleles of the same gene, but may be sufficient to differentiate two genes, as the complete sequence of the *Drosophila* and *A. gambiae* genomes subsequently showed. This causes a nomenclature problem, when two distinct genes are prematurely described as allelic variants of a single gene. The second problem is that the P450 fragments obtained by PCR, one third of the full sequence or less, make the calculation of percentage identity (the base of the nomenclature rules) difficult. In the *A. albimanus* study (Scott *et al.*, 1994) this problem was resolved by establishing a function that derives identity over the full length P450 from the percentage identity of the PCR product. Further validations of this approach have been presented (Danielson *et al.*, 1999; Fogleman and Danielson, 2000). Nonetheless, the practice of bestowing an official CYP designation to a short, partial P450 sequence should be abandoned.

4.1.2.1.4. Genome sequences: diversity revealed
The *D. melanogaster* complete genome sequence was published in 2000 (Adams *et al.*, 2000). Annotation of the P450 sequences (Tijet *et al.*, 2001) was done by multiple Basic Local Alignment Search Tool (BLAST) searches, taking into account intronexon structure, pairwise and multiple alignments, EST sequences, and known features of well-characterized P450. This task gave 90 sequences, of which 83 appeared to code for potentially functional P450s. Seven sequences were either partial sequences or obvious pseudogenes. Forty genomic sequences were not represented by ESTs. With release 3 of the *Drosophila* genome (Celniker *et al.*, 2002), the "official count" of potentially functional P450 genes is 85 with five pseudogenes. A similar annotation of the P450 genes from the complete *A. gambiae* genome published in 2002 (Holt *et al.*, 2002) gave 111 genes (Ranson *et al.*, 2002a) of which five are thought to represent pseudogenes. The need for "manual" annotation remains critical because of the error-prone gene- and transcript-calling programs, as noted for P450s by Gotoh (1998). These difficulties remain a major challenge to the "completion" of a genome program (Misra *et al.*, 2002). Potentially confusing is the proliferation of GenBank accessions for transcripts identified *in silico*, with only scant notice that they represent nothing but software-digested genomic sequences. Some of the sources of errors noted for the *Drosophila* annotation (Misra *et al.*, 2002) were plainly evident with the *A. gambiae* genome sequence as well. They typically include the fusion of two neighboring P450 genes into one, or the truncation of a gene. Comparison of the intron/exon structure of closely related genes and alignments with EST sequences when these are available can facilitate P450 gene annotation in a majority of cases. Independent evidence, such as the cloning of full-length cDNAs, or functional expression is rarely available to resolve annotation problems, so that the complete description of the P450 gene complement of any

species remains an ongoing task. A website presents information on the P450 sequences from completely sequenced genomes (*D. melanogaster, A. gambiae*) as well as links to other P450 websites. The complete sequence of an increasing number of insect genomes (*Drosophila pseudoobscura, Apis mellifera, Bombyx mori, Tribolium castaneum*) and the continuous addition of P450 sequences in EST projects from various species is redefining the approach to P450 research.

4.1.2.2. Genomic Organization: Clusters

The presence of P450 genes in clusters was revealed in early studies with *Drosophila* and the housefly (Frolov and Alatortsev, 1994; Cohen and Feyereisen, 1995). Further evidence was obtained by *in situ* hybridization to polytene chromosomes of *Drosophila* (Dunkov *et al*., 1996) and ultimately by the analysis of P450 genes in the complete genome sequences of *Drosophila* (Tijet *et al*., 2001) and *A. gambiae* (Ranson *et al*., 2002a, 2002b). The largest *Drosophila* P450 cluster carries nine genes (eight *CYP6A* genes and *CYP317A1*) at 51D on the right arm of chromosome 2. Six of these genes (*Cyp6a17* to *Cyp6a21*) are coordinately regulated during the circadian rhythm (Ueda *et al*., 2002). In *A. gambiae*, the largest cluster carries 14 P450 genes of the *CYP6* family at 30A on the right arm of chromosome 3. A large cluster of *CYP325* genes in *A. gambiae* contains 12 genes and 2 pseudogenes. In *A. gambiae*, only 22 of the 111 genes are present as singletons, with 16 clusters of 4 or more genes (Ranson *et al*., 2002a) (**Figure 2**). Gene clusters are thought to arise by sequential gene duplication events, the principal and initial mechanism of P450 diversification. In *Papilio glaucus, CYP6B4v2* and *CYP6B5v1* are clustered within 10 kb of each other (Hung *et al*., 1996). They are 99.3% identical at the nucleotide level, with 98% identity of their single 732-bp intron and 95% identity over 616 bp of the promoter region. The proteins they encode differ by just one amino acid. *CYP6B4* and *CYP6B5* are thus recently duplicated genes that have not yet diverged substantially in sequence (Hung *et al*., 1996). Similar close pairs are found in *Drosophila*, e.g., *Cyp12a4/Cyp12a5*; *Cyp9b1/Cyp9b2*; *Cyp28d1/Cyp28d2*. *Cyp12d1* and *Cyp12d2* are 2 kb apart and differ by only three nucleotides leading to three changes at the amino acid level. In *A. gambiae, CYP6AF1* and *CYP6AF2* are 99.8% identical at the nucleotide level and differ by just one amino acid (Ranson *et al*., 2002a). *CYP6D1* and *CYP6D3* are clustered on chromosome 1 in the housefly (Kasai and Scott, 2001a). Although they are only 50% identical in their 500 nt UTR and their product 80% identical at the amino acid level, both genes are phenobarbital-inducible and are constitutively overexpressed in the LPR strain (Kasai and Scott, 2001b).

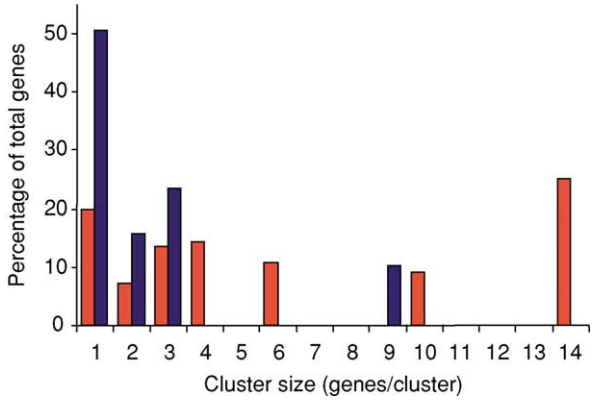


Figure 2 Clusters of P450 genes and singletons (cluster size = 1) genes in the *Drosophila melanogaster* (blue) and *Anopheles gambiae* (red) genomes. (Reproduced with permission from Ranson, H., Claudianos, C., Ortelli, F., Abgrall, C., Hemingway, J., *et al.*, **2002a**. Evolution of supergene families associated with insecticide resistance. *Science 298*, 179–181; © AAAS.)

In the housefly, a cluster of six *CYP6* genes within 24 kb of each other on chromosome 5 shows evidence of both gene duplications and chromosomal inversions (Cohen and Feyereisen, 1995). Three genes are transcribed in one direction and the next three are transcribed in the opposite direction. The six genes have a short intron at the same position in the ETLR conserved region (see Section 4.1.3.3.1) as many *CYP6* genes. The *CYP6A6* gene located at one extremity of the cluster is in fact represented only by the second exon. The first exon was not found in 2 kb of DNA upstream of this exon boundary, although the intron length of the five other genes is only 57–125 bp. *CYP6A6* may therefore be a pseudogene generated during a chromosomal inversion whose breakpoint was located in the intron. The other five genes of the cluster are transcribed, with *CYP6A5* being predominantly expressed in larvae (Cohen and Feyereisen, 1995).

Two recently duplicated genes may undergo gene conversion, if they do not diverge fast enough (Walsh, 1987). Gene conversion at the 5′ end of the *CYP6B8* and *CYP6B28* genes of *Helicoverpa zea* (Li *et al*., 2002b) has been suggested by the lower level of nt differences in the first half of the first exon, as compared to the rest of the sequence. Exon specific deficit in variation among pairs of P450 is indicative of a gene conversion event (Matsunaga *et al*., 1990). Changes in regulatory patterns (induction, tissue specific expression) can be observed for recently duplicated genes (Li *et al*., 2002c) and may

lead to their independent evolution. Extreme examples of gene conversion between duplicated P450 genes leading to their concerted evolution, of the kind seen with α-amylase genes (Hickey *et al.*, 1991), has not been reported to date.

Although events such as unequal crossing-over can lead to gene duplications, there are other mechanisms such as retrotransposition. Capture of a spliced mRNA by a retrotransposon can reintroduce an intronless sequence into the genome where it may evolve further or die as a pseudogene. This process generally occurs in germ cells, so it should be limited to genes that are expressed in those cells. The *Drosophila Cyp4g1* gene at the tip of the X chromosome (1B3) may be a case in point. It lacks introns and its closest paralog, *Cyp4g15*, has five introns and is located at 10C2-3. Furthermore, *Cyp4g1* is represented by the largest number of ESTs in *Drosophila*, and yet the apparent N-terminus of CYP4G1 is quite atypical of a microsomal P450 suggesting that an ancient 5′UTR is now part of the open reading frame. Further studies are needed to clarify the status of this gene.

There is little evidence for alternative splicing of insect P450 transcripts as additional means of generating diversity. The *Cyp4d1* gene seems to utilize two alternate first exons, and ESTs for each transcript type, CYP4D1 and CYP4D1*alt*, have been found. The first cDNA cloned (Gandhi *et al.*, 1992) uses a first exon (exon 1 prime) closest to the rest of the gene, whereas several ESTs use a more distal first exon (exon 1) instead. The two predicted proteins differ only from the N-terminal to the beginning of the first SRS (see Section 4.1.3.3.1 and **Figure 8**). The consequences of this alternative splicing in terms of catalytic competence thus remain to be examined.

4.1.2.3. Genomic Variation: Alleles, Pseudogenes

Apparent allelic variants of cloned P450 cDNAs and genes were already described in the earliest studies of insect P450s (Cohen *et al.*, 1992, 1994; Cohen and Feyereisen, 1995). Examples of this variation are most striking in *A. gambiae*, because the sequence released includes several "scaffolds," which are not included in the "golden path" used for genome assembly. These scaffolds represent large tracts of heterozygosity ("dual haplotype regions," Holt *et al.*, 2002), probably resulting from the mosaic nature (contributions from different *A. gambiae* cytotypes) of the strain that was sequenced. Interestingly, some of these scaffolds harbor P450 genes, even P450 gene clusters, and a comparison of the various haplotypes reveals not only considerable differences, but also variations in the total number of P450 genes. In other words, P450 gene duplication events have occured independently in different mosquito cytotypes. A gene recently converted to a pseudogene in one population or strain, as a result of a debilitating mutation or transposable element insertion (see **Chapter 4.12**) may still be active in another population. The total number of P450 genes in a species is therefore a relative number that is genotype dependent.

In *Blattella germanica*, three pseudogenes of *CYP4C21* and two pseudogenes of *CYP9E2* have been described (Wen *et al.*, 2001). These pseudogenes are characterized by deletions and/or the presence of several stop codons in the open reading frame. *CYP9E2P2*, for instance, has just 10 nucleotide differences from *CYP9E2*, but two base deletions render this gene nonfunctional. The processed pseudogene *CYP4W1P* in the cattle tick has a 191 bp deletion, but only three other nt changes in the open reading frame (Crampton *et al.*, 1999). Another tick pseudogene, *CYP319A1P*, contains two DNA insertions in the open reading frame and also appears to be of recent origin (He *et al.*, 2002).

The rate of gene duplication has been reported to be extremely high, making it an event almost as frequent as point mutations at the nucleotide level (Lynch and Conery, 2000). With a rate of 31 gene duplications per genome per million years in *Drosophila*, and a half life of less than 3 million years for a duplicated gene, one can estimate that there is a P450 gene duplication event on average every 5 million years. *Drosophila* is peculiar in its ability to delete genomic DNA at a high rate, and the proportion of pseudogenes in the genome is low – 1 pseudogene for 130 proteins as compared to 1 for 19 in *Caenorhabditis elegans* (Harrison *et al.*, 2003). However the proportion of pseudogenes in the P450 family is high (five pseudogenes) in *Drosophila*, another indication of rapid turnover of P450 genes.

4.1.2.4. P450 Gene Orthologs

With so much gene duplication, are there still any P450 orthologs in related species? (orthology = the "same" gene in different species that has only diverged as a result of speciation). The availability of two completely sequenced genomes made it possible to identify those members of the P450 family that were truly orthologous. This is not as easy as it seems, because in some borderline cases, two formally orthologous genes may be sufficiently related to their closest formal paralog in the same species as to make the evolutionary connection between the two pairs unclear. Nonetheless, it came as a surprise

to find a very small number of 1 : 1 pairs of orthologous P450 genes between *D. melanogaster* and *A. gambiae* (Ranson *et al.*, 2002a). Zdobnov *et al.* (2002) indicate a genome wide level of 44–47% orthologous genes, but for the P450 genes only ten orthologs were found, of which five are mitochondrial P450s (Ranson *et al.*, 2002a). When two P450 genes are recognized as orthologs even though the two species have diverged about 250 million years ago, the most likely explanation is that there is a strong evolutionary constraint that has maintained this orthologous relationship. A similar or identical physiological function for the orthologous pair of P450 enzymes may represent such a constraint. Three of the five mitochondrial P450s have indeed a recognized function that is predicted to be identical in *Drosophila* and *A. gambiae*: CYP302A1, CYP314A1, and CYP315A1 are hydroxylases of the ecdysteroid biosynthetic pathway (Warren *et al.*, 2002; Petryk *et al.*, 2003; and see Section 4.1.4.1.1). The number of pairs of orthologs between *Drosophila* and *A. gambiae* is however higher than the number of P450 enzymes thought to participate in ecdysteroid metabolism (see **Chapter 3.3**), so that several other insect-specific or dipteran-specific conserved functions are probably carried out by the remaining pairs of orthologs. There are several cases where one gene in one species has two "orthologs" in the other – this is the case when a gene duplication event occurred in just one of the two species after speciation and 250 million years of separate evolutionary history. These cases merit special attention because knowledge of the function of one of the P450s may quickly lead to understanding the function of its alter egos. The comparative analysis of the P450s of the two dipteran species showed that the deficit of true pairs of orthologs (10 versus the predicted ±40) was compensated by a most interesting alternative. Orthologous groups of P450 paralogs were seen, i.e., the phylogenetic analysis clearly identified several cases where one ancestral P450 gene underwent several duplication events in each of the two dipteran lineages. The *CYP6A* cluster on chromosome 5 of the housefly (Cohen and Feyereisen, 1995) and a cluster of *CYP6A* genes on the right arm of chromosome 2 in *Drosophila* are probably syntenic, in view of the synteny of these linkage groups (Weller and Foster, 1993), but the clusters have evolved separately for over 100 million years, and orthologous pairs of genes in each cluster can no longer be recognized.

Even when syntenic relationships are not maintained globally (Zdobnov *et al.*, 2002), cases of local microsynteny are observed. For example, *CYP302A1* and *CYP49A1* are very close on chromosome 2L in *A. gambiae*, but their orthologs in *Drosophila* are on 3L and 2R, respectively. Nonetheless, microsynteny is maintained around *CYP49A1*, which is close to the adrenodoxin reductase gene and is located in both species on the negative strand of the first intron of another gene, the G protein 0-α 47A (CG2204) in *Drosophila* and its ortholog in *A. gambiae*.

4.1.2.5. Intron/Exon Organization of *CYP* Genes

The intron/exon organization of P450 genes is a useful tool in the analysis of P450 phylogeny, as shown by the systematic studies in *C. elegans* (Gotoh, 1998), *Arabidopsis thaliana* (Paquette *et al.*, 2000), and *Drosophila* (Tijet *et al.*, 2001) (**Figure 3**). Intron sequence comparisons, as well as sequence comparisons of 5′ flanking sequences have also helped clarify the evolutionary relationships of very closely related *CYP6B* genes of *Papilio* species (Li *et al.*, 2002a). Multiple events of intron loss and gain can be deduced from a comparison of the intron/exon organization of orthologous pairs of P450 genes for *Drosophila* and *A. gambiae* (Ranson *et al.*, 2002a). Intron phase is nonbiased. *CYP* introns follow the GT/AG rule, except the first intron in the *Drosophila Cyp9c1* gene (Tijet *et al.*, 2001) and the first intron in the *Cyp6a8* gene (Maitra *et al.*, 2002). The latter was recognized by comparison with the full length cDNA, it does not conform to usual sequence patterns, and has an AT/TC splice junction. This intron is very short (36 bp) and potentially represents 12 additional in-frame codons (Maitra *et al.*, 2002).

4.1.3. The P450 Enzymatic Complexes

4.1.3.1. Classical Biochemical Approaches

4.1.3.1.1. Subcellular fractions The enzymology of insect P450 can be studied in different types of environments. These are enriched subcellular organelles from insects (microsomes, mitochondria) where multiple P450 interact *in situ* with their redox partners; membranes from cellular expression systems where a cloned recombinant P450 interacts with native or engineered redox partners; and ultimately a reconstituted system of purified recombinant P450 and its redox partners in a defined system devoid of biological membranes. Transgenic expression of P450 genes is another way to study P450 biochemistry, but also regulation.

The classical preparation of microsomal fractions from insect tissue homogenates by differential centrifugation has been described extensively by Hodgson (1985) and Wilkinson (1979). It remains, with minor modifications, the most widely used first step in the biochemical characterization of insect

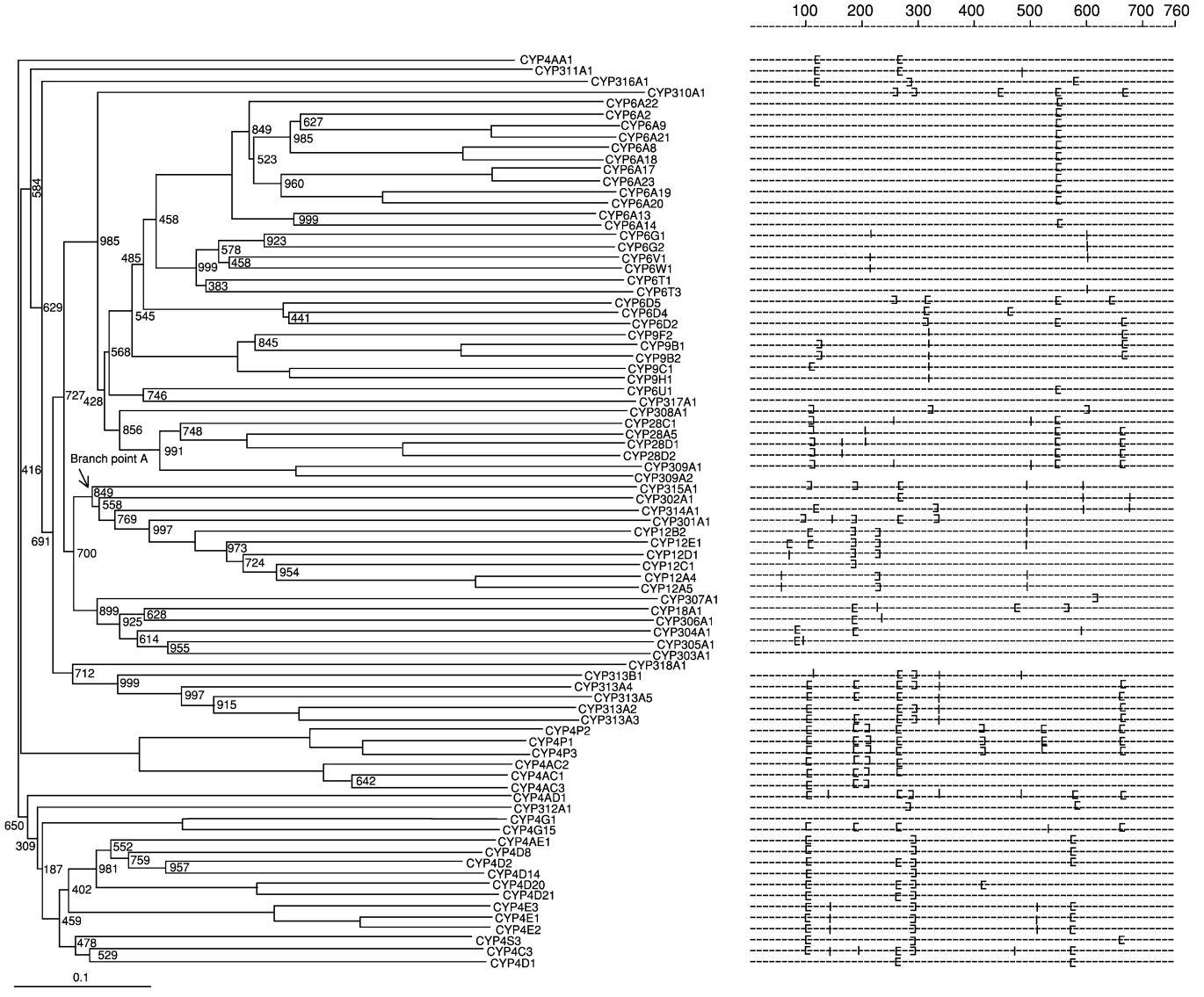


Figure 3 Neighbor-joining tree (with bootstrap value when different from 1000) and intron position of 83 P450 genes of *Drosophila melanogaster*. Branch point A indicates the mitochondrial P450 clade. Intron positions are shown schematically in the aligned open reading frames with their phase (| phase 0; [phase 1 and] phase 2). (Reprinted with permission from Tijet, N., Helvig, C., Feyereisen, R., **2001**. The cytochrome P450 gene superfamily in *Drosophila melanogaster*: annotation, intron-exon organization and phylogeny. *Gene 262*, 189–198. For updated information see [http://P450.antibes.inra.fr].)

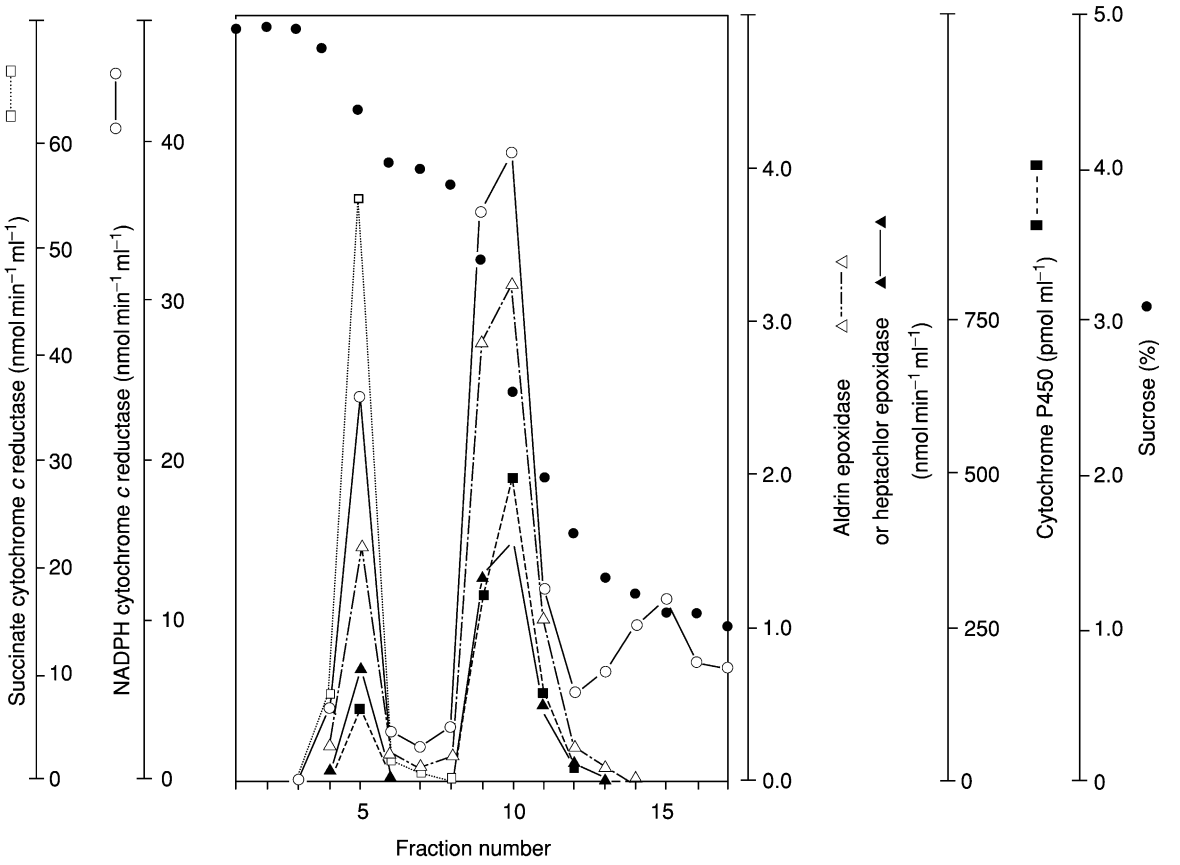


Figure 4 Sucrose density centrifugation separation of subcellular fractions of housefly larval homogenates showing the distribution of marker enzyme activities between mitochondrial and microsomal fractions. Note some P450 reductase activity in the top (soluble) fractions representing proteolytically cleaved enzyme. (Reprinted with permission from Feyereisen, R., **1983**. Polysubstrate monooxygenases (cytochrome P-450) in larvae of susceptible and resistant strains of house flies. *Pestic. Biochem. Physiol. 19*, 262–269; © Elsevier.)

P450 enzymes. Linear or step gradients of sucrose for the centrifugal preparation of microsomes and mitochondria, or $CaCl_2$ precipitation of microsomes have been less favored. In all approaches, the careful use of marker enzymes is critical. A technique for the rapid preparation of microsomal fractions of small tissue samples relying on centrifugation at very high speed on sucrose layers in a vertical rotor has been described (Feyereisen *et al.*, 1985).

The well-documented sedimentation of P450-associated activities at low *g* forces in many early insect studies (reviews: Wilkinson and Brattsten, 1972; Wilkinson, 1979) has been considered a peculiar difficulty of insect biochemistry. At the time, the vertebrate toxicology and endocrinology literature had clearly identified P450 metabolism of xenobiotics as microsomal, whereas mitochondrial P450s were chiefly involved in hormone metabolism. There was therefore little incentive in probing the subcellular distribution of insect P450 activity more carefully. The discovery of insect CYP12 enzymes and their characterization as mitochondrial P450 enzymes capable of metabolizing xenobiotics (Guzov *et al.*, 1998) has shed a new light on the early difficulties in sedimenting insect P450 activities in the "correct" fractions. It is quite probable that at least a part of the P450 activities observed in "mitochondrial" fractions were indeed carried out by CYP12 enzymes. In housefly larvae, 15–20% of the aldrin and heptachlor epoxidase activities were associated with mitochondrial fractions after sucrose density centrifugation (Feyereisen, 1983) (**Figure 4**).

The insect midgut is a particularly rich source of P450 activity (Hodgson, 1985), but the external brush border membrane is a significant source of membrane vesicles (BBMV) upon homogenization of the tissue. Centrifugal methods to separate the BBMV fraction from the microsomes derived from the endoplasmic reticulum have been described (Neal and Reuveni, 1992).

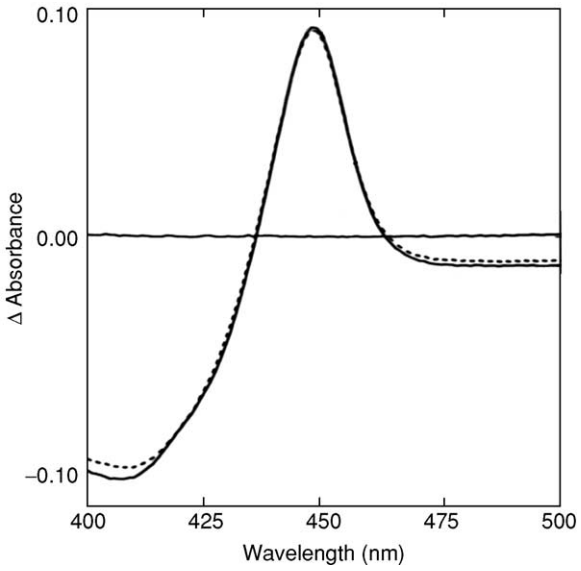


Figure 5 CO-difference spectrum of recombinant CYP12A1. The P450 was reduced with either sodium dithionite (solid line) or with bovine adrenodoxin, adrenodoxin reductase, and NADPH (dashed line).

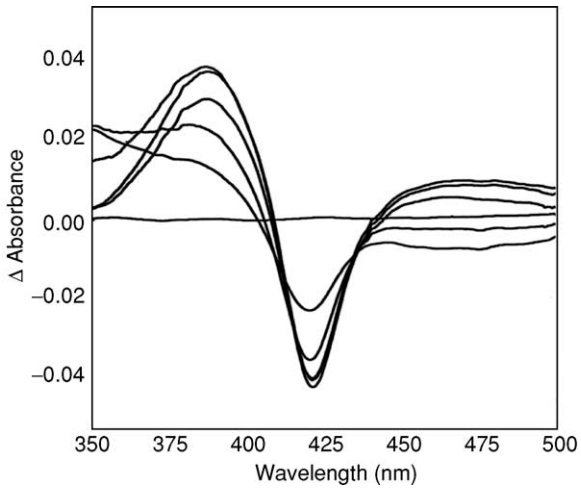


Figure 6 Type I substrate-induced difference spectrum of recombinant CYP12A1 with increasing concentrations of progesterone.

4.1.3.1.2. Spectral characterization and ligand binding The analysis of P450 levels in subcellular fractions follows the original procedure of Omura and Sato (1964). A difference spectrum between reduced microsomes and reduced microsomes after gentle bubbling of CO readily displays the famous redshifted Soret peak at 450 nm (**Figure 5**). The concentration of P450 can be calculated from the ΔOD between 490 and 450 nm and Omura and Sato's extinction coefficient $\varepsilon = 91\,M^{-1}\,cm^{-1}$. This measure gives the total concentration of all forms of P450 present in the preparation. Individual P450 proteins may have peaks that are 1 or 2 nm off the 450 nm norm, and when they represent a large portion of the total P450, the total P450 peak may be shifted as a consequence. The degradation of P450 to the inactive P420 form may interfere with the measurement of P450, as already reviewed by Hodgson (1985), and the respiratory chain pigments interfere with the measurement of P450 in mitochondrial fractions. The classical Omura and Sato procedure remains the procedure of choice to measure the amount and purity of P450 proteins produced in heterologous systems (see below).

Ligand-induced spectral changes also follow classical procedures detailed in a useful review (Jefcoate, 1978). Type I spectra (peak at 380–390 nm, trough at 415–425 nm) result from ligand in the substrate binding site displacing water as sixth ligand to the heme iron (see Section 4.1.3.4.1, **Figure 6**). Type I spectra are concentration dependent, giving a spectral dissociation constant (Ks) and this titration is correlated with a shift of the iron from low spin to high spin. Not all type I ligands are substrates, and not all substrates are type I ligands, so this useful tool must be used with caution. A type I spectral Ks is not an enzymatic Kd.

Type II spectra (peak at 425–435 nm, trough at 390–405 nm) result from the binding of a strong ligand to the heme iron, typically the nitrogen coordination of compounds such as pyrimidines, azoles, or n-octylamine. Type II spectral titration is correlated with a shift from high spin to low spin and is a hallmark of strong inhibitors such as imidazoles (**Figure 6**). Other, less frequently studied spectral changes induced by ligands or their metabolism will not be discussed here (e.g., type III spectra, Hodgson, 1985; spectra of phenyl–iron complexes, Andersen *et al.*, 1997).

4.1.3.1.3. Assays and substrates Measurement of P450 activity is a special challenge because of the large number of different P450 enzymes, each catalyzing the metabolism of a specific (broad or narrow) range of substrates. There is therefore a very large number of assays for P450 activity. Direct assays of product appearance or substrate disappearance rely on all the tools of analytical chemistry. Indirect assays (e.g., activity of a P450 product in an enzyme or bioassay) can be useful but the strength of the claim for P450 activity depends on the purpose of the assay. The assay of a P450 produced in a heterologous system can be straightforward, but the assay of a P450 in its native microsomal or mitochondrial membrane, where it is mixed with an undetermined number and amount of other P450s, is more problematic. Metabolism of compound *M*

to product *N* in microsomes is the sum of the contributions of all P450 enzymes that catalyze the *M* to *N* reaction (and sometimes non-P450 enzymes can catalyze the same reaction!). Selective inhibitors (chemicals or antibodies) of one P450 can, by substraction, indicate the relative contribution of that particular P450 to the reaction being measured (e.g., Wheelock and Scott, 1992a; Hatano and Scott, 1993; Korytko *et al*., 2000b). This indirect inference is only as valid as the inhibitor is selective. Substrates that are selective for one P450 and that are easily assayed have been the object of considerable research in biomedical toxicology. The relative success of this quest (e.g., nifedipine as model probe for CYP3A4) is a result of both the limited number of major P450 expressed in human liver and the heavy investment in their study. The large number of insect P450s times the large number of insect species under study divided by the investment in their research makes a similar quest seem quixotic. By default then, but mostly by inertia of the historical development of insect P450 research, a certain number of assays have taken their place in the literature as some measure of "global" P450 activity. Most authors are now fully aware that the microsomal activity of, e.g., aldrin epoxidation, *p*-nitroanisole *O*-demethylation or 7-ethoxycoumarin *O*-deethylation is only a measure of those P450 enzymes catalyzing these reactions. But this awareness is only as recent as our understanding that there are really many P450 enzymes, and that their individual catalytic competence may be broad or narrow, overlapping with other P450 enzymes or not. Thus, the pioneers who used aldrin epoxidation as an assay did so at a time when P450 research was still strongly influenced by the dichotomy between steroid metabolism by specific P450 enzymes and drug metabolism by two major forms of liver P450. They used an analytical tool then readily available in pesticide toxicology laboratories (GC with electron capture detection for the sensitive detection of organochlorine pesticide residues), but they didn't use it without a caveat. Quoting the classical Krieger *et al*. (1971) study: "to the extent that the rate of epoxidation of aldrin to dieldrin typifies the activity of the enzymes toward a wider range of substrates…" The current and still widespread use of a "model" substrate to explore P450 activities in insect subcellular fractions can be useful. If the metabolism of a randomly chosen P450 substrate (e.g., aldrin, aminopyrine, 7-methoxyresorufin) is quantitatively different between insect strains, in different tissues, following induction, etc., then this substrate has provided a clue that the qualitative or quantitative complement of P450 enzymes is changing. It is

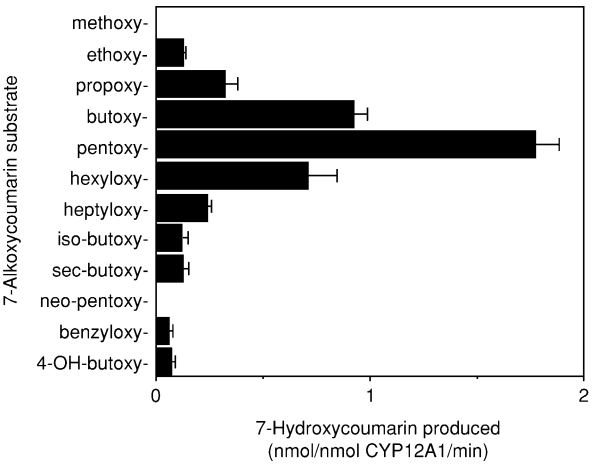


Figure 7 *O*-Dealkylation activity of *E. coli*-expressed recombinant housefly CYP12A1 in a reconstituted system. (Reprinted with permission from Guzov, V.M., Unnithan, G.C., Chernogolov, A.A., Feyereisen, R., **1998**. CYP12A1, a mitochondrial cytochrome P450 from the house fly. *Arch. Biochem. Biophys. 359*, 231–240; © Elsevier.)

up to the investigator to follow up on this clue. Alkoxycoumarins and alkoxyresorufins are useful substrates for sensitive fluorometric assays and have largely replaced organochlorines as model substrates. They can be used to "map" the catalytic competence of a heterologously expressed P450 (e.g., the preference of CYP12A1 for pentoxycoumarin, Guzov *et al*., 1998) (**Figure 7**). Steroids such as testosterone have multiple sites of attack by P450 enzymes and can likewise be used to characterize the activity of subcellular fractions or of heterologously expressed P450s (Amichot *et al*., 1998; Cuany *et al*., 1990; M.B. Murataliev, V.M. Guzov, R. Feyereisen, unpublished data).

Another approach to the study of P450 activity is to follow the consumption of the other substrates, O_2 or NADPH. This approach is certainly valid when the stoichiometry of the reaction (see eqn [1]) is under study, but its use to monitor metabolism by subcellular fractions is fraught with difficulties. Other enzymes consume O_2 and NADPH as well, and the background activity can be very high.

4.1.3.1.4. P450 assays in individual insects The design of assays suitable for assessing P450 activities in single insects has accompanied the need to study variations in P450 activities from the individual to the population level. Such assays allow the presentation of frequency histograms of activity levels in field-collected samples or in laboratory-selected populations and are ideally adapted to microtiter plate format. The NADPH-dependent conversion of *p*-nitroanisole to *p*-nitrophenol was followed in individual homogenates of *H. virescens* and

Pseudoplusia includens larvae (Kirby *et al*., 1994; Rose *et al*., 1995; Thomas *et al*., 1996). This assay has a relatively low sensitivity, but clearly distinguished individuals from susceptible and insecticide-resistant strains. Cut abdomens of adult *Drosophila* in buffer containing 7-ethoxycoumarin can be used to measure 7-hydroxycoumarin formation in a 96-well microtiter plate format (de Sousa *et al*., 1995). This rapid technique allows for instance the monitoring of individual variability over the course of a selection regime (Bride *et al*., 1997). In another example, 10 000 g supernatants of individual homogenates of *Chironomus riparius* larvae were assayed for 7-ethoxyresorufin *O*-deethylation activity (Fisher *et al*., 2003). The classical aldrin epoxidase assay adapted on single larvae of *Spodoptera frugiperda* has also been reported, but this assay is not adapted to the 96-well format (Yu, 1991, 1992).

A simple assay based on the peroxidase activity of the heme group with tetramethylbenzidine was developed for use in single mosquitoes (Brogdon, 1997). This assay, easily developed on a microtiter plate format, is an indirect assay measuring total heme content of the insect homogenate rather than P450 activity, and therefore needs to be carefully validated.

4.1.3.1.5. Solubilization and purification Solubilization and purification of insect P450 has generally followed the advances pioneered in the purification of vertebrate P450 (Agosin, 1985; Hodgson, 1985) and relatively a few studies since 1985 have pursued this difficult task (Ronis *et al*., 1988; Wheelock and Scott, 1989). P450 purification from microsomes of mixed tissues (e.g., fly abdomens) can be sufficient to obtain a protein fraction suitable for antibody generation or peptide sequencing (Feyereisen *et al*., 1989; Wheelock and Scott, 1990; Scott and Lee, 1993b).

Sequential chromatography on octylamino agarose, DEAE-cellulose, and hydroxyapatite was used to purify sodium cholate-solubilized P450s from *Drosophila* (Sundseth *et al*., 1990). The two protein fractions obtained, P450 A and B, had only a very low 7-ethoxycoumarin *O*-deethylase activity (0.01 nmol/nmol P450/min), but the proteins were useful in generating monoclonal antibodies that allowed the subsequent cloning of CYP6A2 (Waters *et al*., 1992).

An original approach was the purification of a locust P450 by affinity chromatography with type II and type I ligands (Winter *et al*., 2001) that led to cloning of CYP6H1 (Winter *et al*., 1999). In this approach, microsomes from larval *Locusta migratoria* Malpighian tubules were first treated with the detergent synperonic NP10 to solubilize P450. The extract was then chromatographed on ω-octylamino agarose then hydroxylapatite. The third and less classical step was chromatography on a triazole agarose affinity column. The affinity ligand was a derivative of the fungicide difenoconazole, which has an affinity for ecdysone 20-monoxygenase of the same level as that of the substrate ecdysone (0.5 versus 0.2 μM). This substituted triazole is a typical type II ligand (active site liganding of the heme) and its use on the affinity matrix led to the adsorption of all the P450 loaded on the column (Winter *et al*., 2001). Elution from the affinity column was done by replacing the immobilized type II ligand with a soluble type I ligand, ecdysone (see **Chapter 3.3**). A major protein band of 60 kDa was thus obtained in 4% yield, with a P450 specific activity of 13.1 nmol/mg protein. Unfortunately, biochemical evidence that this P450 is in fact an ecdysone 20-monooxygenase was not obtained in this study (Winter *et al*., 2001), so the nature of CYP6H1 (Winter *et al*., 1999) remains conjectural.

Another variant on the classical purification schemes has been the use of immobilized artificial membrane high performance liquid chromatography (IAM-HPLC) of microsomal proteins (Scharf *et al*., 1998). This technique allowed the 70-fold purification of a P450 from the German cockroach and the subsequent production of antibodies with this 49 kDa protein as antigen.

4.1.3.2. Heterologous Expression Systems

Biochemical characterization of a P450 protein and its substrate selectivity remains a *sine qua non* condition of its functional identification. Only a few P450 enzymes are characterized well enough (e.g., steroid metabolizing P450s in vertebrates) that sequence comparison can reasonably predict activity. For most other P450s, the sequence does not provide a clue to the activity, and there are now innumerable papers describing how one or a few mutations can change substrate selectivity (review: in Domanski and Halpert, 2001). Only three mutations are needed to confer to *Drosophila* CYP6A2 the ability to metabolize DDT (Bergé *et al*., 1998; Amichot *et al*., 2004). In the absence of significant studies on the activity of P450 proteins purified directly from insect tissues, it is the expression of P450 cDNAs in heterologous systems that has become the standard way of characterizing insect P450 proteins. A number of such expression systems have been developed over the last few years (**Table 1**), and the techniques are essentially similar to those used for the production of P450 proteins from mammalian or plant tissues.

Table 1 Heterologous expression systems for insect P450

Expression system	*P450 produced*	*Substrate metabolized*	*Reference*
Escherichia coli	CYP4C7	Sesquiterpenoids	Sutherland *et al.* (1998)
	CYP6A1	Aldrin, heptachlor	Andersen *et al.* (1994)
		Sesquiterpenoids	Andersen *et al.* (1997)
		Diazinon	Sabourault *et al.* (2001)
		7-Propoxycoumarin	[a]
		1-Bromochlordene, chlordene, 1-hydroxychlordene, isodrin	[b]
		Testosterone, progesterone, androstenedione	[c]
		Pisatin	[d]
		Chlorfenapyr	[e]
	CYP6A2	DDT, testosterone	Amichot *et al.* (2004)
	CYP6A5	Benzphetamine, *p*-chloro-*N*-methylanilin, methoxyresorufin	[f]
	CYP9E1	Sesquiterpenoids	[g]
	CYP12A1	Aldrin, heptachlor, diazinon, azinphosmethyl, amitraz, progesterone, testosterone, 7-alkoxycoumarins	Guzov *et al.* (1998)
	CYP15A1	*t t*, methyl farnesoate	Helvig *et al.* (2004)
Baculovirus	CYP6A2	Aldrin, heptachlor, diazinon	Dunkov *et al.* (1997)
	CYP6B1	Furanocoumarins	Ma *et al.* (1994), Hung *et al.* (1997), Wen *et al.* (2003)
	CYP6B4	Furanocoumarins, ethoxycoumarin	Hung *et al.* (1997), Li *et al.* (2003)
	CYP6B17	Furanocoumarins, ethoxycoumarin	Li *et al.* (2003)
	CYP6B21	Furanocoumarins, ethoxycoumarin	Li *et al.* (2003)
	CYP6B25	Furanocoumarins	Li *et al.* (2003)
Yeast	CYP6A2	AflatoxinB1, 7,12-dimethylbenz[a]anthracene, 3-amino-1-methyl-5H-pyrido[4,3-b]-indole	Saner *et al.* (1996)
	CYP6D1	Methoxyresorufin	Smith and Scott (1997)
Transfected S2 cells	CYP302A1	2,22-Dideoxyecdysone	Warren *et al.* (2002)
	CYP314A1	Ecdysone	Petryk *et al.* (2003)
	CYP315A1	2-Deoxyecdysone, 2,22-dideoxyecdysone	Warren *et al.* (2002)

[a]V.M. Guzov and R. Feyereisen, unpublished data.
[b]J. Walding, J.F. Andersen, and R. Feyereisen, unpublished data.
[c]M.B. Murataliev, V.M. Guzov, and R. Feyereisen, unpublished data.
[d]V.M. Guzov, H. VanEtten, and R. Feyereisen, unpublished data.
[e]V.M. Guzov, M. Kao, B.C. Black, and R. Feyereisen, unpublished data.
[f]J.L. Stevens, J.F. Andersen, and R. Feyereisen, unpublished data.
[g]J.F. Andersen and R. Feyereisen, unpublished data.

4.1.3.2.1. ***Escherichia coli*** Bacterial production (*Escherichia coli*) of insect P450s has required several modifications of the sequence. At the 5′ end of the cDNA, mutations are introduced to optimize expression (Barnes *et al.*, 1991). The second codon is replaced by Ala (Andersen *et al.*, 1994; Guzov *et al.*, 1998; Sutherland *et al.*, 1998) and silent substitutions are introduced to increase the A/T content (Sutherland *et al.*, 1998). In some cases, introduction of 4–6 His codons just before the stop codon directs the translational production of a C-terminal "histidine tag" (Guzov *et al.*, 1998; Sutherland *et al.*, 1998). P450 production can be enhanced by the addition of δ-aminolevulinic acid (a precursor for heme biosynthesis) to the culture broth (Sutherland *et al.*, 1998). The P450 produced in bacteria is found mostly in a membrane fraction, and sometimes in a fraction of inclusion bodies that are difficult to extract. In some cases (Andersen *et al.*, 1994), a significant amount of P450 is produced as a soluble form. The *E. coli* membrane fraction carrying the recombinant P450 protein is generally suitable for analysis by difference spectroscopy for either P450 content by the Omura and Sato (1964) procedure, or for ligand binding (type I binding for potential substrates or type II binding for azoles). Although some P450 proteins (e.g., CYP17) expressed in *E. coli* can utilize an endogenous flavodoxin

reductase/flavodoxin system for catalysis, none of the insect P450 proteins tested thus far have been catalytically active in *E. coli* membrane fractions in the absence of a P450 reductase. Therefore, P450 produced in bacteria needs to be solubilized and purified by classical methods. The proteins produced with a histidine tag, once solubilized, are purified by nickel chelate affinity chromatography. Extensive dialysis is needed in both procedures to remove excess detergent or imidazole used for elution from the nickel affinity column. The P450 obtained is then suitable for reconstitution with redox partners. These partners (microsomal or mitochondrial redox partners, see Section 4.1.3.3) are themselves produced in *E. coli* and purified (Guzov *et al.*, 1998).

Reconstitution of a catalytically active enzyme system is then tedious or artistic, depending on one's degree of patience. It requires attention to the details of concentrations of phospholipids, detergents, proteins, and their order of addition, mixing and dilution (Sutherland *et al.*, 1998). The advantages of bacterial expression are the low cost of production of large amounts of P450, and the possibility to work with a precisely defined *in vitro* system with highly purified enzymes and their partners. A thorough characterization of the enzyme can be undertaken. The disadvantage of this formal biochemical approach is that purification and reconstitution are difficult and time-consuming, and is probably not suitable for when the goal is simply a survey of the catalytic competence of the P450, or the comparison of a large number of P450s or P450 mutants. The host organism, *E. coli*, is a rare organism devoid of P450 genes of its own while other bacteria can carry over 20.

4.1.3.2.2. Baculovirus Expression of P450 in lepidopteran cells by the baculovirus system requires no modification of sequence and is a widely used method for the production of proteins in an eukaryotic system. It has the potential of producing large amounts of P450 proteins for subsequent purification, but studies with insect P450 expressed with this system (Ma *et al.*, 1994; Dunkov *et al.*, 1997; Hung *et al.*, 1997; Chen *et al.*, 2002; Wen *et al.*, 2003) have relied instead on the advantage that the protein is present in a suitable milieu, the endoplasmic reticulum of an insect cell. Thus, cell lysates, briefly centrifuged to pellet cell debris, are used as enzyme source. Difference spectroscopy or immunological methods (Dunkov *et al.*, 1997) can be used to assess the amount of P450 produced. The host cells provide their endogenous P450 reductase to support the activity of the heterologous P450 when the cell lysates are incubated with an NADPH regenerating system. Although the level of P450 reductase is sufficient to allow the measurement of a number of P450-dependent activities (Dunkov *et al.*, 1997), the stoichiometry of endogenous P450 reductase, and cytochrome b_5, to heterologously expressed P450 is probably not optimal. The activities measured do not represent the full potential of the P450 under study. For instance, a thirty-fold increase in CYP6A2 activity was observed when purified housefly P450 reductase and cytochrome b_5 were added to lysates of cells expressing *Cyp6a2* (Dunkov *et al.*, 1997). An improvement of the baculovirus expression system has therefore been designed, wherein the cells are coinfected with a virus engineered to carry the P450 and a virus engineered to carry a P450 reductase (housefly P450 reductase, Wen *et al.*, 2003). Optimal conditions were sought, and a significant increase (33-fold) in CYP6B1 activity towards the substrate xanthotoxin was achieved with the improved P450 reductase/P450 ratio. In fact, the improved conditions allowed the measurement of angelicin metabolism that was barely detectable in the absence of additional P450 reductase. Thus, in the baculovirus system, insect P450s can be studied in an insect membrane environment, without need for purification. Those are great advantages over the *E. coli* expression system. However, the interactions with its redox partners are not manipulated as easily (Wen *et al.*, 2003). The total amounts of P450 produced are also smaller, although addition of hemin to the culture medium can increase the amount of P450 produced (Dunkov *et al.*, 1997; Wen *et al.*, 2003). The total amount of P450 produced is less important in the baculovirus system than in the *E. coli* system as purification is not required for most applications, and as the highest activity of cell lysates is achieved at the correct P450/P450 reductase ratio, not at the maximal P450 production level (Wen *et al.*, 2003). The level of endogenous P450 in the control experiments, i.e., uninfected cells or cells infected with a virus carrying a non-P450 "control" cDNA, are virtually undetectable.

4.1.3.2.3. Transfection in cell lines Heterologous expression in transfected mammalian COS cells was first established in 1986 for bovine CYP17 (Zuber *et al.*, 1986), but it is not until later that an insect P450 was similarly expressed in an insect cell line. Thus, *Drosophila* Schneider 2 cells have been transfected with *Drosophila* P450 cDNAs (Warren *et al.*, 2002). Expression under control of the actin 5C promoter produced sufficient P450 for activity measurements. The advantage of the method is its simplicity. When the expression of the P450 is coupled

with a very sensitive assay, the method can rapidly provide qualitative data on the catalytic competence of the enzyme. However, the quantitative determination of P450 levels is more difficult to achieve, and the interaction with redox partners cannot be optimized except by coinfection. It is interesting that the CYP302A1 and CYP315A1 expressed by this method are mitochondrial P450s and the S2 cell homogenates were able to provide adequate redox partners. As used so far, it has not allowed a measurement of the amounts of P450 produced, nor have the redox partners been characterized or optimized. Cell transfection does not have the potential of the baculovirus system for large-scale production of P450 proteins.

4.1.3.2.4. Yeast Yeast expression systems have only started to be exploited for the production of insect P450 proteins. *Saccharomyces cerevisiae* has three P450 genes that are fully characterized, and are expressed at low levels so that inducible expression of an exogenous P450 is not hindered by the endogenous P450. Coproduction of CYP6A2 from *Drosophila* and of human P450 reductase in yeast (Saner *et al*., 1996) generated a cell system capable of activating several procarcinogens to active metabolites that induced mitotic gene conversion or cytotoxicity. Housefly CYP6D1 was also produced in yeast but methoxyresorufin demethylation was the only marker activity obtained with microsomes of the transformed yeast (Smith and Scott, 1997). Insect P450 production in yeast has not yet achieved the success seen with plant P450 production in yeast (Schuler and Werck-Reichhart, 2003). P450 cDNAs may need to be engineered to recode the N-terminus of the protein. This has been done successfully with plant P450s to conform with the yeast codon usage (Hehn *et al*., 2002). The replacement of the yeast P450 reductase gene with an insect P450 reductase gene by homologous recombination (Pompon *et al*., 1996) should increase the usefulness of this yet underutilized expression system. Indeed, yeast combines the advantages of *E. coli* inducible production of large amounts of protein with the advantage of the eukaryotic cell system in which P450 enzymes can be studied in a normal membrane environment.

4.1.3.2.5. Transgenic insects The use of transgenic insects to study P450 function (or regulation, see Sections 4.1.4.5.3 and 4.1.5.2.2) has until now been restricted to *Drosophila*. In the first report, Gandhi *et al*. were unable to rescue by transgenesis the lethality of two complementation groups in the *Cyp4d1* region (Gandhi *et al*., 1992). Heterologous expression of vertebrate P450s was achieved in studies aimed at developing *Drosophila* as a genotoxicity model organism. The rat *CYP2B1* gene was expressed under control of the *Drosophila* LSP1a promoter (Jowett *et al*., 1991). This promoter ensures high levels of expression in third instar larvae. Transgenic flies expressed functional CYP2B1, as shown by increased CYP2B1-specific metabolism of 7-benzyloxyresorufin and by increased sensitivity to cyclophosphamide, a procarcinogenic drug activated by CYP2B1. In similar experiments, canine *CYP1A1* was expressed under the control of the *Drosophila* heatshock inducible *hsp70* promoter. Small amounts of CYP1A1 were produced after heatshock, sufficient to increase the sensitivity of the flies to 7,12-dimethylbenz[a]anthracene, a polycyclic aromatic hydrocarbon that is metabolized by CYP1A1 to a genotoxic metabolite (Komori *et al*., 1993). Housefly CYP6D1 was produced in *Drosophila* under control of the heatshock inducible *hsp70* promoter, and this led to a significant increase in benzo[a]pyrene hydroxylation (Korytko *et al*., 2000a), though heat shock decreased total P450 levels in whole body microsomes. Transgenic expression of CYP6G1 has been an important piece of evidence in demonstrating its role in DDT and neonicotinoid resistance (Daborn *et al*., 2002; Le Goff *et al*., 2003) as discussed below (Section 4.1.4.5.5). Transformation of other insects (see **Chapter 4.13**), notably with the piggyBac vector as in *Bombyx mori* (Tamura *et al*., 2000) will undoubtedly increase the applications of transgenesis to P450 research.

4.1.3.3. P450 Enzymes and Their Redox Partners

4.1.3.3.1. P450 proteins The sequence identity of distantly related P450 proteins can be as low as that predicted from the random assortment of two sets of 500 or so amino acids. This is because there are very few absolutely conserved amino acids. In insect sequences available to date, these are found in five conserved motifs of the protein (**Figure 8**), the WxxxR motif, the GxE/D **T**T/S motif, the **E**x**LR** motif, the PxxFx**PE**/**DR**F motif and the PFxxGx**R**x**C**xG/A motif. Despite this tremendous overall sequence diversity, the increasing number of crystal structures for P450 proteins, mostly soluble forms from bacteria (Poulos *et al*., 1995), reveals a quite high conservation of the three-dimensional structure. The description of this structure essentially follows the nomenclature of the P450cam protein, the camphor hydroxylase of *Pseudomonas putida* (Poulos *et al*., 1985). The first motif WxxxR is located in the C-helix, and the Arg is thought to form a charge pair with the propionate of the heme. This

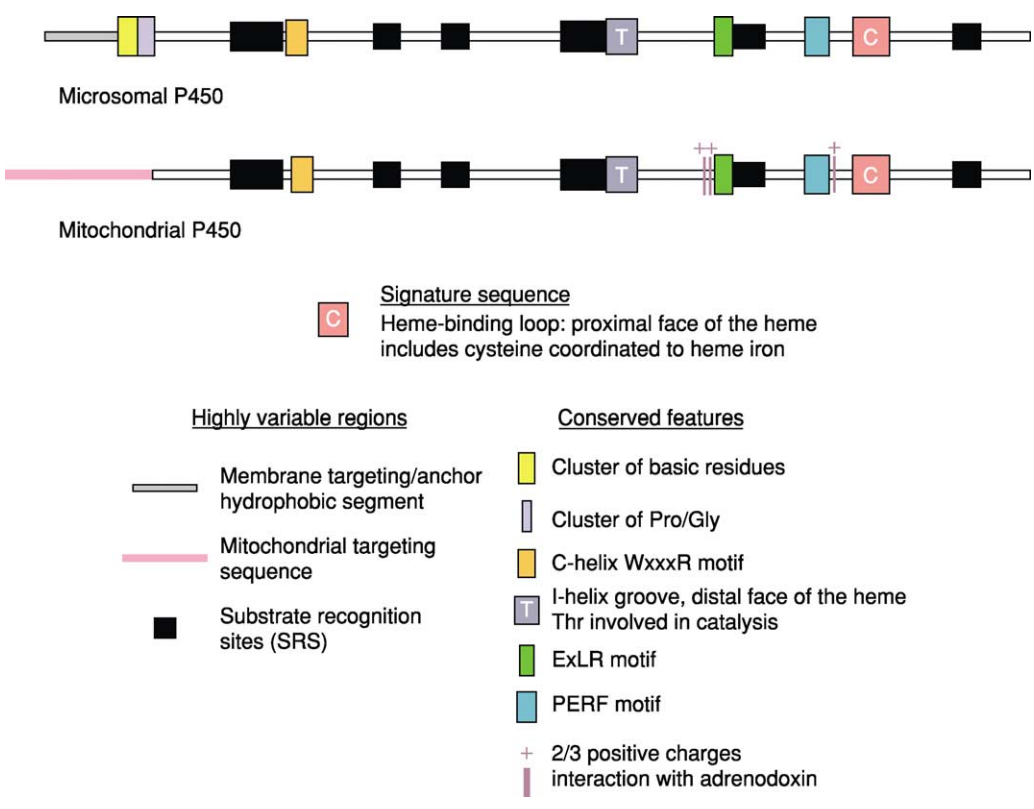


Figure 8 Conserved and variable regions of P450 proteins illustrated over their primary structure (sequence). (Adapted with permission from Werck-Reichhart, D., Feyereisen, R., **2000**. Cytochromes P450: a success story. *Genome Biol. 1*, 3003.1–3003.9; © GenomeBiology.)

motif is not easily discernible, except in multiple alignments. The second conserved motif GxE/DTT/S surrounds a conserved threonine in the middle of the long helix I that runs on top of the plane of the heme, over pyrrole ring B. The third conserved motif ExLR is located in helix K. It is thought to stabilize the overall structure through a set of salt bridge interactions (E-R-R) with the fourth conserved motif PxxFxPE/DRF (often PERF, but R is sometimes replaced by H or N) that is located after the K′ helix in the "meander" facing the ExLR motif (Hasemann *et al*., 1995). The fifth conserved motif PFxxGxRxCxG/A precedes helix L and carries the cysteine (thiolate) ligand to the heme iron on the opposite side of helix I. The cysteine ligand is responsible for the typical 450 nm (hence P450) absorption of the Fe^{II}–CO complex of P450 (Mansuy and Renaud, 1995). This heme binding loop is the most conserved portion of the protein, often considered as "signature" for P450 proteins. Deviations from the consensus sequences of these five motifs deserve special attention. For instance, the CYP301A1 of both *Drosophila* and *A. gambiae* has a very unusual Tyr instead of Phe in the canonical PFxxGxRxCxG/A motif around the Cys axial ligand to the heme. These deviations may denote an atypical catalytic function for the P450 enzyme, as seen in P450 enzymes that are not monooxygenases, such as plant allene oxide synthase (CYP74A) or vertebrate thromboxane synthase (CYP5A1) whose I helix lacks the conserved Thr. In the bacterial hydroxylase P450eryF (CYP107A), the Thr is replaced by Ala. A water molecule and a hydroxyl group of the substrate have become functional equivalents of the Thr hydroxyl (Poulos *et al*., 1995).

P450 proteins are also characterized by their N-terminal sequence (**Figure 9**). Those targeted to the endoplasmic reticulum have a stretch of about 20 hydrophobic amino acids. These precede one or two charged residues that serve as halt-transfer signal and a short motif of prolines and glycines. The latter serves as a "hinge" that slaps the globular domain of the protein onto the surface of the membrane while the N-terminus is anchored through it. The presence of the PGPP hinge is necessary for proper heme incorporation and assembly of functional P450s in the cell (Yamazaki *et al*., 1993; Chen *et al*., 1998). A hydrophobic region between helices F and G is thought to penetrate the lipid bilayer, thus increasing the contact of the P450 with the hydrophobic environment from which many substrates can enter the active site (Williams *et al*., 2000).

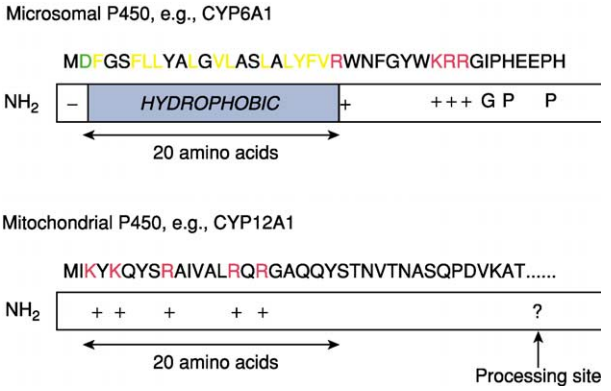


Figure 9 Scheme of the N-terminal sequence of microsomal and mitochondrial P450 proteins.

The N-terminal sequence of P450 proteins targeted to mitochondria is usually somewhat longer, and shows several charged residues (**Figure 9**). The mature mitochondrial protein is proteolytically cleaved at a position that has not been formally recognized for insect mitochondrial P450s to date, but is known for several mitochondrial P450s of vertebrate species. Mitochondrial P450 proteins are also characterized by a pair of charged amino acids in the K helix, R391 and K395 in CYP12A1. Homologous amino acids in mammalian P450scc (K377 and K381) are responsible for the high affinity to the ferrodoxin-type (adrenodoxin) electron donor (Wada and Waterman, 1992; Pikuleva *et al.*, 1999). An additional positively charged residue (R454 of CYP12A1) is homologous to R418 of CYP27A1 shown to increase affinity to adrenodoxin even further (Pikuleva *et al.*, 1999). The insect CYP12, 49, 301, 302, 314, and 315 proteins are most closely related to the mammalian mitochondrial P450 (CYP11, CYP24, and CYP27 families), to CYP44A1 from *C. elegans*, and to the pond snail CYP10. Subcellular localization of CYP12A1 by immunogold histochemistry with antibodies raised against the CYP12A1 protein produced in bacteria established the mitochondrial nature of CYP12A1 (Guzov *et al.*, 1998).

There is evidence that some vertebrate microsomal P450s, e.g., CYP1A1 and CYP2E1, are cleaved *in vivo* of their N-terminal anchor, thus revealing a cryptic mitochondrial targeting sequence, and the shortened protein is enzymatically active in mitochondria (Anandatheerthavarada *et al.*, 1999). Thus, the 1–44 residues of CYP1A1 serve a dual targeting role, with 1–32 targeting the protein to the ER cotranslationally, whereas cleavage and exposure of three basic amino acids in residues 33–44 direct the posttranslational transport to mitochondria (Bhagwat *et al.*, 1999). Whether some insect microsomal P450 proteins (e.g., CYP314A1, see Section 4.1.4.1.1) behave in this fashion is currently unknown.

Interspersed throughout the globular domain of the P450 proteins are six regions with a low degree of sequence similarity, covering about 16% of the total length of the protein (**Figure 8**). Initially recognized in CYP2 proteins by Gotoh (1992), these are called SRS (substrate recognition sites) and this designation has been generically extended to other P450s.

P450 enzymes, whether microsomal or mitochondrial, need to interact with redox partners for their supply of reducing equivalents from NADPH. **Figure 10** schematically illustrates the two types of electron transfer complexes thus formed, and the following sections provide a description of the redox partners.

4.1.3.3.2. NADPH cytochrome P450 reductase

P450 reductase (EC 1.6.2.4) belongs to a family of flavoproteins utilizing both FAD and FMN as cofactors. These diflavin reductases emerged from the ancestral fusion of a gene coding for a ferredoxin reductase with its NADP(H) and FAD binding domains with a gene coding for a flavodoxin with its FMN domain. This origin of the enzyme was first proposed by Porter and Kasper (1986) based on their analysis of the rat P450 reductase sequence. The fusion is dramatically illustrated by the three-dimensional structure of P450 reductase (Wang *et al.*, 1997) where the domains are clearly distinguished (**Figure 11**). The architecture of this domain fusion has been found in a handful of other enzymes (Murataliev *et al.*, 2004a). In some cases, further fusion with a P450 gene has led to self-sufficient P450 proteins, e.g., the fatty acid hydroxylase of *Bacillus megaterium*, P450BM3 (Nahri and Fulco, 1986) and of *Fusarium oxysporum*, P450foxy (Nakayama *et al.*, 1996).

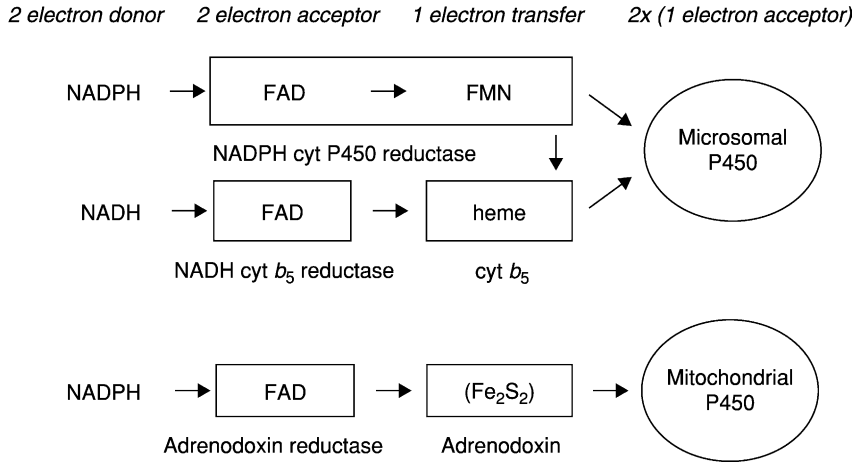


Figure 10 Mitochondrial and microsomal P450 redox partners.

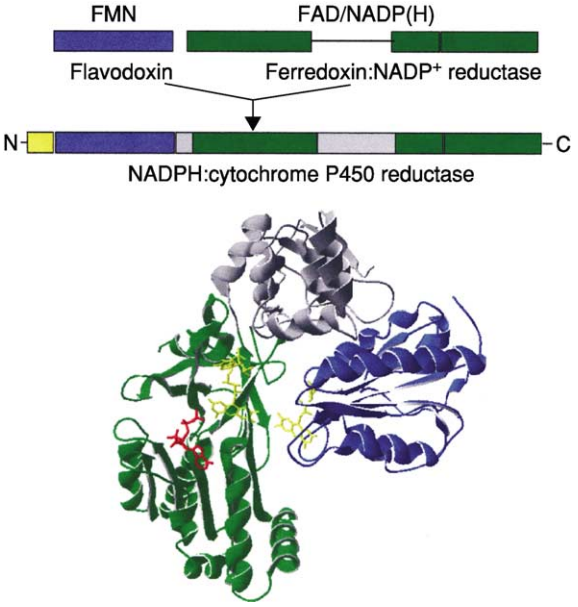


Figure 11 Structure of NADPH cytochrome P450 reductase. Top: evolutionary origin of the FMN (blue), FAD and NADP(H) (green) binding domains of the protein (yellow: membrane anchor; gray: connecting domain). Bottom: three-dimensional structure of the enzyme with the domains identified by color. The bound substrate NADPH (red) and cofactors FAD and FMN (yellow) are indicated. (Reprinted with permission from Murataliev, M.B., Feyereisen, R., Walker, F.A., **2004a**. Electron transfer by diflavin reductases. *Biochem. Biophys. Acta 1698*, 1–26; © Elsevier.)

The insect P450 reductases sequenced to date are clearly orthologous to the mammalian P450 reductases, with an overall amino acid sequence identity of 54% for the housefly P450 reductase, first cloned and sequenced in 1993 (Koener *et al.*, 1993). The housefly P450 reductase gene codes for a protein of 671 amino acids, and was mapped to chromosome III. The P450 reductases of other insects are very similar to the housefly enzyme – 82% identity for the *D. melanogaster* enzyme (Hovemann *et al.*, 1997), 57% identity for the *Bombyx mori* enzyme (Horike *et al.*, 2000), and 75% identity for the *A. gambiae* P450 reductase (Nikou *et al.*, 2003). The insect, mammalian, and yeast enzymes are functionally interchangeable in reconstituted systems of the purified proteins or in heterologous expression systems. However, no detailed study has documented how *well* a mammalian or yeast P450 reductase can support the activity of an insect P450 when compared to the cognate insect P450 reductase.

Early attempts to purify and characterize the enzyme from microsomes of housefly abdomens were hampered by the facile proteolytic cleavage of the N-terminal portion of the protein. This hydrophobic peptide anchors the reductase in the membrane, and its removal abolishes the ability of the remainder of the protein ("soluble" or "tryptic" reductase) to reduce P450s. The proteolytically cleaved reductase nonetheless retains the ability to reduce artificial electron acceptors such as cytochrome *c*, DCPIP or ferricyanide (Hodgson, 1985). Heterologous expression of the cloned P450 reductase has been achieved in *E. coli* (Andersen *et al.*, 1994) and in the baculovirus expression system (Wen *et al.*, 2003) and a purification scheme (Murataliev *et al.*, 1999) has produced quantities of enzyme sufficient for a detailed catalytic characterization of the enzyme's functioning and of its reconstitution with redox partners (**Figure 12**).

P450 reductase is an obligatory partner of microsomal P450 enzymes. Antisera to *Spodoptera eridania* or housefly P450 reductase inhibit all P450-dependent activities tested (Crankshaw *et al.*,

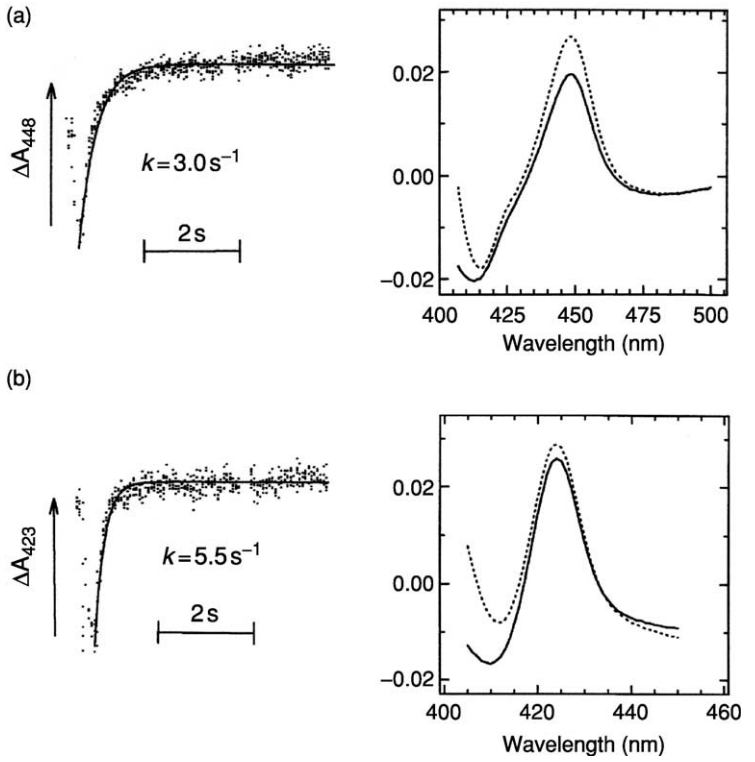


Figure 12 Reduction of housefly CYP6A1 (a) and cytochrome b_5 (b) by NADPH cytochrome P450 reductase. The recombinant proteins expressed in *E. coli* were reconstituted *in vitro*. On the left, the kinetics of reduction measured by stopped-flow spectrophotometry are shown with the calculated first-order rate constants. On the right, the end point difference spectra of CYP6A1 and cytochrome b_5 after reduction by P450 reductase (solid line) and sodium dithionite (dotted line). (Adapted from Guzov, V.M., Houston, H.L., Murataliev, M.B., Walker, F.A., Feyereisen, R., **1996**. Molecular cloning, overexpression in *Escherichia coli*, structural and functional characterization of house fly cytochrome b_5. *J. Biol. Chem. 271*, 26637–26645.)

1981; Feyereisen and Vincent, 1984). Immunoinhibition with P450 reductase antibodies serves as a strong indication of microsomal P450 involvement in NADPH-dependent activities, such as ecdysone 20-hydroxylation in the cockroach and in *B. mori* eggs (Halliday *et al.*, 1986; Horike and Sonobe, 1999; Horike *et al.*, 2000) and (*Z*)-9-tricosene biosynthesis in the housefly (Reed *et al.*, 1994). P450 reductase immunoinhibition could serve as a tool to distinguish P450 dependent activities from flavin monooxygenase (FMO) activities in microsomes from insect sources. P450 reductase also transfers electrons to cytochrome b_5 (see below) and to other microsomal enzymes such as heme oxygenase.

4.1.3.3.3. Cytochrome b_5 In contrast to P450 reductase, the role of cytochrome b_5 as partner in P450 dependent reactions is considerably more complex (see Section 4.1.3.3.3). The housefly cytochrome b_5 is a 134 amino acid protein (Guzov *et al.*, 1996) with 48% sequence identity with the orthologous rat cytochrome b_5. Its N-terminal domain of about 100 residues is the heme-binding domain that is about 60% identical to that of the vertebrate cytochrome b_5. Its C-terminal portion is a hydrophobic membrane anchor. A probable fatty acid desaturase-cyt b_5 fusion protein has been misidentified as the *Drosophila* cytochrome b_5 (Kula *et al.*, 1995; Scott, 1999; Kula and Rozek, 2000), but the correct ortholog is 76% identical to the housefly cytochrome b_5. The *H. armigera* cytochrome b_5 (Ranasinghe and Hobbs, 1999a) is 127 amino acids in length and 51% identical to the housefly cytochrome b_5, and the *A. gambiae* cytochrome b_5 is 54% identical (Nikou *et al.*, 2003). The known insect cytochrome b_5 sequences differ at their C-terminal from both the vertebrate microsomal and outer mitochonrial membrane cytochrome b_5 sequences, so that inferences about the subcellular targeting of the insect protein (Wang *et al.*, 2003) would seem premature.

The housefly cytochrome b_5 protein was produced in *E. coli*, purified and fully characterized (Guzov *et al.*, 1996). Absorption spectroscopy and EPR revealed properties very similar to cytochromes b_5 from vertebrates. NMR spectra indicated that the orientation of the heme in the protein relative to its α, γ *meso* axis is about 1 : 1. This means that the

protein is present in two forms of approximately equal abundance, that result from two modes of insertion of the noncovalently bound heme in the protein between the two coordinating histidines (face up and face down). Expression of the heme-binding domain in *E. coli* revealed that the heme is kinetically trapped in a 1.2 : 1 ratio of the two isomers, and that this orientation results from the selective binding of heme by the apoprotein (Wang *et al.*, 2003). A redox potential of −26 mV was measured by cyclic voltammetry on a treated gold electrode in the presence of hexamminechromium(III) chloride, and was verified by classical electrochemical titration. Stopped flow spectrophotometry showed that the cytochrome b_5 is reduced by housefly P450 reductase at a high rate ($5.5\ s^{-1}$) (Guzov *et al.*, 1996) (**Figure 12**).

Cytochrome b_5 can also be reduced by its own reductase, an NADH-dependent FAD flavoprotein, and can therefore provide either NADH- or NADPH-derived electrons to P450 enzymes. NADH-cytochrome b_5 reductase (EC 1.6.2.2) has been studied in *Ceratitis capitata* and *M. domestica* (Megias *et al.*, 1984; Zhang and Scott, 1996a). The N-terminus sequence of the purified housefly enzyme aligns to an internal sequence of the *Drosophila* enzyme (CG5946) indicating that it represents a proteolytically processed form. NADH-cytochrome b_5 reductase and cytochrome b_5 are also known to provide electrons to other acceptors, such as fatty acid desaturases and elongases.

4.1.3.3.4. Redox partners of mitochondrial P450 The redox partners of mitochondrial P450s are adrenodoxin reductase, an NADPH-dependent FAD flavoprotein and adrenodoxin, a [2Fe-2S] ferredoxin-type iron sulfur protein. These are named for the two redox partners of mammalian adrenal mitochondrial P450s, and this designation has been liberally bestowed on proteins from animals that don't have adrenals. Insect adrenodoxin reductase and adrenodoxin have not been functionally characterized, but their bovine orthologs are capable of supporting the activity of an insect mitochondrial P450, housefly CYP12A1 (Guzov *et al.*, 1998). The reduction of CYP12A1 is rapid and efficient with bovine adrenodoxin reductase/adrenodoxin while under the same conditions housefly microsomal P450 reductase is only marginally effective.

A fragment of a *Drosophila* adrenodoxin-like open reading frame is in GenBank on a stretch of DNA that also encodes a heatshock gene at 67B on the right arm of chromosome 3 (Pauli and Tonka, 1987). This stretch of 95 amino acids is similar (about 46%) to vertebrate adrenodoxin, and was initially identified as the *Drosophila* adrenodoxin ortholog (GenBank X06542). But the correct adrenodoxin ortholog was revealed by the complete genome sequence at 64B1 as a 152 amino acid protein, 45% identical to the bovine protein. EPR spectroscopic evidence for the presence of an adrenodoxin-like protein in fat body mitochondria of *Spodoptera littoralis* has been presented (Shergill *et al.*, 1995).

The *Drosophila* and *A. gambiae* adrenodoxin reductase and adrenodoxin genes have been annotated. The *Drosophila* P-element induced mutant *dare1* for *d*efective in the *a*voidance of *re*pellents was found to encode *D*rosophila *a*drenodoxin *re*ductase (Freeman *et al.*, 1999). Strong *dare* mutants undergo developmental arrest, and this phenotype is largely rescued by feeding 20-hydroxyecdysone. Decreasing by half the wild-type expression of *dare* blocks the olfactory response. The gene is expressed at low levels in all tissues of the adult fly, including the brain and the antennae. Highest expression is found in the prothoracic gland portion of the ring gland of third instar larvae, as well as in the nurse cells of adult ovaries. These tissues are known to require mitochondrial P450s for ecdysteroid production. The 55 kDa protein encoded by *dare* is 42% identical to the human enzyme.

4.1.3.4. Catalytic Mechanisms

4.1.3.4.1. P450 reactions Little work on insect P450 has focused on the catalytic cycle, and the mechanism derived from our understanding of the bacterial and mammalian P450 enzymes (Ortiz de Montellano, 1995b; Schlichting *et al.*, 2000) will be briefly summarized (**Figure 13**). The oxidized P450 is a mixture of two forms: a low spin (Fe^{III}) form with water as the sixth coordinated ligand on the opposite side of the Cys thiolate ligand, and a high spin (Fe^{III}) pentacoordinated form. Substrate binding displaces water from the sixth liganding position, leading to a shift to high spin. This shift can be observed (type I spectrum) and is accompanied by a decrease in the redox potential of P450. The P450–substrate complex receives a first electron from a redox partner (P450 reductase or adrenodoxin), and ferrous P450 (Fe^{II}) then binds O_2. At this step CO can compete with O_2 for binding to P450, its binding leads to a stable complex, with absorption maximum at 450 nm (**Figure 5**), that is catalytically inactive. CO can be displaced by light irradiation at 450 nm. The P450–O_2–substrate complex in the form of a ferric peroxide complex then accepts a second electron (from P450 reductase or in some cases cytochrome b_5, or from adrenodoxin). Different types of activated oxygen forms of the same

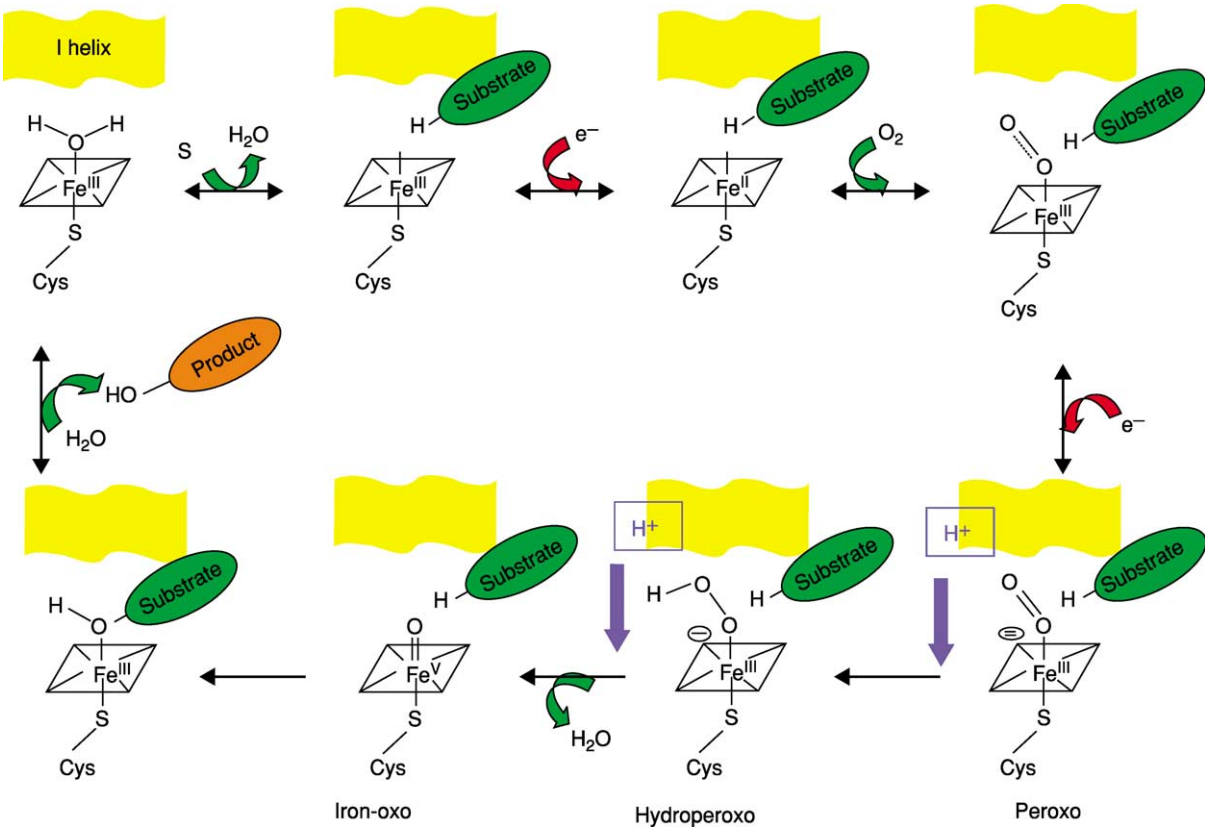


Figure 13 Catalytic cycle of P450 enzymes in monooxygenation reactions. Three possible forms of the activated oxygen species are shown. See text for details. Other reactions (reduction, isomerization, dehydration) can be catalyzed by oxygen-free forms of the enzyme. (Adapted with permission from Werck-Reichhart, D., Feyereisen, R., **2000**. Cytochromes P450: a success story. *Genome Biol. 1*, 3003.1–3003.9; © GenomeBiology.)

P450 enzyme can then be formed, depending on the protonation state of the complex and on the homolytic or heterolytic cleavage of the reduced dioxygen. Although the formal reaction is the insertion of an atom of oxygen into the substrate, the other atom being reduced to water (hence the term "mixed-function oxidase"), the nature of the oxidizing species can vary. A P450 $(Fe^{III}–O–O)^{2-}$ peroxo-iron form, a P450 $(Fe^{III}–O–OH)^{-}$ hydroperoxo form, and a P450 $(Fe^{V}=O)$ or P450 iron-oxo form are the preferred descriptions of the activated oxygen forms (Ortiz de Montellano, 1995b; Schlichting *et al.*, 2000; Newcomb *et al.*, 2003). The type of reaction catalyzed then depends on the substrate and substrate binding site and varies from hydroxylation to epoxidation, *O*-, *N*-, and *S*-dealkyation, *N*- and *S*-oxidations, or at least 60 different chemical reactions. The types of reactions currently known to be catalyzed by insect P450 enzymes are listed in **Table 2**. The P450(Fe^{II})–substrate complex can function as a reductase, and the P450–O_2–substrate complex can also function as an oxidase, releasing superoxide, hydrogen peroxide, or water. Under experimental conditions, NADPH and molecular oxygen can be substituted by organic hydroperoxides, sodium periodate, etc., in what is called the peroxide shunt (Ortiz de Montellano, 1995a).

The obligatory role of P450 reductase in catalysis of the microsomal P450 has been proven in reconstitution experiments, but the role of phospholipids is less clear and has not been specifically studied. The role of cytochrome b_5 is discussed below. Activity of the mitochondrial CYP12A1 also showed absolute dependence on reconstitution in the presence of (bovine) mitochondrial redox partners (Guzov *et al.*, 1998).

The stoichiometry (eqn [1])

$$\mathrm{RH} + \mathrm{O_2} + \mathrm{NADPH} + \mathrm{H^+} \rightarrow \mathrm{ROH} + \mathrm{H_2O} + \mathrm{NADP^+}$$

that commonly describes the monooxygenase (*sensu* Hayaishi) or mixed-function oxidation (*sensu* Mason) reaction of P450 has not been confirmed experimentally for any insect P450. A more complex stoichiometry would take into account the "leakage" of activated oxygen species as superoxide, hydrogen

Table 2 Enzymatic reactions catalyzed by insect P450 enzymes

Reaction catalyzed	*P450*
Oxidase activity	
O_2 to H_2O, H_2O_2, O_2^-	CYP6A1 (and probably most P450)
Monooxygenations	
Aliphatic hydroxylation	
C–H hydroxylation	CYP4C7, CYP6A1, CYP6A2, CYP6A8, CYP12A1, CYP302A1, CYP312A1, CYP314A1, CYP315A1
O-dealkylation	CYP6A1, CYP6D1[a], CYP12A1, CYP6A5, CYP6B4, CYP6B17, CYP6B21
Dehalogenation	CYP6A2 (DDT to DDA, DDD)
Epoxidation	CYP6A1, CYP6A2, CYP12A1, CYP15A1, CYP9E1
Aromatic hydroxylation	CYP6D1[a]
Heteroatom oxidation and dealkylation	
Phosphorothioate ester oxidation	CYP6A1, CYP6A2, CYP12A1, CYP6D1[a]
N-dealkylation	CYP12A1, CYP6A5
N-oxidation	+ (nicotine)
S-oxidation	+ (phorate)
Aldehyde oxidation	+ (C-26 hydroxyecdysteroids)
Complex and atypical reactions	
Carbon–carbon cleavage	+? (sterols, ecdysteroid)
Decarbonylation with C–C cleavage	+ (P450hyd)
Aromatization	+ (defensive steroids)
Dehydrogenation	+ (cholesterol)
Dehydration	
Aldoxime dehydration	+ (R-CN biosynthesis)
Reduction	−
Endoperoxide isomerization	−

[a]Inference from immunoinhibition experiments.
+, indicates metabolism by microsomal P450, but specific enzyme not identified (substrate indicated), −, indicates no evidence to date.

peroxide, and water at the expense of NADPH during catalysis. In a "well coupled" reaction as in (eqn [1]) these by-products would not be formed. It is assumed, but not generally proven, that a specialized P450 metabolizing its favorite substrate would follow stoichiometry (eqn [1]). An approximate balance for CYP6A1 epoxidation of heptachlor (*Hept*) gives the following results: (M.B. Mutataliev, V.M. Guzov, R. Feyereisen, unpublished data).

$$1\ Hept + \underline{13\ O_2} + \underline{11.2\ NADPH} \rightarrow \underline{1\ Hept\ epoxide} + 1.5\ H_2O + 11.2\ NADP^+ + \underline{9.8\ H_2O_2} \quad [2]$$

and for testosterone (*Tst*) hydroxylation under the same experimental conditions:

$$1\ Tst + \underline{6\ O_2} + \underline{4.7\ NADPH} \rightarrow \underline{1\ Tst\text{-}OH} + 0.5\ H_2O + 4.7\ NADP^+ + \underline{4.2\ H_2O_2} \quad [3]$$

In eqns [2] and [3], the parameters that were measured are underlined. These stoichiometries show that a P450 such as CYP6A1 can be simultaneously an oxidase and a monooxygenase. These coupling stoichiometries (M.B. Mutataliev, V.M. Guzov, R. Feyereisen, inpublished data) are dependent on the ratio of P450 and P450 reductase, as well as on the presence or absence of cytochrome b_5 (see below). Note that those stoichiometries are not balanced, reflecting experimental error in the measurements and that the addition of superoxide dismutase did not change the amount of H_2O_2, indicating either no superoxide production, or lack of success in measuring it. The uncoupling of monooxygenation is highly likely to be a common feature of insect P450 enzymes that metabolize xenobiotics of synthetic or plant origin. The generation of reactive oxygen species is a corrolary of active P450 metabolism.

4.1.3.4.2. P450 reductase

4.1.3.4.2.1. P450 interaction with P450 reductase

As seen above, the proper functioning of P450 enzymes depends on an efficient electron supply. In insect microsomes, the ratio of P450 enzymes to P450 reductase is about 6–18 to 1 (Feyereisen *et al.*, 1990). In this ratio, all P450 enzymes are summed, so that the actual ratio of one specific P450 enzyme to P450 reductase is probably smaller. The rate of the overall microsomal P450 reaction (two transfers of one electron) is relatively slow so that dissociation of the P450–P450 reductase complex is possible between the first and the second electron transfer. Indeed, cytochrome b_5 can replace P450 reductase

for the supply of the second electron in some cases (see below). The effect of varying the P450/P450 reductase ratio on catalytic rates was measured in a reconstituted system for heptachlor epoxidation by CYP6A1. The rate of epoxidation was determined by the concentration of the binary complex of P450 and P450 reductase, with the same high rate being observed in the presence of an excess of either protein (**Figure 14**). The half-saturating concentration of either protein was about 0.1 μM in the presence of cytochrome b_5 (M.B. Murataliev, V.M. Guzov, R. Feyereisen, unpublished data). This is in good agreement with the Km of 0.14 and 0.5 μM for P450 reductase in the presence and absence of cytochrome b_5 measured previously (Guzov *et al.*, 1996). Coinfection of Sf9 cells with baculoviruses carrying CYP6B1 and P450 reductase has revealed that highest catalytic activity was achieved at an equivalent, moderate, multiplicities of infection for the two viruses (Wen *et al.*, 2003). Higher enzymatic activities of cell lysates towards furanocoumarins was not achieved when either protein was produced in excess. This result can be explained in part by documented limitations of the cell's ability to host, fold, and provide cofactors for both P450 and reductase (Wen *et al.*, 2003), but it also supports the idea that highest activity is achieved for the highest concentration of the binary complex of the two partners.

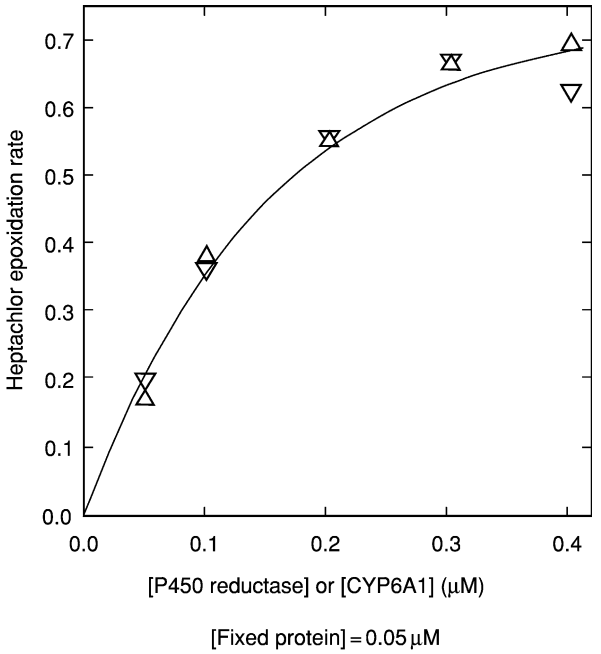


Figure 14 Heptachlor epoxidation by *E. coli*-expressed recombinant housefly CYP6A1 and NADPH cytochrome P450 reductase. A reconstituted system containing variable concentration of CYP6A1 (△) or P450 reductase (▽) and a fixed concentration (0.05 μM) of the reciprocal partner, and a cytochrome b_5 concentration of 1.0 μM was incubated in the presence of NADPH. (M.B. Murataliev, V.M. Guzov, R. Feyereisen, unpublished data)

4.1.3.4.2.2. P450 reductase functioning P450 reductase accepts two electrons from NADPH; more precisely, it accepts a hydride ion (one hydrogen plus one electron), and donates two electrons, one at a time, to P450 enzymes. P450 reductase is therefore an enzyme with two substrates: NADPH and the electron acceptor (P450 or artificial acceptor such as cyt *c*), and two products: $NADP^+$ and the reduced electron acceptor. With two bound flavins and a pathway of electron transfer NADPH > FAD > FMN > P450 (or cyt *c*), its reduction state during catalysis can theoretically vary between the fully oxidized state (0 el.) and the fully reduced state (4 el.). Studies with the purified recombinant housefly P450 reductase (Murataliev *et al.*, 1999; Murataliev and Feyereisen, 1999; Murataliev and Feyereisen, 2000; review: Murataliev *et al.*, 2004a) have shed light on two questions posed by this electron transfer function: what is the kinetic mechanism of this two-substrate enzyme (Ping-Pong or sequential Bi-Bi) and what are the respective reduction states of the two flavins during catalysis?

In the ping-pong mechanism, the first product of the reaction must be released before the second substrate binds to the enzyme, and no ternary complex is formed. In sequential Bi-Bi mechanisms both substrates bind to the enzyme to form a ternary complex. Although several kinetic mechanisms have been proposed (Hodgson, 1985), a careful study of the recombinant housefly P450 reductase clearly established a sequential random Bi-Bi mechanism (Murataliev *et al.*, 1999). The great sensitivity of the enzyme to ionic strength hampers the comparison of different studies (Murataliev *et al.*, 2004a). The formation of a ternary complex of NADPH, P450 reductase, and the electron acceptor suggested a role for reduced nucleotide binding in the catalysis of fast electron transfer. The rate of cytochrome *c* reduction was shown to equal the rate of hydride ion transfer from the nucleotide donor to FAD (Murataliev *et al.*, 1999). A faster electron transfer rate was observed with NADPH as compared to NADH (Murataliev *et al.*, 1999) and the 2′-phosphate was shown to contribute to more than half of the free energy of binding (Murataliev and Feyereisen, 2000). The affinity of the oxidized P450 reductase was ten times higher for NADPH than for $NADP^+$ (Murataliev *et al.*, 1999), and a conformational change induced by NADPH binding and important for fast catalysis was suggested by these studies.

The state of reduction of the flavins of P450 reductase during catalysis was deduced from kinetic

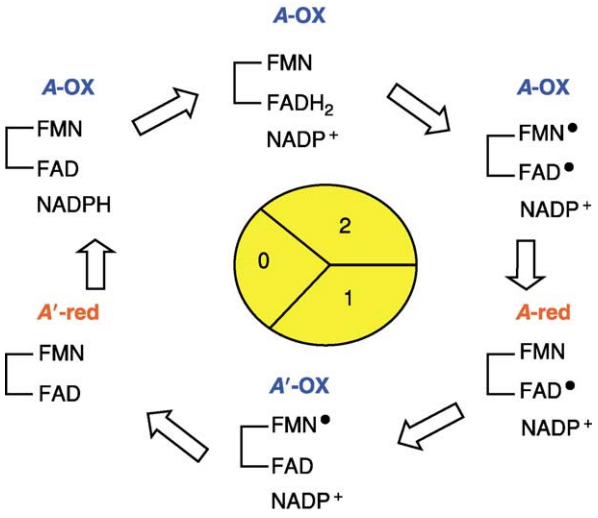


Figure 15 Reduction state of NADPH cytochrome P450 reductase during catalysis, the ''0-2-1-0'' cycle. The enzyme cycles between a fully oxidized state and a 2-electron reduced state. The electron acceptor (A and A′ = P450 or cytochrome *c*) receives one electron at a time from an FMN semiquinone form ($FMN^{\bullet}$) of the enzyme. The release of $NADP^{+}$ is shown here to occur at the last step but may occur earlier. Implicit additional steps are not shown, for clarity. See text for details and Murataliev *et al.* (2004a) for review.

experiments, rates of NADPH oxidation and EPR measurements of flavin semiquinone (free radical) levels. These revealed the existence of a catalytically competent FMN semiquinone, different from the "blue" neutral FMN semiquinone, known as the air-stable semiquinone that is not a catalytically relevant form of the enzyme (Murataliev and Feyereisen, 1999). Furthermore, the detailed studies of housefly P450 reductase led to a proposed catalytic cycle where the reduction state of the enzyme does not exceed 2 el., and where an FMN semiquinone, and not an FMN hydroquinone, serves as electron donor to the acceptor P450 or cytochrome *c*. This "0-2-1-0 cycle" (**Figure 15**) likely represents the general mechanism of P450 reductases, with strong evidence that it operates in P450BM3 and in the human P450 reductase as in the fly P450 reductase (Murataliev *et al.*, 2004a).

4.1.3.4.3. Role of cytochrome b_5 Depending on the P450 enzyme and on the reaction catalyzed, cytochrome b_5 may be either inhibitory, without effect, or its presence may be obligatory. Cytochrome b_5 can have a quantitative effect on overall reaction rates, and/or a qualitative role on the type of reaction catalyzed and the ratio of the reaction products. The role of cytochrome b_5 may or may not depend on its redox (electron transfer) properties. It can also influence the overall stoichiometry of the P450 reaction, in particular the "coupling rate," i.e., the utilization and fate of electrons from NADPH relative to monooxygenation. Cytochrome b_5 should therefore be regarded as an important *modulator* of microsomal P450 systems. General reviews of the role of cytochrome b_5 in P450 reactions are available (Porter, 2002; Schenkman and Jansson, 2003) and known examples of this modulator role in insect systems follow.

The relative contribution of NADH in P450 reactions, but more importantly the NADH synergism of NADPH-dependent reactions that is occasionally observed (e.g., Ronis *et al.*, 1988; Feng *et al.*, 1992), is probably attributable to cytochrome b_5 as redox partner. Indeed, the Km of the P450 reductase for NADH is a thousand-fold higher than for NADPH, and the V_{max} tenfold lower (Murataliev *et al.*, 1999), so that the contribution of NADH under normal conditions is probably channeled by NADH-cytochrome b_5 reductase and cytochrome b_5.

An anticytochrome b_5 antiserum severely inhibited (up to 90%) methoxycoumarin and ethoxycoumarin *O*-dealkylation and benzo[a]pyrene hydroxylation, but not methoxyresorufin and ethoxyresorufin *O*-dealkylation when assayed in microsomes of the housefly LPR strain (Zhang and Scott, 1994). This antiserum also inhibits cypermethrin 4′-hydroxylation by these CYP6D1-enriched microsomes (Zhang and Scott, 1996b).

Housefly cytochrome b_5 stimulates heptachlor epoxidation and steroid hydroxylation when reconstituted with cytochrome P450 reductase, housefly CYP6A1, and phospholipids (Guzov *et al.*, 1996; Murataliev *et al.*, 2004a). Stimulation of cyclodiene epoxidation and diazinon metabolism were also observed with *Drosophila* CYP6A2 expressed with the baculovirus system (Dunkov *et al.*, 1997). Cytochrome b_5 is efficiently reduced by P450 reductase (**Figure 12**), but it does not increase the rate of P450 reduction by P450 reductase. Because of its small redox potential (see above), cytochrome b_5 is unlikely to play an important role in delivering the first electron to P450 catalysis, and its stimulatory role probably involves an increased rate of transfer of the second electron. Cytochrome b_5 decreases the apparent Km for P450 reductase and increases the V_{max} for epoxidation at constant CYP6A1 concentrations (Guzov *et al.*, 1996). The results suggest a role for cytochrome b_5 in the P450 reductase–P450 interactions.

Whereas heptachlor epoxidation by CYP6A1 was increased two- to threefold by the addition of cytochrome b_5, the hydroxylation of testosterone, androstenedione, and progesterone was stimulated

seven- to tenfold. The addition of cytochrome b_5 increased the ratio of 2β-hydroxylation over 15β-hydroxylation of testosterone. This suggests that cytochrome b_5 can have an effect on CYP6A1 conformation, probably altering the interaction of the binding site with either the C-17 hydroxyl group (decreased) or the C-3 carbonyl (increased). Interestingly, the effect of cytochrome b_5 on hydroxylation regioselectivity was also obtained with apo-b_5 (cytochrome b_5 depleted of heme and therefore redox incompetent), whereas the effect on turnover number was only much smaller with apo-b_5. The effect of apo-b_5 is not due to heme transfer from P450 to apo-b_5 and, in fact, both apo-b_5 and (holo) cytochrome b_5 were shown to stabilize the ferrous–CO complex of CYP6A1, decreasing the rate of its conversion to P420 (M.B. Murataliev, V.M. Guzov, R. Feyereisen, unpublished data).

Cytochrome b_5 increases the coupling stoichiometry of CYP6A1 catalysis. In the heptachlor epoxidation assay, coupling (NADPH or O_2 used/heptachlor epoxide formed) increased from <8% to over 25%. This effect is even more pronounced for testosterone, where coupling efficiency in the presence of cytochrome b_5 can reach 84%. The effect of cytochrome b_5 results in a decrease in H_2O_2 production in both assays. The exact site of H_2O_2 production (P450 reductase or CYP6A1) is not known.

Coordinate induction and/or overexpression of cytochrome b_5 and P450 genes has been reported (Liu and Scott, 1996; Kasai *et al.*, 1998b; Ranasinghe and Hobbs, 1999a, 1999b; Nikou *et al.*, 2003) and this indicates that the effects of cytochrome b_5 seen *in vitro* may have significance *in vivo* as well.

4.1.3.4.4. Mechanisms and specificity of P450 inhibitors The common features of electron transfer, ligand binding, and catalysis described above are the features that determine the relative success of P450 inhibitors. Compounds that act as electron sinks and are readily autooxidizable can inhibit P450 reactions by inhibiting electron transfer by the respective redox partners. This mechanism is typical of the eye pigment xanthommatin that was identified as "endogenous inhibitor" in early studies (review: Hodgson, 1985). Several flavonoids may act in this way and care must be taken to distinguish P450 inhibition *per se* from inhibition of electron transfer.

Insecticide synergists (**Figure 16**) are among the most interesting inhibitors of P450 because of their widespread commercial use, in particular piperonyl butoxide. A landmark review paper on synergists remains that of Casida (1970). Synergists (Hodgson, 1985; Bernard and Philogene, 1993) as well as P450 inhibitors in general (Ortiz de Montellano and Correia, 1995) have been covered in other insightful reviews as well. The synergism of carbaryl by piperonyl butoxide has been used in a survey of 54 insect species to estimate P450 activity *in vivo* (Brattsten and Metcalf, 1970). Although the synergistic ratio is most often presented, the usefulness of a synergistic difference has also been proposed (Brindley, 1977). The mode of action of piperonyl butoxide and other related methylenedioxyphenyl (MDP) compounds (or benzodioxole compounds) involves an initial metabolic activation by the P450 enzyme, leading to the formation of a carbene–iron complex that is virtually irreversible (Ortiz de Montellano and Correia, 1995). It follows that those P450 enzymes with (1) low affinity for the MDP compound and/or (2) low capacity to metabolize it to the carbene inhibitory form will not be inhibited, and thus piperonyl butoxide and other MDP compounds are not universal inhibitors of all P450 enzymes. Self-catalyzed destruction of P450 enzymes by terminal acetylenes or olefins and other "suicide substrates" is also, and for the same reasons, not equally effective for all P450 enzymes. The phenylpropynyl ether synergists such as TCPPE fall into this category, as well as the newer synergist verbutin (Bertok *et al.*, 2003). This nongenerality of inhibition has been well documented *in vivo*. For instance, TCPPE can be an effective synergist when piperonyl butoxide

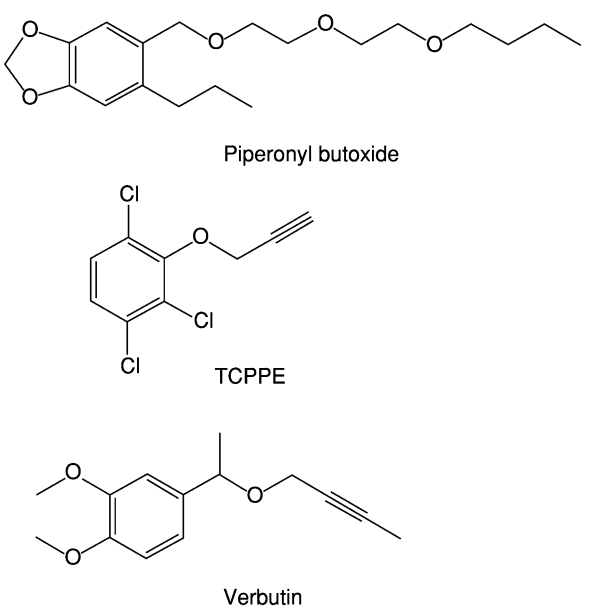


Figure 16 Structures of the synergists and P450 inhibitors piperonyl butoxide, a typical methylene dioxyphenyl (MDP) compound, TCPPE (trichlorophenylpropynyl ether) and verbutin.

cannot (Brown *et al.*, 1996; Zhang *et al.*, 1997). In the case of 1-aminobenzotriazole (ABT), which is metabolized to benzyne that covalently binds to the prosthetic heme, *in vitro* P450 destruction and formation of the porphyrin adduct have been measured in the housefly (Feyereisen *et al.*, 1984). P450 protein labeling by P450 inhibitors has also been achieved (Andersen *et al.*, 1995; Cuany *et al.*, 1995). The NADPH-dependent decrease in P450 caused by ABT, TCPPE, or piperonyl butoxide differs according to the induction status and fly strain, suggesting selectivity (Feyereisen *et al.*, 1984). This selectivity can be harnessed into the design of useful "suicide" inhibitors of, e.g., ecdysone biosynthesis (Luu and Werner, 1996). The synthesis of the MDP moiety seen in many plant natural products is itself dependent on a P450 activity (CYP719, Ikesawa *et al.*, 2003), and it is postulated that such compounds may be a legacy of evolutionary interactions with insect and other enemies (Berenbaum and Neal, 1987).

Another class of powerful P450 inhibitors are heterocyclic compounds with an sp2 hybridized nitrogen as in pyridines, azoles, and imidazoles. These compounds bind simultaneously to the heme iron (type II ligands) and to a hydrophobic binding site of the P450 for its substrate. This "two-point binding" has therefore an intrisinc potential for selectivity, with the substrate mimic moiety of the inhibitor targeting the specific P450 and the type II-ligand moiety coordinating the heme, and inhibiting the enzyme. This reasoning has led to the design of potent inhibitors and photoaffinity labels (Andersen *et al.*, 1995) for a specific insect P450 (see Section 4.1.4.1.2 below). The commercial importance of this type of P450 inhibitors is emphasized by the fungicides and CYP51 inhibitors miconazole and ketoconazole.

4.1.3.4.5. Substrate selectivity and structure/function of P450 enzymes The substrate recognition sites (SRS) of CYP2 proteins were first described by Gotoh (1992) as highly variable regions. Subsequently, a large number of site-directed mutagenesis studies have focused on these regions to explore substrate specificity of mammalian and bacterial P450 enzymes (e.g., review: Domanski and Halpert, 2001). Although multiple sequence alignments usually show that SRS1 is highly polymorphic, the SRS1 of the CYP6B sequences of Lepidoptera are highly conserved (Berenbaum *et al.*, 1996). This suggested that in this case SRS1 may contribute to the recognition of furanocoumarins – substrates of most CYP6B enzymes studied to date. Mutagenesis of six residues of CYP6B1v1 from *Papilio polyxenes* provided some evidence for this hypothesis (Chen *et al.*, 2002). Seven mutants at the Phe116 position in SRS1 led to incorrectly folded or assembled P450 proteins that were catalytically inactive and that had absorption maxima at 420 nm instead of 450 nm in their CO-difference spectrum. One mutant, Phe116 to Trp, had a severely reduced catalytic activity. The Phe116 to Tyr mutant (insertion of a hydroxyl group on the aromatic side chain) was catalytically active and showed an altered substrate specificity towards furanocoumarins. Xanthotoxin metabolism was drastically reduced, bergapten and isopimpinellin metabolism were cut in half whereas metabolism of trioxsalen or psoralen were not affected. Homology modeling of the CYP6B1v1 structure based on a CYP102 crystal structure indicated that the Phe116 side chain projects into the active site above the plane of the heme. Four mutants at the neighboring His117 were catalytically inactive as were nine mutants at the Phe484 position, which is located in SRS6. Val368 is located in SRS5, opposite Phe116 and also in the vicinity of the active site. Its replacement by Phe resulted in a dead enzyme, whereas the mutations of Val368 to Ala or Leu did not affect CYP6B1v1 activity or specificity (Chen *et al.*, 2002). Two further replacements were either lethal (Phe206 to Leu) or without effect (His204 to Leu). These residues are outside the SRS regions, but indicated by homology modeling to be located near the substrate access channel. The overall result of this extensive set of experiments on 33 variants indicates that CYP6B1v1 is very sensitive to changes at sites that control access and geometry of the active site. Only the Phe166 to Tyr mutation in SRS1 showed a clear effect on substrate selectivity towards furanocoumarins. Only three of the inactive mutants at the Phe484 position showed a normal CO difference spectrum, and thus may have lost activity towards furanocoumarins while retaining or gaining activity towards other substrates. This residue in SRS6 is not conserved among CYP6B enzymes, however (Chen *et al.*, 2002). Further analyses of *Helicoverpa* and *Papilio* CYP6B sequences have attempted to describe ancestral CYP6B sequences and to model the geometry of their active site (Li *et al.*, 2003). The crystal structure of a CYP6B protein would contribute enormously to anchor such studies on a firm basis.

Results obtained with naturally occurring mutations situated outside the SRS regions show that substrate specificity is not encoded in the SRS alone. Three point mutations in *Drosophila* CYP6A2 (Bergé *et al.*, 1998), when combined, confer to this P450 the ability to metabolize DDT

without modifying the metabolism of testosterone (Amichot *et al.*, 2004). These mutations found in a DDT-resistant strain (see Section 4.1.4.5.3 below) are located between SRS4 and SRS5 in the J helix (Arg335 to Ser and Leu336 to Val) and at the end of helix L, before SRS6 (Val476 to Leu). The three mutations are located towards the "top" of helix I and may together influence the positioning of this helix and hence access to the active site (Amichot *et al.*, 2004). Significantly, the recombinant enzyme carrying only the Arg335 to Ser mutation has only a fraction of the DDT metabolizing capacity of the triple (naturally occurring) mutant, and is unstable. A better understanding of the structure–activity relationships in insect P450 enzymes is still very distant, but the diversity of insect P450 sequences may prove useful in the engineering of P450 as biocatalysts. For instance, the replacement of portions of SRS1 of P450BM3 (CYP102, a fatty acid ω-1 hydroxylase) by the homologous portions of SRS1 from CYP4C7 (a terpenoid ω-hydroxylase, see Section 4.1.4.1.2) modifies the regioselectivity of hydroxylation of fatty acids and terpenoids of the P450BM3 enzyme (Murataliev *et al.*, 2004b).

Information on substrate access to the active site and active site topology can also be inferred from spectral studies (Section 4.1.3.1.2) and biochemical studies (**Figure 7**). The active site topology of CYP6A1 was studied by a technique developed in Ortiz de Montellano's group (Ortiz de Montellano and Graham-Lorence, 1993). The enzyme is first incubated with phenyldiazene to form a phenyl–iron complex. Ferricyanide-induced *in situ* migration of the phenyl group to the porphyrin nitrogens causes the formation of covalent adducts, which can then be separated by HPLC (**Figure 17**). The *N*-phenyl protoporphyrin IX adducts of CYP6A1 were formed in a 17 : 25 : 33 : 24 ratio of the $N_B : N_A : N_C : N_D$ isomers (Andersen *et al.*, 1997). Thus in the native protein, all four pyrrole groups are somewhat exposed, whereas in several other P450s, labeling is more specific. Specific labeling indicates that the protein encumbers more space on top of the heme prosthetic group, e.g., leaving only one pyrrole ring exposed as in P450scc (Pikuleva *et al.*, 1995). The type of labeling seen with CYP6A1 indicates less encumbrance by the protein on top of the heme as in CYP3A4 (Schrag and Wienkers, 2000). These experiments suggest that the active site of CYP6A1 is relatively accessible and not severely constrained. Indeed CYP6A1 metabolizes flat steroids, bulky cyclodiene insecticides, as well as a variety of sesquiterpenoids (**Table 1**) (Andersen *et al.*, 1994, 1997; M.B. Murataliev, V.M. Guzov, R. Feyereisen, unpublished data). High uncoupling of electron transfer relative to monooxygenation may be a result of this broad substrate specificity.

4.1.4. P450 Functions

Ever since the pioneering work of Agosin and of Terriere in the early sixties on microsomal enzymes that hydroxylate DDT and naphthalene (review: Agosin, 1985), the enzymes now recognized as P450 have been best known for their role in xenobiotic metabolism in insects. They are often denoted as "detoxification enzymes." This designation not only neglects the importance of P450 enzymes in basic physiological processes, as pointed out for drug-metabolizing enzymes in general by Nebert (1991). It also neglects the conceptual difference between metabolism (what the enzyme does to the chemical) and

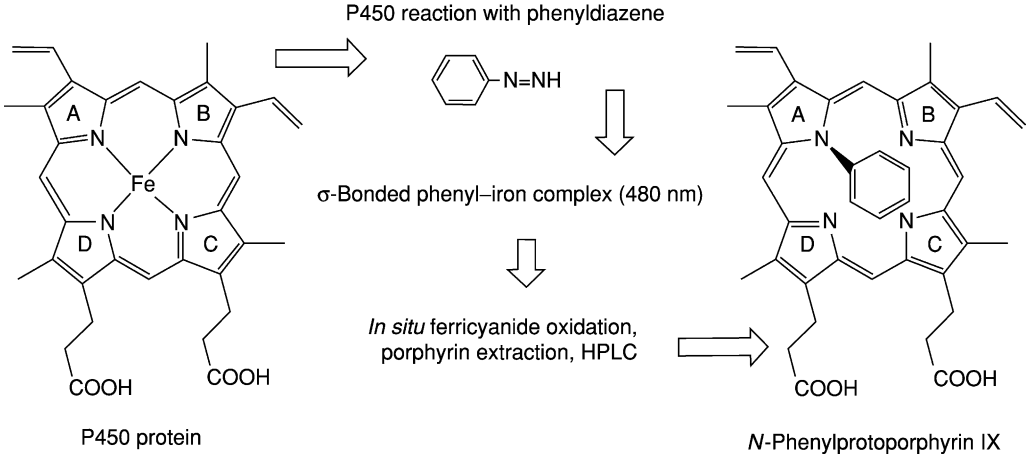


Figure 17 Active site topology of a P450 enzyme tested by formation of phenyl-porphyrin adducts. In this example, an adduct with pyrrole ring A is formed. The ratio of the four possible adducts indicates the degree of encumbrance in the native P450 protein.

toxicity (what the chemical does to the biological system). As noted before (Feyereisen, 1999):

> there are many cases where P450 enzymes act as anything but "detoxification enzymes." The easy dichotomy between biosynthetic and detoxification functions of P450s reflects more the teleological tendencies of the observer than the phylogeny or biochemistry of the enzymes.

In most studies, the P450 substrates tested are either endogenous compounds and their analogs, *or* xenobiotics. Rarely are both types of substrates tested on the same P450 and, indeed, the thermodynamic reality of enzyme–substrate interaction makes the quest for a formally complete description of the chemical diversity of substrates of any enzyme illusory. A future understanding of P450 functions perhaps will be based more on their evolutionary trajectories, some P450s being presently constrained by their proven physiological role (but how long have they played this role?) and some P450s being positively selected for their present adaptive role in detoxification (of plant chemicals in evolutionary time, or of insecticides in historical time). Our knowledge of insect P450 function is not sufficient to classify individual P450s or whole CYP families rationally along such evolutionary lines, so the following description may be found to be arbitrary.

4.1.4.1. Metabolism of Signal Molecules

4.1.4.1.1. Ecdysteroid metabolism

4.1.4.1.1.1. Ecdysone 20-monooxygenase activity (E20MO) The conversion of ecdysone to 20-hydroxyecdysone does not occur in the prothoracic glands but occurs in many peripheral tissues, such as the fat body, midgut, and Malpighian tubules (see **Chapter 3.3**). The P450 nature of the enzyme catalyzing the 20-hydroxylation of ecdysone was well established in 1985 (Smith, 1985) following the initial reports of 1977 (Bollenbacher *et al.*, 1977; Feyereisen, 1977; Johnson and Rees, 1977). An NADPH and O_2-dependent enzyme system, inhibited by typical pharmacological P450 inhibitors, was studied in several insect species. The evidence for P450 involvement was strengthened by the light-sensitive carbon monoxide inhibition observed in some of the most thorough studies (Feyereisen and Durst, 1978; Smith *et al.*, 1979; Greenwood and Rees, 1984). The agreement on the P450 nature of the reaction was accompanied by a lack of consensus on the subcellular localization of E20MO. Some studies showed a microsomal activity, other studies showed a mitochondrial activity, and yet other studies indicated the presence of both microsomal and mitochondrial activities in the same tissue (Smith, 1985). The differences in species, tissues, and experimental conditions, while real, did not permit glossing over the dual subcellular localization of E20MO (**Chapter 3.3**).

Studies past 1985 continued to present evidence for either the mitochondrial or microsomal, or both, localizations of the E20MO activity. For instance, it has been reported to be mostly microsomal in imaginal discs of *Pieris brassicae* (Blais and Lafont, 1986) and in *Gryllus bimaculatus* midgut (Liebrich and Hoffmann, 1991). In the midgut of *Diploptera punctata* (Halliday *et al.*, 1986) and in embryos of *Bombyx mori* (Horike and Sonobe, 1999), the activity is essentially microsomal, and can be inhibited by antibodies to the insect P450 reductase. This is clear evidence that the enzyme derives its reducing equivalents from the usual microsomal redox partner in those cases.

In *Spodoptera littoralis* fat body, E20MO activity is predominantly mitochondrial, with a small amount of microsomal activity (Hoggard and Rees, 1988). The mitochondrial E20MO activity was reportedly inhibited by antibodies to vertebrate P450scc (CYP11A), P45011β (CYP11B), adrenodoxin, and adrenodoxin reductase (Chen *et al.*, 1994), despite the considerable sequence divergence predicted between the vertebrate and insect proteins. The immunodetection of polypeptides significantly larger than predicted (e.g., P450 at 82 kDa and adrenodoxin at 73 kDa) was also a surprising feature of that study. In whole body homogenates of third-instar *Drosophila*, and in the midgut of larval *Spodoptera frugiperda*, the E20MO activity is distributed 1 : 3 between mitochondrial and microsomal fractions (Smith, 1985; Yu, 1995), and it is also distributed in both fractions in larvae of the housefly, and of the flesh fly *Neobellieria bullata* (Darvas *et al.*, 1993). Weirich *et al.* (1996) compared the apparent Km and specific activities of the mitochondrial and microsomal E20MO activities of *Manduca sexta* larval midgut. They concluded that at physiological ecdysone titers, despite a higher specific activity, the mitochondrial E20MO would contribute less than one-eighth the activity of the microsomal form.

The probable existence of several genes encoding P450s with E20MO activity was suggested (Feyereisen, 1999). This would account for the dual localization and complex regulation of the enzyme's activity. On the other hand, alternative splicing or posttranslational modifications of a single gene product could also lead to the microsomal and mitochondrial forms. Interestingly, both microsomal and mitochondrial activities of the *Spodoptera littoralis* fat body E20MO were reported to be

reversibly activated by phosphorylation (Hoggard and Rees, 1988; Hoggard *et al*., 1989). Cases of posttranslational modifications of P450 enzymes by reversible phosphorylation are not very common (Jansson, 1993; Oesch-Bartlomowicz and Oesch, 2003), and this observation would suggest that the *S. littoralis* fat body E20MOs are products of the same gene. The ecdysteroid 26-hydroxylases of *M. sexta* midgut and mitochondria are also both regulated by phosphorylation (Williams *et al*., 2000b).

Petryk *et al*. (2003) identified *Drosophila* CYP314A1 as the product of the *shade* (*shd*) gene, a member of the Halloween group of developmental mutants. Expression of CYP314A1 in *Drosophila* S2 cells enabled the NADPH-dependent hydroxylation of ecdysone to 20-hydroxyecdysone by cell homogenates. RNA *in situ* hybridization shows that the *shd* gene is not expressed in the ring glands, but expression is seen in the gut, fat body, and Malpighian tubules. In embryos, *shd* is expressed primarily in epidermal cells by the time of germ band extension. Embryonic lethality of *shd* mutants indicates that *Cyp314a1* encodes the only significant E20MO activity at that stage in *Drosophila*. A CYP314A1 protein modified at the C-terminus by the addition of three copies of the hemaglutinin epitope was targeted to mitochondria of S2 cells (Petryk *et al*., 2003). This study represents a breakthrough in our understanding of E20MO molecular genetics. *Cyp314a1* has a single clear ortholog in *A. gambiae*, but the sequence of both predicted proteins is unusual. They are clearly members of the mitochondrial P450 clade and have several intron positions in common with other mitochondrial P450 genes (Ranson *et al*., 2002a). However, they lack two of the three positively charged residues thought to confer high affinity to adrenodoxin (Pikuleva *et al*., 1999). Their exact N-terminal sequence is also somewhat unclear in the absence of EST sequences or proteomic data in either species to confirm the annotation prediction of the N-terminus.

Regulation of E20MO activity, developmental changes, induction, and inhibition is discussed by Lafont *et al*. in **Chapter 3.3**.

4.1.4.1.1.2. Ecdysone biosynthesis: biochemistry The hydroxylation of 2-deoxyecdysone to ecdysone in *Locusta migratoria* was observed in the prothoracic glands as expected but in the Malpighian tubules, midgut, fat body, and epidermal tissues as well (Kappler *et al*., 1986). The C-2 hydroxylation was characterized as a mitochondrial P450 activity in larval Malpighian tubules and in ovarian follicle cells of vitellogenic females (Kappler *et al*., 1986) as well as in the prothoracic glands (Kappler *et al*., 1988). This P450 activity is peculiar in its very low sensitivity to CO inhibition, but virtually stoichiometric incorporation of one atom of molecular oxygen was demonstrated (Kabbouh *et al*., 1987). The biochemical characterization of this P450 activity is supportive of the idea that the same gene product is responsible for C-2 hydroxylation in the Malphighian tubules and in the prothoracic glands. Two further hydroxylase activities have been characterized in *L. migratoria* prothoracic glands as typical P450 enzymes (Kappler *et al*., 1988). With 2,22, 25-trideoxyecdysone as substrate for C-25 hydroxylation and 2,22-dideoxyecdysone as substrate for C-22 hydroxylation, these two activities were traced to the microsomal and mitochondrial fractions, respectively (Kappler *et al*., 1988). The specificity of the three enzymes is suggested by their low Km (0.5–2.5 μM) but significant competitive inhibition of the C-2- and C-22-hydroxylations by several ecdysteroids indicates that their place and substrate(s) in a grid or in a linear pathway is still conjectural, as noted by Rees (1995). Evidence towards the P450 nature of the C-25, C-22, and C-2 hydroxylases and their subcellular localization in *M. sexta* prothoracic glands has been presented (Grieneisen *et al*., 1993). That study also suggested that the 7,8-dehydrogenation of cholesterol was a microsomal, NADPH-dependent enzyme. Inhibition of the reaction in "mildly disrupted" prothoracic glands by fenarimol (EC50 = 0.1 mM) and carbon monoxide (63% inhibition at $CO:O_2$ of 19 : 1) was taken as evidence for P450 involvement. There is to date no convincing evidence for the involvement of P450 enzymes in the biosynthetic steps that occur in Dennis Horn's famed "black box" between the 7,8-dehydrogenation and the ultimate hydroxylations.

Furthermore, there is equally little biochemical information on the enzymes responsible for the dealkylation of phytosterols to cholesterol. The reactions involved (desaturation, epoxidation, C–C bond cleavage) have been elegantly described in terms of their chemistry (Ikekawa *et al*., 1993). They are well within the catalytic competence of P450 enzymes, and whether these dealkylation reactions are catalyzed by P450 enzymes or not, this remains a challenging area of research in view of their central role in the nutritional physiology of phytophagous species.

4.1.4.1.1.3. Ecdysone biosynthesis: molecular genetics The study of *Drosophila* Halloween mutants that identified CYP314A1 as an E20MO also identified the C-2 and the C-22 hydroxylase, and may identify further P450 enzymes of the

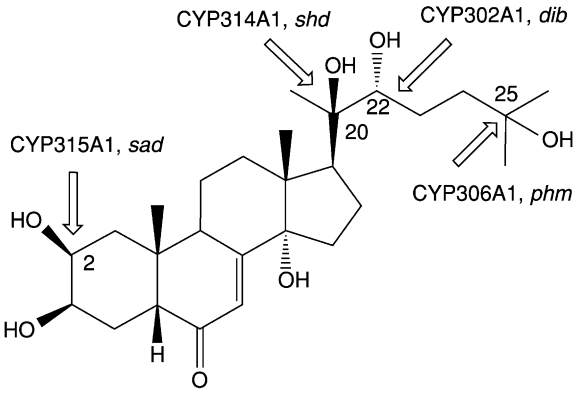


Figure 18 The insect molting hormone 20-hydroxyecdysone with the sites of P450 hydroxylation of its precursors identified to date in *Drosophila melanogaster*. P450 and gene names are shown.

ecdysteroid pathway (**Figure 18**). The *disembodied* (*dib*) gene encodes CYP302A1 identified as the C22 hydroxylase (Chavez *et al.*, 2000; Warren *et al.*, 2002). Similarly, the *shadow* (*sad*) gene encodes CYP315A1, the C2 hydroxylase (Warren *et al.*, 2002). The *dib* gene was identified by classical molecular genetics approaches starting from the genetic locus, ultimately leading to a mitochondrial P450 sequence (Chavez *et al.*, 2000). The identification of the *sad* gene (as that of *shd*, see above) exploited the match between genetic map and the sequence from the *Drosophila* genome project (Warren *et al.*, 2002). In all three cases, mutations leading to stop codons in the sequence confirmed the identification of the gene. Both *dib* and *sad* genes are expressed in embryos, as well as in the larval ring gland and follicle cells of adult ovaries. Transient expression in S2 cells coupled with the use of radiolabeled substrates, 2-deoxyecdysone and 2,22-dideoxyecdysone, was used to identify the reaction catalyzed (Warren *et al.*, 2002). Additional mutants such as *phantom* (*phm*) and *spook* (*spo*) that have a similar phenotype are thought to encode additional P450s involved in ecdysone biosynthesis, CYP306A1, a 25-hydroxylase, and CYP307A1, respectively (J.T. Warren *et al.*, personal communication). The *without children* (*woc*) mutant has identified a transcription factor that controls the 7,8-dehydrogenation of cholesterol or 25-hydroxycholesterol (Warren *et al.*, 2001) and may help identify the enzyme involved in this early step of ecdysone biosynthesis. Little is known of P450 expression in ecdysteroidogenic tissues of other insects. Interestingly, an RT-PCR fragment of a CYP4G-like transcript (GenBank AY635178) was isolated from *Manduca sexta* prothoracic glands. Northern hybridization shows high expression of a 2.3-kb message between days 2 and 5 of the last larval instar (before and during the first peak of ecdysteroid synthesis), with no detectable expression in fat body, midgut, nerve cord, or ovary (M.J. Snyder, V.M. Guzov, J.L. Stevens, and R. Feyereisen, unpublished data). The function of this member of the CYP4 family in prothoracic glands is unknown, but may be related to that of crayfish CYP4C15 (Aragon *et al.*, 2002) or *Drosophila* CYP4G15 (Maibeche-Coisne *et al.*, 2000).

4.1.4.1.1.4. Ecdysteroid catabolism The C-26 hydroxylation of 20-hydroxyecdysone has been characterized as a typical microsomal P450 activity in an epithelial cell line of *Chironomus tentans* (Kayser *et al.*, 1997). However, a dual localization (microsomal and mitochondrial) was reported for ecdysteroid 26-hydroxylase in *M. sexta* midgut (Williams *et al.*, 1997, 2000b). The reaction has a low Km for 20-hydroxyecdysone (~1 μM). The *C. tentans* cell line metabolizes 20, 26-dihydroxyecdysone further to two less polar metabolites produced in a constant 3 : 1 ratio. This metabolism is NADPH-dependent and inhibited by azole compounds. The two metabolites are diastereomers of a cyclic hemiacetal formed (nonenzymatically) by the reaction of the C-22 hydroxyl group with a C-26 aldehyde (Kayser *et al.*, 2002). It is not clearly established whether the P450 conversion of the C-26 hydroxyl to the C-26 aldehyde is carried out by the same enzyme that initially hydroxylates 20-hydroxyecdysone. The C-26 aldehyde is a presumed intermediate in the conversion of C-26 ecdysteroids to C-26 ecdysonoic acids, a common inactivation product of ecdysteroids, but phosphate conjugation of the C-26 hydroxyl is observed in some insects (Rees, 1995). P450-mediated hydroxylations followed by oxidations to the carboxylic acid are known, e.g. CYP701A3 of *A. thaliana* converting *ent*-kaurene to *ent*-kaurenoic acid or CYP27 in the biosynthesis of bile acids.

Side-chain cleavage of ecdysteroids has been reported as an inactivation route, but despite the analogy to the reaction catalyzed by vertebrate P450scc (CYP11A), i.e., the C–C bond cleavage at a vicinal C-20,C-22 diol, there is currently no information on the enzymology of this reaction. P450scc is a unique enzyme in that it catalyzes not just the C–C bond cleavage, but also the two preceding hydroxylations of cholesterol at C-20 and C-22 (in that order).

4.1.4.1.2. P450 and juvenile hormone metabolism (see also Chapter 3.7) The landmark chemical feature of juvenile hormones (JHs) is the presence of the epoxide group, and by 1985 it was already well

established that epoxidation of the JH precursors in the corpora allata (CA) was catalyzed by a P450 enzyme. The biochemical evidence was presented in two studies on the allatal enzyme, one with *Blaberus giganteus* CA homogenates (Hammock, 1975) and the other with *L. migratoria* CA (Feyereisen *et al*., 1981) that established its microsomal nature. A number of compounds that took advantage of the P450 nature of the epoxidase have been tested as inhibitors of juvenile hormone biosynthesis (Hammock and Mumby, 1978; Brooks *et al*., 1985; Pratt *et al*., 1990; Unnithan *et al*., 1995). The allatal epoxidase was viewed as an attractive new target for "biorational" insecticides, because it is an insect-specific target, and because its inhibition should lead to desirable effects – precocious metamorphosis and adult sterility akin to the effects of precocenes (Bowers *et al*., 1976). A serious pursuit of this goal would only be possible with a molecular characterization of the epoxidase and its heterologous expression that would permit screening at a scale not possible by the laborious dissection of the glands. Photoaffinity labeling was first developed as a technique that could facilitate purification and subsequent cloning of the epoxidase. Bifunctional compounds with a substituted imidazole to coordinate the heme iron and a diazirine or a benzophenone group to label the substrate binding site were synthesized and tested (Andersen *et al*., 1995). These compounds were potent inhibitors of methyl farnesoate epoxidation to JH III in the CA of *D. punctata*, and selectively labeled a 55 kDa protein.

4.1.4.1.2.1. CYP15A1 – the epoxidase However, eventual cloning of the epoxidase was achieved by a less subtle route, the sequencing of ESTs from *D. punctata* CA. Sequencing of >900 ESTs from the CA of vitellogenic females yielded three ESTs matching a P450 sequence by BLAST analysis. A full-length cDNA of this P450 was expressed in *E. coli* (Helvig *et al*., 2004). This P450 termed CYP15A1, has a high affinity for methyl farnesoate, showing a type I spectrum with a Ks of 6 μM. The enzyme, when reconstituted with fly P450 reductase, catalyzed the NADPH-dependent epoxidation of 2*E*, 6*E*-methyl farnesoate to JH III. The epoxidation is highly stereoselective (98 : 2) to the natural 10*R* enantiomer over its diastereomer. The enzyme also has a high substrate specificity, epoxidizing the 2*E*,6*E* isomer preferentially over the 2*Z*,6*E* isomer, and accepting no other substrate tested including farnesoic acid. CYP15A1 is expressed selectively in the CA, and the rank order of inhibition of CYP15A1 activity by substituted imidazoles is identical to that of JH biosynthesis by isolated CA (Helvig *et al*., 2004).

4.1.4.1.2.2. CYP4C7 – ω-hydroxylase Expression of the *CYP15A1* gene is high when JH synthetic levels are high, but another P450 gene, *CYP4C7*, also selectively expressed in the CA of *D. punctata* (Sutherland *et al*., 1998), has an expression pattern that mirrors the pattern of JH synthesis. The recombinant CYP4C7 enzyme produced in *E. coli* metabolized a variety of sesquiterpenoids but not mono- or diterpenes. In addition to metabolizing JH precursors, farnesol, farnesal, farnesoic acid, and methyl farnesoate, it also metabolized JH III to a major metabolite identified as (10*E*)-12-hydroxy-JH III (Sutherland *et al*., 1998). Although this ω-hydroxylated JH III has not been identified as a product of the CA, *L. migratoria* CA are known to produce several hydroxylated JHs in the radiochemical assay *in vitro* (Darrouzet *et al*., 1997; Mauchamp *et al*., 1999). These hydroxy-JHs, 8′-OH-, 12′-OH (Darrouzet *et al*., 1997), and 4′-OH-JH III (Mauchamp *et al*., 1999) may be major products of the CA after JH III itself and their role is unknown. Their presence suggests that locust CA have a P450 homologous to cockroach CYP4C7 that has a lower regioselectivity. Indeed, the hydroxy-JHs can be synthesized by locust CA from JH III, and this hydroxylation is inhibited by CO and piperonyl butoxide (Couillaud *et al*., 1996).

The tight physiological regulation of *CYP4C7* expression in adult female *D. punctata* (Sutherland *et al*., 2000) indicates that the terpenoid ω-hydroxylase has an important function to play at the end of vitellogenesis and at the time of impending chorionation and ovulation. It was hypothesized (Sutherland *et al*., 1998, 2000) that this hydroxylation was a first step in the inactivation of the very large amounts of JH and JH precursors present in the CA of this species after the peak of JH synthesis.

The study of JH metabolism has long been dominated by esterases and epoxide hydrolases (see **Chapter 3.7**), but early work on insecticide-resistant strains of the housefly revealed oxidative metabolism as well (review: Hammock, 1985). Evidence that housefly CYP6A1 efficiently metabolizes sesquiterpenoids, including JH to its 6,7-epoxide, confirms these early studies (Andersen *et al*., 1997). Hydroxylation and epoxidation of JHs are thus confirmed P450-mediated metabolic pathways for these hormones.

A comparison of four P450 enzymes that metabolize methyl farnesoate (**Table 3**) shows their

Table 3 Specificity of insect P450 enzymes towards methyl farnesoate (MF)

P450	*Substrate*	*Reaction*	*Product formed*	*Reference*
CYP15A1	2*E*, 6*E*-MF	Epoxidation	10*R*-epoxide	Helvig *et al.* (2004)
CYP4C7	Sesquiterpenoids	ω-Hydroxylation	12*E*-OH	Sutherland *et al.* (1998)
CYP9E1	2*Z* > 2*E*-MF (1.5 : 1)	Epoxidation	10-Epoxide 10R : 10S (1 : 1)	[a]
CYP6A1	MF or JH all 4 isomers	Epoxidation	6 or 10-epoxide 10S : 10R (3 : 1)	Andersen *et al.* (1997)

[a]J.F. Andersen, G.C. Unnithan, J.F. Koener, and R. Feyereisen, unpublished data. Data from Helvig, C., Koener, J.F., Unnithan, G.C., Feyereisen, R., **2004**. CYP15A1, the cytochrome P450 that catalyzes epoxidation of methyl farnesoate to juvenile hormone III in cockroach corpora allata. *Proc. Natl Acad. Sci. USA 101*, 4024–4029; © National Academy of Sciences, USA.

catalytic versatility. One is extremely substrate specific (CYP15A1), another is not (CYP6A1). Three of them are epoxidases, one is a hydroxylase on this substrate (CYP6A1 also has activity as a hydroxylase – but not on this substrate). One is a stereoselective epoxidase (CYP15A1), another lacks product enantioselectivity (CYP9E1).

4.1.4.1.3. Biosynthesis of long-chain hydrocarbons The cuticle of insects can be characterized by its hydrocarbon composition, and this can serve as a subtle tool in chemical taxonomy (Lockey, 1991). In some insects, cuticular or exocrine alkanes and alkenes can serve as allomones, or even pheromones (see **Chapter 3.14**). To cite but one example, the Dufour gland secretions of the leaf-cutting ant *Atta laevigata* are deposited on foraging trails. This trail pheromone comprises n-heptadecane, (*Z*)-9-nonadecene, 8,11-nonadecadiene, and (*Z*)-9-tricosene (Salzemann *et al.*, 1992). Despite their apparently simple structure, the biosynthesis of these hydrocarbons is complex and the enzymes involved are still under intense study. Schematically (**Figure 19**), long chain (18–28 carbons) fatty acid CoA esters are first reduced to their aldehyde by an acyl-CoA reductase, and they are subsequently shortened to form C_{n-1} hydrocarbons (Reed *et al.*, 1994). The conversion of (*Z*)-tetracosenoyl-CoA (from a C24 : 1 fatty acid) to (*Z*)-9-tricosene (a C23 : 1 alkene) has been characterized in housefly epidermal microsomes. The role of the aldehyde tetracosenal as an intermediate is evidenced by trapping experiments with hydroxylamine. The next step is NADPH- and O_2-dependent, and is truly an oxidative decarbonylation reaction, which releases the terminal carbon as CO_2 and not as CO (Reed *et al.*, 1994). Inhibition by carbon monoxide and anti-P450 reductase antibodies are strongly indicative of a P450 reaction, and this most peculiar enzyme was called P450hyd (Reed *et al.*, 1994). The reaction does not involve a terminal desaturation, because both C-2 protons of the aldehyde are retained in the hydrocarbon product, and the aldehydic proton is transfered to the product (Reed *et al.*, 1995). NADPH and molecular oxygen can be replaced by "peroxide shunt" donors, suggesting that this reaction involves a perferryl iron-oxene species of activated oxygen (Ortiz de Montellano, 1995a), and that it proceeds through an alkyl radical intermediate (Reed *et al.*, 1995). This type of oxidative decarbonylation reaction may be widespread in insects because a mechanistically identical conversion of octadecanal to heptadecane has been documented in flies, cockroaches, crickets, and termites (Mpuru *et al.*, 1996).

4.1.4.1.4. Pheromone metabolism In female houseflies, the (*Z*)-9-tricosene produced by P450hyd is a major component of the sex pheromone, being responsible for inducing the courtship ritual and the males' striking activity (see **Chapter 3.14**). Additional components of the pheromone are the sex recognition factors (*Z*)-9,10-epoxytricosane and (*Z*)-14-tricosen-10-one. These compounds are obviously derived from (*Z*)-9-tricosene by oxidative metabolism (Blomquist *et al.*, 1984). A microsomal P450 was shown to oxidize the alkene from either side at a distance of 9/10 carbons in-chain (Ahmad *et al.*, 1987), but the structural requirements of the enzyme for its substrate are otherwise strict (Latli and Prestwich, 1991). This P450 activity is found in various tissues of the male and female, including the epidermis and fat body, but the highest specific activity is found in male antennae (Ahmad *et al.*, 1987; **Chapter 3.15**). The C23 epoxide and ketone are absent internally in females, but accumulate on the surface of the cuticle (Mpuru *et al.*, 2001), suggesting the localization of this P450 activity in epidermal cells of female flies.

More generally, P450 enzymes are probably involved in the biosynthesis of many insect pheromones and allomones, e.g., epoxides of polyunsaturated hydrocarbons in arctiid moths, disparlure of the gypsy moth (Brattsten, 1979a), or monoterpenes

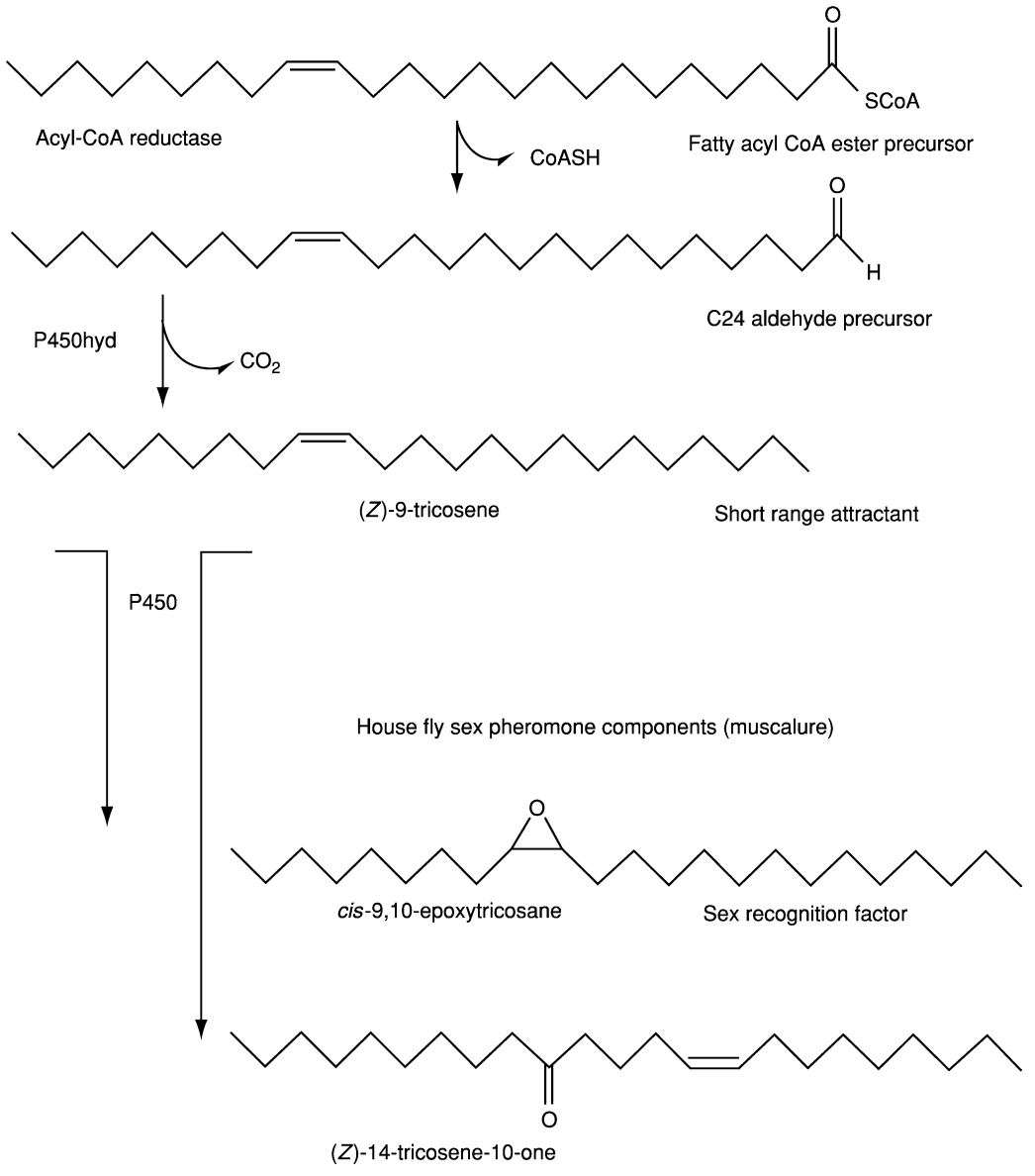


Figure 19 Biosynthesis of (*Z*)-9-tricosene and components of the housefly sex pheromone. P450hyd acts as a decarbonylase to produce (*Z*)-9-tricosene from an aldehyde precursor. Another P450 attacks (*Z*)-9-tricosene from either side to give the two components of muscalure.

in bark beetles (Brattsten, 1979a; White *et al*., 1979; Hunt and Smirle, 1988). In the honeybee, caste-specific ω and ω-1 hydroxylations of fatty acids to mandibular pheromones (see **Chapter 3.13**) have all the characteristics of P450 reactions (Plettner *et al*., 1998). These P450s involved in pheromone biosynthesis are likely to be exquisitely specific.

Evidence for pheromone catabolism by P450 enzymes is also accumulating. As shown for the metabolism of (*Z*)-9-tricosene in the housefly, a distinction between biosynthesis and catabolism can be purely semantic in the case of a biogenetic succession of chemicals that have different signaling functions. Compound B that is a metabolite of compound A may have less or none of compound A's activity, but may have its own specific biological activity. Nonetheless, pheromones as signal molecules need to be metabolically inactivated and this catabolism may occur in the antennae themselves. Several P450 mRNAs have been identified in insect antennae. In *Drosophila*, a partial P450 cDNA, along with a UDP-glycosyltransferase and a short chain dehydrogenase/reductase, were found by northern blot analysis to be preferentially expressed in the third antennal segments, with lower expression in legs (Wang *et al*., 1999). This P450 cDNA

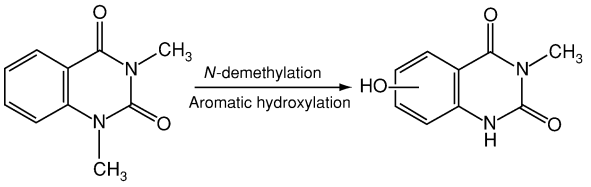


Figure 20 P450-dependent *N*-demethylation and ring hydroxylation of the *Phyllopertha diversa* sex pheromone by antennal microsomes of male scarab beetles. See text and Wojtasek and Leal (1999) for details.

corresponds to *Cyp6w1* for which many ESTs have been identified in a head cDNA library. In *Mamestra brassicae*, two P450 cDNAs were cloned from antennae, CYP4L4 and CYP4S4 (Maibeche-Coisne *et al*., 2002). Both genes are strongly expressed in the sensilla trichodea as shown by *in situ* hybridization. Whereas *CYP4S4* expression is restricted to the antennae, *CYP4L4* is also expressed in proboscis and legs. Five P450 ESTs representing three P450 genes were found in a small EST project using male *Manduca sexta* antennae (Robertson *et al*., 1999), which is further evidence that antennae may harbor several P450 enzymes. High levels of P450 reductase expression have been noted in *Drosophila* antennae (Hovemann *et al*., 1997), and adrenodoxin reductase is also expressed in antennae (Freeman *et al*., 1999). Antennal P450s are therefore an active enzyme system, complete with their redox partners. The physiological role of the P450s thus found in antennae is still formally unknown, but the presence of other enzymes generally associated with detoxification processes (Robertson *et al*., 1999; Wang *et al*., 1999) are indeed suggestive of a large category of odorant-metabolizing enzymes. In addition to the presence of P450 transcripts in antennae, there is at least one report where a P450 reaction has been characterized in antennae, but the CYP gene coding for the P450 catalyzing this reaction is still unknown. Antennal microsomes of the pale brown chafer, *Phyllopertha diversa*, metabolize the alkaloid sex pheromone by a P450 enzyme (Wojtasek and Leal, 1999). This enzyme is specifically produced in male antennae, and its activity is not detected in other scarab beetles or lepidopteran species (**Figure 20**). P450 enzymes may have found in insects a role as odorant degrading enzymes (ODE) along with esterases, aldehyde oxidases, glucosyl-transferases, etc. (review: **Chapter 3.15**).

4.1.4.1.5. Metabolism of fatty acids and related compounds P450-catalyzed fatty acid ω-hydroxylation has been reported in insects (Feyereisen and Durst, 1980; Ronis *et al*., 1988; Clarke *et al*., 1989; Cuany *et al*., 1990; Rose *et al*., 1991). Clofibrate selectively induces ω-hydroxylation (and not ω-1) of lauric acid by housefly and *Drosophila* microsomes, just as it induces the rat CYP4A1 ω-hydroxylase (Clarke *et al*., 1989; Amichot *et al*., 1998). No clear peroxisome proliferator activated receptor (PPAR, Issemann and Green, 1990) gene ortholog but several paralogs are present in the *Drosophila* genome. In contrast to clofibrate, phenobarbital induces *Drosophila* ω, ω-1, and ω-2 hydroxylations (Amichot *et al*., 1998), indicating the presence of several fatty acid hydroxylases. By homology to vertebrate CYP4 enzymes it is thought that cockroach CYP4C1 has a ω-hydroxylase function as well, but evidence is lacking. CYP4C7, CYP6A1, and CYP12A1 lack lauric acid hydroxylase activity (Andersen *et al*., 1997; Guzov *et al*., 1998; Sutherland *et al*., 1998), but CYP6A8 has laurate ω-1 hydroxylase activity (C. Helvig, personal communication). Although the role of vertebrate P450 enzymes in the metabolism of arachidonic acid and eicosanoids is well established (Capdevila *et al*., 2002), there is currently no indication for a similar function of P450 in insects.

The biosynthesis of volicitin, an elicitor of plant volatile production found in the oral secretions of caterpillars occurs in the insect from the plant-derived fatty acid linolenic acid (Pare *et al*., 1998). This biosynthesis involves C-17 hydroxylation and glutamine conjugation. It is likely that this hydroxylation will be shown to be catalyzed by a P450 enzyme.

4.1.4.1.6. Defensive compounds The tremendous variety of chemicals used for defense of insects against predation is well documented, but the enzymes involved in their *de novo* synthesis or in their transformation from ingested precursors are less studied. It is very likely that P450 enzymes may have found there a fertile ground for their chemical prowess. Just two examples are described.

The biosynthesis of cyanogenic compounds found as defensive compounds in many insect species (Nahrstedt, 1988) may well involve P450 enzymes. For instance, the cyanogenic glucosides linamarin and lotaustralin found in *Heliconius* butterflies and *Zygaena* moths, are clearly derived from the amino acids valine and isoleucine. The pathway probably involves N-hydroxylation of the amino acids, further metabolism to the aldoximes and nitriles, and final C-hydroxylation before conjugation to the glucoside (**Figure 21**). The aldoximes and nitriles are efficiently incorporated *in vivo* (Davis and Nahrstedt, 1987; Holzkamp and Nahrstedt, 1994).

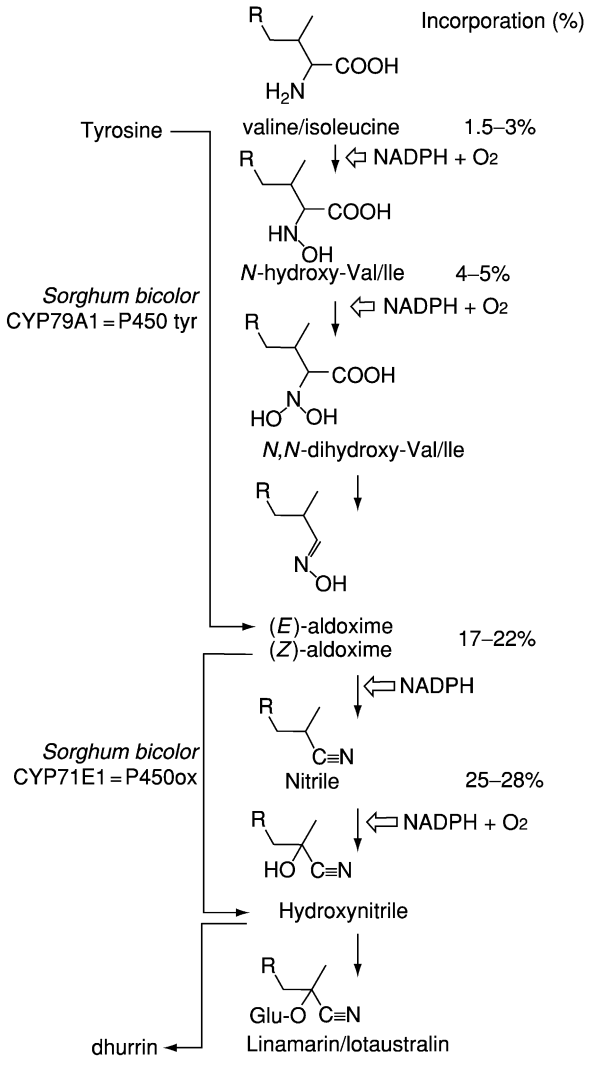


Figure 21 Biosynthesis of linamarin and lotaustralin from valine (R=H) and isoleucine ($R=CH_3$), indicating possible sites of P450 metabolism. The efficiency of incorporation of intermediates *in vivo* is shown on the right. The homologous reactions catalyzed by two multifunctional Sorghum P450 enzymes converting tyrosine to dhurrin are shown on the left.

The pathway resembles that found in plants, where two multifunctional P450 enzymes are sufficient to convert the amino acid to the hydroxynitrile substrate of the conjugating enzyme (**Figure 21**). For dhurrin biosynthesis from tyrosine in sorghum, these are CYP79A1 and CYP71E1 (Kahn *et al.*, 1997). Whereas the plant pathway appears to "channel" the substrates through the two P450s with little escape of the aldoxime intermediate, nothing is known of the number and functioning of the cyanogenic pathway enzymes used by Lepidoptera. Characterization of the P450 enzymes involved should allow a comparison of the plant and insect solutions to this biosynthetic challenge.

Another comparison of interest will be that of enzymes involved in the biosynthesis of insect defensive steroids with the well characterized steroid metabolizing enzymes of vertebrates. A number of aquatic Coleoptera (Dysticidae) and Hemiptera (Belastomatidae) synthesize a variety of steroids, mostly pregnanes (Scrimshaw and Kerfoot, 1987). With cholesterol as the presumed precursor in these carnivorous insects, one may envisage the evolution of an insect side-chain cleavage enzyme, of a C-21 hydroxylase and of a C17-C21 lyase that would catalyze reactions identical to those of CYP11A, CYP21, and CYP17. Some defensive steroids also have an aromatic A-ring, 7α, or 15α hydroxyl groups. A pregnene-3β, 20β-diol glucoside is synthesized from cholesterol in female pupae of *M. sexta* (Thompson *et al.*, 1985). The role of this compound is unknown, but its synthesis strongly suggests the existence of a C20-C22 side chain cleavage enzyme in Lepidoptera as well.

4.1.4.2. Xenobiotic Metabolism: Activation and Inactivation

4.1.4.2.1. Natural products The metabolism of plant toxins by insects has been reviewed extensively (see, e.g., Brattsten, 1979b; Dowd *et al.*, 1983; Ahmad, 1986; Ahmad *et al.*, 1986; Mullin, 1986; Yu, 1986). Relatively a few studies have directly assessed the role of P450 enzymes in the metabolism of natural compounds by more than one or two criteria such as microsomal localization, NADPH and O_2 dependence, and inhibition by piperonyl butoxide. In most cases, metabolism is associated with detoxification, e.g., the metabolism of xanthotoxin in *Papilio polyxenes*, *S. frugiperda*, and *Depressaria pastinacella* (Bull *et al.*, 1986; Nitao, 1990), the metabolism of α-terthienyl in larvae of three lepidopteran species (Iyengar *et al.*, 1990), and the metabolism of nicotine in *Manduca sexta* larvae (Snyder *et al.*, 1993). Alpha- and beta-thujones are detoxified by P450 as evidenced by synergism of their toxicity by three P450 inhibitors in *Drosophila* and by the lower toxicity of six of their metabolites (Hold *et al.*, 2001). Studies with flavone (Wheeler *et al.*, 1993), monoterpenes (Harwood *et al.*, 1990) and the alkaloid carnegine (Danielson *et al.*, 1995) show clearly however that evidence for metabolism by P450 *in vitro* may not be sufficient to define an *in vivo* toxicological outcome.

Natural products in the diet can act as inducers of P450 as well of other enzymes (e.g., glutathione S-transferases) and as a result of this induction (see below), the metabolism of the inducing compound or of coingested plant compounds can change dramatically over time (Brattsten *et al.*, 1977).

Differences between acute and chronic toxicity are thus often the result of altered expression patterns (quantitative and qualitative) of P450 genes. This was demonstrated, for instance, in studies on *Spodoptera* larvae (Brattsten, 1983; Gunderson *et al.*, 1986). The monoterpene pulegone and its metabolite menthofuran are more acutely toxic to *S. eridania* than to *S. frugiperda*, but the reverse is true for chronic toxicity.

A study by Yu (1987) compared the metabolism of a large number of plant chemicals of different chemical classes by *S. frugiperda* and *Anticarsia gemmatalis* (velvetbean caterpillar) microsomes. Two indirect methods were used, on the one hand, the NADPH-dependent decrease in substrate and on the other hand, the substrate-induced NADPH oxidation. This metabolism is inhibited by piperonyl butoxide and by carbon monoxide, and induced by a number of chemicals, particularly indole-3-carbinol, strongly suggesting P450 involvement. Such indirect methods are very useful as screening tools, as a first step towards a more thorough characterization of metabolism. However, they give no qualitative indication of the chemical fate of the substrate, nor quantitative indication of the levels of metabolism, as NADPH consumption is correlated to the coupling rate of the reaction, rather than to the rate of product formation (see Section 4.1.3.4.1). Clearly, insect P450 enzymes as a whole are capable of metabolizing a tremendous variety of naturally occurring chemicals, but the role of individual enzymes and their catalytic competence still needs a better description.

Heterologously expressed P450 enzymes of the CYP6B subfamily from *Papilio* species (see Section 4.1.4.3 and **Table 1**) are well characterized for their ability to metabolize furanocoumarins (Hung *et al.*, 1997; Li *et al.*, 2003; Wen *et al.*, 2003). They fit the description of enzymes with "broad and overlapping specificity" towards these compounds. Their range of catalytic competence is quite variable. For instance, CYP6B21 and CYPB25 metabolize the angular furanocoumarin angelicin at a similar rate (0.4–0.5 nmol/min/nmol P450), but whereas CYP6B21 also metabolizes 7-ethoxycoumarin at a similar rate (0.5 nmol/min/nmol P450), CYP6B25 does not have appreciable 7-dealkylation activity (Li *et al.*, 2003).

Natural products are not just an endless catalog of P450 substrates and inducers, but they also comprise a varied and complex set of inhibitors of P450 enzymes. These inhibitors range from "classical" reversible inhibitors to substrates that are activated to chemically reactive, cytotoxic forms (e.g., Neal and Wu, 1994).

4.1.4.2.2. Insecticides and other xenobiotics The metabolism of insecticides by P450 enzymes is very often a key factor in determining toxicity to insects and to nontarget species, but it can also represent a key step in the chain of events between contact, penetration, and interaction at the target site. The classical example is probably the metabolism of phosphorothioate insecticides. In many cases, the active ingredients of organophosphorus insecticides are phosphorothioate (P=S) compounds (a.k.a. phosphorothionates), whereas the molecule active at the acetylcholinesterase target site is the corresponding phosphate (P=O). It has long been recognized that the P=S to P=O conversion is a P450-dependent reaction. In the case of diazinon, this desulfuration has been studied for three heterologously expressed insect P450 enzymes (Dunkov *et al.*, 1997; Guzov *et al.*, 1998; Sabourault *et al.*, 2001). All three P450s metabolized diazinon not just to diazoxon, the metabolite resulting from desulfuration, but also to a second metabolite resulting from oxidative ester cleavage. Similarly, antibodies to housefly CYP6D1 inhibit the microsomal desulfuration of chlorpyriphos as well as its oxidative ester cleavage (Hatano and Scott, 1993). The mechanism of P450-dependent desulfuration is believed to involve the initial attack of the P=S bond by an activated oxygen species of P450, leading to an unstable and therefore hypothetical phosphooxythiirane product (**Figure 22**). The collapse of this product can lead to two possible outcomes: (1) the replacement of sulfur by oxygen in the organophosphate product with the release of a reactive form of sulfur; and (2) the cleavage of the phosphate ester (or thioester) link with the substituent of highest electron-withdrawing properties, the "leaving group."

Outcome (1) can be viewed as "activation" because the P=S to P=O desulfuration produces an inhibitor of acetylcholinesterase often several orders of magnitude more potent than the P=S parent compound. However, the fate of this product of "activation" depends on the histological proximity to the target, sequestration, excretion, and further metabolism of the phosphate (by oxidative or hydrolytic enzymes). Kinetic evidence with the heterologously expressed CYP6A1, CYP6A2, and CYP12A1, as well as immunological evidence with CYP6D1 indicate that these P450 enzymes do not metabolize the P=O product of the parent P=S compound they metabolize.

Outcome (2) or "dearylation" is without question a detoxification because the oxidative cleavage of the "leaving group" yields compounds unable to inhibit acetylcholinesterase, the dialkylphosphorothioate,

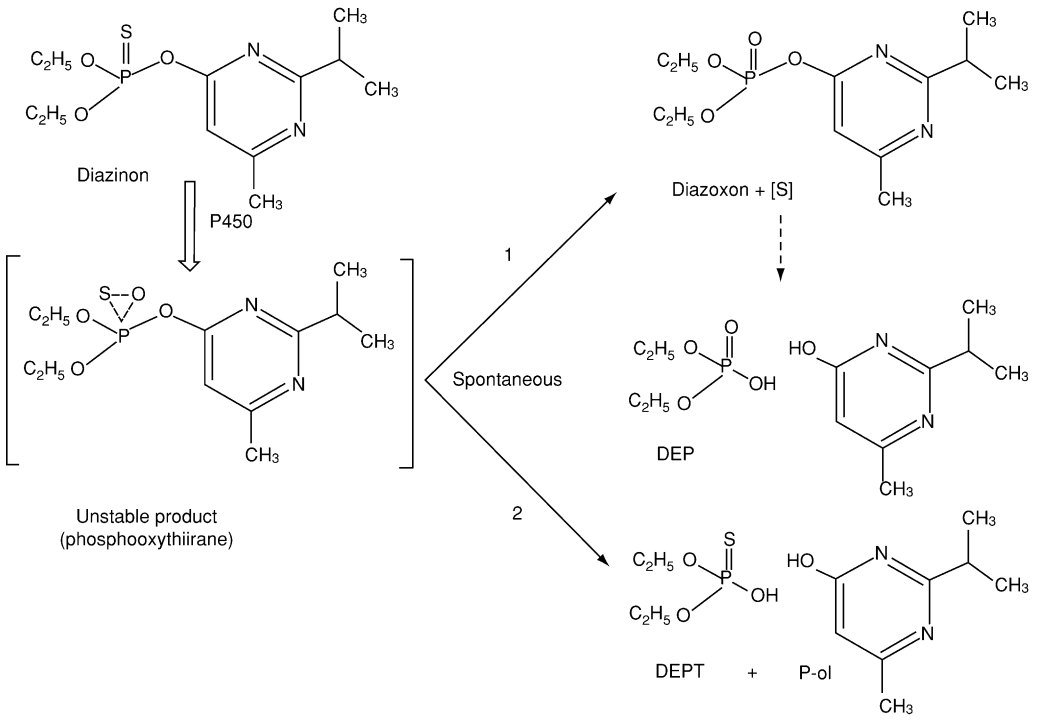


Figure 22 Metabolism of diazinon by cytochrome P450. Following an insertion of oxygen into the substrate, a reactive intermediate collapses (1) by desulfuration or (2) by cleavage of the ester linkage (dearylation). DEP: diethylphosphate, DEPT: diethylphosphorothioate, P-ol: 2-isopropopoxy-4-methyl-6-hydroxypyrimidine, [S]: reactive form of sulfur released during the reaction. Diazoxon can be further converted to DEP and P-ol by the same or another P450 in a subsequent reaction, or by a phosphotriester hydrolase. DEP may also be formed by spontaneous degradation of the initial product of diazinon monooxygenation. The ratio of outcomes 1 and 2 and the fate of the reactive sulfur depends on the P450 enzyme and on the type of OP substrate (see **Table 4**).

and/or dialkylphosphate. The ratio of outcome (1) and (2) appears P450-specific and substrate specific (**Table 4**) suggesting that the collapse of the unstable initial product of P450 attack is influenced by the active site environment. Theoretically, some P450 enzymes may very strongly favor outcome (1) or (2) or vice versa, thus qualifying as relatively "clean" activators or detoxifiers, but there is to date little direct evidence from the insect toxicological literature for such P450 enzymes (see however Oi *et al.*, 1990).

Furthermore, the sulfur released by the reaction can bind covalently either to neighboring proteins thus leading to cellular damage, or to the P450 protein itself (at least in vertebrate liver where this specific aspect has been studied, Kamataki and Neal, 1976). Parathion causes NADPH-dependent inhibition of methoxyresorufin *O*-demethylation activity (a P450Lpr-selective activity) in the housefly whereas chlorpyriphos does not (Scott *et al.*, 2000). Therefore, it is not just the fate of the initial P450 metabolite of the P=S compound that depends on the P450 and the OP, but it is also the fate of the sulfur released that can vary.

Changes in the level of expression of P450 genes or P450 point mutations may be sufficient to change this delicate balance between activation and inactivation *in vivo*. For instance, fenitrothion resistance in the Akita-f strain of the housefly is related to an increase in oxidative ester cleavage over desulfuration measured in abdominal microsomes (Ugaki *et al.*, 1985). In *H. virescens*, methyl parathion resistance in the NC-86 strain, which has an unchanged level of total P450, is related to a replacement of a set of P450 enzymes with high desulfuration activity by a set of P450 enzymes that metabolize less parathion, and do so with a lower desulfuration/oxidative ester cleavage ratio (Konno and Dauterman, 1989) (see **Table 4**).

P450 enzymes that metabolize OPs can metabolize other insecticides as well and this sometimes leads to potentially useful interactions. Thus, enhanced detoxification of dicofol in spider mites can lead to enhanced chlorpyriphos activation, hence negative cross-resistance (Hatano *et al.*, 1992). Similarly, permethrin resistance in horn flies is suppressible by piperonyl butoxide and negatively related to diazinon toxicity (Cilek *et al.*, 1995). In *H. armigera*

Table 4 Desulfuration and oxidative ester cleavage of organophosphorus insecticides by P450 enzymes and microsomes

P450 enzyme	*Ratio of OP desulfuration/oxidative ester cleavage*	*Reference*
Human CYP2C19	8.5 d, 1.30 p, 0.14 c	Kappers *et al.* (2001), Tang *et al.* (2001), Mutch *et al.* (2003)
Human CYP3A4	3.0 d, 0.50 p, 0.66 c	Kappers *et al.* (2001), Mutch *et al.* (2003)
Human CYP2B6	0.7 d, 0.01 p, 3.38 c	Kappers *et al.* (2001), Mutch *et al.* (2003)
Human liver microsomes	0.29 d, 0.37 p, 0.57c	Kappers *et al.* (2001), Mutch *et al.* (2003)
Housefly CYP6A1	0.37 d	Sabourault *et al.* (2001)
Drosophila CYP6A2	0.92 d	Dunkov *et al.* (1997)
Housefly CYP12A1	0.69 d	Guzov *et al.* (1998)
Housefly CYP6D1	2.0 c	Hatano and Scott (1993)
Housefly *CSMA* microsomes	0.95 f	Ugaki *et al.* (1985)
Housefly *Akita-f*[a] microsomes	0.59 f	Ugaki *et al.* (1985)
Heliothis virescens microsomes	1.90 mp	Konno and Dauterman (1989)
Heliothis virescens NC-86[a] microsomes	1.32 mp	Konno and Dauterman (1989)

[a]Resistant strain.
d, diazinon; p, ethyl parathion; c, chlorpyriphos; f, fenitrothion; mp, methyl parathion.

populations from West Africa, triazophos shows negative cross-resistance with pyrethroids, and in this case the synergism shown by the OP towards the pyrethroids appears to be due to an enhanced activation to the oxon form (Martin *et al.*, 2003). These interactions were observed *in vivo* or with microsomes, but it is likely that they do reflect the properties of single P450 enzymes with broad substrate specificity rather than the fortuitous coordinate regulation of different P450 enzymes with distinct specificities.

Organophosphorus compounds such as disulfoton and fenthion can also be activated by thioether oxidation (formation of sulfoxide and sulfone), but it is not clear whether these reactions are catalyzed in insects by a P450 or by a flavin monooxygenase (FMO). Further examples of oxidative bioactivation of organophosphorus compounds have been discussed (Drabek and Neumann, 1985).

The toxicity of fipronil to house flies is increased sixfold by the synergist piperonyl butoxide, whereas the desulfinyl photodegradation product is not detoxified substantially by P450 (Hainzl and Casida, 1996; Hainzl *et al.*, 1998). Conversion of fipronil to its sulfone appears to be catalyzed by a P450 enzyme in *Ostrinia nubilalis* (Durham *et al.*, 2002) and in *Diabrotica virgifera* (Scharf *et al.*, 2000). In the latter, the toxicity of fipronil sulfone is about the same as that of the parent compound, and piperonyl butoxide has only a marginal effect as synergist. In contrast, synergists antagonize the toxicity of fipronil in *Blattella germanica*, suggesting that oxidation to the sulfone represents an activation step in this species (Valles *et al.*, 1997).

The now banned cyclodiene insecticides aldrin, heptachlor, and isodrin are epoxidized by P450 enzymes to environmentally stable, toxic epoxides, dieldrin, heptachlor epoxide, and endrin (Brooks, 1979; Drabek and Neumann, 1985). Recombinant CYP6A1, CYP6A2, and CYP12A1 can catalyze these epoxidations (see **Table 1**). Examples of proinsecticide metabolism include the activation of chlorfenapyr by *N*-dealkylation (Black *et al.*, 1994) and of diafenthiuron by *S*-oxidation (Kayser and Eilinger, 2001). In each case, the insect P450-dependent activation is a key in the selective toxicity of these proinsecticides that target mitochondrial respiration. Recombinant housefly CYP6A1 catalyzes the activation of chlorfenapyr (V.M. Guzov, M. Kao, B.C. Black, and R. Feyereisen, unpublished data). In *H. virescens*, toxicity of chlorfenapyr is negatively correlated with cypermethrin toxicity (Pimprale *et al.*, 1997). Genetic analysis indicates that a single factor is involved so the same P450 that activates chlorfenapyr may also detoxify cypermethrin in this species (**Figure 23**). A similar case of negative cross-resistance of chlorfenapyr in a pyrethroid-resistant strain has been reported in the hornfly *Haematobia irritans* (Sheppard and Joyce, 1998).

The metabolism of imidacloprid is also of interest in this respect. Although not extensively studied to date, there is evidence that piperonyl butoxide can synergize the toxicity of imidacloprid, but P450-dependent metabolism can also lead to several bioactive metabolites in some insects. How these are further metabolized and how resistance can be caused by P450 attack on this molecule remains unclear (see however Section 4.1.4.5.5).

In vivo synergism by piperonyl butoxide, a typical inhibitor of P450 enzymes (see Section 4.1.4.5.1), is often used to implicate a P450-mediated detoxification, and there are innumerable such examples in

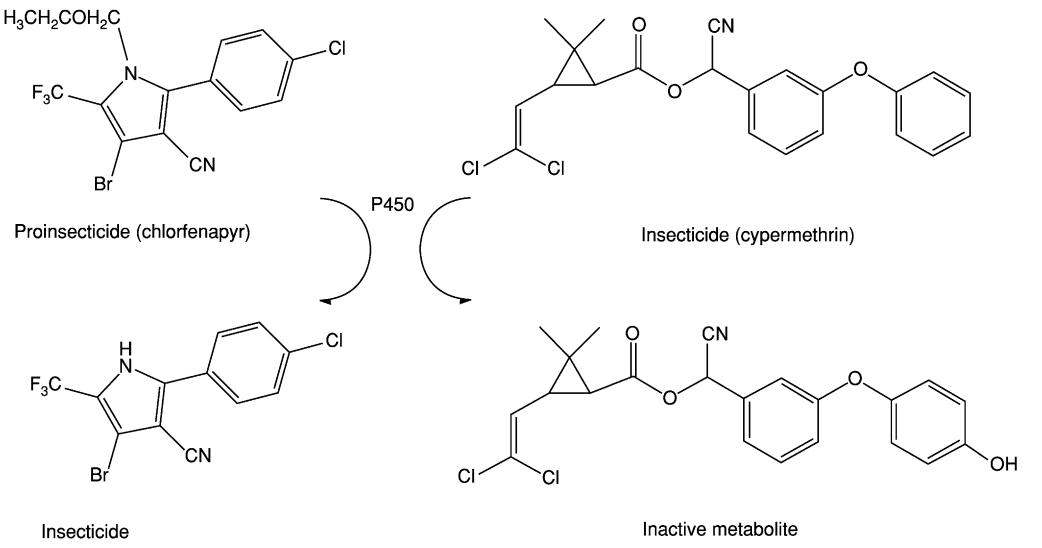


Figure 23 Chlorfenapyr and cypermethrin metabolism. The same P450 in *Heliothis virescens* probably activates the pyrrole and inactivates the pyrethroid, resulting in negative cross-resistance.

the literature. The inference is much stronger when two unrelated synergists are used *in vitro*, and when metabolites of the pesticide are identified. For instance, pyriproxifen is hydroxylated by fat body and midgut microsomes of larval house flies to 4′-OH-pyriproxyfen and 5″-OH-pyriproxyfen and these activities are inhibited by PB and TCPPE (Zhang *et al.*, 1998).

The study of xenobiotic metabolism by individual P450 enzymes expressed in heterologous systems has barely begun (**Table 1**). Whereas the CYP6 enzymes clearly comprise some enzymes with "broad and overlapping" substrate specificity, even closely related enzymes of this family can differ substantially in their catalytic competence. The task of predicting which xenobiotic or natural product will be metabolized by which type of P450 is currently not possible.

4.1.4.3. P450 and Host Plant Specialization

4.1.4.3.1. The Krieger hypothesis and beyond The interactions of plants and insects, and more specifically the role of plant chemistry on the specialization of phytophagous insects have generated a vast literature. "Secondary" plant substances are variously seen to regulate insect behavior and/or to serve as weapons in a coevolutionary "arms race" (Dethier, 1954; Fraenkel, 1959; Ehrlich and Raven, 1964; Jermy, 1984; Bernays and Graham, 1988). In chemical ecology alone, "no other area is quite so rife with theory" (Berenbaum, 1995). Many of the theories and some of the experiments implicitly or explicitly deal with the insect's ability to metabolize plant secondary substances by P450 and other enzymes. In the case of behavioral cues, we are far from understanding the true importance of P450 enzymes in the integration of chemosensory information, e.g., as "odorant degrading enzymes." In the case of detoxification, however, the landmark paper of Krieger *et al.* (1971) can be seen as echoing the Fraenkel (1959) paper, by exposing the *raison d'être* of P450 enzymes. They stated that "higher activities of midgut microsomal oxidase enzymes in polyphagous than in monophagous species indicates that the natural function of these enzymes is to detoxify natural insecticides present in the larval food plants." In that 1971 study, aldrin epoxidation was measured in gut homogenates of last instar larvae from 35 species of Lepidoptera. Polyphagous species had on average a 15 times higher activity than monophagous species. This trend was seen in sucking insects as well, with a 20-fold lower aldrin epoxidase activity in the oleander aphid *Aphis nerii* when compared to the potato aphid *Myzus euphorbiae* or to the green peach aphid *M. persicae* (Mullin, 1986). The former is a specialist feeder on two plant families, Asclepiadaceae and Apocynaceae, whereas the latter two are generalists found on 30–72 plant families. The concept extended to other detoxification enzymes and was broadened to cover prey/predator, e.g., in mites where the predatory mite has a five times lower aldrin epoxidase activity than its herbivorous prey (Mullin *et al.*, 1982). The toxicity of the natural phototoxin α-terthienyl is inversely propotional to the level of its metabolism in Lepidoptera and is related to diet breadth. Metabolism (4.0 nmol/min/nmol P450) is highest in *O. nubilalis* that feeds on numerous

phototoxic Asteraceae, lower in *H. virescens* that has a broad diet, including some Asteraceae that are nonphototoxic, and lowest in *M. sexta*, a specialist of Solanaceae (Iyengar *et al.*, 1990).

The conceptual framework of Krieger *et al.* has been challenged (Gould, 1984) and defended (Ahmad, 1986). An alternative view (Berenbaum *et al.*, 1992) proposes that aldrin epoxidation represents "P450s with broad substrate specificity [that] are most abundant in insects that encounter a wide range of host plant metabolites." A careful repetition of the Krieger experiments on lepidopteran larvae from 58 species of New South Wales failed to show significant differences in aldrin epoxidation between monophagous and polyphagous species (Rose, 1985). High activity in both monophagous and polyphagous species was invariably linked to the presence of monoterpenes in the host diet. The evidence presented in the sections below indicates that polyphagous and oligophagous species alike rely on the ability to draw on a great diversity of P450 genes, encoding a great diversity of specific and less specific enzymes and regulated by a great diversity of environmental sensing mechanisms – induction. The ability to induce P450 enzymes and deal with a wide range of toxic chemicals in the diet has been thought to present a "metabolic load" for polyphagous species, with specialists restricting their "detoxification energy" to one or a few harmful substrates (e.g., Whittaker and Feeny, 1971). However, careful studies in both oligophagous and polyphagous species have refuted this concept of metabolic load (e.g., Neal, 1987; Appel and Martin, 1992).

4.1.4.3.2. Host plant chemistry and herbivore P450 Cactophilic *Drosophila* species from the Sonoran desert are specialized to specific columnar cactus hosts by their dependency on unusual sterols (*D. pachea*) or by their unique ability to detoxify their host's allelochemicals, notably isoquinoline alkaloids and triterpene glycosides (Frank and Fogleman, 1992; Fogleman *et al.*, 1998). P450-mediated detoxification was shown by the loss of larval viability in media that contained both allelochemicals and piperonyl butoxide, and by the induction of total P450 or alkaloid metabolism by the cactus allelochemicals or by phenobarbital (Frank and Fogleman, 1992; Fogleman *et al.*, 1998). Several P450s of the *CYP4*, *CYP6*, *CYP9*, and *CYP28* families are induced by cactus-derived isoquinoline alkaloids and by phenobarbital, but not by triterpene glycosides; only a *CYP9* gene was induced by alkaloids and not by phenobarbital (Danielson *et al.*, 1997, 1998; Fogleman *et al.*, 1998). The capacity to detoxify isoquinoline alkaloids was not related to DDT or propoxur tolerance, and while phenobarbital induced P450s capable of metabolizing the alkaloid carnegine in *D. melanogaster*, this was not sufficient to produce *in vivo* tolerance (Danielson *et al.*, 1995). Selection of *D. melanogaster* with Saguaro alkaloids over 16 generations, however, led to P450-mediated resistance to the cactus alkaloids (Fogleman, 2000). These studies suggest the evolution of specific responses in the cactophilic species involving the recruitment of a phylogenetically unrelated subset of P450 genes in each instance of specialization of a fly species on its host cactus.

The oligophagous tobacco hornworm (*M. sexta*) feeds essentially on Solanaceae and its adaptation to the high levels of insecticidal nicotine found in tobacco depends largely on metabolic detoxification, although other tolerance mechanisms may be contributing as well (Snyder and Glendinning, 1996). Hornworm larvae fed an nicotine-free artificial diet (naive insects) are rapidly poisoned by the ingestion of a nicotine-supplemented diet, but this diet is not deterrent. Poisoning is evidenced by convulsions and inhibition of feeding. The small amount of ingested nicotine induces its own metabolism, so that approximately 36 h later the larvae resume feeding normally, without further signs of poisoning. The inhibition of feeding and its resumption after nicotine exposure is directly related to P450 induction. Indeed, treatment with piperonyl butoxide, which itself has no effect on feeding, inhibits the increase in nicotine-diet consumption that occurs once nicotine metabolism has been induced (Snyder and Glendinning, 1996). Naive insects metabolize nicotine to nicotine 1-N-oxide at a low level, whereas nicotine-fed insects metabolize it further to cotinine-N-oxide at a higher level (Snyder *et al.*, 1994). These reactions are catalyzed by one or more P450 enzymes (Snyder *et al.*, 1993). The effects of nicotine on marker P450 activities are complex: the metabolism of three substrates is induced at low nicotine levels, seven are only induced at higher levels, and three are unaffected (Snyder *et al.*, 1993). *CYP4M1* and *CYP4M3* are moderately induced in the midgut but not in the fat body, but *CYP4M3* and *CYP9A2* are not affected by nicotine (Snyder *et al.*, 1995; Stevens *et al.*, 2000).

P450 induction has also been inferred in the polyphagous spider mite *Tetranychus urticae*, where the performance of a bean-adapted population on tomato was severely compromised by piperonyl butoxide (Agrawal *et al.*, 2002). The P450 inhibitor did not reduce acceptance of tomato as a host, nor did it reduce the performance of the bean-adapted population on bean, strongly suggesting a postingestive induction of P450 as a mechanism of acclimation

to the novel host. In the polyphagous noctuid *Spodoptera frugiperda*, ingestion of indole 3-carbinol increases once the continuous exposure to this toxic compound has induced P450 enzymes (Glendinning and Slansky, 1995).

4.1.4.3.3. *Papilio* species and furanocoumarins The adaptation of specialist herbivores to toxic components of their host plants is best documented in the genus *Papilio*. The black swallowtail, *P. polyxenes*, feeds on host plants from just two families, the Apiaceae (Umbelliferae) and the Rutaceae. These plants are phytochemically similar, particularly in their ability to synthesize furanocoumarins. Biogenetically derived from umbelliferone (7-hydroxycoumarin) the linear furanocoumarins (related to psoralen) and angular furanocoumarins (related to angelicin) are toxic to nonadapted herbivores (Berenbaum, 1990). This toxicity is enhanced by light as furanocoumarins are best known for their UV photoreactivity leading to adduct formation with macromolecules, particularly DNA. *Papilio polyxenes* has become a model in the study of adaptation to dietary furanocoumarins. Xanthotoxin, a linear furanocoumarin, induces its own metabolism in a dose-dependent fashion when added to the diet of *P. polyxenes* larvae (Cohen *et al.*, 1989). This P450-dependent metabolism proceeds probably by an initial epoxidation of the furan ring followed by further oxidative attack and opening of the ring, leading to nontoxic hydroxylated carboxylic acids (Ivie *et al.*, 1983; Bull *et al.*, 1986). Inducible xanthotoxin metabolism is observed in all leaf-feeding stages of *P. polyxenes*, and is higher in early instars (Harrison *et al.*, 2001). Xanthotoxin induces its own metabolism in the midgut, but also in the fat body and integument (Petersen *et al.*, 2001). The metabolism of bergapten and sphondin is also induced by dietary xanthotoxin. Levels of total midgut microsomal P450s are unaffected by xanthotoxin, and photoactivation is not required for induction (Cohen *et al.*, 1989). The metabolism of xanthotoxin is 10 times faster in *P. polyxenes* than in the nonadapted *S. frugiperda* (Bull *et al.*, 1986). *Papilio polyxenes* microsomal P450s are also less sensitive to the inhibitory effects of xanthotoxin. NADPH-dependent metabolism of xanthotoxin leads to an uncharacterized reactive metabolite that can covalently bind P450 or neighboring macromolecules, i.e., xanthotoxin can act as a "suicide substrate" (Neal and Wu, 1994). This NADPH-dependent covalent labeling of microsomal proteins is seven times higher in *M. sexta* than in *P. polyxenes*. Inhibition of aldrin epoxidation and *p*-nitroanisole O-demethylation by xanthotoxin is also 6- and 300-fold higher, respectively, in *M. sexta* than in *P. polyxenes* (Zumwalt and Neal, 1993). Myristicin, a methylene dioxyphenyl compound (see Section 4.1.3.4.4) found in the host plant parsnip is less inhibitory to *P. polyxenes* than to *H. zea* (Berenbaum and Neal, 1985; Neal and Berenbaum, 1989).

A distinct protein band of 55 kDa appears in midgut microsomes of xanthotoxin-treated *P. polyxenes* larvae and its microsequencing led to the cloning of *CYP6B1*, a P450 shown to be inducible by xanthotoxin or parsnip (Cohen *et al.*, 1992). Several variants of *CYP6B1* have been cloned that presumably represent different alleles, v1, v2, and v3. The three variants differ from each other at 3, 6, or 9 amino acid positions (Cohen *et al.*, 1992; Prapaipong *et al.*, 1994). The *CYP6B1* gene is selectively induced by linear furanocoumarins. Initial studies suggested that additional, related P450 transcripts were present in *P. polyxenes* and inducible by angular furanocoumarins (Hung *et al.*, 1995b). Expression in the baculovirus system revealed that CYP6B1 v1 and v2 metabolize the linear furanocoumarins bergapten, xanthotoxin, isopimpinellin, and psoralen (Ma *et al.*, 1994). Little metabolism of the angular furanocoumarin angelicin was observed in this early study but improvements in the heterologous expression system by coexpression of insect P450 reductase increased rates of metabolism sufficiently to confirm the role of CYP6B1 in the metabolism of angelicin as well (Wen *et al.*, 2003). The furanocoumarins were metabolized in the improved expression system with the following preference: xanthotoxin > psoralen > angelicin. The latter is less efficiently metabolized *in vivo* (Li *et al.*, 2003) and *P. polyxenes* is less adapted to it (Berenbaum and Feeny, 1981).

A second P450 was cloned from *P. polyxenes*; it encodes CYP6B3 that is 88% identical to CYP6B1 (Hung *et al.*, 1995a). *CYP6B3* is expressed at lower basal levels than *CYP6B1*, but both *CYP6B1* and *CYP6B3* are inducible by xanthotoxin, sphondin, angelicin, and bergapten in the midgut. *CYP6B3* responds more readily to the angular furanocoumarins than *CYP6B1* (Hung *et al.*, 1995a), and *CYP6B1* is more inducible than *CYP6B3* (Harrison *et al.*, 2001). A later study showed that *CYP6B1* is induced by xanthotoxin in the midgut, fat body, and integument, but *CYP6B3* is induced by xanthotoxin only in the fat body (Petersen *et al.*, 2001).

The presence of CYP6B-like transcripts in species related to *P. polyxenes* was suggested early on by positive signals on northern blots with RNA from *Papilio brevicauda* and *Papilio glaucus* that were treated with xanthotoxin (Cohen *et al.*, 1992). *Papilio brevicauda* is like *P. polyxenes*, a species

that feeds on furanocoumarin-containing Apiaceae, but *Papilio glaucus* is a generalist that encounters furanocoumarin-containing plants (e.g., hoptree, *Ptelea trifoliata*) only occasionally. Xanthotoxin, nevertheless, induces its own metabolism in all three species (Cohen *et al.*, 1992). *Papilio glaucus* is highly polyphagous and is reported to feed on over 34 plant families, and therefore offers an interesting contrast to *P. polyxenes*. Esterase, glutathione *S*-transferase, and P450 activities are highly variable and dependent on the species of deciduous tree foliage that this species feeds on (Lindroth, 1989). *Papilio glaucus* has significant levels of linear and angular furanocoumarin metabolism, that are highly inducible by xanthotoxin (Hung *et al.*, 1997). A series of nine *CYP6B* genes and some presumed allelic variants were cloned from *P. glaucus* (Hung *et al.*, 1996, 1997; Li *et al.*, 2001, 2002a). The first two genes *CYP6B4* and *CYP6B5* are products of a recent gene duplication event, and their promoter region is very similar (Hung *et al.*, 1996). Six additional and closely related members of the *CYP6B* subfamily were cloned from *Papilio canadensis*, another generalist closely related to *P. glaucus* but not known to feed on plants containing furanocoumarins (Li *et al.*, 2001, 2002a). Xanthotoxin induced *CYP6B4*-like and *CYP6B17*-like genes in both species, but the level of furanocoumarin metabolism was lower in *P. canadensis* (Li *et al.*, 2001). This wide spectrum of CYP6B enzymes represents a wide range of activities towards furanocoumarin substrates. Whereas CYP6B4 of *P. glaucus* expressed in the baculovirus system efficiently metabolizes these compounds (Hung *et al.*, 1997), CYP6B17 of *P. glaucus*, and CYP6B21 and CYP6B25 from *P. canadiensis* have a more modest catalytic capacity (Li *et al.*, 2003). *Papilio troilus*, a relative of *P. glaucus* that specializes on Lauraceae that lack furanocoumarins, has undetectable basal or induced xanthotoxin metabolism (Cohen *et al.*, 1992).

The genus *Papilio* thus offers a complete range of situations: (1) specialists that deal efficiently with furanocoumarins by inducible expression of *CYP6B* genes; (2) generalists that also carry related *CYP6B* genes, but whose inducibility and metabolism are less efficient; and (3) nonadapted specialists that appear to have lost the inducible *CYP6B* panoply (Berenbaum *et al.*, 1996; Berenbaum, 1999, 2002; Li *et al.*, 2003).

4.1.4.3.4. Furanocoumarins and other insects The metabolism of furanocoumarins or the inducibility of P450 by these compounds is not restricted to Papilionidae. Xanthotoxin induces its own metabolism in the parsnip webworm, *Depressaria pastinacella* (Nitao, 1989). This species belongs to the Oecophoridae, and is a specialist feeder on three genera of furanocoumarin-containing Apiaceae. It is highly tolerant to these compounds and metabolizes them not just by opening the furan ring, but in the case of sphondin, it is also capable of *O*-demethylation (Nitao *et al.*, 2003). Although furanocoumarin metabolism is inducible, the basal (uninduced) activity is high (Nitao, 1989), and the response is a general one, with little discrimination of the type of furanocoumarin inducer or the type of furanocoumarin metabolized (Cianfrogna *et al.*, 2002). The P450 enzymes involved in furanocoumarin metabolism by *D. pastinacella* are unknown. Low stringency northern hybridization failed to elicit a signal with a CYP6B1 probe (Cohen *et al.*, 1992).

A species that does not encounter furanocoumarins, the solanaceous oligophage *M. sexta*, responds to xanthotoxin by inducing *CYP9A4* and *CYP9A5* (Stevens *et al.*, 2000). The generalist *S. frugiperda* also induces P450 as well as glutathione *S*-transferases in response to xanthotoxin (Yu, 1984; Kirby and Ottea, 1995). It has low basal P450-mediated xanthotoxin metabolism, but this metabolism is inducible by a variety of compounds including terpenes and flavone (Yu, 1987). In the highly polyphagous *H. zea*, a similar situation is encountered. Xanthotoxin metabolism is low, but inducible by itself as well as by phenobarbital and α-cypermethrin (Li *et al.*, 2000b). A number of *CYP6B* genes have been cloned from these *Helicoverpa* species. *CYP6B8* of *H. zea* is very close in sequence to *CYP6B7* from *H. armigera*, and it is inducible by xanthotoxin and phenobarbital (Li *et al.*, 2000a). The high conservation of sequence in the SRS1 region suggests that the CYP6B enzymes of *Helicoverpa* are competent in furanocoumarin metabolism as indeed, these species occasionally encounter furanocoumarins in their diet. The *CYP6B9/B27* and *B8/B28* genes are pairs of recently duplicated genes (Li *et al.*, 2002b). Their tissue and developmental pattern of expression is subtly different as is their pattern of induction by a variety of chemicals (Li *et al.*, 2002c).

4.1.4.4. Host Plant, Induction and Pesticides

The adaptive plasticity conferred by the inducibility of P450 enzymes on different diets can have important consequences for insect control and the bionomics of pest insects. It is far from being just an ecological oddity or an interesting set of tales of insect natural history. It is well recognized that the same insect species fed different (host) plants will show differences in their response to pesticides

(Ahmad, 1986; Yu, 1986; Lindroth, 1991), and that these differences often reflect the induction of P450 enzymes, as well as of other enzymes, glutathione *S*-transferases, epoxide hydrolases, etc. The complexity of plant chemistry makes it difficult to account for the contribution of each individual chemical to this response and key components are often analyzed first (e.g., Moldenke *et al.*, 1992). Similarly, the multiplicity of P450 genes and the range of P450 enzyme specificity makes it difficult to predict the outcome of exposure to a plant chemical. The toxicological importance of the plant diet on the herbivore's P450 status (induction, inhibition) is well recognized in pharmacology where the joint use of chemical therapy and traditional herbs can have unpredicted outcomes (Zhou *et al.*, 2003).

Larvae of the European corn borer, *Ostrinia nubilalis*, fed leaves from corn varieties with increasing DIMBOA content and thus increasing levels of resistance to leaf damage had correspondingly increased levels of total midgut P450 and *p*-nitroanisole O-demethylation activity (Feng *et al.*, 1992). These studies suggest that constitutive host plant resistance may affect the insect response to xenobiotics. In addition, the induction of host plant defense by insect damage may itself be a signal for induction of herbivore P450 enzymes, as shown in *H. zea*. The plant defense signal molecules jasmonate and salicylate induce *CYP6B8*, *B9*, *B27*, and *B28* (Li *et al.*, 2002d) in both fat body and midgut. The response to salicylate is relatively specific as *p*-hydroxybenzoate, but not methylparaben, also acts as an inducer.

Treatment with 2-tridecanone, a toxic allelochemical from trichomes of wild tomato, protects *H. zea* larvae against carbaryl toxicity (Kennedy, 1984) and *H. virescens* larvae against diazinon toxicity (Riskallah *et al.*, 1986b). In *H. virescens* larvae, the compound caused both qualitative and quantitative changes in P450 spectral properties (Riskallah *et al.*, 1986a), an induction confirmed by its effect on specific P450 genes in the gut of *M. sexta* larvae (Snyder *et al.*, 1995; Stevens *et al.*, 2000). Larvae of *H. virescens* with a genetic resistance to 2-tridecanone have increased P450 levels and P450 marker activities (benzphetamine demethylation, benzo[a]-pyrene hydroxylation, phorate sulfoxidation), and these can be further increased by feeding 2-tridecanone (Rose *et al.*, 1991). A laboratory population of *H. zea* can rapidly display increased tolerance to α-cypermethrin by selection of an increased P450 detoxification ability with a high dose of dietary xanthotoxin (Li *et al.*, 2000b).

Beyond the host plant, it is the whole biotic and chemical environment that determines the response of an insect to pesticide exposure. Herbicides and insecticide solvents can serve as inducers (Brattsten and Wilkinson, 1977; Kao *et al.*, 1995; Miota *et al.*, 2000). Aquatic larvae are exposed to natural or anthropogenic compounds that alter their P450 detoxification profile (David *et al.*, 2000, 2002; Suwanchaichinda and Brattsten, 2002). Virus infection affects P450 levels (Brattsten, 1987), and the expression of several P450 genes is affected during the immune response (see Section 4.1.5.1.2).

4.1.4.5. Insecticide Resistance

4.1.4.5.1. Phenotype, genotype, and causal relationships Insecticide resistance is achieved in a selected strain or population by: (1) an alteration of the target site; (2) an alteration of the effective dose of insecticide that reaches the target; or (3) a combination of the two. The resistance phenotypes have long been analyzed according to these useful biochemical and physiological criteria. At the molecular genetic level, several classes of mutations can account for these phenotypes (Taylor and Feyereisen, 1996) and a causal relationship between a discrete mutation and resistance has been clearly established for several cases of target site resistance (ffrench-Constant *et al.*, 1999). The molecular mutations responsible for P450-mediated insecticide resistance are only beginning to be explored. In contrast to CYP51, which is a target for a major class of fungicides, no insect P450 has been recognized as a primary target for a commercial insecticide. Thus, biochemical changes in P450 structure or activity can lead to changes in insecticide sequestration, activation, or inactivation, so that all the classes of molecular mutations (structural, up- or downregulation, see Taylor and Feyereisen, 1996) can be theoretically involved in P450-mediated resistance. When the number of P450 genes is taken into account, it is little wonder that P450 enzymes are so often involved in insecticide resistance, and that it has been so difficult finding, and establishing, the role of resistance mutations for P450 genes.

Traditionally, the first line of evidence for a role of a P450 enzyme in resistance has been the use of an insecticide synergist (e.g., piperonyl butoxide), a suppression or decrease in the level of resistance by treatment with the synergist being diagnostic. In cases too many to list here, this initial and indirect evidence is probably correct, however there are cases where piperonyl butoxide synergism has not been explained by increased detoxification (Kennaugh *et al.*, 1993). Piperonyl butoxide may also be a poor inhibitor of the P450(s) responsible for resistance, so that the use of a second, unrelated synergist may be warranted (Brown *et al.*, 1996; Zhang *et al.*,

1997). In addition, the synergist as P450 inhibitor can decrease the activation of a proinsecticide, so that lack of resistance suppression can be misleading. Chlorpyriphos resistance in *D. melanogaster* from vineyards in Israel maps to the right arm of chromosome 2 (see Section 4.1.4.5.5) and is enhanced by piperonyl butoxide rather than suppressed (Ringo *et al.*, 1995).

An independent and additional line of evidence is the measurement of total P450 levels or metabolism of selected model substrates. An increase in either or both being viewed as diagnostic. Again, such evidence is tantalizing but indirect, and the absence of change uninformative. The validation of a model substrate for resistance studies requires substantial knowledge about the P450(s) involved, and is therefore best assessed a posteriori.

An increase in the metabolism of the insecticide itself in the resistant strain is more conclusive. For instance, permethrin metabolism to 4′-hydroxypermethrin was higher in microsomes from *Culex quinquefasciatus* larvae that are highly resistant to permethrin (Kasai *et al.*, 1998b) than in their susceptible counterparts. Total P450 and cytochrome b_5 levels were 2.5 times higher in the resistant strain. Both permethrin toxicity and metabolism were inhibited by two unrelated synergists, TCPPE and piperonyl butoxide. A similarly convincing approach was taken to show P450 involvement in the resistance of housefly larvae of the YPPF strain to pyriproxifen. Gut and fat body microsomes were shown to metabolize the IGR to 4′-OH-pyriproxyfen and 5″-OH-pyriproxyfen at higher rates than microsomes of the susceptible strains and this metabolism was synergist-suppressible (Zhang *et al.*, 1998). The major, dominant resistance factor was linked to chromosome 2 in that strain (Zhang *et al.*, 1997).

Increased levels of transcripts for one or more P450 genes in insecticide-resistant strains has now been reported in many cases (see **Table 5**). This suggests that overexpression of one or more P450 genes is a common phenomenon of metabolic resistance but does not by itself establish a causal relationship with resistance. In some cases, the increased mRNA levels have been related to increased transcription (Liu and Scott, 1998), or increased protein levels (Liu and Scott, 1998; Sabourault *et al.*, 2001). Genetic linkage between increased mRNA or protein levels for a particular P450 and resistance has been obtained to the chromosome level (*CYP6A1*, *Cyp6a2*, *Cyp6a8*, *CYP6D1*, *CYP9A1*, *CYP12A1*: Cariño *et al.*, 1994; Liu and Scott, 1996; Rose *et al.*, 1997; Guzov *et al.*, 1998; Maitra *et al.*, 2000), and closer to marker genes (*Cyp6g1*, *CYP6A1*: Daborn *et al.*, 2001; Sabourault *et al.*, 2001). Linkage is just the first step in establishing a causal link between a P450 gene and resistance.

The following is a discussion of specific cases of P450 genes associated with insecticide resistance that have been studied in greater detail. Evidence for mutations causing constitutive overexpression in *cis* and *trans*, as well as an example of point mutations in a P450 coding sequence are currently available. The variety of mechanisms, even in a single species in response to the same insecticide, is striking. The paucity of available data on the molecular definition of the resistant genotype and on its causal relationship to resistance is also striking when compared to the wealth of data on target site resistance (ffrench-Constant *et al.*, 1999).

4.1.4.5.2. CYP6A1 and diazinon resistance in the housefly Rutgers strain CYP6A1 was the first insect P450 cDNA to be cloned, and the gene was shown to be phenobarbital-inducible and constitutively overexpressed in the multiresistant Rutgers strain (Feyereisen *et al.*, 1989). A survey of 15 housefly strains (Cariño *et al.*, 1992) showed that *CYP6A1* is constitutively overexpressed at various degrees in eight resistant strains, but not in all resistant strains, notably R-Fc known to possess a P450-based resistance mechanism. Thus, the first survey with a P450 molecular probe confirmed the results of the first survey of housefly strains with marker P450 activities (aldrin epoxidation and naphthalene hydroxylation; Schonbrod *et al.*, 1968): there is no simple relationship between resistance and a molecular marker, here the level of expression of a single P450 gene. That different P450 genes would be involved in different cases of insecticide resistance was a sobering observation (Cariño *et al.*, 1992), even before the total number of P450 genes in an insect genome was known. The constitutive overexpression of *CYP6A1* was observed in larvae and in adults of both sexes. Overexpression was shown in both developmental stages to be linked to a semidominant factor on chromosome 2 (Cariño *et al.*, 1994), but the *CYP6A1* gene was mapped to chromosome 5 (Cohen *et al.*, 1994). The gene copy number being identical between Rutgers and a standard susceptible strain (*sbo*), gene amplification could not be invoked to explain overexpression (Cariño *et al.*, 1994), and the existence of a chromosome 2 *trans*-acting factor(s) differentially regulating *CYP6A1* expression in the Rutgers and *sbo* strains was implied. Competitive ELISA using purified recombinant CYP6A1 protein as standard showed that the elevated mRNA levels were indeed translated into elevated protein levels (Sabourault

Table 5 P450 overexpression in insecticide-resistant strains

P450 overexpressed	*Strain*	*Resistance pattern*	*Reference*
Musca domestica			
CYP6A1	Rutgers and other strains	OP, carbamates, IGR	Feyereisen *et al.* (1989), Cariño *et al.* (1992)
CYP6A1[a]	Rutgers		Sabourault *et al.* (2001)
CYP6D1	LPR	Pyrethroids	Liu and Scott (1996)
P450Lpr/CYP6D1[a]	LPR		Liu and Scott (1996)
CYP6D1/CYP6D1[a]	NG98, Georgia	Pyrethroids	Kasai and Scott (2000)
CYP6D1	YPER	Pyrethroids	Shono *et al.* (2002)
CYP6D3	LPR		Kasai and Scott (2001b)
CYP12A1	Rutgers		Guzov *et al.* (1998)
Drosophila melanogaster			
Cyp4e2	RaleighDDT	DDT	Amichot *et al.* (1994)
Cyp6a2	91R	DDT	Waters *et al.* (1992)
	RaleighDDT	DDT	Amichot *et al.* (1994)
	MHIII-D23	Malathion	Maitra *et al.* (2000)
CYP6A2[a]	Several strains	DDT, pyrethroids	Bride *et al.* (1997)
Cyp6a8	91R		Maitra *et al.* (1996)
	MHIII-D23		Maitra *et al.* (2000)
	Wis-1lab	DDT	Le Goff *et al.* (2003)
Cyp6g1	Hikone R and 20 strains	DDT	Daborn *et al.* (2001, 2002)
	WC2	Lufenuron, propoxur	Daborn *et al.* (2002)
	EMS1	Imidacloprid	Daborn *et al.* (2002)
	Wisconsin-1, 91-R	DDT	Brandt *et al.* (2002)
Cyp12d1/2	Wisconsin-1, 91-R	DDT	Brandt *et al.* (2002), Le Goff *et al.* (2003)
Drosophila simulans			
CYP6G1[b]	OV1	DDT, imidacloprid, malathion	Le Goff *et al.* (2003)
Heliothis virescens			
CYP9A1	Macon Ridge	Thiodicarb	Rose *et al.* (1997)
Helicoverpa armigera			
CYP4G8		Pyrethroids	Pittendrigh *et al.* (1997)
CYP6B7		Pyrethroids	Ranasinghe and Hobbs (1998)
Lygus lineolaris			
CYP6X1		Permethrin	Zhu and Snodgrass (2003)
Anopheles gambiae			
CYP6Z1	PSP	Pyrethroids	Nikou *et al.* (2003)
Culex quinquefasciatus			
CYP6F1	JPal-per	Pyrethroids	Kasai *et al.* (2000)
Culex pipiens pallens			
CYP4H21, H22, H23 CYP4J4, CYP4J6	RR	Deltamethrin	Shen *et al.* (2003)
Diabrotica virgifera			
CYP4		Me-parathion, carbaryl	Scharf *et al.* (2001)
Blattella germanica			
P450MA[a]		Chlorpyriphos	Scharf *et al.* (1999)

[a]P450 protein level increased.
[b]Presumed CYP6G1 ortholog of *D. simulans*.

et al., 2001; see **Table 6** for comparison of Rutgers and *sbo*). A comparison of the coding sequence of CYP6A1 between Rutgers and two susceptible strains showed no (*sbo*) or little (*aabys*) sequence variation. Five amino acid changes were noted in aabys, two at the far N-terminus and three at the far C-terminus, well outside the regions (SRS) thought to influence enzyme activity (Cohen *et al.*, 1994). The lack of CYP6A1 sequence difference between Rutgers and *sbo* indicated that if CYP6A1 was implicated in diazinon resistance in the Rutgers strain, it was through elevation of enzyme activity alone. Reconstitution of recombinant CYP6A1 expressed in *E. coli* with its redox partners (Sabourault *et al.*, 2001) provided the conclusive evidence for its role in diazinon resistance, as CYP6A1 metabolizes the insecticide with a high turnover (18.7 pmol/pmol CYP6A1/min), and a

Table 6 Comparison of a susceptible and a resistant strain of the housefly

Fly strain	*sbo (S)*	*Rutgers (R)*
Diazinon contact toxicity: LC50 (μg/pint jar)	4.4	167.8
Resistance ratio	1	37.8
Diazinon topical toxicity: LD50 (μg/fly)	0.059[a]	7.1
Resistance ratio	1	120
P450 level (nmol/mg protein)	0.14	0.29
Aldrin epoxidation (pmol/min/pmol P450)	4.4[a]	15.6
CYP6A1 mRNA relative level	1	27.5
CYP12A1 mRNA relative level	1	15
CYP6A1 protein level (fmol/abdomen)	36	565
Diazinon metabolized by CYP6A1		
Oxidative cleavage (pmol/min/fly)	0.5	7.7
Desulfuration to oxon (pmol/min/fly)	0.2	2.9
OP oxon metabolized by mutant ali-esterase (pmol/min/fly)	0.0	2.2–2.5
NADPH-dep. diazinon metabolism by microsomes (pmol/min/abdomen)	2.9[a]	15.5

[a]NAIDM susceptible strain.

favorable ratio (2.7) between oxidative ester cleavage and desulfuration (see Section 4.1.4.2.2 and Table 4).

The nature of the chromosome 2 *trans*-acting factor and of the mutation leading to resistance in the Rutgers strain has long remained enigmatic despite considerable circumstantial evidence for a major resistance factor on chromosome 2 (Plapp, 1984). Diazinon resistance and high CYP6A1 protein levels could not be separated by recombination in the short distance between the *ar* and *car* genes (3.3–12.4 cM). This region carries an ali-esterase gene (*MdαE7*) implicated as its *Lucilia cuprina* ortholog in organophosphorus insecticide resistance (Newcomb *et al*., 1997; Claudianos *et al*., 1999). A Gly137 to Asp mutation in this ali-esterase abolishes carboxylesterase activity towards model compounds such as methylthiobutyrate, and confers a measurable phosphotriester hydrolase activity towards an organophosphate ("P=O"), chlorfenvinphos. Chromosome 2 of the Rutgers strain carries this Gly137 to Asp mutation, and low CYP6A1 protein levels are correlated with the presence of at least one wild-type (Gly137) allele of *MdαE7*. Recombination in the *ar-car* region could not dissociate diazinon susceptibility, low CYP6A1 protein level, and the presence of a Gly137 allele of the ali-esterase (Sabourault *et al*., 2001). It was therefore hypothesized that the wild-type ali-esterase metabolizes an (unknown) endogenous substrate into a negative regulator of *CYP6A1* transcription. Removal of this regulator (by loss-of-function of the ali-esterase) would increase CYP6A1 production and, hence, diazinon metabolism. Nature seems to have found the optimal loss-of-function mutation (Gly137 to Asp) as the Rutgers haplotype has swept through global populations of the housefly (C. Claudianos, J. Brownlie, V. Taskin, M. Kence, R.J. Russell, J.G. Oakeshott, personal communication). The mutant ali-esterase probably helps clearing the activated form (P=O) of the insecticide (Sabourault *et al*., 2001). The negative regulation by the Gly137 allele and the diazinon resistance-enhancing effect of the Asp137 allele may explain the incomplete dominance of the diazinon resistance trait. Housefly strains that are susceptible or that are not known/shown to overexpress *CYP6A1* predictably carry at least one Gly137 allele (Scott and Zhang, 2003). The LPR strain that overexpresses *CYP6A1* (Cariño *et al*., 1992) and has increased OP metabolism (Hatano and Scott, 1993), as well as another resistant strain (NG98) carry other alleles (Scott and Zhang, 2003) that have impaired ali-esterase activity, Trp251 to Ser or Leu (Campbell *et al*., 1998; Claudianos *et al*., 2001). Corroborating, but indirect evidence for the hypothesis is the predicted pleiotropic effect of a *trans*-acting regulator. Constitutive overexpression of *CYP12A1* whose product metabolizes diazinon as well (Guzov *et al*., 1998) and *GST-1* in the Rutgers strain are also controlled by a chromosome 2 factor, possibly the same as the one controlling *CYP6A1* expression.

Although alternative hypotheses can be advanced, such as the fortuitous genetic closeness between the ali-esterase and a factor that increases the level of CYP6A1, a diazinon metabolizing enzyme, such hypotheses do not have the benefit of elegance. The mutant ali-esterase cannot account for the carbamate and JHA resistance so is not the sole factor of resistance seen in many housefly strains. Diazinon-resistant *L. cuprina* have provided evidence for Oppenoorth's mutant ali-esterase hypothesis (Newcomb *et al*., 1997), but the role of P450 in OP resistance cannot be ignored. Indeed, the Q strain of the sheep blowfly is more resistant to parathion than to paraoxon (Hughes and Devonshire, 1982) and indirect evidence for a P450 involvement in *L. cuprina* diazinon resistance has also been presented (Kotze, 1995; Kotze and Sales, 1995).

4.1.4.5.3. *Cyp6a2* in *Drosophila:* overexpression and mutant enzyme Insecticide-resistant strains of *D. melanogaster* have been studied at the molecular genetic level and the *Cyp6a2* gene has been implicated in the metabolic resistance of several of

them. *Cyp6a2* is constitutively overexpressed in the 91R strain that is resistant to malathion and DDT by a factor of 20–30 relative to the susceptible 91C strain (Waters *et al*., 1992). DDT resistance maps to 56 cM on the left arm of chromosome 2 in the 91R strain (Dapkus, 1992), which is at or near the chromosomal location of *Cyp6a2* (43A1-2). Initially, the presence of a 96 bp insertion in the 3′ UTR of the gene was proposed to confer a low level of expression to the *Cyp6a2* gene (Waters *et al*., 1992), but this insertion (or rather, the lack of it) was neither correlated with DDT resistance (Delpuech *et al*., 1993) nor confirmed to be linked to overexpression in resistant strains (Dombrowski *et al*., 1998). *Cyp6a8* is also constitutively overexpressed in the 91R strain (Maitra *et al*., 1996) and the expression of both *Cyp6a2* and *Cyp6a8* is repressed in 91R/91C hybrids (Maitra *et al*., 1996; Dombrowski *et al*., 1998; Maitra *et al*., 2000). Flies (rosy^{506}, insecticide susceptible and with constitutively low expression of *Cyp6a2*) were transformed with a P-element carrying the *Cyp6a2* gene of the 91R strain driver by its own promoter. These flies express the transgene at higher levels than the endogenous *Cyp6a2* but at lower levels than in their native 91R background (Dombrowski *et al*., 1998). The expression of both *Cyp6a2* and *Cyp6a8* is also constitutively higher in the MHIII-D23 strain initially selected for malathion resistance (Maitra *et al*., 2000). Genetic crosses and chromosome substitution experiments conclusively showed that the expression of both genes (located on the 2R chromosome) is repressed by factors on the third chromosome of the insecticide-susceptible 91C and rosy^{506} strains. In contrast, the third chromosome of the MHIII-D23 and 91R strains carries a loss-of-function mutation for this negative *trans* regulator, allowing the constitutive overexpression of the two genes (Maitra *et al*., 2000). Further careful examination of the promoter activity of the *Cyp6a8* gene by fusion with a luciferase reporter gene in transgenic flies identified a −11/−761 bp region of the *Cyp6a8* gene of the 91R strain that was sufficient to respond to the negative regulation by the rosy^{506} (wild-type) *trans* acting factor (Maitra *et al*., 2002). CYP6A2, similar to CYP6A1, appears to have a broad substrate specificity (see **Table 1**). Whether CYP6A2 of the 91R strain is capable of metabolizing DDT is unknown, as its sequence differs from that of the baculovirus produced CYP6A2 from the *iso-1* strain, which does not metabolize DDT, and from that of the Raleigh-DDT strain, which does (see below). Nonetheless, the studies with the 91R and MHIII-D23 strains are clear indications for loss-of-function mutations in gene(s) encoding negative regulators of P450 gene expression on chromosome 3. Genetic analyses of malathion resistance and of P450 expression (electrophoretic bands and marker activities) (Hallstrom, 1985; Hallstrom and Blanck, 1985; Houpt *et al*., 1988; Waters and Nix, 1988) have pointed to one or more loci between 51 and 61 cM on the right arm of chromosome 3, and it is interesting that this chromosome arm is thought to be orthologous to chromosome 2 of the housefly (Weller and Foster, 1993). The *Drosophila* ortholog of the *MdαE7* gene (*Est23*) is located at 84D9.

Resistance to DDT in the RaleighDDT strain offers a different picture. This strain has very high DDT resistance (>10 000-fold; Cuany *et al*., 1990). Its piperonyl butoxide-suppressible resistance is polyfactorial but the major, dominant resistance factor maps to 55 cM on the second chromosome as in the 91R strain. At least two P450 genes, *Cyp6a2* and *Cyp4e2*, are constitutively overexpressed in this strain (Amichot *et al*., 1994). The genetic localization of resistance matches the locus of *Cyp6a2* (A. Brun-Barale, S. Tares, J.M. Bride, A. Cuary, J.B. Bergé, M. Amichot, personal communication), and the RaleighDDT allele was sequenced. Three point mutations, Arg335 to Ser, Leu336 to Val, and Val476 to Leu, were found (Bergé *et al*., 1998). Overexpression was separated from the point mutations by repeated backcrossing to a marked susceptible strain and DDT selection. The resulting strain called 152 retained the mutant *Cyp6a2* gene and high monofactorial resistance to DDT (>1000-fold), which also maps to 54.4 cM. Both RaleighDDT and 152 flies are characterized by elevated *in vitro* DDT metabolism, as well as elevated ethoxycoumarin and ethoxyresorufin *O*-deethylase activities. Strain 152 lost the constitutive expression of *Cyp6a2*, which was therefore caused by an unlinked *trans*-acting factor (A. Brun-Barale, S. Tares, J.M. Bride, A. Cuary, J.B. Bergé, M. Amichot, personal communication).

The wild-type CYP6A2 and five engineered versions (each individual mutation, a double mutant carrying the first two and a triple mutant corresponding to the RaleighDDT allele) were expressed in *E. coli* and assayed for activity. The mutant enzymes were all characterized by a decreased stability when compared to the wild-type enzyme. The enzyme production in *E.coli* was significantly lower and the ratio of holoenzyme produced (measured by the Fe^{II}–CO complex) to apoenzyme produced (measured with an anti-CYP6A2 antibody) was also lower. However, the triple mutant was uniquely capable of metabolizing DDT to DDA, DDD, and dicofol at rates that were

9–13 times higher than the wild-type enzyme whose DDT metabolism was barely measurable. In addition to the triple mutant, only the Arg335 to Ser single mutant had the capacity to hydroxylate DDT to dicofol at a rate significantly different from wild-type. In contrast to DDT metabolism, the 16β hydroxylation of testosterone was not affected in the various single and multiple mutants. Thus, the three mutations found in *Cyp6a2* of the RaleighDDT (and its derivative 152) strain collectively confer DDT-metabolizing ability to the mutant CYP6A2 enzyme. Genetic localization of DDT resistance to the *Cyp6a2* gene locus is therefore explained. When the point mutations of the 152 strain are combined with overexpression as in the parent RaleighDDT strain, the very high level of resistance can be rationalized. This is the first example of point mutations in an insect P450 enzyme that contribute to insecticide resistance.

4.1.4.5.4. CYP6D1, the housefly LPR strain, and pyrethroid resistance The LPR strain of the housefly is highly resistant to pyrethroids (see **Chapter 6.1**) with a phenoxybenzyl moiety. This permethrin-selected strain has several resistance mechanisms, with P450-based detoxification as a major contributor. An abundant form of P450 (P450Lpr) was purified from abdomens of adult LPR flies, and immunological data indicated that P450Lpr represents 67% of the P450 in microsomes from LPR flies, a tenfold (Wheelock and Scott, 1990) increase over the reference-susceptible strain. Peptide sequences from the purified protein allowed PCR amplification and cloning of the P450Lpr gene, *CYP6D1* (Tomita and Scott, 1995). This gene is located on chromosome 1 of the housefly (Liu *et al*., 1995), and is constitutively overexpressed by about tenfold in the LPR strain. This overexpression is not caused by gene amplification, but by increased transcription. It has been claimed that increased transcription of *CYP6D1* causes insecticide resistance (Liu and Scott, 1998), but transgenic expression of *CYP6D1* in *Drosophila* (Korytko *et al*., 2000a) has not been reported to confer resistance, and CYP6D1 produced in yeast (Smith and Scott, 1997) has not been reported to metabolize pyrethroids. Instead, the evidence for the role of CYP6D1 in pyrethroid resistance is based on the inhibition of microsomal deltamethrin and cypermethrin metabolism by anti-P450Lpr antibodies (Wheelock and Scott, 1992b; Korytko and Scott, 1998). The metabolism of pyrethroids by CYP6D1 has been studied in its microsomal environment of the LPR strain. Deltamethrin is metabolized (<0.5pmol/min/pmol CYP6D1) preferentially at the gem-dimethyl group on the acid moiety (Wheelock and Scott, 1992b) whereas cypermethrin is mainly hydroxylated at the 4′ position on the alcohol moiety, at a very low rate (4.17 pmol/min/pmol CYP6D1, or less than 1 CYP6D1 turnover per 2 h) (Zhang and Scott, 1996b). Whether overexpression or point mutations of *CYP6D1* or both (as for *Cyp6a2* of *Drosophila* RaleighDDT) are involved in pyrethroid resistance in the LPR strain is yet unknown. Indeed, the *CYP6D1* gene sequence from 5 strains shows a high polymorphism, with 57 variable sites in the coding region alone, of which 12 are nonsilent (Tomita *et al*., 1995). Six amino acid changes are specific to the LPR strain (CYP6D1v1) when compared to pyrethroid-susceptible strains and several of these mutations appear to align with SRS3.

The characterization of the resistance mutation(s) is made difficult by its almost completely recessive nature, and by the important contribution of the genes *pen* (for reduced penetration) and *kdr* (for target site resistance) of chromosome 3 (Liu and Scott, 1995). The absence of this resistant chromosome 3 decreases permethrin resistance from 16 000-fold to 170-fold. The remaining resistance is entirely suppressible by piperonyl butoxide, and is conferred by a combination of the resistant chromosomes 1 and 2 from the parent LPR strain in the homozygous condition (Liu and Scott, 1995). There is no substantial resistance or CYP6D1 overproduction conferred by isolated LPR chromosomes 1 or 2, or by their subsequent combination (Liu and Scott, 1995, 1996). Thus, permethrin selection of a P450-mediated resistance requires both copies of the LPR chromosomes 1 and 2. The resistance and *CYP6D1* overexpression linked to chromosome 1 are dominant, whereas the contributions of chromosome 2 are mostly recessive (Liu and Scott, 1996, 1997a). Furthermore, cytochrome b_5 is essential for CYP6D1-metabolism of cypermethrin (Zhang and Scott, 1996b), and the overexpression of cytochrome b_5 in LPR maps to the same chromosomes as permethrin resistance and *CYP6D1* overexpression (Liu and Scott, 1996). These data suggest a unique combination in the LPR strain of chromosome 2 *trans*-acting factor(s) with at least a matched *cis*-factor on chromosome 1. Sequence differences between the 5′UTR of the *CYP6D1* gene of LPR and of other strains have been documented (Scott *et al*., 1999).

CYP6D1 is also overexpressed, but cytochrome b_5 is not, in the permethrin-resistant strain NG98 from Georgia, USA (Kasai and Scott, 2000). These permethrin-resistant flies from Georgia carry virtually the same *CYP6D1v1* haplotype as LPR, with

only 2 nt changes in introns in a sequence spanning from 658 nt upstream of the ATG to the stop codon (Seifert and Scott, 2002). However, in a strain (ALHF) from neighboring Alabama with high permethrin resistance, chromosomes 1 and 2 play little role but chromosome 5 plays a major piperonyl butoxide-suppressible role (Liu and Yue, 2001). *CYP6D1* is overexpressed by about 2.4-fold and cytochrome b_5 by about 1.5-fold in the Japanese strain YPER (Shono *et al.*, 2002). This strain does not carry the *CYP6D1v1* allele, and chromosome 2 has a major role in this permethrin-resistant strain. Permethrin resistance thus appears largely recessive in several housefly strains, with multiple mutations, both regulatory and structural, potentially playing a role in the P450 contribution to the multifactorial resistance.

A gene closely linked to *CYP6D1* on chromosome 1 codes for a similar (78% identity) P450, CYP6D3 (Kasai and Scott, 2001a). It is 12-fold overexpressed in adult flies of the LPR strain, but it is also expressed in larvae (Kasai and Scott, 2001b), as opposed to *CYP6D1*, which has an adult-specific pattern of expression.

4.1.4.5.5. *Drosophila Cyp6g1*, the *Rst(2)DDT* gene at 64.5 cM on chromosome 2 The power of *Drosophila* genetics coupled with the tools made possible by the complete genome sequence has provided the most detailed, yet complex molecular genetic detail about a P450-based insecticide resistance mechanism. The resistance gene *Rst(2)DDT* has been genetically characterized for over 40 years (reviews: Daborn *et al.*, 2001; Wilson, 2001). The position of this gene around 64.5 cM on the left arm of chromosome 2 has become almost mythical, as a number of phenotypes were linked to this locus, from dominant DDT resistance to phenylthiourea susceptibility, from organophosphorus to carbamate resistance, from various P450-dependent activities to vinyl chloride activation. EMS mutagenesis of a wild-type stock and selection with imidacloprid led to two strains with moderate imidacloprid resistance and moderate cross-resistance to DDT (Daborn *et al.*, 2001). Conversely, two DDT-resistant strains (Hikone-R and Wisconsin-1) were shown to be cross-resistant to imidacloprid. Fine scale mapping of this dominant resistance localized *Rst(2)DDT* to a region from 48D5-6 to 48F3-6 on the polytene chromosome map. Of three candidate P450 genes in this region, *Cyp6g1*, *Cyp6g2*, and *Cyp6t3*, only the first one showed constitutive overexpression in the DDT- and imidacloprid-resistant strains tested (Daborn *et al.*, 2001). A DNA microarray comprising probes for all the *Drosophila* P450 genes was addressed with target cDNAs from susceptible strains and from the DDT-resistant Hikone-R strain and the propoxur-resistant WC2 strain. In both cases, *Cyp6g1* was the only P450 gene showing constitutive overexpression (Daborn *et al.*, 2002). Microarray data thus offer, at the transcriptional level, a complete analysis of all P450 genes, thereby facilitating the further rational study of one or more P450 genes (**Figure 24**). This is in constrast to the classical approaches that address single genes simply because they are the only ones for which a molecular probe is available. Overexpression of *Cyp6g1* was confirmed by quantitative (RT)PCR in 20 strains, and DDT, imidacloprid, nitenpyram, and lufenuron resistances were all independently mapped to the *Cyp6g1* locus in the Hikone-R and WC2 strains. The insertion of a terminal direct repeat of the transposable element *Accord* was systematically found in the 5′ UTR of 20 different resistant strains from across the globe. Phylogenetic analysis of the first intron sequence of the gene showed a unique haplotype in resistant strains versus a large diversity of susceptible haplotypes, suggesting that a selective sweep had occurred in global *Drosophila* populations (Daborn *et al.*, 2002). Transgenic flies producing CYP6G1 under control of a heatshock promoter in the GAL4/UAS system showed both increased transcription of *Cyp6g1* and survival to a discriminating dose of DDT after heatshock. Similarly, transgenic expression under the tubulin promoter showed overexpression of only the *Cyp6g1* gene, and larval survival to discriminating doses of acetamiprid, imidacloprid, and nitenpyram (Le Goff *et al.*, 2003). *Cyp6g1* overexpression was confirmed in the DDT-resistant Wisconsin and 91R strains, and the adjacent *Cyp12d1* (or *Cyp12d2*) gene is overexpressed, as well as DDT-inducible in both strains (Brandt *et al.*, 2002). In the Wisconsin-1 strain, *Cyp6g1* and *Cyp12d1/2* are the only P450 genes that are overexpressed, whereas in the Wis1lab strain, in which DDT selection was applied after genetic removal of the *Cyp6g1* region, only *Cyp6a8* is overexpressed (Le Goff *et al.*, 2003). In all the strains discussed above, DDT resistance is significant but low compared to the RaleighDDT strain suggesting that while *Cyp6g1* (overexpressed in the RaleighDDT strain; S. Tares, personal communication) may constitute a first line of defense seen in field populations, further insecticide pressure in the laboratory may select additional mechanisms.

In a Brazilian strain of *D. simulans* resistant to DDT, imidacloprid, and malathion, only the *Cyp6g1* probable ortholog is overexpressed (Le Goff *et al.*, 2003). In a California population of

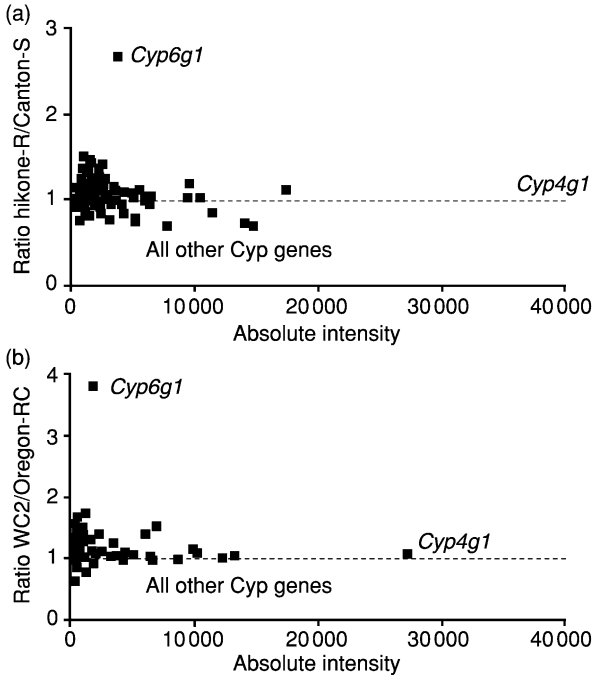


Figure 24 *Cyp6g1* overexpression in two insecticide-resistant strains, Hikone-R(a) and WC2(b), of *Drosophila melanogaster*. DNA microarray analysis revealed overexpression of just one P450 gene. (Reproduced with permission from Daborn, P.J., Yen, J.L., Bogwitz, M.R., Le Goff, G., Feil, E., *et al.*, **2002**. A single P450 allele associated with insecticide resistance in *Drosophila*. *Science 297*, 2253–2256; © AAAS.)

D. simulans the 5′-flanking sequence of the *Cyp6g1* ortholog is nearly fixed for a *Doc* transposable element insertion. This insertion is absent from African populations and is associated with increased transcript abundance of *Cyp6g1* and resistance in a remarkable analogy with the Accord case of *D. melanogaster Cyp6g1* (Schlenke and Begun, 2004).

4.1.4.5.6. P450-mediated resistance: evolution Resistance-conferring mutations in insecticide target sites (Rdl, AchE, sodium channel) have a remarkable pattern of orthology in widely divergent species. The hypothesis that a few paralogous, perhaps orthologous P450 genes would repeatedly be found to be responsible for metabolic resistance in various insect species has not been confirmed, beyond the single example of *Cyp6g1* in *D. melanogaster* and *D. simulans* (that are only 2.5 million years apart). Furthermore, when resistance due to constitutive overexpression of a P450 is caused by a mutation is *trans*, the effect can be pleiotropic with more than one P450 gene from more than one family being overexpressed (e.g., *CYP12A1* and *CYP6A1*). In Diptera as well as in Lepidoptera, members of the *CYP4* family also appear to be involved in resistance, and there is evidence for involvement of *CYP9* genes in the latter (Rose *et al.*, 1997). Therefore, an alternative hypothesis is that the multitude of P450 genes, whose expression is inducible and therefore not strongly expressed in most developmental stages/tissues, constitute a "reservoir" in which mutations affecting expression levels can be selected by insecticide exposure. In the field, this can lead to selective "sweeps" of these most adapted mutations as seen for the global predominance of the Rutgers and *Cyp6g1*/Accord haplotypes in house flies and fruit flies. These mutations would typically be loss-of-function mutations, which inactivate the fine level regulation of expression and therefore increase overall expression of a random P450 gene. If its product happens to metabolize the insecticide, even marginally, this may constitute a selective advantage for the organism. Loss-of-function mutations in the large target of negative regulatory sequences are predicted to be more frequent than gain-of-function mutations (in the smaller open reading frame) that would improve the catalytic efficiency of a P450 enzyme towards the insecticide (Taylor and Feyereisen, 1996). However, polymorphisms in the P450 sequence may be less detrimental to fitness than (often pleiotropic) changes in expression (see, e.g., Halpern and Morton, 1987), so that more examples of point mutations in P450 coding sequences are likely to emerge. The study of insecticide resistance has

clearly entered the age of genomics (Oakeshott *et al.*, 2003) and much progress is expected in the coming years.

4.1.5. Regulation of P450 Gene Expression

4.1.5.1. Spatial and Temporal Patterns of P450 Gene Expression

4.1.5.1.1. Enzyme activities and P450 expression P450-dependent enzyme activities had been detected in virtually all insect tissues and developmental stages studied as early as the first edition of this series (Hodgson, 1985). Many studies have continued to document these patterns and changes (e.g., Feyereisen and Farnsworth, 1985; Ahmad, 1986; Gunderson *et al.*, 1986; Scott and Lee, 1993a). With molecular probes, the expression patterns of individual P450 genes have now been presented in different tissues, developmental stages or induction regimes with varying degrees of detail, and with probes of varying degree of specificity. Early studies relied on northern blots and dilution dot-blots for quantification (e.g., Cariño *et al.*, 1992, 1994), with later ones relying on the highly specific ribonuclease protection assays (e.g., Hung *et al.*, 1995b; Sutherland *et al.*, 1998). **Table 7** shows a summary of those studies reporting a specific pattern of expression. The many studies in which the expression of a particular P450 was studied episodically or by simply rounding up the usual suspects (midgut, fat body, etc.) are not specifically listed.

In situ mRNA hybridization for the *Cyp6a2* gene has been used in *Drosophila* to show that in adults treated with phenobarbital, this gene is transcribed in the midgut, the pericuticular fat bodies, and the Malpighian tubules (Brun *et al.*, 1996). *Cyp6a2* expression is seen in a large number of tissues by immunohistochemistry (Saner *et al.*, 1996), and peaks in the third larval and pupal stages as shown by northern blots. The *Cyp6a2* promoter linked to a GFP reporter gene has been used in transgenic flies to document the expression of this gene (G.C. Unnithan, unpublished data). *In situ* hybridization has been used to document the expression patterns of *sad*, *dib*, and *shd* (Warren *et al.*, 2002; Petryk *et al.*, 2003). Systematic surveys (Tomancak *et al.*, 2002) will eventually document the expression patterns of all P450 genes.

Many unanswered questions remain about P450 expression patterns. For instance, sublethal infection by cytoplasmic polyhedrosis viruses (see **Chapter 6.10**) in *H. virescens*, *M. sexta*, and *S. frugiperda* depresses P450 levels and P450-dependent activities, and increases the toxicity of insecticides (Brattsten, 1987). What are the mechanisms, the P450 genes involved and the significance of these observations? Changes in P450 levels are observed after a blood meal in *Culex pipiens* (Baldridge and Feyereisen,

Table 7 Specific spatial or temporal expression of P450 genes

P450	*Species*	*Tissue/comments*	*Reference*
CYP4C7	*Diploptera punctata*	Corpora allata	Sutherland *et al.* (1998)
CYP4G15	*Drosophila melanogaster*	Larval brain	Maibeche-Coisne *et al.* (2000)
CYP4D21	*Drosophila melanogaster*	Head fat cells in males under control of *transformer* and *doublesex*	Fujii and Amrein (2002)
CYP4L4	*Mamestra brassicae*	Antennae, proboscis, legs	Maibeche-Coisne *et al.* (2002)
CYP4S4	*Mamestra brassicae*	Antennae	Maibeche-Coisne *et al.* (2002)
CYP6B2	*Helicoverpa armigera*	Larval specific	Ranasinghe *et al.* (1997)
CYP6D1	*Musca domestica*	Adult specific	Scott *et al.* (1996)
CYP6D1	*Musca domestica*	Thoracic ganglia	Korytko and Scott (1998)
CYP6L1	*Blattella germanica*	Adult male (probably accessory glands, may be testes as well)	Wen and Scott (2001a)
CYP6Z1	*Anopheles gambiae*	Adult specific	Nikou *et al.* (2003)
CYP15A1	*Diploptera punctata*	Corpora allata	Helvig *et al.* (2004)
CYP302A1	*Drosophila melanogaster*	(Embryo) prothoracic gland portion of the ring gland, follicle cells	Chavez *et al.* (2000); Warren *et al.* (2002)
CYP310A1	*Drosophila melanogaster*	Overexpressed in embryos of Toll1oB mutants (Dorsal target in the mesoderm)	Stathopoulos *et al.* (2002)
CYP312A1	*Drosophila melanogaster*	Male specific	Kasai and Tomita (2003)
CYP314A1	*Drosophila melanogaster*	(Embryo) ovary, gut, Malpighian tubules, fat body	Petryk *et al.* (2003)
CYP315A1	*Drosophila melanogaster*	(Embryo) prothoracic gland portion of the ring gland, follicle cells, nurse cells	Warren *et al.* (2002)
CYP315A1	*Drosophila melanogaster*	One of 30 significantly downregulated genes in eyes overexpressing TIGR/MYOC	Borras *et al.* (2003)

1986), but too little is known of P450 expression patterns in blood-feeding or phloem-feeding insects. What is the significance of the patterns of expression seen in *Drosophila* where about 20% of the genes are in groups of 10–30 genes spread over 20–200 kb that are coordinately expressed (Spellman and Rubin, 2002). Are the P450s present in some of these 200 groups, for instance the *CYP6A* cluster at 55A on chromosome 2R (Ueda *et al.*, 2002), functionally linked?

4.1.5.1.2. DNA microarrays and P450 gene expression The genomic approach to biology has transformed the study of genes one by one to a study of very large numbers of genes in parallel, and ultimately to the study of the whole transcriptome of an organism (see **Chapters 4.12** and **4.13**). This approach, using tools such as DNA microarrays and SAGE, will rapidly change our understanding of the role of P450 in the life of insects. The interpretation of transcriptome studies is in its infancy and "data mining" will require a considerable interplay of experimental design and *in silico* analysis. In the case of P450 genes, the initial EST-based DNA microarrays (e.g., White *et al.*, 1999; Arbeitman *et al.*, 2002) were biased by the poor representation of P450 genes in the early EST collections (Tijet *et al.*, 2001). Careful attention to the potential for cross hybridization between closely related genes is needed (Xu *et al.*, 2001), but the design and validation of thematic arrays (e.g., with all the P450 genes) can allay this concern. The massive amount of data can reveal a link between a P450 gene and a physiological response; it is easy to overstate the relationship that may be a remote effect in a cascade of events. Genomic physiology has nonetheless provided much descriptive information on P450 gene expression already, and a few *Drosophila* examples are listed here. An exhaustive, P450-focused study will be possible in the near future, so these examples are just an exciting preview.

The earliest developmental study in which 534 sequences showed developmental variations (White *et al.*, 1999, Arbeitman *et al.*, 2002) revealed four genes repressed by premature expression of the ecdysone-inducible DHR3 nuclear receptor (*Cyp12a2, Cyp6a2, Cyp4d2*, and *Cyp4ad1*). The link between *Cyp6a2* and the ecdysteroid cascade confirms earlier experimental (Spiegelman *et al.*, 1997) and *in silico* (Dunkov *et al.*, 1997; Dombrowski *et al.*, 1998) data. Two other genes, *Cyp28d1* and *Cyp9f2*, were repressed at the onset of metamorphosis (White *et al.*, 1999). *Cyp310a1* was overexpressed in Toll10B mutant embryos (Furlong *et al.*, 2001), identified as a Dorsal target (Stathopoulos *et al.*, 2002) and its pattern of expression confirmed in a high-throughput *in situ* hybridization project (Tomancak *et al.*, 2002). The Toll pathway represses seven P450 genes (*Cyp316a1, Cyp4ac1, Cyp6g1, Cyp18a1, Cyp28d1, Cyp6w1*, and *Cyp4d14*) in response to bacterial infection in adults, while *Cyp4e2* is also repressed in the acute response (De Gregorio *et al.*, 2001, 2002). Profiles of circadian rhythms (see **Chapter 4.11**) have revealed the coregulation of six *Cyp6a* genes that are adjacent in a cluster at 51D5 on chromosome 2R (Ueda *et al.*, 2002) as well as the control of *Cyp4e2* expression by the *Clk* gene. A number of other P450 genes including *Cyp18a1* and *Cyp4p1* are also cycling with different phases (Claridge-Chang *et al.*, 2001; McDonald and Rosbash, 2001), with *Cyp4d21, Cyp6a21*, and *Cyp304a1* identified in both studies (**Figure 25**). Aging and oxidative stress (paraquat treatment) have revealed the responsiveness of a number of P450 genes, with *Cyp6a17* and *Cyp28a5* upregulated by paraquat and *Cyp6g1*

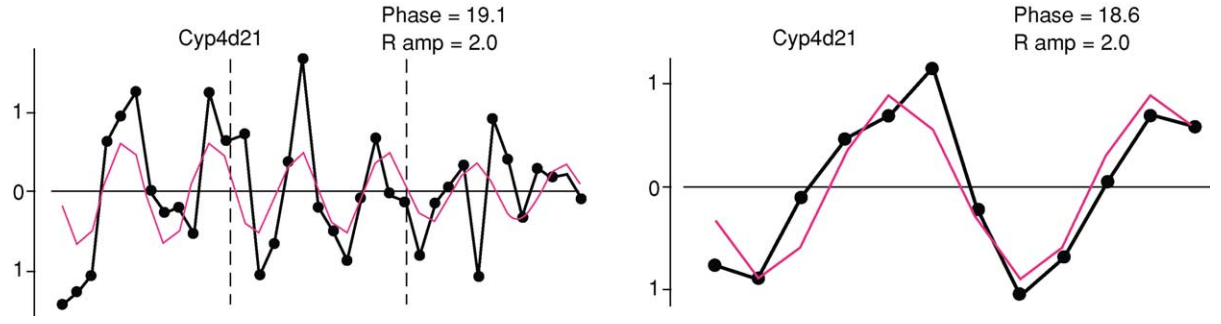


Figure 25 Circadian rhythmicity of *Cyp4d21* expression in *Drosophila melanogaster* expressed as a log ratio of change. Left: 36 time points collected over 6 days (Affymetrix microarray data); right: 12 time points collected over 2 days. Estimated phases and log ratio amplitudes are indicated. Red lines are 24-h guidelines. (Reprinted with permission from Claridge-Chang, A., Wijnen, H., Naef, F., Boothroyd, C., Rajewsky, N., *et al.*, **2001**. Circadian regulation of gene expression systems in the *Drosophila* head. *Neuron. 32*, 657–671; © Elsevier.)

downregulated under those conditions (Zou *et al.*, 2000). In another study, 18 P450 genes were found to respond to paraquat treatment by either increased or decreased expression. Those increased by paraquat treatment were also increased by H_2O_2 treatment. Four genes are increased by treatment with tunicamycin that blocks N-glycosylation and thus causes a stress at the level of the endoplasmic reticulum (ER). *Cyp28a5* is strongly induced by the three treatments, suggesting it may be a general stress-responsive gene (F. Girardot, V. Monnier, and H. Tricoire, personal communication). The subtle interactions of sex, age, and genotype have been discussed (Jin *et al.*, 2001), with *Cyp4c3* and *Cyp6g1* as two examples. *Cyp4c3* is upregulated by starvation in larvae, an effect that can be attributed to lack of sugar (Zinke *et al.*, 2002). This effect points to a conserved physiological function of some *CYP4C* genes, as cockroach *CYP4C1* is also induced by starvation and is upregulated by hypertrehalosemic hormone (Lu *et al.*, 1995).

DNA microarray analysis of the transcriptome of transgenic flies has also led to the validation of *Drosophila* as a model for human disease (Borras *et al.*, 2003). Flies transformed with a human gene associated with glaucoma show a phenotype of distorted ommatidia and fluid discharge. Fifty transcripts have an altered expression profile in the head and, interestingly, *Cyp315a1* is suppressed significantly in these experiments (Borras *et al.*, 2003). Specific or thematic microarrays have also started to change research in insecticide resistance (Oakeshott *et al.*, 2003; see Section 4.1.4.5.5).

Serial analysis of gene expression (SAGE) analysis identified *Cyp4d21* as one of three genes expressed preferentially in the fat cells of the head, and specifically expressed in males under control of the sex determination genes *transformer* and *doublesex* (Fujii and Amrein, 2002). The male-specific expression of *Cyp312a1* in microarray experiments has also been reported and confirmed by real-time PCR (Kasai and Tomita, 2003).

4.1.5.2. P450 Induction, Inducers and Signal Transduction

4.1.5.2.1. Induction by hormones *Drosophila Cyp18* was discovered as the ecdysone-inducible *Eig17-1* gene (Hurban and Thummel, 1993). The expression of this gene clearly pulses closely after each ecdysteroid peak (Bassett *et al.*, 1997) (**Figure 26**), but the mechanism of its ecdysone inducibility has not been studied in detail. The *Cyp6a2* gene is also inducible by ecdysone. The arrest of ecdysone production decreases CYP6A2 levels and slightly increases DDT toxicity (Spiegelman *et al.*, 1997). Ecdysone inducibility of *Cyp6a2* may be mediated by Broad-Complex Z1 and Z4 transcription factor binding sites or ecdysone response elements (EcRE) (see **Chapter 3.5**) seen in the promoter of the gene (Dunkov *et al.*, 1997; Dombrowski *et al.*, 1998).

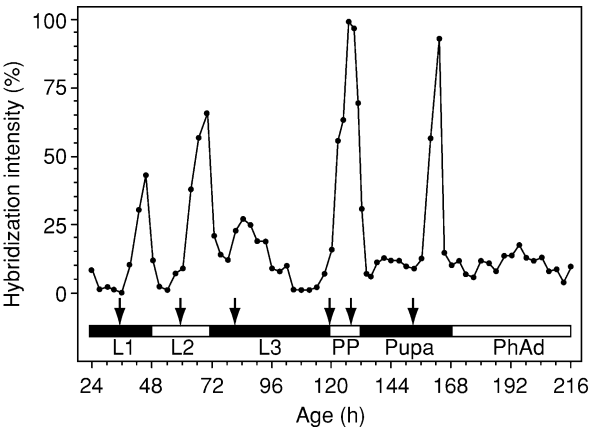


Figure 26 Developmental pattern of *Cyp18* expression in *Drosophila*. Peaks of *Cyp18* expression follow each surge in endogenous molting hormone level. (Reprinted with permission from Bassett, M.H., McCarthy, J.L., Waterman, M.R., Sliter, T.J. **1997**. Sequence and developmental expression of Cyp18, a member of a new cytochrome P450 family from *Drosophila*. *Mol. Cell. Endocrinol.* *131*, 39–49; © Elsevier.)

CYP4C1 is inducible by hypertrehalosemic hormone (HTH) (see **Chapter 3.10**) and starvation in the fat body of *Blaberus discoidalis* (Bradfield *et al.*, 1991). The dose-response for *CYP4C1* induction (ED50 = 3 pmol/insect, 8 h after injection) corresponds to the physiological range of HTH-dependent biosynthesis of trehalose and this induction appears to be a direct effect of the hormone (Lu *et al.*, 1995). JH inhibits *CYP4C1* expression in adult females, but not in males; this JH inhibition controls the decrease in *CYP4C1* transcript levels in the fat body. Yet HTH can override the JH inhibition (Lu *et al.*, 1999). The dual regulation of *CYP4C1* expression is indicative of a fine regulation of this P450, presumed to be involved in fatty acid ω-hydroxylation (by analogy to the function of mammalian CYP4 enzymes). If confirmed, such a function might be related to gluconeogenesis from fatty acid oxidation (an HTH stress and starvation response) as opposed to (and thus inhibited by) the JH-controlled metabolism of the fat body cell towards vitellogenesis (Lu *et al.*, 1999) (see **Chapters 3.7** and **3.9**).

4.1.5.2.2. Xenobiotic-inducible genes One remarkable feature of P450 genes is that the transcription rate of many of them (sometimes gene batteries

that also include genes other than P450s) is induced by foreign chemicals or "xenobiotics." Genetic models have provided much of our current understanding of induction of the mammalian *CYP1* genes by aryl hydrocarbons, such as dioxin, through the Ah receptor pathway (Denison and Nagy, 2003). For instance, dioxin and other *a*romatic *h*ydrocarbons bind to the Ah receptor, which itself dimerizes with a protein called Arnt. The dimer of these two bHLH-PAS proteins is thought to induce a battery of genes including the *CYP1A1* gene. However, the molecular mechanism is not well understood for many other classes of inducers. Nuclear receptors play a key role at the interface between physiological responses and environmental responses (Chawla *et al*., 2001; Honkakoski *et al*., 2003) (see **Chapter 3.6**).

Our understanding of the molecular genetics of induction in insects will benefit from the power of the *Drosophila* model. Phenobarbital (see below) is a known inducer of many P450 activities in *Drosophila* (Hallstrom and Grafstrom, 1981; Bigelow *et al*., 1985; Fuchs *et al*., 1994; Amichot *et al*., 1998). Other inducers of one or more marker activities in *Drosophila* include: β-naphthoflavone or polychlorinated biphenyls (Hallstrom and Grafstrom, 1981), but not benzo[a]anthracene or dioxin (TCDD) (Bigelow *et al*., 1985); butylated hydroxytoluene and rifampicin (Zijlstra *et al*., 1984); *trans*-stilbene oxide, triphenyldioxane, benzo[a]pyrene, but not TCPOBOP, a potent inducer of *CYP2B* genes in mice but not in rats (Fuchs *et al*., 1994); prochloraz, aminopyrine and clofibrate (Amichot *et al*., 1998). Thus, the major types of inducers known in the vertebrate toxicology literature, if not each specific chemical, have shown to be active as inducers in *Drosophila*. Specific genes induced are shown in **Table 8**. Biochemical techniques for the measurement of an Ah receptor in *Drosophila* have detected characteristic TCDD binding to a cytosolic fraction (Bigelow *et al*., 1985, but see Hahn, 2002). The *Drosophila* genome carries 12 paralogs of the vertebrate *Ah* receptor gene and a probable ortholog, *spineless* (Hahn, 2002). The mechanism of aryl hydrocarbon induction in insects remains uncertain.

4.1.5.2.3. Phenobarbital-inducible genes Phenobarbital-like inducers include a variety of chemicals with widely divergent physicochemical properties. In mammals, these inducers are mostly tumor promoters (phenobarbital, polychlorinated biphenyls, chlordane) but they are not genotoxic. They must affect some important homeostatic process in cells, as phenobarbital induction is observed in plants, nematodes, insects and mammals, and some bacteria. Thus phenobarbital may not be an ecologically relevant inducer of environmental response genes *sensu* Berenbaum (2002), but a physiologically relevant inducer *sensu* Nebert (1991). It highlights a pathway shared by many chemically unrelated compounds.

In vitro studies with vertebrate phenobarbital-inducible genes have identified a phenobarbital-responsive enhancer module (PBREM) in the distal regulatory element of *CYP2B* genes. Heterodimers of the nuclear receptor CAR and of the retinoid X receptor activate the PBREM of the mouse *Cyp2b10* gene (Honkakoski and Negishi, 2000; Wei *et al*., 2000). Both the CAR and the PXR nuclear receptors are involved in the induction of drug metabolizing enzymes in vertebrates. In the case of phenobarbital induction in *Drosophila*, the definition of *cis*- and *trans*-regulatory elements has begun. Transfection experiments with *Cyp6a2*/luciferase reporter constructs showed that dose-dependent phenobarbital induction can be studied in this system. Promoter elements sufficient for directing both basal and phenobarbital-inducible expression are located within 428 bp of 5′ of the start codon (Dunkov *et al*., 1997). Elements further upstream (984 and 1328 bp) are needed for higher basal activity. *In vivo* experiments with transgenic flies showed that (pheno)barbital induction was functional with 1331 or 985 bp of 5′ upstream DNA, whereas 129 bp upstream of the start codon, while conferring a low level of basal expression, does not support induction (Dombrowski *et al*., 1998). Further experiments with transgenic flies carrying the luciferase reporter gene driven by promoter sequences of the *Cyp6a8* gene showed low basal activity of a −11/−199 bp construct, with higher basal activities seen with 761 and 3100 bp of upstream DNA (Maitra *et al*., 2002). Phenobarbital inducibility was apparent with the three types of constructs, the highest level of induced activity being achieved with the −11/−761 bp promoter region. These studies thus consistently identify a small upstream region of the promoter that appears necessary and sufficient for phenobarbital inducibility, and all point to the presence of barbie box-like sequences in the promoter region of these inducible genes (Dunkov *et al*., 1997; Dombrowski *et al*., 1998; Maitra *et al*., 2002). Barbie box sequences are difficult to define structurally in insects as the only functionally verified sequences are bacterial (Shaw and Fulco, 1993). However, their distribution is not random, and has been observed upstream of phenobarbital-inducible P450 genes of other species

Table 8 Inducers of P450 genes in insects

Inducers	*P450 induced*	*Reference*
Hormones		
HTH	*CYP4C1*	Bradfield *et al.* (1991); Lu *et al.* (1995, 1999)
Ecdysteroids	*Cyp6a2, Cyp18*	Hurban and Thummel (1993); Spiegelman *et al.* (1997)
Ovarian hormone[a]	*CYP4C7*	Sutherland *et al.* (2000)
Alkaloids		
Nicotine	*CYP4M1, CYP4M3*	Snyder *et al.* (1995a)
Senita/saguaro cactus alkaloids	*CYP28A1, A2, A3, CYP4D10*	Danielson *et al.* (1997, 1998); Fogleman *et al.* (1998)
Terpenoids		
Monoterpenes (peppermint oil)	*CYP6B2*	Ranasinghe *et al.* (1997)
α-Pinene	*CYP6B2, CYP6B7*	Ranasinghe *et al.* (1997); Ranasinghe and Hobbs (1999b)
Menthol	*CYP6B2*	Ranasinghe *et al.* (1997)
Gossypol	*CYP6B27*	Li *et al.* (2002c)
Derived from phenylpropanoid pathway		
Salicylic acid	*CYP6B8, B9, B27, B28*	Li *et al.* (2002d)
Chlorogenic acid	*CYP6B8, B9, B27, B28*	Li *et al.* (2002c)
Coumarins		
Coumarin	*CYP6B27*	Li *et al.* (2002c)
Xanthotoxin	*CYP6B1, B3, B4, B8, B9, B17, B27, B28, CYP9A2, A4, A5*	Cohen *et al.* (1992); Hung *et al.* (1995b); Stevens *et al.* (2000); Li *et al.* (2000a, 2001, 2002c)
Bergapten	*CYP6B3*	Hung *et al.* (1995a)
Angelicin	*CYP6B3*	Hung *et al.* (1995a)
Sphondin	*CYP6B3*	Hung *et al.* (1995a)
Flavonoids		
Flavone	*CYP6B8, B9, B27, B28*	Li *et al.* (2002c)
Quercetin	*CYP6B8*	Li *et al.* (2002c)
Rutin	*CYP6B8, B27, B28*	Li *et al.* (2002c)
Various natural products		
Ethanol	*CYP6A1*	Cariño *et al.* (1992)
2-tridecanone	*CYP4M3, CYP6A2*	Snyder *et al.* (1995); Stevens *et al.* (2000)
2-undecanone	*CYP4M1, M3, CYP9A2, A4, A5*	Snyder *et al.* (1995); Stevens *et al.* (2000)
Jasmonic acid	*CYP6B8, B9, B27, B28*	Li *et al.* (2002d)
Indole-3-carbinol	*CYP6B8, B9, B27, B28, CYP9A2*	Stevens *et al.* (2000); Li *et al.* (2002c)
Synthetic chemicals		
Phenobarbital	*CYP4D10, Cyp4e2, CYP4L2, CYP4M1, CYP4M3, CYP6A1, Cyp6a2, CYP6B7, CYP6B8, CYP6B9, CYP6B27, CYP6B28, CYP6D1, CYP6D3, CYP9A2, CYP12A1, CYP28A1, A2, A3*	Snyder *et al.* (1995); Brun *et al.* (1996); Dunkov *et al.* (1996); Danielson *et al.* (1997, 1998); Guzov *et al.* (1998); Ranasinghe and Hobbs (1999a); Li *et al.* (2000a, 2002c), Kasai and Scott (2001b)
Barbital	*Cyp6a2, Cyp6a8, Cyp6a9*	Maitra *et al.* (1996), Dombrowski *et al.* (1998)
DDT	*Cyp12d1/2*	Brandt *et al.* (2002)
Alkylbenzenes (pentamethyl benzene)	*CYP4*	Scharf *et al.* (2001)
Butylated hydroxyanisole	*CYP6B2*	Ranasinghe *et al.* (1997)
Piperonyl butoxide	*CYP6A1, CYP6B2*	Cariño *et al.* (1992); Ranasinghe *et al.* (1997)
Clofibrate	*CYP4M1, M3, CYP9A2, 4*	Snyder *et al.* (1995); Stevens *et al.* (2000)
p-Hydroxybenzoate	*CYP6B8, B9, B27, B28*	Li *et al.* (2002d)
Permethrin	*CYP6X1*	Zhu and Snodgrass (2003)
Fenvalerate	*CYP6B7*	Ranasinghe and Hobbs (1999b)
Cypermethrin	*CYP6B7, B27, B28*	Ranasinghe and Hobbs (1999b); Li *et al.* (2002c)

[a]Identity of ovarian hormone unknown.

such as *CYP6A1, CYP6D1* (Dunkov *et al*., 1997; Scott *et al*., 1999), and of *CYP6B* genes (Hung *et al*., 1996), of which some are phenobarbital-inducible (Li *et al*., 2002c). The role of these sequences in regulating insect P450 genes thus remain conjectural. Microarray experiments show that less than 10% of the Drosophila P450 genes are inducible by phenobarbital, and this approach will facilitate the rational study of P450 induction in this model species.

Inducibility is under genetic control (Hallstrom, 1987), but is not trivial to quantify. Several pioneers have noted that induction in insecticide-resistant strains was modified (e.g., Terriere and Yu, 1974), often lower (in fold induction) than in susceptible strains in the housefly and in *Drosophila*. Non-inducibility maps to the resistance gene (Hallstrom *et al*., 1982). Resistant strains have thus been called "constitutive mutants" (Hallstrom, 1985). It has been claimed that phenobarbital induction of *CYP6D1* is *due* to a *trans* acting factor on chromosome 2 of the housefly (Liu and Scott, 1997b). However, this claim is based on genetic crosses between an inducible multimarker strain *aabys*, and the pyrethroid-resistant strain LPR. In that strain, CYP6D1 transcript levels are higher than those achieved in phenobarbital-treated *aabys* flies, so that lack of further induction cannot be characterized as refractoriness to induction (a term which should be reserved for a low, basal level that is not changed by phenobarbital treatment – a volunteer is not refractory to the draft). That study therefore has no relevance to the genetic control of phenobarbital inducibility, but merely confirms that a specific interaction of LPR chromosomes 1 and 2 is needed for high expression of *CYP6D1v1* (Liu and Scott, 1995, 1996).

4.1.5.2.4. Furanocoumarin-inducible genes The inducibility of *CYP6B* genes by furanocoumarins has been studied extensively in *Papilio* species (see Section 4.1.4.3.3). The *CYP6B1* and *CYP6B3* genes are inducible by the linear furanocoumarin xanthotoxin, and variably induced by bergapten, angelicin, and sphondin (Prapaipong *et al*., 1994; Hung *et al*., 1995a, 1995b). This variability may be the result of individual polymorphism in inducibility. The 5′ flanking sequence of the *CYP6B1v3* gene comprising nt −838 to +22 (relative to the transcription start site) was fused to the reporter chloramphenicol acetyl transferase (CAT) gene and this construct was transfected into Sf9 cells (Prapaipong *et al*., 1994). In this system, the promoter region of the *CYP6B3* gene had a low basal activity, and this was maximally induced (about twofold) by 2 μg xanthotoxin/ml culture. Thus, at least some of the sequences required for induction are present in this upstream region, and the xanthotoxin signaling cascade is present in these cells from the generalist herbivore *Spodoptera frugiperda*. Analysis of the 5′-upstream region of the *CYP6B1, B3, B4*, and *B5* genes from *P. polyxenes* and *P. glaucus* indicated a high degree of similarity, with several putative regulatory elements being noted (Hung *et al*., 1996). These include sequences similar to the XRE (xenobiotic response element) of the Ah receptor of vertebrates, the barbie box, or the ARE (antioxidant RE) of rodent GST genes. With the cloning of additional *CYP6B* genes, a number of additional sequences, conserved among *CYP6B* genes and with similarities to known regulatory sequences of other genes, have been recorded (Petersen *et al*., 2001; Li *et al*., 2002a; Petersen *et al*., 2003). A region (−136 to −119 of the *CYP6B1* gene) named XRE-xan (for XRE-xanthotoxin) has reportedly been identified as a being required for basal transcription and xanthotoxin inducibility and is conserved among *CYP6B* genes (results quoted by Hung *et al*., 1996; Berenbaum, 2002). Evidence for the function of this XRE-xan was presented in an extensive series of experiments with the promoter region of the *CYP6B1v3* gene driving expression of CAT in Sf9 cells (Petersen *et al*., 2003). Serial deletion experiments from −5 kb to −97 indicated the presence of a negative regulatory element necessary for basal transcription between nt −228 and −146 and a strong positive element between −146 and +22. Xanthotoxin inducibility was similarly mapped to a region between −146 and −97. This region includes the CAAT box basal promoter element. Substitution mutagenesis of tracts of DNA within this region identified nt −136 to −119 (TGACTGGCAATTTTTTTT) as the element called XRE-xan. Sequences similar to the ecdysone RE (EcRE) and an ARE overlap on the 5′ end of the XRE-xan. A region −37 to +22 was found to be necessary for both basal and induced activity and thus forms part of the core promoter of the *CYP6B1* gene. Mutagenesis of the EcRE portion 5′ of the XRE-xan element slightly diminished basal and induced expression, but mutagenesis of the TGAC sequence common to the EcRE, ARE, and XRE-xan abolished basal and induced expression. However, this sequence is not conserved in the same region of the xanthotoxin-inducible *CYP6B3* gene. The induction pattern of the *CYP6B1v3* promoter fragment (−380 to +22) driving CAT expression in Sf9 cells revealed a significant induction by flavone, a coumarin, and angular furanocoumarins, in contrast to the pattern seen for the *CYP6B1* gene *in vivo* (Petersen *et al*., 2003). This lack of fidelity may be caused by a different specificity of the receptor(s)

expressed in Sf9 cells from that of the *P. polyxenes* midgut receptor(s). It may also be due to the absence of additional regulatory elements upstream of −380 in the test system. The definition of xenobiotic-responsive elements of well-characterized P450 genes may facilitate their detection in other inducible genes (e.g., lepidopteran *CYP9A* genes; Stevens *et al.*, 2000), and will help identify the DNA-binding proteins that bind to them.

4.1.6. Conclusion and Prospects

Some examples of the multiple functions of insect P450 enzymes, of their complex biochemistry and of their toxicological and physiological importance have been presented. Is there a common thread and does the hopeless chaos of this chapter simply reflect our fragmentary knowledge? With evolution as the guide, we can now summarize the relative positions of the insect P450s that have been discussed on the phylogenetic tree of the P450 superfamily. Insect P450 sequences can be distinguished in four major clades, which fall into four subclasses of Gotoh's classification of P450 families (Gotoh, 1993) and this is schematically illustrated in **Figure 27**:

1. A large group of insect P450s comprising the CYP6, CYP9, CYP28 families as well as the CYP308–310 and CYP321 families is most closely related to vertebrate CYP3 and CYP5 families. Because of the initial "lumping" of CYP6B1 into the CYP6 family and later "splitting" of the CYP300s, this group is rather heterogeneous with regard to CYP families and subfamilies. Few of the CYP6 enzymes have been characterized, the CYP6A1 and 2 of Diptera are capable of metabolizing xenobiotics, and several CYP6B of Lepidoptera are known to metabolize a variety of furanocoumarins. Genes from this group appear to share the characteristics of "environmental response genes" as defined by Berenbaum (2002), specifically: (1) very high diversity; (2) proliferation by duplication events; (3) rapid rates of evolution; (4) occurrence in gene clusters; and (5) tissue- or temporal-specific expression. Of course these characteristics are not independent of each other, and they are difficult to measure objectively. Nonetheless, the multiple paralogs from these families can be dated to no more than 150–200 MYA and this gives a good idea of the dynamic nature of this clade's evolution.
2. Another clade includes the large CYP4 family that has members from vertebrates and insects, but also several P450 families from *C. elegans*, as well as the insect CYP311, 312, 313, 316, and 325 families initially discovered with the sequence of the *Drosophila* and *Anopheles* genomes. This group of sequences is highly diversified, perhaps even more so than the CYP6/9/28 group, and some *CYP4* genes are clearly inducible by xenobiotics. However, specialized physiological functions are recognized (Sutherland *et al.*, 1998) or proposed (Bradfield *et al.*, 1991) for some enzymes of this group.
3. The mitochondrial P450s of vertebrates and insects (as well as CYP10 of *Lymnea stagnalis* and CYP44A1 of *C. elegans*, both presumed mitochondrial P450 sequences) are monophyletic. Within the mitochondrial clade, the CYP12 family appears to behave like the CYP6/9/28 clade, as a rapidly evolving group of paralogous genes. Some of the other families of mitochondrial P450s are now clearly linked to the ecdysteroid metabolism pathway. Mitochondrial P450s of insects thus evolved differently from the vertebrate mitochondrial P450s and subcellular localization in mitochondria can no longer be considered as evidence for a role in endocrine physiology. These sequences are all derived from an ancestral microsomal P450, and are only distantly related to soluble bacterial P450s that rely on a similar tandem of redox partners (i.e., adrenodoxin and adrenodoxin reductase and their paralogs). It is likely that mitochondrial P450s evolved as a result of the mistargeting of a microsomal P450 after mutations affected the N-terminal sequence. Drastic changes in the redox partner interactions were probably not necessary as rabbit CYP2B4 activity is supported by the bacterial ferredoxin and ferredoxin reductases of the P450cam and P450lin systems (Bernhardt and Gunsalus, 1992). Also, truncated CYP1A1 targeted to mitochondria interacts productively with either adrenodoxin, P450 reductase, or bacterial flavodoxin (Anandatheerthavarada *et al.*, 2001).
4. In fact, mitochondrial sequences are more closely related to a group of sequences that include vertebrate microsomal CYP1, CYP2, CYP17, and CYP21 as well as insect CYP15, CYP18, and the CYP303–307 series. Several of these insect sequences represent enzymes involved in essential physiological functions (e.g., CYP15), as are CYP17 and CYP21 of vertebrates. The CYP2, however, are widely considered as "environmental response genes" in mammalian species.

No clear pattern emerges from this current knowledge of insect P450 evolution. There is not one

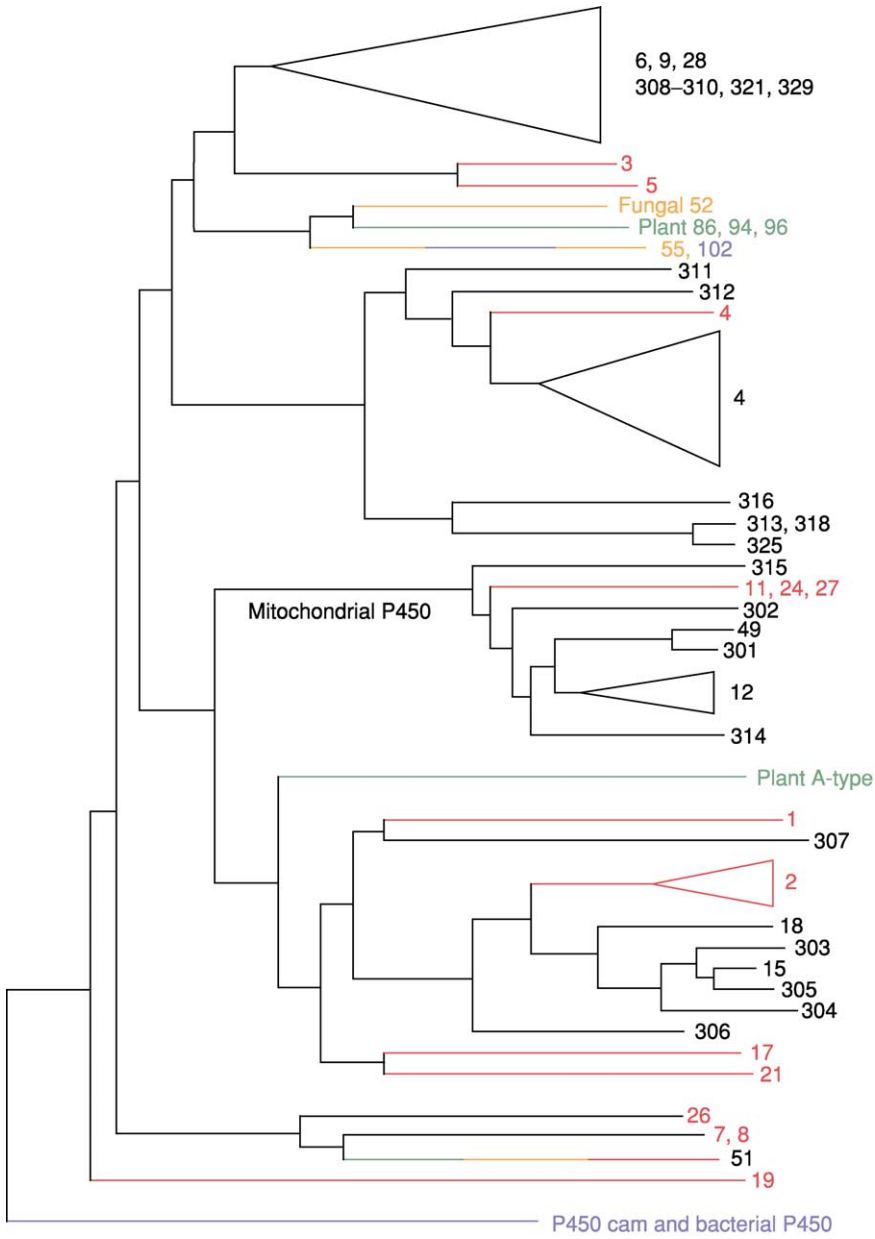


Figure 27 Schematic phylogeny of insect P450 genes and relationships with major P450 families from other organisms. The color code is: black, insects; red, vertebrates; brown, fungi; green, plants; blue, bacteria. Triangles represent families with large numbers of genes. Plant A-type P450 genes are found only in plants. CYP55 and CYP102 are P450foxy and P450BM3, fatty acid hydroxylases fused with a P450 reductase domain. CYP51 is the sterol 14α-demethylase, the only P450 ortholog found in different phyla. The phylogeny is not drawn to scale. The deep branch topology may be revised and there are two main sources of bias: most P450 sequences are from Diptera, and the molecular clock is not constant. This is seen for instance in the deep branching of the steroid aromatase (CYP19).

class of P450s involved in physiological processes and another distinct class involved mostly in xenobiotic metabolism. Physiological functions are not restricted to one branch of the P450 evolutionary tree, and the ramifications that seem typical of "environmental response genes" are found in both microsomal and mitochondrial P450 encoding genes. Will this fuzzy image clear up with the availability of more insect genome sequences, or will it be clouded even more by the discovery of dispersed physiological functions throughout the phylogeny?

There are many physiological functions for P450s that are not known or even suspected. In humans, mutations in the *CYP1B1* gene were linked to congenital glaucoma (Stoilov *et al.*, 1997), a clear physiological effect, still unexplained by what is known

of CYP1B1 biochemistry. In insects, CYP4C7 was shown to be selectively expressed in the corpora allata and to metabolize JH and its precursors to new metabolites (Sutherland *et al*., 1998). Both the presence of this P450 in corpora allata and the existence of these metabolites were unexpected findings. Consequently, the search for new functions of P450 enzymes should involve broad, rational screens, and the technology is now available to perform such functional screens.

When significantly more data are obtained on the catalytic competence of a wide variety of insect P450 enzymes, it will become easier to understand the way in which insects maintain a wide repertoire of P450 genes. If positive selection of a few P450 genes can lead to specialized enzymes in oligophagous species (Li *et al*., 2003), is this an evolutionary dead-end? Do the P450 enzymes with "broad and overlapping" specificity serve as a perpetual reservoir where some genes, because of their pattern of expression or inducibility or catalytic competence, can then serve as templates for the evolution of a new branch of specialized enzymes? Does this "primordial soup" perpetuate itself simply by a neutral process of intense gene duplication or are new chemical insults of the environment frequent enough to positively select for a minimal number of "jack-of-all-trades" P450 enzymes? How do P450 genes get recruited into physiological networks and biosynthetic pathways? The gap between "endogenous" and "xenobiotic" is being filled with new data and insights from P450 research (Nebert, 1991). Every 20 years or so researchers are discovering and rediscovering that treatments with inducers and inhibitors that cause major imbalances in P450 levels are deleterious to fitness in insects (Mitlin and Konecky, 1955; Yu and Terriere, 1974; Darvas *et al*., 1992; Fuchs *et al*., 1993). It is unlikely that these effects result solely from an interference with the handful of P450 enzymes involved in ecdysteroid and JH metabolism. Other regulatory pathways and other signal molecules remain to be discovered through a better understanding of insect P450 enzymes.

At the time of the first edition of this series, it seemed difficult enough to fully comprehend the function and role, the catalytic competence and the regulation of a single insect P450 enzyme. In the era of genomics, this task has now been multiplied by a hundred. It is hoped that interest in insect P450s will also grow by as much, and that new knowledge on insect P450 will continue to contribute to all branches of entomology, from toxicology to physiology and ecology.

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., *et al*., **2000**. The genome sequence of *Drosophila melanogaster*. *Science 287*, 2185–2195.

Agosin, M., **1985**. Role of microsomal oxidations in insecticide degradation. In: Kerkut, G.A., Gilbert, L.I. (Eds.), Comprehensive Insect Physiology, Biochemistry and Pharmacology, vol. 12. Pergamon, Oxford, pp. 647–712.

Agrawal, A.A., Vala, F., Sabelis, M.W., **2002**. Induction of preference and performance after acclimation to novel hosts in a phytophagous spider mite: adaptive plasticity? *Am. Nat. 159*, 553–565.

Ahmad, S., **1986**. Enzymatic adaptations of herbivorous insects and mites to phytochemicals. *J. Chem. Ecol. 12*, 533–560.

Ahmad, S., Brattsten, L.B., Mullin, C.A., Yu, S.J., **1986**. Enzymes involved in the metabolism of plant allelochemicals. In: Brattsten, L.B., Ahmad, S. (Eds.), Molecular Aspects of Insect–Plant Interactions. Plenum, New York, pp. 73–151.

Ahmad, S., Kirkland, K.E., Blomquist, G.J., **1987**. Evidence for a sex pheromone metabolizing cytochrome P-450 mono-oxygenase in the housefly. *Arch. Insect Biochem. Physiol. 6*, 21–40.

Amichot, M., Brun, A., Cuany, A., Helvig, C., Salaun, J.P., *et al*., **1994**. Expression study of CYP genes in *Drosophila* strains resistant or sensitive to insecticides. In: Lechner, M.C. (Ed.), Cytochrome P450. John Libbey Eurotext, Paris, pp. 689–692.

Amichot, M., Brun, A., Cuany, A., Souza, G.D., Mouel, T.L., *et al*., **1998**. Induction of cytochrome P450 activities in *Drosophila melanogaster* strains susceptible or resistant to insecticides. *Comp. Biochem. Physiol. 121C*, 311–319.

Amichot, M., Tares, S., Brun-Barale, A., Arthaud, L., Bride, J.M., *et al*., **2004**. Point mutations associated with insecticide resistance in the *Drosophila* cytochrome P450 CYP6A2 enable DDT metabolism. *Eur. J. Biochem. 271*, 1250–1257.

Anandatheerthavarada, H.K., Amuthan, G., Biswas, G., Robin, M.A., Murali, R., *et al*., **2001**. Evolutionary divergent electron donor proteins interact with P450MT2 through the same helical domain but different contact points. *EMBO J. 20*, 2394–2403.

Anandatheerthavarada, H.K., Vijayasarathy, C., Bhagwat, S.V., Biswas, G., Mullick, J., *et al*., **1999**. Physiological role of the N-terminal processed P4501A1 targeted to mitochondria in erythromycin metabolism and reversal of erythromycin-mediated inhibition of mitochondrial protein synthesis. *J. Biol. Chem. 274*, 6617–6625.

Andersen, J.F., Ceruso, M., Unnithan, G.C., Kuwano, E., Prestwich, G.D., *et al*., **1995**. Photoaffinity labeling of methyl farnesoate epoxidase in cockroach corpora allata. *Insect Biochem. Mol. Biol. 25*, 713–719.

Andersen, J.F., Utermohlen, J.G., Feyereisen, R., **1994**. Expression of housefly CYP6A1 and NADPH-cytochrome P450 reductase in *Escherichia coli* and

reconstitution of an insecticide-metabolizing P450 system. *Biochemistry 33*, 2171–2177.

Andersen, J.F., Walding, J.K., Evans, P.H., Bowers, W.S., Feyereisen, R., **1997**. Substrate specificity for the epoxidation of terpenoids and active site topology of housefly cytochrome P450 6A1. *Chem. Res. Toxicol. 10*, 156–164.

Appel, H.M., Martin, M., **1992**. Significance of metabolic load in the evolution of host specificity of *Manduca sexta*. *Ecology 73*, 216–228.

Aragon, S., Claudinot, S., Blais, C., Maibeche, M., Dauphin-Villemant, C., **2002**. Molting cycle-dependent expression of CYP4C15, a cytochrome P450 enzyme putatively involved in ecdysteroidogenesis in the crayfish, *Orconectes limosus*. *Insect Biochem. Mol. Biol. 32*, 153–159.

Arbeitman, M.N., Furlong, E.E., Imam, F., Johnson, E., Null, B.H., *et al.*, **2002**. Gene expression during the life cycle of *Drosophila melanogaster*. *Science 297*, 2270–2275.

Baldridge, G.D., Feyereisen, R., **1986**. Blood meal and cytochrome P-450 monooxygenases in the northern house mosquito, *Culex pipiens*. *Pestic. Biochem. Physiol. 25*, 407–413.

Barnes, H.J., Arlotto, M.P., Waterman, M.R., **1991**. Expression and enzymatic activity of recombinant cytochrome P450 17 alpha-hydroxylase in *Escherichia coli*. *Proc. Natl Acad. Sci. USA 88*, 5597–5601.

Bassett, M.H., McCarthy, J.L., Waterman, M.R., Sliter, T.J., **1997**. Sequence and developmental expression of Cyp18, a member of a new cytochrome P450 family from *Drosophila*. *Mol. Cell. Endocrinol. 131*, 39–49.

Berenbaum, M.R., **1990**. Evolution of specialization in insect-umbellifer associations. *Annu. Rev. Entomol. 35*, 319–343.

Berenbaum, M.R., **1995**. The chemistry of defense: theory and practice. *Proc. Natl Acad. Sci. USA 92*, 2–8.

Berenbaum, M.R., **1999**. Animal–plant warfare: molecular basis for cytochrome P450-mediated natural adaptation. In: Puga, A., Wallace, K.B. (Eds.), Molecular Biology of the Toxic Response. Taylor & Francis, Philadelphia, pp. 553–571.

Berenbaum, M.R., **2002**. Postgenomic chemical ecology: from genetic code to ecological interactions. *J. Chem. Ecol. 28*, 873–896.

Berenbaum, M.R., Cohen, M.B., Schuler, M.A., **1992**. Cytochrome P450 monooxygenase genes in oligophagous lepidoptera. *ACS Symp. Ser. 505*, 114–124.

Berenbaum, M.R., Favret, C., Schuler, M.A., **1996**. On defining key innovations in an adaptive radiation – cytochrome P450s and Papilionidae. *Am. Nat. 148*, S139–S155.

Berenbaum, M.R., Feeny, P., **1981**. Toxicity of angular furanocoumarins to swallowtails: escalation in the coevolutionary arms race. *Science 212*, 927–929.

Berenbaum, M.R., Neal, J.J., **1985**. Synergism between myristicin and xanthotoxin, a naturally co-occurring plant toxicant. *J. Chem. Ecol. 11*, 1349–1358.

Berenbaum, M.R., Neal, J.J., **1987**. Interactions among allelochemicals and insect resistance in crop plants. *ACS Symp. Ser. 330*, 416–430.

Bergé, J.B., Feyereisen, R., Amichot, M., **1998**. Cytochrome P450 monooxygenases and insecticide resistance in insects. *Philos. Trans. R. Soc. Lond. B Biol. Sci. 353*, 1701–1705.

Bernard, C.B., Philogene, B.J., **1993**. Insecticide synergists: role, importance, and perspectives. *J. Toxicol. Environ. Health 38*, 199–223.

Bernays, E., Graham, M., **1988**. On the evolution of host specificity in phytophagous arthropods. *Ecology 69*, 886–892.

Bernhardt, R., Gunsalus, I.C., **1992**. Reconstitution of cytochrome P4502B4 (LM2) activity with camphor and linalool monooxygenase electron donors. *Biochem. Biophys. Res. Commun. 187*, 310–317.

Bertok, B., Pap, L., Arvai, G., Bakonyvari, I., Kuruczne Ribai, Z., **2003**. Structure-activity relationship study of alkynyl ether insecticide synergists and the development of MB-599 (verbutin). *Pestic. Manag. Sci. 59*, 377–392.

Bhagwat, S.V., Biswas, G., Anandatheerthavarada, H.K., Addya, S., Pandak, W., *et al.*, **1999**. Dual targeting property of the N-terminal signal sequence of P4501A1. Targeting of heterologous proteins to endoplasmic reticulum and mitochondria. *J. Biol. Chem. 274*, 24014–24022.

Bigelow, S.W., Zijlstra, J.A., Vogel, E.W., Nebert, D.W., **1985**. Measurements of the cytosolic Ah receptor among four stains of *Drosophila melanogaster*. *Arch. Toxicol. 56*, 219–225.

Black, B.C., Hollingworth, R.M., Ahammadsahib, K.I., Kukel, C.D., Donovan, S., **1994**. Insecticidal action and mitochondrial uncoupling activity of AC303,630 and related halogenated pyrroles. *Pestic. Biochem. Physiol. 50*, 115–128.

Blais, C., Lafont, R., **1986**. Ecdysone 20-hydroxylation in imaginal wing discs of *Pieris brassicae* (Lepidoptera): correlations with ecdysone and 20-hydroxyecdysone titers in pupae. *Arch. Insect Biochem. Physiol. 3*, 501–512.

Blomquist, G.J., Dillwith, J.W., Pomonis, J.G., **1984**. Sex pheromone of the housefly. Metabolism of (Z)-9-tricosene to (Z)-9,10-epoxytricosane and (Z)-14-tricosen-10-one. *Insect Biochem. 14*, 279–284.

Bollenbacher, W.E., Smith, S.L., Wielgus, J.J., Gilbert, L.I., **1977**. Evidence for an α-ecdysone cytochrome P-450 mixed function oxidase in insect fat body mitochondria. *Nature 268*, 660–663.

Borras, T., Morozova, T.V., Heinsohn, S.L., Lyman, R.F., Mackay, T.F., *et al.*, **2003**. Transcription profiling in *Drosophila* eyes that overexpress the human glaucoma-associated trabecular meshwork-inductible glucocorticoid response protein/myocilin (TIGR/MYOC). *Genetics 163*, 637–645.

Bowers, W.S., Ohta, T., Cleere, J.S., Marsella, P.A., **1976**. Discovery of insect anti-juvenile hormones in plants. *Science 193*, 542–547.

Bradfield, J.Y., Lee, Y.H., Keeley, L.L., **1991**. Cytochrome P450 family 4 in a cockroach: molecular cloning and regulation by hypertrehalosemic hormone. *Proc. Natl Acad. Sci. USA 88*, 4558–4562.

Brandt, A., Scharf, M., Pedra, J.H., Holmes, G., Dean, A., *et al.*, **2002**. Differential expression and induction of two *Drosophila* cytochrome P450 genes near the Rst(2)DDT locus. *Insect Mol. Biol. 11*, 337–341.

Brattsten, L.B., **1979a**. Ecological significance of mixed-function oxidations. *Drug Metab. Rev. 10*, 35–58.

Brattsten, L.B., **1979b**. Biochemical defense mechanisms in herbivores against plant allelochemicals. In: Rosenthal, G.A., Janzen, D.H. (Eds.), Herbivores, their Interaction with Secondary Plant Metabolites. Academic Press, New York, pp. 199–270.

Brattsten, L.B., **1983**. Cytochrome P-450 involvement in the interactions between plant terpenes and insect herbivores. In: Hedin, P.A. (Ed.), Plant Resistance to Insects. American Chemical Society, pp. 173–195.

Brattsten, L.B., **1987**. Sublethal virus infection depresses cytochrome P-450 in an insect. *Experientia 43*, 451–454.

Brattsten, L.B., Metcalf, R.L., **1970**. The synergistic ratio of carbaryl with piperonyl butoxide as an indicator of the distribution of multifunction oxidases in the insecta. *J. Econ. Entomol. 63*, 101–104.

Brattsten, L.B., Wilkinson, C.F., **1977**. Insecticide solvents: interference with insecticidal action. *Science 196*, 1211–1213.

Brattsten, L.B., Wilkinson, C.F., Eisner, T., **1977**. Herbivore–plant interactions: mixed-function oxidases and secondary plant substances. *Science 196*, 1349–1352.

Bride, J.M., Cuany, A., Amichot, M., Brun, A., Babault, M., *et al.*, **1997**. Cytochrome P-450 field insecticide tolerance and development of laboratory resistance in grape vine populations of *Drosophila melanogaster* (Diptera: Drosophilidae). *J. Econ. Entomol. 90*, 1514–1520.

Brindley, W.A., **1977**. Synergist differences as an alternate interpretation of carbaryl-piperonyl butoxide toxicity data. *Environ. Entomol. 6*, 885–888.

Brogdon, W.G., **1997**. Heme peroxidase activity measured in single mosquitoes identifies individuals expressing an elevated oxidase for insecticide resistance. *J. Am. Mosq. Control Assoc. 13*, 233–237.

Brooks, G.T., **1979**. The metabolism of xenobiotics in insects. In: Bridges, J.W., Chasseaud, L. (Eds.), Progress in Drug Metabolism. Wiley, London, pp. 151–214.

Brooks, G.T., Pratt, G.E., Mace, D.W., Cocks, J.A., **1985**. Inhibitors of juvenile hormone biosynthesis in *corpora allata* of the cockroach *Periplaneta amaericana* (L.) *in vitro*. *Pestic. Sci. 16*, 132–142.

Brown, T.M., Bryson, P.K., Payne, G.T., **1996**. Synergism by propynyl aryl ethers in permethrin-resistant tobacco budworm larvae. *Pestic. Sci. 46*, 323–331.

Brun, A., Cuany, A., Le Mouel, T., Berge, J., Amichot, M., **1996**. Inducibility of the *Drosophila melanogaster* cytochrome P450 gene, CYP6A2, by phenobarbital in insecticide susceptible or resistant strains. *Insect Biochem. Mol. Biol. 26*, 697–703.

Bull, D.L., Ivie, G.W., Beier, R.C., Pryor, N.W., **1986**. In vitro metabolism of a linear furanocoumarin (8-methoxypsoralen, xanthotoxin) by mixed-function oxidases of larvae of black swallowtail butterfly and fall armyworm. *J. Chem. Ecol. 12*, 885–892.

Campbell, P.M., Newcomb, R.D., Russell, R.J., Oakeshott, J.G., **1998**. Two different amino acid substitutions in the ali-esterase, E3, confer alternative types of organophosphorus insecticide resistance in the sheep blowfly, *Lucilia cuprina*. *Insect Biochem. Mol. Biol. 28*, 139–150.

Capdevila, J.H., Harris, R.C., Falck, J.R., **2002**. Microsomal cytochrome P450 and eicosanoid metabolism. *Cell. Mol. Life Sci. 59*, 780–789.

Cariño, F., Koener, J.F., Plapp, F.W., Jr., Feyereisen, R., **1992**. Expression of the cytochrome P450 gene CYP6A1 in the housefly, *Musca domestica*. *ACS Symp. Ser. 505*, 31–40.

Cariño, F.A., Koener, J.F., Plapp, F.W., Jr., Feyereisen, R., **1994**. Constitutive overexpression of the cytochrome P450 gene CYP6A1 in a housefly strain with metabolic resistance to insecticides. *Insect Biochem. Mol. Biol. 24*, 411–418.

Casida, J.E., **1970**. Mixed-function oxidase involvement in the biochemistry of insecticide synergists. *J. Agric. Food Chem. 18*, 753–771.

Celniker, S.E., Wheeler, D.A., Kronmiller, B., Carlson, J.W., Halpern, A., *et al.*, **2002**. Finishing a whole-genome shotgun: Release 3 of the *Drosophila melanogaster* euchromatic genome sequence. *Genome Biol. 3*, 79.1–79.14.

Chavez, V.M., Marques, G., Delbecque, J.P., Kobayashi, K., Hollingsworth, M., *et al.*, **2000**. The *Drosophila* disembodied gene controls late embryonic morphogenesis and codes for a cytochrome P450 enzyme that regulates embryonic ecdysone levels. *Development 127*, 4115–4126.

Chawla, A., Repa, J.J., Evans, R.M., Mangelsdorf, D.J., **2001**. Nuclear receptors and lipid physiology: opening the X-files. *Science 294*, 1866–1870.

Chen, J.S., Berenbaum, M.R., Schuler, M.A., **2002**. Amino acids in SRS1 and SRS6 are critical for furanocoumarin metabolism by CYP6B1v1, a cytochrome P450 monooxygenase. *Insect Mol. Biol. 11*, 175–186.

Chen, C.D., Doray, B., Kemper, B., **1998**. A conserved proline-rich sequence between the N-terminal signal-anchor and catalytic domains is required for assembly of functional cytochrome P450 2C2. *Arch. Biochem. Biophys. 350*, 233–238.

Chen, J.H., Hara, T., Fisher, M.J., Rees, H.H., **1994**. Immunological analysis of developmental changes in ecdysone 20-monooxygenase expression in the cotton leafworm, *Spodoptera littoralis*. *Biochem. J. 299*, 711–717.

Cianfrogna, J.A., Zangerl, A.R., Berenbaum, M.R., **2002**. Dietary and developmental influences on induced

detoxification in an oligophage. *J. Chem. Ecol. 28*, 1349–1364.

Cilek, J.E., Dahlman, D.L., Knapp, F.W., **1995**. Possible mechanism of diazinon negative cross-resistance in pyrethroid-resistant horn flies (Diptera: Muscidae). *J. Econ. Entomol. 88*, 520–524.

Claridge-Chang, A., Wijnen, H., Naef, F., Boothroyd, C., Rajewsky, N., *et al.*, **2001**. Circadian regulation of gene expression systems in the *Drosophila* head. *Neuron 32*, 657–671.

Clarke, S.E., Brealey, C.J., Gibson, G.G., **1989**. Cytochrome P-450 in the housefly: induction, substrate specificity and comparison to three rat hepatic isoenzymes. *Xenobiotica 19*, 1175–1180.

Claudianos, C., Crone, E., Coppin, C., Russell, R., Oakeshott, J.G., **2001**. A genomics perspective on mutant aliesterases and metabolic resistance to organophosphates. In: Clark, J.M., Yamaguchi, I. (Eds.), Agrochemical Resistance: Extent, Mechanism and Detection. Oxford Press, Washington, pp. 90–101.

Claudianos, C., Russell, R.J., Oakeshott, J.G., **1999**. The same amino acid substitution in orthologous esterases confers organophosphate resistance on the housefly and a blowfly. *Insect Biochem. Mol. Biol. 29*, 675–686.

Cohen, M.B., Berenbaum, M.R., Schuler, M.A., **1989**. Induction of cytochome P450-mediated detoxification of xanthotoxin in the black swallowtail. *J. Chem. Ecol. 15*, 2347–2355.

Cohen, M.B., Feyereisen, R., **1995**. A cluster of cytochrome P450 genes of the CYP6 family in the housefly. *DNA Cell Biol. 14*, 73–82.

Cohen, M.B., Koener, J.F., Feyereisen, R., **1994**. Structure and chromosomal localization of CYP6A1, a cytochrome P450-encoding gene from the housefly. *Gene 146*, 267–272.

Cohen, M.B., Schuler, M.A., Berenbaum, M.R., **1992**. A host-inducible cytochrome P-450 from a host-specific caterpillar: molecular cloning and evolution. *Proc. Natl Acad. Sci. USA 89*, 10920–10924.

Couillaud, F., Debernard, S., Darrouzet, E., Rossignol, F., **1996**. Hidden face of juvenile hormone metabolism in the African locust. *Arch. Insect Biochem. Physiol. 32*, 387–397.

Crampton, A.L., Baxter, G.D., Barker, S.C., **1999**. Identification and characterisation of a cytochrome P450 gene and processed pseudogene from an arachnid: the cattle tick, *Boophilus microplus*. *Insect Biochem. Mol. Biol. 29*, 377–384.

Crankshaw, D.L., Hetnarski, K., Wilkinson, C.F., **1981**. The functional role of NADPH-cytochrome *c* reductase in southern armyworm (*Spodoptera eridania*) midgut microsomes. *Insect Biochem. 11*, 515–522.

Cuany, A., Helvig, C., Amichot, M., Pflieger, P., Mioskowski, C., *et al.*, **1995**. Fate of a terminal olefin with *Drosophila* microsomes and its inhibitory effects on some P-450 dependent activities. *Arch. Insect Biochem. Physiol. 28*, 325–338.

Cuany, A., Pralavorio, M., Pauron, D., Berge, J.B., Fournier, D., *et al.*, **1990**. Characterization of microsomal oxidative activities in a wild-type and in a DDT resistant strain of *Drosophila melanogaster*. *Pestic. Biochem. Physiol. 37*, 293–302.

Daborn, P., Boundy, S., Yen, J., Pittendrigh, B., ffrench-Constant, R., **2001**. DDT resistance in *Drosophila* correlates with *Cyp6g1* over-expression and confers cross-resistance to the neonicotinoid imidacloprid. *Mol. Genet. Genomics 266*, 556–563.

Daborn, P.J., Yen, J.L., Bogwitz, M.R., Le Goff, G., Feil, E., *et al.*, **2002**. A single P450 allele associated with insecticide resistance in *Drosophila*. *Science 297*, 2253–2256.

Danielson, P.B., Fogleman, J.C., **1997**. Isolation and sequence analysis of cytochrome P450 12B1: the first mitochondrial insect P450 with homology to 1 alpha, 25 dihydroxy-D3 24-hydroxylase. *Insect Biochem. Mol. Biol. 27*, 595–604.

Danielson, P.B., Foster, J.L., Cooper, S.K., Fogleman, J.C., **1999**. Diversity of expressed cytochrome P450 genes in the adult Mediterranean fruit fly, *Ceratitis capitata*. *Insect Mol. Biol. 8*, 149–159.

Danielson, P.B., Foster, J.L., McMahill, M.M., Smith, M.K., Fogleman, J.C., **1998**. Induction by alkaloids and phenobarbital of Family 4 Cytochrome P450s in *Drosophila*: evidence for involvement in host plant utilization. *Mol. Gen. Genet. 259*, 54–59.

Danielson, P.B., Letman, J.A., Fogleman, J.C., **1995**. Alkaloid metabolism by cytochrome P-450 enzymes in *Drosophila melanogaster*. *Comp. Biochem. Physiol. 110B*, 683–688.

Danielson, P.B., MacIntyre, R.J., Fogleman, J.C., **1997**. Molecular cloning of a family of xenobiotic-inducible drosophilid cytochrome P450s: evidence for involvement in host–plant allelochemical resistance. *Proc. Natl Acad. Sci. USA 94*, 10797–10802.

Dapkus, D., **1992**. Genetic localization of DDT resistance in *Drosophila melanogaster* (Diptera: Drosophilidae). *J. Econ. Entomol. 85*, 340–347.

Darrouzet, E., Mauchamp, B., Prestwich, G.D., Kerhoas, L., Ujvary, I., *et al.*, **1997**. Hydroxy juvenile hormones: new putative juvenile hormones biosynthesized by locust corpora allata *in vitro*. *Biochem. Biophys. Res. Commun. 240*, 752–758.

Darvas, B., Rees, H.H., Hoggard, N., **1993**. Ecdysone 20-monooxygenase systems in flesh-flies (Diptera: Sarcophagidae), *Neobellieria bullata* and *Parasarcophaga argyrostoma*. *Comp. Biochem. Physiol. 105B*, 765–773.

Darvas, B., Rees, H.H., Hoggard, N., Tag El-Din, M.H., Kuwano, E., *et al.*, **1992**. Cytrochrome P-450 inducers and inhibitors interfering with ecdysone 20-monooxygenases and their activities during postembryonic development of *Neobellieria bullata* Parker. *Pestic. Sci. 36*, 135–142.

David, J.P., Rey, D., Cuany, A., Amichot, M., Meyran, J.C., **2000**. Comparative ability to detoxify alder leaf

litter in field larval mosquito collections. *Arch. Insect Biochem. Physiol. 44*, 143–150.

David, J.P., Rey, D., Cuany, A., Bride, J.M., Meyran, J.C., **2002**. Larvicidal properties of decomposed leaf litter in the subalpine mosquito breeding sites. *Environ. Toxicol. Chem. 21*, 62–66.

Davis, R.H., Nahrstedt, A., **1987**. Biosynthesis of cyanogenic glucosides in butterflies and moths. *Insect Biochem. 17*, 689–693.

De Gregorio, E., Spellman, P.T., Rubin, G.M., Lemaitre, B., **2001**. Genome-wide analysis of the *Drosophila* immune response by using oligonucleotide microrrays. *Proc. Natl Acad. Sci. USA 98*, 12590–12595.

De Gregorio, E., Spellman, P.T., Tzou, P., Rubin, G.M., Lemaitre, B., **2002**. The Toll and Imd pathways are the major regulators of the immune response in *Drosophila*. *EMBO J. 21*, 2568–2579.

Delpuech, J.M., Aquadro, C.F., Roush, R.T., **1993**. Non-involvement of the long terminal repeat of transposable element 17.6 in insecticide resistance in *Drosophila*. *Proc. Natl Acad. Sci. USA 90*, 5643–5647.

De Mot, R., Parret, A.H.A, **2002**. A novel class of self-sufficient cytochrome P450 monooxygenases in prokaryotes. *Trends Microbiol. 10*, 502–508.

Denison, M.S., Nagy, S.R., **2003**. Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals. *Ann. Rev. Pharmacol. Toxicol. 43*, 309–334.

de Sousa, G., Cuany, A., Brun, A., Amichot, M., Rahmani, R., *et al.*, **1995**. A microfluorometric method for measuring ethoxycoumarin-O-deethylase activity on individual *Drosophila melanogaster* abdomens: interest for screening resistance in insect populations. *Anal. Biochem. 229*, 86–91.

Dethier, V.G., **1954**. Evolution of feeding preferences in phytophagous insects. *Evolution 8*, 33–54.

Domanski, T.L., Halpert, J.R., **2001**. Analysis of mammalian cytochrome P450 structure and function by site-directed mutagenesis. *Curr. Drug Metab. 2*, 117–137.

Dombrowski, S.M., Krishnan, R., Witte, M., Maitra, S., Diesing, C., *et al.*, **1998**. Constitutive and barbital-induced expression of the *Cyp6a2* allele of a high producer strain of CYP6A2 in the genetic background of a low producer strain. *Gene 221*, 69–77.

Dowd, P.F., Smith, C.M., Sparks, T.C., **1983**. Detoxification of plant toxins by insects. *Insect Biochem. 13*, 453–468.

Drabek, J., Neumann, R., **1985**. Proinsecticides. In: Hutson, D.H., Roberts, T.R. (Eds.), Insecticides. Wiley, London, pp. 35–86.

Dunkov, B.C., Guzov, V.M., Mocelin, G., Shotkoski, F., Brun, A., *et al.*, **1997**. The *Drosophila* cytochrome P450 gene *Cyp6a2*: structure, localization, heterologous expression, and induction by phenobarbital. *DNA Cell Biol. 16*, 1345–1356.

Dunkov, B.C., Rodriguez-Arnaiz, R., Pittendrigh, B., ffrench-Constant, R.H., Feyereisen, R., **1996**. Cytochrome P450 gene clusters in *Drosophila melanogaster*. *Mol. Gen. Genet. 251*, 290–297.

Durham, E.W., Siegfried, B.D., Scharf, M.E., **2002**. *In vivo* and *in vitro* metabolism of fipronil by larvae of the European corn borer *Ostrinia nubilalis*. *Pest Manag. Sci. 58*, 799–804.

Ehrlich, P.R., Raven, P.H., **1964**. Butterflies and plants: a study in coevolution. *Evolution 18*, 586–608.

Estabrook, R.W., **1996**. The remarkable P450s: a historical overview of these versatile hemeprotein catalysts. *FASEB J. 10*, 202–204.

Feng, R., Houseman, J.G., Downe, A.E.R., **1992**. Effect of ingested meridic diet and corn leaves on midgut detoxification processes in the European corn borer, *Ostrinia nubilalis*. *Pestic. Biochem. Physiol. 42*, 203–210.

Feyereisen, R., **1977**. Cytochrome P-450 et hydroxylation de l'ecdysone en ecdysterone chez *Locusta migratoria*. *C.R. Acad. Sci. Paris 284*, 1831–1834.

Feyereisen, R., **1983**. Polysubstrate monooxygenases (cytochrome P-450) in larvae of susceptible and resistant strains of house flies. *Pestic. Biochem. Physiol. 19*, 262–269.

Feyereisen, R., **1999**. Insect P450 enzymes. *Annu. Rev. Entomol. 44*, 507–533.

Feyereisen, R., Baldridge, G.D., Farnsworth, D.E., **1985**. A rapid method for preparing insect microsomes. *Comp. Biochem. Physiol. 82B*, 559–562.

Feyereisen, R., Durst, F., **1978**. Ecdysterone biosynthesis: a microsomal cytochrome-P-450-linked ecdysone 20-monooxygenase from tissues of the African migratory locust. *Eur. J. Biochem. 88*, 37–47.

Feyereisen, R., Durst, F., **1980**. Development of microsomal cytochrome P-450 monooxygenases during the last larval instar of the locust, *Locusta migratoria*: correlation with the hemolymph 20-hydroxyecdysone titer. *Mol. Cell. Endocrinol. 20*, 157–169.

Feyereisen, R., Farnsworth, D.E., **1985**. Developmental changes of microsomal cytochrome P-450 monooxygenases in larval and adult *Diploptera punctata*. *Insect Biochem. 15*, 755–761.

Feyereisen, R., Koener, J.F., Cariño, F.A., Daggett, A.S., **1990**. Biochemistry and molecular biology of insect cytochrome P450. In: Hagedorn, H.H., Hildebrand, J.G., Kidwell, M.G., Law, J.H. (Eds.), Molecular Insect Science. Plenum, New York, pp. 263–272.

Feyereisen, R., Koener, J.F., Farnsworth, D.E., Nebert, D.W., **1989**. Isolation and sequence of cDNA encoding a cytochrome P-450 from an insecticide-resistant strain of the housefly, *Musca domestica*. *Proc. Natl Acad. Sci. USA 86*, 1465–1469.

Feyereisen, R., Langry, K.C., Ortiz de Montellano, P.R., **1984**. Self-catalyzed destruction of insect cytochrome P-450. *Insect Biochem. 14*, 19–26.

Feyereisen, R., Pratt, G.E., Hamnett, A.F., **1981**. Enzymic synthesis of juvenile hormone in locust corpora allata: evidence for a microsomal cytochrome P-450 linked methyl farnesoate epoxidase. *Eur. J. Biochem. 118*, 231–238.

Feyereisen, R., Vincent, D.R., **1984**. Characterization of antibodies to housefly NADPH-cytochrome P-450 reductase. *Insect Biochem. 14*, 163–168.

ffrench-Constant, R.H., Park, Y., Feyereisen, R., **1999**. Molecular biology of insecticide resistance. In: Puga, A., Wallace, K.B. (Eds.), Molecular Biology of the Toxic Response. Taylor & Francis, Philadelphia, pp. 533–551.

Fisher, T., Crane, M., Callaghan, A., **2003**. Induction of cytochrome P450 activity in individual *Chironomus riparius* Meigen larvae exposed to xenobiotics. *Ecotoxicol. Environ. Saf. 54*, 1–6.

Fogleman, J.C., **2000**. Response of *Drosophila melanogaster* to selection for P450-mediated resistance to isoquinoline alkaloids. *Chem. Biol. Interact. 125*, 93–105.

Fogleman, J.C., Danielson, P.B., **2000**. Analysis of fragment homology among DNA sequences from cytochrome P450 families 4 and 6. *Genetica 110*, 257–265.

Fogleman, J.C., Danielson, P.B., MacLntyre, R.J., **1998**. The molecular basis of adaptation in *Drosophila*: the role of cytochrome P450s. In: Hecht, M., MacIntyre, R., Clegg, M. (Eds.), Evolutionary Biology, vol. 30. Plenum, New York, pp. 15–77.

Fraenkel, G.S., **1959**. The raison d'etre of secondary plant substances. *Science 129*, 1466–1470.

Frank, M.R., Fogleman, J.C., **1992**. Involvement of cytochrome P450 in host–plant utilization by Sonoran Desert *Drosophila*. *Proc. Natl Acad. Sci. USA 89*, 11998–12002.

Freeman, M.R., Dobritsa, A., Gaines, P., Segraves, W.A., Carlson, J.R., **1999**. The dare gene: steroid hormone production, olfactory behavior, and neural degeneration in *Drosophila*. *Development 126*, 4591–4602.

Frolov, M.V., Alatortsev, V.E., **1994**. Cluster of cytochrome P450 genes on the X chromosome of *Drosophila melanogaster. DNA Cell Biol. 13*, 663–668.

Fuchs, S., Spiegelman, V.S., Belitsky, G.A., **1993**. The effect of the cytochrome P-450 system inducers on the development of *Drosophila melanogaster. J. Biochem. Toxicol. 8*, 83–88.

Fuchs, S.Y., Spiegelman, V.S., Belitsky, G.A., **1994**. Inducibility of various cytochrome P450 isozymes by phenobarbital and some other xenobiotics in *Drosophila melanogaster. Biochem. Pharmacol. 47*, 1867–1873.

Fujii, S., Amrein, H., **2002**. Genes expressed in the *Drosophila* head reveal a role for fat cells in sex-specific physiology. *EMBO J. 21*, 5353–5363.

Fuji-Kuriyama, Y., Mizukami, Y., Kawajiri, K., Sogawa, K., Muramatsu, M., **1982**. Primary structure of a cytochrome P-450: coding nucleotide sequence of phenobarbital-inducible cytochrome P-450 cDNA from rat liver. *Proc. Natl. Acad. Sci. USA 79*, 2793–2797.

Furlong, E.E., Andersen, E.C., Null, B., White, K.P., Scott, M.P., **2001**. Patterns of gene expression during *Drosophila* mesoderm development. *Science 293*, 1629–1633.

Gandhi, R., Varak, E., Goldberg, M.L., **1992**. Molecular analysis of a cytochrome P450 gene of family 4 on the *Drosophila* X chromosome. *DNA Cell Biol. 11*, 397–404.

Glendinning, J.I., Slansky, F., Jr., **1995**. Consumption of a toxic food by caterpillars increases with dietary exposure: support for a role of induced detoxification enzymes. *J. Comp. Physiol. A 176*, 337–345.

Gordon, H.T., **1961**. Nutritional factors in insect resistance to chemicals. *Annu. Rev. Entomol. 6*, 27–54.

Gotoh, O., **1992**. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J. Biol. Chem. 267*, 83–90.

Gotoh, O., **1993**. Evolution and differentiation of P-450 genes. In: Omura, T., Ishimura, Y., Fuji-Kuriyama, Y. (Eds.), Cytochrome P-450, 2nd edn. Kodansha, Tokyo, pp. 255–272.

Gotoh, O., **1998**. Divergent structures of *Caenorhabditis elegans* cytochrome P450 genes suggest the frequent loss and gain of introns during the evolution of nematodes. *Mol. Biol. Evol. 15*, 1447–1459.

Gould, F., **1984**. Mixed function oxidases and herbivore polyphagy: the devil's advocate position. *Ecol. Entomol. 9*, 29–34.

Greenwood, D.R., Rees, H.H., **1984**. Ecdysone 20-monooxygenase in the desert locust, *Schistocerca gregaria*. *Biochem. J. 223*, 837–847.

Grieneisen, M.L., Warren, J.T., Gilbert, L.I., **1993**. Early steps in ecdysteroid biosynthesis: evidence for the involvement of cytochrome P-450 enzymes. *Insect Biochem. Mol. Biol. 23*, 13–23.

Gunderson, C.A., Brattsten, L.B., Fleming, J.T., **1986**. Microsomal oxidase and glutathione transferase as factors influencing the effects of pulegone in southern and fall armyworm larvae. *Pestic. Biochem. Physiol. 26*, 238–249.

Guzov, V.M., Houston, H.L., Murataliev, M.B., Walker, F.A., Feyereisen, R., **1996**. Molecular cloning, overexpression in *Escherichia coli*, structural and functional characterization of housefly cytochrome b_5. *J. Biol. Chem. 271*, 26637–26645.

Guzov, V.M., Unnithan, G.C., Chernogolov, A.A., Feyereisen, R., **1998**. CYP12A1, a mitochondrial cytochrome P450 from the housefly. *Arch. Biochem. Biophys. 359*, 231–240.

Hahn, M.E., **2002**. Aryl hydrocarbon receptors: diversity and evolution. *Chem. Biol. Interact. 141*, 131–160.

Hainzl, D., Casida, J.E., **1996**. Fipronil insecticide: novel photochemical desulfinylation with retention of neurotoxicity. *Proc. Natl Acad. Sci. USA 93*, 12764–12767.

Hainzl, D., Cole, L.M., Casida, J.E., **1998**. Mechanisms for selective toxicity of fipronil insecticide and its sulfone metabolite and desulfinyl photoproduct. *Chem. Res. Toxicol. 11*, 1529–1535.

Halliday, W.R., Farnsworth, D.E., Feyereisen, R., **1986**. Hemolymph ecdysteroid titer and midgut ecdysone 20-monooxygenase activity during the last larval stage of *Diploptera punctata*. *Insect Biochem. 16*, 627–634.

Hallstrom, I., **1985**. Genetic regulation of the cytochrome P-450 system in *Drosophila melanogaster*: II. Localization of some genes regulating cytochrome P-450 activity. *Chem. Biol. Interact. 56*, 173–184.

Hallstrom, I., **1987**. Genetic variation in cytochrome P-450-dependent demethylation in *Drosophila melanogaster*. *Biochem. Pharmacol. 36*, 2279–2282.

Hallstrom, I., Blanck, A., **1985**. Genetic regulation of the cytochrome P-450 system in *Drosophila melanogaster*: I. Chromosomal determination of some cytochrome P-450-dependent reactions. *Chem. Biol. Interact. 56*, 157–171.

Hallstrom, I., Grafstrom, R., **1981**. The metabolism of drugs and carcinogens in isolated subcellular fractions of *Drosophila melanogaster*. II. Enzyme induction and metabolism of benzo[*a*]pyrene. *Chem. Biol. Interact. 34*, 145–159.

Hallstrom, I., Magnusson, J., Ramel, C., **1982**. Relation between the somatic toxicity of dimethylnitrosamine and a genetically determined variation in the level and induction of cytochrome P450 in *Drosophila melanogaster*. *Mutat. Res. 92*, 161–168.

Halpern, M.E., Morton, R.A., **1987**. Reproductive and developmental defects in a malathion-resistant laboratory-selected population of *Drosophila melanogaster*. *Pestic. Biochem. Physiol. 28*, 44–56.

Hammock, B.D., **1975**. NADPH dependent epoxidation of methyl farnesoate to juvenile hormone in the cockroach *Blaberus giganteus* L. *Life Sci. 17*, 323–328.

Hammock, B.D., **1985**. Regulation of juvenile hormone titer: degradation. In: Kerkut, G.A., Gilbert, L.I. (Eds.), Comprehensive Insect Physiology, Biochemistry and Pharmacology, vol. 7. Pergamon, Oxford, pp. 431–472.

Hammock, B.D., Mumby, S.M., **1978**. Inhibition of epoxidation of methyl farnesoate to juvenile hormone III by cockroach corpus allatum homogenates. *Pestic. Biochem. Physiol. 9*, 39–47.

Harrison, P.M., Milburn, D., Zhang, Z., Bertone, P., Gerstein, M., **2003**. Identification of pseudogenes in the *Drosophila melanogaster* genome. *Nucleic Acids Res. 31*, 1033–1037.

Harrison, T.L., Zangerl, A.R., Schuler, M.A., Berenbaum, M.R., **2001**. Developmental variation in cytochrome P450 expression in *Papilio polyxenes* in response to xanthotoxin, a hostplant allelochemical. *Arch. Insect Biochem. Physiol. 48*, 179–189.

Harwood, S.H., Moldenke, A.F., Berry, R.E., **1990**. Toxicity of peppermint monoterpenes to the variegated cutworm, *Peridroma saucia* Hubner (Lepidoptera: Noctuidae). *J. Econ. Entomol. 83*, 1761–1767.

Hasemann, C.A., Kurumbail, R.G., Boddupalli, S.S., Peterson, J.A., Deisenhofer, J., **1995**. Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Structure 3*, 41–62.

Hatano, R., Scott, J.G., **1993**. Anti-P450lpr antiserum inhibits the activation of chlorpyrifos to chlorpyrifos oxon in housefly microsomes. *Pestic. Biochem. Physiol. 45*, 228–233.

Hatano, R., Scott, J.G., Dennehy, T.J., **1992**. Enhanced activation is the mechanism of negative cross-resistance to chlorpyriphos in the Dicofol-IR strain of *Tetranychus urticae* (Acari: Tetranychidae). *J. Econ. Entomol. 85*, 1088–1091.

He, H., Chen, A.C., Davey, R.B., Ivie, G.W., **2002**. Molecular cloning and nucleotide sequence of a new P450 gene, *CYP319A1*, from the cattle tick, *Boophilus microplus*. *Insect Biochem. Mol. Biol. 32*, 303–309.

Hehn, A., Morant, M., Werck-Reichhart, D., **2002**. Partial recoding of P450 and P450 reductase cDNAs for improved expression in yeast and plants. *Methods in Enzymol. 357*, 343–351.

Helvig, C., Koener, J.F., Unnithan, G.C., Feyereisen, R., **2004**. CYP15A1, the cytochrome P450 that catalyzes epoxidation of methyl farnesoate to juvenile hormone III in cockroach corpora allata. *Proc. Natl Acad. Sci. USA 101*, 4024–4029.

Hickey, D.A., Bally-Cuif, L., Abukashawa, S., Patyant, V., Benkel, B.F., **1991**. Concerted evolution of duplicated protein coding genes in *Drosophila*. *Proc. Natl Acad. Sci. USA 88*, 1611–1615.

Hodgson, E., **1985**. Microsomal monooxygenases. In: Kerkut, G.A., Gilbert, L.I. (Eds.), Comprehensive Insect Physiology, Biochemistry and Pharmacology, vol. 11. Pergamon, Oxford, pp. 225–331.

Hoggard, N., Fisher, M.J., Rees, H.H., **1989**. Possible role for covalent modification in the reversible activation of ecdysone 20-monooxygenase activity. *Arch. Insect Biochem. Physiol. 10*, 241–253.

Hoggard, N., Rees, H.H., **1988**. Reversible activation-inactivation of mitochondrial ecdysone 20-mono-oxygenase: a possible role for phosphorylation-dephosphorylation. *J. Insect Physiol. 34*, 647–653.

Hold, K.M., Sirisoma, N.S., Casida, J.E., **2001**. Detoxification of alpha- and beta-Thujones (the active ingredients of absinthe): site specificity and species differences in cytochrome P450 oxidation in vitro and in vivo. *Chem. Res. Toxicol. 14*, 589–595.

Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., *et al.*, **2002**. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science 298*, 129–149.

Holzkamp, G., Nahrstedt, A., **1994**. Biosynthesis of cyanogenic glucosides in the lepidoptera. Incorporation of [U-14C]-2-methylpropanealdoxime, 2S-[U-14C]-nethylbutanealdoxime and -[U-14C]-N-hydroxyisoleucine into linamarin and lotaustralin by the larvae of *Zygaena trifolli*. *Insect Biochem. Mol. Biol. 24*, 161–165.

Honkakoski, P., Negishi, M., **2000**. Regulation of cytochrome P450 (CYP) genes by nuclear receptors. *Biochem. J. 347*, 321–337.

Honkakoski, P., Sueyoshi, T., Negishi, M., **2003**. Drug-activated nuclear receptors CAR and PXR. *Ann. Med. 35*, 172–182.

Horike, N., Sonobe, H., **1999**. Ecdysone 20-monooxygenase in eggs of the silkworm, *Bombyx mori*: enzymatic properties and developmental changes. *Arch. Insect Biochem. Physiol. 41*, 9–17.

Horike, N., Takemori, H., Nonaka, Y., Sonobe, H., Okamoto, M., **2000**. Molecular cloning of NADPH-

cytochrome P450 oxidoreductase from silkworm eggs. Its involvement in 20-hydroxyecdysone biosynthesis during embryonic development. *Eur. J. Biochem. 267*, 6914–6920.

Houpt, D.R., Pursey, J.C., Morton, R.A., **1988**. Genes controlling malathion resistance in a laboratory-selected population of *Drosophila melanogaster. Genome 30*, 844–853.

Hovemann, B.T., Sehlmeyer, F., Malz, J., **1997**. *Drosophila melanogaster* NADPH-cytochrome P450 oxidoreductase: pronounced expression in antennae may be related to odorant clearance. *Gene 189*, 213–219.

Hughes, P.B., Devonshire, A.L., **1982**. The biochemical basis of resistance to organophosphorus insecticides in the sheep blowfly, *Lucilia cuprina*. *Pestic. Biochem. Physiol. 18*, 289–297.

Hung, C.F., Berenbaum, M.R., Schuler, M.A., **1997**. Isolation and characterization of CYP6B4, a furanocoumarin-inducible cytochrome P450 from a polyphagous caterpillar (Lepidoptera:papilionidae). *Insect Biochem. Mol. Biol. 27*, 377–385.

Hung, C.F., Harrison, T.L., Berenbaum, M.R., Schuler, M.A., **1995a**. CYP6B3: a second furanocoumarin–inducible cytochrome P450 expressed in *Papilio polyxenes*. *Insect Mol. Biol. 4*, 149–160.

Hung, C.F., Holzmacher, R., Connolly, E., Berenbaum, M.R., Schuler, M.A., **1996**. Conserved promoter elements in the CYP6B gene family suggest common ancestry for cytochrome P450 monooxygenases mediating furanocoumarin detoxification. *Proc. Natl Acad. Sci. USA 93*, 12200–12205.

Hung, C.F., Prapaipong, H., Berenbaum, M.R., Schuler, M.A., **1995b**. Differential induction of cytochrome P-450 transcripts in *Papilio polyxenes* by linear and angular furanocoumarins. *Insect Biochem. Mol. Biol. 25*, 89–99.

Hunt, D.W.A., Smirle, M.J., **1988**. Partial inhibition of pheromone production in *Dendroctonus ponderosae* (Coleoptera: Scolytidae) by polysubstrate monooxygenase inhibitors. *J. Chem. Ecol. 14*, 529–536.

Hurban, P., Thummel, C.S., **1993**. Isolation and characterization of fifteen ecdysone-inducible *Drosophila* genes reveal unexpected complexities in ecdysone regulation. *Mol. Cell Biol. 13*, 7101–7111.

Ikezawa, N., Morisaki, M., Fujimoto, Y., **1993**. Sterol metabolism in insects: dealkylation of phytosterol to cholesterol. *Acc. Chem. Res. 26*, 139–146.

Ikezawa, N., Tanaka, M., Nagayoski, M., Shinkyo, R., Sakaki, T., *et al.*, **2003**. Molecular cloning and characterization of CYP719, a methylenedioxy bridge-forming enzyme that belongs to a novel P450 family, from cultured *Coptis japonica* cells. *J. Biol. Chem. 278*, 38557–38565.

Issemann, I., Green, S., **1990**. Activation of a member of the steroid hormone receptor superfamily by peroxisome proliferators. *Nature 347*, 645–650.

Ivie, G.W., Bull, D.L., Beier, R.C., Pryor, N.W., Oertli, E.H., **1983**. Metabolic detoxification: mechanism of insect resistance to plant psoralens. *Science 221*, 374–376.

Iyengar, S., Arnason, J.T., Philogene, B.J.R., Werstiuk, N.H., Morand, P., **1990**. Comparative metabolism of the phototoxic allelochemical α-terthienyl in three species of Lepidopterans. *Pestic. Biochem. Physiol. 37*, 154–164.

Jansson, I., **1993**. Posttranslational modification of cytochrome P450. In: Schenkman, J.B., Greim, H. (Eds.), Cytochrome P450. Springer, Berlin, pp. 561–580.

Jefcoate, C.F., **1978**. Measurement of substrate and inhibitor binding to microsomal cytochrome P-450 by optical-difference spectroscopy. *Methods Enzymol. 52*, 258–279.

Jermy, T., **1984**. Evolution of insect/host plant relationships. *Am. Naturalist. 124*, 609–630.

Jin, W., Riley, R.M., Wolfinger, R.D., White, K.P., Passador-Gurgel, G., *et al.*, **2001**. The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster. Nat. Genet. 29*, 389–395.

Johnson, P., Rees, H.H., **1977**. The mechanism of C-20 hydroxylation of α-ecdysone in the desert locust, *Schistocerca gregaria*. *Biochem. J. 168*, 513–520.

Jowett, T., Wajidi, M.F., Oxtoby, E., Wolf, C.R., **1991**. Mammalian genes expressed in *Drosophila*: a transgenic model for the study of mechanisms of chemical mutagenesis and metabolism. *EMBO Journal 10*, 1075–1081.

Kabbouh, M., Kappler, C., Hetru, C., Durst, F., **1987**. Further characterization of the 2-deoxyecdysone C-2 hydroxylase from *Locusta migratoria*. *Insect Biochem. 17*, 1155–1161.

Kahn, R.A., Bak, S., Svendsen, I., Halkier, B.A., Moller, B.L., **1997**. Isolation and reconstitution of cytochrome P450ox and in vitro reconstitution of the entire biosynthetic pathway of the cyanogenic glucosidedhurrin from sorghum. *Plant Physiol. 115*, 1661–1670.

Kamataki, T., Neal, R.A., **1976**. Metabolism of diethyl p-nitrophenyl phosphorothionate (parathion) by a reconstituted mixed-function oxidase enzyme system: studies of the covalent binding of the sulfur atom. *Mol. Pharmacol. 12*, 933–944.

Kao, L.M., Wilkinson, C.F., Brattsten, L.B., **1995**. In vivo effects of 2,4–D and atrazine on cytochrome P-450 and insecticide toxicity in southern armyworm (*Spodoptera eridania*) larvae. *Pestic. Sci. 45*, 331–334.

Kappers, W.A., Edwards, R.J., Murray, S., Boobis, A.R., **2001**. Diazinon is activated by CYP2C19 in human liver. *Toxicol. Appl. Pharmacol. 177*, 68–76.

Kappler, C., Kabbouh, M., Durst, F., Hoffmann, J.A., **1986**. Studies on the C-2 hydroxylation of 2-deoxyecdysone in *Locusta migratoria*. *Insect Biochem. 16*, 25–32.

Kappler, C., Kabbouh, M., Hetru, C., Durst, F., Hoffmann, J.A., **1988**. Characterization of three hydroxylases involved in the final steps of biosynthesis of the steroid hormone ecdysone in *Locusta migratoria* (Insecta, Orthoptera). *J. Steroid Biochem. 31*, 891–898.

Kasai, S., Scott, J.G., **2000**. Overexpression of cytochrome P450 CYP6D1 is associated with monooxygenase-mediated pyrethroid resistance in house flies from Georgia. *Pestic. Biochem. Physiol. 68*, 34–41.

Kasai, S., Scott, J.G., **2001a**. Cytochrome P450s CYP6D3 and CYP6D1 are part of a P450 gene cluster on autosome 1 in the housefly. *Insect Mol. Biol. 10*, 191–196.

Kasai, S., Scott, J.G., **2001b**. Expression and regulation of CYP6D3 in the housefly, *Musca domestica* (L.). *Insect Biochem. Mol. Biol. 32*, 1–8.

Kasai, S., Shono, T., Yamakawa, M., **1998a**. Molecular cloning and nucleotide sequence of a cytochrome P450 cDNA from a pyrethroid-resistant mosquito, *Culex quinquefasciatus* say. *Insect Mol. Biol. 7*, 185–190.

Kasai, S., Tomita, T., **2003**. Male specific expression of a cytochrome P450 (*Cyp312a1*) in *Drosophila melanogaster. Biochem. Biophys. Res. Commun. 300*, 894–900.

Kasai, S., Weerasinghe, I.S., Shono, T., **1998b**. P450 monooxygenases are an important mechanism of permethrin resistance in *Culex quinquefasciatus* Say larvae. *Arch. Insect Biochem. Physiol. 37*, 47–56.

Kasai, S., Weerashinghe, I.S., Shono, T., Yamakawa, M., **2000**. Molecular cloning, nucleotide sequence and gene expression of a cytochrome P450 (CYP6F1) from the pyrethroid-resistant mosquito, *Culex quinquefasciatus* Say. *Insect Biochem. Mol. Biol. 30*, 163–171.

Kayser, H., Eilinger, P., **2001**. Metabolism of diafenthiuron by microsomal oxidation: procide activation and inactivation as mechanisms contributing to selectivity. *Pest Manag. Sci. 57*, 975–980.

Kayser, H., Ertl, P., Eilinger, P., Spindler-Barth, M., Winkler, T., **2002**. Diastereomeric ecdysteroids with a cyclic hemiacetal in the side chain produced by cytochrome P450 in hormonally resistant insect cells. *Arch. Biochem. Biophys. 400*, 180–187.

Kayser, H., Winkler, T., Spindler-Barth, M., **1997**. 26-Hydroxylation of ecdysteroids is catalyzed by a typical cytochrome P-450-dependent oxidase and related to ecdysteroid resistance in an insect cell line. *Eur. J. Biochem. 248*, 707–716.

Kennaugh, L., Pearce, D., Daly, J.C., Hobbs, A.A., **1993**. A piperonyl butoxide synergizable resistance to permethrin in *Helicoverpa armigera* which is not due to increased detoxification by cytochrome P450. *Pestic. Biochem. Physiol. 45*, 234–241.

Kennedy, G.G., **1984**. 2-Tridecanone, tomatoes and *Heliothis zea*: potential incompability of plant antibiosis with insecticidal control. *Entomol. Exp. Appl. 35*, 305–311.

Kirby, M.L., Ottea, J.A., **1995**. Multiple mechanisms for enhancement of glutathione S-transferase activities in *Spodoptera frugiperda* (Lepidopter: Noctuidae). *Insect Biochem. Mol. Biol. 25*, 347–353.

Kirby, M.L., Young, R.J., Ottea, J.A., **1994**. Mixed-function oxidase and glutathione S-transferase activities from field-collected larval and adult tobacco budworms, *Heliothis virescens* (F.). *Pestic. Biochem. Physiol. 49*, 24–36.

Koener, J.F., Cariño, F.A., Feyereisen, R., **1993**. The cDNA and deduced protein sequence of housefly NADPH-cytochrome P450 reductase. *Insect Biochem. Mol. Biol. 23*, 439–447.

Komori, M., Kitamura, R., Fukuta, H., Inoue, H., Baba, H., *et al.*, **1993**. Transgenic *Drosophila* carrying mammalian cytochrome P-4501A1: an application to toxicology testing. *Carcinogenesis 14*, 1683–1688.

Konno, T.E.H., Dauterman, W.C., **1989**. Studies on methyl parathion resistance in *Heliothis virescens*. *Pestic. Biochem. Physiol. 33*, 189–199.

Korytko, P.J., MacLntyre, R.J., Scott, J.G., **2000a**. Expression and activity of a house-fly cytochrome P450, CYP6D1, in *Drosophila melanogaster. Insect Mol. Biol. 9*, 441–449.

Korytko, P.J., Scott, J.G., **1998**. CYP6D1 protects thoracic ganglia of houseflies from the neurotoxic insecticide cypermethrin. *Arch. Insect Biochem. Physiol. 37*, 57–63.

Korytko, P.J., Quimby, F.W., Scott, J.G., **2000b**. Metabolism of phenanthrene by housefly CYP6D1 and dog liver cytochrome P450. *J. Biochem. Mol. Toxicol. 14*, 20–25.

Kotze, A.C., **1995**. Induced insecticide tolerance in larvae of *Lucilia cuprina* (Wiedemann) (Diptera: Calliphoridae) following dietary phenobarbital treatment. *J. Austral. Entomol. Soc. 34*, 205–209.

Kotze, A.C., Sales, N., **1995**. Elevated in vitro monooxygenase activity associated with insecticide resistances in field-strain larvae of the Australian sheep blowfly (Diptera: Calliphoridae). *J. Econ. Entomol. 88*, 782–787.

Krieger, R.I., Feeny, P.P., Wilkinson, C.F., **1971**. Detoxication enzymes in the guts of caterpillars: an evolutionary answer to plant defenses? *Science 172*, 579–581.

Kula, M.E., Rozek, C.E., **2000**. Expression and translocation of *Drosophila* nuclear encoded cytochrome b(5) proteins to mitochondria. *Insect Biochem. Mol. Biol. 30*, 927–935.

Kula, M.E., Allay, E.R., Rozek, C.E., **1995**. Evolutionary divergence of the cytochrome b_5 gene of *Drosophila*. *J. Mol. Evol. 41*, 430–439.

Latli, B., Prestwich, G.D., **1991**. Metabolically blocked analogs of housefly sex pheromone: I. Synthesis of alternative substrates for the cuticular monooxygenases. *J. Chem. Ecol. 17*, 1745–1768.

Le Goff, G., Boundy, S., Daborn, P.J., Yen, J.L., Sofer, L., *et al.*, **2003**. Microarray analysis of cytochrome P450 mediated insecticide resistance in *Drosophila*. *Insect Biochem. Mol. Biol. 33*, 701–708.

Li, X., Berenbaum, M.R., Schuler, M.A., **2000a**. Molecular cloning and expression of CYP6B8: a xanthotoxin-inducible cytochrome P450 cDNA from *Helicoverpa zea*. *Insect Biochem. Mol. Biol. 30*, 75–84.

Li, X., Berenbaum, M.R., Schuler, M.A., **2002c**. Plant allelochemicals differentially regulate *Helicoverpa zea* cytochrome P450 genes. *Insect Mol. Biol. 11*, 343–351.

Li, W., Berenbaum, M.R., Schuler, M.A., **2001**. Molecular analysis of multiple CYP6B genes from polyphagous

Papilio species. *Insect Biochem. Mol. Biol. 31*, 999–1011.

Li, X., Berenbaum, M.R., Schuler, M.A., **2002b**. Cytochrome P450 and actin genes expressed in *Helicoverpa zea* and *Helicoverpa armigera*: paralogy/orthology identification, gene conversion and evolution. *Insect Biochem. Mol. Biol. 32*, 311–320.

Li, W., Petersen, R.A., Schuler, M.A., Berenbaum, M.R., **2002a**. CYP6B cytochrome P450 monooxygenases from *Papilio canadensis* and *Papilio glaucus*: potential contributions of sequence divergence to host plant associations. *Insect Mol. Biol. 11*, 543–551.

Li, X., Schuler, M.A., Berenbaum, M.R., **2002d**. Jasmonate and salicylate induce expression of herbivore cytochrome P450 genes. *Nature 419*, 712–715.

Li, W., Schuler, M.A., Berenbaum, M.R., **2003**. Diversification of furanocoumarin-metabolizing cytochrome P450 monooxygenases in two papilionids: specificity and substrate encounter rate. *Proc. Natl Acad. Sci. USA 100*, 14593–14598.

Li, X., Zangerl, A.R., Schuler, M.A., Berenbaum, M., **2000b**. Cross-resistance to α-cypermethrin after xanthotoxin ingestion in *Helicoverpa zea* (Lepidoptera: Noctuidae). *J. Econ. Entomol. 93*, 18–25.

Liebrich, W., Hoffmann, K.H., **1991**. Ecdysone 20-monooxygenase in a cricket, *Gryllus bimaculatus* (Ensifera, Gryllidae): characterization of the microsomal midgut steroid hydroxylase in adult females. *J. Comp. Physiol. B 161*, 93–99.

Lindigkeit, R., Biller, A., Buch, M., Schiebel, H.M., Boppre, M., *et al.*, **1997**. The two facies of pyrrolizidine alkaloids: the role of the tertiary amine and its N-oxide in chemical defense of insects with acquired plant alkaloids. *Eur. J. Biochem. 245*, 626–636.

Lindroth, R.L., **1989**. Host plant alteration of detoxication activity in *Papilio glaucus glaucus*. *Entomol. Exp. Appl. 50*, 29–35.

Lindroth, R.L., **1991**. Differential toxicity of plant allelochemicals to insects: roles of enzymatic detoxication systems. In: Bernays, E.A. (Ed.), Insect–Plant Interactions, vol. 3. CRC Press, Boca Raton, pp. 1–33.

Lindsley, D.L., Zimm, G.G., **1992**. The Genome of *Drosophila melanogaster*. Academic Press, San Diego.

Liu, N., Scott, J.G., **1995**. Genetics of resistance to pyrethroid insecticides in the housefly, *Musca domestica*. *Pestic. Biochem. Physiol. 52*, 116–124.

Liu, N., Scott, J.G., **1996**. Genetic analysis of factors controlling high-level expression of cytochrome P450, CYP6D1, cytochrome b_5, P450 reductase, and monooxygenase activities in LPR house flies, *Musca domestica*. *Biochem. Genet. 34*, 133–148.

Liu, N., Scott, J.G., **1997a**. Inheritance of CYP6D1-mediated pyrethroid resistance in housefly (Diptera: Muscidae). *J. Econ. Entomol. 90*, 1478–1481.

Liu, N., Scott, J.G., **1997b**. Phenobarbital induction of CYP6D1 is due to a trans acting factor on autosome 2 in house flies, *Musca domestica*. *Insect Mol. Biol. 6*, 77–81.

Liu, N., Scott, J.G., **1998**. Increased transcription of CYP6D1 causes cytochrome P450-mediated insecticide resistance in housefly. *Insect Biochem. Mol. Biol. 28*, 531–535.

Liu, N., Tomita, T., Scott, J.G., **1995**. Allele-specific PCR reveals that CYP6D1 is on chromosome 1 in the housefly, *Musca domestica*. *Experientia 51*, 164–167.

Liu, N., Yue, X., **2001**. Genetics of pyrethroid resistance in a strain (ALHF) of house flies (Diptera: Muscidae). *Pestic. Biochem. Physiol. 70*, 151–158.

Liu, N., Zhang, L., **2002**. Identification of two new cytochrome P450 genes and their 5′-flanking regions from the housefly, *Musca domestica*. *Insect Biochem. Mol. Biol. 32*, 755–764.

Lockey, K.H., **1991**. Insect hydrocarbon classes: implications for chemotaxonomy. *Insect Biochem. 21*, 91–97.

Lu, K.H., Bradfield, J.Y., Keeley, L.L., **1995**. Hypertrehalosemic hormone-regulated gene expression for cytochrome P4504C1 in the fat body of the Cockroach, *Blaberus discoidalis*. *Arch. Insect Biochem. Physiol. 28*, 79–90.

Lu, K.H., Bradfield, J.Y., Keeley, L.L., **1999**. Juvenile hormone inhibition of gene expression for cytochrome P4504C1 in adult females of the cockroach, *Blaberus discoidalis*. *Insect Biochem. Mol. Biol. 29*, 667–673.

Luu, B., Werner, F., **1996**. Sterols that modify moulting in insects. *Pestic. Sci. 46*, 49–53.

Lynch, M., Conery, J.S., **2000**. The evolutionary fate and consequences of duplicate genes. *Science 290*, 1151–1155.

Ma, R., Cohen, M.B., Berenbaum, M.R., Schuler, M.A., **1994**. Black swallowtail (*Papilio polyxenes*) alleles encode cytochrome P450s that selectively metabolize linear furanocoumarins. *Arch. Biochem. Biophys. 310*, 332–340.

Maibeche-Coisne, M., Jacquin-Joly, E., Francois, M.C., Nagnan-Le Meillour, P., **2002**. cDNA cloning of biotransformation enzymes belonging to the cytochrome P450 family in the antennae of the noctuid moth *Mamestra brassicae*. *Insect Mol. Biol. 11*, 273–281.

Maibeche-Coisne, M., Monti-Dedieu, L., Aragon, S., Dauphin-Villemant, C., **2000**. A new cytochrome P450 from *Drosophila melanogaster*, CYP4G15, expressed in the nervous system. *Biochem. Biophys. Res. Commun. 273*, 1132–1137.

Maitra, S., Dombrowski, S.M., Basu, M., Raustol, O., Waters, L.C., *et al.*, **2000**. Factors on the third chromosome affect the level of *Cyp6a2* and *Cyp6a8* expression in *Drosophila melanogaster*. *Gene 248*, 147–156.

Maitra, S., Dombrowski, S.M., Waters, L.C., Ganguly, R., **1996**. Three second chromosome-linked clustered Cyp6 genes show differential constitutive and barbital-induced expression in DDT-resistant and susceptible strains of *Drosophila melanogaster*. *Gene 180*, 165–171.

Maitra, S., Price, C., Ganguly, R., **2002**. *Cyp6a8* of *Drosophila melanogaster*: gene structure, and sequence and

functional analysis of the upstream DNA. *Insect Biochem. Mol. Biol. 32*, 859–870.

Mansuy, D., **1998**. The great diversity of reactions catalyzed by cytochrome P450. *Comp. Biochem. Physiol. 121C*, 5–14.

Mansuy, D., Renaud, J.P., **1995**. Heme-thiolate proteins different from cytochromes P450 catalyzing monooxygenations. In: Ortiz de Montellano, P.R. (Ed.), Cytochrome P450. Plenum, New York and London, pp. 537–574.

Martin, T., Ochou, O.G., Vaissayre, M., Fournier, D., **2003**. Oxidases responsible for resistance to pyrethroids sensitize *Helicoverpa armigera* (Hubner) to triazophos in West Africa. *Insect Biochem. Mol. Biol. 33*, 883–887.

Matsunaga, E., Umeno, M., Gonzalez, F.J., **1990**. The rat P450 IID subfamily: complete sequences of four closely linked genes and evidence that gene conversions maintained sequence homogeneity at the heme-binding region of the cytochrome P450 active site. *J. Mol. Evol. 30*, 155–169.

Mauchamp, B., Darrouzet, E., Malosse, C., Couillaud, F., **1999**. 4′-OH-JH-III: an additional hydroxylated juvenile hormone produced by locust corpora allata in vitro. *Insect Biochem. Mol. Biol. 29*, 475–480.

McDonald, M.J., Rosbash, M., **2001**. Microarray analysis and organization of circadian gene expression in *Drosophila*. *Cell 107*, 567–578.

Megias, A., Saborido, A., Muncio, A.M., **1984**. NADH-cytochrome b_5 reductase from the insect *Ceratitis capitata*. Enzyme properties and membrane binding capacity. *Comp. Biochem. Physiol. 77B*, 679–685.

Miota, F., Siegfried, B.D., Scharf, M.E., Lydy, M.J., **2000**. Atrazine induction of cytochrome P450 in *Chironomus tentans* larvae. *Chemosphere 40*, 285–291.

Misra, S., Crosby, M.A., Mungall, C.J., Matthews, B.B., Campbell, K.S., *et al.*, **2002**. Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. *Genome Biol. 3*, 83.1–83.22.

Mitlin, N., Konecky, M.S., **1955**. The inhibition of development in the housefly by piperonyl butoxide. *J. Econ. Entomol. 48*, 93–94.

Moldenke, A.F., Berry, R.E., Miller, J.C., Kelsey, R.G., Wernz, J.G., *et al.*, **1992**. Carbaryl susceptibility and detoxication enzymes in gypsy moth (Lepidoptera: Lymantriidae): influence of host plant. *J. Econ. Entomol. 85*, 1628–1635.

Mpuru, S., Blomquist, G.J., Schal, C., Roux, M., Kuenzli, M., *et al.*, **2001**. Effect of age and sex on the production of internal and external hydrocarbons and pheromones in the housefly, *Musca domestica*. *Insect Biochem. Mol. Biol. 31*, 139–155.

Mpuru, S., Reed, J.R., Reitz, R.C., Blomquist, G.J., **1996**. Mechanism of hydrocarbon biosynthesis from aldehyde in selected insect species: requirement for O_2 and NADPH and carbonyl group released as CO_2. *Insect Biochem. Mol. Biol. 26*, 203–208.

Mullin, C.A., **1986**. Adaptive divergence of chewing and sucking arthropods to plant allelochemicals. In: Brattsten, L.B., Ahmad, S. (Eds.), Molecular Aspects of Insect–Plant Interactions. Plenum, New York, pp. 175–209.

Mullin, C.A., Croft, B.A., Strickler, K., Matsumura, F., Miller, J.R., **1982**. Detoxification enzyme differences between a herbivorous and predatory mite. *Science 217*, 1270–1272.

Murataliev, M.B., Ariño, A., Guzov, V.M., Feyereisen, R., **1999**. Kinetic mechanism of cytochrome P450 reductase from the housefly (*Musca domestica*). *Insect Biochem. Mol. Biol. 29*, 233–242.

Murataliev, M.B., Feyereisen, R., **1999**. Mechanism of cytochrome P450 reductase from the housefly: evidence for an FMN semiquinone as electron donor. *FEBS Lett. 453*, 201–204.

Murataliev, M.B., Feyereisen, R., **2000**. Interaction of NADP(H) with oxidized and reduced P450 reductase during catalysis. Studies with nucleotide analogues. *Biochemistry 39*, 5066–5074.

Murataliev, M.B., Feyereisen, R., Walker, F.A., **2004a**. Electron transfer by diflavin reductases. *Biochem. Biophys. Acta 1698*, 1–26.

Murataliev, M.B., Trinh, L.N., Moser, L.V., Bates, R.B., Feyereisen, R., *et al.*, **2004b**. Chimeragenesis of the fatty acid binding site of cytochrome P450BM3. Replacement of residues 73–84 with the homologous residues from the insect cytochrome P450 CYP4C7. *Biochemistry 43*, 1771–1780.

Mutch, E., Daly, A.K., Leathart, J.B., Blain, P.G., Williams, F.M., **2003**. Do multiple cytochrome P450 isoforms contribute to parathion metabolism in man? *Arch. Toxicol. 77*, 313–320.

Nahri, L.O., Fulco, A.J., **1986**. Characterization of a catalytically self-sufficient 119,000-dalton cytochrome P450 monooxygenase induced by barbiturates in *Bacillus megaterium*. *J. Biol. Chem. 261*, 7160–7169.

Nahrstedt, A., **1988**. Cyanogenesis and the role of cyanogenic compounds in insects. *Ciba Found. Symp. 140*, 131–150.

Nakayama, N., Takemae, A., Shoun, H., **1996**. Cytochrome P450foxy, a catalytically self-sufficient fatty acid hydroxylase of the fungus *Fusarium oxysporum*. *J. Biochem. 119*, 435–440.

Naumann, C., Hartmann, T., Ober, D., **2002**. Evolutionary recruitment of a flavin-dependent monooxygenase for the detoxication of host plant-acquired pyrrolizidine alkaloid-defended arctiid moth *Tyria Jacobaeae*. *Proc. Natl Acad. Sci. USA 99*, 6085–6090.

Neal, J.J., **1987**. Metabolic costs of mixed-function oxidase induction in *Heliothis zea*. *Entomol. Exp. Appl. 43*, 175–179.

Neal, J.J., Berenbaum, M.R., **1989**. Decreased sensitivity of mixed-function oxidases from *Papilio polyxenes* to inhibitors in host plants. *J. Chem. Ecol. 15*, 439–446.

Neal, J.J., Reuveni, M., **1992**. Separation of cytochrome P450 containing vesicles from the midgut microsomal fraction of *Manduca sexta*. *Comp. Biochem. Physiol. 102C*, 77–82.

Neal, J.J., Wu, D., **1994**. Inhibition of insect cytochromes P450 by furanocoumarins. *Pestic. Biochem. Physiol. 50*, 43–50.

Nebert, D.W., **1991**. Proposed role of drug-metabolizing enzymes: regulation of steady state levels of the ligands that effect growth, homeostasis, differentiation, and neuroendocrine functions. *Mol. Endocrinol. 5*, 1203–1214.

Nebert, D.W., Adesnik, M., Coon, M.J., Estabrook, R.W., Gonzalez, F.J., *et al.*, **1987**. The P450 gene superfamily: recommended nomenclature. *DNA 6*, 1–11.

Nebert, D.W., Nelson, D.R., Coon, M.J., Estabrook, R.W., Feyereisen, R., *et al.*, **1991**. The P450 superfamily: update on new sequences, gene mapping, and recommended nomenclature [published erratum appears in DNA Cell Biol. 1991 Jun;10(5):397–398]. *DNA Cell Biol. 10*, 1–14.

Nelson, D.R., **1998**. Metazoan cytochrome P450 evolution. *Comp. Biochem. Physiol. 121C*, 15–22.

Nelson, D.R., Kamataki, T., Waxman, D.J., Guengerich, F.P., Estabrook, R.W., *et al.*, **1993**. The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. *DNA Cell Biol. 12*, 1–51.

Nelson, D.R., Koymans, L., Kamataki, T., Stegeman, J.J., Feyereisen, R., *et al.*, **1996**. P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics 6*, 1–42.

Newcomb, R.D., Campbell, P.M., Ollis, D.L., Cheah, E., Russell, R.J., *et al.*, **1997**. A single amino acid substitution converts a carboxylesterase to an organophosphorus hydrolase and confers insecticide resistance on a blowfly. *Proc. Natl Acad. Sci. USA 94*, 7464–7468.

Newcomb, M., Hollenberg, P.F., Coon, M.J., **2003**. Multiple mechanisms and multiple oxidants in P450-catalyzed hydroxylations. *Arch. Biochem. Biophys. 409*, 72–79.

Nikou, D., Ranson, H., Hemingway, J., **2003**. An adult-specific CYP6 P450 gene is overexpressed in a pyrethroid-resistant strain of the malaria vector, *Anopheles gambiae*. *Gene 318*, 91–102.

Nitao, J.K., **1989**. Enzymatic adaptation in a specialist herbivore for feeding on furanocoumarin-containing plants. *Ecology 70*, 629–635.

Nitao, J.K., **1990**. Metabolism and excretion of the furanocoumarin xanthotoxin by parsnip webworm, *Depressaria pastinacella*. *J. Chem. Ecol. 16*, 417–428.

Nitao, J.K., Berhow, M., Duval, S.M., Weisleder, D., Vaughn, S.F., *et al.*, **2003**. Characterization of furanocoumarin metabolites in parsnip webworm, *Depressaria pastinacella*. *J. Chem. Ecol. 29*, 671–682.

Oakeshott, J.G., Horne, I., Sutherland, T.D., Russell, R.J., **2003**. The genomics of insecticide resistance. *Genome Biol. 4*, 202.

Oesch-Bartlomowicz, B., Oesch, F., **2003**. Cytochrome P450 phosphorylation as a functional switch. *Arch. Biochem. Biophys. 409*, 228–234.

Oi, M., Dauterman, W.C., Motoyama, N., **1990**. Biochemical factors responsible for an extremely high level of diazinon resistance in a housefly strain. *J. Pestic. Sci. 15*, 217–224.

Omura, T., **1993**. History of cytochrome P450. In: Omura, T., Ishimura, Y., Fujii-Kuriyama, Y. (Eds.), Cytochrome P-450. Kodansha, Tokyo, pp. 1–15.

Omura, T., Ishimura, Y., Fujii-Kuriyama, Y. (Eds.), **1993**. Cytochrome P-450. Kodansha, Tokyo.

Omura, T., Sato, R., **1964**. The carbon monoxide-binding pigment of liver microsomes I. Evidence for its hemoprotein nature. *J. Biol. Chem. 239*, 2370–2378.

Ortiz de Montellano, P.R., **1995a**. Oxygen activation and reactivity. In: Ortiz de Montellano, P.R. (Ed.), Cytochrome P450. Plenum, New York and London, pp. 245–303.

Ortiz de Montellano, P.R. (Ed.), **1995b**. Cytochrome P450, Structure, Mechanism, and Biochemistry. Plenum, New York and London.

Ortiz de Montellano, P.R., Correia, M.A., **1995**. Inhibition of cytochrome P450 enzymes. In: Ortiz de Montellano, P.R. (Ed.), Cytochrome P450. Plenum, New York and London, pp. 305–364.

Ortiz de Montellano, P.R., Graham-Lorence, S.E., **1993**. Structure of cytochrome P450: heme-binding site and heme reactivity. In: Schenkman, J.B., Greim, H. (Eds.), Cytochrome P450. Springer, Berlin, pp. 169–181.

Paquette, S.M., Bak, S., Feyereisen, R., **2000**. Intron-exon organization and phylogeny in a large superfamily, the paralogous cytochrome P450 genes of *Arabidopsis thaliana*. *DNA Cell Biol. 19*, 307–317.

Pare, P.W., Alborn, H.T., Tumlinson, J.H., **1998**. Concerted biosynthesis of an insect elicitor of plant volatiles. *Proc. Natl Acad. Sci. USA 95*, 13971–13975.

Pauli, D., Tonka, C.H., **1987**. A *Drosophila* heat shock gene from locus 67B is expressed during embryogenesis and pupation. *J. Mol. Biol. 198*, 235–240.

Petersen, R.A., Niamsup, H., Berenbaum, M.R., Schuler, M.A., **2003**. Transcriptional response elements in the promoter of CYP6B1, an insect P450 gene regulated by plant chemicals. *Biochimica et Biophysica Acta (BBA) – General Subjects 1619*, 269–282.

Petersen, R.A., Zangerl, A.R., Berenbaum, M.R., Schuler, M.A., **2001**. Expression of CYP6B1 and CYP6B3 cytochrome P450 monooxygenases and furanocoumarin metabolism in different tissues of *Papilio polyxenes* (Lepidoptera: Papilionidae). *Insect Biochem. Mol. Biol. 31*, 679–690.

Petryk, A., Warren, J.T., Marques, G., Jarcho, M.P., Gilbert, L.I., *et al.*, **2003**. Shade is the *Drosophila* P450 enzyme that mediates the hydroxylation of ecdysone to the steroid insect molting hormone 20-hydroxyecdysone. *Proc. Natl Acad. Sci. USA 100*, 13773–13778.

Pikuleva, I.A., Cao, C., Waterman, M.R., **1999**. An additional electrostatic interaction between adrenodoxin and P450c27 (CYP27A1) results in tighter binding than between adrenodoxin and P450scc (CYP11A1). *J. Biol. Chem. 274*, 2045–2052.

Pikuleva, I.A., Mackman, R.L., Kagawa, N., Waterman, M.R., Ortiz de Montellano, P.R., **1995**. Active-site

topology of bovine cholesterol side-chain cleavage cytochrome P450 (P450scc) and evidence for interaction of tyrosine 94 with the side chain of cholesterol. *Arch. Biochem. Biophys. 322*, 189–197.

Pimprale, S.S., Besco, C.L., Bryson, P.K., Brown, T.M., **1997**. Increased susceptibility of pyrethroid-resistant tobacco budworm (Lepidoptera: Noctuidae) to chlorfenapyr. *J. Econ. Entomol. 90*, 49–54.

Pittendrigh, B., Aronstein, K., Zinkovsky, E., Andreev, O., Campbell, B., *et al.*, **1997**. Cytochrome P450 genes from *Helicoverpa armigera*: expression in a pyrethroid-susceptible and resistant strain. *Insect Biochem. Mol. Biol. 27*, 507–512.

Plapp, F.W., Jr., **1984**. The genetic basis of insecticide resistance in the housefly: evidence that a single locus plays a major role in metabolic resistance to insecticides. *Pestic. Biochem. Physiol. 22*, 94–201.

Plettner, E., Slessor, K.N., Winston, M.L., **1998**. Biosynthesis of mandibular acids in honey bees (*Apis mellifera*): De novo synthesis, route of fatty acid hydroxylation and caste selective β-oxidation. *Insect Biochem. Mol. Biol. 28*, 31–42.

Pompon, D., Louerat, B., Bronine, A., Urban, P., **1996**. Yeast expression of animal and plant P450s in optimized redox environments. *Methods Enzymol. 272*, 51–64.

Porter, T.D., **2002**. The roles of cytochrome b_5 in cytochrome P450 reactions. *J. Biochem. Mol. Toxicol. 16*, 311–316.

Porter, T.D., Kasper, C.B., **1986**. NADPH-cytochrome P-450 oxidoreductase: flavin mononucleotide and flavin adenine dinucleotide domains evolved from different flavoproteins. *Biochemistry 25*, 1682–1687.

Poulos, T.L., Cupp-Vickery, J., Li, H., **1995**. Structural studies on prokaryotic cytochromes P450. In: Ortiz de Montellano, P.R. (Ed.), Cytochrome P450. Plenum, New York and London, pp. 125–150.

Poulos, T.L., Finzel, B.C., Gunsalus, I.C., Wagner, G.C., Kraut, J., **1985**. The 2.6-A crystal structure of *Pseudomonas putida* cytochrome P-450. *J. Biol. Chem. 260*, 16122–16130.

Prapaipong, H., Berenbaum, M.R., Schuler, M.A., **1994**. Transcriptional regulation of the *Papilio polyxenes* CYP6B1 gene. *Nucleic Acids Res. 22*, 3210–3217.

Pratt, G.E., Kuwano, E., Farnsworth, D.E., Feyereisen, R., **1990**. Structure/activity studies on 1,5-disubstituted imidazoles as inhibitors of juvenile hormone biosynthesis in isolated corpora allata of the cockroach, *Diploptera punctata*. *Pestic. Biochem. Physiol. 38*, 223–230.

Ranasinghe, C., Headlam, M., Hobbs, A.A., **1997**. Induction of the mRNA for CYP6B2, a pyrethroid inducible cytochrome P450, in *Helicoverpa armigera* (Hubner) by dietary monoterpenes. *Arch. Insect Biochem. Physiol. 34*, 99–109.

Ranasinghe, C., Hobbs, A.A., **1998**. Isolation and characterization of two cytochrome P450 cDNA clones for CYP6B6 and CYP6B7 from *Helicoverpa armigera* (Hubner): possible involvement of CYP6B7 in pyrethroid resistance. *Insect Biochem. Mol. Biol. 28*, 571–580.

Ranasinghe, C., Hobbs, A.A., **1999a**. Isolation and characterisation of a cytochrome b_5 cDNA clone from *Helicoverpa armigera* (Hubner): possible involvement of cytochrome b_5 in cytochrome P450 CYP6B7 activity towards pyrethroids. *Insect Biochem. Mol. Biol. 29*, 145–151.

Ranasinghe, C., Hobbs, A.A., **1999b**. Induction of cytochrome P450 CYP6B7 and cytochrome b_5 mRNAs from *Helicoverpa armigera* (Hubner) by pyrethroid insecticides in organ culture. *Insect Mol. Biol. 8*, 443–447.

Ranson, H., Claudianos, C., Ortelli, F., Abgrall, C., Hemingway, J., *et al.*, **2002a**. Evolution of supergene families associated with insecticide resistance. *Science 298*, 179–181.

Ranson, H., Nikou, D., Hutchinson, M., Wang, X., Roth, C.W., *et al.*, **2002b**. Molecular analysis of multiple cytochrome P450 genes from the malaria vector, *Anopheles gambiae*. *Insect Mol. Biol. 11*, 409–418.

Ravichandran, K.G., Boddupalli, S.S., Haseman, C.A., Peterson, J.A., Deisenhofer, J., **1993**. Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. *Science 261*, 731–736.

Reed, J.R., Quilici, D.R., Blomquist, G.J., Reitz, R.C., **1995**. Proposed mechanism for the cytochrome P450-catalyzed conversion of aldehydes to hydrocarbons in the housefly, *Musca domestica*. *Biochemistry 34*, 16221–16227.

Reed, J.R., Vanderwel, D., Choi, S., Pomonis, J.G., Reitz, R.C., *et al.*, **1994**. Unusual mechanism of hydrocarbon formation in the housefly: cytochrome P450 converts aldehyde to the sex pheromone component (Z)-9-tricosene and CO_2. *Proc. Natl Acad. Sci. USA 91*, 10000–10004.

Rees, H.H., **1995**. Ecdysteroid biosynthesis and inactivation in relation to function. *Eur. J. Entomol. 92*, 9–39.

Ringo, J., Jona, G., Rockwell, R., Segal, D., Cohen, E., **1995**. Genetic variation for resistance to chlorpyrifos in *Drosophila melanogaster* (Diptera: Drosophilidae) infesting grapes in Israel. *J. Econ. Entomol. 88*, 1158–1163.

Riskallah, M.R., Dauterman, W.C., Hodgson, E., **1986a**. Nutritional effects on the induction of cytochrome P-450 and glutathione transferase in larvae of the tobacco budworm, *Heliothis virescens* (F.). *Insect Biochem. 16*, 491–499.

Riskallah, M.R., Dauterman, W.C., Hodgson, E., **1986b**. Host plant induction of microsomal monooxygenases activity in relation to diazinon metabolism and toxicity in larvae of the tobacco budworm *Heliothis virescens* (F.). *Pestic. Biochem. Physiol. 25*, 233–247.

Robertson, H.M., Martos, R., Sears, C.R., Todres, E.Z., Walden, K.K., *et al.*, **1999**. Diversity of odourant binding proteins revealed by an expressed sequence tag project on male *Manduca sexta* moth antennae. *Insect Mol. Biol. 8*, 501–518.

Ronis, M.J.J., Hodgson, E., Dauterman, W.C., **1988**. Characterization of multiple forms of cytochrome P-450 from an insecticide resistant strain of housefly (*Musca domestica*). *Pestic. Biochem. Physiol. 32*, 74–90.

Rose, H.A., **1985**. The relationship between feeding specialization and host plants to aldrin epoxidase activities of midgut homogenates in larval Lepidoptera. *Ecol. Entomol. 10*, 455–467.

Rose, R.L., Barbhaiya, L., Roe, R.M., Rock, G.C., Hodgson, E., **1995**. Cytochrome P450-associated insecticide resistance and the development of biochemical diagnostic assays in *Heliothis virescens*. *Pestic. Biochem. Physiol. 51*, 178–191.

Rose, R.L., Goh, D., Thompson, D.M., Verma, K.D., Heckel, D.G., *et al.*, **1997**. Cytochrome P450 (CYP)9A1 in *Heliothis virescens*: the first member of a new CYP family. *Insect Biochem. Mol. Biol. 27*, 605–615.

Rose, R.L., Gould, F., Levi, P.E., Hodgson, E., **1991**. Differences in cytochrome P450 activities in tobacco budworm larvae as influenced by resistance to host plant allelochemicals and induction. *Comp. Biochem. Physiol. 99B*, 535–540.

Sabourault, C., Guzov, V.M., Koener, J.F., Claudianos, C., Plapp, F.W., Jr., *et al.*, **2001**. Overproduction of a P450 that metabolizes diazinon is linked to a loss-of-function in the chromosome 2 ali-esterase (MdαE7) gene in resistant house flies. *Insect Mol. Biol. 10*, 609–618.

Salzemann, A., Nagnan, P., Tellier, F., Jaffe, K., **1992**. Leaf-cutting ant *Atta laevigata* (Formicidae: Attini) marks its territory with colony-specific Dufour gland secretion. *J. Chem. Ecol. 18*, 183.

Saner, C., Weibel, B., Wurgler, F.E., Sengstag, C., **1996**. Metabolism of promutagens catalyzed by *Drosophila melanogaster* CYP6A2 enzyme in *Saccharomyces cerevisiae*. *Environ. Mol. Mutagen. 27*, 46–58.

Scharf, M.E., Lee, C.Y., Neal, J.J., Bennett, G.W., **1999**. Cytochrome P450 MA expression in insecticide-resistant German cockroaches (Dictyoptera: Blattellidae). *J. Econ. Entomol. 92*, 788–793.

Scharf, M.E., Neal, J.J., Marcus, C.B., Bennett, G.W., **1998**. Cytochrome P450 purification and immunological detection in an insecticide resistant strain of German cockroach (*Blattella germanica*, L.). *Insect Biochem. Mol. Biol. 28*, 1–9.

Scharf, M.E., Parimi, S., Meinke, L.J., Chandler, L.D., Siegfried, B.D., **2001**. Expression and induction of three family 4 cytochrome P450 (CYP4)* genes identified from insecticide-resistant and susceptible western corn rootworms, *Diabrotica virgifera virgifera*. *Insect Mol. Biol. 10*, 139–146.

Scharf, M., Siegrfried, B.D., Meinke, L.J., Chandler, L.D., **2000**. Fipronil metabolism, oxidative sulfone formation and toxicity among organophosphate- and carbamate-resistant and susceptible western corn rootworm populations. *Pest Manag. Sci. 56*, 757–766.

Schenkman, J.B., Greim, H. (Eds.), **1993**. Cytochrome P450. Springer, Berlin.

Schenkman, J.B., Jansson, I., **2003**. The many roles of cytochrome b_5. *Pharmacol. Ther. 97*, 139–152.

Schlenke, T.A., Begun, D.J., **2004**. Strong selective sweep associated with a transposon insertion in *Drosophila simulans*. *Proc. Natl Acad. Sci. USA 101*, 1626–1631.

Schlichting, I., Berendzen, J., Chu, K., Stock, A.M., Maves, S.A., *et al.*, **2000**. The catalytic pathway of cytochrome P450cam at atomic resolution. *Science 287*, 1615–1622.

Schonbrod, R.D., Khan, M.A.Q., Terriere, L.C., Plapp, F.W., Jr., **1968**. Microsomal oxidases in the housefly: a survey of fourteen strains. *Life Sci. 7*, 681–688.

Schrag, M.L., Wienkers, L.C., **2000**. Topological alteration of the CYP3A4 active site by the divalent cation $Mg^{(2+)}$. *Drug Metab. Disposition 28*, 1198–1201.

Schuler, M.A., Werck-Reichhart, D., **2003**. Functional genomics of P450s. *Ann. Rev. Plant Biol. 54*, 629–667.

Scott, J.G., **1999**. Cytochromes P450 and insecticide resistance. *Insect Biochem. Mol. Biol. 29*, 757–777.

Scott, J.A., Collins, F.H., Feyereisen, R., **1994**. Diversity of cytochrome P450 genes in the mosquito, *Anopheles albimanus*. *Biochem. Biophys. Res. Commun. 205*, 1452–1459.

Scott, J.G., Foroozesh, M., Hopkins, N.E., Alefantis, T.G., Alworth, W.L., **2000**. Inhibition of cytochrome P450 6D1 by alkynylarenes, methylenedioxyarenes, and other subsituted aromatics. *Pestic. Biochem. Physiol. 67*, 63–71.

Scott, J.G., Lee, S.S., **1993a**. Tissue distribution of microsomal cytochrome P-450 monooxygenases and their inducibility by phenobarbital in the insecticide resistant LPR strain of housefly, *Musca domestica* L. *Insect Biochem. Mol. Biol. 23*, 729–738.

Scott, J.G., Lee, S.S., **1993b**. Purification and characterization of a cytochrome P-450 from insecticide susceptible and resistant strains of housefly, *Musca domestica* L., before and after phenobarbital exposure. *Arch. Insect Biochem. Physiol. 24*, 1–19.

Scott, J.G., Liu, N., Wen, Z., Smith, F.F., Kasai, S., *et al.*, **1999**. House-fly cytochrome P450 CYP6D1: 5′ flanking sequences and comparison of alleles. *Gene 226*, 347–353.

Scott, J.G., Sridhar, P., Liu, N., **1996**. Adult specific expression and induction of cytochrome P450lpr in house flies. *Arch. Insect Biochem. Physiol. 31*, 313–323.

Scott, J.G., Zhang, L., **2003**. The housefly aliesterase gene (MdαE7) is not associated with insecticide resistance or P450 expression in three strains of housefly. *Insect Biochem. Mol. Biol. 33*, 139–144.

Scrimshaw, S., Kerfoot, W.C., **1987**. Chemical defenses of freshwater organisms: beetles and bugs. In: Kerfoot, W.C., Sih, A. (Eds.), Predation: Direct and Indirect Impacts on Aquatic Communities. University Press of New England, London, pp. 240–262.

Seifert, J., Scott, J.G., **2002**. The CYP6D1v1 allele is associated with pyrethroid resistance in the housefly, *Musca domestica*. *Pestic. Biochem. Physiol. 72*, 40–44.

Shaw, G.C., Fulco, A.J., **1993**. Inhibition by barbiturates of the binding of Bm3R1 repressor to its operator site

on the barbiturate-inducible cytochrome P450BM-3 gene of *Bacillus megaterium*. *J. Biol. Chem. 268*, 2997–3004.

Shen, B., Dong, H.Q., Tian, H.S., Ma, L., Li, X.L., *et al.*, **2003**. Cytochrome P450 genes expressed in the deltamethrin-susceptible and -resistant strains of *Culex pipiens pallens*. *Pestic. Biochem. Physiol. 75*, 19–26.

Sheppard, D.G., Joyce, J.A., **1998**. Increased susceptibility of pyrethroid-resistant horn flies (Diptera: Muscidae) to chlorfenapyr. *J. Econ. Entomol. 91*, 398–400.

Shergill, J.K., Cammack, R., Chen, J.H., Fisher, M.J., Madden, S., *et al.*, **1995**. EPR spectroscopic characterization of the iron-sulphur proteins and cytochrome P-450 in mitochondria from the insect *Spodoptera littoralis* (cotton leafworm). *Biochem. J. 307*, 719–728.

Shono, T., Kasai, S., Kamiya, E., Kono, Y., Scott, J.G., **2002**. Genetics and mechanisms of permethrin resistance in the YPER strain of housefly. *Pestic. Biochem. Physiol. 73*, 27–36.

Smith, S.L., **1985**. Regulation of ecdysteroid titer: synthesis. In: Kerkut, G.A., Gilbert, L.I. (Eds.), Comprehensive Insect Physiology, Biochemistry and Pharmacology, vol. 7. Pergamon, Oxford, pp. 295–341.

Smith, S.L., Bollenbacher, W.E., Cooper, D.Y., Schleyer, H., Wielgus, J.J., *et al.*, **1979**. Ecdysone 20-monooxygenase: characterization of an insect cytochrome P-450 dependent steroid hydroxylase. *Mol. Cell. Endocrinol. 15*, 111–133.

Smith, F.F., Scott, J.G., **1997**. Functional expression of housefly (*Musca domestica*) cytochrome P450 CYP6D1 in yeast (*Saccharomyces cerevisiae*). *Insect Biochem. Mol. Biol. 27*, 999–1006.

Snyder, M.J., Glendinning, J.I., **1996**. Causal connection between detoxification enzyme activity and consumption of a toxic plant compound. *J. Comp. Physiol. A 179*, 255–261.

Snyder, M.J., Hsu, E.-L., Feyereisen, R., **1993**. Induction of cytochrome P450 activities by nicotine in the tobacco hornworm, *Manduca sexta*. *J. Chem. Ecol. 19*, 2903–2916.

Snyder, M.J., Scott, J.A., Andersen, J.F., Feyereisen, R., **1996**. Sampling P450 diversity by cloning polymerase chain reaction products obtained with degenerate primers. *Methods Enzymol. 272*, 304–312.

Snyder, M.J., Stevens, J.L., Andersen, J.F., Feyereisen, R., **1995**. Expression of cytochrome P450 genes of the CYP4 family in midgut and fat body of the tobacco hornworm, *Manduca sexta*. *Arch. Biochem. Biophys. 321*, 13–20.

Snyder, M.J., Walding, J.K., Feyereisen, R., **1994**. Metabolic fate of the allelochemical nicotine in the tobacco hornworm, *Manduca sexta*. *Insect Biochem. Mol. Biol. 24*, 837–846.

Spellman, P.T., Rubin, G.M., **2002**. Evidence for large domains of similarly expressed genes in the *Drosophila* genome. *J. Biol. 1*, 5.

Spiegelman, V.S., Fuchs, S.Y., Belitsky, G.A., **1997**. The expression of insecticide resistance-related cytochrome P450 forms is regulated by molting hormone in *Drosophila melanogaster*. *Biochem. Biophys. Res. Commun. 232*, 304–307.

Stathopoulos, A., Van Drenth, M., Erives, A., Markstein, M., Levine, M., **2002**. Whole-genome analysis of dorsal–ventral patterning in the *Drosophila* embryo. *Cell 111*, 687–701.

Stevens, J.L., Snyder, M.J., Koener, J.F., Feyereisen, R., **2000**. Inducible P450s of the CYP9 family from larval *Manduca sexta* midgut. *Insect Biochem. Mol. Biol. 30*, 559–568.

Stoilov, I., Akarsu, A.N., Sarfarazi, M., **1997**. Identification of three different truncating mutations in cytochrome P4501B1 (CYP1B1) as the principal cause of primary congenital glaucoma (Buphthalmos) in families linked to the GLC3A locus on chromosome 2p21. *Hum. Mol. Genet. 6*, 641–647.

Sundseth, S.S., Nix, C.E., Waters, L.C., **1990**. Isolation of insecticide resistance-related forms of cytochrome P-450 from *Drosophila melanogaster*. *Biochem. J. 265*, 213–217.

Sutherland, T.D., Unnithan, G.C., Andersen, J.F., Evans, P.H., Murataliev, M.B., *et al.*, **1998**. A cytochrome P450 terpenoid hydroxylase linked to the suppression of insect juvenile hormone synthesis. *Proc. Natl Acad. Sci. USA 95*, 12884–12889.

Sutherland, T.D., Unnithan, G.C., Feyereisen, R., **2000**. Terpenoid omega-hydroxylase (CYP4C7) messenger RNA levels in the corpora allata: a marker for ovarian control of juvenile hormone synthesis in *Diploptera punctata*. *J. Insect Physiol. 46*, 1219–1227.

Suwanchaichinda, C., Brattsten, L.B., **2002**. Induction of microsomal cytochrome P450s by tire-leachate compounds, habitat components of *Aedes albopictus* mosquito larvae. *Arch. Insect Biochem. Physiol. 49*, 71–79.

Tamura, T., Thibert, C., Royer, C., Kanda, T., Abraham, E., *et al.*, **2000**. Germline transformation of the silkworm *Bombyx mori* L. using a piggyBac transposon-derived vector. *Nat. Biotechnol. 18*, 81–84.

Tares, S., Berge, J., Amichot, M., **2000**. Cloning and expression of cytochrome P450 genes belonging to the CYP4 family and to a novel family, CYP48, in two hymenopteran insects, *Trichogramma cacoeciae* and *Apis mellifera*. *Biochem. Biophys. Res. Commun. 268*, 677–682.

Taylor, M., Feyereisen, R., **1996**. Molecular biology and evolution of resistance of toxicants. *Mol. Biol. Evol. 13*, 719–734.

Terriere, L.C., Yu, S.J., **1974**. The induction of detoxifying enzymes in insects. *J. Agric. Food Chem. 22*, 366–373.

Thomas, J.D., Ottea, J.A., Boethel, D.J., Ibrahim, S., **1996**. Factors influencing pyrethroid resistance in a permethrin-resistant strain of the soybean looper, *Pseudoplusia includens* (Walker). *Pestic. Biochem. Physiol. 55*, 1–9.

Thompson, M., Svoboda, J., Lusby, W., Rees, H., Oliver, J., *et al.*, **1985**. Biosynthesis of a C21 steroid conjugate in an insect. The conversion of [^{14}C]cholesterol

to 5-[^{14}C]pregnen-3 beta, 20 beta-diol glucoside in the tobacco hornworm, *Manduca sexta*. *J. Biol. Chem. 260*, 15410–15412.

Tijet, N., Helvig, C., Feyereisen, R., **2001**. The cytochrome P450 gene superfamily in *Drosophila melanogaster*: annotation, intron-exon organization and phylogeny. *Gene 262*, 189–198.

Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., *et al.*, **2002**. Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol. 3*, 88.1–88.14.

Tomita, T., Liu, N., Smith, F.F., Sridhar, P., Scott, J.G., **1995**. Molecular mechanisms involved in increased expression of a cytochrome P450 responsible for pyrethroid resistance in the housefly, *Musca domestica*. *Insect Mol. Biol. 4*, 135–140.

Tomita, T., Scott, J.G., **1995**. cDNA and deduced protein sequence of CYP6D1: the putative gene for a cytochrome P450 responsible for pyrethroid resistance in housefly. *Insect Biochem. Mol. Biol. 25*, 275–283.

Ueda, H.R., Matsumoto, A., Kawamura, M., Iino, M., Tanimura, T., *et al.*, **2002**. Genome-wide transcriptional orchestration of circadian rhythms in *Drosophila*. *J. Biol. Chem. 277*, 14048–14052.

Ugaki, M., Shono, T., Fukami, J.I., **1985**. Metabolism of fenitrothion by organophosphorus-resistant and susceptible house flies, *Musca domestica* L. *Pestic. Biochem. Physiol. 23*, 33–40.

Unnithan, G.C., Andersen, J.F., Hisano, T., Kuwano, E., Feyereisen, R., **1995**. Inhibition of juvenile hormone biosynthesis and methyl farnesoate epoxidase activity by 1,5-disubstituted imidazoles in the cockroach, *Diploptera punctata*. *Pestic. Sci. 43*, 13–19.

Valles, S.M., Koehler, P.G., Brenner, R.J., **1997**. Antagonism of fipronil toxicity by piperonyl butoxide and *S*,*S*,*S*-tributyl phosphorotrithioate in the German cockroach (Dictyoptera: Blattelliidae). *J. Econ. Entomol. 90*, 1254–1258.

Wada, A., Waterman, M.R., **1992**. Identification by site-directed mutagenesis of two lysine residues in cholesterol side chain cleavage cytochrome P450 that are essential for adrenodoxin binding. *J. Biol. Chem. 267*, 22877–22882.

Walsh, J.B., **1987**. Sequence-dependent gene conversion: can duplicated genes diverge fast enough to escape conversion? *Genetics 117*, 543–557.

Wang, L., Bieber Urbauer, R.J., Urbauer, J.L., Benson, D.R., **2003**. Housefly cytochrome b_5 exhibits kinetically trapped hemin and selectivity in hemin binding. *Biochem. Biophys. Res. Commun. 305*, 840–845.

Wang, Q., Hasan, G., Pikielny, C.W., **1999**. Preferential expression of biotransformation enzymes in the olfactory organs of *Drosophila melanogaster*, the antennae. *J. Biol. Chem. 274*, 10309–10315.

Wang, X.P., Hobbs, A.A., **1995**. Isolation and sequence analysis of a cDNA clone for a pyrethroid inducible cytochrome P450 from *Helicoverpa armigera*. *Insect Biochem. Mol. Biol. 25*, 1001–1009.

Wang, M., Roberts, D.L., Paschke, R., Shea, T.M., Masters, B.S., *et al.*, **1997**. Three-dimensional structure of NADPH-cytochrome P450 reductase: prototype for FMN- and FAD-containing enzymes. *Proc. Natl Acad. Sci. USA 94*, 8411–8416.

Warren, J.T., Petryk, A., Marques, G., Jarcho, M., Parvy, J.P., *et al.*, **2002**. Molecular and biochemical characterization of two P450 enzymes in the ecdysteroidogenic pathway of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA 99*, 11043–11048.

Warren, J.T., Wismar, J., Subrahmanyam, B., Gilbert, L.I., **2001**. Woc (without children) gene control of ecdysone biosynthesis in *Drosophila melanogaster*. *Mol. Cell. Endocrinol. 181*, 1–14.

Waters, L.C., Nix, C.E., **1988**. Regulation of insecticide resistance-related cytochrome P-450 expression in *Drosophila melanogaster*. *Pestic. Biochem. Physiol. 30*, 214–227.

Waters, L.C., Zelhof, A.C., Shaw, B.J., Ch'ang, L.Y., **1992**. Possible involvement of the long terminal repeat of transposable element 17.6 in regulating expression of an insecticide resistance-associated P450 gene in *Drosophila*. [published erratum appears in Proc. Natl Acad. Sci. USA 1992, 89, 12209]. *Proc. Natl Acad. Sci. USA 89*, 4855–4859.

Wei, P., Zhang, J., Egan-Hafley, M., Liang, S., Moore, D.D., **2000**. The nuclear receptor CAR mediates specific xenobiotic induction of drug metabolism. *Nature 407*, 920–923.

Weirich, G.F., Williams, V.P., Feldlaufer, M.F., **1996**. Ecdysone 20-hydroxylation in *Manduca sexta* midgut: kinetic parameters of mitochondrial and microsomal ecdysone 20-monooxygenases. *Arch. Insect Biochem. Physiol. 31*, 305–312.

Weller, G.L., Foster, G.G., **1993**. Genetic maps of the sheep blowfly *Lucilia cuprina*: linkage-group corelations with other dipteran genera. *Genome 36*, 495–506.

Wen, Z., Horak, C.E., Scott, J.G., **2001**. CYP9E2, CYP4C21 and related pseudogenes from German cockroaches, *Blattella germanica*: implications for molecular evolution, expression studies and nomenclature of P450s. *Gene 272*, 257–266.

Wen, Z., Pan, L., Berenbaum, M.B., Schuler, M.A., **2003**. Metabolism of linear and angular furanocoumarins by *Papilio polyxenes* CYP6B1 co-expressed with NADPH cytochrome P450 reductase. *Insect Biochem. Mol. Biol. 33*, 937–947.

Wen, Z., Scott, J.G., **2001a**. Cytochrome P450 CYP6L1 is specifically expressed in the reproductive tissues of adult male German cockroaches, *Blattella germanica* (L.). *Insect Biochem. Mol. Biol. 31*, 179–187.

Wen, Z., Scott, J.G., **2001b**. Cloning of two novel P450 cDNAs from German cockroaches, *Blattella germanica* (L.): CYP6K1 and CYP6J1. *Insect Mol. Biol. 10*, 131–137.

Werck-Reichhart, D., Feyereisen, R., **2000**. Cytochromes P450: a success story. *Genome Biol. 1*, 3003.1–3003.9.

Wheeler, G.S., Slansky, F., Jr., Yu, S.J., **1993**. Fall armyworm sensitivity to flavone: limited role of constitutive

and induced detoxifying enzyme activity. *J. Chem. Ecol. 19*, 645–667.

Wheelock, G.D., Scott, J.G., **1989**. Simultaneous purification of a cytochrome P-450 and cytochrome b_5 from the housefly, *Musca domestica* L. *Insect Biochem. 19*, 481–488.

Wheelock, G.D., Scott, J.G., **1990**. Immunological detection of cytochrome P450 from insecticide resistant and susceptible house flies (*Musca domestica*). *Pestic. Biochem. Physiol. 38*, 130–139.

Wheelock, G.D., Scott, J.G., **1992a**. Anti-P450lpr antiserum inhibits specific monooxygenase activities in LPR housefly microsomes. *J. Exp. Zool. 264*, 153–158.

Wheelock, G.D., Scott, J.G., **1992b**. The role of cytochrome P450 in deltamethrin metabolism by pyrethroid-resistant and susceptible strains of house flies. *Pestic. Biochem. Physiol. 43*, 67–77.

White, J.R.A., Franklin, R.T., Agosin, M., **1979**. Conversion of α-pinene to α-pinene oxide by rat liver and the bark beetle *Dendroctonus terebrans* microsomal fractions. *Pestic. Biochem. Physiol. 10*, 233–242.

White, K.P., Rifkin, S.A., Hurban, P., Hogness, D.S., **1999**. Microarray analysis of *Drosophila* development during metamorphosis. *Science 286*, 2179–2184.

Whittaker, R.H., Feeny, P.P., **1971**. Allelochemics: chemical interactions between species. *Science 171*, 757–770.

Wilkinson, C.F., **1979**. The use of insect subcellular components for studying the metabolism of xenobiotics. *ACS Symp. Ser. 97*, 249–284.

Wilkinson, C.F., Brattsten, L.B., **1972**. Microsomal drug metabolizing enzymes in insects. *Drug Metab. Rev. 1*, 153–228.

Williams, D.R., Chen, J.H., Fisher, M.J., Rees, H.H., **1997**. Induction of enzymes involved in molting hormone (ecdysteroid) inactivation by ecdysteroids and an agonist, 1,2-dibenzoyl-1-*tert*-butylhydrazine (RH-5849). *J. Biol. Chem. 272*, 8427–8432.

Williams, P.A., Cosme, J., Sridhar, V., Johnson, E.F., McRee, D.E., **2000a**. Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. *Mol. Cell 5*, 121–131.

Williams, D.R., Fisher, M.J., Rees, H.H., **2000b**. Characterization of ecdysteroid 26-hydroxylase: an enzyme involved in molting hormone inactivation. *Arch. Biochem. Biophys. 376*, 389–398.

Wilson, T.G., **2001**. Resistance of *Drosophila* to toxins. *Annu. Rev. Entomol. 46*, 545–571.

Winter, J., Bilbe, G., Richener, H., Sehringer, B., Kayser, H., **1999**. Cloning of a cDNA encoding a novel cytochrome P450 from the insect *Locusta migratoria*: CYP6H1, a putative ecdysone 20-hydroxylase. *Biochem. Biophys. Res. Commun. 259*, 305–310.

Winter, J., Eckerskorn, C., Waditschatka, R., Kayser, H., **2001**. A microsomal ecdysone-binding cytochrome P450 from the insect *Locusta migratoria* purified by sequential use of type-II and type-I ligands. *Biol. Chem. 382*, 1541–1549.

Wojtasek, H., Leal, W.S., **1999**. Degradation of an alkaloid pheromone from the pale-brown chafer, *Phyllopertha diversa* (Coleoptera: Scarabaeidae), by an insect olfactory cytochrome P450. *FEBS Lett. 458*, 333–336.

Xu, W., Bak, S., Decker, A., Paquette, S.M., Feyereisen, R., *et al.*, **2001**. Microarray-based analysis of gene expression in very large gene families: the cytochrome P450 gene superfamily of *Arabidopsis thaliana*. *Gene 272*, 61–74.

Yamazaki, S., Sato, K., Suhara, K., Sakaguchi, M., Mihara, K., *et al.*, **1993**. Importance of the proline-rich region following signal-anchor sequence in the formation of correct conformation of microsomal cytochrome P-450s. *J. Biochem. (Tokyo) 114*, 652–657.

Yu, S.J., **1984**. Interactions of allelochemicals with detoxication enzymes of insecticide-susceptible and resistant fall armyworms. *Pestic. Biochem. Physiol. 22*, 60–68.

Yu, S.J., **1986**. Consequences of induction of foreign compound-metabolizing enzymes in insects. In: Brattsten, L.B., Ahmad, S. (Eds.), Molecular Aspects of Insect–Plant Interactions. Plenum, New York, pp. 211–255.

Yu, S.J., **1987**. Microsomal oxidation of allelochemicals in generalist (*Spodoptera frugiperda*) and semispecialist (*Anticarsia gemmatalis*) insects. *J. Chem. Ecol. 13*, 423–436.

Yu, S.J., **1991**. Insecticide resistance in the fall armyworm, *Spodoptera frugiperda* (J.E. Smith). *Pestic. Biochem. Physiol. 39*, 84–91.

Yu, S.J., **1992**. Detection and biochemical characterization of insecticide resistance in fall armyworm (Lepidoptera: Noctuidae). *J. Econ. Entomol. 85*, 675–682.

Yu, S.J., **1995**. Allelochemical stimulation of ecdysone 20-monoxygenase in fall armyworm larvae. *Arch. Insect Biochem. Physiol. 28*, 365–375.

Yu, S.J., Terriere, L.C., **1974**. A possible role for microsomal oxidases in metamorphosis and reproduction in the housefly. *J. Insect Physiol. 20*, 1901–1912.

Zdobnov, E.M., von Mering, C., Letunic, I., Torrents, D., Suyama, M., *et al.*, **2002**. Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science 298*, 149–159.

Zhang, L., Harada, K., Shono, T., **1997**. Genetic analysis of pyriproxifen resistance in the housefly, *Musca domestica* L. *Appl. Ent. Zool. 32*, 217–226.

Zhang, L., Kasai, S., Shono, T., **1998**. In vitro metabolism of pyriproxyfen by microsomes from susceptible and resistant housefly larvae. *Arch. Insect Biochem. Physiol. 37*, 215–224.

Zhang, M., Scott, J.G., **1994**. Cytochrome b_5 involvement in cytochrome P450 monooxygenase activities in housefly microsomes. *Arch. Insect Biochem. Physiol. 27*, 205–216.

Zhang, M., Scott, J.G., **1996a**. Purification and characterization of cytochrome b_5 reductase from the housefly, *Musca domestica*. *Comp. Biochem. Physiol. 113B*, 175–183.

Zhang, M., Scott, J.G., **1996b**. Cytochome b_5 is essential for cytochome P450 6D1-mediated cypermethrin resistance in LPR house flies. *Pestic. Biochem. Physiol. 55*, 150–156.

Zhou, S., Gao, Y., Jiang, W., Huang, M., Xu, A., *et al.*, **2003**. Interactions of herbs with cytochrome P450. *Drug Metab. Rev. 35*, 35–98.

Zhu, Y.C., Snodgrass, G.L., **2003**. Cytochrome P450 CYP6X1 cDNAs and mRNA expression levels in three strains of the tarnished plant bug *Lygus lineolaris* (Heteroptera: Miridae) having different susceptibilities to pyrethroid insecticide. *Insect Mol. Biol. 12*, 39–49.

Ziegler, D.M., **2002**. An overview of the mechanism, substrate specificities, and structure of FMOs. *Drug Metab. Rev. 34*, 503–511.

Zijlstra, J.A., Vogel, E.W., Breimer, D.D., **1984**. Strain-differences and inducibility of microsomal oxidative enzymes in *Drosophila melanogaster* flies. *Chem. Biol. Interact. 48*, 317–338.

Zinke, I., Schutz, C.S., Katzenberger, J.D., Bauer, M., Pankratz, M.J., **2002**. Nutrient control of gene expression in *Drosophila*: microarray analysis of starvation and sugar-dependent response. *EMBO J. 21*, 6162–6173.

Zou, S., Meadows, S., Sharp, L., Yan, L.Y., Jan, Y.N., **2000**. Genome-wide study of aging and oxidative stress response in *Drosophila melanogaster. Proc. Natl Acad. Sci. USA 97*, 13726–13731.

Zuber, M.X., Simpson, E.R., Waterman, M.R., **1986**. Expression of bovine 17α-hydroxylase cytochrome P450 cDNA in non-steroidogenic (COS1) cells. *Science 234*, 1258–1261.

Zumwalt, J.G., Neal, J.J., **1993**. Cytochromes P450 from *Papilio polyxenes*: adaptations to host plant allelochemicals. *Comp. Biochem. Physiol. 106C*, 111–118.

Relevant Websites

http://www.imm.ki.se – Human cytochrome P450 (CYP) allele nomenclature committee. A complete referenced and annotated list of human P450 alleles.

http://P450.antibes.inra.fr – The insect P450 site at INRA. A searchable annotation of insect cytochrome P450 sequences.

http://dinelson.utmem.edu – David Nelson's cytochrome P450 homepage. Listing of P450 genes from a variety of organisms.

4.2 Cuticular Proteins

J H Willis, University of Georgia, Athens, GA, USA
V A Iconomidou, University of Athens, Athens, Greece
R F Smith, Northern Arizona University, Flagstaff, AZ, USA
S J Hamodrakas, University of Athens, Athens, Greece

© 2005, Elsevier BV. All Rights Reserved.

4.2.1. Introduction

In the previous edition of this series, Silvert (1985) outlined several major areas of uncertainty regarding cuticular proteins. The questions raised were: Were proteins extracted from cuticle authentic cuticular proteins or might some be contaminants of adhering cells and hemolymph? Was the epidermis the sole site of synthesis of cuticular proteins or were some synthesized in other tissues and transported to the cuticle? What was the relation among cuticular proteins of various developmental stages? Did cuticular proteins share common structural features?

That article presented all the cuticular protein sequence data then available – four complete and three partial sequences from *Drosophila melanogaster* and one partial sequence from *Sarcophaga bullata*. The *considerable* sequence similarity seen with those limited data indicated that cuticular protein genes belonged to multigene families, and the even more limited genomic information revealed that similar genes were adjacent on a chromosome.

Progress in less than two decades has been spectacular, but not surprising given the advances in relevant techniques. Elegant immunolocalization analyses have solved the problem of the sources of cuticular proteins. Over 300 cuticular protein sequences are now available from six orders and over 20 species of insects. Listed in **Tables 1** and **2** are all but those that come exclusively from the annotation of the genomes of *D. melanogaster* and *Anopheles gambiae* or SilkBase, an extensive expressed sequence tag (EST) project in *Bombyx mori* (Mita *et al.*, 1999; SilkBase, 2003). Throughout this chapter, proteins are referred to by the names used in **Tables 1** and **2**. These tables also provide a gi number or an identifier from SwissProt that provides access to the complete sequence and the relevant references. Recombinant proteins have

Table 1 Characteristics of ''structural'' cuticular proteins that have been completely sequenced

Order/species	*Protein name*	*Number of amino acids*[a]	*Type*	*AAP(A/V) Repeats*[b]	*Other features*[c]	*Sequence method*[d]	*Identifier*[e]
Coleoptera							
Apriona germari	LCP10.7	87	RR-1			CT	16226511
	LCP12.3	114	RR-1			CT	21617523
	LCP12.6	120	RR-1			CT	21617525
Tenebrio molitor	TM-LCP-A1A	174	RR-2	2N 1C	2[AAP(I/L)]	DS	1706191
	TM-LCP-A2B	117	RR-2	1N 1C		DS	1706192
	TM-LCP-A3A	134	RR-2	1N 1C	1[AAP(I/L)]	DS	1706194
	TM-PCP-C1B	161		9	51 aa motif 1[AAP(I/L)]	DS	1706197
	TM-E1A	243		22		DS	913040
	TM-F1A	154		3	5 [AAP(I/L)]	DS	998953
	TM1-F1B	158		4	4 [AAPL]	DS	998954
	TM-F1C	162		4	5 [AAP(I/L)]	DS	998955
	TM-PCP-G1A	211		5	51 aa motif	DS	1706198
	TM-H1C = TMLPCP22	195		4	51 aa motif 1[AAP(I/L)]	CT DS	3123202*
	PCP5.8	53				DS	7511760
	PCP9.2	92				DS	7511761
	PCP15.6	161		5	1[AAP(I/L)]	DS	7441995
	PCP16.7	166				DS	7511759
	TMLPCP-23	214		2	51 aa motif 1[AAP(I/L)]	CT	3121955*
	TMLPCP29	276	RR-3	11N	1 (18 residue motif) 1[AAP(I/L)]	CT	2275132*
	ACP17	167		2	High G	CT	1078986*
	ACP20	191	RR-2		High G	CT	102879
	ACP-22	180	RR-2		2(6G), 1(8G), 1(15G)	CT	113012*
Dictyoptera							
Blaberus craniifer	BC-NCP1	87			Has 6 Cys, forming 3 potential S-S	DS	3023587
	BC-NCP2	99			8 As; 2 (18 residue motif)	DS	3023589
	BC-NCP4	127			2 (18 residue motif)	DS	3023588
	BC-NCP5	145			2 (18 residue motif)	DS	P82118
	BC-NCP6	139	RR-3		1 (18 residue motif)	DS	P82119
	BC-NCP7	145	RR-3		1 (18 residue motif)	DS	P82120
	BC-NCP8	195	RR-2			DS	P82121
	BC-NCP9	34				DS	P82122
Diptera							
Anopheles gambiae	AGCP2A	214	RR-2	1N	4 [(S/A)APIAH]	CT	2961109*
	AGCP2B	222	RR-2	1N	5 [(S/A)APIAH]	CT	2961110*
	AGCP2C	214	RR-2	2N	5 [(S/A)APIAH]	CT	2961111*
	AGCP2D	215	RR-2	2N	4 [(S/A)APIAH]	CT	2961113
Drosophila melanogaster	Dacp-1	116		3		CT	7580461*
	LCP-1	114	RR-1			DS CT	17380379*
	LCP-2	110	RR-1			DS CT	117634*
	LCP-3	96	RR-1			DS CT	117635*
	LCP-4	96	RR-1			DS CT	117636*
	PCP = GART INTRON	166	RR-1			CT	157483*
	ACP65A	85	RR-1			CT	1857602*
	LCP65Aa	83	RR-1			CT	1857600*
	LCP65Ab1	86	RR-1			CT pDS	1857597*
	LCP65Ab2	86	RR-1			CT pDS	1857595*
	LCP65Ac	91	RR-1			CT	1857593*
	LCP65Ad	90	RR-1			CT	1857495*
	LCP65Ae	83	RR-1			CT	1857604*
	LCP65Af	84	RR-1			CT	1857606*
	LCP65Ag1	87	RR-1			CT pDS	1857608*
	LCP65Ag2	87	RR-1			CT pDS	1857610*
	EDG-78	106	RR-1			CT	117639*
	EDG-84	171	RR-2	1C		CT	117640*

Table 1 Continued

Order/species	*Protein name*	*Number of amino acids*[a]	*Type*	*AAP(A/V) Repeats*[b]	*Other features*[c]	*Sequence method*[d]	*Identifier*[e]
	EDG91	138			8 (2G) 6(3G) 3 (4G)	CT	17380419*
	Ccp84Aa	188	RR-2	1N 6C		CT	4389433*
	Ccp84Ab	204	RR-2	1N 7C		CT	4389434*
	Ccp84Ac	199	RR-2	2C		CT	4389435*
	Ccp84Ad	182	RR-2	1N 6C		CT	4389436*
	Ccp84Ae	191	RR-2	1N 3C		CT	4389437*
	Ccp84Af	134	RR-2	1N		CT	4389438*
	Ccp84Ag	173	RR-2	6C		CT	4389439*
	Cry	457	RR-2		7Q, internal M	CT	22946279*
Drosophila miranda	LCP-1	122	RR-1			CT	3023591*
	LCP-2	110	RR-1			CT pDS	3023592
	LCP-3	96	RR-1			CT	231917*
	LCP-4	96	RR-1			CT	231917*
	LCP-3Y	96	RR-1			CT pDS	386246
Drosophila pseudoobscura	CP = GART INTRON	177	RR-1			CT	435017*
Drosophila simulans	DS-06238.4-like	188	RR-2			CT	9966434*
Drosophila yakuba	DS-06238.4-like	194	RR-2			CT	9966436*
Lucilia cuprina	CUT1	102	RR-1			CT	2565392
	CUT12	89	RR-1			CT	2565394
Hemiptera							
Myzus persicae	CP	208	RR-2	1N 4C	[(S/K)APAY] 1N 10C	CT	16798648
Aphis fabae	CP	208	RR-2	3C	[(S/K)APAY] 1N 12C	CT	29124938
Aphis gossypii	CP	203	RR-2	1N 3C	[(S/K)APAY] 10C	CT	29124934
Brevicoryne brassicae	CP	208	RR-2	1N 2C	[(S/K)APAY] 1N 12C	CT	29124932
Lipaphis erysimi	CP	208	RR-2	1N 3C	[(S/K)APAY] 1N 11C	CT	29124930
Rhopalosiphum maidis	CP	210	RR-2	2N 2C	[(S/K)APAY] 13C	CT	29124936
Lepidoptera							
Bombyx mori	PCP	235		3	3 (18 residue motif)	CT	1169137*
	BMWCP1A	205	RR-2	5N		CT	12862579
	BMWCP1B	205	RR-2	5N		CT	12862581
	BMWCP2	231	RR-2	9N		CT	12862583
	BMWCP3	215	RR-2	7N		CT	12862585
	BMWCP4	226	RR-2	3N		CT	12862587
	BMWCP5	274	RR-2	2N 1C		CT	12862589
	BMWCP6	186	RR-2	2N 2C		CT	
	BMWCP7A	157	RR-2	1N		CT	12862593
	BMWCP7B	157	RR-2	1N		CT	12862595
	BMWCP8	221	RR-2			CT	12862597
	BMWCP9	158	RR-1			CT	12862599
	BMWCP10	295	RR-1		Internal M	CT	23096118
	EDG84A homologue	180	RR-2	1N 1C		CT	3608259
	BMCP17	127	RR-1	1N		CT pDS	2204069
	BMCP18	89	RR-1	1N		CT pDS	5360249*
	BMCP22	158	RR-1			CT pDS	2204071
	BMCP30	223	RR-1			CT pDS	6634056*
	GCP1	149			17(GGY)	CT	15146344
Galleria melonella	PCP-52	338		1	Has 2 Cys; 21% Ala	CT	1086154
Helicoverpa armigera	LCP-1	95	RR-1		VVVV	CT	3913261
Hyalophora cecropia	HCCP12	89	RR-1			CT pDS	1169129*

Continued

Table 1 Continued

Order/species	*Protein name*	*Number of amino acids*[a]	*Type*	*AAP(A/V) Repeats*[b]	*Other features*[c]	*Sequence method*[d]	*Identifier*[e]
	HCCP66	112	RR-2			CT pDS	1169133*
Manduca sexta	LCP-14	109	RR-1			CT	117623
	CP14.6	90	RR-1			CT	3121956*
	LCP16/17	123	RR-1			CT	3121953*
	CP20	182	RR-1		7(GG) 3(GGG); glycosylated	CT DS	19548965
	CP27	165	RR-1		Glycosylated	CT DS	19743772
	CP36	327	RR-1		32(GG) 1(GGG); glycosylated	CT DS	22000820
Orthoptera							
Locusta migratoria	LM-ACP-abd4	116	RR-1		3 glycosylated forms	DS	461860
	LM-ACP-abd5	82	RR-1		Glycosylated	DS	3913394
	LM-ACP7	131	RR-2	2N 3C		DS	998751
	LM-ACP8	148	RR-2	5N		DS	84730
	LM-ACP19	157	RR-2	6N 1C	GYL motif	DS	1345864
	LM-ACP21	169	RR-2	6N 3C		DS	3287770
	LM-ACP38	163		14	GYL motif	DS	72263
	LM-NCP55	33		2		DS	446069
	LM-NCP62	88				DS	446070
	LM-ACP63	157		8	GYL motif	DS	1169130
	LM-ACP64	152		7	GYL motif	DS	1169131
	LM-ACP65	145		7	GYL motif	DS	1169132
	LM-ACP67a	98		6	51 aa motif	DS	416850
	LM-ACP67b	104		6	51 aa motif	DS	542520
	LM-ACP70	88		4	GYL motif	DS	416852
	LM-ACP76	139		6	GYL motif; 51 aa motif	DS	1169134
	LM-ACP79a	131		3	5 [GGG(L/Y)]	DS	1168721
	LM-ACP79b	131		3	5 [GGG(L/Y)]	DS	1168722
	LM-NCP4.9	46		1		DS	P82168
	LM-NCP5.1	48		1		DS	P82169
	LM-NCP6.4	62				DS	P82170
	LM-NCP9.5	90				DS	P82171
	LM-NCP18.7	193			4 (18 residue motif)	DS	P82165
	LM-NCP21.3	200		10		DS	P82167
	LM-NCP19.8	200	RR-2	3N 5C		DS	P82166
Schistocerca gregaria	SGAbd-1	184	RR-1		Glycosylated; (PPPPPPP)	DS	7511754
	SGAbd-2	135	RR-1		Glycosylated	DS	7511755
	SGAbd-3	119	RR-1		Glycosylated	DS	7441999
	SGAbd-4	116	RR-1		Glycosylated	DS	7441997
	SGAbd-5	82	RR-1		Glycosylated	DS	3913395
	SGAbd-6	82	RR-1		Glycosylated	DS	7511756
	SGAbd-8	139	RR-1		Glycosylated; internal M	DS	7511757
	SGAb-9	129	RR-1		Glycosylated	DS	7441998

[a]Sequence length of mature peptide; signal peptides were deleted using data from authors or SignalP V2.0 (http://www.cbs.dtu.dk/services/SignalP-2.0/).

[b]If a protein has an R&R Consensus, location of the AAP(A/V) repeats is given relative to the consensus.

[c]See Section 4.2.3.2.2 for description of these features.

[d]DS, direct sequencing of protein, a p indicates only a partial (generally N-terminal) sequence was obtained; CT, conceptual translation of a cDNA, genomic region, or EST product.

[e]Protein sequences and additional annotation can be found at http://www3.ncbi.nlm.nih.gov/Entrez/index.html. Sequences that have an identifier that begins with a letter can be found at http://us.expasy.org. An asterisk indicates that genomic sequence information is available.

Table 2 Characteristics of some nonstructural proteins that have been found in cuticle

Species	*Protein name*	*Number of amino acids*[a]	*Function*	*Sequence method*[b]	*Identifier*[c]
Schistocerca gregaria	Putative carotene binding protein	250	Transfers carotene into cuticle	DS	13959527
Caliphora vicinia	ARYLPHORIN A4	743	Found in cuticle	CT	114232
	ARYLPHORIN C223	743		CT	114236
Drosophila melanogaster	YELLOW	520	Positions melanin pigment in cuticle	CT	140623
Bombyx mori	CECROPIN A	41	Defense protein	CT	2493573
	CECROPIN B	41	Defense protein	CT	1705754
	PROPHENOLOXIDASE	675	Melanization enzyme	CT	13591614
Calpodes ethlius	CECP22	169	Cuticle digestion	CT	4104409
Manduca sexta	ARYLPHORINα	684		CT	114240
	ARYLPHORINβ	687		CT	1168527
	INSECTICYANIN A	189	Blue pigment	CT	102968
	INSECTICYANIN B	189	Blue pigment	CT	124527
	SCOLEXIN A	279	Serine protease immune protein	CT	4262357
	SCOLEXIN B	279	Serine protease immune protein	CT	4262359

[a]Sequence length of mature peptide; signal peptides were deleted using data from authors or SignalP V2.0 (http://www.cbs.dtu.dk/services/SignalP-2.0/).
[b]DS, direct sequencing of protein; CT, conceptual translation of a cDNA, genomic region, or EST product.
[c]Protein sequences and additional annotation can be found at: http://www3.ncbi.nlm.nih.gov/Entrez/index.html.

revealed a function for a highly conserved domain that was present in those first protein sequences discussed by Silvert. Structural predictions have elucidated the basis for this function. It is these developments that will be the focus of this review.

4.2.2. Cuticle Structure and Synthesis

4.2.2.1. Cuticle Morphology

4.2.2.1.1. Terminology The descriptive terms used here to describe the regions of cuticle have been simplified according to Locke's (2001) cogent suggestions for new nomenclature. He proposes the use of the term "envelope" to describe the outermost layer of cuticle, rather than the previous term "cuticulin." At the start of each molt cycle, the smooth apical plasma membrane forms microvilli with plaques at their tips where the new envelope assembles. This discrete layer of 10–30 nm not only serves to protect the underlying epidermis from molting fluid enzymes that begin to digest the old cuticle, but, as Locke points out, affects "resistance to abrasion and infection, penetration of insecticides, permeability, surface reflectivity, and physical colors." The sequences and properties of its constituent proteins remain unknown.

Next formed is the epicuticle, about 1 μm in thickness. This chitin-free layer (but see Section 4.2.2.1.3) is stabilized by quinones. It was formerly referred to as the "inner epicuticle" with cuticulin being the outer.

Former arguments about the precise distinction between *exo*- and *endo*-cuticle are eliminated by Locke's lumping of the inner regions of the cuticle under the term "procuticle," encompassing both preecdysial and postecdysial secretions. The procuticle, then, is the region that combines chitin and cuticular proteins in various combinations and becomes sclerotized (see **Chapter 4.4**) and pigmented to varying degrees. This is the region depicted in electron micrographs showing stacks of precisely oriented lamellae. According to Locke (2001), apical microvilli bend in concert across the epithelial sheet and this movement serves to orient the laminae that will form lamellae. While knowledge of the process of secreting and assembling such a highly ordered structure is limited, details about the proteins associated with the lamellae are now voluminous.

4.2.2.1.2. Growth of the cuticle within an instar Central to the issue of cuticle structure is the important fact that considerable cuticle growth can occur during an intermolt period (Williams, 1980), some of it by a smoothing out of macro- and microscopic folds and pleats (Carter and Locke, 1993). During intrainstar growth, new cuticular proteins are interspersed among the old, necessitating a model of chitin–protein and protein–protein

interactions that will permit such intussusception (Condoulis and Locke, 1966; Wolfgang and Riddiford, 1986).

4.2.2.1.3. Localization of cuticular proteins within the cuticle Precise localization of cuticular proteins within the cuticle and even within cellular organelles has been made possible with immunogold labeling of electron microscopic sections. Here a specific primary antibody is bound to the sections and visualized with a secondary antibody conjugated to colloidal gold particles.

Antibodies have been raised against extracts of whole cuticle or isolated electrophoretic bands and the specificity of each antibody ascertained with Western blots. While each polyclonal antibody raised against a single band was specific for the immunizing protein, monoclonals raised against cuticular extracts frequently reacted with more than one electrophoretic band.

One concern with immunolocalization is that as cuticular proteins become modified in the cuticle by binding to chitin or by becoming sclerotized, the immunizing epitopes might become masked, a problem that should be more serious with monoclonal than polyclonal antibodies. All groups recognized that while the presence of an antigen is significant, its absence may reflect no more than such masking.

This concern is significant when one considers results of immunolocalization in the assembly zone, the region of cuticle directly above the microvilli. It is here that chitin secreted from the tips of the microvilli interacts with cuticular proteins secreted into the perimicrovillar space. Immunolocalization studies revealed only a few of the cuticular proteins within the perimicrovillar space but the same ones and others were abundant in the assembly zone directly above it (Locke *et al*., 1994; Locke, 1998). The authors' conclusion was that the assembly zone "is where we should expect proteins to unravel and expose most epitopes in preparation for assuming a new configuration as they stabilize in the maturing cuticle." Wolfgang *et al*. (1986, 1987) found two *D. melanogaster* cuticular proteins exclusively in this zone and suggested they might function in cuticle assembly. Locke *et al*. (1994) point out that it was common for antibodies raised against *Calpodes ethlius* proteins to react more strongly with the assembly zone than with more mature regions of cuticle where sclerotization and chitin binding might mask epitopes. Thus more substantial evidence than the failure to detect a protein in more mature regions is needed to confirm that it belonged exclusively to the assembly zone.

It was known from earlier work on protein and mRNA distribution that cuticles from different metamorphic stages and different anatomical regions had different cuticular proteins and that there may be a change in cuticular proteins synthesized by a single cell within a molt cycle (review: Willis, 1996). Such a transition in proteins synthesized is especially apparent at the time of ecdysis, and, in some insects, late in the instar. Consistent with this, immunolocalization revealed different proteins in morphologically distinct early and late lamellae in *D. melanogaster* pupae, and *Tenebrio molitor* and *Manduca sexta* larvae (Doctor *et al*., 1985; Fristrom *et al*., 1986; Wolfgang and Riddiford, 1986; Wolfgang *et al*., 1986; Lemoine *et al*., 1989, 1993; Bouhin *et al*., 1992a, 1992b; Rondot *et al*., 1996). Only two proteins with known sequence are among this group, TMACP22 and TMLPCP22.

Csikos *et al*. (1999) have used immunohistochemistry to follow some of *Manduca*'s cuticular proteins throughout the molt cycle. These proteins are obviously in a dynamic state as they move from epidermis to cuticle to molting fluid to fat body and then apparently back to cuticle via the hemolymph. More detailed studies are needed to learn if the same molecules make the return trip, and whether their initial passage from molting fluid into the hemolymph is solely via uptake and then basal secretion by the epidermis or whether the midgut plays a role, since lepidopteran larvae drink their molting fluid (Cornell and Pan, 1983).

The findings with epicuticle, the first region to be secreted beneath the envelope, were complex. None of the monoclonal antibodies that recognized *Tenebrio* cuticular proteins reacted with epicuticle (Lemoine *et al*., 1990). On the other hand, arylphorin from *Calpodes* has been localized to epicuticle and no other cuticular region (Leung *et al*., 1989) and several proteins, of unknown sequence, were found both in the epicuticle and in the lamellar regions of the procuticle in *D. melanogaster* (Fristrom *et al*., 1986) and *Calpodes* (Locke *et al*., 1994). This finding of cuticular proteins in both epicuticle and lamellar regions was surprising, since the epicuticle had always been described as lacking chitin (cf. Fristrom *et al*., 1986; Fristrom and Fristrom, 1993) and thus was expected to have unique proteins.

In addition to temporal differences in the secretion of cuticular proteins by single cells, there may be regional differences in the cuticle secreted by single cells. Individual epidermal cells of the articulating membranes (intersegmental membranes) in *Tenebrio* secrete a cuticle with sclerotized cones

embedded in softer cuticle. Two of the classes of monoclonal antibodies raised against *Tenebrio*'s larval and pupal cuticular proteins recognized proteins in these cones. The same antibodies recognized proteins in cuticles in other regions that were destined to be sclerotized. Different antibodies recognized the proteins in the softer cuticle (Lemoine *et al.*, 1990, 1993).

Locke *et al.* (1994) were able, using carefully reconstructed sections of *Calpodes* larval cuticle, to distinguish one protein (C36) that was found with the same distribution as the chitin microfibrils that had been visualized with wheat germ agglutinin, a lectin that recognizes *N*-acetylglucosamine, while other antigens failed to show this distribution. Notably, only C36 isolated from cuticle reacted with wheat germ agglutinin on lectin blots. Based on this evidence Locke *et al.* (1994) suggest that the isolated protein may have obtained its *N*-acetylglucosamine from chitin.

4.2.2.1.4. Cuticles formed following disruption of normal metamorphosis Treatment of many insects with juvenile hormone (JH) causes them to resynthesize a cuticle with a morphology characteristic of the current metamorphic stage, rather than the next (see **Chapter 3.7**). Thus, in *Tenebrio*, treatment of pupae with JH prior to pupal–adult apolysis causes the formation of a second pupa rather than an adult. Earlier work revealed that these second pupae had proteins with the same electrophoretic mobility as those extracted from normal pupae (Roberts and Willis, 1980b; Lemoine *et al.*, 1989). A combination of Northern analysis and *in situ* hybridization demonstrated that second pupae have the same cuticular protein mRNAs and protein localization as normal pupae (Lemoine *et al.*, 1993; Rondot *et al.*, 1996). Adult cuticular proteins are not deposited in these cuticles and the adult mRNAs do not appear (Lemoine *et al.*, 1989, 1993; Bouhin *et al.*, 1992a, 1992b; Charles *et al.*, 1992). Some JH-treated *Tenebrio* pupae form two cuticles, the first pupalike in morphology and the second with adult features. The adultlike cuticle was shown with immunolocalization to have ACP22 (Bouhin *et al.*, 1992a). If JH is applied too late to form a perfect second pupa, the next cuticle formed will be a composite with morphological features of two metamorphic stages (Willis *et al.*, 1982). Bouhin *et al.* (1992b) found that all the epidermal cells laying down such a composite cuticle had mRNAs for ACP22.

Zhou and Riddiford (2002) used Northern analysis to characterize the somewhat nondescript cuticles made by *D. melanogaster* that had been manipulated by misexpressing the gene, *broad*, that codes for a transcription factor that appears before the larval–pupal molt in flies and moths. By following mRNAs for the adult cuticular protein ACP65A or the pupal cuticular protein Edg78E, they were able to demonstrate the essential role of *broad* in directing pupal development and thereby helped clarify the perplexing action of juvenoids in the higher Diptera.

4.2.2.2. The Site of Synthesis of Cuticular Proteins

One of the unresolved issues addressed in Silvert's (1985) review was the site of synthesis of cuticular proteins. This might appear to be a trivial issue, for one would expect that the epidermis that underlies the cuticle would synthesize the cuticular proteins. There are, however, reports in the literature that proteins found in the hemolymph were present in cuticle and even that labeled proteins injected into the hemolymph would appear in cuticle. Silvert discussed the possibility that the injected protein had been broken down and resynthesized so that the cuticular protein was labeled solely because its constituent amino acids had come from a labeled pool.

Five methods have now provided data that address the site of synthesis of cuticular proteins. The most common is to use Northern analysis to learn in what tissues and at which stages mRNA is present for a particular cuticular protein. This method is so common that specific examples will not be given. The second method is to incubate epidermis or integument *in vitro* with radioactive amino acids, separate the proteins, and compare the electrophoretic mobility of the labeled proteins to proteins isolated from cleaned cuticles. A third method is to isolate mRNAs from tissues and translate these *in vitro* with commercially available wheat germ extracts or rabbit reticulocytes and compare the translation products to known cuticular proteins. The fourth method is *in situ* hybridization, and the fifth immunolocalization to visualize proteins within the endoplasmic reticulum and Golgi apparatus.

The first three methods suffer from the possibility that tracheae and adhering tissues, fat body, muscles, hemocytes, contribute to the mRNA pool. Both labeling methods suffer from the problem that cuticular proteins are notoriously sensitive to solubilizing buffer and gel conditions (pH, urea concentration) (Cox and Willis, 1987a) and unless cuticular protein standards and labeled translation products are mixed prior to electrophoresis, they may not show identical electrophoretic mobility even in adjacent lanes. Some workers have precipitated labeled

translation products with antibodies raised against extracts of cuticle or individual cuticular proteins, then solublized the precipitate, run it on a gel, and detected the labeled product with fluorography. Csikos *et al*. (1999) used Western blots of translation products to identify cuticular proteins. Since cuticular proteins are destined for secretion from cells, they have a signal peptide that is cleaved before the protein is secreted into the cuticle. Hence, translation products made *in vitro* will be larger than the protein extracted from cuticle. There are two methods to circumvent this problem. The translation products can have their signal peptides cleaved by adding a preparation of canine microsomes, or antibodies against cuticular proteins (specific or against an extract) can be used to precipitate the translation products before they are solubilized and run on a gel. Either method allows some certainty in the comparison of these *in vitro* translation products with authentic cuticular proteins. It was also found that some commercial preparations of wheat germ extract have endogenous signal peptide processing activity (Binger and Willis, 1990).

Frequently, ^{35}S-methionine was used for metabolic labeling of integument and for *in vitro* translation. This is an unfortunate choice as almost all mature cuticular proteins lack methionine residues (see Section 4.2.3.2.1). The initiator methionine will be lost along with the entire signal peptide. Clear differences in labeling patterns with ^{35}S-methionine and ^{3}H-leucine have been found, with none of the major proteins from pharate adult cuticle of *D. melanogaster* or from larval cuticles of *Hyalophora cecropia* showing methionine labeling (Roter *et al*., 1985; Willis, 1999). Why then did several studies find all of the known cuticular proteins labeled with methionine? Perhaps the finding that ^{35}S-methionine can donate its label to a variety of amino acids in preformed proteins (Browder *et al*., 1992; Kalinich and McClain, 1992) explains its appearance and suggests that it needs to be used with caution for such studies with cuticular proteins.

The fourth method is *in situ* hybridization, where specific mRNAs can be identified in the epidermis. Results from several studies are summarized in **Table 3**. *In situ* hybridization allows one to be somewhat more discerning about the site of synthesis of a cuticular protein because it is possible to monitor the presence or absence of a particular mRNA at the level of an individual cell. With this technique, integument is fixed and sectioned and then probed with a labeled cDNA or cRNA allowing the identification of particular regions of the epidermis by examining the morphology of the overlying cuticle. With most detection methods, contaminating tissues and precise regions of the epidermis can be identified and the presence of the particular mRNA in them can be assessed. Thus this technique identifies the location of the mRNAs recognized by the specific probe used. It was this technique that revealed the precision with which mRNAs are produced, for abrupt boundaries of expression occur between sclerites and intersegmental membranes (Rebers *et al*., 1997) or at muscle insertion zones (Horodyski and Riddiford, 1989) or next to specialized epidermal cells (Horodyski and Riddiford, 1989; Rebers *et al*., 1997). This technique even revealed the presence of mRNA for cuticular proteins in epithelia of imaginal disc from young larvae (Gu and Willis, 2003). A limitation of the technique is that cRNA probes sometimes bind to the cuticle itself, possibly obscuring detection of mRNA in the underlying epidermis (Fechtel *et al*., 1989; Gu and Willis, 2003). Fechtel *et al*. (1989) found this artifact to be cuticle-type as well as strand- and probe-specific. Results from several species are summarized in **Table 3**.

The fifth method, immunolocalization, was described earlier in conjunction with localization of specific proteins within the cuticle, but it can also be used to identify the site of synthesis by looking for a particular protein within the endoplasmic reticulum or Golgi apparatus (Sass *et al*., 1994a, 1994b).

The results from Northern analyses, metabolic tissue labeling, and *in vitro* translations reveal that all cuticular proteins with known sequences or for which specific probes are available are synthesized by the integumental preparations. Different proteins are synthesized at different times in a molt cycle and in different anatomical regions and there are some cuticular proteins whose synthesis is stage-specific. Differences in the presence of mRNA parallel the appearance of labeled proteins indicating that much of the temporal and spatial control of cuticular protein synthesis is at the level of transcription. As mentioned above, however, all three of these methods are limited by the possible contamination of tissues by nonepidermal cells and by their inability to address heterogeneity of cell types within the epidermis.

Studies that have combined tissue labeling or *in vitro* translations with immunolocalization have at last clarified the relationship between hemolymph and cuticular proteins with identical electrophoretic and immunological properties. The most comprehensive studies of protein trafficking, carried out in *Calpodes*, revealed four classes of exported proteins that are handled by the epidermis.

Table 3 Evidence for the association of location or type of cuticle and sequence class of some cuticular proteins

Species	*Protein*	*Sequence class*	*Localization*[a]	*Nature of evidence*[b]	*When deposited*	*Reference*
Bombyx mori	BMLCP18	RR-1	Imaginal discs	EST		Gu and Willis (2003)
Drosophila melanogaster	EDG-78	RR-1	Larval and imaginal cells of prepupa	ISH		Fechtel *et al.* (1989)
Drosophila melanogaster	EDG-84	RR-2	Imaginal disc cells	ISH		Fechtel *et al.* (1989)
Drosophila melanogaster	PCP	RR-1	Prepupal thorax and abdomen	ISH		Henikoff *et al.* (1986)
Hyalophora cecropia	HCCP12	RR-1	Soft cuticle; imaginal discs	CD and ISH		Cox and Willis (1985), Gu and Willis (2003)
Hyalophora cecropia	HCCP66	RR-2	Hard cuticle	CD and ISH		Cox and Willis (1985), Gu and Willis (2003)
Locusta migratoria	LM-ACP7	RR-2	Hard cuticle	CD		Andersen *et al.* (1995a)
Locusta migratoria	LM-ACP8	RR-2	Hard cuticle	CD		Andersen *et al.* (1995a)
Locusta migratoria	LM-ACP19	RR-2	Hard cuticle	CD		Andersen *et al.* (1995a)
Manduca sexta	CP14.6	RR-1	Soft cuticle	ISH		Rebers *et al.* (1997)
Manduca sexta	LCP16/17	RR-1	Soft cuticle	ISH		Horodyski and Riddiford (1989)
Tenebrio molitor	ACP17	Glycine-rich	Hard cuticle	ISH	Strongest post-ecdysis	Mathelin *et al.* (1995, 1998)
Tenebrio molitor	ACP20	RR-2	Hard cuticle	ISH	Primarily pre-ecdysis	Charles *et al.* (1992)
Tenebrio molitor	ACP-22	RR-2	Hard cuticle	ISH, mAB	Pre-ecdysis	Bouhin *et al.* (1992a, 1992b)
Tenebrio molitor	TMLPCP22	51 aa motif	Hard and soft cuticle pre-ecdysis, then only soft cuticle	ISH, mAB	Primarily pre-ecdysis	Rondot *et al.* (1998)
Tenebrio molitor	TMLPCP23	51 aa motif	Hard and soft cuticle	ISH	Only pre-ecdysis	Rondot *et al.* (1998)
Tenebrio molitor	TMLPCP29	RR-3 and 18-residue motif	Hard and soft cuticle, except not posterior borders of sclerites	ISH	Post-ecdysis	Mathelin *et al.* (1998)

[a]For *in situ* hybridization, cuticle type was determined by nature of cuticle overlying the epidermis.
[b]CD, careful dissection prior to extraction of proteins; ISH, *in situ* hybridization used to localize mRNA; mAB, monoclonal antibody immunolocalization; EST, from *Bombyx* EST project (Mita *et al.* 1999).

These findings are so important that the experimental methodology is worth discussing. The first approach used was to seal sheets of final instar integument into a bathing chamber so there could be no leakage from the cut edges of the tissue and then find what proteins were made in a 2 h exposure to ^{35}S-methionine. Three classes of proteins were identified with this procedure. One was secreted exclusively into the cuticle (C class), a second appeared in the bathing fluid, hence has been secreted basally (B class) while the third was secreted in both directions (BD class) (Palli and Locke, 1987). Immunolocalization of numerous other *Calpodes* proteins (of unknown sequence) confirmed the existence of these three routing classes of epidermal proteins. A fourth, T class, for proteins transported into cuticle, but not synthesized by the epidermis, was identified. Its presence eliminated any concerns that the classes might be artifacts from labeling with ^{35}S-methionine (Sass *et al.*, 1993).

One member of the T class (T66) was studied in more detail. It was localized by immunogold throughout the cuticle, and although found in epidermal cells was not found in association with the Golgi apparatus, confirming its transcellular transport, rather than synthesis by the epidermis. A subsequent study identified the exclusive site of its synthesis as spherulocytes (Sass *et al.*, 1994a).

Whether the BD proteins are secreted from both apical and basal borders of epidermal cells is still not clear. Locke (1998, 2003) now favors the possibility that all secretion is apical, where the Golgi are concentrated, and that the secreted proteins are subsequently taken back into the cell from the perimicrovillar space and transported in vesicles to the basal surface where the contents are released into the hemolymph.

In conclusion, it is now clear that the epidermis can synthesize both cuticular and hemolymph proteins. It can also transport proteins made in tissues other than epidermis from hemolymph to cuticle.

4.2.2.3. Tracheal Cuticular Proteins

An often-neglected source of cuticle in insects is the tracheal system. Since tracheae are associated with all insect tissues, caution is needed in interpreting the significance of the presence of mRNAs or cuticular proteins from nonintegumental tissues. Cox and Willis (1985) recognized that some of the proteins from tracheae had the same isoelectric points as proteins isolated from integumentary cuticle. A further study was carried out a decade later by Sass *et al.* (1994b), combining electrophoretic analysis with immunogold labeling. Chitin was localized with wheat germ agglutinin and found in all regions of tracheae and tracheoles except the taenidial cushion. Antibodies that had been raised against individual electrophoretic bands from integumentary extracts represented proteins from all four classes of integumentary peptides. Some C proteins, those from the surface cuticle, were found associated with chitin but only in taenidia, other C proteins were in the general matrix with and without chitin. The B and BD peptides were only found in the taenidial cushion, the region lacking chitin. It appears that hemolymph peptides that are synthesized by the epidermis may be tracheal cuticle precursors. The one T protein studied (T66, made in spherulocytes) was also found in the general matrix. An important insight from this study was the conclusion that: "The extremely thin tracheal epithelium suggests that transepithelial transport might supply proteins to the tracheal cuticle more evenly than Golgi complex secretions" (Sass *et al.*, 1994b).

4.2.3. Classes of Proteins Found in Cuticles

4.2.3.1. Nonstructural Proteins

Nonstructural proteins that have been identified in cuticle are listed in **Table 2**.

4.2.3.1.1. Pigments Proteins from three classes of pigments used in cuticle – insecticyanins and two different yellow proteins – have been sequenced. The insecticyanins are blue pigments made by the epidermis and secreted into both hemolymph and cuticle. They are easily extracted from cuticle with aqueous buffers. Members of the lipocalin family, they are present as tetramers with the gamma isomer of biliverdin IX situated in a hydrophobic pocket (see **Chapter 4.8**). In the cuticle, in cooperation with carotenes, they confer green coloration. Their structure has been determined to 2.6 Å by X-ray diffraction (Holden *et al.*, 1987), making them structurally the best characterized cuticular proteins. Two genes code for insecticyanins in *Manduca* (Li and Riddiford, 1992).

The yellow protein in *D. melanogaster* has been localized with immunocytochemistry in cuticles destined to become melanized (Kornezos and Chia, 1992). Thus it was found in association with larval mouth hooks, denticle belts, and Keilin's organs. Mutants of the gene *yellow* lack black pigment in the affected cuticular region. Mutant analysis revealed two classes of mutants, those that affect all types of cuticle at all stages, and those affecting only particular areas of specific stages. At least 40 different adult cuticular structures could express their color independently (Nash, 1976), and the regulatory regions responsible for some of the stage and regional specificity have been identified (Geyer and Corces, 1987). The yellow protein has been described as a structural component of the cuticle that interacts with products from the gene *ebony*, a β-alanyl-dopamine synthase, to allow melanin to be deposited. Flybase (2003) reports that 740 different alleles of *yellow* have been described, in 542 references beginning in 1916. The complete sequence of *yellow* has been determined for 13 species of *Drosophila* in addition to *D. melanogaster*. An examination of *yellow* expression revealed that both *cis*- and *trans*-regulation are responsible for differences in pigmentation patterns among different species (Wittkopp *et al.*, 2002). There is no evidence for a known chitin-binding domain in the yellow protein; the only domain recognized is pfam03022 (major royal jelly protein). Although the sequence for yellow is 37% identical and 56% similar to a dopachrome conversion enzyme from *Aedes aegypti* that is involved in the melanotic encapsulation immune response, yellow itself evidently is devoid of enzyme activity (Han *et al.*, 2002) (see **Chapter 4.4**).

Another cuticular protein implicated in pigmentation, putatively β-carotene binding, has been isolated from extracts of cuticle from mature adult

Schistocerca gregaria using column chromatography to isolate a protein that was yellow in color. It bears significant sequence similarity to various insect JH-binding proteins (see **Chapter 3.7**), as well as odorant-binding proteins (see **Chapter 3.15**). Wybrandt and Andersen (2001) suggest that it is involved in the transport of carotenes into epidermis and then the cuticle.

4.2.3.1.2. Enzymes Some of the enzymes involved in sclerotization have been identified in cuticle. Since they are discussed by Andersen (see **Chapter 4.4**) they will not be considered here.

Some enzymes that belong to the molting fluid become evident as the electrophoretic banding pattern of cuticular proteins changes as *Calpodes* initiates molting at the end of the fifth instar, with the most conspicuous change being the appearance of a band of 19 kDa. Antibodies raised against this protein were used to isolate a cDNA from a library cloned in an expression vector. The conceptual translation revealed a protein (CECP22). Its sequence suggested it might have amidase activity. Further analysis revealed that the protein was present in the cuticle before each molt, and was also found in molting fluid. Marcu and Locke (1998, 1999) present evidence that this protein may be activated by proteolysis and speculate that it may function to cleave an amidic bond between *N*-acetylglucosamine from chitin and amino acids in cuticular proteins.

Enzymes involved in digesting the old cuticle are temporary residents in cuticle. These include proteases and chitinases. Their interaction is discussed by Marcu and Locke (1998).

4.2.3.1.3. Defense proteins Also found in the cuticle are components of the insect defense system. In one study, cuticle was removed from *Bombyx* larvae 24 h after they had been abraded with emery paper and exposed to bacteria. The antibacterial peptide cecropin was purified from the cuticles (Lee and Brey, 1994). Both prophenoloxidase and a zymogen form of a serine protease capable of activating it have been extracted from *Bombyx* larval cuticle. Colloidal gold secondary antibodies revealed that the prophenoloxidase was localized throughout the epicuticle and procuticle, and in a conspicuous orderly array on the basal side of the helicoidal chitin lamellae. An extraepidermal source is likely for this enzyme since no labeling was found in the epidermis, nor was mRNA detected in the epidermal cells. It is assumed to function in the melanization that occurs in response to injury (Ashida and Brey, 1995).

Molnar *et al.* (2001) presented immunological evidence for a protein related to the defense protein scolexin in the cuticle of *Manduca*. This protein exists in two forms in *Manduca*, but the antibody used did not distinguish between them.

4.2.3.1.4. Arylphorins The final class of nonstructural proteins is the arylphorins, proteins with high content of aromatic amino acids and some lipid. These proteins, assumed to be hemolymph proteins, have been of special interest since the discovery by Scheller *et al.* (1980) that although calliphorin (the arylophorin from *Calliphora*) was found in cuticle, it seemed to come from the hemolymph, because radioactively labeled calliphorin injected into the hemolymph appeared in cuticle. But there is also evidence that the epidermis is capable of synthesizing arylphorins, for Riddiford and Hice (1985) had detected arylphorin mRNA in the epidermis of *Manduca*.

Palli and Locke (1987) used an anti-arylphorin antibody to identify an 82 kDa protein made in *Calpodes* integumental sheets *in vitro* that appeared in both cuticle and media. Thus arylphorin appeared to be a bidirectionally secreted integumentary protein. Next, colloidal gold secondary antibodies were used to visualize the location of anti-arylphorin in ultrathin sections of various tissues (Leung *et al.*, 1989). The resolution afforded by this method made it possible to recognize arylphorin in epicuticle (but not lamellar cuticle) and in the Golgi complexes of the fat body, and to show by quantifying gold particles that it was also found in Golgi complexes of epidermis, midgut, pericardial cells, and hemocytes as well as the meshwork of fibrous cuticle in tracheae. Thus, while the possibility remains that some arylphorin is transported from hemolymph to cuticle, it need not be, for the epidermis itself is capable of synthesizing and secreting this protein. These studies further demonstrated that a given protein can be synthesized by multiple tissues. Whether it is the same gene that functions in all tissues remains to be determined.

The role of arylphorin remains unknown. It is generally assumed to be participating in sclerotization because of its high tyrosine content. Is it degraded in the cuticle so that its constituent amino acids are released or does it remain an integral part of the cuticle? The latter is favored by the available evidence because calliphorin has been shown to bind strongly to chitin *in vitro* (Agrawal and Scheller, 1986) and no breakdown products were detected after injection of labeled calliphorin (Konig *et al.*, 1986).

4.2.3.2. Structural Proteins

4.2.3.2.1. Overview Slightly less than a decade ago, a comprehensive and insightful review of cuticular proteins presented the complete sequence and full citation for all 40 cuticular proteins known at that time and identified features that remain their hallmarks (Andersen *et al.*, 1995a). As of June 2003, in addition to the nonstructural cuticular proteins discussed above, there are now 139 sequences available for what are postulated to be structural proteins. These numbers do not include almost 200 more that have been identified by protein prediction programs used to annotate the *D. melanogaster* and *Anopheles gambiae* genomes. These have been omitted because their annotation is still in a state of flux. All 139 sequences and some of their key features are listed in **Table 1**. Marcu and Locke (1998) have also published tabular summaries of a smaller number of cuticular proteins.

Unfortunately, cuticular protein terminology is not uniform. Most workers have included the initials of the genus and species. Some have named their proteins based on their molecular mass, others on the order in which they obtained them. Many *D. melanogaster* proteins have been designated by the chromosomal band to which a gene-specific probe hybridized. Some have been named after their sequence similarity to a particular cuticular protein from another species. Capitalization and the use of hyphens are erratic. Two different groups have worked on proteins from *Tenebrio*, and given two different names to one protein. Some of the names have included a designator for genus, species, and metamorphic stage (e.g., TMLCP-A1A). Although this designation is an accurate indication of the stage from which the protein was purified, it can inadvertently support the misconception that a particular cuticular protein is stage-specific. In *Tenebrio*, larval and pupal cuticular proteins are indistinguishable electrophoretically (Andersen, 1975; Roberts and Willis, 1980a; Lemoine and Delachambre, 1986; Andersen *et al.*, 1995b), and molecular analyses of several *Tenebrio* cDNAs found that all expressed in pupae are also expressed in larvae (Mathelin *et al.*, 1998; Rondot *et al.*, 1998). Furthermore, proteins that are a major component of the cuticle of one stage can be a minor component of another (Cox and Willis, 1985; Willis, 1986). A final complication is whether two almost identical proteins are allelic variants or distinct proteins. In some cases an "isoform" has been described. Genomic sequences, however, have revealed that stretches coding for proteins with very similar or indeed identical sequence may be linked on a chromosome (Charles *et al.*, 1997; Dotson *et al.*, 1998) (see Section 4.2.4.2). Thus the finding of "isoforms" may reflect distinct genes and hence distinct proteins.

Table 1 includes proteins from discrete genes even when two or more may have the same amino acid sequence. In **Table 1**, proteins isolated from cleaned cuticles were counted as cuticular proteins, as were proteins whose nucleic acid sequences were obtained using partial protein sequences or antibodies raised against cuticular protein to select corresponding cDNAs. In addition, **Table 1** contains numerous proteins that had been classified as cuticular proteins by their "discoverers" because their sequence, or a part thereof, was similar to a cuticular protein already in the databases. For many of those in the latter category, the source of the cDNA that led to the sequence came from integumental epidermis or imaginal discs or a cDNA hybridized to epidermal RNA in a Northern analysis. For some, especially those from *Tenebrio* studied by Delachambre's group, confirmation came from *in situ* hybridization of specific probes (see Section 4.2.2.1.3 and **Table 3**). But for many, sequence similarity served as the sole criterion.

Most of the structural cuticular proteins whose sequences were known in 1995 came from the efforts of Svend Andersen and his group, and a significant fraction (42%) still does. All of their sequences come from direct sequencing of purified cuticular proteins. Most of the other protein sequences come from sequencing cDNA or genomic DNA. For these, the length of the mature proteins can only be deduced by subtracting the amino acids of the signal peptide. In a few cases, N-terminal sequence data is available to assure that the signal peptide has been correctly identified. In cases where this information was not available, or when the original submissions did not provide this information, it was determined using the program SignalP (Nielsen *et al.*, 1997; Nielsen and Krogh, 1998). All lengths in **Table 1** represent the mature, processed protein.

One notable feature of the structural cuticular proteins is that almost all lack cysteine and methionine residues in the mature protein; the five exceptions to this are indicated in **Table 1** and **Figure 4**. Andersen (see **Chapter 4.4**) suggests that the reactivity of cystine and cysteine with *ortho*-quinones could interfere with sclerotization.

Most of the cuticular proteins identified to date are quite short. Those less than 100 amino acids account for 27% of the 139 sequences in **Table 1**, while those between 100 and 199 account for an additional 52%. Only three proteins have more than 300 amino acids. The largest is the gene for

a *D. melanogaster* corneal lens protein (Cry, drosocrystallin) with 457 amino acids, and a perfect RR-2 consensus (Janssens and Gehring, 1999). (See Section 4.2.3.2.3 for discussion of this consensus.) A cDNA of the appropriate size has been described (gi:2143072). The next largest is for the only cuticular protein characterized from *Galleria* (Kollberg *et al.*, 1995). This protein is unusual in that its only resemblance to known cuticular proteins is an abundance of alanine residues, and it is unique in having two cysteine residues. Yet its cDNA was selected with a polyclonal antibody raised against pupal cuticular proteins. The third largest protein is MSCP36; this is a high glycine protein that has a RR-1 consensus.

4.2.3.2.2. Motifs found in cuticular proteins The review by Andersen *et al.* (1995a) was the first to assemble a variety of motifs found in cuticular proteins. The occurrence of such motifs is given in **Table 1** and summarized in **Table 4**. Most common of these is a 28-residue region, first recognized by Rebers and Riddiford (1988) in seven cuticular proteins that is commonly referred to as the R&R Consensus. The original R&R Consensus is part of a longer conserved sequence – pfam00379 – and it is now apparent that there are three distinct forms of the extended R&R Consensus. These matters are discussed in detail below (see Section 4.2.3.2.3).

After the R&R Consensus, the next most common motifs were repeats of A-A-P-(A/V). These repeats were found in cuticular proteins both with and lacking the R&R Consensus; in sequences with the R&R Consensus they may occur N- or C-terminal to the extended Consensus. They are found in 46% of the sequences in **Table 1**. Thus, while abundant in cuticular proteins, they certainly are not diagnostic for this class of protein.

Andersen *et al.* (1995a) recognized several sequences with stretches of glycine, leucine, and tyrosine, beginning G-Y-G-L- or G-L-L-G. In **Table 1**, they are combined under the designation, G(Y/L) motifs. Other cuticular proteins are also high in glycine, but with less regular motifs; these are designated by the number of consecutive Gs. Proteins enriched in glycine residues are found in a variety of structures such as plant cell walls, cockroach ootheca, and silk (see Bouhin *et al.* (1992a) for discussion). Subsequent to their 1995 review, Andersen and his colleagues recognized two additional motifs. There is an 18-residue motif found in seven cuticular proteins from four orders of insects (and two crustaceans), and its consensus has been described (Andersen, 2000). It occurs in proteins with and without the R&R Consensus. Also reported was a 51-residue motif so far identified only in cuticular proteins from *Locusta* and *Tenebrio*. It has not been found in proteins with the R&R Consensus (Andersen *et al.*, 1997). Other short repeats have been found in a limited number of proteins, from a single species. Proteins with the various motifs are identified in **Table 1**.

The basic sequences of the three long repeats are:

Original R&R Consensus: [G-x(8)-G-x(6)-Y-x(2)-A-x-E-x-G-F-x(7)-P-x-P.]
18 amino acid repeat: [(PV)-x-D-T-P-E-V-A-A-A-(KR)-A-A-(HF)-x-A-A-(HY).]
51 amino acid repeat: [x(6)-A-x(9)-R-S-x-G-x(4)-V-S-x-Y-x-K-x(2)-D-x(3)-S-S-V-x-K-x-D-x-R-x(2)-N-x(3).]

With this nomenclature, x is any amino acid, the number in parentheses represents the number of amino acids, and multiple letters in parentheses indicate that either of two amino acids may be present. This is the format used by MOTIF (2003), a resource that lets you search a given motif against various databases.

Andersen (2000) presented a model where proteins with the R&R Consensus bind to chitin and the other structural proteins remain free in the interfilament space.

4.2.3.2.3. Proteins with the R&R consensus The R&R Consensus is a common feature of cuticular proteins from all six orders of insects examined to date and it has also been recognized in cuticular proteins from arachnids and crustaceans (review: Willis, 1999).

Three distinct forms of the consensus have been recognized and named by Andersen (1998, 2000) RR-1, RR-2, and RR-3 (**Figures 1–3**). RR-1 is present in 51 (37%) of the proteins in **Table 1**.

Table 4 Summary characteristics of the 139 ''structural'' cuticular proteins that have been completely sequenced

	Number with AAP(AV)	*Number lacking AAP(AV)*	*Number 18-residue motif*	*Number 51-residue motif*	*Mean number H+K in extended R&R region (range)*	*Number with Met/Cys residues*	*Total proteins in class*
RR-1	2	49	0	0	3.7 (0–9)	3/0	51
RR-2	36	8	0	0	7.4 (2–19)	0/0	44
RR-3	1	2	3	0	6.3 (5–8)	0/0	3
Not RR	26	15	4	7		0/2	41

RR-1-bearing proteins have been isolated from flexible cuticles, while RR-2 proteins have been associated with hard cuticle. This generalization, based on relatively few cases (**Table 3**), has been used to link proteins to cuticle types in the absence of any other evidence.

The RR-1 proteins have the essential features of the original consensus (**Figure 1**). The vast majority have the short sequence -Y-x-A-x-E-x-G-(FY)-x(7)-P. N-terminal to this region the sequences diverge, but most have a series of three aromatic residues, such as Y-x-F-x-Y, that begins about 32 amino acids N-terminal to the start of the R&R Consensus (**Figure 1**). Another distinguishing feature defined by Andersen (1998) is a glutamic acid residue found in a conserved position [Y-x-A-x-**E**-G-(FY)] in 90% of the sequences in **Figure 1**.

The RR-2 proteins have a considerably extended consensus, first recognized by Bouhin *et al.* (1992a) and Charles *et al.* (1992). What is extraordinary about the RR-2 consensus is its conservation across six orders of insects. Only two single amino acid gaps are required to accommodate all 44 RR-2 sequences with this variant listed in **Table 1**. Twenty-two of the 70 residues in the extended consensus (31%) are virtually invariant and an additional 21 are represented by a single amino acid in over half of the proteins (**Figure 2**). The first few RR-2 sequences identified suggested that the residues G-F-N-A-V-V would be diagnostic (Andersen, 1998). The identification of more RR-2 sequences revealed that that region of the consensus is not perfectly conserved. Rather, all of the sequences have G-F-x-A-x-V, a configuration found in none of the RR-1 sequences.

There are other differences between RR-1- and RR-2-bearing proteins. Most of the RR-2 (82 %) have at least one A-A-P-(AV) motif, while only 2 (4%) of the RR-1 type have this motif (**Table 1**). Histidine and lysine residues can be more abundant in the extended consensus region of RR-2 proteins (**Table 4**, **Figures 1** and **2**). Only one RR-1 protein has been found with more than six histidine plus lysines in this region, while seven or more were present in 20 of the 44 RR-2 sequences. There is an invariant lysine in all RR-2 sequences and several other positions appear to be favorable for either of these amino acids. Over half of the RR-2 proteins, but only a quarter of the RR-1 proteins, have histidine as their final or penultimate C-terminal amino acid. Histidines and lysines are known to be reactive sites for sclerotizing agents (see **Chapter 4.4**), so it is possibly significant that proteins from "hard" sclerotized cuticles would have these amino acids in abundance. The number of potential sclerotization sites may also be related to whether a cuticle can grow by intussusception, something that would be impossible if the proteins were extensively cross-linked. The six aphid cuticular proteins, while all of the RR-2 type, have relatively few histidines plus lysines (only four or five); this paucity may reflect the need for cuticular expansion with the type of feeding and brooding of progeny that occurs in these animals.

An RR-3 form of the consensus has been based on three sequences from insects and two from other arthropods (Andersen, 2000). A tentative consensus (**Figure 3**) was constructed from the sequence alignment in Andersen (2000).

Whole genome sequencing has led to the need to classify annotated sequences. A valuable website, Pfam (2003a) has used hidden Markov modeling to define motifs characteristic of particular classes of proteins (Bateman *et al.*, 2002). When a protein sequence is searched against all known (and predicted) proteins using the BLAST server (Blast, 2003), the first information that is presented is an indication of matches to Pfam entries. The Pfam sequence that allows annotators to classify a protein as a cuticular protein is Pfam00379, a 68 amino acid sequence that includes the extended R&R Consensus. It also goes under the name "chitin_bind_4," for reasons that will become apparent later (see Section 4.2.5.4). The pfam00379 was obviously based on proteins of both RR-1 and RR-2 classes, for it matches neither particularly well (**Figure 3**). This makes it particularly useful for a preliminary classification of a putative cuticular protein sequence.

The pfam00379 is found in 70% of the 139 cuticular proteins in **Table 1**, i.e., all the RR-1, RR-2, and RR-3 sequences. A complete listing of all sequences with pfam00379 can be found at Pfam (2003b) or at ENTREZ (2003), where you search Domains for pfam00379. There are no nonarthropod sequences with this consensus. Now that two species of insects (*D. melanogaster* and *A. gambiae*) have had their genomes completely sequenced and pfam00379 is being used to recognize cuticular proteins, the representation of proteins bearing this motif will be disproportionate. As of August 2003, 90 *D. melanogaster* sequences beyond those listed in **Table 1** have been found to have pfam00379, and using the ENTREZ site, *A. gambiae* had over 100 beyond those in **Table 1**.

Pfam00379 so far has been found to occur only once in a given protein, with the notable exception of a protein from the tailfin of the prawn *Penaeus japonicus*. The entire sequence of this protein is made up of 14 consecutive pfam00379 motifs (Ikeya *et al.*, 2001). In the insect proteins, this motif has been found near the N- or C-terminus, or within the protein.

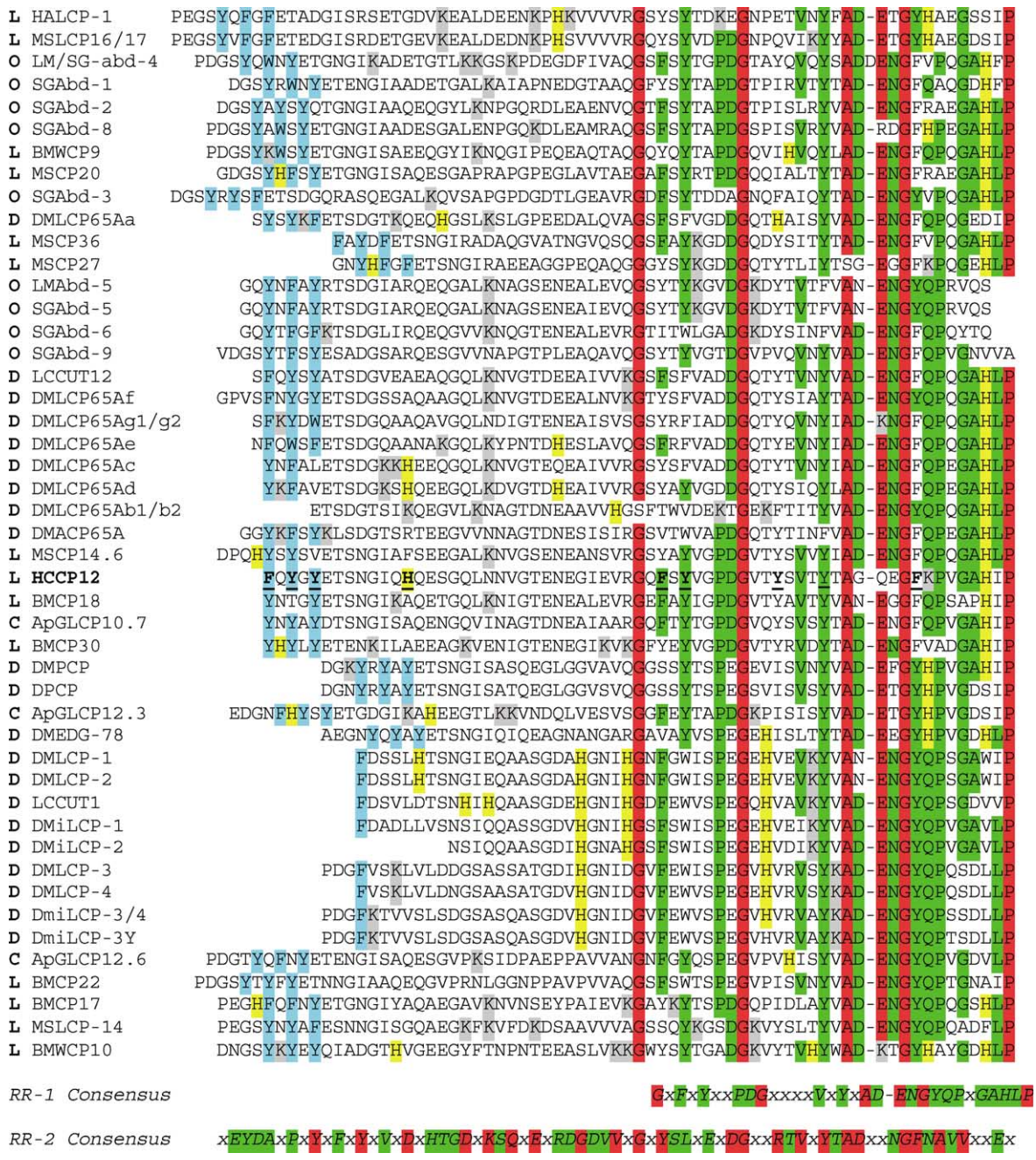


Figure 1 Alignment of the pfam00379 regions of 51 proteins with the RR-1 consensus. The pfam 00379 regions for RR-1 proteins were aligned with ClustalW (http://clustalw.genome.ad.jp/); only one internal gap was used to allow direct comparison with the RR-2 consensus. Orders of insects are: (**C**) Coleoptera, (**D**) Diptera, (**L**) Lepidoptera, (**O**) Orthoptera. Abbreviations for proteins as in **Table 1**. Four pairs of identical sequences are each presented on a single line. Red represents amino acids present in at least 95% of the proteins, green in the majority. Histidines are shown in yellow, lysines in light gray. A common triad of aromatic residues is shown in light blue. Bolded and underlined are several residues from HCCP12 that are shown in **Figure 5a–c**. An RR-1 consensus based on these sequences is given. The bottom line gives the consensus for RR-2 proteins from **Figure 2**, except that two single amino acid gaps needed to accommodate three atypical sequences were eliminated.

The wealth of information on cuticular protein sequences and the unraveling of how the structure of some contributes to the interaction of chitin and protein (see Section 4.2.5) is only a beginning. Essential properties of cuticle remain to be explained, and important questions raised in the older literature about various means of achieving cuticle plasticity and the importance of hydration in cuticle stabilization must not be forgotten (Vincent, 2002 and references therein).

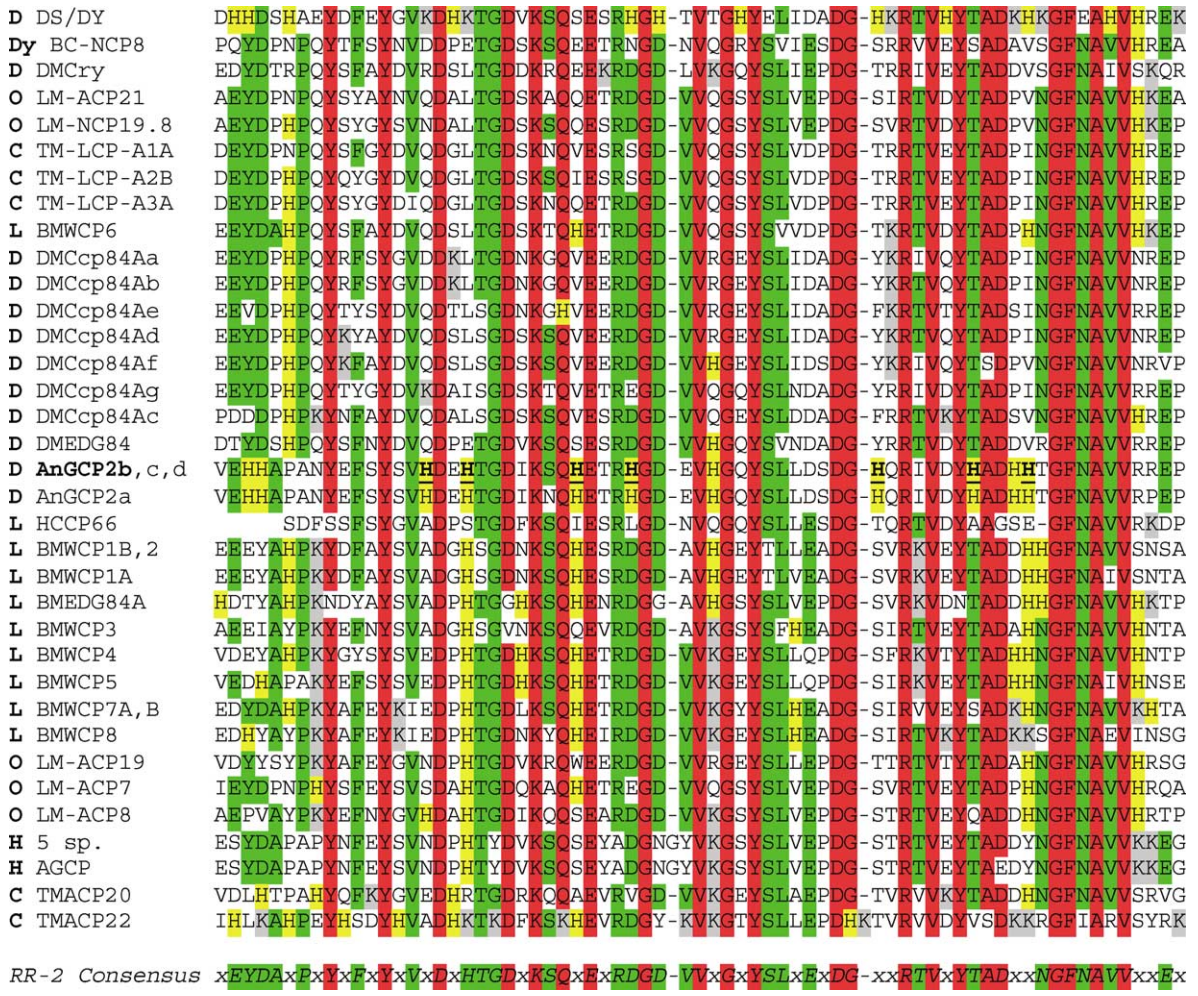


Figure 2 Alignment of the pfam00379 region of 44 cuticular proteins with the RR-2 consensus. Abbreviations as in **Figure 1** plus Dictyoptera (**Dy**) and Hemiptera (**H**). All hemipteran proteins except the one from *Aphis gossypii* are indicated by [***H*** 5 sp]. The order of the proteins was based on alignment by ClustalW. Red represents amino acids present in at least 95% of the proteins, green in the majority. Histidines are shown in yellow, lysines in light gray. The seven histidines in AGCP2b that are discussed in the text (see Section 4.2.5.4) and modeled in **Figure 5d** are bolded and underlined, beginning with residue 99. An RR-2 consensus is given.

Figure 3 Consensus regions from the three types of RR cuticular proteins plus pfam00379. For pfam00379, the three aromatic residues found in RR-1 and RR-2 sequences (**Figures 1** and **2**) are shown in light blue. Indicated in red is the original R&R Consensus; the dash (-) was inserted to facilitate alignment with the other sequences. The green F was a Y in the original. RR-1 and RR-2 consensuses are from **Figures 1** and **2**, respectively. Red residues were found in at least 95% of the sequences, green in at least 50%. RR-3 was taken from Andersen (2000) using the sequences from the three insects and two arthropods arachnid and crustacean, that was identified as RR-3. Red residues were found in all five RR-3 sequences, green in three or four. The first and third dashes were inserted to allow alignment among all RR-3 sequences; the second was necessary to allow alignment with the RR-2 sequence.

4.2.3.2.4. Resilin The name resilin has been given to the rubberlike proteins responsible for the elasticity of jumping fleas and vibrating wings. Resilins are characterized by a high percentage of glycine (35–40%) and proline (7–10%). They are cross-linked with di- and tri-tyrosine residues (Andersen and Weis-Fogh, 1964). An intact protein corresponding to resilin has never been isolated from cuticle, presumably reflecting its insolublility after cross-linking. Ardell and Andersen (2001) used short peptide sequences that they had obtained from resilin-bearing regions of *Schistocerca* cuticle to probe the annotated *D. melanogaster* genome. Two candidate proteins had good matches to the locust peptides and to some that Lombardi and Kaplan (1993) had obtained from resilin in *Periplaneta americana*.

Ardell and Andersen concluded that predicted protein CG15920 (gi:24654243) of 620 amino acids was most likely to be a true proresilin (the non-cross-linked version), for in addition to the peptide matches, it had 35% glycine and 11% proline. Its 18 N-terminal copies of a 15-residue repeat and 13 C-terminal copies of a 13-residue repeat were predicted to contribute to a β-spiral, a common form for proteins with elastic properties (Ardell and Andersen, 2001). A cDNA (gi:27820115) is available that corresponds to most of the genomic sequence, but it lacks 45 internal amino acids. These correspond to the predicted second exon in the genomic sequence and contain almost the entire match to the locust proresilin peptides as well as most of the match to pfam00379. All of the repeat regions, however, are present. Thus, in order for CG15920 to match the locust resilin, it would have to be coded by an alternatively spliced form.

The second *D. melanogaster* gene that had "resilin peptides" was CG9036. Ardell and Andersen considered it to be a less likely candidate because it lacked both a predicted signal peptide and features expected for elastic properties. The original version they described has been replaced (gi:19922620) and a cDNA identical to this new sequence has been obtained. The new version has a predicted signal peptide of 19 amino acids. The mature protein of 198 amino acids has a pfam00379 region (occupying one-third of the mature protein) and is 20% glycine and 10% proline; it has neither of the repeats found in the other candidate protein.

Proof that either of these sequences is proresilin will require localization of the mRNA or protein to the tendons shown to have resilin in Diptera (Andersen and Weis-Fogh, 1964). The special properties of resilin justify further work to establish its sequence.

4.2.3.2.5. Glycosylation of cuticular proteins Glycosylation of cuticular proteins was first reported by Trim (1941) and in limited subsequent reports (review: Cox and Willis, 1987b). In recent years, posttranslational modifications of cuticular proteins have been determined by staining gels with periodic acid Schiff (PAS), by using labeled lectins to probe blots of electrophoretically separated proteins or by discovering discrepancies in masses of peptide fragments experimentally determined by matrix-assisted laser desorption ionization – mass spectrometry (MALDI-MS) analysis and calculated from Edman sequencing.

Most of the major cuticular proteins seen on gels stained with Coomassie Blue are not recognized by PAS or lectins, while some minor ones are glycosylated. This was true for *H. cecropia* where PAS staining revealed glycosylated proteins in extracts of flexible cuticles and a screen with eight lectins revealed the presence of mannose and *N*-acetylgalactosamine, with more limited binding to *N*-acetylglucosamine, galactose, and fucose, in a few of the proteins from all stages (Cox and Willis, 1987b). A comparable study in *Tenebrio* revealed one major band of water-soluble larval and pupal cuticular proteins that had *N*-acetylglucosamine; a few other bands were weakly visualized with lectins; none of the proteins from adult cuticle reacted with the lectins (Lemoine *et al.*, 1990). In another coleopteran, *Anthonomus grandis*, glycosylation was found in cuticular proteins extracted from all three metamorphic stages (Stiles, 1991). In *Calpodes*, all the BD peptides (see Section 4.2.2.2) extracted from the cuticle were associated with α-D-glucose and α-D-mannose, just like most of the hemolymph proteins but very few of the C class proteins. Some of each class appeared to be modified with *N*-acetylglucosamine. T66, a protein synthesized in spherulocytes, transported to epidermis, and then secreted into the cuticle, however, was not glycosylated. In none of these species is the amino acid sequence of a glycosylated protein known.

Sequence-related information about glycosylation is available for cuticular proteins isolated from locusts and *Manduca* (see **Table 1**) where the direct analysis of residues had been used. In *Locusta migratoria*, one to three threonine residues were modified in the protein LM-ACP-abd4. In each case, the modification was with a moiety with a mass of 203, identified as *N*-acetylglucosamine (Talbo *et al.*, 1991). Each of the three threonine residues occurred in association with proline (FPTPPP, LATLPPTPE). All eight of the cuticular proteins that have been sequenced from *Schistocerca gregaria* nymphs had evidence for glycosylation with a moiety with a mass of 203, all at a threonine residue found in a cluster of prolines (Andersen, 1998). Three proteins recently isolated from *Manduca* were similarly shown to be

glycosylated on threonines also in proline-rich regions. Surprisingly in these cases, masses of the adducts were varied (184, 188, and 189) and their nature was not determined (Suderman *et al.*, 2003). In all of these cases, the available evidence indicates that the threonine residues had been *O*-glycosylated. The significance of such glycosylation awaits further elucidation.

4.2.4. Genomic Information

4.2.4.1. Introduction

The first four cuticular proteins whose complete sequences were determined were also the first to have their genes described (Snyder *et al.*, 1982). The wealth of experimental detail and thoughtful discussion in that paper make it a classic in the cuticular protein literature. These four genes were for *D. melanogaster* cuticular proteins LCP-1, -2, -3, and -4, and were found to occupy 7.9 kb of DNA, along with what appeared to be a pseudogene. Each gene had a single intron and that intron interrupted the protein-coding region between the third and fourth amino acid. *LCP-1* and *-2* were in the opposite orientation of *LCP-3* and *-4*. The nucleic acid sequences in the protein coding regions for LCP-1 and -2 were 91% identical, for LCP-3 and -4, 85%, with similarity between the two groups about 60%. For the noncoding regions of the mRNAs, the 5′ upstream regions had more sequence similarity than the 3′ downstream. A consensus poly(A) addition site, AATAAA, was found for two of the genes, 110 bp from the stop codon, while similar but not identical sequences (AATACA, AGTAAA), were found for the other two. The four genes were all expressed in the third instar and several short, shared elements were found in their 5′ regions upstream from the transcription start site. Snyder *et al.* (1982) also speculated on the origin of the cluster through gene duplication and inversion. These features of those four genes (coding for RR-1 proteins) have turned out to be the common elements of most of the cuticular protein genes that are known – hence linkage, shared and divergent orientation, an unusually placed intron that interrupts the signal peptide, presence of a pseudogene in the cluster, atypical poly(A) addition sites, and divergence of 3′-untranslated regions have been found for cuticular protein genes in Diptera, Lepidoptera, and Coleoptera.

4.2.4.2. Chromosomal Linkage of Cuticular Proteins Genes

In addition to the four *D. melanogaster* genes discussed in the previous section, several more instances of linked cuticular proteins genes were described prior to sequencing entire genomes. In some cases, the evidence for these genes was restricted to cross-hybridization of the genomic fragment, and complete sequences are not known for all the members. A summary of such linked genes is presented in **Table 5**. Many more instances will become known as annotation of the *A. gambiae* and *D. melanogaster* genomes is completed.

A detailed analysis of the cluster of genes at 65A allowed Charles *et al.* (1997, 1998) to describe important features that most likely contributed to the multiplication and diversification of cuticular protein genes. Twelve genes (**Table 5**) were identified in a stretch of 22 kb with the direction of transcription, or more accurately the strand used, was

$$><<<<<<<\quad>>>>>>>.$$

The third gene in the cluster appeared to be a pseudogene. Several important features were found. First, the number of *Lcp-b* genes within the cluster was variable among different strains of *D. melanogaster*, two in the original line (with identical coding sequences), one in another, and three in a third. On the other hand, three copies of the *Lcp-g*

Table 5 Linked cuticular protein genes

Species	*Length of DNA examined*	*Number of genes found*	*Protein names*	*Reference*
Anopheles gambiae	17.4 kb	3	AGCP2a, 2b, 2c	Dotson *et al.* (1998)
Drosophila melanogaster	9 kb	6	DM-LCP1, 1ψ, 2, 3, 4 and one other	Snyder *et al.* (1982)
Drosophila melanogaster	20.5 kb	8	EDG-84, 84Aa, Ab, Ac, Ad, Ae, Af, Ag	Apple and Fristrom (1991), Kaufman *et al.* (1990), genome annotation
Drosophila melanogaster	22 kb	12	ACP65A, LCP65Aa, aψ, b1, b2, c, d, e, f, g1, g2, g3	Charles *et al.* (1997, 1998)
Manduca sexta	20 kb	3	LCP16/17 + 2 not named	Horodyski and Riddiford (1989)
Tenebrio molitor	3.9 kb	2	TMLPCP 22, 23	Rondot *et al.* (1998)

genes (also with 100% sequence identity) were found in all three strains. Comparison of cDNA sequences and genes revealed that *Lcp-b1* and *-b2* lacked introns. Both *Lcp-b1* and *-b2* had tracts of As at the 3′ end of the genes, as well as short flanking direct repeats. These features are consistent with the *Lcp-b* genes arising in this cluster by retrotransposition. The sequence data also indicated that *Acp* and *Lcp-a* lacked introns. The rest of the genes had introns but not the common one interrupting the signal peptide (see Section 4.2.4.3).

Evidence for gene conversion between the *Lcp-c* gene and those on the right side of the cluster was also found after a careful analysis of the sequences (Charles *et al.*, 1997).

The consequences of gene duplication in terms of gene expression are an important issue. It could be that duplicated genes were preserved to boost the amount of product made in the short period that the single-layer epidermis is secreting cuticle. Alternatively, duplication may allow for precise regulation of expression of genes both spatially and temporally. Subtle difference in protein sequence may be advantageous for particular structures. A detailed analysis of mRNA levels with Northern blot analysis demonstrated that some members of the *65A* cuticular protein cluster have quite different patterns of expression. *Acp* was expressed only in adults. Expression was not detected for *Lcp-a;* all other *Lcp* genes were expressed in all larval stages, and all but *Lcp-b* and *-f* also contributed to pupal cuticle (Charles *et al.*, 1998).

4.2.4.3. Intron Structure of Cuticular Protein Genes

Genomic sequence data is available for 45 cuticular proteins in **Table 1**. None of these has more than two introns; most have only one, and these introns are in a very conserved positions. Thirty-five of these proteins have an intron that interrupts the signal peptide. In 25 of these sequences, interruption occurs after four amino acids (12 bp); in the remainder, two to eight amino acids are coded for before the intron begins. The PSORT tool (Psort, 2003) calculates the location of discrete regions of a potential signal peptide using modifications of McGeoch's method (McGeoch, 1985) and reports this information as "PSG" data. These short stretches, confined to the first exon, were shown to be identical to the N-terminal positively charged region in 20 cases; all but four of the rest were but one amino acid longer. Whether this correlation of the coding region of the first exon with the N-region of the signal peptide is because it is so short, or because it reflects something more fundamental awaits further exploration. Most of the putative cuticular protein genes in the annotated *A. gambiae* database are missing their initiator methionine, probably because it resides in a short exon, coding for these few amino acids of the signal plus 5′ untranslated nucleotides. The programs, unfortunately, are not yet trained to recognize a configuration with such a short open reading frame.

Genes for four of the proteins listed in **Table 1** (HCCP12 and MSLCP 16/17, MSCP14.6, and TMACP22) have two introns, one interrupting the signal peptide, the other occurring shortly after the beginning of the pfam00379 region. The intron-bearing *D. melanogaster* genes in the cluster at 65A (see above) have their sole intron at the internal position. This led Charles *et al.* (1997) to postulate that the primitive condition for introns in insect cuticular proteins would be two; over time, some genes lost one, some the other, and some lost both or arrived in the genome by retrotransposition.

There is also a *Drosophila* cuticular protein whose gene is located within the region corresponding to the first intron of *Gart*, a gene that encodes three proteins involved in the purine pathway. The gene for this RR-1 protein (*Gart Intron*) is read off the opposite strand and has its own intron, conventionally placed interrupting the signal peptide (Henikoff *et al.*, 1986). A comparably placed gene with 70% amino acid sequence identity is found in *D. pseudoobscura* (Henikoff and Eghtedarzadeh, 1987).

4.2.4.4. Regulatory Elements

One of the attractions of studying cuticular proteins is that they are secreted at precise times in the molt cycle and are thus candidates for genes under hormonal control (Riddiford, 1994). It would be expected, therefore, that some might have hormone response elements (see **Chapter 3.5**). Imperfect matches to ecdysteroid response elements (EcREs) from *D. melanogaster* have been found on two of its cuticular protein genes: EDG78 and EDG84 (Apple and Fristrom, 1991). These genes are activated in imaginal discs exposed to a pulse of ecdysteroids, but if exposed to continuous hormone, no message appears. The two cuticular protein genes that have been studied in *H. cecropia* have regions close to their transcription start sites that resemble EcREs (Binger and Willis, 1994; Lampe and Willis, 1994) and upstream from MSCP14.6 are also two regions that match (Rebers *et al.*, 1997).

Both *Bombyx* PCP and *H. cecropia* HCCP66 have response elements for members of the POU

family of receptors (Nakato *et al*., 1992; Lampe and Willis, 1994). POU proteins are transcription factors used for tissue-specific regulation in mammals (Scholer, 1991). Gel mobility shift experiments established that there was a protein in epidermal cells that could bind to this element (Lampe and Willis, 1994).

As more genomic sequence information becomes available, identification of regulatory elements and verification of their action is certain to be productive.

4.2.5. Interaction of Cuticular Proteins with Chitin

Ever since the R&R Consensus was recognized in 1988, it has been predicted that it must be playing an important role in cuticle. As more and more sequences were discovered with the Consensus, and as it was learned that it also is present in cuticles formed by arachnids and crustaceans, this prediction became more likely. Several workers suggested that the role of the R&R Consensus might be to bind to chitin (Bouhin *et al*., 1992a; Charles *et al*., 1992; Andersen *et al*., 1995a).

Four complementary routes have been followed to learn more about the function of this consensus region. The first was to analyze it with appropriate programs to generate predictions of secondary structure. The second approach was to use spectroscopic techniques on cuticular components to gain information about the conformation of their protein constituents *in situ*. Third, the tertiary structure of the extended Consensus has been modeled, and the fourth route was a direct experimental approach to test whether the extended Consensus could bind to chitin.

4.2.5.1. Secondary Structure Predictions

Prediction of secondary structure was carried out on the extended R&R Consensus region (67–68 amino acids, the pfam00379 region) of cuticular proteins representing different metamorphic stages and four different orders (Iconomidou *et al*., 1999). For each protein, individual predictions of α-helix, β-sheet, and β-turn/coil/loops were carried out using several different predictions programs. These predictions on individual proteins were combined to produce joint prediction histograms for the two classes of proteins. (See Iconomidou *et al*. (1999) for details of proteins analyzed, programs used, and pictorial representation of results.)

The results indicated that the extended R&R domain of cuticular proteins has a considerable proportion of β-pleated sheet structure and a total absence of α-helix. There appeared to be four β-strands in the RR-2 proteins and only three in the RR-1. Three other features were immediately apparent:

1. The three invariant glycines of the original R&R Consensus correspond exactly at the maxima of β-turn/loop predictions, and it is well known that glycines are good turn/loop formers (Chou and Fasman, 1974a, 1974b).
2. With both classes of cuticular proteins, the sheets showed an amphipathic character, i.e., one face is polar, the other nonpolar. Alternating residues along a strand point in the opposite direction on the two faces of a β-sheet. With these proteins, it is the aromatic or hydrophobic amino acids that alternate with other, sometimes hydrophilic, residues. The aromatic rings are thus positioned to stack against faces of the saccharide rings of chitin. This type of interaction is fairly common in protein–saccharide complexes (Vyas, 1991; Hamodrakas *et al*., 1997; Tews *et al*., 1997).
3. The turn/loop regions frequently contained histidines. This would place them "exposed" at the "edges" of a β-pleated sheet. Histidines are involved in cuticular sclerotization (see **Chapter 4.4**) and are involved in the variations of the water-binding capacity of cuticle and the interactions of its constituent proteins. This occurs because small changes of pH can affect the ionization of their imidazole group (Andersen *et al*., 1995a).

The suggestion that cuticular proteins adopt a β-sheet configuration is not new. Fraenkel and Rudall (1947) provided evidence from X-ray diffraction that the protein associated with chitin in insect cuticle has a β-type of structure.

4.2.5.2. Experimental Studies of Cuticular Protein Secondary Structure

The next step in probing the structure of cuticular proteins involved direct measurements on intact cuticles, on proteins extracted from them with a strong denaturing buffer with 8 M guanidine hydrochloride, and on the extracted cuticle. The cuticles came from the flexible abdominal cuticle of larvae of *H. cecropia*, and extracts have HCCP12, a RR-1 protein, as a major constituent (Cox and Willis, 1985). The same prediction programs described above were used on the sequence for HCCP12, and it indicated that the entire protein had a considerable proportion of β-pleated sheet and total absence of α-helix. Fourier-transform Raman spectroscopy

(FT-Raman), attenuated total reflectance infrared spectroscopy (ATR-FT-IR), and circular dichroism spectroscopy (CD) were carried out on these preparations (Iconomidou *et al.*, 2001). These techniques eliminated problems that had been found previously with more conventional laser-Raman spectra due to the high fluorescent background associated with cuticle.

The FT-Raman spectra of both the intact and extracted cuticle were dominated by the contribution of bands due to chitin. Certain features of the Raman spectrum of the intact cuticle signified the presence of proteins. The protein contribution to the spectrum of intact cuticle was revealed by subtracting the spectrum of the extracted cuticle, after scaling the discrete chitin bands of both preparations. The comparison of this difference spectra to that from the isolated proteins revealed striking similarities suggesting that the former gave a reliable physical picture of the cuticle protein vibrations in the native state. While Iconomidou *et al.* (2001) presented a detailed analysis of the spectra and the basis for each assignment, only a few features will be reviewed here. Several of the bands could be attributed to side-chain vibrations of amino acids with aromatic rings, tyrosine, phenylalanine, and tryptophan. Bands in the amide I region (1600–1700 cm^{-1}) of the Raman spectra of the extracted cuticle proteins and of the difference spectrum exhibited a well-defined maximum at 1669 cm^{-1}, typical of β-sheet structure. The absence of bands at ~1650 cm^{-1} indicates that α-helical structures are not favored. The amide III range (1230–1320 cm^{-1}) is relatively free from side group vibrations and, thus, highly diagnostic of secondary structure. The extracted proteins had a doublet at 1241 and 1268 cm^{-1}; the former can be assigned to β-sheet and the latter to β-turns or coil.

Results from ATR-FT-IR spectra from the extracted proteins were in good agreement with their FT-Raman spectra. These spectra had been obtained on lyophilized samples: the CD spectrum, on the other hand, was obtained with proteins solubilized in water. Detailed analysis of the CD spectrum indicated a high percentage (54%) of β-sheet conformation with a small contribution of α-helix (~13%). The contributions of β-turns/loops and random coil were estimated as 24% and 9% respectively (Iconomidou *et al.*, 2001). These results demonstrated that the main structural element of cuticle protein is the antiparallel β-pleated sheet. Comparable results were obtained from lyophilized proteins and intact cuticles and from proteins in solutions, thus negating the concern that lyophilization might increase the β-sheet content of proteins as discussed by Griebenow *et al.* (1999). These direct measurements confirm the results from secondary structure prediction discussed above (see Section 4.2.5.1).

These findings are in accord with the prediction of Atkins (1985) that the antiparallel β-pleated sheet part of cuticular proteins would bind to α-chitin. His proposal was based mainly on a two-dimensional lattice matching between the surface of α-chitin and the antiparallel β-pleated sheet structure of cuticular proteins.

There seem to have been several independent solutions in nature whereby chitin binds to protein; in all surface aromatic residues appear to be significant (Shen and Jacobs-Lorena, 1999). In several cases β-sheets have been implicated. The chitin-binding motifs of two lectins studied at atomic resolution contain a two-stranded β-sheet (Suetake *et al.*, 2000). In bacterial chitinases, an antiparallel β-sheet barrel has also been postulated to play an important role in "holding" the chitin chain in place to facilitate catalysis. Four conserved tryptophans on the surface of the β-sheet are assumed to interact firmly with chitin, "guiding" the long chitin chains towards the catalytic "groove" (Perrakis *et al.*, 1997; Uchiyama *et al.*, 2001).

4.2.5.3. Modeling of Chitin-Binding Domains of Cuticular Proteins

Secondary structure prediction and experimental data summarized above (see Sections 4.2.5.1 and 4.2.5.2) indicated that β-pleated sheet is most probably the underlying molecular conformation of a large part of the extended R&R Consensus, especially the part which contains the R&R Consensus itself, and that this conformation is most probably involved in β-sheet/chitin–chain interactions of the cuticular proteins with the chitin filaments (Iconomidou *et al.*, 1999, 2001). Can this information be translated into a three-dimensional model?

Unexpectedly, a distant (20%) sequence similarity was found between RR-1-bearing cuticular proteins and the crystallographically determined C-terminal, β-sheet barrel portion, of bovine plasma retinol-binding protein (RBP). When, following alignment, both conservative substitutions and identities were combined, the similarity rises to 60% of the total HCCP12 sequence (Hamodrakas *et al.*, 2002). This similarity allowed the construction, by "homology" modeling, of a structural model of the "extended R&R consensus" (Hamodrakas *et al.*, 2002). This modeling was successful even though it is clear that RBP and the R&R Consensus-bearing cuticular

proteins are not strictly homologous, for HCCP12 lacks the N-terminal region that is conserved in members of the lipocalin superfamily to which RBP belongs. The original model (**Figure 5a**) comprises the C-terminal 66 residues (out of 89 in total) of HCCP12 and corresponds to the "extended R&R consensus" (see Section 4.2.3.2.3).

Does this model fit both major classes of RR proteins? Stereo plots of this model of HCCP12 (**Figure 4a**) can be compared to comparable models of two RR-2 proteins (HCCP66 and AGCP2b) (**Figure 4b and c**). These models demonstrated that the extended R&R Consensus of both "soft" and "hard" cuticle proteins may easily adopt the proposed conformation.

How would this proposed structure interact with chitin? A low-resolution docking experiment of an extended *N*-acetylglucosamine tetramer to the model of HCCP12, utilizing the docking program GRAMM (Vakser, 1996) revealed that the proposed model for cuticle proteins accommodates, rather comfortably, at least one extended chitin chain (**Figure 4d**) (Hamodrakas *et al.*, 2002). The features revealed by secondary structure predictions (see Section 4.2.5.1) and by experimental spectroscopic analysis (see Section 4.2.5.2) work exceedingly well with this model. It is an antiparallel β-sheet structure with a "cleft" full of conserved aromatic residues that form "flat" hydrophobic surfaces on one "face," perfectly positioned to stack against faces of the saccharide rings of chitin. One unpredicted feature in the model is a short (seven-residue) two-turn α-helix at the C-terminus of the extended R&R Consensus of HCCP12, starting and ending with two proline residues, present in 60% of the "soft" cuticle proteins in **Figure 1**. This C-terminal part of the model is reminiscent in some respects of the chitin-binding domain of an invertebrate chitin-binding lectin, a two-stranded β-sheet followed by a helical turn (Suetake *et al.*, 2000). The structures of these two different chitin-binding proteins cannot be superimposed, however, and show no sequence similarity.

For this review, the proposed half-barrel model (see **Chapter 4.8**). has been used as a basis for more detailed docking experiments. A "high-resolution" docking experiment with the same tetramer was performed and the results are displayed and discussed in **Figure 5**. A new possibility emerges from this "high-resolution" experiment: the chitin chains can run either parallel to the β-strands of the half β-barrel model (**Figure 5b and c**), in good agreement with the observations of Atkins (1985), or perpendicular to the β-strands (**Figure 5a**). **Figure 5b and c** also provide an instructive view as to how a twisted helicoidal structure might arise from a close packing interaction of half-β-barrel models of cuticle

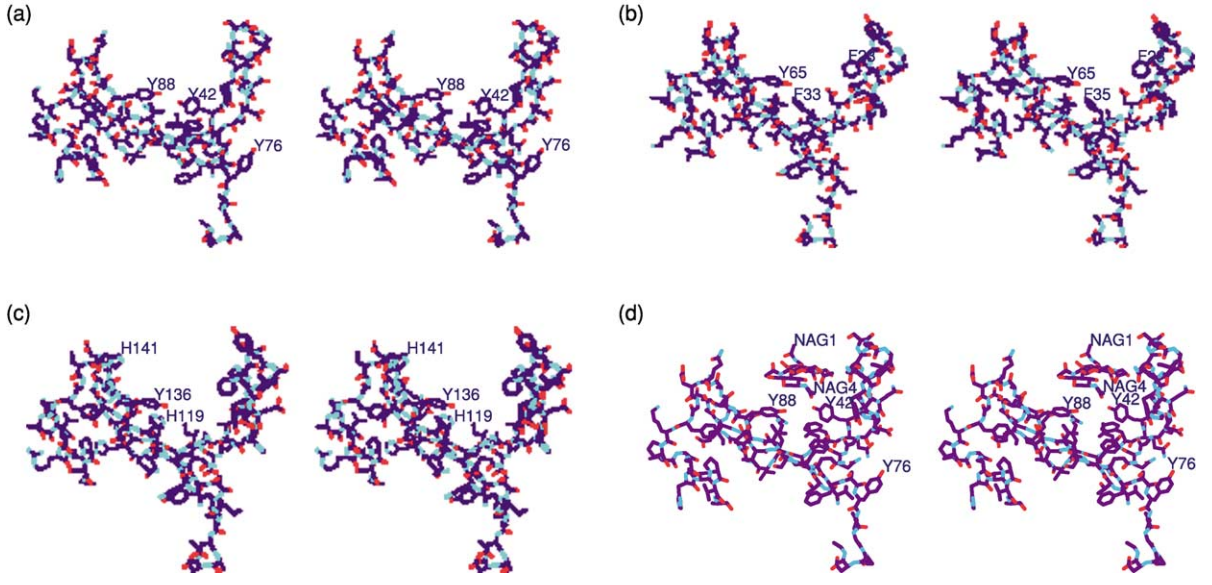


Figure 4 Stero pairs of cuticular proteins and their interaction with chitin. Stero pairs of cuticular proteins drawn with the program O (Jones *et al.*, 1991). The numbering scheme used is that of the unprocessed proteins. (a) View of the ''soft'' cuticle protein HCCP12. (b) View of the ''hard'' cuticle protein HCCP66. The terminal Ile83 residue could not be modeled and is not shown. (c) View of ''hard'' cuticle protein AGCP2b. His140 and the terminal Val155 residues could not be modeled and are not shown. (d) HCCP12 shown with an *N*-acetyl glucosamine (NAG) tetramer in an extended conformation. The complex was derived from a ''low-resolution'' docking experiment of a NAG tetramer, in an extended conformation, with the model of HCCP12, utilizing the docking program GRAMM (Vakser, 1996) and the default parameters of the program. (Reprinted with permission from Hamodrakas, S.J., Willis, J.H., Iconomidou, V.A., **2002.** A structural model of the chitin-binding domain of cuticle proteins. *Insect Biochem. Mol. Biol. 32*, 1577–1583, © Elsevier.)

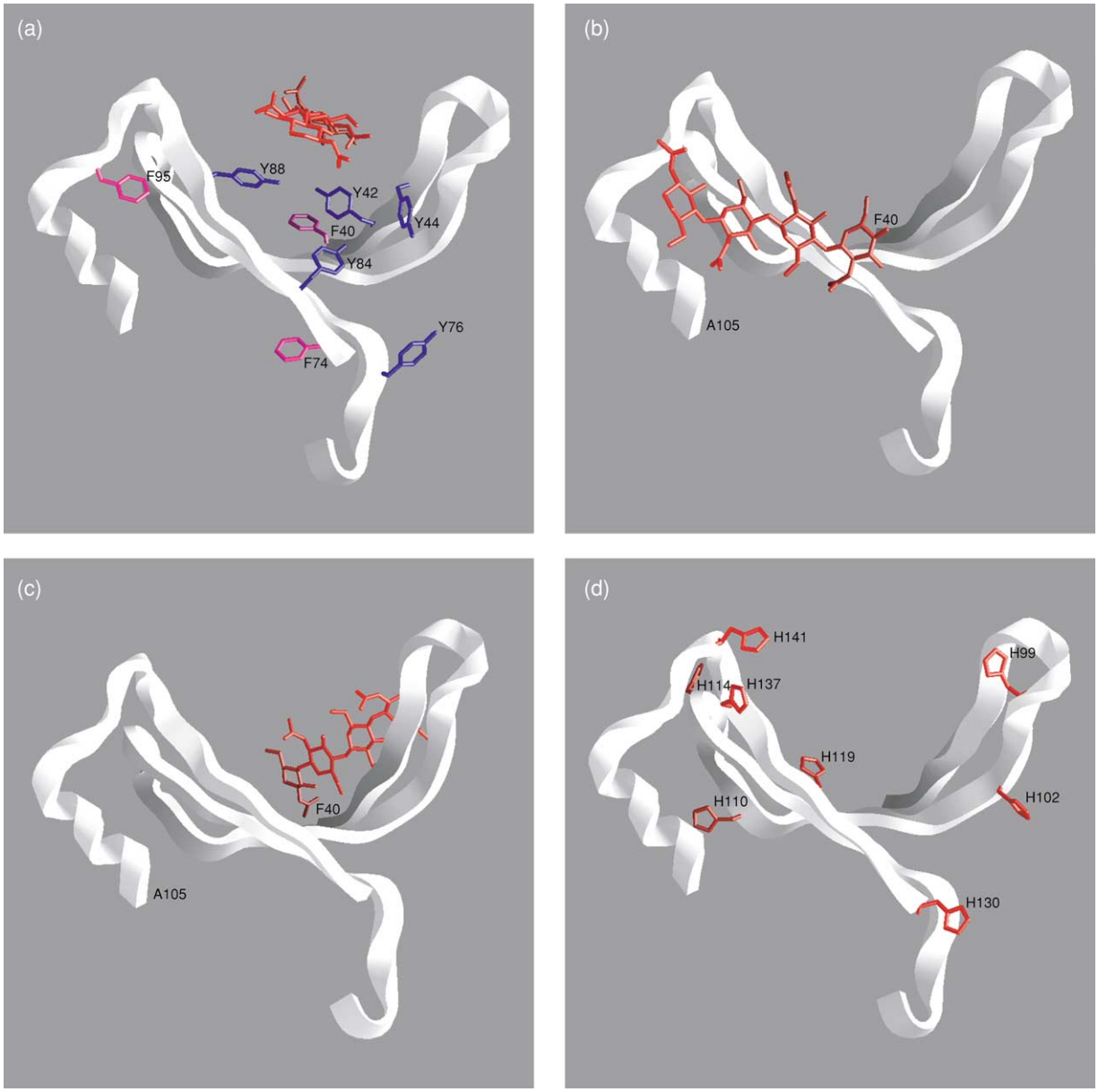


Figure 5 Ribbon models of cuticular proteins derived from homology modeling. (a) A ribbon model of cuticle protein structure, displayed using GRASP (Nicholls *et al.*, 1991). The structure of the representative ''soft'' cuticle protein HCCP12 was modeled on that of bovine retinol binding protein (RBP; PDB code 1FEN) (Zanotti *et al.*, 1994) utilizing the program WHAT IF (Vriend, 1990). Further details are in Hamodrakas *et al.* (2002). The side chains of several aromatic residues are shown and numbered, following the numbering scheme of the unprocessed HCCP12 sequence. These are: F40, Y42, Y44, H52, F74, Y76, Y84, Y88, and F95, underlined and bolded in **Figure 1**. The model structure has a ''cleft'' full of aromatic residues, which form ''flat'' surfaces of aromatic rings (upper side), ideally suited for cuticle protein–chitin chain interactions, and an outer surface (lower side) which should be important for protein–protein interactions in cuticle. The model is a complex of HCCP12 with an *N*-acetyl glucosamine (NAG) tetramer in an extended conformation. The complex was derived from a ''low-resolution'' docking experiment of a NAG tetramer, in an extended conformation, with the model of HCCP12, utilizing the docking program GRAMM (Vakser, 1996) and the default parameters of the program (a view similar to **Figure 4d**). (b) and (c) Two more possible complexes of HCCP12 with an NAG tetramer in an extended conformation derived from a ''high-resolution'' docking experiment, utilizing the program GRAMM (Vakser, 1996) and the default parameters of the program for ''high resolution.'' The two models presented in (b) and (c) are the two ''top on the list,'' most favorable complexes, whereas third on the list is a structure similar to that of (a). The one in (b) has the NAG tetramer more or less parallel to the last β-strand of the HCCP12 half β-barrel model, whereas that in (c) has the NAG tetramer more or less parallel to the first β-strand of the HCCP12 half β-barrel model. Note that, both in (b) and (c) the chitin chain runs parallel to the β-strands, whereas in (a) the chain is arranged perpendicular to the β-strands. (d) A display of a model of the ''hard cuticle'' protein AGCP2b. The numbering is that of the unprocessed protein. Histidine (H) side chains are shown as ''ball and sticks,'' in red, with their corresponding numbering. The corresponding residues are underlined in the sequence for AGCP2b in **Figure 2**.

proteins with chitin chains. It can be seen that the chitin chains, although more or less parallel to the β-strands, are forming an angle of the order of 10–15 degrees with the β-strands, and this, together with the inherent twist of the β-strands in the β-barrel, could provide the basis for the twisted helicoidal structure of the cuticle in general. Thus, both the inherent twist of the half-barrel β-sheet of the cuticle proteins and its packing arrangement at an angle with the chitin chains may provide a molecular basis for the morphological observation of a helicoidal twist in cuticle.

The model proposed by Hamodrakas *et al.* (2002) was subjected to a further test, namely that it should provide for the right positioning of histidine residues in the "hard cuticle" proteins, so that these histidines might play a significant role in cuticle sclerotization (Neville, 1975; Andersen *et al.*, 1995a (see **Chapter 4.4**). Histidines are a common feature of the extended R&R Consensus of many RR-2 proteins and many of their positions are conserved (**Figure 2**) (see Section 4.2.3.2.3). If they are to function in protein cross-linking by sclerotizing agents, they must reside on opposite faces to the aromatic residues that were postulated to interact with chitin (Iconomidou *et al.*, 1999; Hamodrakas *et al.*, 2002). In **Figure 5d**, a model of the "hard cuticle" protein AGCP2b is shown, similar in orientation to **Figure 5a**, indicating the positions of the histidine side chains. The relevant histidines are underlined and in bold in the sequence for AGCP2b in **Figure 2**. Three histidines, H102, H110, and H141 (the second, third, and last) are at sites where histidines are common. Such an interspecies conservation of these histidines, most probably signifies their very important structural and functional role (see below).

All the bolded and underlined histidines occupy "exposed" positions either in turns (like H99, H114, H137, H141), or at the "edges" of the half-β-barrel or its periphery (like H102, H110, H130), in excellent positions to be involved in cuticular sclerotization, readily reacting with activated *N*-acetyldopamine residues. Alternatively, they could be involved in the variations of the water-binding capacity of cuticle and the interactions of its constituent proteins, because small changes of pH can affect the ionization of their imidazole groups (Andersen *et al.*, 1995a).

These observations are in excellent agreement with the predictions made several years ago for the role of histidines from secondary structure predictions (Iconomidou *et al.*, 1999) and strengthen further the value of the model previously proposed both for "soft" and "hard" cuticle proteins (Hamodrakas *et al.*, 2002).

4.2.5.4. Fusion Proteins Establish a Role for the Extended R&R Consensus

Predictions of secondary and tertiary structure and experimental evidence supporting them (discussed above in Sections 4.2.5.1–4.2.5.3) established that the extended R&R Consensus has the properties to serve as a chitin-binding motif. In particular, the planar surfaces of the predicted β-sheets will expose aromatic residues positioned for protein–chitin interaction. The ultimate test of these predictions would be to show that the extended consensus region is sufficient to confer chitin binding on a protein.

Rebers and Willis (2001) investigated this possibility by creating fusion proteins using the extended R&R Consensus from the *A. gambiae* putative cuticular protein, AGCP2b. First they expressed this protein in *Escherichia coli* and isolated it from cell lysates. The construct used coded for the complete protein minus the predicted signal peptide and had a histidine-tag added to the N-terminus to facilitate purification (Dotson *et al.*, 1998). AGCP2b is a protein of 222 amino acids, with a RR-2 type of consensus. The purified protein bound to chitin beads and could be eluted from these beads with 8 M urea or boiling SDS. This established unequivocally that AGCP2b was a chitin-binding protein. Chitin binding previously had been obtained with mixtures of protein extracted from cuticles of two beetles and *D. melanogaster* (Hackman, 1955; Fristrom *et al.*, 1978; Hackman and Goldberg, 1978).

The next, and essential, step was to create a fusion protein uniting a protein that did not bind to chitin with the extended R&R Consensus region. Such a fusion was created between glutathione-*S*-transferase (GST) and 65 amino acids for AGCP2b, covering the region of pfam00379, as shown in **Figure 6**.

The GST and the fusion protein were each affinity purified using a glutathione–sepharose column. GST alone did not bind to chitin but the fusion protein did, requiring denaturing agents for release.

Other experiments defined in more detail the requirements for converting GST into a chitin-binding protein. A shorter fragment of AGCP2b,

APANYEFSYSVHDEH**T**G**D**IKSQHETRHGDEVH*GQYSLLDSDGHQRIVD**Y**HADHHTG**F**NAVVREP*

Figure 6 The pfam00379 region of AGcP26 used to construct fusion proteins. For details, see text.

40 amino acids (underlined in **Figure 6**) with the strict R&R Consensus (shown in italics) did not bind chitin. Nor did the full construct when either the Y and F (bolded and highlighted) of the strict R&R Consensus or the T and D (highlighted) of the extended consensus were "mutated" to alanine (Rebers and Willis, 2001).

These experiments established, at last, that the extended R&R Consensus is sufficient to confer chitin-binding properties on a protein and thereby resolved years of speculation on the importance of this region.

Chitinase, some lectins, and proteins from peritrophic membranes all bind chitin (review: Shen and Jacobs-Lorena, 1999). What is unique about the extended R&R Consensus is that it lacks cysteine residues. These residues serve essential roles in the other types of chitin-binding proteins, forming disulfide bonds that hold the protein in the proper configuration for binding. While these other chitin-binding proteins have weak sequence similarities to one another, they do not approach the sequence conservation seen in the R&R Consensus throughout the arthropods. Rebers and Willis (2001) suggested that this conservation (see **Figures 1** and **2**) could well be due to the need to preserve a precise conformation of the chitin-binding domain in the absence of stabilizing disulfide bonds.

In addition to establishing a function of the extended R&R Consensus, these experiments also provided confirmation of key elements in the models discussed above (see Section 4.2.5.4). Substitution of the two conserved aromatic residues abolished chitin binding. With the TD "mutations," alanines were substituted for two other conserved residues. These flank a glycine that is conserved in position in the "extended consensus" of all hard and many soft cuticles (Iconomidou *et al.*, 1999). According to the proposed model (**Figure 5a**), these two polar residues would point away from the hydrophobic "cleft" and thus should not participate in chitin binding. It should be noted, however, that this glycine is located at a sharp turn, at the end of the second β-strand (in the vicinity of H102 of **Figure 5d**). The substitution of two polar residues by two alanines may result in destruction of this turn and to improper folding, thus leading to a structure not capable of binding chitin.

4.2.5.5. Summary

Four different types of data have been presented above (see Section 4.2.5) analyzing the extended R&R Consensus: secondary structure predictions of antiparallel β-sheets (see Section 4.2.5.1), experimental spectroscopic evidence from cuticles and cuticle extracts for the predominance of such β-sheets in cuticular protein conformation (see Section 4.2.5.2), models showing organization of the consensus into a half β-barrel with a groove that can accommodate chitin (see Section 4.2.5.3), and direct demonstration that the extended consensus is sufficient to confer chitin binding on a protein (see Section 4.2.5.4). These four types of data are all in agreement that the highly conserved amino acid sequence of the extended R&R Consensus forms a novel chitin-binding domain, albeit one that displays an essential feature of other proteins that interact with chitin, namely the presentation of aromatic residues in a planar surface. Crystal structures of the cuticular protein–chitin complex are needed to assure that these inferences are correct.

4.2.6. Comparison of Cuticle and Chorion: Structure and Proteins

Silkmoth and fish chorions (eggshells) and cuticle are known to have a helicoidal architecture (Neville, 1975; Hamodrakas, 1992). Excellent reviews on helicoidal architecture and its appearance in biological systems have been made by Bouligand (1972, 1978a, 1978b) and Neville (1975, 1981, 1986). These works describe, in a beautiful and most comprehensive way, how helicoids are identified, how widespread they are, and the basic molecular principles of their formation as well as their geometrical, physical, and biological properties. The close analogy between the helicoidal structures of (usually extracellular) biological materials and the structure of cholesteric liquid crystals suggests that these structures self-assemble according to a mechanism that is very similar to the process allowing materials to form liquid crystals. Apparently, helicoids should pass through a liquid crystalline phase before solidifying. It is assumed that this occurs in the assembly zone during cuticle formation. Self-assembling systems are important in biology, as they are economical in energy terms, requiring neither enzymatic control nor the expenditure of energy-rich bonds. They are particularly appropriate for building extracellular skeletal structures outside of the cells that secrete the components (Bouligand, 1978a, 1978b; Neville, 1986).

Silkmoth chorion is produced by the follicular cells that surround the oocyte (Regier and Kafatos, 1985 and references therein). Fish eggshell is mainly produced by the oocyte, with minor contributions from the follicular cells (Hamodrakas, 1992 and references therein) and cuticle is produced by the epidermis.

Natural helicoidal composites occur in several combinations such as polysaccharide fibers in a polysaccharide matrix (plant cell walls), polysaccharide fibers in a protein matrix (arthropod cuticle), and protein fibers in a protein matrix (insect and fish eggshells). In all cases, principles of molecular recognition and weak intermolecular interactions should govern the self-assembly mechanisms (Neville, 1986).

In silkmoth chorion, disulfide bonds, and in fish eggshell, isopeptide bonds between the side chains of R-K and D-E, are major contributors to stabilization. These covalent bonds, however, are totally absent in cuticle, where stabilization occurs via protein–chitin interaction and by cross-linking by sclerotization compounds (see **Chapter 4.4**).

It is clear that the main characteristic of chorion proteins is the presence of exact, tandemly repeating hexapeptide motifs that adopt a characteristic antiparallel β-pleated sheet structure. This is the main structural unit of silkmoth chorion fibrils and, apparently, the molecular denominator, which dictates formation of the helicoidal architecture (Hamodrakas, 1992). The ellipsoidal shape of silkmoth chorions is, most probably, due to the fact that the basic building-blocks, chorion protein fibrils, are so uniform in shape. By contrast, in cuticle, despite the fact that there are regions of the molecules rich in tandem repeats of certain motifs (see Section 4.2.3.2.2), the sequences are mainly characteristic of globular proteins, and cuticle may adopt all sorts of shapes depending on the local needs of the arthropods producing it. The majority of cuticular proteins contain a conserved domain, rich in a characteristic antiparallel β-pleated sheet structure, a half β-barrel (see Section 4.2.5.3) which again should serve as the molecular denominator determining the helicoidal structure of cuticle, interacting with chitin crystalline chains and giving rise to a plethora of architectural plans as needed locally.

Apparently, an antiparallel β-pleated sheet type of structure is the common molecular denominator, that dictates the helicoidal architecture adopted by the chorion of Lepidoptera and fish and also by the arthropod cuticle.

4.2.7. Summary and Future Challenges

This review has summarized the wealth of information about cuticular proteins amassed since Silvert's review in 1985. Most striking is that the 35-fold increase in sequences for structural cuticular proteins has revealed that the majority has a conserved domain (pfam00379) that is an extended version of the R&R Consensus. A group of proteins that appears to contribute to hard cuticles have a highly conserved extended consensus (RR-2). It is now known that RR-2 proteins interact with chitin and we can predict in some detail the features of their sequence that confer this property. It is not known whether the RR-1 proteins are as effective in binding chitin. We have not yet begun to analyze how the regions outside the consensus contribute to cuticular properties, nor have we learned how the proteins lacking the consensus but with other conserved features contribute to cuticle structure.

Cuticular proteins with pfam00379 are one of the largest multigene families found in *Drosophila* (Lespinet *et al.*, 2002). We need to learn whether this multiplicity serves to allow rapid synthesis of cuticle or whether different genes are used to construct cuticles in different regions. If the latter, the question becomes whether subtle differences in sequence are important for different cuticular properties, or if gene multiplication has been exploited to allow precise temporal and spatial control. The elegant immunolocalization studies that have been carried out were done with antibodies against proteins whose sequences for the most part are unknown. Now that we recognize that several genes may have almost identical sequences, we have to be very careful in designing specific probes for use in Northern analyses, for *in situ* hybridization, and for immunolocalization, if our goal is to learn the use to which each individual gene is put.

Cuticular protein sequences are certain to be described in ever-increasing numbers as more insect genomes are analyzed. Describers need to be careful to submit to databases an indication of whether assignment as a cuticular protein is based on sequence alone or on some type of corroborating evidence. It would be helpful if there were a more consistent system for naming cuticular proteins. At the very least, each protein should have a designation of genus and species and a unique number.

A wealth of information is available already but many challenges lie ahead for those who wish to continue to further our understanding of how the diverse forms and properties of cuticle are constructed extracellularly as these proteins self-assemble in proximity to chitin.

Acknowledgments

We appreciate the helpful comments of John Rebers, Augustine Dunn, and John S. Willis.

References

Agrawal, O.P., Scheller, K., **1986**. The formation of the chitin-arylphorin complex *in vitro*. In: Muzzarelli, R., Jeuniaux, C., Gooday, G.W. (Eds.), Chitin in Nature and Technology. Plenum Press, New York, pp. 316–320.

Andersen, S.O., **1975**. Cuticular sclerotization in the beetles *Pachynoda epphipiata* and *Tenebrio molitor*. *J. Insect Physiol. 21*, 1225–1232.

Andersen, S.O., **1998**. Amino acid sequence studies on endocuticular proteins form the desert locust, *Schistocerca gregaria*. *Insect Biochem. Mol. Biol. 28*, 421–434.

Andersen, S.O., **2000**. Studies on proteins in post-ecdysial nymphal cuticle of locust, *Locusta migratoria*, and cockroach, *Blaberus cranifer*. *Insect Biochem. Mol. Biol. 30*, 569–577.

Andersen, S.O., Hojrup, P., Roepstorff, P., **1995a**. Insect cuticular proteins. *Insect Biochem. Mol. Biol. 25*, 153–176.

Andersen, S.O., Rafn, K., Krogh, T.N., Hojrup, P., Roepstorff, P., **1995b**. Comparison of larval and pupal cuticular proteins in *Tenebrio molitor*. *Insect Biochem. Mol. Biol. 25*, 177–187.

Andersen, S.O., Rafn, K., Roepstorff, P., **1997**. Sequence studies of proteins from larval and pupal cuticle of the yellow meal worm, *Tenebrio molitor*. *Insect Biochem. Mol. Biol. 27*, 121–131.

Andersen, S.O., Weis-Fogh, T., **1964**. Resilin: a rubber-like protein in arthropod cuticle. *Adv. Insect Physiol. 2*, 1–65.

Apple, R.T., Fristrom, J.W., **1991**. 20-Hydroxyecdysone is required for, and negatively regulates, transcription of *Drosophila* pupal cuticle protein genes. *Devel. Biol. 146*, 569–582.

Ardell, D.H., Andersen, S.O., **2001**. Tentative identification of a resilin gene in *Drosophila melanogaster*. *Insect Biochem. Mol. Biol. 31*, 965–970.

Ashida, M., Brey, P.T., **1995**. Role of the integument in insect defense: pro-phenol oxidase cascade in the cuticular matrix. *Proc. Natl Acad. Sci. USA 92*, 10698–10702.

Atkins, E.D.T., **1985**. Conformations in polysaccharides and complex carbohydrates. *Proc. Int. Symp. Biomol. Struct. Interactions, Suppl. J. Biosci. 8*, 375–387.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., *et al.*, **2002**. The Pfam protein families database. *Nucleic Acids Res. 30*, 276–280.

Binger, L.C., Willis, J.H., **1990**. *In vitro* translation of epidermal RNAs from different anatomical regions and metamorphic stages of *Hyalophora cecropia*. *Insect Biochem. 20*, 573–583.

Binger, L.C., Willis, J.H., **1994**. Identification of the cDNA, gene and promoter for a major protein from flexible cuticles of the giant silkmoth *Hyalophora cecropia*. *Insect Biochem. Mol. Biol. 24*, 989–1000.

BLAST, **2003**. http://www.ncbi.nlm.nih.gov/BLAST/

Bouhin, H., Charles, J.-P., Quennedey, B., Courrent, A., Delachambre, J., **1992a**. Characterization of a cDNA clone encoding a glycine-rich cuticular protein of *Tenebrio molitor*: developmental expression and effect of a juvenile hormone analogue. *Insect Mol. Biol. 1*, 53–62.

Bouhin, H., Charles, J.-P., Quennedey, B., Delachambre, J., **1992b**. Developmental profiles of epidermal mRNAs during the pupal–adult molt of *Tenebrio molitor* and isolation of a cDNA encoding an adult cuticular protein: effects of a juvenile hormone analogue. *Devel. Biol. 149*, 112–122.

Bouligand, Y., **1972**. Twisted fibrous arrangements in biological materials and cholesteric mesophases. *Tissue and Cell 4*, 189–217.

Bouligand, Y., **1978a**. Cholesteric order in biopolymers. *ACS Symposium Series 74*, 237–247.

Bouligand, Y., **1978b**. Liquid crystalline order in biological materials. In: Blumstein, A. (Ed.), Liquid Crystalline Order in Polymers. Academic Press, London, pp. 261–297.

Browder, L.W., Wilkes, J., Rodenhiser, D.I., **1992**. Preparative labeling of proteins with [^{35}S]methionine. *Anal. Biochem. 204*, 85–89.

Carter, D., Locke, M., **1993**. Why caterpillars do not grow short and fat. *Int. J. Insect Morphol. Embryol. 22*, 81–102.

Charles, J.-P., Bouhin, H., Quennedey, B., Courrent, A., Delachambre, J., **1992**. cDNA cloning and deduced amino acid sequence of a major, glycine-rich cuticular protein from the coleopteran *Tenebrio molitor*: temporal and spatial distribution of the transcript during metamorphosis. *Eur. J. Biochem. 206*, 813–819.

Charles, J.-P., Chihara, C., Nejad, S., Riddiford, L.M., **1997**. A cluster of cuticle protein genes of *Drosophila melanogaster* at 65A: sequence, structure and evolution. *Genetics 147*, 1213–1226.

Charles, J.-P., Chihara, C., Nejad, S., Riddiford, L.M., **1998**. Identification of proteins and developmental expression of RNAs encoded by the 65A cuticle protein gene cluster in *Drosophila melanogaster*. *Insect Biochem. Mol. Biol. 28*, 131–138.

Chou, P., Fasman, G.D., **1974a**. Conformational parameters for amino acids in helical, β-sheet and random coil regions calculated from proteins. *Biochemistry 13*, 211–221.

Chou, P., Fasman, G.D., **1974b**. Prediction of protein conformation. *Biochemistry 13*, 222–245.

Condoulis, W.V., Locke, M., **1966**. The deposition of endocuticle in an insect, *Calpodes ethlius* Stoll (Lepidoptera, Hesperiidae). *J. Insect Physiol. 12*, 311–323.

Cornell, J.C., Pan, M.L., **1983**. The disappearance of moulting fluid in the tobacco hornworm, *Manduca sexta*. *J. Exp. Biol. 107*, 501–504.

Cox, D.L., Willis, J.H., **1985**. The cuticular proteins of *Hyalophora cecropia* from different anatomical regions and metamorphic stages. *Insect Biochem. 15*, 349–362.

Cox, D.L., Willis, J.H., **1987a**. Analysis of the cuticular proteins of *Hyalophora cecropia* with two dimensional electrophoresis. *Insect Biochem. 17*, 457–468.

Cox, D.L., Willis, J.H., **1987b**. Post-translational modifications of the cuticular proteins of *Hyalophora cecropia*

from different anatomical regions and metamorphic stages. *Insect Biochem. 17*, 469–484.

Csikos, G., Molnar, K., Borhegyi, N.H., Talian, G.C., Sass, M., **1999**. Insect cuticle, an *in vivo* model of protein trafficking. *J. Cell Sci. 112*, 2113–2124.

Doctor, J., Fristrom, D., Fristrom, J.W., **1985**. The pupal cuticle of *Drosophila*: biphasic synthesis of pupal cuticle proteins *in vivo* and *in vitro* in response to 20-hydroxyecdysone. *J. Cell Biol. 101*, 189–200.

Dotson, E.M., Cornel, A.J., Willis, J.H., Collins, F.H., **1998**. A family of pupal-specific cuticular protein genes in the mosquito *Anopheles gambiae*. *Insect Biochem. Mol. Biol. 28*, 459–472.

ENTREZ, **2003**. http://www3.ncbi.nlm.nih.gov/Entrez/index.html

Fechtel, K., Fristrom, D.K., Fristrom, J.W., **1989**. Prepupal differentiation in *Drosophila*: distinct cell types elaborate a shared structure, the pupal cuticle, but accumulate transcripts in unique patterns. *Development 106*, 649–656.

Flybase, **2003**. Genetic data on *Drosophila*. http://flybase.bio.indiana.edu/

Fraenkel, G., Rudall, K.M., **1947**. The structure of insect cuticles. *Proc. Roy. Soc. B 134*, 111–143.

Fristrom, D., Doctor, J., Fristrom, J.W., **1986**. Procuticle proteins and chitin-like material in the inner epicuticle of the *Drosophila* pupal cuticle. *Tissue and Cell 18*, 531–543.

Fristrom, D., Fristrom, J.W., **1993**. The metamorphic development of the adult epidermis. In: Bate, M., Martinez Arias, A. (Eds.), The Development of *Drosophila melanogaster*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 843–897.

Fristrom, J.W., Hill, R.J., Watt, F., **1978**. The procuticle of *Drosophila*: heterogeneity of urea-soluble proteins. *Biochemistry 17*, 3917–3924.

Geyer, P.G., Corces, V.G., **1987**. Separate regulatory elements are responsible for the complex pattern of tissue-specific and developmental transcription of the *yellow* locus in *Drosophila melanogaster*. *Genes and Devel. 1*, 996–1004.

Griebenow, K., Santos, A.M., Carrasquillo, K.G., **1999**. Secondary structure of proteins in the amorphous dehydrated state probed by FTIR spectroscopy: dehydration-induced structural changes and their prevention. *Int. J. Vibr. Spectrosc. 3*, 1–34.

Gu, S., Willis, J.H., **2003**. Distribution of cuticular protein mRNAs in silk moth integument and imaginal discs. *Insect Biochem. Mol. Biol. 33*, 1177–1188.

Hackman, R.H., **1955**. Studies on chitin. 3. Absorption of proteins to chitin. *Austral. J. Biol. Sci. 8*, 530–536.

Hackman, R.H., Goldberg, M., **1978**. The non-covalent binding of two insect cuticular proteins by a chitin. *Insect Biochem. 8*, 353–357.

Hamodrakas, S.J., **1992**. Molecular architecture of helicoidal proteinaceous eggshells. In: Case, S.T. (Ed.), Results and Problems in Cell Differentiation, vol. 19. Springer-Verlag, Berlin, pp. 116–186.

Hamodrakas, S.J., Kanellopoulos, P.N., Pavlou, K., Tucker, P.A., **1997**. The crystal structure of the complex of concanavalin A with 4-methylumbelliferyl-α-D-glucopyranoside. *J. Struct. Biol. 118*, 23–30.

Hamodrakas, S.J., Willis, J.H., Iconomidou, V.A., **2002**. A structural model of the chitin-binding domain of cuticle proteins. *Insect Biochem. Mol. Biol. 32*, 1577–1583.

Han, Q., Fang, J., Ding, H., Johnson, J.K., Christensen, B.M., *et al.*, **2002**. Identification of *Drosophila melaogaster* yellow-f and yellow-f2 proteins as dopachrome-conversion enzymes. *Biochem. J. 368*, 333–340.

Henikoff, S., Eghtedarzadeh, M.K., **1987**. Conserved arrangement of nested genes at the *Drosophila Gart* locus. *Genetics 117*, 711–725.

Henikoff, S., Keene, M.A., Fechtel, K., Fristrom, J.W., **1986**. Gene within a gene: nested *Drosophila* genes encode unrelated proteins on opposite DNA strands. *Cell 44*, 33–42.

Holden, H.M., Rypniewski, W.R., Law, J.H., Rayment, I., **1987**. The molecular structure of insecticyanin from the tobacco hornworm *Manduca sexta* L. at 2.6 A resolution. *EMBO J. 6*, 1565–1570.

Horodyski, F.M., Riddiford, L.M., **1989**. Expression and hormonal control of a new larval cuticular multigene family at the onset of metamorphosis of the tobacco hornworm. *Devel. Biol. 132*, 292–303.

Iconomidou, V.A., Chryssikos, G.D., Gionis, V., Willis, J.H., Hamodrakas, S.J., **2001**. "Soft"-cuticle protein secondary structure as revealed by FT-Raman, ATR-FT-IR and CD spectroscopy. *Insect Biochem. Mol. Biol. 31*, 877–885.

Iconomidou, V.A., Willis, J.H., Hamodrakas, S.J., **1999**. Is β-pleated sheet the molecular conformation which dictates formation of helicoidal cuticle? Insect Biochem. *Mol. Biol. 29*, 285–292.

Ikeya, T., Persson, P., Kono, M., Watanabe, T., **2001**. The *DD5* gene of the decapod crustacean *Penaeus japonicus* encodes a putative exoskeletal protein with a novel tandem repeat structure. *Comp. Biochem. Physiol. B 128*, 379–388.

Janssens, H., Gehring, W.J., **1999**. Isolation and characterization of *drosocrystallin*, a lens crystallin gene of *Drosophila melanogaster*. *Devel. Biol. 207*, 204–214.

Jones, T.A., Zou, J.Y., Cowan, S.W., Kjeldgaard, M., **1991**. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A47*, 110–119.

Kalinich, J.F., McClain, D.E., **1992**. An *in vitro* method for radiolabeling proteins with ^{35}S. *Anal. Biochem. 205*, 208–212.

Kaufman, T.C., Seeger, M.A., Olsen, G., **1990**. Molecular and genetic organization of the Antennapedia gene complex of *Drosophila melanogaster*. *Adv. Genet. 27*, 309–362.

Kollberg, U., Obermaier, B., Hirsch, H., Kelber, G., Wobert, P., **1995**. Expression cloning and characterization of a pupal cuticle protein cDNA of *Galleria mellonella* L. Insect Biochem. *Mol. Biol. 25*, 355–363.

Konig, M., Agrawal, O.P., Schenkel, H., Scheller, K., **1986**. Incorporation of calliphorin into the cuticle of the developing blowfly, *Calliphora vicinia*. *Roux's Arch. Devel. Biol. 195*, 296–301.

Kornezos, A., Chia, W., **1992**. Apical secretion and association of the *Drosophila yellow* gene product with developing larval cuticle structures during embryogenesis. *Mol. Gen. Genet. 235*, 397–405.

Lampe, D.J., Willis, J.H., **1994**. Characterization of a cDNA and gene encoding a cuticular protein from rigid cuticles of the giant silkmoth, *Hyalophora cecropia*. *Insect Biochem. Mol. Biol. 24*, 419–435.

Lee, W.-J., Brey, P.T., **1994**. Isolation and identification of cecropin antibacterial peptides from the extracellular matrix of insect integument. *Anal. Biochem. 217*, 231–235.

Lemoine, A., Delachambre, J., **1986**. A water-soluble protein specific to the adult cuticle in *Tenebrio*: its use as a marker of a new programme expressed by epidermal cells. *Insect Biochem. 16*, 483–489.

Lemoine, A., Millot, C., Curie, G., Delachambre, J., **1989**. A monoclonal antibody against an adult-specific cuticular protein of *Tenebrio molitor* (Insecta, Coleoptera). *Devel. Biol. 136*, 546–554.

Lemoine, A., Millot, C., Curie, G., Delachambre, J., **1990**. Spatial and temporal variations in cuticle proteins as revealed by monoclonal antibodies, immunoblotting analysis and ultrastructural immunolocalization in a beetle, *Tenebrio molitor. Tissue and Cell 22*, 177–189.

Lemoine, A., Millot, C., Curie, G., Massonneau, V., Delachambre, J., **1993**. Monoclonal antibodies recognizing larval- and pupal-specific cuticular proteins of *Tenebrio molitor* (Insecta, Coleoptera). *Roux's Arch. Devel. Biol. 203*, 92–99.

Lespinet, O., Wolf, Y.I., Koonin, E.V., Aravind, L., **2002**. The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Res. 12*, 1048–1059.

Leung, H., Palli, S.R., Locke, M., **1989**. The localization of arylphorin in an insect, *Calpodes ethlius*. *J. Insect Physiol. 35*, 223–231.

Li, W., Riddiford, L.M., **1992**. Two distinct genes encode two major isoelectric forms of insecticyanin in the tobacco hornworm, *Manduca sexta*. *Eur. J. Biochem. 205*, 491–499.

Locke, M., **1998**. Epidermis. In: Harrison, F.W., Locke, M. (Eds.), Microscopic Anatomy of Invertebrates, vol. 11A. Wiley–Liss, New York, pp. 75–138.

Locke, M., **2001**. The Wigglesworth Lecture: Insects for studying fundamental problems in biology. *J. Insect Physiol. 47*, 495–507.

Locke, M., **2003**. Surface membranes, golgi complexes and vacuolar systems. *Annu. Rev. Entomol. 48*, 1–27.

Locke, M., Kiss, A., Sass, M., **1994**. The cuticular localization of integument peptides from particular routing categories. *Tissue and Cell 26*, 707–734.

Lombardi, E.C., Kaplan, D.L., **1993**. Preliminary characterization of resilin isolated from the cockroach, *Periplanata americana*. *Mat. Res. Soc. Proc. 292*, 3–7.

Marcu, O., Locke, M., **1998**. A cuticular protein from the moulting stages of an insect. *Insect Biochem. Mol. Biol. 28*, 659–669.

Marcu, O., Locke, M., **1999**. The origin, transport and cleavage of the molt-associated cuticular protein CECP22 from *Calpodes ethlius* (Lepidoptera, Hesperridae). *J. Insect Physiol. 45*, 861–870.

Mathelin, J., Bouhin, H., Quennedey, B., Courrent, A., Delachambre, J., **1995**. Identification, sequence and mRNA expression pattern during metamorphosis of a cDNA encoding a glycine-rich cuticular protein in *Tenebrio molitor. Gene 156*, 259–264.

Mathelin, J., Quennedey, B., Bouhin, H., Delachambre, J., **1998**. Characterization of two new cuticular genes specifically expressed during the post-ecdysial molting period in *Tenebrio molitor. Gene 211*, 351–359.

McGeoch, D.J., **1985**. On the predictive recognition of signal peptide sequences. *Virus Res. 3*, 271–286.

Mita, K., Morimyo, M., Okano, K., Shimada, T., Maeda, S., **1999**. The construction of EST database for genome analysis of *Bombyx mori*. *RIKEN Rev. 22*, 63–67.

Molnar, K., Borhegyi, N.H., Csikos, G., Sass, M., **2001**. The immunoprotein scolexin and its synthesizing sites: the midgut epithelium and the epidermis. *Acta Biol. Hungarica 52*, 473–484.

MOTIF, **2003**. http://motif.genome.ad.jp/MOTIF2.html

Nakato, H., Izumi, S., Tomino, S., **1992**. Structure and expression of gene coding for a pupal cuticle protein of *Bombyx mori*. *Biochim. Biophys. Acta 1132*, 161–167.

Nash, W.G., **1976**. Patterns of pigmentation color states regulated by the *y* locus in *Drosophila melanogaster. Devel. Biol. 48*, 336–343.

Neville, A.C., **1975**. Biology of the Arthropod Cuticle. Springer-Verlag, Berlin.

Neville, A.C., **1981**. Cholesteric proteins. *Mol. Cryst. Liq. Cryst. 76*, 279–286.

Neville, A.C., **1986**. The physics of helicoids: multidirectional "plywood" structures in biological systems. *Phys. Bull. 37*, 74–76.

Nicholls, A., Sharp, K.A., Honig, B., **1991**. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet. 11*, 281–296.

Nielsen, H., Engelbrecht, J., Brunak, S., von Heijne, G., **1997**. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng. 10*, 1–6.

Nielsen, H., Krogh, A., **1998**. Prediction of signal peptides and signal anchors by a hidden Markov model. *Proc. Int. Conf. Intell. Syst. Mol. Biol. 6*, 122–130.

Palli, S.R., Locke, M., **1987**. The synthesis of hemolymph proteins by the larval epidermis of an insect *Calpodes ethlius* (Lepidoptera: Hesperiidae). *Insect Biochem. 17*, 711–722.

Perrakis, A., Ouzounis, C., Wilson, K.S., **1997**. Evolution of immunoglobulin-like modules in chitinases: their structural flexibility and functional implications. *Folding and Design 2*, 291–294.

Pfam, **2003a**. http://www.sanger.ac.uk/Software/Pfam/
Pfam, **2003b**. http://pfam.wustl.edu/cgi-bin/getdesc?acc=PF00379
Psort, **2003**. http://psort.ims.u-tokyo.ac.jp/
Rebers, J.E., Niu, J., Riddiford, L.M., **1997**. Structure and spatial expression of the *Manduca sexta* MSCP14.6 cuticle gene. *Insect Biochem. Mol. Biol. 27*, 229–240.
Rebers, J.F., Riddiford, L.M., **1988**. Structure and expression of a *Manduca sexta* larval cuticle gene homologous to *Drosophila* cuticle genes. *J. Mol. Biol. 203*, 411–423.
Rebers, J.E., Willis, J.H., **2001**. A conserved domain in arthropod cuticular proteins binds chitin. *Insect Biochem. Mol. Biol. 31*, 1083–1093.
Regier, J.C., Kafatos, F.C., **1985**. Molecular aspects of chorion formation. In: Kerkut, G.A., Gilbert, L.I. (Eds.), Comprehensive Insect Physiology, Biochemistry, and Pharmacology, vol. 1. Pergamon Press, Oxford, pp. 113–151.
Riddiford, L.M., **1994**. Cellular and molecular actions of juvenile hormone I: general considerations and premetamorphic actions. *Adv. Insect Physiol. 24*, 213–274.
Riddiford, L.M., Hice, R.H., **1985**. Developmental profiles of the mRNAs for *Manduca* arylphorin and two other storage proteins during the final larval instar of *Manduca sexta*. *Insect Biochem. 15*, 489–502.
Roberts, P.E., Willis, J.H., **1980a**. The cuticular proteins of *Tenebrio molitor*. 1. Electrophoretic banding patterns during postembryonic development. *Devel. Biol. 75*, 59–69.
Roberts, P.E., Willis, J.H., **1980b**. Effects of juvenile hormone, ecdysterone, actinomycin D, and mitomycin C on the cuticular proteins of *Tenebrio molitor*. *J. Embryol. Exp. Morph. 56*, 107–123.
Rondot, I., Quennedey, B., Courrent, A., Lemoine, A., Delachambre, J., **1996**. Cloning and sequencing of a cDNA encoding a larval–pupal-specific cuticular protein in *Tenebrio molitor* (Insecta, Coleoptera): developmental expression and effect of a juvenile hormone analogue. *Eur. J. Biochem. 235*, 138–143.
Rondot, I., Quennedey, B., Delachambre, J., **1998**. Structure, organization and expression of two clustered cuticle protein genes during the metamorphosis of an insect *Tenebrio molitor*. *Eur. J. Biochem. 254*, 304–312.
Roter, A.H., Spofford, J.B., Swift, H., **1985**. Synthesis of the major adult cuticle proteins of *Drosophila melanogaster* during hypoderm differentiation. *Devel. Biol. 107*, 420–431.
Sass, M., Kiss, A., Locke, M., **1993**. Classes of integument peptides. *Insect Biochem. Mol. Biol. 23*, 845–857.
Sass, M., Kiss, A., Locke, M., **1994a**. Integument and hemocyte peptides. *J. Insect Physiol. 40*, 407–421.
Sass, M., Kiss, A., Locke, M., **1994b**. The localization of surface integument peptides in tracheae and tracheoles. *J. Insect Physiol. 40*, 561–575.
Scheller, K., Zimmermann, H.-P., Sekeris, C.E., **1980**. Calliphorin, a protein involved in the cuticle formation of the blowfly, *Calliphora vicinia*. *Z. Naturforsch. 35c*, 387–389.
Scholer, H.R., **1991**. Octamania: the POU factors in murine development. *Trends Genet. 7*, 323–328.
Shen, Z., Jacobs-Lorena, M., **1999**. Evolution of chitin-binding proteins in invertebrates. *J. Mol. Evol. 48*, 341–347.
SilkBase, **2003**. Expressed sequence tags in *Bombyx mori*. http://www.ab.a.u-tokyo.ac.jp/silkbase
Silvert, D.J., **1985**. Cuticular proteins during postembryonic development. In: Kerkut, G.A., Gilbert, L.I. (Eds.), Comprehensive Insect Physiology, Biochemistry, and Pharmacology, vol. 2. Pergamon Press, Oxford, pp. 239–254.
Snyder, M., Hunkapiller, M., Yuen, D., Silvert, D., Fristrom, J., *et al.*, **1982**. Cuticle protein genes of *Drosophila*: structure, organization and evolution of four clustered genes. *Cell 29*, 1027–1040.
Stiles, B., **1991**. Cuticle proteins of the boll weevil, *Anthonomus grandis*, abdomen: structural similarities and glycosylation. *Insect Biochem. 21*, 249–258.
Suderman, R.J., Andersen, S.O., Hopkins, T.L., Kanost, M.R., Kramer, K.J., **2003**. Characterization and cDNA cloning of three major proteins from pharate pupal cuticle of *Manduca sexta*. *Insect Biochem. Mol. Biol. 33*, 331–343.
Suetake, T., Tsuda, S., Kawabata, S., Miura, K., Iwanaga, S., *et al.*, **2000**. Chitin-binding proteins in invertebrates and plants comprise a common chitin-binding structural motif. *J. Biol. Chem. 275*, 17929–17932.
Talbo, G., Hojrup, P., Rahbek-Nielsen, H., Andersen, S.O., Roepstorff, P., **1991**. Determination of the covalent structure of an N- and C-terminally blocked glycoprotein from endocuticle of *Locusta migratoria*: combined use of plasma desorption mass spectrometry and Edman degradation to study post-translationally modified proteins. *Eur. J. Biochem. 195*, 495–504.
Tews, I., Scheltiga, T., Perrakis, A., Wilson, K.S., Dijkstra, B.W., **1997**. Substrate-assisted catalysis unifies 2 families of chitinolytic enzymes. *J. Am. Chem. Soc. 119*, 7954–7959.
Trim, A.R.H., **1941**. Studies in the chemistry of insect cuticle. 1. Some general observations on certain arthropod cuticles with special reference to the characterization of the proteins. *Biochem. J. 35*, 1088–1098.
Uchiyama, T., Katouno, F., Nikaidou, N., Nonaka, T., Sugiyama, J., *et al.*, **2001**. Roles of the exposed aromatic residues in crystalline chitin hydrolysis by chitinase A from *Serratia marcescens* 2170. *J. Biol. Chem. 276*, 41343–41349.
Vakser, I.A., **1996**. Low-resolution docking: prediction of complexes for undetermined structures. *Biopolymers 39*, 455–464.
Vincent, J.F.V., **2002**. Arthropod cuticle: a natural composite shell system. *Composites: Part A 33*, 1311–1315.
Vriend, G., **1990**. WHAT IF: A molecular modeling and drug design package. *J. Mol. Graph. 8*, 52–56.
Vyas, N.K., **1991**. Atomic features of protein–carbohydrate interactions. *Curr. Opin. Struct. Biol. 1*, 723–740.

Williams, C.M., **1980**. Growth in insects. In: Locke, M., Smith, D.S. (Eds.), Insect Biology in the Future. Academic Press, London, pp. 369–383.

Willis, J.H., **1986**. The paradigm of stage-specific gene sets in insect metamorphosis: time for revision! *Arch. Insect Biochem. Physiol.* (Suppl. 1), 47–57.

Willis, J.H., **1996**. Metamorphosis of the cuticle, its proteins, and their genes. In: Gilbert, L.I., Atkinson, B.G., Tata, J. (Eds.), Metamorphosis/Post-Embryonic Reprogramming of Gene Expression in Amphibian and Insect Cells. Academic Press, London, pp. 253–282.

Willis, J.H., **1999**. Cuticular proteins in insects and crustaceans. *Amr. Zool. 39*, 600–609.

Willis, J.H., Rezaur, R., Sehnal, F., **1982**. Juvenoids cause some insects to form composite cuticles. *J. Embryol. Exp. Morph. 71*, 25–40.

Wittkopp, P.J., Vaccaro, K., Carroll, S.B., **2002**. Evolution of *yellow* gene regulation and pigmentation in *Drosophila*. *Curr. Biol. 12*, 1547–1556.

Wolfgang, W.J., Fristrom, D., Fristrom, J.W., **1986**. The pupal cuticle of *Drosophila*: differential ultrastructural immunolocalization of cuticle proteins. *J. Cell Biol. 102*, 306–311.

Wolfgang, W.J., Fristrom, D., Fristrom, J.W., **1987**. An assembly zone antigen of the insect cuticle. *Tissue and Cell 19*, 827–838.

Wolfgang, W.J., Riddiford, L.M., **1986**. Larval cuticular morphogenesis in the tobacco hornworm, *Manduca sexta*, and its hormonal regulation. *Devel. Biol. 113*, 305–316.

Wybrandt, G.B., Andersen, S.O., **2001**. Purification and sequence determination of a yellow protein from sexually mature males of the desert locust, *Schistocerca gregaria*. *Insect Biochem. Mol. Biol. 31*, 1183–1189.

Zanotti, G., Marcello, M., Malpeli, G., Folli, C., Sartori, G., *et al.*, **1994**. Crystallographic studies on complexes between retinoids and plasma retinol-binding protein. *J. Biol. Chem. 269*, 29613–29620.

Zhou, X., Riddiford, L.M., **2002**. Broad specifies pupal development and mediates the "status quo" action of juvenile hormone on the pupal–adult transformation in *Drosophila* and *Manduca*. *Development 129*, 2259–2269.

Relevant Website

http://bioinformatics2.biol.uoa.gr – A relational database of arthropod cuticular proteins established by C.K. Magkrioti, I.C. Spyropoulos, V.A. Iconomidou, J.H. Willis, and S.J. Hamodrakas.

4.3 Chitin Metabolism in Insects

K J Kramer, US Department of Agriculture, Manhattan, KS, USA
S Muthukrishnan, Kansas State University, Manhattan, KS, USA

© 2005, Elsevier BV. All Rights Reserved.

4.3.1. Introduction

"Chitin Metabolism in Insects" was the title of one of the chapters in the original edition of *Comprehensive Insect Physiology, Biochemistry, and Pharmacology* series published in 1985 (Kramer *et al*., 1985). Since that time substantial progress in gaining an understanding of this topic has occurred, primarily through the application of techniques of molecular genetics and biotechnology to assorted studies on insect chitin metabolism. Several other reviews have also been published, which have reported on some of the advances that have taken place (Kramer and Koga, 1986; Cohen, 1987, 2001; Koga *et al*., 1999; Fukamizo, 2000). Thus, in this chapter we will highlight some of the more important findings since 1985, with an emphasis on results obtained from studies conducted on the two enzymes primarily responsible for chitin synthesis and degradation, namely chitin synthase (CHS) and chitinase (CHI).

4.3.2. Chitin Structure and Occurrence

Chitin is widely distributed in animals and represents the skeletal polysaccharide of several phyla such as the Arthropoda, Annelida, Mollusca, and Coelenterata. In several groups of fungi, chitin replaces cellulose as the structural polysaccharide. In insects, it is found in the body wall, gut lining, cuticle, salivary glands, trachea, mouth parts, and muscle attachment points.

In the course of evolution, insects have made excellent use of the rigidity and chemical stability of the polymeric chitin to assemble extracellular structures such as the cuticle (exoskeleton) and gut lining (peritrophic membrane (PM)), both of which enable insects to be protected from the environment while allowing growth, mobility, respiration, and communication. Several genes and gene products are involved in chitin metabolism in insects. In general there are two primary extracellular structures in

which chitin deposition occurs. Those are the cuticle and the PM where both synthesis and degradation of chitin take place at different developmental stages.

Chitin is the major polysaccharide present in insects and many other invertebrates and several microbes. Structurally, it is the simplest of the glycosaminoglycans, being a β (1→4) linked linear homopolymer of *N*-acetylglucosamine (GlcNAc, $(C_8H_{13}O_5N)_{n\gg1}$). It is usually synthesized as the old endocuticle and PM are resorbed and the digested materials are recycled. Because of the intractable nature of insect sclerotized structures such as cuticle, there was very little quantitative data available about chemical composition until recently when solid-state nuclear magnetic resonance (NMR) was utilized for analyses. The cuticle and PM are composed primarily of a mixture of protein and chitin, with the former usually predominating (Kramer *et al*., 1995). Chitin contents vary substantially depending on the type of cuticle. For example, in the sclerotized puparial cuticle from the housefly, *Musca domestica*, the chitin content is approximately 45% of the wet weight, whereas in the mineralized puparial cuticle of the face fly, *Musca autumnalis*, the chitin content is only about 19% (Roseland *et al*., 1985; Kramer *et al*., 1988). In larval, pupal, and adult cuticles of the tobacco hornworm, *Manduca sexta*, the chitin content is approximately 14%, 25%, and 7%, respectively (Kramer *et al*., 1995). In newly ecdysed pupal cuticle, there is only about 2% chitin prior to sclerotization, but that amount increases more than 10-fold after sclerotization. When cuticular protein and chitin are mixed, they form a matrix in which the components of lower abundance, such as water, catechols, lipids, and minerals, are interspersed. The PM of the tobacco hornworm is made up primarily of protein (60%) and chitin (40%) (Kramer *et al*., 1995). Although primarily composed of poly-GlcNAc, chitin also can contain a small percentage of unsubstituted (or *N*-deacetylated) glucosamine (GlcN) residues (Fukamizo *et al*., 1986). When the epidermal and gut cells synthesize and secrete a particular form of chitin consisting of antiparallel chains, α-chitin, the chains are formed into sheets. As layers are added, the sheets become cross-oriented to one another, which can contribute to the formation of an extremely strong plywood-like material. The origin of proteins in the cuticle is unknown, but some hemolymph proteins are deposited in cuticle. Thus, apparently the epidermal cells do not need to supply all of the component parts of the exoskeleton. The cells lining the gut produce some of the PM-associated proteins and these proteins are referred to as the peritrophins (Tellam *et al*., 1999; Wang and Granados, 2000a; Bolognesi *et al*., 2001; Eisemann *et al*., 2001). Analysis of expressed sequence tags in the cat flea, *Ctenocephalides felis*, demonstrated that some peritrophins are produced exclusively by hindgut and Malpighian tubule tissues (Gaines *et al*., 2002).

The last step in cuticle formation, tanning, involves modification of the free amino acid tyrosine that is sequestered as a conjugate with glucose in the hemolymph. Tyrosine is first hydroxylated to 3,4-dihydroxyphenylalanine (DOPA), and then decarboxylated to 3,4-dihydroxyphenethylamine (dopamine) (Hopkins and Kramer, 1992). Dopamine is *N*-acylated with acetate or β-alanine in the epidermal cells and sequestered in the hemolymph as conjugates with glucose, sulfate, or another hydrophilic compound. The *N*-acylated dopamine conjugates then are delivered through pore canals to the epicuticle where the conjugates are hydrolyzed and then converted by phenoloxidases to very highly reactive quinones and quinone methides. These transient compounds then cross-link proteins to form tanned proteins in a process known as sclerotization (see **Chapter 4.4**). These cross-linked proteins and chitin make up most of the exocuticle. Chitin chains also may become cross-linked with cuticular proteins, but the evidence for that is not definitive.

Chitin oligosaccharides that are produced during degradation of chitin by chitinases appear to play an important role in insect immunity towards microorganisms. The basic immune strategy against microbial infection in insects appears to be similar to the strategy used by plants against fungal infection. These oligosaccharides are known to activate chitinase genes in plants, which are actively involved in the plant defense response against fungal infection (Nichols *et al*., 1980). In the silkworm, *Bombyx mori*, chitin oligomers trigger expression of three different antibacterial proteins – cecropin, attacin, and lebocin – in the fat body and hemocytes (Furukawa *et al*., 1999).

4.3.3. Chitin Synthesis

Relatively little additional biochemical data on the enzymes of the chitin biosynthetic pathway have been generated since the previous review was published (Kramer *et al*., 1985). The paucity of information concerning the biochemical properties of these enzymes is due to the inability to obtain soluble preparations of CHSs and the instability of the glutamine-fructose-6-phosphate aminotransferase (GFAT), the enzyme that provides

the GlcN precursor of the chitin biosynthetic pathway. However, CHSs have been identified in a variety of organisms, including nematodes, fungi, and insects. Amino acid sequence similarities have been the principal tools used for identifying CHSs, which form a subfamily within a larger group (family GT2) of the glycosyltransferases that catalyze the transfer of a sugar moiety from an activated sugar donor onto saccharide or nonsaccharide acceptors (Coutinho and Henrissat, 1999; Coutinho *et al*., 2003; CAZY, 2004). During the past 3 years, there has been a sudden increase in research in the area of chitin synthesis. The impetus for this enhanced interest has come predominantly from cloning of genes for the two key enzymes of the pathway, GFAT and CHS, from insects.

CHS has not been an easy enzyme to assay, which has made its study rather difficult. Traditionally, CHS activity was measured by a radioactive assay using [^{14}C]UDP-GlcNAc as the substrate followed by quantification of insoluble ^{14}C-labeled chitin after acid precipitation. Recently, however, a high throughput nonradioactive assay has been developed (Lucero *et al*., 2002). The procedure involves binding of synthesized chitin to a wheat germ agglutinin (WGA)-coated surface followed by detection of the polymer with a horseradish peroxidase–WGA conjugate. This nonradioactive assay should facilitate greater progress in CHS studies in the future.

4.3.3.1. Precursors of the Chitin Biosynthetic Pathway

Early studies on chitin synthesis using whole insects or isolated tissues demonstrated that in addition to whole animals, a variety of tissues including larval and pupal epidermis, abdomen, integument, gut, imaginal discs, leg regenerates, hypodermis, and oocytes were capable of synthesizing chitin (review: Kramer *et al*., 1985). An assortment of compounds, including glycogen, glucose, glucosamine, fructose, and GlcNAc could serve as biosynthetic precursors of chitin in these tissues. These early studies also identified several compounds that inhibited the pathway. This list includes substrate analogs such as tunicamycin, polyoxin D, nikkomycin, and uridine diphosphate (UDP), as well as several compounds belonging to the benzoylphenylurea class of insect growth regulators whose exact mode of action has not yet been established. Results of these studies also indicated that ecdysone may influence chitin synthesis either directly or indirectly. However, the details of such a regulation remain unclear.

4.3.3.2. Sites of Chitin Biosynthesis

The epidermis and the midgut are two major tissues where chitin synthesis occurs in insects. Epidermal cells are responsible for the deposition of new cuticle during each molt and the midgut cells are generally associated with the formation of the PM during feeding. Chitin is associated with other tissues as well, including the foregut, hindgut, trachea, wing hinges, salivary gland, and mouth parts of adults and/or larvae (Wilson and Cryan, 1997). In general, it is assumed that the cells closest to the site where chitin is found are responsible for its biosynthesis. However, this interpretation is complicated by the fact that assembly of chitin microfibrils occurs in the extracellular space and is influenced by the presence or absence of associated proteins. This is particularly true in the gut where some cells around the cardia may be contributing to chitin synthesis and secretion, whereas other cells in different parts of the gut may be responsible for synthesis of PM-associated proteins (Wang and Granados, 2000a). Visible PM may appear at sites remote from the original site of synthesis of either chitin or PM proteins.

4.3.3.3. Light and Electron Microscopic Studies of Peritrophic Membrane Synthesis

The most detailed picture of chitin synthesis and its association with proteins to form the composite PM has emerged from observations using light microscopy as well as transmission and scanning electron microscopy (SEM) of PM synthesis in the three lepidopteran insects, *Ostrinia nubilalis* (European corn borer), *Trichoplusia ni* (cabbage looper), and *M. sexta* (Harper and Hopkins, 1997; Harper *et al*., 1998; Harper and Granados, 1999; Wang and Granados, 2000a; Hopkins and Harper, 2001). The presence of chitin in nascent PM can be followed by staining with gold-labeled WGA, which binds to GlcNAc residues in chitin and glycoproteins. This method was used to show that chitin-containing fibrous material appears first at the tips of the microvilli of the midgut epithelial cells of *O. nubilalis* just past the stomadeal valves and is rapidly assimilated into a thin PM surrounding the food bolus (Harper and Hopkins, 1997). The PM becomes thicker and multilayered in the middle and posterior regions of the mesenteron. The orthogonal lattice of chitin meshwork is slightly larger than the diameter of the microvilli. SEM and light microscopic studies revealed that the PM delaminates from the tips of the microvilli. This observation suggests that microvilli serve as sites and possibly as templates for the organization of the PM by laying down a matrix of chitin microfibrils onto

which some PM proteins are deposited. A similar pattern of delamination of PM containing both chitin and intestinal mucins was demonstrated in larvae of *T. ni* (Harper and Granados, 1999; Wang and Granados, 2000a).

Incorporating WGA into the diet can interrupt formation of the PM. WGA-fed *O. nubilalis* larvae had an unorganized PM, which was multilayered and thicker than the normal PM (Hopkins and Harper, 2001). WGA was actually associated with the PM as well as with the microvillar surface as revealed by immunostaining with antibodies specific for WGA. Because there was very little WGA within the epithelial cells, the action of WGA appears to be extracellular. Presumably, WGA interferes with the formation of the organized chitin network and/or the association of PM proteins with the chitin network, leading to a reduced protein association with the PM (Harper *et al.*, 1998). There was also extensive disintegration of the microvilli and the appearance of dark inclusion bodies as well as apparent microvillar fragments within the thickened multilayered PM. Insects such as *M. sexta*, which secrete multiple and thickened PMs that are somewhat randomly organized, tolerated WGA better and sequestered large amounts of WGA within the multilayered PM (Hopkins and Harper, 2001).

4.3.3.4. *In Situ* Hybridization and Immunological Studies

In situ hybridizations with a DNA probe for the catalytic domain of a CHS revealed that high levels of transcripts for this gene are present in apical regions of the columnar cells of the anterior midgut of *M. sexta* larvae (Zimoch and Merzendorfer, 2002). Lesser amounts of CHS transcripts were detected in the posterior midgut. An antibody to the catalytic domain of *M. sexta* CHS also detected the enzyme in midgut brush border membranes at the extreme apical ends of microvilli, suggestive of some special compartment or possibly apical membrane-associated vesicles. Staining was also seen in apical membranes of tracheal and salivary gland cells. Materials reacting with CHS antibody also were detected underneath the epidermal cuticle, even though it could not be specifically assigned to the apical membrane of epidermal cells due to loss of structural integrity of these cells during cryosectioning. These *in situ* hybridization and immunochemical studies are in agreement with earlier observations about chitin synthesis in *Calpodes ethlius* (larger canna leafroller), which indicated the involvement of specialized structures called plasma membrane plaques found in apical portions of epidermal cells (Locke and Huie, 1979). Comparable electron microscope (EM) and immunological localization of CHS associated with epidermis during cuticle deposition have not been reported primarily because of technical difficulties with the handling of cuticular samples. In *Drosophila melanogaster* the chitin synthase gene (*kkv*) is expressed predominantly in developmental stages 13-14 in the embryonic ventral and dorsal epidermis, foregut and in the larval tracheal system (see the "Patterns of gene expression in *Drosophila* embryogenesis" at the Berkeley Drosophila Genome Project (BDGP)).

4.3.3.5. Chitin Biosynthetic Pathway

It has been assumed that the pathway of chitin biosynthesis in insects would be similar or identical to the pathway that has been worked out extensively in fungi and other microbes (**Figure 1**). This appears to be the case except for some minor details (Palli and Retnakaran, 1999). The source of the sugar residues for chitin synthesis can be traced to fat body glycogen, which is acted upon by glycogen phosphorylase. Glucose-1-P produced by this reaction is converted to trehalose, which is released into the hemolymph. Trehalose, the extracellular source of sugar in many insects, is acted upon by a trehalase, which is widely distributed in insect tissues including the epidermis and gut to yield intracellular glucose (Becker *et al.*, 1996). The conversion of glucose to fructose-6-P needed for chitin synthesis involves two glycolytic enzymes present in the cytosol. These enzymes are hexokinase and glucose-6-P isomerase, which convert glucose to fructose-6-P. From the latter, the chitin biosynthetic pathway branches off, with the first enzyme catalyzing this branch being GFAT, which might be thought of as the first committed step in amino sugar biosynthesis. The conversion of fructose-6-P to GlcNAc phosphate involves amination, acetyl transfer, and an isomerization step, which moves the phosphate from C-6 to C-1 (phosphoacetylglucoasmine mutase). The conversion of this compound to the nucleotide sugar derivative follows the standard pathway and leads to the formation of a UDP-derivative of GlcNAc, which serves as the substrate for CHS. The entire chitin biosynthetic pathway is outlined in **Figure 1**.

The involvement of dolichol-linked GlcNAc as a precursor for chitin was proposed quite some time ago (Horst, 1983), but it has received very limited experimental support (Quesada-Allue, 1982). At this point, this possibility remains unproven. Similarly, the requirement for a primer to which the

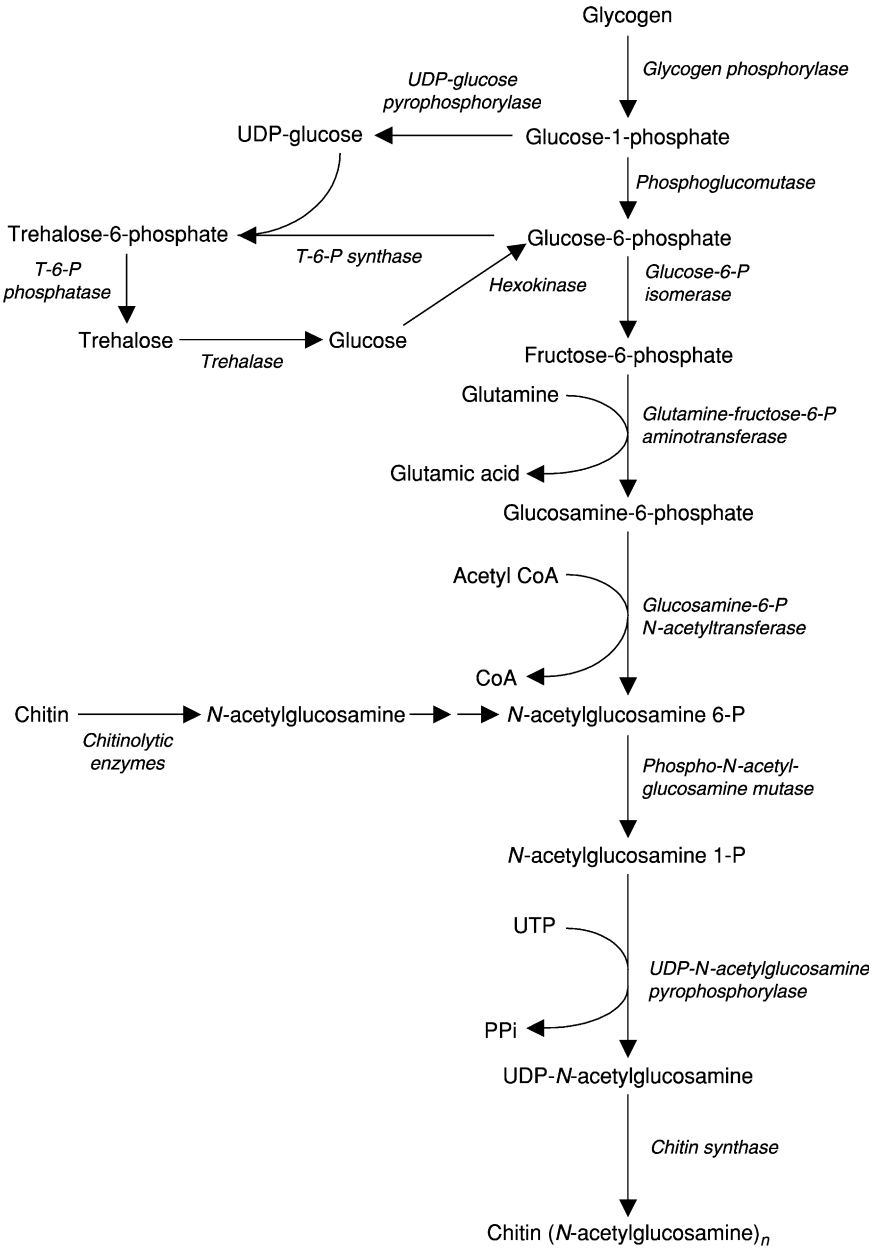


Figure 1 Biosynthetic pathway for chitin in insects starting from glycogen, trehalose, and recycled chitin.

GlcNAc residues can be transferred also remains speculative. Based on the model for glycogen biosynthesis, which requires glycogenin as the primer (Gibbons *et al*., 2002), CHS or an associated protein may fulfill this priming function. Because each sugar residue in chitin is rotated ~180° relative to the preceding sugar, which requires CHS to accommodate a alternating "up/down" configuration, another precursor, UDP-chitobiose, has been proposed to be a disaccharide donor during biosynthesis (Chang *et al*., 2003). However, evaluation of radiolabeled UDP-chitobiose as a CHS substrate in yeast revealed that it was not a viable one. Even at elevated concentrations, no incorporation of radioactivity above background was observed using membranous preparations of CHS from the yeast *Saccharomyces cerevisiae* (Chang *et al*., 2003).

4.3.3.5.1. Key enzymes The biosynthetic pathway of chitin can be thought of as consisting of two segments. The first set of reactions leads to the formation of the amino sugar, GlcNAc, and the second set of reactions leads to the synthesis of the polymeric chitin from the amino sugar. The

rate-limiting enzyme in the first segment appears to be GFAT (also known as glucosamine-fructose-6-phosphate aminotransferase (GFAT, EC 2.6.1.16), which is found in the cytosol. The critical enzyme in the second segment is CHS (EC 2.4.1.16), which is localized in the plasma membrane. Not surprisingly, these two enzymes appear to be major sites of regulation of chitin synthesis.

4.3.3.5.2. Regulation of glutamine-fructose-6-phosphate aminotransferase synthesis

4.3.3.5.2.1. Drosophila GFAT Two genes encoding GFAT (*Gfat1* and *Gfat2*) have been identified in *Drosophila* (Adams *et al*., 2000; Graack *et al*., 2001). Both of these genes are on chromosome 3, but they are at different locations. Their intron–exon organizations are different as are the amino acid sequences of the encoded proteins. GFAT consists of two separate domains, an N-terminal domain that has both glutamine binding and aminotransferase motifs identified in GFATs from other sources and a C-terminal domain with both fructose-6-phosphate binding and isomerase motifs. *Gfat1* is expressed in embryos in the developing trachea and in cuticle-forming tissues including the chitinous mouth armature of the developing first instar larva. In the last larval stadium, *Gfat1* is expressed in the corpus cells of salivary glands, but this synthesis may be related to the production of the highly glycosylated Sgs glue proteins (Graack *et al*., 2001). The major regulation of GFAT1 appears to be posttranslational. When *Gfat1* was expressed in yeast cells, the resulting enzyme was feedback inhibited by UDP-GlcNAc and was stimulated by protein kinase A. Even though it has not been demonstrated that there is a phosphorylated form of GFAT1 that is susceptible to feedback inhibition by UDP-GlcNAc, this possibility remains viable. The expression and regulation of the other GFAT isozyme (GFAT2) has not yet been reported.

4.3.3.5.2.2. Aedes aegypti GFAT The gene and cDNA for the mosquito *Aedes aegypti* GFAT1 have been cloned (Kato *et al*., 2002). The mosquito gene has no introns and the promoter appears to contain sequences related to ecdysteroid response elements (EcRE) as well as E74 and Broad complex Z4 elements. E74 and Broad complex Z4 proteins are transcription factors known to be upregulated by ecdysone (Thummel, 1996). Two *Gfat1* transcripts with different sizes were observed in Northern blot analyses of RNA from adult females and their levels increased further after blood-feeding (Kato *et al*., 2002). Since ecdysteroid titers increase following blood-feeding, it is possible that this gene is under the control of ecdysteroid either directly or indirectly. Feedback inhibition by UDP-GlcNAc has not been reported, but the *Aedes* enzyme is likely to be regulated in a manner similar to the *Drosophila* enzyme by this effector and possibly by a phosphorylation/dephosphorylation mechanism as well.

4.3.3.5.3. CHS gene number and organization

CHS genes from numerous fungi have been isolated and characterized (Munrow and Gow, 2001). However, the complete sequence of a cDNA clone for an insect CHS (sheep blowfly, *Lucilia cuprina*) was reported only recently (Tellam *et al*., 2000). Since then, the sequences of several other full-length cDNAs and genes for CHSs from other insects and nematodes have been reported. The nematode CHSs were from two filarial pathogens, *Brugia malayi*, and *Dirofilaria immitis*, and the plant parasite *Meloidogyne artiellia* (Harris *et al*., 2000; Veronico *et al*., 2001; Harris and Fuhrman, 2002). The other insect species from which CHS cDNAs have been isolated are *A. aegypti* (Ibrahim *et al*., 2000), *M. sexta* (Zhu *et al*., 2002) and the red flour beetle, *Tribolium castaneum* (Arakane *et al*., 2004). DNA sequencing of polymerase chain reaction (PCR)-amplified fragments encoding a highly conserved region in the catalytic domains of insect CHSs indicates a high degree of sequence conservation (Tellam *et al*., 2000). In addition, a search of the databases in light of the sequence data from these cDNAs has allowed identification of open reading frames (ORFs) from CHS genes from *Drosophila*, *Anopheles, Aedes* and the nematode *Caenorhabditis elegans* (Tellam *et al*., 2000; Gagou *et al*., 2002; Arakane *et al*., 2004). **Table 1** lists the properties of insect CHSs encoded by these genes/cDNAs. Insect species typically have two genes for CHSs. Among the nematodes, the *C. elegans* genome contains two CHS genes, but so far there is evidence for only one gene in the plant parasitic nematode *M. artiellia*, and in the filarial nematodes *B. malayi* and *D. immitis* (Harris *et al*., 2000; Veronico *et al*., 2001; Harris and Fuhrman, 2002). Fungi, on the other hand, exhibit a wide range in the number of genes for CHS (Munrow and Gow, 2001).

The two *Tribolium* CHS genes, *TcCHS1* and *TcCHS2*, have ten and eight exons, respectively (Arakane *et al*., 2004). The organizations of the two genes in *Tribolium* are quite different, with some introns occurring in identical positions in both genes, whereas others are at variable positions. The introns ranged in length from 46 bp to more than 3000 bp. The most interesting difference between the two genes was the presence of two

Table 1 Properties of insect chitin synthases and their genes

Species	Number of amino acids	Expressed in	Alt. Exon	Coiled-coil	CHS class	GI no.	Reference
Lucilia cuprina	1592	Epidermis	Yes[a]	Yes	A	9963823	Tellam *et al.* (2000)
Drosophila melanogaster	1615	Epidermis/gut/tracheal	Yes	Yes	A	24644218	Adams *et al.* (2000); Fly base – http://www.flybase.bio.indiana.edu; Berkeley Drosophila genome project (Drosophila EST database) – http://www.fruitfly.org
	1674		Yes	Yes	A	24644220	
	1416	ND	No	No	B	24668460	
Anopheles gambiae	1578		Yes	Yes	A		
	1583		No	No	B		
Aedes aegypti	1564	Midgut	No	No	B	22773456	Ibrahim *et al.* (2000)
Tribolium castaneum	1558	ND	Yes	Yes	A		
	1558	ND	Yes	Yes	A		Arakane *et al.* (2004)
	1464	ND	No	No	B		
Manduca sexta	1563	Epidermis/gut	Yes	Yes	A	24762312	Zhu *et al.* (2002)
	1563	Epidermis/gut	Yes	Yes	A		H. Merzendorfer (unpublished data)
	1524	Gut	No	No	B		D. Hogenkamp *et al.* (unpublished data)

[a]Predicted.
ND, not determined.

nonidentical copies of exon 8 (named 8a and 8b) in *TcCHS1*, whereas *TcCHS2* has only one copy of this region as a part of exon 6. An analysis of genomic sequences from the *D. melanogaster* and *Anopheles gambiae* genome projects, partial sequencing of cDNAs available as separate sequence files submitted to GenBank, and "TBLASTN" queries were used to determine the organization of CHS genes in these insects (**Figure 2**). These analyses revealed that the sequences and organization of CHS genes of *D. melanogaster* (Tellam *et al.*, 2000) and *A. gambiae* were similar to those of *TcCHS1* and *TcCHS2* (Arakane *et al.*, 2004). One major difference between the two exons that are alternately spliced is that all of the B forms code for segments that have a site for *N*-linked glycosylation just before the transmembrane helix, whereas none of the A forms do. The physiological significance of alternate exon usage and potential glycosylation in *CHS* expression is unknown even though it is clear that there is developmental regulation of alternate exon usage (see Section 4.3.3.5.6).

4.3.3.5.4. Modular structure of chitin synthases CHSs are members of family GT2 of the glycosyltransferases (Coutinho *et al.*, 2003), which generally utilize a mechanism where inversion of the anomeric configuration of the sugar donor occurs. The protein fold (termed GT-A) for this family is considered to be two associated β/α/β domains that form a continuous central sheet of at least eight β-strands. The GT-A enzymes share a common ribose/metal ion-coordinating motif (termed DxD motif) as well as another carboxylate residue that acts as a catalytic base. The general organization of CHSs has been deduced from a comparison of amino acid sequences of these enzymes from several insects, nematodes and yeasts (Zhu *et al.*, 2002; Arakane *et al.*, 2004). These enzymes have three distinguishable domains: an N-terminal domain with moderate sequence conservation among different species and containing several transmembrane segments, a middle catalytic domain that is highly conserved even among CHSs from different kingdoms, and a C-terminal module with multiple transmembrane segments (**Figure 3**). The catalytic domain contains several stretches of highly conserved amino acid sequences including the following: CATMWHXT at the beginning of the catalytic domain, FEYAIGHW and VQYDDQGEDRW in the middle of the catalytic domain, and the presumed catalytic site, EFYNQRRRW, at the end of the catalytic domain. While the transmembrane segments in the N-terminal domain show different patterns among different insect species, the transmembrane segments in the C-terminal domain are remarkably conserved both with respect to their location and the spacing between adjacent transmembrane segments. Particularly striking is the fact that five such transmembrane segments are found in a cluster immediately following the catalytic domain and two more segments are located closer to the

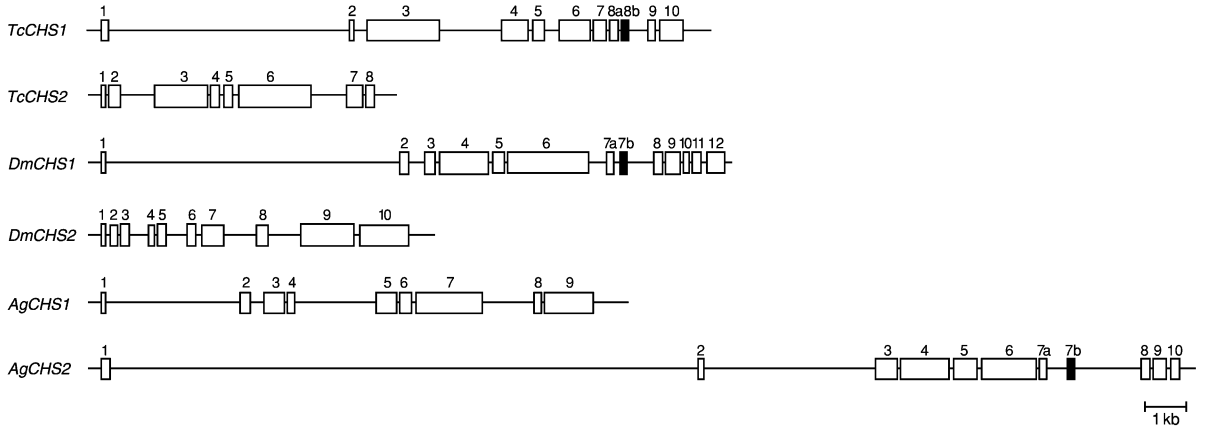


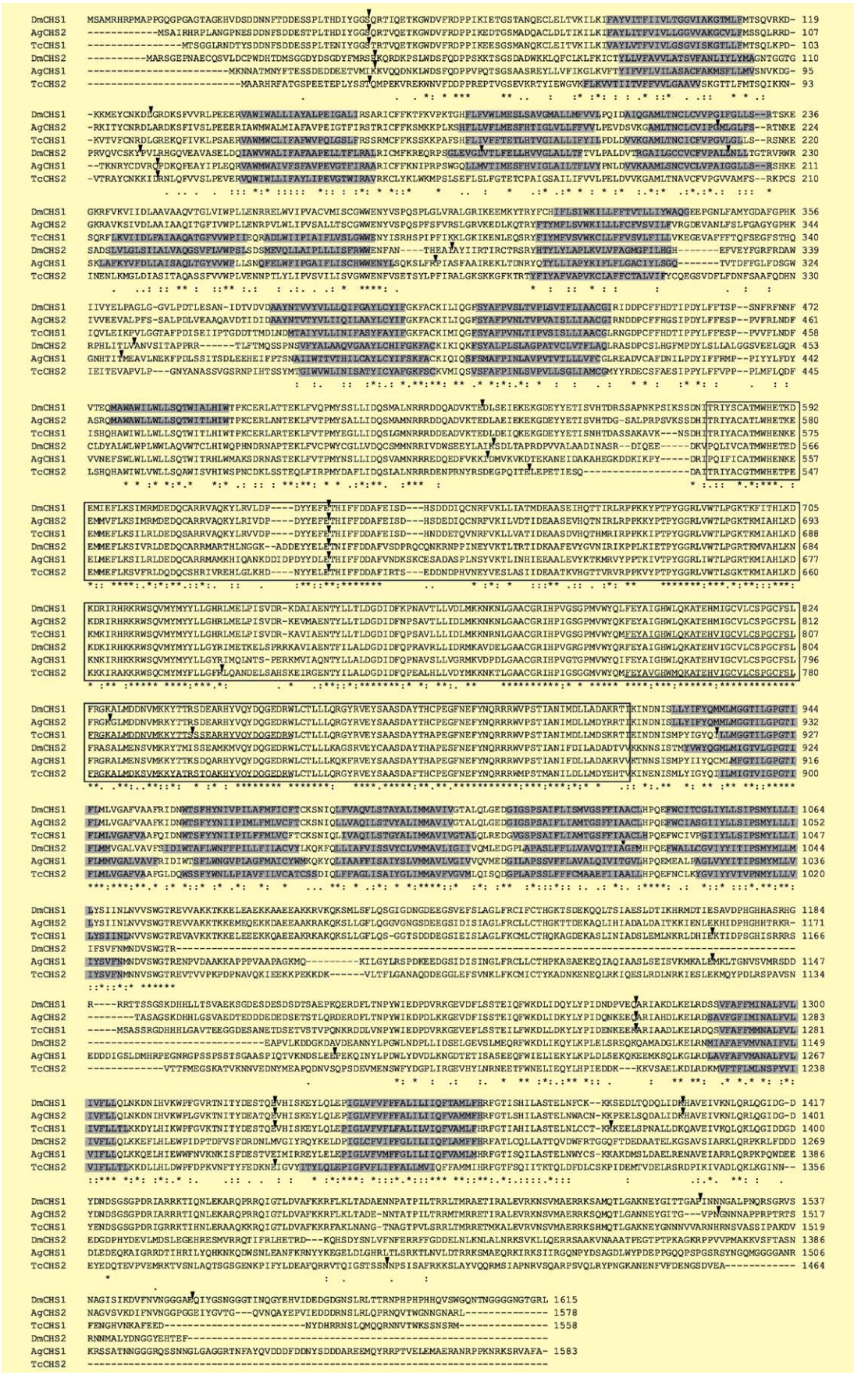
Figure 2 Schematic diagram of the organization of the *TcCHS1*, *TcCHS2*, *DmCHS1, DmCHS2, AgCHS1*, and *AgCHS2* genes. Boxes indicate exons. Lines indicate introns. The second of the two alternative exons (8b) of *TcCHS1*, *DmCHS1 (7b),* and *AgCHS2 (6b)* are indicated as closed boxes. About 9 kb of the *TcCHS1* and *TcCHS2* gDNA sequences were compared to their respective cDNA sequences to define the exons and introns. The exon–intron organization of the other four CHS genes was deduced partially from comparisons of available cDNA and genomic sequences. (Reprinted with permission from Arakane, Y., Hogenkamp, D., Zhu, Y.C., Kramer, K.J., Specht, C.A., *et al.*, **2004**. Chitin synthase genes of the red flour beetle, *Tribolium castaneum*: characterization, expression, linkage mapping and alternate exon usage. *Insect Biochem. Mol. Biol. 34*, 291–304.)

C-terminus. The 5-transmembrane cluster, known as 5-TMS, has been suggested to be involved in the extrusion of the polymerized chitin chains across the plasma membrane to the exterior of the cell as proposed for extrusion of cellulose (Richmond, 2000).

The CHSs of insects characterized so far can be broadly grouped into two classes, A and B, based on amino acid sequence identities. The class A proteins were predicted to have a coiled-coil region immediately following the 5-TMS region (Zhu *et al.*, 2002; Arakane *et al.*, 2004). Also, all of the genes encoding the class A CHSs have two alternate exons (corresponding to alternate exon 7 of *D. melanogaster*, exon 8 of *T. castaneum*, exon 6 of *A. gambiae*, and an unnumbered exon of *M. sexta* CHS-A gene) (see **Table 1**). The alternate exons are located on the C-terminal side of the 5-TMS region and encode the next transmembrane segment and flanking sequences. The alternate exon-encoded regions of the CHS proteins differ in sequence by as much as 30% and most of these differences are in the regions flanking the transmembrane segment. This finding suggests that the proteins may differ in their ability to interact with cytosolic or extracellular proteins, which might regulate chitin synthesis and/or transport. An attractive hypothesis is that these flanking sequences may influence the plasma membrane location of a CHS by interacting with cytoskeletal elements or perhaps by generation of extracellular vesicles involved in chitin assembly.

4.3.3.5.5. Regulation of chitin synthase gene expression The two insect genes encoding CHSs appear to have different patterns of expression during development. The high degree of sequence identity of the catalytic domains and the absence of antibodies capable of discriminating between the two isoforms have complicated the interpretation of experimental data to some extent. In some cases, the technical difficulties associated with isolation of specific tissues free of other contaminating tissues have precluded unambiguous assignment of tissue specificity of expression. Nonetheless, the following conclusions can be reached from the analyses of expression of CHS genes in several insect species. CHS genes are expressed at all stages of insect growth including embryonic, larval, pupal, and adult stages. *CHS1* genes (coding for class A CHS proteins) are expressed over a wider range of developmental stages (Tellam *et al.*, 2000; Zhu *et al.*,

Figure 3 Alignment of deduced amino acid sequences of TcCHS1, TcCHS2, DmCHS1, DmCHS2, AgCHS1, and AgCHS2 using ClustalW software. Transmembrane regions predicted using TMHMM software (v. 2.0) are shaded. Shaded arrowheads indicate the positions in the protein sequences of TcCHS1 and TcCHS2 where coding regions are interrupted by introns. Intron 1 of *TcCHS1* lies in the 5′-UTR region two nucleotides 5′ of the translation start site and is not indicated in this figure. The putative catalytic domains are boxed. Symbols below the aligned amino acid sequences indicate identical (*), highly conserved (:), and conserved residues (.). The regions in TcCHS1 and TcCHS2 corresponding to the PCR probe made from two degenerate primers representing two highly conserved sequences in CHSs are underlined. (Reprinted with permission from Arakane, Y., Hogenkamp, D., Zhu, Y.C., Kramer, K.J., Specht, C.A., *et al.*, **2004**. Chitin synthase genes of the red flour beetle, *Tribolium castaneum*: characterization, expression, linkage mapping and alternate exon usage. *Insect Biochem. Mol. Biol. 34*, 291–304.)


DmCHS1
AgCHS2
TcCHS1
DmCHS2
AgCHS1
TcCHS2

2002). *CHS2* genes (coding for class B CHSs) are not expressed in the embryonic or pupal stages but are expressed in the larval stages, especially during feeding in the last instar and in the adults including blood-fed mosquitoes (Ibrahim *et al*., 2000; Zimoch and Merzendorfer, 2002; Arakane *et al*., 2004). The finding that both classes of CHS genes are expressed at high levels 3 h after pupariation in *Drosophila* suggests that both enzymes are required for postpuparial development (Gagou *et al*., 2002).

CHS genes also show tissue-specific expression patterns. In *L. cuprina*, *CHS1* (coding for a class A CHS) is expressed only in the carcass (larva minus internal tissues) and trachea but not in salivary gland, crop, cardia, midgut or hindgut (Tellam *et al*., 2000). In blood-fed female mosquitoes, a CHS gene encoding a class B CHS is expressed in the epithelial cells of the midgut (Ibrahim *et al*., 2000). In *M. sexta*, *CHS1* (coding for a class A CHS) is expressed in the epidermal cells of larvae and pupae (Zhu *et al*., 2002). Transcripts specific for class B CHS were detected only in the gut tissue (D. Hogenkamp *et al*., unpublished data). As discussed above, in *Drosophila,* both classes of CHS genes were shown to be upregulated after the ecdysone pulse had ceased in the last larval instar, but the tissue specificity of expression of each gene was not determined. In *T. castaneum*, the *CHS1* gene (coding for a class A CHS) was expressed in embryos, larvae and pupae, and in young adults, but not in mature adults (Arakane *et al*., 2004). Even though unequivocal data are not available for each of these insect species, the following generalizations may be made. Class A CHS proteins are synthesized by epidermal cells when cuticle deposition occurs in embryos, larvae, pupae, and young adults, whereas the class B CHS proteins are expressed by the midgut columnar epithelial cells facing the gut lumen in the larval and adult stages and is probably limited to feeding stages.

4.3.3.5.6. Developmental control of alternate exon usage Insect CHS genes characterized so far have eight or more exons. The genes encoding *Drosophila*, *Anopheles*, *Tribolium*, and *Manduca* class A CHSs, but not the genes encoding class B CHSs, have two alternate exons, each encoding a 59 amino acid long segment following the 5-TMS region (**Table 1**). This segment contains a 20 amino acid long transmembrane region and flanking sequences. In addition, the presence of a predicted coiled-coil region immediately following the 5-TMS region in the CHSs encoded by those genes that have the alternate exons suggests a link between these two structural features and the possibility of regulation of alternate exon usage. In agreement with this idea, transcripts containing either one of these exons have been detected in *T. castaneum* and *M. sexta* (Arakane *et al*., 2004; D. Hogenkamp *et al*., unpublished data). In *T. castaneum* embryos, transcripts with either exon 8a or 8b were detected, whereas in last instar larvae and prepupae, only exon 8a transcripts were present. By the pupal stage, however, transcripts with exon 8a or exon 8b were abundant along with trace amounts of a transcript with both exons. In mature adults, none of these transcripts was detectable, whereas *TcCHS2* transcripts were easily detected especially in females (Arakane *et al*., 2004). In *Drosophila*, transcripts containing either exon 7a or both exons 7a and 7b (but not those containing exon 7b alone) have been reported (Drosophila EST Database).

It appears that the TcCHS1 with the exon 8b-encoded segment is needed during cuticle deposition in the pupal and embryonic stages but not at other stages of development. Similar results were observed with fifth instar *M. sexta* larvae (Hogenkamp *et al*., unpublished data). The biochemical basis for a specific requirement of the TcCHS1 with the exon 8b-encoded segment is unknown.

4.3.3.6. Chitin Synthesis during Development

4.3.3.6.1. Effect of chitin inhibitors Chitin synthesis occurs during embryonic, larval, pupal, and adult stages for cuticle deposition and for production of the PM in larvae and adults. The inhibition of chitin synthesis using chemical inhibitors or by introduction of mutations affects insect development at different developmental stages and to varying degrees. Studies with "chitin inhibitors" have provided some insights concerning the role of chitin in development and its biological function. The use of lufenuron, a member of the class of insecticides known as benzoylphenylureas, has provided substantial information on chitin synthesis during *Drosophila* development (Wilson and Cryan, 1997). The effects of this insect growth regulator were complex and variable depending on the developmental stage and dose at which the insects were exposed to this agent. When newly hatched larvae were reared on a diet containing very low concentrations of lufenuron, the larvae did not die until the second or third instar and usually pupariated even though the pupae were abnormally compressed. Pharate adults either failed to eclose or died shortly after emergence and had deformed legs. The flight ability of the emerged adults was also affected when the larvae were exposed to very low concentrations of lufenuron. First and second instar larvae fed higher concentrations of lufenuron

had normal growth and physical activity for several hours, but the insects died at about the time of the next ecdysis. Third instar larvae fed high concentrations of lufenuron underwent pupariation, but the puparia had an abnormal appearance. The anterior spiracles failed to evert. Thus, insect development is affected by lufenuron at all stages when chitin synthesis occurs. Another aspect of insect development affected by this compound was egg hatching even though oviposition was normal. The embryos completed development but failed to rupture the vitelline membrane. These results indicated that maternally derived lufenuron can affect egg hatching, which requires the use of chitinous mouth parts by the newly ecdysed larvae. The adults showed no mortality and had no flight disability even when fed high levels of lufenuron, indicating that once all chitin-containing structures had been formed, this "chitin inhibitor" had very little effect on adult morphology and function. However, the benzoylphenylureas may not be affecting CHS activity directly because diflubenzuron did not inhibit incorporation of UDP-GlcNAc into chitin microfibrils in an *in vitro* assay using a microsomal preparation from *T. castaneum* (Cohen and Casida, 1980). It is more likely that the benzoylphenylurea class of insecticides interferes with a step in the assembly of the cuticle and/or PM rather than chitin synthesis *per se*.

4.3.3.6.2. Genetics of chitin synthesis Several *Drosophila* genes involved in controlling cuticle morphology have been characterized (Jurgens *et al*., 1984; Nusslein-Volhard *et al*., 1984; Wiechaus *et al*., 1984; Ostrowski *et al*., 2002). These genes are *krotzkopf verkehrt* (*kkv*), *knickkopf* (*knk*), *grainy head (grh), retroactive (rtv)*, and *zepellin* (*zep*). All of these mutations result in poor cuticle integrity and reversal of embryo orientation in the egg to varying degrees. The homozygous mutant embryos failed to hatch. When these mutant embryos were mechanically devitellinized, the cuticles became grossly enlarged, yielding the "blimp" phenotype. Ostrowski *et al*. (2002) characterized the *kkv* gene and identified it as a *CHS*-like gene. Interestingly, embryos derived from wild-type females treated with high concentrations of lufenuron displayed a similar "blimp" phenotype when devitellinized, indicating that either genetic or chemical disruption of chitin deposition leads to this phenotype. The *knk* gene codes for a protein with sequence similarity to a protein component of the nuclear spindle matrix and is located on chromosome 3 close to the *kkv* gene near the centromere. The *knk* and *kkv* functions are not additive and *kkv* appears to be epistatic to *knk*, which is expressed at very low levels compared to the *kkv* gene as indicated by mRNA levels. The *knk* and *zep* genes appear to function in the epidermis prior to cuticle deposition because they exacerbate the effect of a heterozygous *shotgun* (*shg*) mutation, which codes for an E-cadherin-like protein. The *shg* gene is recessive, but in a *knk/knk* or *zep/zep* background, the cuticle is fragmented suggesting that the protein products of these genes interact with cadherin to reinforce the cuticle by promoting adhesion of the epithelia. Thus, products of all of the "blimp" class of genes, including *kkv*, control the integrity of the embryonic cuticle. It is also possible that some of these genes, whose functions have not been identified yet, may be involved directly or indirectly in extrusion or polymerization of chitin microfibrils. Alternatively, these proteins may reinforce chitin–chitin or chitin–protein interactions. For example, the *grh* gene encodes a GATA family transcription factor that regulates the expression of a DOPA decarboxylase needed for the production of precursors of cuticular protein cross-linking agents (Bray and Kafatos, 1991). It is also possible that some of these proteins are involved in vesicular trafficking and/or targeting CHS to plasma membrane plaques that are associated with chitin synthesis (Locke and Huie, 1979).

4.3.4. Chitin Degradation

Chitinases are among a group of proteins that insects use to digest the structural polysaccharide in their exoskeletons and gut linings during the molting process (Kramer *et al*., 1985; Kramer and Koga, 1986; Kramer and Muthukrishnan, 1997; Fukamizo, 2000). Chitin is digested in the cuticle and PM to GlcNAc by a binary enzyme system composed of a chitinase (CHI) and a β-*N*-acetylglucosaminidase (Fukamizo and Kramer, 1985; Filho *et al*., 2002). The former enzyme from molting fluid hydrolyzes chitin into oligosaccharides, whereas the latter, which is also found in the molting fluid, further degrades the oligomers to the monomer from the nonreducing end. This system also probably operates in the gut during degradation of chitin in the PM or in digestion of chitin-containing prey.

Chitinase (EC 3.2.1.14, endochitinase) is defined as an enzyme that catalyzes the random hydrolysis of *N*-acetyl-β-D-glucosaminide β-1,4-linkages in chitin and chitodextrins. Chitinases are found in a variety of organisms besides insects including bacteria, fungi, plants, and marine and land animals (Watanabe and Kono, 2002). Many genes encoding chitinolytic enzymes including several from insects

(Table 2) have been cloned and characterized. Some chitinases are now being used for biotechnological applications in agriculture and healthcare (Patil *et al*., 2000).

Chitinases are members of the superfamily of *O*-glycoside hydrolases, which hydrolyze the glycosidic bond in polysaccharides or between a sugar and a noncarbohydrate moiety. The International Union for Biochemistry and Molecular Biology enzyme nomenclature of glycoside hydrolases is based on their substrate specificity and occasionally based on their molecular mechanism. Such a classification, however, does not reflect the structural features of these enzymes. Another classification of glycoside hydrolases into families is based on amino acid sequence similarities. This classification is expected to: (1) reflect the structural features of these enzymes better than their sole substrate specificity; (2) help to reveal the evolutionary relationships between these enzymes; and (3) provide a convenient tool to derive mechanistic information (Henrissat and Bairoch, 1996). There are 91 families of glycosylhydrolases and to date all mechanistically characterized insect chitinases belong to family 18 (Coutinho and Henrissat, 1999; CAZY, 2004). Unlike family 19 chitinases that are found almost exclusively in plants, members of family 18 have been found in a wide variety of sources including bacteria, yeasts and other fungi, nematodes, arthropods, and even vertebrates such as mice, chickens, and humans (Nagano *et al*., 2002). The vertebrate proteins probably function as defensive proteins against chitin-containing pathogenic organisms.

4.3.4.1. Insect *N*-Acetylglucosaminidases

Beta-*N*-acetylglucosaminidases (EC 3.2.1.30) have been defined as enzymes that release β-*N*-acetylglucosamine residues from the nonreducing end of chitooligosaccharides and from glycoproteins with terminal *N*-acetylglucosamine. Insect β-*N*-acetylglucosaminidases are members of family 20 of the glycosylhydrolases (Coutinho and Henrissat, 1999; CAZY, 2004). These enzymes have been detected in the molting fluid, hemolymph, integument, and gut tissues of several species of insects (Kramer and Koga, 1986 and references therein). A β-*N*-acetylglucosaminidase also has been detected in the gut of *A. aegypti*, where its activity increased dramatically after blood feeding (Filho *et al*., 2002). Beta-*N*-acetylglucosaminidases also hydrolyze synthetic substrates such as *p*-nitrophenyl *N*-acetylglucosamine and 4-methylumbelliferyl oligo-β-*N*-acetylglucosamines. These two substrates have proven to be very useful in assays of these enzymes.

During development, β-*N*-acetylglucosaminidase activities are the highest in hemolymph a few days prior to larval or pupal ecdysis and in molting fluid from pharate pupae (Kimura, 1976, 1977; Turner *et al*., 1981). Two different enzymes with different physical and kinetic properties have been purified from the lepidopterans *B. mori* and *M. sexta*. The first enzyme (EI), which is found in larval and pharate pupal molting fluid and in pupal hemolymph, is probably involved in the turnover of chitobiose and possibly chitooligosaccharides because it has a lower K_m for these substrates than does the second (EII) enzyme. EII is found in larval and pupal hemolymph and has a lower K_m for pNpGlcNAc. The role of the enzyme (EII) is unclear, but its natural substrates may be glycoproteins containing terminal *N*-acetylglucosamines. However, this specificity remains to be proven.

4.3.4.2. Catalytic Mechanism of Insect *N*-Acetylglucosaminidases

N-acetylglucosaminidases have lower K_m values for substrates containing *N*-acetylglucosamine than those with *N*-acetylgalactosamine residues. They release monosaccharides from the nonreducing end by an exocleavage mechanism. Two ionizable groups with pKa values of 3.8 and 8.1 are involved in catalysis (Koga *et al*., 1982). Studies with competitive inhibitors such as δ-lactone derivatives of *N*-acetylglucosamine and *N*-acetylgalactosamine suggested that the active site of enzyme EI consists of subsites that bind larger substrates than does the active site of the EII enzyme. EI has a lower K_m than EII for the chitooligosaccharides and a larger K_m for pNpβGlcNAc, properties that are consistent with the two enzymes having different endogenous substrate specificities.

4.3.4.3. Cloning of cDNAs for Insect *N*-Acetylglucosaminidases

cDNAs for epidermal β-*N*-acetylglucosaminidases of *B. mori* (GenBank accession no. S77548), *B. mandarina* (accession no. AAG48701), *T. ni* (accession no. AAL82580), and *M. sexta* (accession no. AY368703) have been isolated and characterized (Nagamatsu *et al*., 1995; Zen *et al*., 1996; Goo *et al*., 1999). A search of the *Drosophila* and *Anopheles* genome databases also revealed the presence of closely related genes encoding β-*N*-acetylglucosaminidases. These genes encode closely related proteins (70-75% amino acid sequence identity between the *Manduca* and *Bombyx* enzymes) of approximately 68 kDa. The conceptual proteins contain leader peptides of 22-23 amino acids followed by stretches of

Table 2 Properties of insect chitinases

Species	*Common name*	*Tissue source*	*Number of amino acids*	*Domain structure*[a]	*GI no.*	*Reference*
Aedes aegypti	Yellow fever mosquito	ND	574	Cat-linker-ChBD 3ChBDs-3Cats	2564719	de la Vega *et al.* (1998)
Anopheles gambiae	Malaria mosquito	Gut	525	Cat-linker-ChBD	2654602	Shen and Jacobs-Lorena (1997)
Bombyx mori	Silkworm	Epidermis/gut	565	Cat-linker-ChBD	1841851, 10119784	Kim *et al.* (1998), Mikitani *et al.* (2000), Abdel-Banat and Koga (2001)
Chelonus sp. venom	Wasp	Venom gland	483	Cat-linker-ChBD	1079185	Krishnan *et al.* (1994)
Chironomus tentans	Midge	Cell line	475	Cat	2113832	Feix *et al.* (2000)
Choristoneura fumiferana	Spruce budworm	Epidermis/fat body	557	Cat-linker-ChBD	21913148	Zheng *et al.* (2002)
Drosophila melanogaster	Fruit fly	ND	508	Cat	17647257	de la Vega *et al.* (1998), Adams *et al.* (2000)
		ND	484	Cat	24655584	
		ND	458	ChBD-Cat	17647259	
Glossina morsitans	Tsetse fly	Fat body	460	Cat-ChBD	18201665	Yan *et al.* (2002)
Hyphantria cunea	Fall webworm	Epidermis	553	Cat-linker-ChBD	1841853	Kim *et al.* (1998)
Lutzomyia longipalpis	Sand fly	Midgut	474	Cat-linker-ChBD	28863959	Ramalho-Ortigão and Traub-Csekö (2003)
Manduca sexta	Tobacco hornworm	Epidermis/gut	554	Cat-linker-ChBD	1079015	Kramer *et al.* (1993), Choi *et al.* (1997)
Phaedon cochleariae	Mustard beetle	Gut	405	Cat	4210812	Girard and Jouanin (1999)
Spodoptera litura	Common cutworm	Epidermis	552	Cat-linker-ChBD	9971609	Shinoda *et al.* (2001)
Tenebrio molitor	Yellow mealworm	ND	2838	5 Cats+5 linkers+4 ChBDs+2 Mucs	21038943	Royer *et al.* (2002)

[a]Cat, catalytic domain; linker, linker region; ChBD, chitin-binding domain; Muc, mucin-like domain.
ND, not determined.

mature N-terminal amino acid sequences experimentally determined from *N*-acetylglucosaminidases purified from either the molting fluid or integument of these two species. The amino acid sequences include two regions that are highly conserved among *N*-acetylglucosaminidases from a variety sources including bacteria, yeast, mouse, and humans (Zen *et al*., 1996). The *M. sexta* gene was expressed most abundantly in epidermal and gut tissues prior to metamorphosis and was induced by 20-hydroxyecdysone. The inductive effect of molting hormone was suppressed by juvenoids (Zen *et al*., 1996).

4.3.4.4. Cloning of Genes Encoding Insect Chitinases

A chitinase from *M. sexta*, which is a 535 amino acid long glycoprotein (Chi535), as well as the cDNA and gene that encode it (*MsCHI*, accession no. AAC04924) were the first insect chitinase and gene to be isolated and characterized (Koga *et al*., 1983; Kramer *et al*., 1993; Choi *et al*., 1997; Kramer and Muthukrishnan, 1997). They represent the most extensively studied chitinase enzyme–gene system in any insect species and they have become a model for study of other insect chitinases and their genes. Since the cloning of the *M. sexta* gene in 1993, cDNAs or genomic clones for several other insect chitinases have been isolated and sequenced (**Table 2**). The organization of most of these genes is very similar to that of *M. sexta* and most of the proteins display a domain architecture consisting of catalytic, linker, and/or chitin-binding domains similar to MsCHI. These genes/enzymes include epidermal chitinases from the silkworm, *B. mori* (Kim *et al*., 1998; Abdel-Banat and Koga, 2001), the fall webworm, *Hyphantria cunea* (Kim *et al*., 1998), wasp venom (*Chelonus* sp.) (Krishnan *et al*., 1994), the common cutworm, *Spodoptera litura* (Shinoda *et al*., 2001), a molt-associated chitinase from the spruce budworm, *Choristoneura fumiferana* (Zheng *et al*., 2002), and midgut-associated chitinases from the malaria mosquito, *A. gambiae* (Shen and Jacobs-Lorena, 1997), yellow fever mosquito, *A. aegypti* (de la Vega *et al*., 1998), the beetle *Phaedon cochleariae* (Girard and Jouanin, 1999), and the sand fly, *Lutzomyia longipalpis* (Ramalho-Ortigão and Traub-Csekö, 2003), and several deduced from the *Drosophila* genome data. A smaller linkerless fatbody-specific chitinase from the tsetse fly, *Glossina morsitans* (Yan *et al*., 2002) and a very large epidermal chitinase with five copies of the catalytic-linker-chitin binding domain from the yellow mealworm, *Tenebrio molitor* (Royer *et al*., 2002) have also been described.

Recently, a gene encoding another type of chitinase from the silkworm, *BmChi-h*, has been reported (Daimon *et al*., 2003). The encoded chitinase shared extensive similarities with microbial and baculoviral chitinases (73% amino acid sequence identity to *Serratia marcescens* chitinase and 63% identity to *Autographa californica* nuclear polyhedrosis virus chitinase). Even though this enzyme had the signature sequence characteristic of family 18 chitinases, it had a rather low percentage of sequence identity with the family of insect chitinases listed in **Table 2**. It was suggested that an ancestral species of *B. mori* acquired this chitinase gene via horizontal gene transfer from *Serratia* or a baculovirus. Unlike the chitinases listed in **Table 2**, which typically have a leader peptide, catalytic domain, a serine/threonine(S/T)-rich domain and a C-terminal chitin-binding domain, BmChi-h chitinase has a leader peptide, one copy of module w1 domain that is found only in bacterial and baculoviral chitinases (Perrakis *et al*., 1994; Henrissat, 1999), and a catalytic domain. Apparently, *B. mori* is not alone among insects possessing such a chitinase of bacterial origin. A protein in the molting fluid of *M. sexta*, which cross-reacted with an antibody to *M. sexta N*-acetylglucosaminidase, was found to have an N-terminal amino acid sequence closely resembling that of *Serratia* chitinase (Zen *et al*., 1996). The N-terminal sequence of this protein was identical to that of BmChi-h up to the 25th amino acid residue, which strongly suggested that an ortholog of this chitinase gene exists in *M. sexta* as well. It will be interesting to investigate in the future whether this enzyme is widespread and found in other insect species. A search of the *Drosophila* and *Anopheles* genome databases, however, failed to identify any chitinase-like protein with an amino acid sequence identity to BmChi-h of greater than 40% (S. Muthukrishnan *et al*., unpublished data).

Reports of multiple forms of insect chitinases, which can be generated by several mechanisms, have appeared. Some of these proteins are no doubt products of multiple genes as described in the previous paragraph. Others are likely the result of posttranslational modifications that are caused by glycosylation and/or proteolysis, which can lead to larger glycosylated forms and smaller truncated forms (Koga *et al*., 1983; Wang *et al*., 1996; Gopalakrishnan *et al*., 1995; Arakane *et al*., 2003). Another cause can be alternative splicing of mRNA. In *B. mori*, alternative splicing of the primary transcript from a single chitinase gene generates heterogeneity within the products (Abdel-Banat and Koga, 2002). Larger chitinase-like proteins have been observed in the mosquito *Anopheles* and

it has been proposed that these zymogenic proteins are activated via proteolysis by trypsin (Shen and Jacobs-Lorena, 1997). However, Filho *et al.* (2002) found no evidence for such activation in the mosquito *Aedes* because high levels of chitinase activity were observed early after a blood meal and even in the guts of unfed insects. Putative zymogenic forms have been reported in other insects as well (Koga *et al.*, 1992; Bhatnagar *et al.*, 2003). However, the existence of a chitinase zymogen is still speculative in most cases because all of the fully characterized cDNAs encoding full-length insect chitinases apparently have the mature catalytic domains immediately following their leader peptides and there is no indication of the presence of pre-proproteins (**Table 2**). Preliminary evidence suggests that most, if not all, of the larger proteins reacting with chitinase antibodies are multimeric forms that are enzymatically inactive and produced as a result of intermolecular disulfide pairing. These larger forms appear after long periods of storage of the monomeric enzyme and they can be reconverted to enzymatically active monomeric forms by treatment with thiol reagents (Y. Arakane *et al.*, unpublished data).

4.3.4.5. Modular Structure of Insect Chitinases

A multidomain structural organization is generally observed in polysaccharide-degrading enzymes where one or more domains are responsible for hydrolysis and other domains are responsible for associating with the solid polysaccharide substrate. In addition, there usually are linker regions between the two types of domains, which also may be responsible, at least in part, for some functional properties of the enzymes. For example, the first chitinases shown to contain catalytic, linker, and chitin-binding or fibronectin-like domains were isolated from the bacterium *Bacillus circulans* (Watanabe *et al.*, 1990), the yeast *S. cereviseae* (Kuranda and Robbins, 1991), and the parasitic nematode *B. malayi* (Venegas *et al.*, 1996). Insect chitinases possess a similar structural organization, as do some other nematode, microbial, and plant chitinases as well as fungal cellulases. Observed in all of these enzymes is a multidomain architecture that may include a signal peptide and one or more of the following domains: catalytic domains, cysteine-rich chitin-binding domains, fibronectin-like domains, mucin-like domains, and S/T-rich linker domains, with the latter usually being rather heavily glycosylated (Tellam, 1996; Henrissat, 1999; Suzuki *et al.*, 1999). For example, chitinases from the bacterium *S. marcescens*, fall into three classes with sizes ranging from 36 to 52 kDa, which are composed of different combinations of catalytic domains, fibronectin type-III-like domains, and N- or C-terminal chitin-binding domains (Suzuki *et al.*, 1999). A novel multidomain structure exhibited by an insect chitinase is that of the yellow mealworm beetle, *T. molitor* (Royer *et al.*, 2002). This protein is unusually large, with a calculated molecular mass of approximately 320 kDa. It contains five catalytic domains, five S/T-rich linker domains, four chitin-binding domains, and two mucin-like domains. Gene duplication and domain deletion mechanisms have probably generated the diversity and multiplicity of chitinase genes in insects, as was demonstrated previously in bacteria (Saito *et al.*, 2003).

The structure of the catalytic domain of insect chitinase is a $(\beta\alpha)_8$ TIM (triose phosphate isomerase) barrel fold, which is one of the most common folds found in proteins (Nagano *et al.*, 2001, 2002). During protein evolution, domain shuffling has allowed this fold to acquire a large number of specific catalytic functions such as enzymes with a glycosidase activity like insect chitinase. The presence of additional domains such as linker and chitin-binding domains appears to further enhance the catalytic properties of these enzymes.

Figure 4 shows a phylogenetic tree of 16 insect chitinases inferred from an amino acid sequence alignment. All five of the lepidopteran enzymes and only one dipteran chitinase reside in the upper portion of the tree, whereas the other seven dipteran, one hymenopteran, and two coleopteran enzymes appear in the lower part.

Manduca sexta CHI is much smaller than the *Tenebrio* enzyme and much less complex in domain structure with only a single N-terminal catalytic domain (376 amino acids long), a linker domain (about 100 amino acids long), and a C-terminal chitin-binding domain (ChBD, 58 amino acids long) (Arakane *et al.*, 2003). Alternate domain arrangements occur in other glycosylhydrolases. For example, class I, class IV, and class VII plant chitinases contain an N-terminal ChBD and a G/P-rich linker preceding the catalytic domain (Raikhel *et al.*, 1993; Neuhaus, 1999), whereas fungal cellulases, like insect chitinase, possess a threonine/serine/proline-rich linker between the N-terminal catalytic domain and the C-terminal cellulose-binding domain (Srisodsuk *et al.*, 1993). The *Manduca* CHI linker region that is rich in T and S residues is also rich in P, D, and E residues, which qualifies it as a PEST sequence-containing protein according to Rogers *et al.* (1986). That composition suggested that insect chitinase might be rapidly degraded via the intracellular ubiquitin-conjugating enzymes/proteosome system, which recognizes the PEST

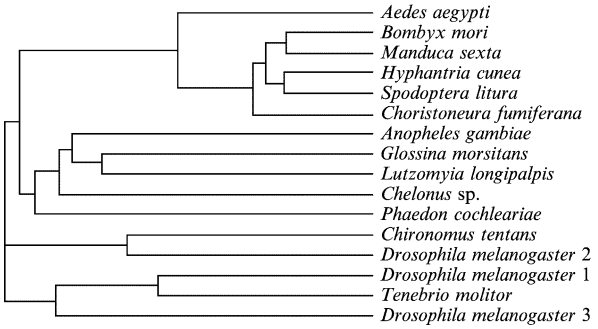


Figure 4 Phylogenetic tree of insect chitinases inferred from an amino acid sequence alignment of 16 enzymes from *Aedes aegypti*, *Bombyx mori*, *Manduca sexta*, *Hyphantria cunea*, *Spodoptera litura*, *Choristoneura fumiferana*, *Anopheles gambiae*, *Glossina morsitans*, *Lutzomyia longipalpis*, *Chelonus* sp., *Phaedon cochleariae*, *Chironomus tentans*, *Drosophila melanogaster*, and *Tenebrio molitor*. The GI numbers are listed in **Table 2**. Multiple sequence alignment was performed using Clustal W software (Thompson *et al.*, 1994) and the tree was built using the neighbor-joining method (Saitou and Nei, 1987).

sequence so that proteosomes can digest the conjugated protein when it is localized intracellularly. However, since insect chitinase is a secreted protein, it would be exposed to intracellular proteases or the ubiquitin-conjugating system only for a relatively short period of time. Instead, the linker apparently helps to optimize interactions with the insoluble substrates and to stabilize proteins, and perhaps also helps to protect protease-susceptible bonds in the catalytic domains from hydrolysis. Recombinant chitinases that contain this linker region were more stable in the presence of midgut digestive proteases than recombinant proteins lacking the linker region (Arakane *et al.*, 2003). The linker domain also may have another function involving protein trafficking. Recombinant forms of *Manduca* CHI lacking amino acid residues beyond position 376 accumulated intracellularly during expression in the baculovirus-insect cell line, whereas all of the forms that had an additional ten amino acids or longer stretches of the linker domain were secreted into the media (Arakane *et al.*, 2003). We concluded, therefore, that for secretion of recombinant protein to the outside of the insect cells to occur, the N-terminal portion of the linker region (residues 377–386) must be present, in addition to the 19 amino acid long N-terminal leader peptide. For secretion, the linker region may also need to be *O*-glycosylated because when glycosylation was inhibited by the addition of tunicamycin, insect chitinase accumulated intracellularly in an insect cell line (Gopalakrishnan *et al.*, 1995). Some of the critical residues for secretion/glycosylation, therefore, may involve residues between amino acids 376 and 386 (which includes two threonines) because the truncated Chi376 accumulated intracellularly, whereas Chi386 was secreted. Site-directed mutagenesis of these residues might help to answer the question about what residues in the linker region are required for secretion.

Peptides linking protein domains are very common in nature and some, unlike the insect chitinase linker, are believed to join domains rather passively without disturbing their function or affecting their susceptibility to cleavage by host proteases (Argos, 1990; Gilkes *et al.*, 1991). Linker peptides with G, T, or S residues are most common, perhaps because those residues are relatively small with G providing flexibility and T and S being uncharged but polar enough to interact with solvent or by their ability to hydrogen bond to water or to the protein backbone to achieve conformational and energetic stability. The interdomain linker peptide of a fungal cellobiohydrolase apparently has a dual role in providing the necessary distance between the two functional domains and also facilitating the dynamic adsorption process led by the cellulose-binding domain (Srisodsuk *et al.*, 1993). Solution conformation studies of a fungal cellulase with two domains revealed that its linker exhibited an extended conformation leading to maximum distance between the two domains and that heterogeneous glycosylation of the linker was likely a key factor defining its extended conformation (Receveur *et al.*, 2002). Since the domain structure of *M. sexta* CHI is similar to that of this fungal cellulase, these two enzymes may have similar global structural characteristics. Circular dichroism (CD) spectra of the wild-type and truncated insect chitinases were consistent with the hypothesis that whereas the catalytic and ChBDs possess secondary structure, the linker region itself does not (Arakane *et al.*, 2003).

Mammalian chitinase is similar in structure to *M. sexta* chitinase in both the catalytic domain and ChBD, but it lacks a linker domain (Tjoelker *et al.*, 2000). The absence of the ChBD does not affect the ability of the human enzyme to hydrolyze soluble oligosaccharides but does abolish hydrolysis of the insoluble substrate, a result consistent with the hypothesis that the function of the ChBD is to facilitate heterogeneous catalysis on insoluble substrates. One of the basic functions of carbohydrate-binding domains (CBD) is thought to be to help localize the enzyme on the insoluble substrate to enhance the efficiency of degradation (Linder and Teeri, 1997). These domains aid in recognition and hydrolysis of substrates that can exist in several physical states, i.e., contain both crystalline and noncrystalline forms. In general, for many glycosylhydrolases,

the binding specificity of the carbohydrate-binding domain mirrors that of the catalytic domain and these two domains are usually in relatively close association. Such is not the case for *Manduca* CHI, which has a very long linker of over several hundred angstroms.

Like their cognate catalytic domains, CBDs are classified into families of related amino acid sequences. The ChBD of insect chitinases belongs to carbohydrate-binding module family 14, which consists of approximately 70 residues (Coutinho and Henrissat, 1999; CAZY, 2004). Only three subfamilies of chitin-binding modules have been identified to date and the ChBD of *M. sexta* CHI is a member of subfamily 1 (Henrissat, 1999). Such a carbohydrate-binding function has been demonstrated in several other carbohydrolases and carbohydrate-binding proteins. Other CBD families, family 17 and family 28, both of which recognize cellulose, have been found to act in a cooperative manner either by modifying the action of the catalytic module or by targeting the enzyme to areas of cellulose that differ in susceptibility to hydrolysis (Boraston *et al.*, 2003). ChBDs may play a similar role in chitinases. These domains are attached not only to catalytic domains but also to chitinase-like proteins devoid of enzyme activity. The ChBDs can be either N- or C-terminal and may be present as a single copy or as multiple repeats. They are cysteine-rich and have several highly conserved aromatic residues (Shen and Jacobs-Lorena, 1999). The cysteine residues help to maintain protein folding by forming disulfide bridges and the aromatic residues interact with saccharides in the ligand-binding pocket. The PM proteins, mucins, which have affinity for chitin, also have a six-cysteine-containing peritrophin-A/mucin consensus sequence that is similar to ChBD sequences in chitinases (Tellam *et al.*, 1999; Morlais and Severson, 2001).

When fused with the catalytic domain of *M. sexta* CHI, both insect and rice ChBDs promoted the binding to and hydrolysis of chitin (Arakane *et al.*, 2003). The influence of extra substrate-binding domains has been examined previously using a fungal chitinase that was constructed to include plant and fungal carbohydrate-binding domains (Limón *et al.*, 2001). The addition of those domains increased the substrate-binding capacity and specific activity of the enzyme toward insoluble substrates of high molecular mass such as ground chitin or chitin-rich fungal cell walls. On the other hand, removal or addition of cellulose-binding domains can reduce or enhance, respectively, the ability of cellulases to degrade crystalline cellulose (Chhabra and Kelly, 2002). When a second cellulose-binding domain was fused to *Trichoderma reesei* cellulase, the resulting protein had a much higher affinity for cellulose than the protein with only a single binding domain (Linder *et al.*, 1996). Likewise, the *M. sexta* CHI catalytic domain fused with two ChBDs associated with chitin more strongly than any of the single ChBD-containing proteins or the protein devoid of a ChBD (Arakane *et al.*, 2003). This domain apparently helps to target the secreted enzyme to its insoluble substrate.

The chitin-binding domain of insect chitinase not only has the function of associating with insoluble chitin, but it may also help to direct the chitin chain into the active site of the catalytic domain in a manner similar to the processive hydrolysis mechanism proposed for *S. marcescens* chitinase A (ChiA), which has a very short ChBD (Uchiyama *et al.*, 2001). However, whether such an extended linker like that of *M. sexta* chitinase can direct the substrate into the active site in a manner similar to that proposed for a shorter linker is unknown. Catalytically, the full-length *M. sexta* CHI was two- to fourfold more active in hydrolyzing insoluble colloidal chitin than any of the other truncated enzymes with an intact catalytic domain, but all of the enzymes were comparable in turnover rate when two soluble substrates, carboxymethyl-chitin-remazol-brilliant-violet (CM-chitin-RBV), which is a chromogenic chitin derivative that is *O*-carboxymethylated, and MU-$(GlcNAc)_3$, a fluorogenic oligosaccharide substrate, were hydrolyzed (Arakane *et al.*, 2003). A moderate increase in catalytic efficiency of hydrolysis of insoluble substrate was observed when the catalytic domain was fused with the ChBD. When the C-terminal ChBD was deleted from a bacterial chitinase (*Aeromonas caviae*), this truncated chitinase was active also, but it liberated longer oligosaccharide products than did the full-length enzyme (Zhou *et al.*, 2002). Thus, as was observed with other carbohydrolases such as xylanases (Gill *et al.*, 1999), the ChBD of insect chitinase facilitates hydrolysis of insoluble, but not soluble, substrates, and also influences the size of the oligosaccharide products generated. The linker region also can influence the functionality of the carbohydrate-binding domain. When a fungal cellulose-binding domain was fused with a fungal S/T-rich linker peptide, the fusion protein adsorbed to both crystalline and amorphous cellulose. However, deletion of the linker peptide caused a decrease in cellulose adsorption and a higher sensitivity to protease digestion (Quentin *et al.*, 2002). The addition of a carbohydrate-binding module to a catalytic domain via a linker domain may increase the catalytic efficiency for degradation of the insoluble

polysaccharide and may modify the finely tuned binding specificity of the enzyme (McLean *et al.*, 2002; Lehtio *et al.*, 2003).

Figure 5 shows a theoretical model structure for *M. sexta* chitinase that is complexed with chitin oligosaccharides in both the catalytic domain and ChBD at a time subsequent to hydrolysis of a larger oligosaccharide. What is perhaps most striking is the very long linker (>200 Å) between the other domains. Apparently, the enzyme is tethered to the cuticle by the ChBD, which anchors the catalytic domain to the insoluble substrate and localizes the hydrolysis of chitin to an area with a radius of several hundred angstroms. The use of such a tethered enzyme would help to prevent diffusion of the soluble enzyme from the insoluble polysaccharide. In the case of *Tenebrio* chitinase, which consists of five catalytic, five linker, and six chitin-binding domains (Royer *et al.*, 2002), one could envision a situation where a much wider area of the chitin–protein matrix undergoes intensive degradation by a much larger tethered enzyme.

4.3.4.6. Mechanism of Catalysis

Insect chitinases are members of family 18 of the glycosylhydrolases (CAZY, 2004), which generally utilize a mechanism where retention of the anomeric configuration of the sugar donor occurs via a substrate-assisted catalysis, rather than a mechanism similar to lysozyme, which involves a proton donor and an electrostatic stabilizer (Fukamizo, 2000). However, a recent kinetic study of bacterial family 18 chitinases demonstrated that substrates lacking the *N*-acetyl group and thus incapable of anchimeric assistance were nevertheless hydrolyzed, suggesting that the reaction mechanism of family 18 chitinases cannot be fully explained by the substrate-assisted catalysis model (Honda *et al.*, 2003). Therefore, additional studies are still required to understand fully the reaction mechanism of family 18 chitinases.

The interaction of insect chitinases with insoluble chitin in the exoskeleton and PM is rather complex and believed to be a dynamic process that involves adsorption via a substrate-binding domain, hydrolysis, desorption, and repositioning of the catalytic domain on the surface of the substrate. This degradative process apparently requires a coordinated action of multiple domains by a mechanism that is not well understood. In addition to the catalytic events, the mechanism of binding of the enzyme onto the heterogeneous surface of native chitin is poorly characterized. Hydrolysis of chitin to GlcNAc is accomplished by a binary enzyme system composed of a chitinase and a β-*N*-acetylglucosaminidase (Fukamizo and Kramer, 1985;

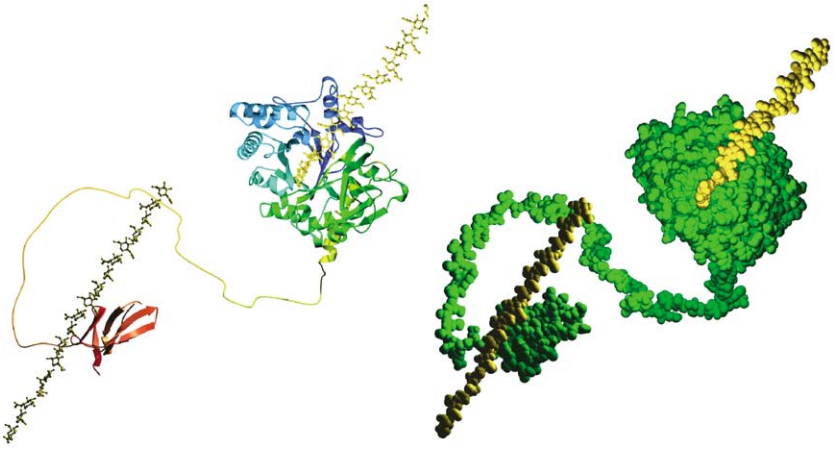


Figure 5 Ribbon (left) and space-filling (right) model structures of *Manduca sexta* chitinase with the catalytic and chitin-binding domains shown in complexes with chitin oligosaccharides (yellow). In the ribbon representation, the polypeptide chain is color-coded, beginning with blue at the N-terminus and proceeding through the rainbow to red at the C-terminus. The catalytic domain structure (top) was modeled using the program SOD (Kleywegt *et al.*, 2001) with human chitotriosidase (PDB entry code 1LG1) (Fusetti *et al.*, 2002) serving as the template. The chitin-binding domain (bottom) was similarly obtained using tachychitin (PDB entry 1DQC) (Suetake *et al.*, 2000) as the template. The linker region is shown as a random coil as predicted by secondary structure prediction software and supported by circular dichroism data. The oligosaccharides are shown as stick models (left) and space-filling models (right). Substrate binding to the catalytic domain was modeled using the available structures of complexes from glycosyl hydrolase family 18, while binding to the chitin-binding domain was modeled based on sequence conservation within the subfamily. *M. sexta* chitinase is a glycoprotein that is glycosylated especially in the linker region; however, no carbohydrate is shown in the model. (The model was constructed by Wimal Ubhayasekera and Dr. Sherry Mowbray, Swedish University of Agricultural Sciences, Uppsala.)

Filho *et al.*, 2002). The former enzyme hydrolyzes the insoluble polymer into soluble oligosaccharides, whereas the latter further degrades the oligomers to the monomer from the nonreducing end. Mechanistically, chitinases of family 18 hydrolyze chitin with retention of the anomeric configuration at the cleavage site, involving a double-displacement mechanism where a substrate-assisted catalysis occurs (Tomme *et al.*, 1995; Henrissat, 1999; Zechel and Withers, 2000; Brameld *et al.*, 2002). *B. mori* chitinase utilizes a retaining mechanism, yielding products that retain the β-anomeric configuration (Abdel-Banat *et al.*, 1999). All of the enzymes of this family are inhibited by allosamidin, a transition state analog inhibitor which apparently is diagnostic for enzymes that utilize the retaining mechanism (Koga *et al.*, 1987; Bortone *et al.*, 2002; Brameld *et al.*, 2002; Lu *et al.*, 2002). Analysis of the products from the hydrolysis of chitin oligosaccharides by the family 18 chitinase from *S. marcescens* revealed variable subsite binding preferences, anomeric selectivity, and the importance of individual binding sites for the processing of short oligosaccharides compared to the cumulative recognition and processive hydrolysis mechanism used to digest the polysaccharide (Aronson *et al.*, 2003).

Polysaccharide-hydrolyzing enzymes are known to exhibit nonideal kinetic behavior because they often are susceptible to inhibition by both substrates and products (Väljamäe *et al.*, 2001). All insect chitinases examined were found to be susceptible to inhibition by oligosaccharide substrates but to varying extents (Fukamizo and Kramer, 1985; Fukamizo *et al.*, 1995; Fukamizo, 2000). Apparently, the oligosaccharide substrate molecules can bind to these enzymes in such a manner that none of the target bonds is properly exposed to the functional groups of catalytic amino acids or the substrate may bind in only noncatalytic subsites of the larger active site, forming nonproductive instead of productive complexes. Cellulose is also degraded by the synergistic action of cellulolytic enzymes, which also display this characteristic substrate inhibition (Väljamäe *et al.*, 2001). Site-directed mutagenesis studies involving amino acids present in the putative catalytic site of *M. sexta* CHI have identified residues required for catalysis (Huang *et al.*, 2000; Lu *et al.*, 2002; Zhang *et al.*, 2002). Aspartic acids 142 and 144, tryptophan 145, and glutamic acid 146 were identified as residues very important for catalysis and also for extending the pH range of enzyme activity into the alkaline pH range. Acidic and aromatic residues in other family 18 chitinases also are important for substrate binding and catalysis (Watanabe *et al.*, 1993, 1994; Uchiyama *et al.*, 2001; Bortone *et al.*, 2002). Some of these residues are essential only for crystalline chitin hydrolysis, whereas others are important not only for crystalline chitin hydrolysis but for other substrates as well (Watanabe *et al.*, 2003).

4.3.4.7. Glycosylation of Insect Chitinases

Manduca sexta CHI is moderately *N*-glycosylated in the catalytic domain and heavily *O*-glycosylated in the linker region (Arakane *et al.*, 2003). The insect cell line TN-5B1-4 (Hi 5), which is routinely used for expression of recombinant foreign glycoprotein, synthesizes proteins with both *N*- and *O*-linked oligosaccharides (Davidson *et al.*, 1990; Davis and Wood, 1995; Jarvis and Finn, 1995; Hsu *et al.*, 1997). Results of experiments investigating the effects of the *N*-glycosylation inhibitor tunicamycin on recombinant expression of insect chitinases in these cells indicated that the proteins were glycosylated prior to being secreted by the cells (Gopalakrishnan *et al.*, 1995; Zheng *et al.*, 2002). Direct chemical and enzymatic analyses confirmed that *M. sexta* CHI was both *N*- and *O*-glycosylated. Prolonged deglycosylation with a mixture of *N*- and *O*-glycosidases resulted in a protein that was smaller by about 6 kDa accounting for about 30 sugar residues per mole of protein (Arakane *et al.*, 2003). Because *N*-linked oligosaccharides in insects typically have six or seven residues, two of which are GlcNAc (Paulson, 1989; Kubelka *et al.*, 1995), the best estimate of the distribution of *N*-glycosylation indicated a single or possibly two sites of *N*-glycosylation in the catalytic domain and *O*-glycosylation of between 10 and 20 serine or threonine residues in the linker region. *O*-glycosylation may involve mainly addition of galactose and *N*-acetylgalactosamine.

The chitinase from *B. mori* also is probably glycosylated because this protein and its breakdown product (65 kDa) stain with periodic acid–Schiff reagent. Further, the apparent mobility of the protein in sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is 88 kDa, whereas the molecular weight of the mature protein predicted from the cDNA sequence is only 60 kDa (Koga *et al.*, 1997). This protein has an S/T-rich linker similar to the *M. sexta* chitinase. On the other hand, the chitinase from wasp venom which has only a short linker region and is low in serine and threonine has nearly the same molecular weight as the one predicted from the cDNA sequence, suggesting that this protein may not be glycosylated (Krishnan *et al.*, 1994). Thus, there is a good

correlation between the presence of an S/T-rich linker and extensive glycosylation (predominantly *O*-glycosylation) of the chitinolytic proteins.

Glycosylation of the linker region may help to prevent proteolytic cleavage(s) at sites between the catalytic and chitin-binding domains. Such a functional role of glycosylated regions has been observed in some bacterial cellulases (Langsford *et al*., 1987). The full-length and near full-length *O*-glycosylated forms of *M. sexta* CHI were the most stable proteins when incubated with the midgut proteinases of the hornworm (Arakane *et al*., 2003). Protein modeling studies using the crystal structures of other family 18 glycosylhydrolases as templates suggested that the catalytic domain of *M. sexta* CHI has a $(\beta\alpha)_8$-triose phosphate isomerase (TIM) barrel structure (Kramer and Muthukrishnan, 1997; Nagano *et al*., 2002). The ChBD probably exhibits a multistranded β-sheet structure based on similarity to tachycitin (Suetake *et al*., 2000). We know of no structures computed or proposed for linker domains, which may be very hydrophilic and rather flexible as well as potentially susceptible to proteolytic degradation unless they are protected by glycosylation. The CD spectrum of the linker domain was consistent with the lack of any secondary structure in this domain (**Figure 5**). It is conceivable that during the developmental period of maximum chitinase activity, the enzyme is fully glycosylated. When required, a glycosidase(s) could be produced that would remove sugar residues, thus exposing several more peptide bonds for proteolytic cleavage. Alternatively, proteolytic cleavage may be reduced because of glycosylation. Consistent with this notion is the finding that analysis of molting fluid from *M. sexta* and *B. mori* revealed the presence of truncated forms of catalytically active chitinases with sizes ranging from 50 to 60 kDa (Kramer and Koga, 1986; Koga *et al*., 1997; Abdel-Banat *et al*., 1999). We also detected similar truncated forms in our insect cell recombinant chitinase expression system, especially several days subsequent to infection with the recombinant baculovirus (Gopalakrishnan *et al*., 1995).

4.3.4.8. Antigenicity of Insect Chitinases

Invertebrate chitinases have been reported to elicit allergies in mammals. For example, a high prevalence of IgE antibodies to a tick chitinase was identified in canine atopic dermatitis with the chitinase formally designated Der f 15 (McCall *et al*., 2001). In ticks, this chitinase was localized in the proventriculus and intestine, indicating that it has a digestive, rather than molting-related, function. Like insect chitinase, tick chitinase is extensively *O*-glycosylated on multiple sites along the 84 amino acid long S/T-rich sequence in the molecule. The transmission blocking antibody MF1 from the blood of gerbils infected with the nematode *B. malayi* was found to be directed against a microfilarial chitinase (Fuhrman *et al*., 1992). This antibody mediates the clearance of peripheral microfileremia in gerbils, indicating that chitinase is indeed a potent antigen. Even though it is unclear which region of the nematode chitinase is highly antigenic, the most probable one is the S/T–rich region known to be *O*-glycosylated.

The primary epitope recognized by antibodies elicited by *Manduca* chitinases is the highly glycosylated S/T-rich linker region (Arakane *et al*., 2003). Other highly immunogenic insect proteins that also are extensively *O*-glycosylated in S/T-rich domains similar to the linker region of *Manduca* CHI are peritrophins-55 and -95 from the sheep blowfly, *L. cuprina* (Tellam *et al*., 2000, 2003). The sera of sheep vaccinated with these peritrophins exhibited a strong immune response that also inhibited growth of blowfly larvae (Casu *et al*., 1997; Tellam *et al*., 2003).

4.3.4.9. Other Possible Enzymes of Chitin Metabolism

Chitin deacetylases and chitosanases are two other enzymes that play major roles in chitin catabolism in other types of organisms. Chitin deacetylase catalyzes the removal of acetyl groups from chitin. This enzyme is widely distributed in microorganisms and may have a role in cell wall biosynthesis and in counteracting plant defenses (Tsigos *et al*., 2000). There is one report of an insect chitin deacetylase in physogastric queens of the termite *Macrotermetes estherae* (Sundara Rajulu *et al*., 1982). However, there have been no follow-up studies about this enzyme in other insect species. To our knowledge, there are no reports of chitosanases present in insects.

4.3.5. Nonenzymatic Proteins That Bind to Chitin

There are approximately 32 families of CBDs that are defined as contiguous amino acid sequences within a carbohydrate-active enzyme or noncatalytic analogs, which exhibit a discrete fold having carbohydrate-binding activity (CAZY, 2004). One or more members in families 1, 2, 3, 5, 12, 14, 16, 18, and 19 are reported to bind chitin. Most, if not all, of the insect ChBDs, however, belong only to family 14.

Several chitinase-related proteins have been identified in insects, which are catalytically inactive because they are missing an amino acid residue critical for hydrolytic activity but nonetheless are carbohydrate-binding proteins with either a single copy or multiple repeats of ChBDs. These proteins may act as growth factors or play a defensive function as anti-inflammatory proteins. A chitinase homolog glycoprotein HAIP (hemolymph aggregation inhibiting protein) occurs in hemolymph of the lepidopteran *M. sexta*, which inhibits hemocyte aggregation (Kanost *et al.*, 1994). A similar immunoreactive protein was detected in hemolymph of three other lepidopterans, *B. mori*, *Heliothis zea*, and *Galleria mellonella*. These proteins may have a role in modulating adhesion of hemocytes during defensive responses. Another glycoprotein, Ds47, which is produced *in vitro* by a *Drosophila* embryo-derived cell line and by fat body and hemocytes, may play a role in promoting the growth of imaginal discs (Kirkpatrick *et al.*, 1995; Bryant, 2001). Another chitinase-related protein is induced together with a chitinase and β-*N*-acetylglucosaminidase by ecdysteroid in the anterior silk gland of *B. mori* at molting and at metamorphosis (Takahashi *et al.*, 2002). The former is rather large in size and has a novel structure consisting of tandemly repeated catalytic domain-like plus linker sequences, but it has only one ChBD located in the middle of the protein. All of these proteins are evolutionarily related to chitinases, but they apparently have acquired a new growth-promoting or infection-resistance function that does not require catalytic activity. Evidently, chitinases have evolved into these lectin-like proteins by mutation of key residues in the active site, which abolishes enzyme activity and fine tunes the ligand-binding specificity.

Chitin-binding proteins in vertebrates, invertebrates, and plants share a common structural motif composed of one to eight disulfide bonds and several aromatic residues, apparently the result of convergent evolution (Shen and Jacobs-Lorena, 1999; Suetake *et al.*, 2000). A chitin-binding antifungal peptide from the coconut rhinoceros beetle, *Oryctes rhinoceros*, scarabaecin, is only 36 residues in length and contains only one disulfide bond (Hemmi *et al.*, 2003). It shares significant tertiary structural similarity with ChBDs of other invertebrates and plants that have multiple disulfide bonds, even though there is no overall sequence similarity. Other invertebrate proteins that contain one or more ChBDs include the peritrophins (Tellam *et al.*, 1999), mucins (Casu *et al.*, 1997; Wang and Granados, 1997; Tellam *et al.*, 1999; Rayms-Keller *et al.*, 2000; Sarauer *et al.*, 2003), and tachycitin (Suetake *et al.*, 2000).

Other proteins that bind to chitin include several lectins and cuticular proteins (see **Chapter 4.2**). The lectins are related to ChBDs found in PM and chitinases. Many insect cuticular proteins contain an amino acid sequence motif of approximately 35 residues known as the R&R consensus sequence (Rebers and Willis, 2001). This sequence, however, has no similarity to the cysteine-rich ChBDs found in chitinases, some PM proteins, and lectins. There are no or very few cysteine residues in the cuticular protein ChBDs (noncysChBD). Thus, there are two distinct classes of invertebrate ChBDs, those with the chitin-binding domain found in lectins, chitinases, and PM proteins (cysChBDs) and those with the cuticular protein chitin-binding domain (noncysChBDs). Homology modeling of insect cuticle proteins using the bovine plasma retinol binding protein as a template predicted an antiparallel β-sheet half-barrel structure as the basic folding motif where an almost flat surface consisting of aromatic amino acid side chains interacts with the polysaccharide chains of chitin (Hamodrakas *et al.*, 2002).

In mammals there are several nonenzymatic members of the chitinase protein family. Oviduct-specific glycoprotein (OGP), a member of this family, is believed to be involved in the process of fertilization such as sperm function and gamete interactions (Araki *et al.*, 2003). However, OGP was not essential for *in vitro* fertilization in mice, and so the functionality of OGP remains unknown. The human cartilage protein HCgp-39 is a chitin-specific lectin (Renkema *et al.*, 1998; Houston *et al.*, 2003) that is overexpressed in articular chondrocytes and certain cancers. It is thought to be an anti-inflammatory-response protein and/or to play a role in connective tissue remodeling. In contrast to chitinases, which bind and hydrolyze chitin oligosaccharides but do not undergo large conformational changes, HCgp-39 exhibits a large conformational change upon ligand binding, which appears to signal the presence of chitinous pathogens such as fungi and nematodes (van Aalten, 2003). The murine *Ym1* gene belongs to a family of mammalian genes encoding nonenzymatic proteins that are homologous to the chitinases from lower organisms, such as insects, nematodes, bacteria, and plants (Sun *et al.*, 2001). YKL-40 is a nonenzymatic member of the mammalian family 18 glycosylhydrolases, which is a growth factor for connective tissue cells and stimulates migration of endothelial cells (Johansen *et al.*, 2003). It is secreted in large amounts by

human osteosarcoma cells and murine mammary tumors, and it is also elevated in patients with metastatic breast cancer and colorectal cancer. These homologous mammalian proteins have no demonstrable chitinase activity and, therefore, cannot be considered chitinases. The biological functions of these proteins remain obscure. However, these proteins likely function through binding to carbohydrate polymers and since they are secreted from activated hemocytes, they may have a function in immunity such as a hemocyte inhibition (Falcone *et al*., 2001). Sequence comparison of these nonenzymatic and enzymatic proteins indicates that the enzymatic proteins have evolved into these lectins by the mutation of key residues in the active site and optimization of the substrate-binding specificity (Fusetti *et al*., 2002).

4.3.6. Regulation of Chitin Degradation

The *M. sexta* chitinase and *N*-acetylglucosaminidase genes were shown to be upregulated by ecdysteroid (see **Chapter 3.5**) and down-regulated by the juvenile hormone mimic (see **Chapter 3.7**), phenoxycarb, in larval abdomens cut off from their hormonal sources (Fukamizo and Kramer, 1987; Koga *et al*., 1991; Kramer *et al*., 1993; Zen *et al*., 1996). Differential display was used to show that chitinase expression was regulated not only by ecdysteroid but also by juvenile hormone in the beetle *T. molitor* (Royer *et al*., 2002). Northern blot analysis of RNA from epidermis and 20-hydroxyecdysone-injected pupae showed that chitinase transcripts were correlated with molting hormone levels during metamorphosis. In addition, topical application of a juvenile hormone (JH) analog indirectly induced expression of chitinase mRNA. Thus, the *Tenebrio* chitinase gene is an early direct ecdysteroid-responsive one at the transcriptional level, but unlike *M. sexta* chitinase, it is apparently a direct target of JH as well. In the former case, the level at which JH regulates chitinase mRNA levels remains to be determined. The 20-hydroxyecdysone agonist, tebufenozide, induced expression of *C. fumiferana* chitinase when it was injected into mature larvae. The enzyme was produced 24 h post treatment in both the epidermis and molting fluid (Zheng *et al*., 2003).

4.3.7. Chitin Metabolism and Insect Control

Chitinases have been used in a variety of ways for insect control and other purposes (Kramer *et al*., 1997; Gooday, 1999). Several chitinase inhibitors with biological activity have been identified based on natural products chemistry (Spindler and Spindler-Barth, 1999), such as allosamidin (Rao *et al*., 2003) which mimics the carbohydrate substrate, and cyclic peptides (Houston *et al*., 2002). Although useful for biochemical studies, none of these chitin catabolism inhibitors have been developed for commercial use primarily because of their high cost of production and potential side effects. As we learn more details about chitinase catalysis, it might become more economically feasible to develop and optimize chitinase inhibitors for insect pest management.

Additional uncharacterized steps in chitin synthesis and/or assembly of chitin microfibrils, on the other hand, have proved to be important for developing control chemicals that act selectively on economically important groups of insect pests (Verloop and Ferrell, 1977; Ishaaya, 2001). The benzoylphenylureas have been developed as commercial compounds for controlling agricultural pests. These antimolting insecticides are relatively nontoxic to mammals due to their strong protein binding and extensive metabolization to less toxic compounds (Bayoumi *et al*., 2003). Studies using imaginal discs and cell-free systems indicated that benzoylphenylureas inhibit ecdysteroid-dependent GlcNAc incorporation into chitin (Mikolajczyk *et al*., 1994; Oberlander and Silhacek, 1998). Those results suggest that benzoylphenylureas affect ecdysone-dependent sites, which leads to chitin inhibition. However, the site of action of the benzoylphenylureas still is not well known. Recently, several heteryl nucleoside nonhydrolyzable transition state analogs of UDP-GlcNAc were synthesized and evaluated for fungicidal activity, but they were not assayed for insecticidal activity (Behr *et al*., 2003).

Entomopathogens secrete a plethora of extracellular proteins with potential activity in insect hosts. One of these proteins is chitinase, which is used by fungi such as *Metarhizium anisopliae* to help penetrate the host cuticle and render host tissues suitable for consumption (St. Leger *et al*., 1996; Krieger de Moraes *et al*., 2003). Among the 10 most frequent transcripts in a strain of *M. anisopliae* are three encoding chitinases and one a chitosanase, presumably reflecting a greater propensity to produce chitinases for host cuticle penetration (Freimoser *et al*., 2003a). Expressed sequence tag analysis of *M. anisopliae* may hasten gene discovery to enhance development of improved mycoinsecticides. However, when *M. anisopliae* was transformed to overexpress its native chitinase, the pathogenicity to the tobacco hornworm was unaltered, suggesting that

wild-type levels of chitinase are not limiting for cuticle penetration (Screen *et al*., 2001). Another fungal species, *Conidiobolus coronatus*, also produces both endo- and exo-acting chitinolytic enzymes during growth on insect cuticle (Freimoser *et al*., 2003b). Apparently, both *M. anisopliae* and *C. coronatus* produce a chitinolytic enzyme system to degrade cuticular components.

Both microbial and insect chitinases have been shown to enhance the toxicity of the entomopathogenic bacterium *Bacillus thuringiensis* (Bt) (Regev *et al*., 1996; Tantimavanich *et al*., 1997; Ding *et al*., 1998; Sampson and Gooday, 1998; Wiwat *et al*., 2000). For example, when the chitinolytic activities of several strains of *B. thuringiensis* were compared with their insecticidal activity, it was determined that the enzyme could enhance the toxicity of Bt to *Spodoptera exigua* larvae by more than twofold (Liu *et al*., 2002). Microbial chitinases have been used in mixing experiments to increase the potency of entomopathogenic microorganisms (review: Kramer *et al*., 1997). Synergistic effects between chitinolytic enzymes and microbial insecticides have been reported as early as the 1970s. Bacterial chitinolytic enzymes were first used to enhance the activity of Bt and a baculovirus. Larvae of *C. fumiferana* died more rapidly when exposed to chitinase–Bt mixtures than when exposed to the enzyme or bacterium alone (Smirnoff and Valero, 1972; Morris, 1976; Lysenko, 1976). Mortality of gypsy moth, *Lymantria dispar*, larvae was enhanced when chitinase was mixed with Bt relative to a treatment with Bt alone in laboratory experiments (Dubois, 1977). The toxic effect was correlated positively with enzyme levels (Gunner *et al*., 1985). The larvicidal activity of a nuclear polyhedrosis virus toward *L. dispar* larvae was increased about fivefold when it was administered with a bacterial chitinase (Shapiro *et al*., 1987). Chitin synthesis-inhibiting antifungal agents such as flufenoxuron and nikkomycin were used to promote the infection of silkworms with *B. mori* nucleopolyhedrovirus (Arakawa, 2002, 2003; Arakawa and Sugiyama, 2002; Arakawa *et al*., 2002). The mechanism of viral infection enhancement by these agents is not established, but it may involve destruction of PM structure, which would facilitate tissue invasion.

Inducible chitinolytic enzymes from bacteria cause insect mortality under certain conditions. These enzymes may compromise the structural integrity of the PM barrier and improve the effectiveness of Bt toxin by enhancing contact of the toxin molecules with their epithelial membrane receptors. For example, five chitinolytic bacterial strains isolated from midguts of *Spodoptera littoralis* induced a synergistic increase in larval mortality when combined with Bt spore-crystal suspensions relative to either an individual bacterial strain or a Bt suspension alone (Sneh *et al*., 1983). An enhanced toxic effect toward *S. littoralis* also resulted when a combination of low levels of a truncated recombinant Bt toxin and a bacterial endochitinase was incorporated into a semisynthetic insect diet (Regev *et al*., 1996). Crude chitinase preparations from *B. circulans* enhanced the toxicity of Bt *kurstaki* toward diamondback moth larvae (Wiwat *et al*., 1996). Liu *et al*. (2002) recently reported that several strains of Bt produced their own chitinases, which had synergistic larvicidal activity with the endotoxins.

In biopesticide development research, we used a family 18 insect chitinase as an enhancer protein for baculovirus toxicity and as a host plant resistance factor in transgenic plants. Introduction of an insect chitinase cDNA into *A. californica* multiple nuclear polyhedrosis viral (AcMNPV) DNA accelerated the rate of killing of fall armyworm compared to the wild-type virus (Gopalakrishnan *et al*., 1995). Baculoviral chitinases themselves play a role in liquefaction of insect hosts (Hawtin *et al*., 1997; Thomas *et al*., 2000). A constitutively expressed exochitinase from *B. thuringiensis* potentiated the insecticidal effect of the vegetative insecticidal protein Vip when they were fed to neonate larvae of *S. litura* (Arora *et al*., 2003). Some granuloviruses, on the other hand, do not utilize chitinases in a similar manner, which helps to explain why some granulovirus-infected insects do not lyse at the end of the infection process (Wormleaton *et al*., 2003). Mutagenesis of the AcMNPV chitinase gene resulted in cessation of liquefaction of infected *T. ni* larvae, supporting a role of chitinase in virus spread (Thomas *et al*., 2000). However, the insecticidal activity of insect chitinase was not substantial enough for commercial development. We have attempted with little success to improve the catalytic efficiency and stability of this enzyme so that its pesticidal activity would be enhanced (Lu *et al*., 2002; Zhang *et al*., 2002; Arakane *et al*., 2003). Nevertheless, tobacco budworms were killed when reared on transgenic tobacco expressing a truncated, enzymatically active form of insect chitinase (Ding *et al*., 1998). We also discovered a synergistic interaction between insect chitinase expressed in transgenic tobacco plants and Bt (applied as a spray at sublethal levels) using the tobacco hornworm as the test insect. In contrast to results obtained with the tobacco budworm, studies with the hornworm revealed no consistent differences in larval growth or foliar damage when the insects were reared on

first-generation transgenic chitinase-positive tobacco plants as compared to chitinase-negative control plants. When Bt toxin was applied at levels where no growth inhibition was observed on control plants, chitinase-positive plants had significantly less foliar damage and lower larval biomass production. These results indicated that the insect chitinase transgene did potentiate the effect of sublethal doses of Bt toxin and vice versa (Ding *et al*., 1998). Tomato plants have been transformed with fungal chitinase genes with concomitant enhancement in resistance to insect pests (Gongora *et al*., 2001). Effects observed include reduced growth rates and increased mortality, as well as a decrease in plant height and flowering time with an increase in the number of flowers and fruits (Gongora and Broadway, 2002). Chitinase-secreting bacteria have been used to suppress herbivorous insect pests. A chitinase gene-transformed strain of *Enterobacter cloacae* digested the chitinous membranes of phytophagous ladybird beetles, *Epilachna vigintioctopunctata*, and also suppressed leaf-feeding and oviposition when the beetles ingested transformed bacteria entrapped in alginate microbeads sprayed on tomato seedlings (Otsu *et al*., 2003).

Several GlcNAc-specific lectins from plants have been evaluated for insect toxicity (Harper *et al*., 1998; Macedo *et al*., 2003). These proteins appear to disrupt the integrity of the PM by binding to chitin or glycan receptors on the surface of cells lining the insect gut. They also may bind to glycosylated digestive enzymes and inhibit their activity. Another type of plant chitin-binding protein is the seed storage protein, vicilin, which is actually a family of oligomeric proteins with variable degrees of glycosylation (Macedo *et al*., 1993; Shutov *et al*., 1995). Some vicilins are insecticidal to bruchid beetles and stalk borers (Sales *et al*., 2001; Mota *et al*., 2003). Apparently, these proteins bind to the PM, causing developmental abnormalities and reduced survival rates. To date no carbohydrate-binding protein derived from an insect has been evaluated for biocidal activity. A novel approach has been proposed to develop strategies for insect control by utilizing chitin-binding molecules to specifically target formation of the PM. Calcofluor, a chemical whitener with chitin-binding properties, was used as a model compound in the diet to inhibit PM formation in *T. ni* and also to increase larval susceptibility to baculovirus infection (Wang and Granados, 2000b). It also was effective in suppressing PM formation in *Spodoptera frugiperda* and at the same time in preventing the establishment of a decreasing gradient of proteinases along the midgut tissue (Bolognesi *et al*., 2001).

Another type of hydrolytic enzyme with a ChBD has been shown to exhibit insecticidal activity in plants. Maize accumulates a 33 kDa cysteine protease containing a ChBD in response to insect feeding (Perchan *et al*., 2002). This enzyme apparently damages the insect's PM by utilizing the ChBD to localize itself at the chitin-protein-rich PM, where the PM proteins are digested, rendering the PM dysfunctional. Another protease with a chitin-binding domain has been described from *A. gambiae*, which may be involved in insect defense (Danielli *et al*., 2000). This 147 kDa protein, sp22D, is expressed in a variety of tissues, most strongly in hemocytes, and is secreted into the hemolymph. Upon bacterial infection, the transcripts for this protein increase by about twofold suggesting a role in insect defense. This protein has a multidomain organization that includes two copies of an N-terminal ChBD, a C-terminal protease domain, and additional receptor domains. It binds strongly to chitin and undergoes complex proteolytic processing during pupal to adult metamorphosis. It has been proposed that exposure of this protease to chitin may regulate its activity during tissue remodeling or wounding.

Recently, two synthetic peptides were found to inhibit *A. gambiae* midgut chitinase and also to block sporogonic development of the human malaria parasite, *Plasmodium falciparum*, and avian malaria parasite, *P. gallinaceum*, when the peptides were fed to infected mosquitoes (Bhatnagar *et al*., 2003). The design of these peptides was based on the putative proregion sequence of mosquito midgut chitinase. The results indicated that expression of chitinase inhibitory peptides in transgenic mosquitoes might alter the vectorial capacity of mosquitoes to transmit malaria.

4.3.8. Concluding Remarks

Although chitin was discovered nearly two centuries ago, it remains a biomaterial in waiting because, unlike other natural materials such as collagen and hyaluronic acid, very few technological uses have been developed (Khor, 2002; Tharanathan and Kittur, 2003). There are many unanswered questions about chitin morphology and chitin deposition in the insect cuticle and PM. We do not know how or whether chitin forms covalent interactions with other components in these extracellular matrices. Chitosan, on the other hand, does react with quinones (Muzzarelli and Muzzarelli, 2002; Muzzarelli *et al*., 2003). Thus, if there were any free amino groups in insect chitin, C–N linkages between chitin and catechols would be expected (Schaefer *et al*.,

1987). We do not yet understand how factors such as metal ions affect chitin metabolism. In fungi, ions such as zinc were found to alter chitin deposition and morphology (Lanfranco *et al*., 2002). Perhaps, in insects there is an ionic effect on differential expression of CHS isozymes.

We know much more about insect chitinolytic enzymes than about insect chitin biosynthetic enzymes. Many questions remain about the biosynthesis of insect chitin, not the least of which are why insects have multiple genes for CHS, how many CHSs are required to make an insect, at what developmental stages are the various CHSs produced, and what are the unique properties and functions of each CHS. Of particular interest is the role of alternate splicing in generating different isoforms of CHSs from the same gene. The developmental cues that control alternate splicing and how they affect chitin synthesis and/or deposition will be subjects of future studies. The cloning of CHS genes should soon lead to availability of large amounts of recombinant enzymes or subdomains thereof using appropriate expression systems. Studies with pure proteins and the availability of molecular probes will provide a better understanding of the chitin biosynthetic pathway and its regulation in the future.

Two other major questions about insect chitin biosynthesis are: what is the mechanism of the initiation phase and is there an autocatalytic initiator. Like glycogen synthesis, chitin synthesis probably includes both initiation and elongation phases. As the initiator of glycogen synthesis, glycogenin transfers glucose from UDP-glucose to itself to form an oligosaccharide–protein primer for elongation (Gibbons *et al*., 2002). Like chitin synthase, glycogenin is a glycosyltransferase, which raises the question of whether chitin synthase has an autocatalytic function similar to glycogenin and whether there is a chitinogenin-like protein. Another possibility is the participation of a lipid primer for chitin synthesis. Recently, cellulose synthesis in plants was found to involve the transfer of lipid-linked cellodextrins to a growing glucan chain (Read and Bacic, 2002). The lipid in this case was sitosterol-β-glucoside.

Little is known about the catalytic mechanism of any insect CHS. Once insect CHS-related recombinant proteins are obtained, site-directed mutagenesis can be used to probe for essential residues in the catalytic and regulatory domains. It is likely that acidic amino acids play critical roles in CHS catalysis in a manner comparable to those identified in other glycosyltransferases (Hefner and Stockigt, 2003) and in yeast chitin synthases (Nagahashi *et al*., 1995).

Chitinolytic enzymes are gaining importance for their biotechnological applications in agriculture and healthcare (Patil *et al*., 2000). Additional success in using chitinases for different applications depends on a better understanding of their biochemistry and regulation so that their useful properties can be optimized through genetic and biochemical engineering. Reasons for the rather high multiplicity of domain structures for insect and other chitinases are not fully understood. So far little success has occurred in using chitinase in pest control applications, but it may prove more useful as an enhancer protein in a cocktail with other biopesticides targeted at the cuticle or gut. Also, only a few catalytic domains or chitin-binding domains or various combinations thereof have been evaluated for biocidal activity and thus, further toxicological experimentation is warranted.

Although substantial progress in studies of insect chitin metabolism has occurred since the first edition of *Comprehensive Insect Physiology, Biochemistry, and Pharmacology* was published in 1985, we still do not know much about how chitin is produced and transported across the membrane so that it can interact perfectly with other components for assembly of the supramolecular extracellular structures called the exoskeleton and PM. These materials are still very much biochemical puzzles in which we do not understand well how the various components come together during morphogenesis or are digested apart during the molting process. Hopefully, this chapter will stimulate more effort to understand how insects utilize chitin metabolism for growth and development, and to develop materials that may perturb insect chitin metabolism for pest management purposes.

Acknowledgments

The authors are grateful to Sherry Mowbray, Wimal Ubhayasekera, Yasuyuki Arakane, Qingsong Zhu, David Hogenkamp, Renata Bolognesi, Tamo Fukamizo, Daizo Koga, Walter Terra and Clelia Terra for their help with the preparation and/or comments on various aspects of this review. Supported in part by National Science Foundation grant no. 0316963. Mention of a proprietary product does not constitute a recommendation or endorsement by the US Department of Agriculture. The US Department of Agriculture is an equal opportunity/affirmative action employer and all agency services are available without discrimination. This is contribution no. 04-058-C of the Kansas Agricultural Experiment Station.

References

Abdel-Banat, B.M.A., Kameyama, Y., Yoshioka, T., Koga, D., **1999**. Purification and characterization of a 54 kDa chitinase from *Bombyx mori*. *Insect Biochem. Mol. Biol. 30*, 107–117.

Abdel-Banat, B.M.A., Koga, D., **2001**. A genomic clone for a chitinase gene from the silkworm, *Bombyx mori*: structural organization identifies functional motifs. *Insect Biochem. Mol. Biol. 31*, 497–508.

Abdel-Banat, B.M.A., Koga, D., **2002**. Alternative splicing of the primary transcript generates heterogeneity within the products of the gene for *Bombyx mori* chitinase. *J. Biol. Chem. 277*, 30524–30534.

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., *et al.*, **2000**. The genome sequence of *Drosophila melanogaster. Science 287*, 2185–2195.

Arakane, Y., Hogenkamp, D., Zhu, Y.C., Kramer, K.J., Specht, C.A., *et al.*, **2004**. Chitin synthase genes of the red flour beetle, *Tribolium castaneum*: characterization, expression, linkage mapping and alternate exon usage. *Insect Biochem. Mol. Biol. 34*, 291–304.

Arakane, Y., Zhu, Q., Matsumiya, M., Muthukrishnan, S., Kramer, K.J., **2003**. Properties of catalytic, linker and chitin-binding domains of insect chitinase. *Insect Biochem. Mol. Biol. 33*, 631–648.

Arakawa, T., **2002**. Promotion of nucleopolyhedrovirus infection in larvae of the silkworm, *Bombyx mori* (Lepidoptera: Bombycidae) by flufenoxuron. *Appl. Entomol. Zool. 37*, 7–11.

Arakawa, T., **2003**. Chitin synthesis inhibiting antifungal agents promote nucleopolyhedrovirus infection in silkworm, *Bombyx mori* (Lepidoptera: Bombycidae) larvae. *J. Invert. Pathol. 83*, 261–263.

Arakawa, T., Furuta, Y., Miyazawa, M., Kato, M., **2002**. Flufenoxuron, an insect growth regulator, promotes peroral infection by nucleopolyhedrovirus (BmNPV) budded particles in the silkworm, *Bombyx mori* L. *J. Virol. Methods 100*, 141–147.

Arakawa, T., Sugiyama, M., **2002**. Promotion of nucleopolyhedrovirus infection in larvae of the silkworm, *Bombyx mori* (Lepidoptera: Bombycidae) by an antibiotic, nikkomycin Z. *Appl. Entomol. Zool. 37*, 393–397.

Araki, Y., Nohara, M., Yoshida-Komiya, H., Kuramochi, T., Ito, M., *et al.*, **2003**. Effect of a null mutation of the oviduct-specific glycoprotein gene on mouse fertilization. *Biochem. J. 374*, 551–557.

Argos, P., **1990**. An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. *J. Mol. Biol. 211*, 943–958.

Aronson, N.N., Halloran, B.A., Alexyev, M.F., Amable, L., Madura, J.D., *et al.*, **2003**. Family 18 chitinase-oligosaccharide substrate interaction: subsite preference and anomer selectivity of *S. marcescens* chitinase A. *Biochem. J. 376*, 87–95. Immediate Publication, doi:10.1042/BJ20030273.

Arora, N., Ahmad, T., Rajagopal, R., Bhatnagar, R.K., **2003**. A constitutively expressed 36 kDa exochitinase from *Bacillus thuringiensis* HD-1. *Biochem. Biophys. Res. Commun. 307*, 620–625.

Bayoumi, A.E., Rerez-Pertejo, Y., Zidan, H.Z., Balana-Fouce, R., Ordonez, C., *et al.*, **2003**. Cytotoxic effects of two antimolting insecticides in mammalian CHO-K1 cells. *Ecotoxicol. Environ. Safety 55*, 19–23.

Becker, A., Schloder, P., Steele, J.E., Wegener, G., **1996**. The regulation of insect metabolism in insects. *Experientia 52*, 433–439.

Behr, J.B., Gourlain, T., Helimi, A., Guillerm, G., **2003**. Design, synthesis and biological evaluation of heterylnucleoside derivatives as inhibitors of chitin synthase. *Biorgan. Medicinal Chem. Lett. 13*, 1713–1716.

Bhatnagar, R.K., Arora, N., Sachidanand, S., Shahabuddin, M., Keister, D., *et al.*, **2003**. Synthetic propeptide inhibits mosquito midgut chitinase and blocks sporogonic development of malaria parasite. *Biochem. Biophys. Res. Comm. 304*, 783–787.

Bolognesi, R., Ribeiro, A.F., Terra, W.R., Terra, C., **2001**. The peritrophic membrane of *Spodoptera frugiperda*: secretion of peritrophins and role in immobilization and recycling digestive enzymes. *Arch. Insect Biochem. Physiol. 47*, 62–75.

Boraston, A.B., Kwan, E., Chiu, P., Warren, A.J., Kilburn, D.G., **2003**. Recognition and hydrolysis of noncrystalline cellulose. *J. Biol. Chem. 278*, 6120–6127.

Bortone, K., Monzingo, A.F., Ernst, S., Robertus, J.D., **2002**. The structure of an allosamidin complex with the *Coccidioides immitis* chitinase defines a role for a second acid residue in substrate-assisted mechanism. *J. Mol. Biol. 320*, 293–302.

Brameld, K.A., Shrader, W.D., Imperiali, B., Goddard, W.A., **2002**. Substrate assistance in the mechanism of family 18 chitinases: theoretical studies of potential intermediates and inhibitors. *J. Mol. Biol. 280*, 913–923.

Bray, S.J., Kafatos, F.C., **1991**. Developmental function of *Elf-1*: an essential transcription factor during embryogenesis in *Drosophila*. *Genes Devel. 5*, 1672–1683.

Bryant, P., **2001**. Growth factors controlling imaginal disc growth in *Drosophila*. In: Bock, G., Cardew, G., Goode, J.A. (Eds.), The Cell Cycle and Development. John Wiley, New York, pp. 182–199.

Casu, R., Eisemann, C., Pearson, R., Riding, G., East, I., *et al.*, **1997**. Antibody-mediated inhibition of the growth of larvae from an insect causing cutaneous myiasis in a mammalian host. *Proc. Natl Acad. Sci. USA 94*, 8939–8944.

CAZY, **2004**. Carbohydrate-active enzymes family server: http://afmb.cnrs mrs.fr/~pedro/CAZY/db.html.

Chang, R., Yeager, A.R., Finney, N.S., **2003**. Probing the mechanism of fungal glycosyltransferase essential for cell wall synthesis: UDP-Chitobiose is not a substrate for chitin synthase. *Org. Biomol. Chem. 1*, 39–41.

Chhabra, S.R., Kelly, R.M., **2002**. Biochemical characterization of *Thermotoga maritima* endoglucanase Cel74

with and without a carbohydrate-binding module (CBM). *FEBS Lett. 531*, 375–380.

Choi, H.K., Choi, K.H., Kramer, K.J., Muthukrishnan, S., **1997**. Isolation and characterization of a genomic clone for the gene of an insect molting enzyme, chitinase. *Insect Biochem. Mol. Biol. 27*, 37–47.

Cohen, E., **1987**. Chitin biochemistry: synthesis and inhibition. *Annu. Rev. Entomol. 32*, 71–93.

Cohen, E., **2001**. Chitin synthesis and inhibition: a revisit. *Pest Mgt Sci. 57*, 946–950.

Cohen, E., Casida, J.E., **1980**. Inhibition of *Tribolium* gut chitin synthetase. *Insecticide Biochem. Physiol. 13*, 129–136.

Coutinho, P.M., Deleury, E., Davies, G.J., Henrissat, B., **2003**. An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol. 328*, 307–317.

Coutinho, P.M., Henrissat, B., **1999**. Carbohydrate-active enzymes: an integrated database approach. In: Gilbert, H.H., Davies, G.J., Henrissat, H., Svensson, B. (Eds.), Recent Advances in Carbohydrate Bioengineering. Royal Society of Chemistry, Cambridge, pp. 3–12.

Daimon, T., Hamada, K., Mita, K., Okano, K., Suzuki, M.G., *et al.*, **2003**. A *Bombyx mori* gene, *BmChi-h*, encodes a protein homologous to bacterial and baculovirus chitinases. *Insect Biochem. Mol. Biol. 33*, 749–759.

Danielli, A., Loukeris, T.G., Lagueux, M., Muller, H.M., Richman, A., *et al.*, **2000**. A modular chitin-binding protease associated with hemocytes and hemolymph in the mosquito *Anopheles gambiae*. *Proc. Natl Acad. Sci. USA 97*, 7136–7141.

Davidson, D.J., Fraser, M.J., Castellino, F.J., **1990**. Oligosaccharide processing in the expression of human plasminogen cDNA by lepidopteran insect (*Spodoptera frugiperda*) cells. *Biochemistry 19*, 5584–5590.

Davis, T.R., Wood, H.A., **1995**. Intrinsic glycosylation potentials of insect cell lines and insect larvae. *In Vitro Cell. Devel. Biol. Anim. 31*, 659–663.

de la Vega, H., Specht, C.A., Liu, Y., Robbins, P.W., **1998**. Chitinases are a multi-gene family in *Aedes*, *Anopheles* and *Drosophila*. *Insect Mol. Biol. 7*, 233–239.

Ding, X., Gopalakrishnan, B., Johnson, L., White, F.F., Wang, X., *et al.*, **1998**. Insect resistance of transgenic tobacco expressing an insect chitinase gene. *Transgenic Res. 7*, 77–84.

Dubois, N.R., **1977**. Pathogenicity of selected resident microorganisms of *Lymantria dispar* after induction for chitinase. Ph.D. Dissertation, Univ. Massachusetts, Amherst.

Eisemann, C., Wijffels, G., Tellam, R.L., **2001**. Secretion of type 2 peritrophic matrix protein, peritrophin-15, from the cardia. *Arch. Insect Biochem. Physiol. 47*, 76–85.

Falcone, F.H., Loke, P., Zang, X., MacDonald, A.S., Maizels, R.M., *et al.*, **2001**. A *Brugia malayi* homolog of macrophage migration inhibitory factor reveals an important link between macrophages and eosinophil recruitment during nematode infection. *J. Immunol. 167*, 5348–5354.

Feix, M., Gloggler, S., Londershausen, M., Weidemann, W., Spindler, K.D., *et al.*, **2000**. A cDNA encoding a chitinase from the epithelial cell line of *Chironomus tentans* (Insecta, Diptera) and its functional expression. *Arch. Insect Biochem. Physiol. 45*, 24–36.

Filho, B.P.D., Lemos, F.J.A., Secundino, N.F.C., Pascoa, V., Pereira, S.T., *et al.*, **2002**. Presence of chitinase and beta-*N*-acetylglucosaminidase in the *Aedes aegypti* chitinolytic system involving peritrophic matrix formation and degradation. *Insect Biochem. Mol. Biol. 32*, 1723–1729.

Fly Base, **2004**. http://www.flybase.bio.indiana.edu/

Freimoser, F.M., Screen, S., Bagga, S., Hu, G., St. Leger, R.J., **2003a**. Expressed sequence tag (EST) analysis of two subspecies of *Metarhizium anisopliae* reveals a plethora of secreted proteins with potential activity in insect hosts. *Microbiology 149*, 239–247.

Freimoser, F.M., Screen, S., Hu, G., St. Leger, R.J., **2003b**. EST analysis of genes expressed by the zygomycete pathogen *Conidiobolus coronatus* during growth on insect cuticle. *Microbiology 149*, 1893–1900.

Fuhrman, J.A., Lane, W.S., Smith, R.F., Piessens, W.F., Perler, F.B., **1992**. Transmission-blocking antibodies recognize microfilarial chitinase in brugian lymphatic filariasis. *Proc. Natl Acad. Sci. USA 89*, 1548–1552.

Fukamizo, T., **2000**. Chitinolytic enzymes: catalysis, substrate binding, and their application. *Curr. Protein Peptide Sci. 1*, 105–124.

Fukamizo, T., Koga, D., Goto, S., **1995**. Comparative biochemistry of chitinases: anomeric form of the reaction products. *Biosci. Biotech. Biochem. 59*, 311–313.

Fukamizo, T., Kramer, K.J., **1985**. Mechanism of chitin hydrolysis by the binary chitinase system in insect moulting fluid. *Insect Biochem. 15*, 141–145.

Fukamizo, T., Kramer, K.J., **1987**. Effect of 20-hydroxyecdysone on chitinase and β-*N*-acetylglucosaminidase during the larval-pupal transformation in *Manduca sexta* (L.). *Insect Biochem. 17*, 547–550.

Fukamizo, T., Kramer, K.J., Mueller, D.D., Schaefer, J., Garbow, J., *et al.*, **1986**. Analysis of chitin structure by nuclear magnetic resonance spectroscopy and chitinolytic enzyme digestion. *Arch. Biochem. Biophys. 249*, 15–26.

Furukawa, S., Taniai, K., Yang, J., Shono, T., Yamakawa, M., **1999**. Induction of gene expression of antibacterial proteins by chitin oligomers in the silkworm, *Bombyx mori*. *Insect Mol. Biol. 8*, 145–148.

Fusetti, F., von Moeller, H., Houston, D., Rozeboom, H.J., Dijkstra, B.W., *et al.*, **2002**. Structure of human chitotriosidase: implications for specific inhibitor design and function of mammalian chitinase-like lectins. *J. Biol. Chem. 277*, 25537–25544.

Gagou, M.E., Kapsetaki, M., Turberg, A., Kafetzopoulos, D., **2002**. Stage-specific expression of the chitin synthase *DmeChSA* and *DmeChSB* genes during the onset of *Drosophila* metamorphosis. *Insect Biochem. Mol. Biol. 32*, 141–146.

Gaines, P.J., Brandt, K.S., Eisele, A.M., Wagner, W.P., Bozic, C.M., *et al.*, **2002**. Analysis of expressed

sequence tags from *Ctenocephalides felis* hindgut and Malpighian tubule cDNA libraries. *Insect Mol. Biol. 11*, 299–306.

Gibbons, B.J., Roach, P.J., Hurley, T.D., **2002**. Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin. *J. Mol. Biol. 319*, 463–477.

Gilkes, N.R., Henrissat, B., Kilburn, D.G., Miller, R.C., Warren, R.A.J., **1991**. Domains in microbial β-1,4-glycanases: sequence conservation, function, and enzyme families. *Microbiol. Rev. 55*, 303–315.

Gill, J., Rixon, J.E., Bolam, D.N., McQueen-Mason, S., Simpson, P.J., *et al.*, **1999**. The type II and X cellulose-binding domains of *Pseudomonas* xylanase A potentiate catalytic activity against complex substrates by a common mechanism. *Biochem. J. 342*, 473–480.

Girard, C., Jouanin, L., **1999**. Molecular cloning of a gut-specific chitinase cDNA from the beetle *Phaedon cochleariae*. *Insect Biochem. Mol. Biol. 29*, 549–556.

Gongora, C.E., Broadway, R.M., **2002**. Plant growth and development influenced by transgenic insertion of bacterial chitinolytic enzymes. *Mol. Breeding 9*, 123–135.

Gongora, C.E., Wang, S., Barbehenn, R.V., Broadway, R.M., **2001**. Chitinolytic enzymes from *Streptomyces albidoflavus* expressed in tomato plants: effects on *Trichoplusia ni*. *Entomol. Experimentia Appl. 99*, 193–204.

Goo, T.-W., Hwang, J.-S., Sung, G.-B., Yun, E.-Y., Bang, H.-S., *et al.*, **1999**. Molecular cloning and characterization of a gene encoding a beta-*N*-acetylglucosaminidase homologue from *Bombyx mandarina*. *Korean J. Sericult. Sci. 41*, 147–153.

Gooday, G.W., **1999**. Aggressive and defensive roles for chitinases. In: Jolles, P., Muzzarelli, R.A.A. (Eds.), Chitin and Chitinases. Birkhäuser Verlag, Basel, pp. 157–169.

Gopalakrishnan, B., Muthukrishnan, S., Kramer, K.J., **1995**. Baculovirus-mediated expression of a *Manduca sexta* chitinase gene: properties of the recombinant protein. *Insect Biochem. Mol. Biol. 25*, 255–265.

Graack, H., Cinque, U., Kress, H., **2001**. Functional regulation of glutamine: fructose-6-phosphate aminotransferase 1 (GFAT1) of *Drosophila melanogaster* in a UDP-*N*-acetylglucosamine and cAMP-dependent manner. *Biochem. J. 360*, 401–412.

Gunner, H.P., Met, M.Z., Berger, S., **1985**. In: Gamble, D.G., Lewis, F.B. (Eds.), Microbial Control of Spruce Budworms. US Forest Service GTR-NE-100, Northeastern Forest Experiment Station, Broomall, PA, p. 102.

Hamodrakas, S.J., Willis, J.H., Iconomidou, V.A., **2002**. A structural model of the chitin-binding domain of cuticle proteins. *Insect Biochem. Mol. Biol. 32*, 1577–1583.

Harper, M.S., Granados, R.R., **1999**. Peritrophic membrane structure and formation of larval *Trichoplusia ni* with an investigation on the secretion patterns of a PM mucin. *Tissue and Cell 31*, 201–211.

Harper, M.S., Hopkins, T.L., **1997**. Peritrophic membrane structure and secretion in European corn borer larvae (*Ostrinia nubilalis*). *Tissue and Cell 29*, 461–475.

Harper, M.S., Hopkins, T.L., Czapla, T.H., **1998**. Effect of wheat germ agglutinin on formation and structure of the peritrophic membrane in European corn borer (*Ostrinia nubilalis*) larvae. *Tissue and Cell 30*, 166–176.

Harris, M.T., Fuhrman, J.A., **2002**. Structure and expression of chitin synthase in the parasitic nematode *Dirofilaria immitis*. *Mol. Biochem. Parasitol. 122*, 231–234.

Harris, M.T., Lai, K., Arnold, K., Martinez, H.F., Specht, C.A., *et al.*, **2000**. Chitin synthase in the filarial parasite, *Brugia malayi*. *Mol. Biochem. Parasitol. 111*, 351–362.

Hawtin, R.E., Zarkowska, T., Arnold, K., Thomas, C.J., Gooday, G.W., *et al.*, **1997**. Liquefaction of *Autographa californica* nucleopolyhedrosis-infected insects is dependent on the integrity of virus-encoded chitinase and cathepsin genes. *Virology 238*, 243–253.

Hefner, T., Stockigt, J., **2003**. Probing suggested catalytic domains of glycosyltransferases by site-directed mutagenesis. *Eur. J. Biochem. 270*, 533–538.

Hemmi, H., Ishibashi, J., Tomie, T., Yamakawa, M., **2003**. Structural basis for new pattern of conserved amino acid residues related to chitin binding in the antifungal peptide from the coconut rhinoceros beetle *Oryctes rhinoceros*. *J. Biol. Chem. 278*, 22820–22827.

Henrissat, B., **1999**. Classification of chitinase modules. In: Jolles, P., Muzzarelli, R.A.A. (Eds.), Chitin and Chitinases. Birkhäuser Verlag, Basel, pp. 137–156.

Henrissat, B., Bairoch, A., **1996**. Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J. 316*, 695–696.

Honda, Y., Kitaoka, M., Tokuyasu, K., Sasaki, C., Fukamizo, T., *et al.*, **2003**. Kinetic studies on the hydrolysis of *N*-acetylated and *N*-deacetylated derivatives of 4-methylumbelliferyl chitobioside by the family 18 chitinases ChiA and ChiB from *Serratia marcescens*. *J. Biochem. (Tokyo) 133*, 253–258.

Hopkins, T.L., Harper, M.S., **2001**. Lepidopteran peritrophic membranes and effects on dietary wheat germ agglutinin on their formation and structure. *Arch. Insect Biochem. Physiol. 47*, 100–109.

Hopkins, T.L., Kramer, K.J., **1992**. Insect cuticle sclerotization. *Annu. Rev. Entomol. 37*, 273–302.

Horst, M.N., **1983**. The biosynthesis of crustacean chitin: isolation and characterization of polyprenol-linked intermediates from brine shrimp microsomes. *Arch. Biochem. Biophys. 223*, 254–263.

Houston, D.R., Reeklies, A.D., Krupa, J.C., van Aalten, D.M.F., **2003**. Structure and ligand-induced conformational change of the 39-kDa glycoprotein from human articular chondrocytes. *J. Biol. Chem. 278*, 30206–30212.

Houston, D.R., Shiomi, K., Arai, N., Omura, S., Peter, M.G., *et al.*, **2002**. High-resolution structures of a chitinase complexed with natural product cyclopentapeptide inhibitors: mimicry of carbohydrate substrate. *Proc. Natl Acad. Sci. USA 99*, 9127–9132.

Hsu, T.A., Takahashi, N., Tsukamoto, Y., Kato, K., Shimada, I., *et al.*, **1997**. Differential *N*-glycan patterns of secreted and intracellular IgG produced in *Trichoplusia ni* cells. *J. Biol. Chem. 272*, 9062–9070.

Huang, X., Zhang, H., Zen, K.C., Muthukrishnan, S., Kramer, K.J., **2000**. Homology modeling of the insect chitinase catalytic domain-oligosaccharide complex and the role of a putative active site tryptophan in catalysis. *Insect Biochem. Mol. Biol. 30*, 107–117.

Ibrahim, G.H., Smartt, C.T., Kiley, L.M., Christensen, B.M., **2000**. Cloning and characterization of a chitin synthase cDNA from the mosquito *Aedes aegypti*. *Insect Biochem. Mol. Biol. 30*, 1213–1222.

Ishaaya, I., **2001**. Biochemical processes related to insecticide action: an overview. In: Ishaaya, I. (Ed.), Biochemical Sites of Insecticide Action and Resistance. Springer-Verlag, New York, pp. 1–16.

Jarvis, D.L., Finn, E.E., **1995**. Biochemical analysis of the *N*-glycosylation pathway in baculovirus-infected lepidopteran insect cells. *Virology 212*, 500–511.

Johansen, J.S., Christensen, I.J., Riisbro, R., Greenall, M., Han, C., *et al.*, **2003**. High serum YKL-40 levels in patients with primary breast cancer is related to short recurrence free survival. *Breast Cancer Res. and Treatmt 80*, 15–21.

Jurgens, G., Wieschaus, E., Nusslein-Volhard, C., Kluding, H., **1984**. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster.* 2. Zygotic loci on the third chromosome. *Roux's Arch. Devel. Biol. 193*, 283–295.

Kanost, M.R., Zepp, M.K., Ladendorff, N.E., Andersson, L.A., **1994**. Isolation and characterization of a hemocyte aggregation inhibitor from hemolymph of *Manduca sexta* larvae. *Arch. Insect Biochem. Physiol. 27*, 123–136.

Kato, N., Dasgupta, R., Smartt, C.T., Christensen, B.M., **2002**. Glucosamine: fructose-6-phosphate aminotransferase: gene characterization, chitin biosynthesis and peritrophic matrix formation in *Aedes aegypti*. *Insect Mol. Biol. 11*, 207–216.

Khor, E., **2002**. Chitin: a biomaterial in waiting. *Curr. Opin. Solid State Materials Sci. 2*, 313–317.

Kim, M.G., Shin, S.W., Bae, K.S., Kim, S.C., Park, H.Y., **1998**. Molecular cloning of chitinase cDNAs from the silkworm, *Bombyx mori*, and the fall webworm, *Hyphantria cunea*. *Insect. Biochem. Mol. Biol. 28*, 163–171.

Kimura, S., **1976**. Insect haemolymph exo-β-*N*-acetylglucosaminidase from *Bombyx mori*. *Biochim. Biophys. Acta 446*, 399–406.

Kimura, S., **1977**. Exo-β-*N*-acetylglucosaminidase and chitobiase in *Bombyx mori*. *Insect Biochem. 7*, 237–245.

Kirkpatrick, R.B., Matico, R.E., McNulty, D.E., Strickler, J.E., Rosenberg, M., **1995**. An abundantly secreted glycoprotein from *Drosophila melanogaster* is related to mammalian secretory proteins produced in rheumatoid tissues and by activated macrophages. *Gene 153*, 147–154.

Kleywegt, G.J., Zou, J.Y., Kjeldgaard, M., Jones, T.A., **2001**. In: Around, O., Rossmann, M.G., Arnold, E. (Eds.), International Tables for Crystallography. vol. F. Kluwer Academic, Dordrecht, The Netherlands, pp. 353–356.

Koga, D., Funakoshi, T., Fujimoto, H., Kuwano, E., Eto, M., *et al.*, **1991**. Effects of 20-hydroxyecdysone and KK-42 on chitinase and beta-*N*-acetylglucosaminidase during the larval–pupal transformation of *Bombyx mori*. *Insect Biochem. 21*, 277–284.

Koga, D., Funakoshi, T., Mizuki, K., Ide, A., Kramer, K.J., *et al.*, **1992**. Immunoblot analysis of chitinolytic enzymes in integument and molting fluid of the silkworm, *Bombyx mori* and the tobacco hornworm, *Manduca sexta*. *Insect Biochem. Mol. Biol. 22*, 305–311.

Koga, D., Isogai, A., Sakuda, S., Matsumoto, S., Suzuki, A., *et al.*, **1987**. Specific inhibition of *Bombyx mori* chitinase by allosamidin. *Agric. Biol. Chem. 51*, 471–476.

Koga, D., Jilka, J., Kramer, K.J., **1983**. Insect endochitinases: glycoproteins from moulting fluid, integument and pupal haemolymph. *Insect Biochem. 13*, 295–305.

Koga, D., Mai, M.S., Dziadik-Turner, C., Kramer, K.J., **1982**. Kinetics and mechanism of exochitinase and β-*N*-acetylglucosaminidase from the tobacco hornworm, *Manduca sexta* L. (*Lepidoptera: Sphingidae*). *Insect Biochem. 12*, 493–499.

Koga, D., Mitsutomi, M., Kono, M., Matsumiya, M., **1999**. Biochemistry of chitinases. In: Jolles, P., Muzzarelli, R.A.A. (Eds.), Chitin and Chitinases. Birkhäuser Verlag, Basel, pp. 111–123.

Koga, D., Sasaki, Y., Uchiumi, Y., Hirai, N., Arakane, Y., *et al.*, **1997**. Purification and characterization of *Bombyx mori* chitinase. *Insect Biochem. Mol. Biol. 27*, 757–767.

Kramer, K.J., Corpuz, L., Choi, H., Muthukrishnan, S., **1993**. Sequence of a cDNA and expression of the genes encoding epidermal and gut chitinases of *Manduca sexta*. *Insect Biochem. Mol. Biol. 23*, 691–701.

Kramer, K.J., Dziadik-Turner, C., Koga, D., **1985**. Chitin metabolism in insects. In: Kerkut, G.A., Gilbert, L.I. (Eds.), Comprehensive Insect Physiology, Biochemistry, and Pharmacology, vol. 3. Pergamon Press, Oxford, pp. 75–115.

Kramer, K.J., Hopkins, T.L., Schaefer, J., **1988**. Insect cuticle structure and metabolism. *Adv. Chem. 379*, 160–185.

Kramer, K.J., Hopkins, T.L., Schaefer, J., **1995**. Applications of solids NMR to the analysis of insect sclerotized structures. *Insect Biochem. Mol. Biol. 25*, 1067–1080.

Kramer, K.J., Koga, D., **1986**. Insect chitin: physical state, synthesis, degradation and metabolic regulation. *Insect Biochem. 16*, 851–877.

Kramer, K.J., Muthukrishnan, S., **1997**. Insect chitinases: molecular biology and potential use as biopesticides. *Insect Biochem. Mol. Biol. 27*, 887–900.

Kramer, K.J., Muthukrishnan, S., Johnson, L., White, F., **1997**. Chitinase for insect control. In: Carozzi, N., Koziel, M. (Eds.), Advances in Insect Control: The Role of Transgenic Plants. Taylor and Francis, London, pp. 185–193.

Krieger de Moraes, C., Schrank, A., Vainstein, M.H., **2003**. Regulation of extracellular chitinases and proteases in the entomopathogen and acaricide *Metarhizium anisopliae*. *Curr. Microbiol. 46*, 205–210.

Krishnan, A., Nair, P.N., Jones, D., **1994**. Isolation, cloning, and characterization of new chitinase stored in active form in chitin-lined venom reservoir. *J. Biol. Chem. 269*, 20971–20976.

Kubelka, V., Altmann, F., Marz, L., **1995**. The asparagine-linked carbohydrate of honeybee venom hyaluronidase. *Glycoconjugate J. 12*, 77–83.

Kuranda, M.J., Robbins, P.W., **1991**. Chitinase is required for cell separation during growth of *Saccharomyces cerevisiae*. *J. Biol. Chem. 266*, 19758–19767.

Lanfranco, L., Balsamo, R., Martino, E., Perotto, S., Bonfante, P., **2002**. Zinc ions alter morphology and chitin deposition in an ericoid fungus. *Eur. J. Histochem. 46*, 341–350.

Langsford, M.L., Gilkes, N.R., Singh, B., Moser, B., Miller, R.C., *et al.*, **1987**. Glycosylation of bacterial cellulases prevents proteolytic cleavage between functional domains. *FEBS Lett. 25*, 163–167.

Lehtio, J., Sugiyama, J., Gustavsson, M., Fransson, L., Linder, M., *et al.*, **2003**. The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. *Proc. Natl Acad. Sci. USA 100*, 484–489.

Limón, M.C., Margolles-Clark, E., Benítez, T., Penttilä, M., **2001**. Addition of substrate-binding domains increases substrate-binding capacity and specific activity of a chitinase from *Trichoderma harzianum*. *FEMS Microbiol. Lett. 198*, 57–63.

Linder, M., Salovuori, I., Ruohonen, L., Teeri, T.T., **1996**. Characterization of a double cellulose-binding domain: synergistic high affinity binding to crystalline cellulose. *J. Biol. Chem. 271*, 21268–21272.

Linder, M., Teeri, T.T., **1997**. The roles and function of cellulose-binding domains. *J. Biotechnol. 57*, 15–28.

Liu, M., Cai, Q.X., Liu, H.Z., Zhang, B.H., Yan, J.P., *et al.*, **2002**. Chitinolytic activities in *Bacillus thuringiensis* and their synergistic effects on larvicidal activity. *J. Appl. Microbiol. 93*, 374–379.

Locke, M., Huie, P., **1979**. Apolysis and the turnover of plasma membrane plaques during cuticle formation in an insect. *Tissue and Cell 11*, 277–291.

Lu, Y., Zen, K.C., Muthukrishnan, S., Kramer, K.J., **2002**. Site-directed mutagenesis and functional analysis of active site acidic amino acid residues D142, D144 and E146 in *Manduca sexta* (tobacco hornworm) chitinase. *Insect Biochem. Mol. Biol. 32*, 1369–1382.

Lucero, H.A., Kuranda, M.J., Bulik, D.A., **2002**. A nonradioactive, high throughput assay for chitin synthase activity. *Anal. Biochem. 305*, 97–105.

Lysenko, O., **1976**. Chitinase of *Serratia marcescens* and its toxicity to insects. *J. Invert. Pathol. 27*, 385–386.

Macedo, M.L.R., Andrade, L.B.S., Moraes, R.A., Xavier-Filho, J., **1993**. Vicilin variants and the resistance of cowpea (*Vigna unguiculata*) seeds to the cowpea weevil (*Callosobruchus maculatus*). *Comp. Biochem. Physiol. C 105*, 89–94.

Macedo, M.L.R., Damico, D.C.S., Freire, M. D. G. M., Toyama, M.H., Marangoni, S., *et al.*, **2003**. Purification and characterization of an *N*-acetylglucosamine-binding lectin from *Koelreuteria paniculata* seeds and its effect on the larval development of *Callosobruchus maculatus* (Coleoptera: Bruchidae) and *Anagasta kuehniella* (Lepidoptera: Pyralidae). *J. Agric. Food Chem. 51*, 2980–2986.

McCall, C., Hunter, S., Stedman, K., Weber, E., Hillier, A., *et al.*, **2001**. Characterization and cloning of a major high molecular weight house dust mite allergen (Der f 15) for dogs. *Veter. Immunol. Immunopathol. 78*, 231–247.

McLean, B.W., Boraston, A.B., Brouwer, D., Sanaie, N., Fyfe, C.A., *et al.*, **2002**. Carbohydrate-binding modules recognize fine substructures of cellulose. *J. Biol. Chem. 277*, 50245–50254.

Mikitani, K., Sugasaki, T., Shimada, T., Kobayashi, M., Gustafsson, J.A., **2000**. The chitinase gene of the silkworm, *Bombyx mori*, contains a novel Tc-like transposable element. *J. Biol. Chem. 275*, 37725–37732.

Mikolajczyk, P., Oberlander, H., Silhacek, D., Ishaaya, I., Shaaya, E., **1994**. Chitin synthesis in *Spodoptera frugiperda* wing imaginal discs. 1. Chlorfluzuron, diflubenzuron, and teflubenzuron inhibit incorporation but not uptake of [^{14}C]-*N*-acetyl-D-glucosamine. *Arch. Insect Biochem. Physiol. 25*, 245–258.

Morlais, I., Severson, D.W., **2001**. Identification of a polymorphic mucin-like gene expressed in the midgut of the mosquito, *Aedes aegypti*, using an integrated bulked segregant and differential display analysis. *Genetics 158*, 1125–1136.

Morris, O.N., **1976**. A 2-year study of the efficacy of *Bacillus thuringiensis*–chitinase combinations in spruce budworm (*Choristoneura fumiferana*) control. *Can. Entomol. 108*, 225.

Mota, A.C., DaMatta, R.A., Filho, M.L., Silva, C.P., Xavier-Filho, J., **2003**. Cowpea (*Vigna unguiculata*) vicilins bind to the peritrophic membrane of larval sugarcane stalk borer (*Diatraea saccharalis*). *J. Insect Physiol. 49*, 873–880.

Munrow, C.A., Gow, N.A., **2001**. Chitin synthesis in human pathogenic fungi. *Med. Mycol. 39*(Suppl. 1), 41–53.

Muzzarelli, C., Muzzarelli, R.A.A., **2002**. Reactivity of quinones towards chitosans. Trends Glycosci. *Glycotech. 14*, 223–229.

Muzzarelli, R.A.A., Littarru, G., Muzzarelli, C., Tosi, G., **2003**. Selective reactivity of biochemically relevant quinones towards chitosans. *Carbohydrate Polymers 53*, 109–115.

Nagahashi, S., Sudoh, M., Ono, N., Sawada, R., Yamaguchi, E., *et al.*, **1995**. Characterization of chitin synthase 2 of *Saccharomyces cereviseae*: implications of two highly conserved domains as possible catalytic sites. *J. Biol. Chem. 270*, 13961–13967.

Nagamatsu, Y., Yanagisawa, I., Kimoto, M., Okamoto, M., Koga, D., **1995**. Purification of a chitooligosaccharidolytic β-*N*-acetylglucosaminidase from *Bombyx mori* larvae during metamorphosis and the nucleotide sequence of its cDNA. *Biosci. Biotech. Biochem. 59*, 219–225.

Nagano, N., Orengo, C.A., Thornton, J.M., **2002**. One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J. Mol. Biol. 321*, 741–765.

Nagano, N., Porter, C.T., Thornton, J.M., **2001**. The $(\beta\alpha)_8$ glycosidases: sequence and structure analyses suggest distant evolutionary relationships. *Protein Eng. 14*, 845–855.

Neuhaus, J.M., **1999**. Plant chitinases (PR-3, PR-4, PR-8, PR-11). In: Datta, S.K., Muthukrishnan, S. (Eds.), Pathogenesis-Related Proteins in Plants. CRC Press, New York, pp. 77–106.

Nichols, E.J., Beckman, J.M., Hadwiger, L.A., **1980**. Glycosidic enzyme activity in pea tissue and pea–*Fusariun solani* interactions. *Plant Physiol. 66*, 199–204.

Nusslein-Volhard, C., Wieschaus, E., Kluding, H., **1984**. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster.* 1. Zygotic loci on the second chromosome. *Roux's Arch. Devel. Biol. 193*, 267–282.

Oberlander, H., Silhacek, D., **1998**. New perspectives on the mode of action of benzoylphenylurea insecticides. In: Ishaaya, I., Degheele, D. (Eds.), Insecticides with Novel Modes of Action: Mechanism and Application. Springer-Verlag, New York, pp. 92–105.

Ostrowski, S., Dierick, H.A., Bejsovec, A., **2002**. Genetic control of cuticle formation during embryonic development of *Drosophila melanogaster. Genetics 161*, 171–182.

Otsu, Y., Mori, H., Komuta, K., Shimizu, H., Nogawa, S., *et al.*, **2003**. Suppression of leaf feeding and oviposition of phytophagous ladybird beetles (Coleoptera: Coccinellidae) by chitinase gene-transformed phylloplane bacteria and their specific bacteriophages entrapped in alginate gel beads. *J. Econ. Entomol. 96*, 555–563.

Palli, S.R., Retnakaran, A., **1999**. Molecular and biochemical aspects of chitin synthesis inhibition. In: Jolles, P, Muzzarelli, R.A.A. (Eds.), Chitin and Chitinases. Birkhäuser Verlag, Basel, pp. 85–98.

Patil, R.S., Ghormade, V., Deshpande, M.V., **2000**. Chitinolytic enzymes: an exploration. *Enzyme Microbiol. Technol. 26*, 473–483.

Paulson, J.C., **1989**. Glycoproteins: what are the sugar chains for? *Trends Biochem. Sci. 14*, 272–276.

Perchan, T., Cohen, A., Williams, W.P., Luthe, D.S., **2002**. Insect feeding mobilizes a unique plant defense protease that disrupts the peritrophic matrix of caterpillars. *Proc. Natl Acad. Sci. USA 99*, 13319–13323.

Perrakis, A., Tews, I., Dauter, Z., Oppenheim, A.B., Chet, I., *et al.*, **1994**. Crystal structure of a bacterial chitinase at 2.3 Å resolution. *Structure 2*, 1169–1180.

Quentin, M., Ebbelaar, M., Derksen, J., Mariani, C., Van der Valk, H., **2002**. Description of a cellulose-binding domain and a linker sequence from *Aspergillus* fungi. *Appl. Microbiol. Biotechnol. 58*, 658–662.

Quesada-Allue, L.A., **1982**. The inhibition of insect chitin biosynthesis by tunicamycin. *Biochem. Biophys. Res. Commun. 105*, 312–319.

Raikhel, N.V., Lee, H.-I., Broekaert, W.F., **1993**. Structure and function of chitin-binding proteins. *Annu. Rev. Plant Physiol. Plant Mol. Biol. 44*, 591–615.

Ramalho-Ortigão, J.M., Traub-Csekö, Y.M., **2003**. Molecular characterization of *Llchit1*, a midgut chitinase cDNA from the leishmaniasis vector *Lutzomyia longipalpis*. *Insect Biochem. Mol. Biol. 33*, 279–287.

Rao, F.V., Houston, D.R., Boot, R.G., Aerts, J. M. F. G., Sakuda, S., van Aalten, D.M.F., **2003**. Crystal structures of allosamidin derivatives in complex with human macrophage chitinase. *J. Biol. Chem. 278*, 20110–20116.

Rayms-Keller, A., McGaw, M., Oray, C., Carlson, J.O., Beaty, B.J., **2000**. Molecular cloning and characterization of a metal responsive *Aedes aegypti* mucin cDNA. *Insect Mol. Biol. 9*, 419–426.

Read, S.M., Bacic, T., **2002**. Prime time for cellulose. *Science 295*, 59–60.

Rebers, J.E., Willis, J.H., **2001**. A conserved domain in arthropod cuticular proteins binds chitin. *Insect Biochem. Mol. Biol. 31*, 1083–1093.

Receveur, V., Czjzek, M., Schülein, M., Panine, P., Heinrissat, B., **2002**. Dimension, shape, and conformational flexibility of a two domain fungal cellulase in solution probed by small angle X-ray scattering. *J. Biol. Chem. 277*, 40887–40892.

Regev, A., Keller, M., Strizhov, N., Sneh, B., Prudovsky, E., *et al.*, **1996**. Synergistic activity of a *Bacillus thuringiensis* delta-endotoxin and a bacterial endochitinase against *Spodoptera littoralis* larvae. *Appl. Environ. Microbiol. 62*, 3581–3586.

Renkema, G.H., Boot, R.G., Au, F.L., Donker-Koopman, W.E., Strijland, A., *et al.*, **1998**. Chitotriosidase, a chitinase, and the 39-kDa human cartilage glycoprotein, a chitin-binding lectin, are homologues of family 18 glycosyl hydrolases secreted by human macrophages. *Eur. J. Biochem. 251*, 504–509.

Richmond, T., **2000**. Higher plant cellulose synthases. *Genome Biol. 1*, 3001–3006.

Rogers, S., Wells, R., Rechsteiner, M., **1986**. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science 234*, 364–368.

Roseland, C.R., Grodowitz, M.J., Kramer, K.J., Hopkins, T.L., Broce, A.B., **1985**. Stabilization of mineralized and sclerotized puparial cuticle of muscid flies. *Insect Biochem. 15*, 521–528.

Royer, V., Fraichard, S., Bouhin, H., **2002**. A novel putative insect chitinase with multiple catalytic domains: hormonal regulation during metamorphosis. *Biochem. J. 366*, 921–928.

Saito, A., Fujii, T., Miyashita, K., **2003**. Distribution and evolution of chitinase genes in *Streptomyces* species: involvement of gene-duplication and domain-deletion. *Antonie van Leeuwenhoek 84*, 7–16.

Saitou, N., Nei, M., **1987**. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol. 4*, 406–425.

Sales, M.P., Pimenta, P.P., Paes, N.S., Grossi-de-Sá, M.F., Xavier-Filho, J., **2001**. Vicilins (7S storage globulins) of cowpea (*Vigna unguiculata*) seeds bind to chitinous structures of the midgut of *Callosobruchus maculatus* (Coleoptera: Bruchidae) larvae. *Brazil. J. Med. Biol. Res. 34*, 27–34.

Sampson, M.N., Gooday, G.W., **1998**. Involvement of chitinases of *Bacillus thuringiensis* during pathogenesis in insects. *Microbiology 144*, 2189–2194.

Sarauer, B.L., Gillott, C., Hegedus, D., **2003**. Characterization of an intestinal mucin from the peritrophic matrix of the diamondback moth, *Plutella xylostella*. *Insect Mol. Biol. 12*, 333–343.

Schaefer, J., Kramer, K.J., Garbow, J.R., Jacob, G.S., Stejskal, E.O., *et al.*, **1987**. Aromatic cross-links in insect cuticle: detection by solid-state ^{13}C and ^{15}N NMR. *Science 235*, 1200–1204.

Screen, S.E., Hu, G., St. Leger, R.J., **2001**. Transformants of *Metarhizium anisopliae* sf. *anisopliae* over expressing chitinase from *Metarhizium anisopliae* sf. *acridum* show early induction of native chitinase but are not altered in pathogenicity to *Manduca sexta*. *J. Invert. Pathol. 78*, 260–266.

Shapiro, M., Preisler, H.K., Robertson, J.L., **1987**. Enhancement of baculovirus activity on gypsy moth (Lepidoptera: Limantidae) by chitinase. *J. Econ. Entomol. 80*, 1113–1115.

Shen, Z., Jacobs-Lorena, M., **1997**. Characterization of a novel gut-specific gene from the human malaria vector *Anopheles gambiae*. *J. Biol. Chem. 272*, 28895–28900.

Shen, Z., Jacobs-Lorena, M., **1999**. Evolution of chitin-binding proteins in invertebrates. *J. Mol. Evol. 48*, 341–347.

Shinoda, T., Kobayashi, J., Matsui, M., Chinzei, Y., **2001**. Cloning and functional expression of a chitinase cDNA from the common cutworm, *Spodoptera litura*, using a baculovirus lacking the virus-encoded chitinase gene. *Insect Biochem. Mol. Biol. 31*, 521–532.

Shutov, A.D., Kakhovskaya, I.A., Braun, H., Bäumlein, H., Müntz, K., **1995**. Legumin-like and vicilin-like seed storage proteins: evidence for a common single-domain ancestral gene. *J. Mol. Evol. 41*, 1057–1069.

Smirnoff, W.A., Valero, R., **1972**. Perturbations métaboliques chez *Choristoneura fumiferana* Clemens au cours de l'infection par *Bacillus thuringiensis* seul ou en présence de chitinase. *Rev. Con. Biol. 31*, 163.

Sneh, B., Schuster, S., Gross, S., **1983**. Biological control of *Spodoptera littoralis* (Boisd.) (Lepidoptera: Noctuidae) by *Bacillus thuringiensis* subsp. *entomocidus* and *Bracon hebetor* Say (Hymenoptera:Braconidae). *Z. Ang. Entomol. 96*, 77–83.

Spindler, K.D., Spindler-Barth, M., **1999**. Inhibitors of chitinases. In: Jolles, P., Muzzarelli, R.A.A. (Eds.), Chitin and Chitinases. Birkhäuser Verlag, Basel, pp. 201–209.

Srisodsuk, M., Reinikainen, T., Pentilä, M., Teeri, T., **1993**. Role of the interdomain linker peptide of *Trichoderma reesei* cellobiohydrolase I in its interaction with crystalline cellulose. *J. Biol. Chem. 268*, 20756–20761.

St. Leger, R.J., Joshi, L., Bidochka, M.L., Rizzo, N.W., Roberts, D.W., **1996**. Characterization and ultrastructural localization of chitinases from *Metarhizium anisopliae*, *M. flavoride*, and *Beauveria bassiana* during fungal invasion of host (*Manduca sexta*) cuticle. *Appl. Environ. Microbiol. 62*, 907–912.

Suetake, T., Tsuda, S., Kawabata, S., Miura, K., Iwanaga, S., *et al.*, **2000**. Chitin-binding proteins in invertebrates and plants comprise a common chitin-binding motif. *J. Biol. Chem. 275*, 17929–17932.

Sun, Y.J., Chang, N.C., Hung, S.I., Chang, A.C., Chou, C.C., *et al.*, **2001**. The crystal structure of a novel mammalian lectin, Ym1, suggests a saccharide-binding site. *J. Biol. Chem. 276*, 17507–17514.

Sundara Rajulu, G., Aruchami, A., Gowri, N., **1982**. Natural deacetylation of chitin to chitosan in the abdominal cuticle of the physogastric queen of *Macrotermetes estherae*. In: Proc. 2nd Int. Conf. Chitin and Chitosan, cited in Gooday, G. W., 1985. Chitin deacetylases in invertebrates. In: Muzzerelli, R., Jeuniaux, C., Gooday, G.W. (Eds.), Chitin in Nature and Technology. Plenum Press, New York, pp. 263–268.

Suzuki, K., Taiyoji, M., Sugawara, N., Nikaidou, N., Hernissat, B., *et al.*, **1999**. The third chitinase gene (*ChiC*) of *Serratia marcescens* 2170 and the relationship of its product to other bacterial chitinases. *Biochem. J. 343*, 587–596.

Takahashi, M., Kiuchi, M., Kamimura, M., **2002**. A new chitinase-related gene, *BmChiR1*, is induced in the *Bombyx mori* anterior silk gland at molt and metamorphosis by ecdysteroid. *Insect Biochem. Mol. Biol. 32*, 147–151.

Tantimavanich, S., Pantuwatana, S., Bhumiratana, A., Panbangred, W., **1997**. Cloning of a chitinase gene into *Bacillus thuringiensis* subsp. *aizawai* for enhanced insecticidal activity. *J. Gen. Appl. Microbiol. 43*, 341–347.

Tellam, R.L., **1996**. Protein motifs in filarial chitinases: an alternative view. *Parasitol. Today 12*, 291–292.

Tellam, R.L., Vuocolo, T., Eisemann, C., Briscoe, S., Riding, G., *et al.*, **2003**. Identification of an immunoprotective mucin-like protein, peritrophin-55, from the peritrophic matrix of *Lucilia cuprina* larvae. Insect Biochem. *Mol. Biol. 33*, 239–252.

Tellam, R.L., Vuocolo, T., Johnson, S.E., Jarmey, J., Pearson, R.D., **2000**. Insect chitin synthase cDNA sequence, gene organization and expression. *Eur. J. Biochem. 267*, 6025–6043.

Tellam, R.L., Wijffels, G., Willadsen, P., **1999**. Peritrophic matrix proteins. *Insect Biochem. Mol. Biol. 29*, 87–101.

Tharanathan, R.N., Kittur, F.S., **2003**. Chitin: the undisputed biomolecule of great potential. *Crit. Rev. Food Sci. Nutri. 43*, 61–87.

Thomas, C.J., Gooday, G.W., King, L.A., Possee, R.D., **2000**. Mutagenesis of the active site coding region of the *Autographa californica* nucleopolyhedrovirus *chiA* gene. *J. Gen. Virol. 81*, 1403–1411.

Thompson, J.D., Higgins, D.G., Gibson, T.J., **1994**. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res. 22*, 4673–4680.

Thummel, C.S., **1996**. Flies on steroids: *Drosophila* metamorphosis and the mechanisms of steroid hormone action. *Trends Genet. 12*, 306–310.

Tjoelker, L.W., Gosting, L., Frey, S., Hunter, C.L., Trong, H.L., *et al.*, **2000**. Structural and functional definition of the human chitinase chitin-binding domain. *J. Biol. Chem. 275*, 514–520.

Tomme, P., Warren, R.A., Miller, R.C., Kilburn, D.G., Gilkes, N.R., **1995**. In: Saddler, J.N., Penner, M. (Eds.), Enzymatic Degradation of Insoluble Polysaccharides. American Chemical Society, Washington, DC, pp. 142–163.

Tsigos, I., Martinou, A., Kafetzopoulos, D., Bouriotis, V., **2000**. Chitin deacetylases: new, versatile tools in biotechnology. *Trends Biotechnol. 18*, 305–312.

Turner, C.D., Koga, D., Mai, M.S., Kramer, K.J., **1981**. Purification and characterization of two β-*N*-acetylhexosaminidases from the tobacco hornworm, *Manduca sexta* (L.) (Lepidoptera: Sphingidae). *Arch. Biochem. Biophys. 212*, 546–560.

Uchiyama, T., Katouno, F., Nikaidou, N., Nonaka, T., Sugiyama, J., *et al.*, **2001**. Roles of the exposed aromatic residues in crystalline chitin hydrolysis by chitinase A from *Serratia marcescens* 2170. *J. Biol. Chem. 276*, 41343–41349.

Väljamäe, P., Pettersson, G., Johansson, G., **2001**. Mechanism of substrate inhibition in cellulose synergistic degradation. *Eur. J. Biochem. 268*, 4520–4526.

van Aalten, D. 2003. Ligand-induced signaling and conformational change of the 39 kD glycoprotein from human chondrocytes. Proc. 9th Int. Chitin Chitosan Conf., Montreal, Abstract Ha. 1, p. 6.

Venegas, A., Goldstein, J.C., Beauregard, K., Oles, A., Abdulhayoglu, N., *et al.*, **1996**. Expression of recombinant microfilarial chitinase and analysis of domain function. *Mol. Biochem. Parasitol. 78*, 149–159.

Verloop, A., Ferrell, C.D., **1977**. Benzoylphenylureas: a new group of insecticides interfering with chitin deposition. *Pesticide Chem. 20th Cent. 37*, 237–270.

Veronico, P., Gray, L.J., Jones, J.T., Bazzicalupo, P., Arbucci, S., *et al.*, **2001**. Nematode chitin synthases: gene structure, expression and function in *Caenorhabditis elegans* and the plant parasitic nematode *Meloidogyne artiellia*. *Mol. Genet. Genomics 266*, 28–34.

Wang, P., Granados, R.R., **1997**. Molecular cloning and sequencing of a novel invertebrate intestinal mucin cDNA. *J. Biol. Chem. 272*, 16663–16669.

Wang, P., Granados, R.R., **2000a**. Molecular structure of the peritrophic membrane (PM): identification of potential PM target sites for insect control. *Arch. Insect Biochem. Physiol. 47*, 110–118.

Wang, P., Granados, R.R., **2000b**. Calcofluor disrupts the midgut defense system in insects. *Insect Biochem. Mol. Biol. 30*, 135–143.

Wang, X., Ding, X., Gopalakrishnan, B., Morgan, T.D., Johnson, L., *et al.*, **1996**. Characterization of a 46 kDa insect chitinase from transgenic tobacco. *Insect Biochem. Mol. Biol. 26*, 1055–1064.

Watanabe, T., Ariga, Y., Sato, U., Toratani, T., Hashimoto, M., *et al.*, **2003**. Aromatic residues within the substrate-binding cleft of Bacillus circulans chitinase A1 are essential for crystalline chitin hydrolysis. *Biochem. J. 376*, 237–244. Immediate Publication, doi:10.1042/BJ20030419.

Watanabe, T., Kobori, K., Miyashita, K., Fujii, T., Sakai, H., *et al.*, **1993**. Identification of glutamic acid 204 and aspartic acid 200 in chitinase A1 of *Bacillus circulans* WL-12 as essential residues for chitinase activity. *J. Biol. Chem. 268*, 18567–18572.

Watanabe, T., Kono, M., **2002**. Molecular studies of family 18 chitinases in animals: structural and functional diversity. *Chitin Chitosan Res. 8*, 95–104.

Watanabe, T., Oyangi, W., Suzuki, K., Tanaka, H., **1990**. Chitinase system of *Bacillus circulans* WL-12 and importance of chitinase A1 in chitin degradation. *J. Bacteriol. 172*, 4017–4022.

Watanabe, T., Uchida, M., Kobori, K., Tanaka, H., **1994**. Site-directed mutagenesis of the Asp-197 and Asp-202 residues in chitinase A1 of *Bacillus circulans* WL-12. *Biosci. Biotechnol. Biochem. 58*, 2283–2285.

Wiechaus, E., Nusslein-Volhard, C., Jurgens, G., **1984**. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster.* 1 Zygotic loci on the X chromosome and the fourth chromosome. *Roux's Arch. Devel. Biol. 193*, 296–307.

Wilson, T.G., Cryan, J.R., **1997**. Lufenuron, a chitin-synthesis inhibitor, disrupts development of *Drosophila melanogaster. J. Exp. Biol. 278*, 37–44.

Wiwat, C., Lertcanawanichkul, M., Siwayapran, P., Pantuwatana, S., Bhumiratan, A., **1996**. Expression of chitinase encoding genes from *Aeromonas hydrophila* and *Pseudomonas maltophilia* in *Bacillus thuringiensis* subsp. *israelensis*. *Gene 179*, 119–126.

Wiwat, C., Thaithanun, S., Pantuwatana, S., Bhumiratana, A., **2000**. Toxicity of chitinase-producing *Bacillus thuringiensis* sp. *kurstaki* HD-1 toward *Plutella xylostella*. *J. Invert. Path. 76*, 270–277.

Wormleaton, S., Kuzio, J., Winstanley, D., **2003**. The complete sequence of the *Adoxophyes orana* granulovirus genome. *Virology 311*, 350–365.

Yan, J., Cheng, Q., Narashimhan, S., Li, C.B., Aksoy, S., **2002**. Cloning and functional expression of a fat

body-specific chitinase cDNA from the tsetse fly, *Glossina morsitans morsitans*. *Insect Biochem. Mol. Biol. 32*, 979–989.

Zechel, D.L., Withers, S.G., **2000**. Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc. Chem. Res. 33*, 11–18.

Zen, K.C., Choi, H.K., Nandegama, K., Muthukrishnan, S., Kramer, K.J., **1996**. Cloning, expression and hormonal regulation of an insect β-*N*-acetylglucosaminidase gene. *Insect Biochem. Mol. Biol. 26*, 435–444.

Zhang, H., Huang, X., Fukamizo, T., Muthukrishnan, S., Kramer, K.J., **2002**. Site-directed mutagenesis and functional analysis of an active site tryptophan of insect chitinase. *Insect Biochem. Mol. Biol. 32*, 1477–1488.

Zheng, Y., Zheng, S., Cheng, X., Ladd, T., Lingohr, E.J., *et al.*, **2002**. A molt-associated chitinase cDNA from the spruce budworm, *Choristoneura fumiferana*. *Insect Biochem. Mol. Biol. 32*, 1813–1823.

Zheng, Y.P., Retnakaran, A., Krell, P.J., Arif, B.M., Primavera, M., *et al.*, **2003**. Temporal, spatial and induced expression of chitinase in the spruce budworm, *Choristoneura fumiferana*. *J. Insect Physiol. 49*, 241–247.

Zhou, Y., Wang, F., Xiao, X., **2002**. Structures of the cloned chitinase and its truncant from *A. caviae*. *Prog. Nat. Sci. 12*, 587–592.

Zhu, Y.C., Muthukrishnan, S., Specht, C.A., Dittmer, N., Kanost, M.R., *et al.*, **2002**. Sequence of a cDNA and expression of the gene encoding a putative epidermal chitin synthase of *Manduca sexta*. *Insect Biochem. Mol. Biol. 32*, 1497–1506.

Zimoch, L., Merzendorfer, H., **2002**. Immunolocalization of chitin synthase in the tobacco hornworm. *Cell Tissue Res. 308*, 287–297.

4.4 Cuticular Sclerotization and Tanning

S O Andersen, University of Copenhagen, Copenhagen, Denmark

© 2005, Elsevier BV. All rights reserved.

4.4.1. Introduction

The cuticle covers the complete body of the insects as an effective barrier between the animal and its surroundings; it provides protection against desiccation, microorganisms, and predators, and as an exoskeleton it serves as attachment sites for muscles. Cuticle is typically divided into relatively hard and stiff regions, the sclerites, separated by more flexible and pliable regions, the arthrodial membranes, which make the various forms of locomotion possible. Marked differences in mechanical properties may also be present on the microscopical level; two neighboring epidermal cells can produce cuticle with highly contrasting properties, indicating a precise control of cuticular composition. The mechanical properties of individual cuticular regions correspond closely to their functions and the forces to which they are exposed during the normal life of the animal. Proper flight can thus only be sustained when all wing regions have near-optimal balance between stiffness and flexibility; if the wing material locally is too soft or too stiff, the varying air pressure during the wing strokes will not cause the wings to bend to the shapes needed for generating optimal lift.

The mechanical properties of cuticle are determined by the interplay of many factors, such as cuticular thickness, relative amounts of chitin and proteins, chitin architecture, protein composition, water content, intracuticular pH, and degree of sclerotization and other secondary modifications. Sclerotization of insect cuticle has been reviewed several times in recent years (Sugumaran, 1988, 1998; Andersen, 1990; Hopkins and Kramer, 1992; Andersen *et al.*, 1996), but many aspects of the process are still rather poorly understood. Besides giving an overview of the present knowledge of the sclerotization process attention is drawn to some problems that need to be investigated in more detail.

Cuticular sclerotization is a chemical process whereby certain regions of insect cuticle are transformed irreversibly from a pliant material into a stiffer and harder structure, characterized by decreased deformability, decreased extractability of the matrix proteins, and increased resistance towards enzymatic degradation. During sclerotization the color of the cuticle may change from nearly colorless over various brown and black shades to the completely black. The term tanning is often used synonymously with sclerotization, but sometimes it is specifically used for the processes resulting in light brown (tan) cuticles. Sclerotization often takes place in connection with molting, starting just after the new, yet unsclerotized cuticle has been expanded to its final size and shape, but some specialized cuticular regions are sclerotized while the insect is still in its pharate state inside the old cuticle. Such pre-ecdysially sclerotized regions cannot be expanded post-ecdysially, but may help the insect to escape from the exuvium. The dipteran puparium is an example of a soft larval cuticle, which is sclerotized at the end of the last larval instar to form a hard protective case inside which metamorphosis to pupa and adult can take place. Sclerotization of structural materials in insects is not restricted to cuticle; other materials as well, such as egg cases and silks, may be stabilized by chemical processes closely related to cuticular sclerotization.

4.4.2. Current Model for Cuticular Sclerotization

During the years several models have been proposed for the chemical reactions which occur in the insect cuticle during the sclerotization process, and although many details of the individual steps in the reactions are still controversial or unexplored, there is general agreement concerning the main features of the process. The currently accepted sclerotization model is shown in Figures 1 and 2, and its main features are: the amino acid tyrosine (**1**) is hydroxylated to 3,4-dihydroxyphenylalanine (DOPA, **2**), which by decarboxylation is transformed to dopamine (**3**), a compound of central importance for both sclerotization and melanine formation. Dopamine can be *N*-acylated to either *N*-acetyldopamine (NADA, **4**) or *N*-β-alanyldopamine (NBAD, **5**), and both can serve as precursors in the sclerotization process. They are enzymatically oxidized to the corresponding *o*-quinones (**6**), which can react with available nucleophilic groups, whereby the catecholic structure is regained and the nucleophile is linked to the aromatic ring (**11**). The *o*-quinones of NADA and NBAD may also be enzymatically isomerized to the corresponding *p*-quinone methides (**7**), and the β-position of their side chain react readily with nucleophiles (**12**). The *p*-quinone methides may be enzymatically isomerized to side chain unsaturated catechol derivatives (**8**), dehydro-NADA

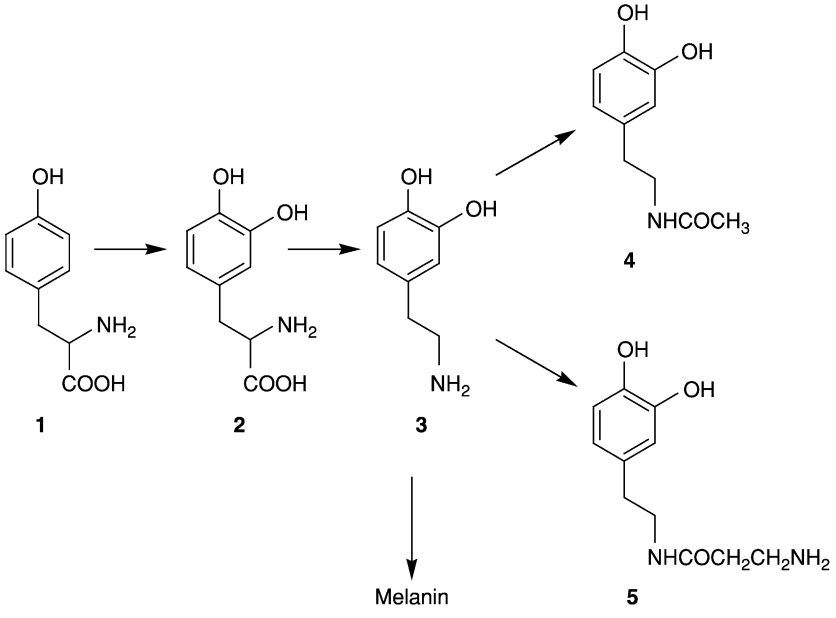


Figure 1 Biosynthesis of the sclerotization precursors NADA and NBAD from tyrosine. The enzyme tyrosine hydroxylase hydroxylates tyrosine (**1**) to 3,4-dihydroxyphenylalanine (DOPA, **2**), which is decarboxylated to dopamine (**3**) by the enzyme dopa-decarboxylase. Dopamine can be enzymatically acylated to either *N*-acetyldopamine (NADA, **4**) or *N*-β-alanyldopamine (NBDA, **5**). Surplus dopamine can be used for melanin synthesis.

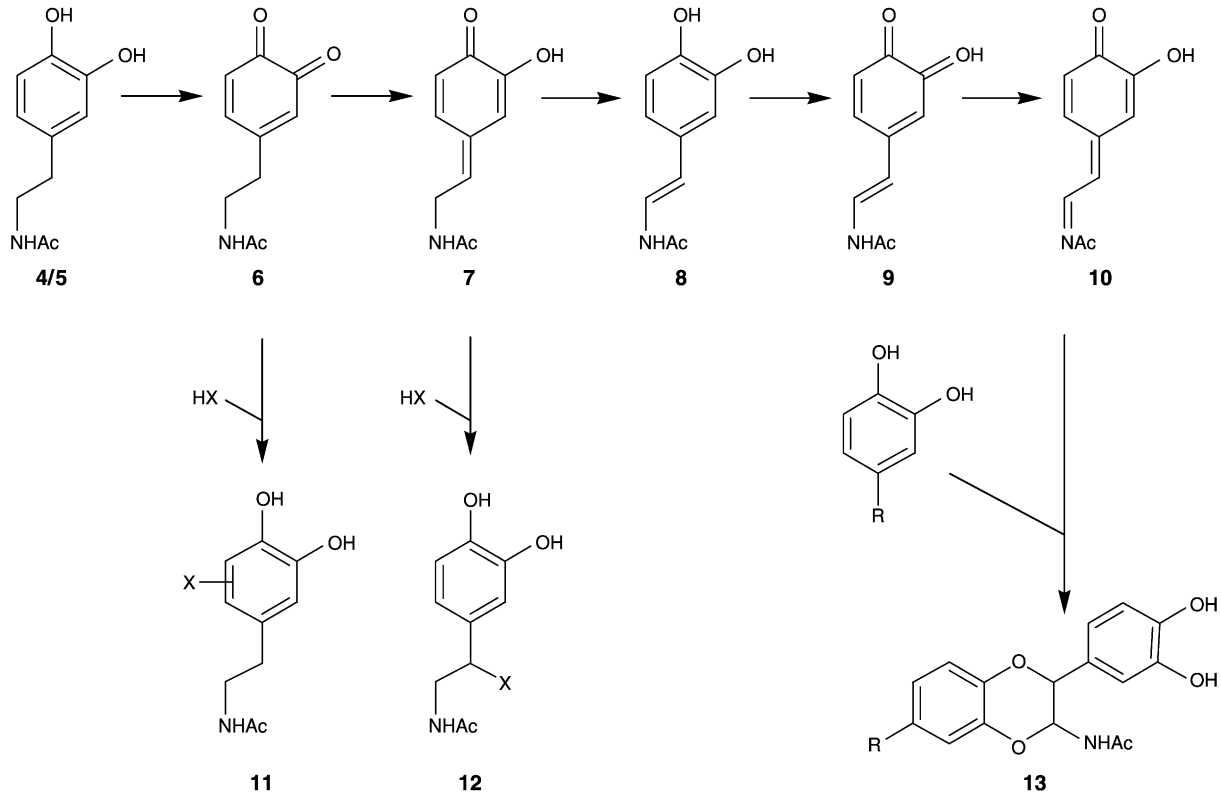


Figure 2 Cuticular oxidation of NADA and NBAD to various sclerotization agents. The acyldopamines are enzymatically oxidized to *o*-quinones (**6**), which can react spontaneously with nucleophilic compounds (HX) to give ring-substituted adducts (**11**), or they can be isomerized to the more reactive *p*-quinone methides (**7**), which can react with nucleophiles to give side chain substituted adducts (**12**). The quinone methides may also be isomerized to dehydro-acyldopamines (**8**), which after oxidation to quinones (**9** and **10**) can react with catechols to give dihydroxyphenyl-dihydrobenzodioxine derivatives (**13**).

and dehydro-NBAD, which after oxidation to unsaturated quinones can react with catechols to form dihydroxyphenyl-dihydrobenzodioxine derivatives (**13**) and with other nucleophilic groups to give yet unidentified compounds.

The *o*-quinones and *p*-quinone methides react preferably with the imidazole group in the cuticular proteins, but may also react with free amino groups, such as terminal amino groups in proteins and ε-amino groups in lysine residues, with catechols, with water, and probably also with hydroxyl groups in the *N*-acetylglucosamine residues in chitin. Depending upon the degree of sclerotization the various reactions between quinones and nucleophilic residues in the cuticular matrix proteins will result in the proteins being more or less covered by aromatic residues; some of these residues may be involved in cross-linking the cuticular proteins and maybe also forming links between proteins and chitin; some will be linked to only a single protein molecule, thereby increasing its hydrophobicity without being part of a covalent cross-link. During sclerotization most of the water-filled spaces between the matrix proteins in the presclerotized cuticle will become filled with polymerized catecholic material. As a result of these processes the interactions between the cuticular components have become stronger, the peptide chains more difficult to deform, and the proteins cannot be moved relative to each other or to the chitin filament system. Together, all these changes contribute to make the material stiffer and more resistant towards degradation.

The various reactions involved in the model will be discussed in more detail in the following sections, with main emphasis on aspects where the evidence is insufficient or missing or where some observations disagree with the scheme, to indicate areas where more research is needed. The appearance and properties of cuticle from different body regions of the same animal can vary widely, and a considerable part of this variation, such as the different coloration and mechanical properties, is probably due to quantitative or qualitative differences in the sclerotization process. There is no compelling reasons to believe that exactly the same stabilization process is used for sclerotization in all types of solid cuticle,

and generalizations based upon results obtained with a single or a few insect species can easily be misleading. A number of cuticular types will have to be analyzed to determine how the individual steps involved in sclerotization are modulated to give the local variation between cuticular regions of a given insect and between cuticles from different insect species. It will also be important to study how the reactions are controlled to give an optimal degree of sclerotization. Most of the results and ideas presented in this chapter have been obtained by studies involving material from only a few insect species, such as cuticle of blowfly larvae, pupae of *Manduca sexta*, and locust femurs, and it is to be expected that detailed studies of cuticle from more species will give a much more varied and fascinating picture of the complexity of cuticular sclerotization.

4.4.3. Sclerotization (Tanning) Precursors

The terms sclerotization agents and tanning agents were originally used for the compounds which are secreted from the epidermal cells into the cuticle, where they are oxidized by enzymes to become sufficiently reactive to form covalent linkages to proteins and chitin. There is now a tendency to restrict the term sclerotization agents to the reactive species directly involved in forming links to the cuticular components. The compounds secreted from the epidermis to be activated in the cuticle shall accordingly be called sclerotization or tanning precursors (Sugumaran, 1998).

4.4.3.1. *N*-Acetyldopamine and *N*-β-Alanyldopamine

The first discovered and most common precursor for cuticular sclerotization is NADA (**4**), which is synthesized by *N*-acetylation of dopamine. The central role of NADA in sclerotization was demonstrated by Karlson's research group during studies of tyrosine metabolism in insects (review: Karlson and Sekeris, 1976). They showed that NADA is incorporated into the puparial cuticle of the blowfly *Calliphora vicina* during its sclerotization, and that radioactively labeled tyrosine was metabolized to NADA when injected into last instar larvae shortly before puparium formation, and degraded when injected into younger larvae. NADA was also shown to be involved in cuticular sclerotization in several other insect species, such as the desert locust, *Schistocerca gregaria* (Karlson and Schlossberger-Raecke, 1962; Schlossberger-Raecke and Karlson, 1964). Incorporation of NADA into cuticle can be a very efficient process; after injection of radioactive NADA into young adult locusts about 80% of the total radioactivity was later recovered from the sclerotized cuticle (Andersen, 1971). NADA appears to be involved in cuticular sclerotization in all insect species investigated.

The amino acid β-alanine has been reported as a constituent of several types of sclerotized cuticle, and was suspected to participate somehow in the sclerotization process (Andersen, 1979a). Hopkins *et al.* (1982) showed that the β-alanyl derivative of dopamine, NBAD (**6**), is a sclerotizing precursor in the cuticle of *M. sexta* pupae, thus accounting for the presence of β-alanine in hydrolysates of the fully sclerotized cuticle. NBAD is also a sclerotization precursor in other cuticles from which β-alanine is released by acid hydrolysis, such as the cuticle of the red flour beetle, *Tribolium castaneum* (Kramer *et al.*, 1984). The synthesis and utilization of NBAD during pupation of *M. sexta* has been reported (Krueger *et al.*, 1989). An enzyme system in the medfly, *Ceratitis capitata* which can catalyze β-alanylation of dopamine to give NBAD has been partially characterized (Pérez *et al.*, 2002).

The two sclerotizing compounds, NADA and NBAD, are used together in many types of cuticles; but the cuticle of some insects, such as the locusts, *S. gregaria* and *Locusta migratoria*, appears to be exclusively sclerotized by NADA, as no β-alanine has been obtained from their acid hydrolysates. No cuticles have yet been reported to be sclerotized exclusively by NBAD. A correlation appears to exist between the intensity of brown color of the fully sclerotized cuticle and the amounts of NBAD taking part in the sclerotization process: cuticles that are sclerotized exclusively by NADA are colorless or very lightly straw-colored, and the more NBAD dominates in the process the darker brown will the cuticle become (Brunet, 1980; Hopkins *et al.*, 1984). Czapla *et al.* (1990) reported that cuticular strength in five differently colored strains of the cockroach *Blattella germanica* correlated well with their concentrations of β-alanine and NBANE, whereas dopamine concentration correlated with melanization. Cuticular strength as well as cuticular concentrations of β-alanine and NBAD increased more rapidly in the rust-red wild-type of *T. castaneum* than in the black mutant strain, whereas cuticular dopamine increased more rapidly in the black strain than in the wild-type (Roseland *et al.*, 1987).

Significant amounts of sclerotization precursors are often present as conjugates before the onset of sclerotization. The conjugates, which can be glucosides, phosphates, or sulfates (Brunet, 1980; Kramer and Hopkins, 1987), are not easily oxidized, and

have to be hydrolyzed to free catechols before they can take part in sclerotization. It is assumed that the catechol conjugates serve as a storage reservoir of catecholamines ready to be used when the need for sclerotization arises (Brunet, 1980). A dopamine conjugate, identified as the 3-*O*-sulfate ester, is present in the hemolymph of newly ecdysed cockroaches, and its concentration decreases rapidly as the cockroach cuticle sclerotizes. The sulfate moiety is not transferred into the cuticle, and removal of sulfate and acylation of the liberated dopamine to NADA and/or NBAD most likely occurs in the epidermal cells (Bodnaryk and Brunet, 1974; Czapla *et al*., 1988, 1989).

Hopkins *et al*. (1984) reported that a large fraction of the various catecholamines in *M. sexta* hemolymph and cuticle is present as acid labile conjugates. In larval and pupal hemolymph these conjugates are mainly 3-*O*-glucosides together with small amounts of the 4-*O*-glucosides, whereas adult hemolymph contains more of the 4-*O*-glucoside than of the 3-*O*-glucoside (Hopkins *et al*., 1995). Both conjugated and unconjugated forms of NADA and NBAD were extracted from *M. sexta* cuticle (Hopkins *et al*., 1984), indicating that the epidermal cells possess transporting systems for conjugated as well as unconjugated catecholamines. It is also likely that a β-glucosidase activity is present inside the *Manduca* cuticular matrix, to catalyze the release of the sclerotization precursors from their conjugates, as the conjugated forms cannot be used directly for sclerotization.

4.4.3.2. Putative Sclerotization Precursors

So far, convincing evidence that they function as cuticular sclerotization precursors has only been obtained for NADA and NBAD, but several other compounds have been described as likely sclerotization precursor candidates, such as dopamine (**3**), *N*-acetyl-norepinephrine (NANE) (**14**), *N*-β-alanyl-norepinephrine (NBANE) (**15**), and 3,4-dihydroxyphenylethanol (DOPET) (**16**) (Figure 3). It is likely that they all have some role to play in sclerotization.

4.4.3.2.1. Dopamine Dopamine is precursor for both NADA and NBAD synthesis, and it can also be precursor for black, insoluble melanins. Melanin-like materials are of common occurrence in cuticular structures, and they can be present either in microscopic granules or homogeneously distributed within the cuticular matrix (Kayser-Wegmann, 1976; Kayser-Wegmann and Kayser, 1983; Hiruma and Riddiford, 1988). The granules are produced within epidermal cells and transported to the subepicuticular space via long cellular projections (Curtis *et al*., 1984; Kayser, 1985), and apparently the melanin in the granules is linked to granular proteins, but not to proteins in the cuticular matrix. The diffusely distributed melanins appears to be formed *in situ* within the cuticular matrix, and covalent links are probably formed between the polymeric melanin and the matrix proteins, thereby rendering the proteins more stable and insoluble, contributing to both darkening and increased mechanical stiffness of the cuticle. It will therefore be difficult in all cases to discern between the process of melanization of cuticle and the process of sclerotization. The observation that cuticular incorporation of radioactive tyrosine is nearly the same in wild-type and albino mutant of *S. gregaria* (Karlson and Schlossberger-Raecke, 1962) shows that melanization does not play a major role in sclerotization, but handling of small samples of melanized or albino cuticle indicates that it may have a minor role, as the melanized samples appear to be more brittle than the unmelanized (S.O. Andersen, unpublished data). A quantitative comparison of the mechanical properties of the two cuticular samples, to see to what extent melanization influences the physical properties of the material, would be interesting.

Melanin can be formed from either DOPA or dopamine by slightly different routes. DOPA can via dopachrome and 5,6-dihydroxyindole carboxylic acid be transformed to 5,6-dihydroxyindole,

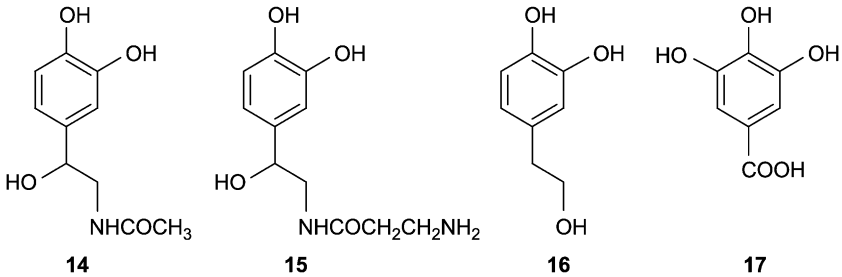


Figure 3 Putative precursors for cuticular sclerotization. **14**, *N*-acetylnorepinephrine (NANE); **15**, *N*-β-alanylnorepinephrine (NBANE); **16**, 3,4-dihydroxyphenylethanol (DOPET); **17**, gallic acid.

which after enzymatic oxidation polymerizes to melanin, and transformation of dopamine to 5,6-dihydroxyindole goes via dopamine chrome. The formation of the two intermediates, dopachrome and dopamine chrome, is catalyzed by different, but related enzymes. Dopamine appears to be a better substrate than DOPA for formation of cuticular melanin, and the melanins formed in hemolymph during defense reactions appear to be formed mainly via the DOPA pathway. In *Drosophila melanogaster* the products of two of the *yellow* genes, *yellow*-f and *yellow*-f2, are dopachrome-conversion enzymes, which have low activities towards dopamine chrome. The product of the gene *yellow*-y, involved in wing and cuticle melanization, has significant sequence similarity to yellow-f and yellow-f2 proteins, but is apparently devoid of dopachrome-conversion activity (Han *et al*., 2002). Whether the yellow-y protein can catalyze the conversion of dopamine chrome to 5,6-dihydroxyindole was not reported.

The black body color of the *D. melanogaster* mutants *black* and *ebony* is due to the inability of these mutants to produce sufficient NBAD for cuticular sclerotization: *black* is defective in the synthesis of β-alanine, and can be rescued by injection of β-alanine, and *ebony* is defective in the enzyme NBAD-synthetase, and cannot be rescued by injection of β-alanine (Wright, 1987). The result is that in both mutants some of the dopamine not used for NBAD synthesis is channeled into cuticular melanin production. Both stiffness and puncture resistance of the cuticle are decreased in the mutants, and electron microscope studies show that the cuticular chitin lamellae are abnormally wide and diffuse (Jacobs, 1978, 1980, 1985), indicating that even if dopamine-derived melanin can take part in cuticular stabilization, the result is inferior to the material obtained by NBAD sclerotization.

The *tan* mutant of *D. melanogaster,* which is characterized by absence of the wild-type cuticular melanin pattern, has low activity of the enzyme *N*-β-alanyldopamine-hydrolase, catalyzing the hydrolysis of NBAD to β-alanine and dopamine (Wright, 1987). The enzyme systems, responsible for NBAD synthesis and NBAD hydrolysis, respectively, are probably located in different compartments, and the presence of melanin in some but not all cuticular regions of wild-type fruitflies could be explained by the local presence of the NBAD-hydrolase within the matrix of the melanizing regions, where it will hydrolyze some of the NBAD secreted from the epidermal cells, creating a dopamine concentration sufficient to stimulate a localized melanin production.

4.4.3.2.2. *N*-acetylnorepinephrine (NANE) and *N*-β-alanylnorepinephrine (NBANE) NANE and NBANE are special cases among the cuticular catechols, as they can be considered both as byproducts of the sclerotization process and precursors for sclerotization. They have been reported to occur both free and as an *O*-glucoside in hemolymph and integument in several insects (Hopkins *et al*., 1984, 1995; Morgan *et al*., 1987; Czapla *et al*., 1989). NANE and NBANE can be generated within the cuticle, when the enzymatically produced *p*-quinone methides of NADA and NBAD react with water instead of reacting with cuticular proteins or isomerizing to dehydro-derivatives, but they may also be produced by hydrolysis of some unidentified products of the sclerotization process. Mild acid treatment of sclerotized cuticles can release some NANE and NBANE from the cuticular structure, probably due to hydrolysis of a bond between the β-position of the catechols and some cuticular constituent. The nature of the bond is uncertain, but it could well be an ether linkage connecting the acyldopamine side chain to chitin. Formation of an ether, β-methoxy-NADA, occurs when isolated pieces of cuticle or extracted cuticular enzymes act upon NADA in the presence of methanol; the compound is acid labile and is readily hydrolyzed to free NANE (Andersen, 1989c; Sugumaran *et al*., 1989b).

NANE can be covalently incorporated into the cuticular matrix during sclerotization, indicating that the compound can serve as a sclerotization precursor (Andersen, 1971), and this is probably also the case for NBANE. When radioactively labeled NANE was injected into newly ecdysed locusts a significant fraction of the radioactivity (about 10%) was incorporated into the cuticle, and hydrolysis of the cuticle was needed to release the activity. Acid hydrolysis of cuticle from locusts injected with labeled norepinephrine resulted in the release of both labeled norepinephrine and arterenone, whereas little radioactivity was present in the neutral ketocatechol fraction. This is in contrast to parallel experiments when labeled dopamine was injected and nearly all the radioactivity was recovered as neutral ketocatechols, indicating that the cuticular enzymes can catalyze the incorporation of norepinephrine and NANE into the cuticular matrix, but not as efficiently and not by the same route as the incorporation of dopamine and NADA.

4.4.3.2.3. Dihydroxyphenylethanol (DOPET) The third putative sclerotization precursor, 3,4-dihydroxyphenylethanol (DOPET), is also present in the hemolymph and integument of insects; it has been obtained from cuticle of the cockroach *Periplaneta*

americana (Atkinson *et al.*, 1973b; Czapla *et al.*, 1988) and the beetle *Pachynoda sinuata* (Andersen and Roepstorff, 1978), and it can function as substrate for the cuticular phenoloxidases. Extraction and acid hydrolysis of sclerotized cuticles have yielded various DOPET derivatives, suggesting that DOPET is transported into the cuticle and incorporated into the cuticular matrix during sclerotization. Adducts of DOPET and histidine have been obtained by acid hydrolysis of sclerotized *Manduca* pupal cuticle and identified by means of mass spectrometry (Kerwin *et al.*, 1999). A dihydroxyphenyl-dihydrobezodioxine-type adduct of DOPET and NADA has also been extracted from sclerotized beetle (*P. sinuata*) cuticle (Andersen and Roepstorff, 1981), but the metabolism of DOPET in insects has not been studied in much detail.

The relative roles of dopamine, NANE, NBAD, and DOPET compared to the two major sclerotization precursors, NADA and NBAD, have never been properly established. The compounds are probably only of minor importance for the mechanical properties of sclerotized cuticles, but their involvement in the sclerotization process may play a role in fine-tuning of the cuticular properties.

4.4.3.2.4. Other sclerotization precursors It has been reported that improved growth of the tree locust, *Anacridium melanorhodon* can be obtained by addition of gallic acid (**17**) and other plant phenols to its food, and that the ingested plant phenols are incorporated into the cuticle and may contribute to its stabilization (Bernays *et al.*, 1980; Bernays and Woodhead, 1982).

4.4.3.3. Noncuticular Sclerotization

Although the sclerotization process has mainly been studied in the cuticle of insects, related processes are used for stabilization of other insect materials, such as silks, egg capsules, and chorions, and their structures are shown in **Figure 4**. Pryor (1940a) showed that 3,4-hydroxybenzoic acid (**18**) is a sclerotization precursor for egg capsules of the cockroach *Blatta orientalis*, and Pau and Acheson (1968) reported that other cockroach species use 3,4-dihydroxybenzyl alcohol (**19**) as a precursor for egg capsule sclerotization. The praying mantid, *Hierodula patellifera* uses various catechol derivatives for stabilizing the egg capsules; besides NADA and NBAD the following catechols have been reported: *N*-malonyldopamine (**20**), *N*-(*N*-acetyl-β-alanyl) dopamine (**21**), and *N*-(*N*-malonyl-β-alanyl)dopamine (**22**) (Kawasaki and Yago, 1983; Yago and Kawasaki, 1984; Yago *et al.*, 1984). The cockroach catechols are stored in the left colleterial gland as 4-*O*-glucosides (Brunet, 1980), and the catechols in praying mantids are stored as 3-*O*-glucosides (Yago

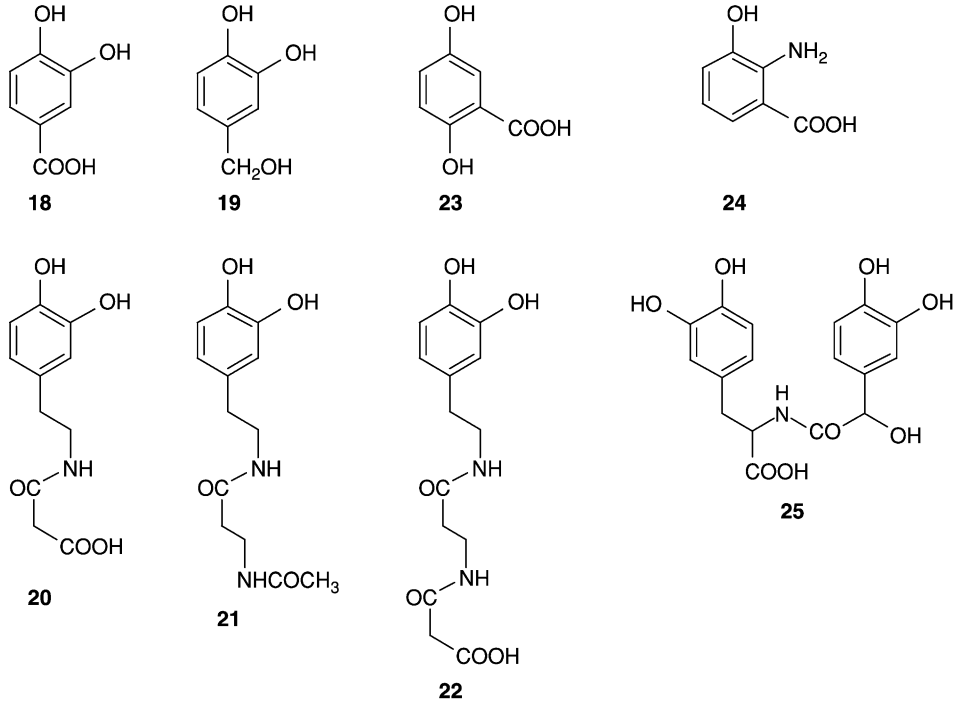


Figure 4 Precursors for noncuticular sclerotization. **18**, 3,4-dihydroxybenzoic acid (protocatechuic acid); **19**, 3,4-dihydroxybenzyl alcohol; **20**, *N*-malonyldopamine; **21**, *N*-(*N*-acetyl-β-alanyl)dopamine; **22**, *N*-(*N*-malonyl-β-alanyl)dopamine; **23**, gentisic acid; **24**, 3-hydroxyanthranilic acid; **25**, *N*-(3,4-dihydroxyphenyllactyl)DOPA.

et al., 1984). The right colleterial gland in both species contains a glucosidase, and when the secretions from the two glands are mixed, the glucosides are hydrolyzed, and the liberated catechols are enzymatically oxidized to *o*-quinones, which in mantids may be further converted to a highly reactive intermediates, presumably *p*-quinone methides (Yago *et al.*, 1990).

The eggs of the grasshopper *Melanoplus sanguinipes* are deposited in soil and covered by a frothy proteinaceous material produced in the female accessory glands. An acid-labile conjugate of 3,4-dihydroxybenzoic acid was extracted from the accessory glands, and unconjugated 3,4-dihydroxybenzoic acid was obtained from the egg pods shortly after their deposition, indicating that it may be involved in sclerotization of the material (Hopkins *et al.*, 1999).

Some moths, such as saturniid silkmoths (*Actias selene, Antheraea pernyi, Hyalophora cecropia, H. gloveri,* and *Samia cynthia*), produce silk cocoons that are sclerotized after spinning. The silks are white when spun and remain white as long as they are in a dry atmosphere: When exposed to humid conditions a sclerotization process is initiated and the cocoons become brown. Glucosides of gentisic acid (**23**) and 3-hydroxyanthranilic acid (**24**) have been obtained from newly spun silks, and β-glucosidases and diphenoloxidases are present, indicating that silk sclerotization follow a pattern similar to that of the other sclerotization processes (Brunet and Coles, 1974). Manthey *et al.* (1992) reported that oxidation of 3-hydroxyanthranilic acid in the presence of proteins leads to formation of adducts between 3-hydroxyanthranilic acid and tyrosine residues in the proteins, and such adducts were isolated from cocoons of *H. gloveri* and *S. cynthia*. It appears that the adducts are formed via a radical–radical coupling mechanism.

The newly spun silk of the Japanese giant silkmoth, *Dictyoploca japonica* contains a diphenoloxidase and a catecholic derivative, which was tentatively identified as *N*-(3,4-dihydroxyphenyllactyl)DOPA (**25**) (Kawasaki and Sato, 1985). It is apparently not glycosylated and no glucosidase activity was found in the silk. The presence of two catecholic groups in this sclerotization precursor may make it a more effective cross-linking agent than precursors with only a single catechol group.

4.4.4. Cuticular Enzymes and Sclerotization

The study of cuticular enzymes has to a large extent been concerned with characterization of enzymes assumed to play a role in cuticular sclerotization, mainly those involved in catechol oxidation and quinone isomerization, and there has been a tendency to neglect the possible presence of other types of enzymes that may play a role in cuticular stabilization. Catechol oxidation is an important step in sclerotization, wound healing, and immune responses in insects, and it is often difficult to decide whether a given cuticular phenoloxidase activity is of importance for sclerotization or whether its main role is to take part in defense reactions. It is therefore necessary to be extremely careful in interpreting the observations reported for cuticular enzymes.

4.4.4.1. *ortho*-Diphenoloxidases

After being transferred from the epidermal cells into the cuticle the catecholic sclerotization precursors may encounter various enzymes, such as *o*-diphenoloxidases, laccases, and peroxidases, capable of oxidizing them to quinones, but the relative roles of these activities are still uncertain. Insects contain inactive proenzymes for *o*-diphenoloxidases both in hemolymph and in the cuticle, and the proenzymes can be activated by a process initiated by wounds or the presence of small amounts of microbial cell wall components and involving limited proteolysis (Ashida and Dohke, 1980; Ashida and Brey, 1995). The *o*-diphenoloxidases can oxidize a wide range of *o*-diphenols, but not *p*-diphenols; they possess low hydroxylating activity towards monophenols, such as tyrosine and tyramine; and they are copper-containing enzymes readily inhibited by thioureas and Na-diethyldithiocarbamate. They have been isolated and characterized from both soft, nonsclerotizing cuticles, such as larval cuticle of *Bombyx mori* (Ashida and Brey, 1995; Asano and Ashida, 2001a), and from blowfly larval cuticles (Barrett, 1987a, 1987b, 1991) and sclerotizing pupal cuticle of *Manduca sexta* (Aso *et al.*, 1984; Morgan *et al.*, 1990). The amino acid sequences of *o*-diphenoloxidases from various insect species have been deduced from the corresponding DNA sequences (Fujimoto *et al.*, 1995; Hall *et al.*, 1995; Kawabata *et al.*, 1995). The insect *o*-diphenoloxidases differ from diphenoloxidases (tyrosinases) from other organisms with regard to both substrate specificity and amino acid sequence (Sugumaran, 1998; Chase *et al.*, 2000).

According to the established sequences of insect prophenoloxidase genes, their gene products do not possess an N-terminal signal peptide sequence (Sugumaran, 1998), which is in contrast to what is commonly observed for proteins destined for export from the cells. The silk moth *B. mori* has genes for two *o*-diphenoloxidase proenzymes, and in the larvae the products of both genes are present in both

hemolymph and cuticle, the only difference between the cuticular and hemolymphal forms being that several methionine residues are intact in the hemolymphal proenzymes, and they are oxidized to methionine sulfoxides in the cuticular proenzymes. The activated cuticular enzymes have nearly the same substrate specificity as the hemolymph enzymes; the main difference between the proenzymes is that in contrast to the unmodified hemolymph form the oxidized cuticular form cannot be transported across the epidermal cell layer, indicating that the epidermal transport of the proenzymes is a one-way traffic, from hemolymph to cuticle (Asano and Ashida, 2001a, 2001b).

The enzymatic properties of the diphenoloxidases purified from hemolymph and from pharate pupal cuticle of *M. sexta* are very similar, suggesting a close relationship between the enzymes (Aso *et al.*, 1985; Morgan *et al.*, 1990), and it seems probable that they, like the *B. mori* enzymes, are derived from the same gene(s).

When activated, both cuticular and hemolymph *o*-diphenoloxidases are very sticky i.e., they tend to aggregate and stick to any available surface and macromolecule, which will hinder diffusion of the active enzymes from the site where they became activated. The ultrastructural localization of *o*-diphenoloxidase activity in the larval cuticle of the blowfly *Lucilia cuprina* was described by Binnington and Barrett (1988) who observed activity in the epicuticular filaments. Activity was also observed in the procuticle, but only when the cuticle had been damaged beforehand, and the activity was limited to the close neighborhood of the wound, indicating that wounding is needed to activate the enzyme, and that the active enzyme remains in the vicinity of the wound.

The prophenoloxidase in larval cuticle of *B. mori* was localized by immunocytochemical methods to the nonlamellate endocuticle, where it was randomly distributed, and to an orderly arrayed pattern in the lamellate endocuticle, but it appeared to be absent from the cuticulin layer and from the epidermal cells (Ashida and Brey, 1995).

The role of the various cuticular *o*-diphenoloxidases in cuticular sclerotization is problematic; their presence as inactive proenzymes, which have to be activated, their close relationship to the hemolymphal phenoloxidases, and their abundance in nonsclerotizing cuticle indicate that their role is to take part in defense against wounding and microorganisms and not to be involved in sclerotization. It is, however, difficult to state with certainty that they do not play some role in sclerotization.

4.4.4.2. Laccases

Laccase-type phenoloxidases have been reported to be present in dipteran larval cuticles shortly before and during puparium sclerotization, such as *Drosophila virilis* (Yamazaki, 1969), *D. melanogaster* (Sugumaran *et al.*, 1992), *Calliphora vicina* (Barrett and Andersen, 1981), *Sarcophaga bullata* (Barrett, 1987a), and *L. cuprina* (Barrett, 1987b), and such enzymes have also been described from pupal cuticles of *B. mori* (Yamazaki, 1972) and *M. sexta* (Thomas *et al.*, 1989) as well as from adult cuticle of the locust *Schistocerca gregaria* (Andersen, 1978). The nucleotide sequences for two laccase genes from *M. sexta* and a laccase gene from the mosquito *Anopheles gambiae* have recently been deposited in the GenBank, and the accession numbers for the corresponding proteins are: AAN1706, AAN1707, and AAN17505, respectively. The insect laccases are structurally related to laccases of plant or fungal origin. In contrast to the insect diphenoloxidases the laccase gene products contain a typical signal peptide sequence, indicating that the enzymes are secreted into the extracellular space.

In larval cuticles of *D. virilis* (Yamazaki, 1969) and *L. cuprina* (Binnington and Barrett, 1988) laccase activity makes its appearance shortly before pupariation. In both species the enzyme activity decreases gradually as puparial sclerotization progresses. Laccase activity can be demonstrated a few days before ecdysis in pharate cuticle of adult locusts, *S. gregaria*; it remains at high levels for at least 2 weeks after ecdysis, and activity has also been demonstrated in nymphal exuviae, indicating that the locust enzyme is not inactivated by sclerotization (S.O. Andersen, unpublished data).

Laccases are active towards a broad spectrum of *o*- and *p*-diphenols: NBAD and NADA are among the best *o*-phenolic substrates tested, and methylhydroquinone is the best *p*-diphenolic substrate. Insect laccases are not inhibited by compounds, such as thiourea, phenylthiourea, and Na-diethyldithiocarbamate, which are effective inhibitors of *o*-diphenoloxidases, but they are inhibited by carbon monoxide and millimolar concentrations of fluorides, cyanides, and azides (Yamazaki, 1972; Andersen, 1978; Barrett and Andersen, 1981; Barrett, 1987a). The laccases are resistant towards treatments inactivating many enzyme activities; the *S. gregaria* laccase remains active after blocking available amino and phenolic groups by dinitrophenylation or dansylation, and it survives temperatures up to about 70 °C, but it is inactivated by treatment with tetranitromethane, which nitrates tyrosine residues (Andersen, 1979b). The laccases

appear to be firmly linked to the cuticular structure; typically they cannot be extracted by conventional protein extractants, but are readily extracted after limited tryptic digestion of the yet-unhardened cuticle (Yamazaki, 1972; Andersen, 1978). The enzyme was obtained from *C. vicina* larval cuticle by prolonged extraction at pH 8 without addition of any protease, but as latent protease activity is present in the cuticle the release of laccase from the cuticular residue may be due to proteolysis (Barrett and Andersen, 1981). The enzyme is not released by tryptic digestion of already sclerotized cuticle.

The ultrastructural localization of laccase activity has been studied in the *L. cuprina* larval cuticle (Binnington and Barrett, 1988) and enzyme activity was observed in the inner epicuticle of late third instar larvae (about to pupariate), but not in epicuticle of younger larvae. The laccase activity in *L. cuprina* larval cuticle could be demonstrated without prior activation, in contrast to the cuticular *o*-diphenoloxidases, indicating that the laccase is not deposited as an inactive precursor in this insect, and neither is an inactive proenzyme likely to be present in pharate locust cuticle since enzyme activity could be demonstrated without any activating treatment. A pro-laccase has been purified and partially characterized from cuticle of newly pupated pupae of *B. mori* (Ashida and Yamazaki, 1990). The inactive pro-laccase could be activated by treatment with various proteolytic enzymes, and the substrate specificities of the laccase variants obtained depended upon the protease used for activation.

4.4.4.3. Cuticular Peroxidases

Several different routes for the oxidation of catechols to *o*-quinones may be of advantage for an insect, especially if they can be regulated independently in various cuticular regions, and peroxidase activity may provide such an alternative route. Peroxidase activity has been demonstrated by histochemical methods in proleg spines of *Calpodes ethlius* larvae (Locke, 1969) and in larval and pupal cuticle of *Galleria mellonella* and *Protophormia terraenovae* (Grossmüller and Messner, 1978; Messner and Janda, 1991; Messner and Kerstan, 1991), and such activity is also observed intracellularly in different cell types in insects. It is not known whether the cuticular peroxidase activities are identical to the intracellular enzymes, as the cuticular activity has never been properly characterized. Proteins can be cross-linked by means of the peroxidase system, and it has been suggested that the enzyme could be involved in cuticular sclerotization (Hasson and Sugumaran, 1987). A peroxidase is likely to be involved in the cross-linking of the rubberlike elastic cuticular protein resilin (Andersen, 1966; Coles, 1966). This cuticular protein is cross-linked by oxidative coupling of tyrosine residues during its extracellular deposition, and the tyrosine radicals needed for the coupling may be formed by a peroxidase catalyzed oxidation process. Peroxidases can also oxidize catechols to semiquinone radicals, two of which readily dismutate to form an *o*-quinone and a catechol. The enzyme needs hydrogen peroxide as one of its substrates, and Candy (1979) reported that locust cuticle contains a glucose oxidase activity, which oxidizes glucose to D-gluconate with concomitant production of hydrogen peroxide. It was suggested that the hydrogen peroxide produced may participate in sclerotization reactions. Candy (1979) also reported the presence in locust cuticle of several other enzymes involved in hydrogen peroxide metabolism, such as peroxidase, catalase, and superoxide dismutase.

Peroxidase activity in solid cuticle may be involved in the production of dityrosine cross-links and in oxidizing catechols to quinones for sclerotization. Dityrosine has so far not been demonstrated in sclerotized cuticle from insects, but dityrosine as well as brominated dityrosines have been obtained from the hardened exocuticle of the crab *Cancer pagurus* (Welinder *et al.*, 1976). The eggshells of *D. melanogaster* are stabilized by formation of dityrosine cross-links between the protein chains (Petri *et al.*, 1976; Mindrinos *et al.*, 1980), and the hardening of *Aedes aegypti* egg chorion includes both peroxidase-mediated protein cross-linking through dityrosine formation and diphenoloxidase-catalyzed chorion melanization (Li *et al.*, 1996). The hydrogen peroxide necessary for dityrosine formation in *A. aegypti* chorion is produced by an enzymatic process by which NADH is oxidized with concomitant reduction of molecular oxygen to hydrogen peroxide. The necessary supply of NADH for this process is provided by enzyme-catalyzed oxidation of malate coupled to reduction of NAD^+ (Han *et al.*, 2000a, 2000b). It is unknown whether a similar system for providing hydrogen peroxide operates during sclerotization of insect cuticles.

4.4.4.4. *ortho*-Quinones and *para*-Quinone Methides

The three types of oxidases, *o*-diphenoloxidases, laccases, and peroxidases, can all oxidize NADA and NBAD to their respective *o*-quinones. The *o*-quinones are reactive compounds, they can spontaneously form adducts by reaction with available nucleophilic groups, and they can serve as substrates for other cuticular enzymes, *o*-quinone and *p*-quinone

methide isomerases, catalyzing isomerization to *p*-quinone methides, which also react readily with nucleophilic compounds. In the traditional sclerotization scheme, suggested by Pryor (1940a, 1940b), *o*-quinones were supposed to react preferably with ε-amino groups from lysine residues, but solid state nuclear magnetic resonance (NMR) studies have shown that the imidazole group in histidine residues is the preferred cuticular target for reaction with quinones (Schaefer *et al*., 1987; Christensen *et al*., 1991).

Incubation of blowfly (*Sarcophaga bullata)* larval cuticle with NADA and *N*-acetylcysteine resulted in formation of an adduct (**26**) where the sulfur atom is linked to the 5-position of the NADA moiety (Sugumaran *et al*., 1989a), indicating that SH-groups are good acceptors for oxidized NADA (**Figure 5**). Electrochemical oxidation of dopamine to dopamine quinone in the presence of *N*-acetylcysteine gave a mixture of C-5 and C-2 (**27**) monoadducts together with some of the disubstituted product, 2,5-S,S′-di(*N*-acetylcysteinyl)dopamine (**28**) (Xu *et al*., 1996a; Huang *et al*., 1998). Since the monoadducts are more readily oxidized to quinones than is the parent catechol, a monoadduct formed between an oxidized catechol and a protein-linked cysteine will be more prone than a free catechol to be reoxidized to quinone, which can react with a cysteine residue in neighboring protein chain to form a covalent cross-link between the proteins. Thus cysteine–catechol based cross-links are possible,

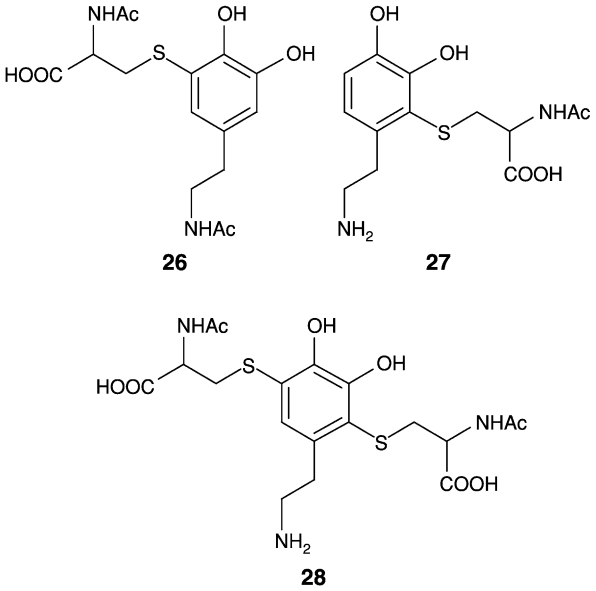


Figure 5 Adducts formed by reaction of *N*-acetylcysteine with the *o*-quinones of NADA and dopamine. **26**, 5-S-(*N*-acetylcysteinyl)-*N*-acetyldopamine; **27**, 2-S-(*N*-acetylcysteinyl)-dopamine; **28**, 2,5-S,S′-di-(*N*-acetylcysteinyl)-dopamine.

but they appear not to play an important role in cross-linking cuticular proteins since these proteins typically contain neither cysteine nor cystine. Among all the cuticular proteins that have been sequenced so far, only a few contain residues of sulfur-containing amino acids, and it seems unlikely that sulfur can have an important role in sclerotization. Perhaps the scarcity of cystine and cysteine in cuticular proteins is related to the readiness with which they react with *o*-quinones. When locust cuticle is incubated with NADA and benzenesulfinic acid, the oxidized NADA is trapped by adduct formation with the sulfinic acid, and before all sulfinic acid has been consumed by adduct formation no *o*-quinone is available for reaction with cuticular proteins or isomerization and further metabolism to polymeric compounds (S.O. Andersen, unpublished data). The presence of significant amounts of free SH-groups in the cuticular matrix proteins could in a similar way delay further metabolism of the *o*-quinones and thereby cause suboptimal sclerotization.

Electrochemical oxidation of NADA to NADA-quinone in the presence of *N*-acetylhistidine gave monoadducts (**Figure 6**), where a nitrogen atom in the imidazole ring is linked to either the C-6 (**29**) or the C-2 ring position (**30**) in NADA, the C-6 position being the preferred position (Xu *et al*., 1996b). Electrochemical characterization of the C-6 and C-2 *N*-acetylhistidine-NADA adducts showed that both adducts are more difficult to oxidize than NADA itself (Xu *et al*., 1996b), indicating that formation of adducts involving two *N*-acetylhistidine residues linked to the same NADA residue is not very likely.

Oxidation of a mixture of NADA plus *N*-acetylhistidine by means of larval cuticle from *H. cecropia* gave a mixture of products, and adduct formation to the C-6 ring position as well as to the β-position of the side chain (**31**) was observed (**Figure 6**) (Andersen *et al*., 1991, 1992c). The formation of a side chain adduct indicates activation of the β-position, probably due to formation of the *p*-quinone methide, demonstrating that cuticular oxidation of catechols is more complex than electrochemical oxidation. From acid hydrolysates of *M. sexta* pupal cuticle, adducts have been obtained where histidine is linked to either the C-6 ring position or the β-position of dopamine or to the corresponding positions in DOPET, demonstrating that adduct formation to histidine residues occurs during *in vivo* sclerotization (Xu *et al*., 1997; Kerwin *et al*., 1999), and confirming that DOPET has a role as a sclerotization precursor. Direct evidence that covalent bonds are formed between acyldopamines and

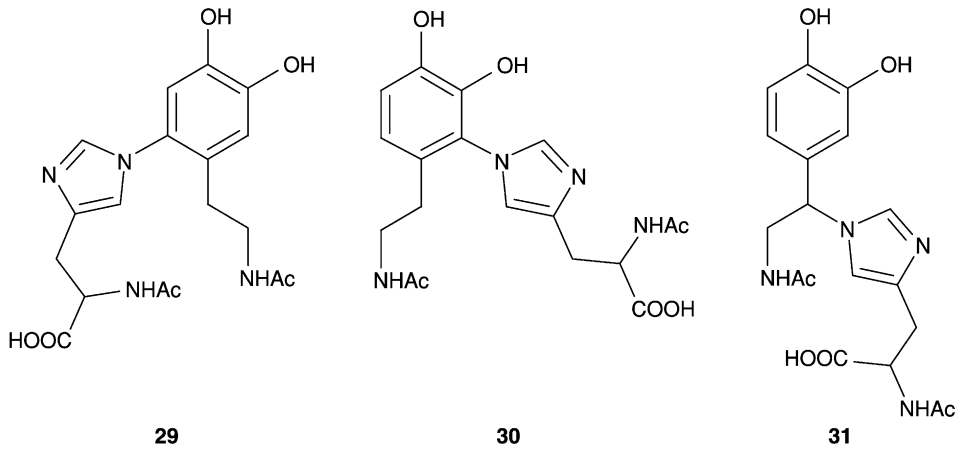


Figure 6 Adducts formed by reaction of *N*-acetylhistidine with NADA-*o*-quinone or NADA-*p*-quinone methide. **29**, 6-(*N*-acetylhistidyl)-*N*-acetyldopamine; **30**, 2-(*N*-acetylhistidyl)-*N*-acetyldopamine; **31**, 7-(*N*-acetylhistidyl)-*N*-acetyldopamine.

histidine residues during sclerotization had previously been obtained by solid state NMR studies, utilizing incorporation of isotopically labeled dopamine, histidine, and β-alanine into sclerotizing pupal cuticle of *M. sexta* (Schaefer *et al.*, 1987; Christensen *et al.*, 1991). The spectra demonstrated the presence of bonds between nitrogen atoms in the imidazole ring of histidine and ring positions and β-position of the side chain of dopamine. The formation of covalent bonds involving the amino group of β-alanine and the ε-amino group of lysine was also indicated. Furthermore, catecholamine-containing proteins have been purified and partially characterized from sclerotizing *M. sexta* pupal cuticle, and NBANE was released from these proteins on mild acid hydrolysis, indicating the presence of a bond between the β-position of the side chain of NBAD and some amino acid residues in the proteins (Okot-Kotber *et al.*, 1996).

Incubation of larval cuticle of *H. cecropia* with NADA together with compounds containing a free amino group, α-*N*-acetyllysine or β-alanine, resulted in the formation of a number of products (**Figure 7**), and adducts were identified with the amino groups linked to the 6-position of the ring in NADA (**32**). These adducts have a quinoid structure, indicating that the primary catecholic product is readily oxidized, in contrast to the adducts formed to the β-position of the side chain (**33**), which are stable in their catecholic form (Andersen *et al.*, 1992b). Adduct formation to either histidine and lysine residues during cuticular oxidation of NADA appears to be so much slower than adduct formation to cysteine that a significant fraction of the *o*-quinones obtained from NADA or NBAD will be available for the isomerase catalyzing further metabolism of the quinones. Prolonged incubation of NADA and

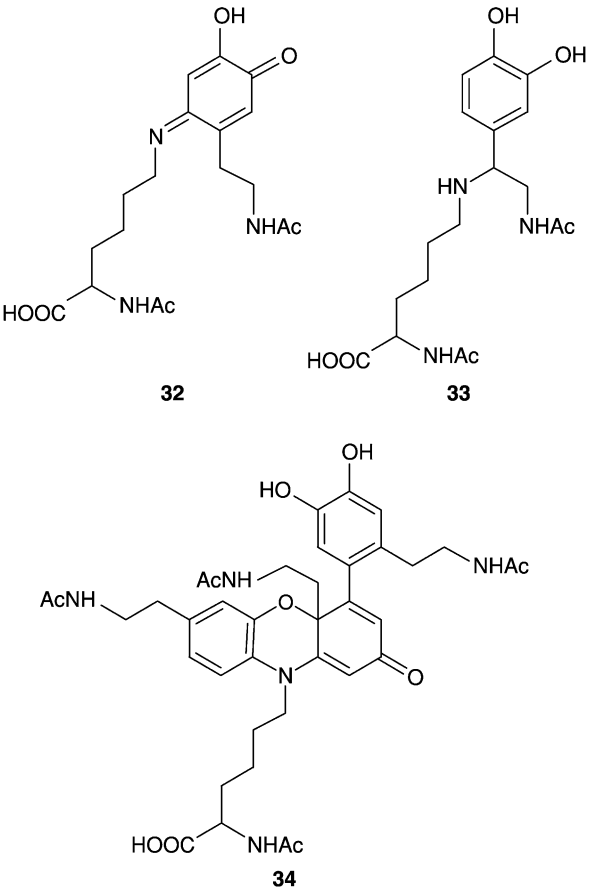


Figure 7 Adducts formed by cuticular oxidation of a mixture of NADA and α-*N*-acetyllysine.

α-*N*-acetyllysine together with cuticle resulted in the formation of 4-phenylphenoxazin-2-ones (**34**), which are composed of three NADA residues joined to one α-*N*-acetyllysine residue (Peter *et al.*, 1992).

The structure of one of the products formed during incubation of insect cuticles with NADA plus *N*-acetylamino acids indicated that *o*-quinone of NADA can react with water to form *N*-acetyl-3,4,6-trihydroxyphenylethylamine (6-hydroxy-NADA), which then couples oxidatively with another NADA residue to give the dimeric compound (**35**) (**Figure 8**) (Andersen *et al.*, 1992a). Corresponding derivatives indicating formation of 6-hydroxy-NADA have so far not been reported from naturally sclerotized cuticle, and its formation *in vitro* may be due to the large excess of water in incubation experiments, in contrast to the relatively low water content in sclerotizing cuticle (30–40% of the cuticular wet weight). The incubation of cuticle together with NADA resulted also in the production of small amounts of a product (**36**) consisting of two NADA-quinones linked together via their 6-positions (Andersen *et al.*, 1992a). A corresponding 2,6′-linked dimer of NADA was suggested as intermediate in the formation of the above mentioned 4-phenylphenoxazin-2-ones (**34**) (Peter *et al.*, 1992).

The side chain β-position in the *p*-quinone methides reacts readily with water, and the lack of stereospecificity in the reaction indicates that the addition of water is nonenzymatic (Peter and Vaupel, 1985). The relative high yields of NANE and NBANE obtained by extraction of various cuticles are probably due to both this reaction and hydrolysis of labile adducts formed with other nucleophiles, such as hydroxyl groups in chitin (see Section 4.4.3.2.2).

The enzyme responsible for quinone isomerization has been partially characterized from larval cuticles from *H. cecropia* (Andersen, 1989c) and the flesh fly *S. bullata* (Sugumaran, 1987; Saul and Sugumaran, 1988), both cuticles belonging to the soft, pliant type. The enzyme is also present in the hemolymph of *S. bullata* (Saul and Sugumaran, 1989a, 1990), where it participates in defense reactions. The enzyme has thus been described from both sclerotizing and nonsclerotizing cuticles and from insect hemolymph, leading to the question whether the same enzyme is involved in both defense reactions and sclerotization. It would be worthwhile to study the temporal and regional distribution of the enzyme in various insect systems.

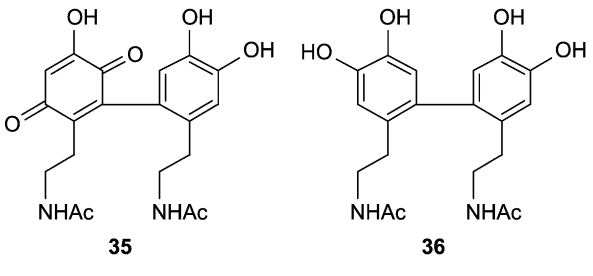


Figure 8 Products obtained by cuticular oxidation of NADA.

4.4.4.5. Dehydro-NADA and Dehydro-NBAD

The *p*-quinone methide formed by isomerization of NADA-*o*-quinone can be further isomerized to dehydro-NADA, a NADA derivative carrying a double bond between the α- and β-carbon atoms in the side chain. The enzyme responsible for this isomerization has been called *N*-acetyldopamine quinone methide/1,2-dehydro-*N*-acetyldopamine tautomerase (Saul and Sugumaran, 1989c). The activity has been reported to be present in larval cuticle of *S. bullata* (Saul and Sugumaran, 1989b, 1989c) and *D. melanogaster* (Sugumaran *et al.*, 1992). A related enzyme activity catalyzing the isomerization of NBAD quinone methide to dehydro-NBAD has been demonstrated in extracts of *C. vicina* larval cuticle (Ricketts and Sugumaran, 1994). It is probably the same enzyme which leads to the formation of dehydro-NADA and dehydro-NBAD. Locust cuticle catalyzes formation of dehydro-NADA from NADA, whereas NBAD is a poor substrate for the locust enzyme (Andersen, 1989d). It appears that several cuticles, which are sclerotized by mixtures of NADA and NBAD, readily convert NADA to the dehydro-derivative, while conversion of NBAD to dehydro-NBAD only occurs to a minor extent, if at all (Andersen, 1989d; Andersen *et al.*, 1996).

It is uncertain whether formation of dehydro-NADA in locust cuticle occurs only via NADA-quinone methide or whether it can also be formed directly from NADA by a route circumventing quinone formation. Inhibition of oxidation of NADA to NADA-quinone resulted in accumulation of dehydro-NADA in locust cuticle, and since it seemed unlikely that accumulation of a product can be caused by decreased production of its precursors, the presence of a specific "NADA-desaturase" was suggested (Andersen and Roepstorff, 1982; Andersen *et al.*, 1996).

The dehydro-NADA formed during cuticular sclerotization can be oxidized by both the *o*-diphenoloxidases and laccases present in cuticle, and the resulting unsaturated quinones react spontaneously with other catechols to give substituted dihydroxyphenyl-dihydrobenzodioxines (XIII) (Andersen and Roepstorff, 1982; Sugumaran *et al.*, 1988). It has not been established whether the unsaturated quinones will also react with nucleophilic residues in cuticular proteins. Locust cuticle will catalyze *in vitro* formation of adducts between dehydro-NADA and catechols, and dihydroxyphenyl-dihydrobenzodioxine derivatives, corresponding to those extracted from naturally sclerotized cuticle, can be formed in this

way (Andersen and Roepstorff, 1982; Andersen, 1985). It appears that the presence of dihydroxyphenyl-benzodioxine derivatives in sclerotized cuticles can be taken as an indication for the presence of a dehydro-NADA forming activity.

It has been suggested that the cuticular diphenol-oxidases, isomerases, and tautomerases occur together in large enzyme complexes, enabling one enzyme to deliver its products directly to the following enzyme for further processing (Andersen *et al.*, 1996; Sugumaran, 1998). The evidence for such complexes is insufficient, but the observation that during *in vitro* incubation with NADA some cuticles convert NADA into benzodioxine dimers with very little formation of NANE, while other cuticles preferably produce NANE, may indicate that the former cuticles contain enzyme complexes with tight coupling between the individual enzymes, so the intermediates will have little chance of reacting with water or the surrounding proteins before they are used as substrates for the next enzyme in the sequence. The sclerotization of some types of cuticle may thus be dominated by reactions involving the dehydro-NADA *o*-quinone, while sclerotization of other types of cuticle is dominated by reactions between the initially generated quinones and the matrix proteins.

The final outcome of the sclerotization process probably depends upon a carefully controlled balance between various enzyme activities and available substrates, a balance that differs between the local cuticular regions according to their mechanical properties. It would be interesting to have quantitative determinations of the relevant enzyme activities in different types of cuticle, and to be able to specifically inhibit the individual enzymes participating in sclerotization to study their individual roles in the total process. Quantitative enzyme activity determinations on pieces of insoluble materials are problematic, due both to slow diffusion of substrates into, and products out of, the pieces and because it is difficult to obtain homogeneous samples for analysis. However, it should be possible to obtain relative values for the activities to be used for comparison of various cuticular regions.

4.4.4.6. Various Catechol Derivatives Obtained from Cuticles

An indication of the quantitative role of the possible sclerotization reaction pathways may be obtained by identification and quantification of the low molecular weight catecholic compounds, which can be extracted from sclerotized cuticles after more or less extensive degradation of the cuticular material. The structures of such compounds may suggest whether they are likely to be products derived from the sclerotizing precursors and perhaps give some indication of the reactions responsible for their generation. The compounds extracted from sclerotized cuticle could represent unused sclerotization precursors, intermediates in the sclerotization process, by-products from the process, and degradation products of protein-bound cross-links and polymers. It is also possible that the extracted compounds have their own distinct functions in the cuticle, quite unrelated to sclerotization; i.e., catechols can thus be precursors for pigments, such as papiliochromes (Umebachi, 1993), or they can function as antioxidants protecting the epicuticular lipids from autoxidation (Atkinson *et al.*, 1973a). 3,4-Dihydroxyphenylacetic acid, which is present in the solid cuticle of many beetle species (Andersen, 1975), could serve the latter purpose, as it apparently takes no part in the sclerotization process (Barrett, 1984, 1990). A detailed study of the formation and fate of the various cuticular compounds is necessary for deciding whether they are related to the sclerotization process.

The mixture of compounds obtained from cuticles by the use of mild extraction methods, such as extraction in distilled water or neutral salt solutions at moderate temperatures, contains compounds that are less modified than those obtained by acidic extraction at elevated temperatures, but a critical and careful interpretation of their structures will be necessary in all cases. Extraction with dilute acids tends to give higher yields of catechols than extraction with water, but the compounds identified in the extracts are often the same (Atkinson *et al.*, 1973b). The higher yield obtained with acids may be due partly to swelling of the cuticular material at low pH values, resulting in easier liberation of trapped compounds, and partly to hydrolysis of acid-labile bonds. Typical extraction products are the sclerotization precursors, NADA and NBAD, and their hydroxylated derivatives, NANE and NBANE, and they can also occur as *O*-glucosyl derivatives. The *O*-glycoside linkage is acid-labile and may not survive prolonged extraction, and the β-hydroxyl group in NANE and NBANE may either be produced by reaction of the *p*-quinone methides of NADA and NBAD with water, or can result from the hydrolysis of the products formed by reaction of a *p*-quinone methide with cuticular proteins and chitin.

3,4-Dihydroxybenzoic acid (**18**) and 3,4-dihydroxybenzaldehyde have been extracted from several types of sclerotized cuticle; they are probably formed by extensive oxidative degradation of the side chain of the sclerotizing precursors, NADA and NBAD. The precise reaction pathway for their

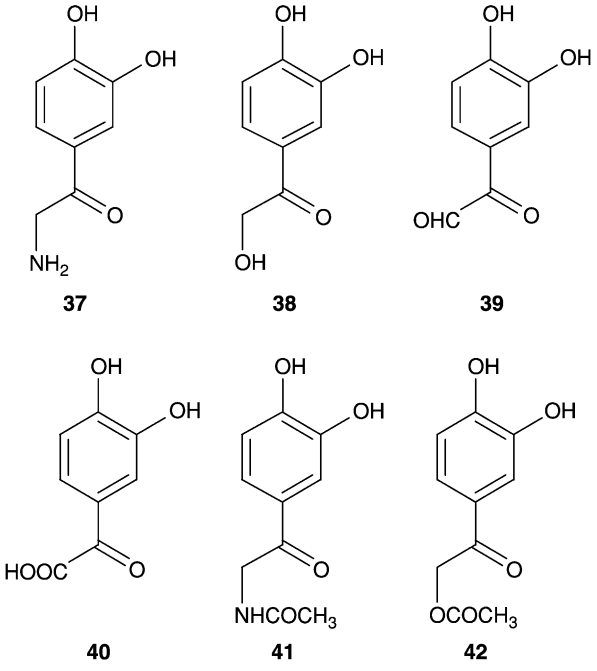


Figure 9 Ketocatechols obtained by mild acid hydrolysis of sclerotized insect cuticle. **37**, arterenone; **38**, 3,4-dihydroxyphenylketoethanol (DOPKET); **39**, 3,4-dihydroxyphenylglyoxal; **40**, 3,4-dihydroxyphenylglyoxylic acid; **41**, *N*-acetylarterenone; **42**, 3,4-dihydroxyphenylketoethylacetate.

formation is not known, but their common occurrence in sclerotized cuticle indicates that the intracuticular environment is highly oxidative during the sclerotization process. The presence of dopamine and norepinephrine in cuticular extracts may be due to deacylation of the *N*-acylated forms, or they may have been transferred directly from epidermal cells to cuticle, either by active transport across the apical cell membrane or by passive leakage through the cell membrane. The black cuticle of the fruitfly mutants *black* and *ebony* contains elevated levels of dopamine (Wright, 1987), which probably have been transferred to the cuticle in the nonacylated state.

Quite large amounts of ketocatechols (Figure 9), such as arterenone (**37**), DOPKET (**38**), and *N*-acetylarterenone (**41**), as well as 3,4-dihydroxyphenylglyoxal (**39**), 3,4-dihydroxyglyoxylic acid (**40**), and 3,4-dihydroxyphenylketoethylacetate (**42**), can be obtained from sclerotized cuticle by treatment with dilute acids (Andersen, 1970, 1971; Andersen and Barrett, 1971; Andersen and Roepstorff, 1978). The yields of ketocatechols can amount to several percent of the cuticular dry weight (Andersen, 1975; Barrett, 1977), and the type of ketocatechols released depends upon the exact conditions of hydrolysis. It has been argued that they are all

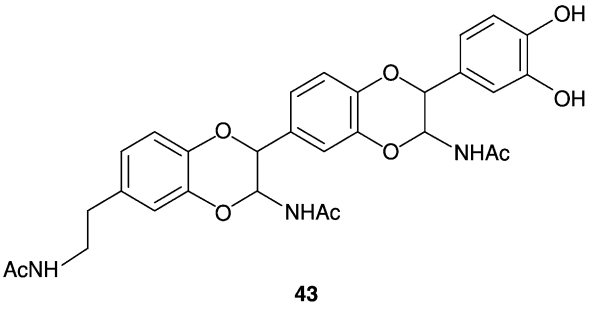


Figure 10 **43**, Trimeric dihydrobenzodioxine derivative from sclerotized locust cuticle.

degradation products of a common precursor in the cuticle (Andersen, 1971).

More complex catechol derivatives can be obtained by extracting sclerotized cuticle at relatively mild conditions, such as concentrated formic acid at ambient temperature or boiling dilute acetic acid. From such extracts a number of dimeric compounds of the dihydroxyphenyl-dihydrobenzodioxine type were isolated and identified (Andersen and Roepstorff, 1981; Roepstorff and Andersen, 1981), and later a trimeric compound (**44**) (**Figure 10**) was obtained by formic acid extraction of sclerotized locust cuticle (Andersen *et al.*, 1992c). Ketocatechols are readily formed when such dimers and oligomers are hydrolyzed with acid (Andersen and Roepstorff, 1981), but the extractable material accounts for only a minor fraction of ketocatechols formed by acid hydrolysis of sclerotized cuticle, and the main fraction of ketocatechols obtainable from cuticle is presumably derived from catecholic material covalently linked to the cuticular proteins and chitin. Since the various dimers can be formed *in vitro* by reaction between oxidized dehydro-NADA and catechols, it is likely that oxidized dehydro-NADA will react with catechols (NADA or NBAD) linked to cuticular proteins to form protein-linked dimers and higher oligomers.

4.4.5. Control of Sclerotization

Sclerotization of insect cuticles varies regionally both with respect to time of initiation, intensity, duration, and balance between NADA- and NBAD-sclerotization, indicating that the process must be controlled locally to give the optimal result.

4.4.5.1. Pre-Ecdysial Sclerotization

In some cuticular regions sclerotization starts in the pharate stage, and these regions attain their final size and shape during the pre-ecdysial deposition of cuticular materials (Cottrell, 1964). Pre-ecdysial

sclerotization may be limited to small, local regions, such as mandibles and spines, or larger regions of cuticle, covering thorax, head, and legs, may be stabilized before ecdysis. In the presclerotized regions sclerotization continues after ecdysis, whereby the already stiffened material is further strengthened. Pre-ecdysial sclerotization has been observed in several cuticular regions in adult honeybees (*Apis mellifera adansonii*) (Andersen *et al*., 1981), and the matrix proteins in these regions resist extraction with solvents that do not degrade the cuticle. Acid hydrolysis of pre-ecdysial sclerotized honeybee pharate cuticle yielded significant amounts of ketocatechols, and still more was obtained by hydrolysis of the corresponding cuticular regions from mature worker bees where endocuticle deposition was complete, indicating that the same type of sclerotization is involved in both pre- and post-ecdysial stabilization. The wings were the only body region where no pre-ecdysial sclerotization was observed. Their sclerotization occurred rapidly after ecdysis and was nearly complete when the bees left the cell in which they had pupated, confirming that insect wings are not stabilized until they have been expanded to their proper size after emergence.

4.4.5.2. Post-Ecdysial Sclerotization

Many regions of the cuticle are often expanded to a new and bigger size after the insect has emerged from the old cuticle (Cottrell, 1964). In some insects the period from ecdysis to fully expanded cuticle can be rather prolonged, for instance in flies where the newly emerged adult has to dig its way through the substratum in which it pupariated, but many insects can start cuticular expansion as soon as they have escaped from the exuvium. Release of the neurohormone bursicon from the central nervous system is a signal for initiating general sclerotization after ecdysis. Bursicon has a pronounced influence on the activities of the epidermal cells. It has been reported that lack of bursicon results in the failure of endocuticle deposition as well as melanin production and sclerotization of the cuticle (Fraenkel and Hsiao, 1965; Fogal and Fraenkel, 1969), and it appears that bursicon is involved in the control of tyrosine hydroxylation to DOPA (Seligman *et al*., 1969).

In some insects sclerotization stops when the preecdysially deposited cuticle has become sclerotized to form exocuticle, although deposition of endocuticle continues for several days, resulting in a sclerotized exocuticle and a nonsclerotized endocuticle. In other insects cuticular sclerotization continues during endocuticle deposition with the result that both exo- and endocuticle become sclerotized, but not to the same degree. Sclerotization of femur cuticle in adult locusts (*Schistocerca gregaria*) continues for at least 12 days after ecdysis, and both exo- and endocuticle are sclerotized (Andersen and Barrett, 1971), in contrast to sclerotization of the femur cuticle of fifth instar nymphs of the same species which lasts for only 1 day, and deposition of unsclerotized endocuticle continues for about 4–5 days (Andersen, 1973). Accordingly, the endocuticular proteins are readily extractable from femurs of mature nymphs, but little protein can be extracted from the femurs of mature adult locusts. The difference in duration of sclerotization in locust nymphs and adults is probably related to the different fate of these two types of cuticle. The nymphal cuticle will to a large extent be degraded in preparation for the next ecdysis, and sclerotized cuticle is more resistant to enzymatic degradation than nonsclerotized cuticle. The adult cuticle has to last for the rest of the life of the animal, and there is no apparent advantage in having an easily degraded endocuticle. The leg cuticle of adult locusts is also exposed to stronger mechanical forces than the cuticle of nymphal legs, and it may therefore be an advantage for adult locusts to have both layers of the leg cuticle stabilized, but not to the same degree. A similar sclerotization difference between adult cuticle and cuticle in younger instars is probably present in other insect species. It has not yet been established how the duration of the post-ecdysial sclerotization is regulated.

4.4.5.3. Puparial Sclerotization

During puparium formation in the higher Diptera the soft, pliable cuticle of the last larval instar is modified to a hard and nondeformable material, its function being mainly to protect the animal during pupal and adult development and not to allow movements. Pronounced regional differentiation in cuticular sclerotization is apparently not essential for puparia. The chemical processes of puparium sclerotization are very similar to those involved in the sclerotization of adult cuticle, but the two systems differ in the way they are controlled (Sekeris, 1991). Puparium sclerotization begins with an increase in ecdysteroid titer, which induces the expression of the enzyme DOPA decarboxylase, catalyzing the decarboxylation of DOPA to dopamine. The latter is then acylated to the sclerotization precursors, NADA and NBAD, a process catalyzed by the enzymes acetyl transferase and β-alanyl transferase, respectively. The precursors are oxidized to *o*-quinones, which are then isomerized to *p*-quinone methides which in turn may be converted to the corresponding dehydro-derivatives (Saul and Sugumaran, 1989a; Sugumaran *et al*., 1992). The flies *Musca autumnalis* and *M. fergusoni* harden

their puparia by the deposition of calcium and magnesium phosphates and not by phenolic sclerotization (Gilby and McKellar, 1976; Darlington *et al*., 1983).

4.4.5.4. Transport of Sclerotization Precursors to the Cuticle

Cuticle of actively moving insects is constantly exposed to external forces which vary in intensity and direction, and to keep the resulting local deformations of the cuticle within functionally acceptable ranges, the mechanical properties of each of the cuticular regions will probably have to be rather precisely regulated, suggesting that the degree of sclerotization of the individual regions is controlled by local mechanisms. In all likelihood such local sclerotization control resides in the underlying epidermal cells. It could operate by regulating the supply of sclerotization precursors to the cuticle, for instance by controlling local synthesis of precursors and/or their uptake from hemolymph into epidermis or by controlling the transport of precursors from epidermal cells to the cuticular matrix where they will be exposed to the enzymes converting them to sclerotization agents. Precursors for sclerotization can be synthesized in the epidermal cells, but it is likely that they can also be produced by other cell types, such as hemocytes and/or fat body cells. Precursors, such as dopamine, NADA, and NBAD, injected into the hemocoel shortly before or during cuticular sclerotization, are rapidly taken up by the epidermal cells and transported into the cuticle, whereas precursors injected several days before the start of sclerotization are mainly degraded or modified by glycosylation or phosphorylation. Such precursor conjugates may either remain in the animal and serve as a reserve pool of sclerotizing material, or they may be excreted via the Malphigian tubules by standard detoxification mechanism.

The *white pupa* mutant of the Mediterranean fruit fly, *Ceratitis capitata,* fails to sclerotize the puparium, but develops normal larval and adult cuticles. The concentrations of the various catecholamines were very low in the mutant puparial cuticle compared to the wild-type strain, whereas the concentrations in the hemolymph of NADA, NBAD, and dopamine were about ten times higher in the mutant than in the wild-type, indicating that the mutant is defective in the system transporting catecholamines from hemolymph to puparial cuticle (Wappner *et al*., 1995).

To control the transport of sclerotizing precursors into the cuticle in the right amounts at the right time the epidermal cells must possess specific transport systems both in the basolateral cell membrane to facilitate uptake from the hemolymph and in the apical cell membrane to control the transport of sclerotizing precursors from the cells into the cuticle. Active diphenoloxidases (laccases) are, at least in some insects, present in unsclerotized pharate cuticle before or at ecdysis (Andersen, 1972, 1979b, and unpublished data), and it appears unlikely that the precursors are transported into these cuticles before sclerotization is initiated following ecdysis. Radioactive NADA injected into fifth instar locust nymphs 1–2 days before ecdysis, when deposition of the pharate adult cuticle occurs, is partly incorporated into the old nymphal cuticle, soon to be discarded, partly retained to be incorporated into the new adult cuticle after ecdysis, and partly excreted as an *O*-glucoside (Andersen, 1974b and unpublished data). The labeled NADA, which after ecdysis was incorporated into the new cuticle, was in the meantime probably stored in the epidermal cells, but other locations, such as the fat body or hemocytes, cannot be disregarded, and the mechanism for transferring NADA into the cuticle is unexplored.

It has been reported that proteins to which catechol derivatives are attached can be transported intact from the hemolymph to the cuticular structure to serve as combined matrix components and sclerotization precursors (Koeppe and Mills, 1972; Koeppe and Gilbert, 1974; Bailey *et al*., 1999), indicating that receptors able to recognize such proteins must be present in the basolateral membrane of the epidermal cells.

4.4.5.5. Balance between Cuticular Enzymes

The type of sclerotization occurring in the local regions will depend in part upon the relative amounts of the two precursor compunds, NADA and NBAD, imported from the epidermal cells, and partly upon the balance between the various enzyme activities in the cuticle. An attempt to study some of these questions was made 30 years ago (Andersen, 1974a, 1974b), but the means available at that time were insufficient to give conclusive answers. NADA labeled with tritium either on the aromatic ring or on the β-position of the side chain was used to study the extent to which the ring and the side chain were involved in adduct formation, based on the assumption that sclerotization via an *o*-quinone will result in tritium release from the ring and that release of tritium from the NADA side chain will occur without intermediate quinone formation. Different types of cuticle were found to differ in their ability to release tritium from the two positions; lightly colored, sclerotized cuticles preferentially released tritium from the side chain, and dark-brown cuticles

released tritium preferentially from the ring system (Andersen, 1974a). It is now evident that tritium release from the β-position of the side chain will occur during formation of the *p*-quinone methide of NADA and will be a measure of the *o*-quinone isomerase activity present, and that tritium release from the aromatic ring will occur during adduct formation between an *o*-quinone and some nucleophile. This will be a measure of the fraction of available NADA that is oxidized by the cuticular diphenoloxidase, but escapes being isomerized to a *p*-quinone methide. Release of tritium from NADA specifically labeled as the α-position of the side chain can presumably be used for obtaining an estimate of the relative importance of the formation of dehydro-NADA during cuticular sclerotization (Andersen, 1991).

Locust (*S. gregaria*) femur cuticle was found to release little tritium from the ring system of NADA, but significant and nearly equal amounts of tritium from the α- and β-positions of the NADA side chain (Andersen, 1974a and unpublished data), indicating that almost all NADA used for sclerotization by this insect is converted to 1,2-dehydro-NADA. Since NADA and NBAD appear to be treated differently during cuticular sclerotization, it will be necessary to use α- as well as β-tritiated forms of both NADA and NBAD to determine the quantitative importance of *p*-quinone methide sclerotization relative to sclerotization involving the dehydro-derivatives. Such determinations have, to the best of our knowledge, not been reported.

4.4.5.6. Intensity of Cuticular Sclerotization

It is difficult to obtain a useful measure of the degree of sclerotization of a given piece of cuticle, but due to its importance for the mechanical properties of cuticles, it would be an advantage to have some means for determining how much NADA and NBAD are incorporated during sclerotization. It is possible that solid state NMR studies can be used for the measurements (Schaefer *et al.*, 1987; Christensen *et al.*, 1991), but such techniques are not generally available. Attempts to determine the degree of sclerotization of various cuticular samples have been made by measuring the amounts of ketocatechols released by acid hydrolysis from cuticular samples (Andersen, 1974b, 1975; Barrett, 1977, 1980), but this method only determines that part of total sclerotization which is due to reactions involving activated dehydro-NADA and dehydro-NBAD. A useful measure for the quantitative role of NBAD in sclerotization might be obtained by determining the amounts of β-alanine released by acid hydrolysis; it appears that some of the amino groups in β-alanine take part in the sclerotization process and cannot be released by hydrolysis (Christensen *et al.*, 1991), but this will probably only represent a minor fraction of the total amount of β-alanine incorporated. Although analysis for ketocatechols as well as β-alanine cannot give absolute values for the types of sclerotization they represent, relative values can be useful for comparing different cuticular samples.

Ketocatechols were not obtained from adult femur cuticle of the locust *S. gregaria* when the samples were taken just after the animals had emerged from their exuvium, as no sclerotization had yet occurred. Later, during maturation of the locusts, a steady increase in the yield of ketocatechols was observed, reaching a constant level after about 1 week, when the ketocatechol yield was about $0.2\,\mu mol\,mg^{-1}$ dry cuticle (Andersen and Barrett, 1971), indicating a close relationship between ketocatechol formation and sclerotization in locust cuticle.

Determination of ketocatechol release from various types of cuticle from mature locusts (*S. gregaria*) showed that all cuticular types yielded some ketocatechols upon hydrolysis: from 10–12 days old locusts the lowest yields were obtained from abdominal intersegmental membranes (0.20% of the dry weight), abdominal sclerites gave 0.38% of the dry weight, and the highest values were obtained from the dorsal mesothorax (5.25% of the dry weight) and the mandibles (3.36% of the dry weight) (Andersen, 1974b), in agreement with the expectation that the regions where strength and hardness are essential for proper function are also the regions giving the highest yields of ketocatechols.

Schistocerca gregaria mature nymphal cuticle gave much lower amounts of ketocatechols than corresponding samples from mature adults. The only exception was the nymphal mandibles; 3.74% of their dry weight was recovered as ketocatechols, comparable to what was obtained from the adult mandibles (Andersen, 1974b). The relative low yield of ketocatechols obtained from nymphal cuticle are probably related to the nymphs weighing less than adults, their leg cuticle being exposed to smaller forces during walking and jumping than the legs of adults, and the nymphal wings and thorax cuticle not being exposed to deforming forces comparable to those for which adult cuticle is exposed to during flight.

Quantitative ketocatechol determinations have been performed on a few other insect species, such as the beetles *Tenebrio molitor* and *Pachynoda epphipiata* (Andersen, 1975). Various parts of the exuviae of a cicada, *Tibicen pruinosa,* have also been analyzed (Barrett, 1977), and most regions

gave ketocatechol values between 6% and 7% of the exuvial dry weight. The exuvial cuticle from the cicada compound eyes gave a ketocatechol yield (11.8% of the dry weight) that was much higher than that obtained from any of the other cuticular regions. The high values obtained from exuviae agree with the notion that sclerotization is generally more pronounced in exocuticle than in endocuticle. The very high values obtained from exuvial cornea of cicada taken together with the low, but significant values from cornea and intersegmental membrane from adult locusts (Andersen, 1974b) indicate that these samples of soft cuticles, often considered to be unsclerotized, are sclerotized in the epicuticle and maybe also in the procuticular layer immediately beneath the epicuticle.

An attempt has been made to obtain a measure of the relative degree of sclerotization by determining the amounts of radioactivity incorporated into the various cuticular regions after injection of a single dose of labeled dopamine or NADA (Andersen, 1974b). This can give a picture of how the various regions compete for the available sclerotization precursor at the time of injection, but it cannot measure the total sclerotization occurring in the regions. Soon after ecdysis labeled dopamine was injected into nymphal and adult *S. gregaria*, and good agreement was observed between the amounts of radioactivity incorporated into the various cuticular regions and the amounts of ketocatechols recovered after acid hydrolysis of corresponding regions from noninjected animals. This indicates that yield of ketocatechols can be a useful measure of sclerotization of locust cuticle. The only exception was that significantly less ketocatechol was obtained from the adult mandibles than expected considering their ability to incorporate radioactive dopamine.

The regional pattern of incorporation of labeled NADA into locust cuticle was the same as that observed for labeled dopamine, and both patterns changed similarly during maturation of the animals (Andersen, 1974b). No convincing correlation was observed between the rate at which the various cuticular regions released tritium from β-labeled NADA *in vitro* and either the uptake of labeled dopamine and NADA *in vivo* or the yield of ketocatechols obtained by hydrolysis of these regions. These results suggest that it is not the amounts of sclerotizing enzymes which are the main determining factor for the degree of sclerotization, but the local availability of sclerotizing precursors. Similar attempts to determine the rate limiting factors for cuticular sclerotization have, to the best of our knowledge, not been performed on other insect species.

4.4.6. Comparative Aspects

4.4.6.1. Cuticular Darkening

It is an old observation that cuticular sclerotization often is accompanied by a darkening of the cuticle, resulting in various shades of brown or black coloration. The black colors are presumably always due to deposition of melanins formed by oxidation of dopamine or DOPA (see Section 4.4.3.2.1), and melanin formation can apparently be blocked without interference with sclerotization. Melanins are not deposited in the cuticle of an albino mutant of the locust *S. gregaria*, but cuticular sclerotization appears not to be affected (Malek, 1957; Karlson and Schlossberger-Raecke, 1962). Further, injection of the phenoloxidase inhibitor phenylthiourea into larvae of *Protophormia terraenovae* prevents melanin deposition during puparium formation but does not affect hardening or the appearance of brown color (Dennell, 1958).

The formation of brown colors appears to be more directly linked to cuticular sclerotization, and various suggestions have been put forward to explain why some cuticles become brown during sclerotization while other cuticles remain colorless. It has been suggested that a correlation exists between the use of NBAD as a sclerotization precursor and the intensity of brown color of the sclerotized cuticle (Brunet, 1980; Morgan *et al.*, 1987; Hopkins and Kramer, 1992), and it has also been suggested that a dark-brown cuticular color indicates that the ring system of the sclerotization precursors is linked directly to cuticular proteins via quinone formation, whereas formation of links between the NADA side chain and the proteins results in a colorless cuticle (Andersen, 1974a). That suggestion was based upon the observation that colorless or lightly colored cuticles preferentially release tritium from the side chain of NADA and that dark-brown cuticles preferentially release tritium from the ring positions.

The colorless benzodioxine-type compounds, which on acid hydrolysis yield ketocatechols, are preferentially formed from NADA and only to a minor extent from NBAD, as NBAD appears to be a poor substrate for the enzymatic activities responsible for introducing a double bond into the side chain (Andersen, 1989d). During sclerotization NBAD can be expected to form links to cuticular proteins via both ring positions and the β-position of the side chain, and some of the β-alanyl amino groups may react with quinones, while most of the NADA residues will be processed to dehydro-NADA, resulting in formation of colorless, protein-linked dihydroxyphenyl-benzodioxine derivatives. The formation of

brown color during sclerotization can thus depend both on the amounts of NBAD available and the balance between quinone-forming enzymes and the enzyme activity catalyzing formation of a double bond in the side chain (Andersen, 1989b).

4.4.6.2. Cuticular Sclerotization in Insects Compared to That in Other Arthropods

The release of ketocatechols during acid hydrolysis can be used as an indication of the involvement of dehydro-NADA in sclerotization, and makes it possible to determine how widespread the occurrence of this variant of cuticular sclerotization is. Ketocatechols have been obtained in varying amounts from cuticular samples of all pterygote insects studied so far, and especially the wings were found to be a good source of ketocatechols. None of the apterygote insects analyzed, representing Thysanura, Collembola, and Diplura, gave any measurable amounts of ketocatechols, and neither did the sclerotized cuticle of noninsectan arthropods, such as Decapoda, Isopoda, Araneae, Xiphosura, and Acarina (Andersen, 1985). The distribution of ketocatechol-yielding material among cuticles indicates that development of the ability to fly occurred in parallel with the development of the use of dehydro-NADA in sclerotization, resulting in a form of sclerotin which combines strength, toughness, and lightness to an optimal degree for flight purposes. The suggestion needs to be investigated in much more detail, and a detailed characterization of the sclerotization process(es) in cuticle of noninsectan arthropods is also needed. Little is known of sclerotization in crustaceans or most other arthropod groups.

4.4.7. Unsolved Problems

The model shown in Figures 1 and 2 for sclerotization of insect cuticle can account for most of the observations and experimental results which have been published during the years of cuticular studies. It appears likely that we have now obtained an understanding of the main features of the chemical processes occurring during sclerotization, but some observations are difficult to reconcile with the suggested scheme. This may be due to faulty observations or errors in interpretation, but could also represent variations of the scheme developed to serve specialized demands in some types of cuticle, or may represent essential, but unrecognized elements in the general sclerotization scheme. In any case, these observations deserve critical study before we can consider our understanding of the chemistry of sclerotization complete. A weakness in the suggested sclerotization scheme is that it is a combination of individual reactions that have been characterized in different types of cuticles, and we have not yet obtained sufficient evidence that all these reactions, and only these reactions, are of general occurrence in cuticular sclerotization.

4.4.7.1. Alternative Pathway for Dehydro-NADA Formation?

In Figure 2 the sclerotization process is depicted as a linear reaction chain, where reactive intermediates formed at different steps in the chain can react with cuticular proteins, and it is supposed that the sclerotization patterns in different cuticular types depend partly upon the absolute and relative amounts of the two sclerotization precursors, NADA and NBAD, entering the chain at the beginning of the process and partly upon the extent to which the intermediate products will react with cuticular proteins or will continue along the reaction chain. If the *o*-quinone isomerase is present in larger amounts than the *p*-quinone methide isomerase one should expect that the quinone methide will be produced more rapidly than it is isomerized, favoring reaction between the quinone methide and proteins, whereas if the opposite is the case only a small fraction of the NADA *p*-quinone methide produced will have an opportunity to react with proteins before being isomerized to dehydro-NADA.

Inhibition of laccase in locust cuticle by Na-azide results in a decrease in the consumption of NADA and an accumulation of dehydro-NADA in the incubation medium. As dehydro-NADA is a better substrate for the cuticular laccase than NADA, it was unexpected that Na-azide inhibits the laccase-catalyzed oxidation of dehydro-NADA much more than oxidation of NADA (Andersen, 1989a). One possible explanation could be that dehydro-NADA can be formed not only by isomerization of the *p*-quinone methide, but directly from NADA by means of a special enzyme, a desaturase, as suggested by Andersen *et al.* (1996). Another possible explanation could be that different enzymes catalyze the oxidation of NADA and dehydro-NADA, and that the dehydro-NADA oxidizing enzyme is much more sensitive to azide inhibition than the NADA oxidizing enzyme. This could result in accumulation of dehydro-NADA, but will demand a strict compartmentalization for the two enzyme activities, as the surplus of dehydro-NADA should not have access to the NADA-oxidizing laccase.

4.4.7.2. Extracuticular Synthesis of Catechol–Protein Conjugates for Sclerotization?

Another question to be studied in more detail is to what extent protein-bound catecholic derivatives

are transferred from hemolymph to cuticle to participate in sclerotization. It has been reported that the epidermal cells can transfer proteins, arylphorins, from hemolymph to the cuticular compartment (Scheller *et al*., 1980; Schenkel *et al*., 1983; König *et al*., 1986; Peter and Scheller, 1991), and apparently such transfer can also occur for proteins to which catechols are covalently bound (Koeppe and Mills, 1972; Koeppe and Gilbert, 1974; Bailey *et al*., 1999). If the protein-bound catechols are oxidized to quinones inside the cuticle during sclerotization, it is reasonable to assume that they will react with residues in the other proteins present in the cuticle and thus participate in sclerotin formation, but so far it is not known for certain how, where, and whether such protein–catechol conjugates are formed, and whether such conjugates after transfer to the cuticle will take part in sclerotization. Diffusion problems and steric hindrance may make it difficult for the catecholic residues to access the cuticular diphenoloxidases, but it is possible that the catechols are oxidized by encountering small, easily diffusable quinones formed by enzyme catalyzed oxidation of free catechols.

4.4.7.3. Importance of Cuticular Dehydration?

Fraenkel and Rudall (1940) reported a significant decrease in cuticular water content in connection with puparium sclerotization in blowflies, and it was later argued that formation of covalent cross-links cannot fully explain the changes in mechanical properties occurring during sclerotization and that controlled dehydration of the cuticular matrix may be the most important factor in the stabilization of cuticle (Hillerton and Vincent, 1979; Vincent and Hillerton, 1979; Vincent, 1980). Dehydration may be caused by increased hydrophobicity of cuticular proteins due to reaction with the enzymatically formed quinones, by filling the initially water-filled interstices between proteins with polymerized catechols, and by water being actively transported out of the cuticle by some transport system residing in the epidermal apical cell membrane. All three mechanisms are probably involved in cuticular dehydration. Water transport coupled to active transport of ions appears to be the process which can be most precisely controlled and may be the most important dehydration mechanism.

4.4.7.4. Lipids and Sclerotization?

Lipids may play an essential role in cuticular sclerotization (Wigglesworth, 1985, 1988), a possibility that should be studied in more detail. The epicuticle consists mainly of proteins and lipids connected to each other to form a thin, extremely resistant, and inextractable layer. It is not known how the lipids and proteins are linked together, but the resistance towards hydrolytic degradation indicates that it is not only ester bonds which are involved, and that stable carbon–carbon bonds between lipid molecules and between lipids and proteins may play an important role. Semiquinones and other free radicals can easily be formed during enzyme catalyzed oxidation of catechols to quinones, and free radicals may react with unsaturated lipids resulting in stable lipophenolic complexes. Such reactions could be part of the stabilization of the epicuticle, and they could also contribute to making the connections between epicuticle and the underlying procuticle more stable and secure.

References

Andersen, S.O., **1966**. Covalent cross-links in a structural protein, resilin. *Acta Physiol. Scand. 66* (Suppl. 263), 1–81.

Andersen, S.O., **1970**. Isolation of arterenone (2-amino-3′,4′-dihydroxyacetophenone) from hydrolysates of sclerotized insect cuticle. *J. Insect Physiol. 16*, 1951–1959.

Andersen, S.O., **1971**. Phenolic compounds isolated from insect hard cuticle and their relationship to the sclerotization process. *Insect Biochem. 1*, 157–170.

Andersen, S.O., **1972**. An enzyme from locust cuticle involved in the formation of cross-links from *N*-acetyldopamine. *J. Insect Physiol. 18*, 527–540.

Andersen, S.O., **1973**. Comparison between the sclerotization of adult and larval cuticle in *Schistocerca gregaria*. *J. Insect Physiol. 19*, 1603–1614.

Andersen, S.O., **1974a**. Evidence for two mechanisms of sclerotization in insect cuticle. *Nature 251*, 507–508.

Andersen, S.O., **1974b**. Cuticular sclerotization in larval and adult locusts, *Schistocerca gregaria*. *J. Insect Physiol. 20*, 1537–1552.

Andersen, S.O., **1975**. Cuticular sclerotization in the beetles, *Pachynoda epphipiata* and *Tenebrio molitor*. *J. Insect Physiol. 21*, 1225–1232.

Andersen, S.O., **1978**. Characterization of trypsin-solubilized phenoloxidase from locust cuticle. *Insect Biochem. 8*, 143–148.

Andersen, S.O., **1979a**. Biochemistry of insect cuticle. *Annu. Rev. Entomol. 24*, 29–61.

Andersen, S.O., **1979b**. Characterization of the sclerotization enzyme(s) in locust cuticle. *Insect Biochem. 9*, 233–239.

Andersen, S.O., **1985**. Sclerotization and tanning of the cuticle. In: Kerkut, G.A., Gilbert, L.I. (Eds.), Comprehensive Insect Physiology, Biochemistry, and Pharmacology, vol. 3. Pergamon, Oxford, pp. 59–74.

Andersen, S.O., **1989a**. Enzymatic activities in locust cuticle involved in sclerotization. *Insect Biochem. 19*, 59–67.

Andersen, S.O., **1989b**. Enzymatic activities involved in incorporation of *N*-acetyldopamine into insect cuticle during sclerotization. *Insect Biochem. 19*, 375–382.

Andersen, S.O., **1989c**. Investigation of an *ortho*-quinone isomerase from larval cuticle of the American silkmoth, *Hyalophora cecropia*. *Insect Biochem. 19*, 803–808.

Andersen, S.O., **1989d**. Oxidation of *N*-beta-alanyldopamine by insect cuticles and its role in cuticular sclerotization. *Insect Biochem. 19*, 581–586.

Andersen, S.O., **1990**. Sclerotization of insect cuticle. In: Ohnishi, E., Ishizaki, H. (Eds.), Molting and Metamorphosis. Japan Scientific Societies Press, Tokyo, pp. 133–155.

Andersen, S.O., **1991**. Sclerotization. In: Binnington, K., Retnakaran, A. (Eds.), Physiology of the Insect Epidermis. CSIRO, Melbourne, pp. 123–140.

Andersen, S.O., Barrett, F.M., **1971**. The isolation of ketocatechols from insect cuticle and their possible role in sclerotization. *J. Insect Physiol. 17*, 69–83.

Andersen, S.O., Jacobsen, J.P., Bojesen, G., Roepstorff, P., **1992a**. Phenoloxidase catalyzed coupling of catechols: identification of novel coupling products. *Biochim. Biophys. Acta 1118*, 134–138.

Andersen, S.O., Jacobsen, J.P., Roepstorff, P., **1992b**. Coupling reactions between amino compounds and *N*-acetyldopamine catalyzed by cuticular enzymes. *Insect Biochem. Mol. Biol. 22*, 517–527.

Andersen, S.O., Jacobsen, J.P., Roepstorff, P., Peter, M.G., **1991**. Catecholamine–protein conjugates: isolation of an adduct of *N*-acetylhistidine to the side chain of *N*-acetyldopamine from an insect-enzyme catalyzed reaction. *Tetrahedron Lett. 32*, 4287–4290.

Andersen, S.O., Peter, M.G., Roepstorff, P., **1992c**. Cuticle-catalyzed coupling between *N*-acetylhistidine and *N*-acetyldopamine. *Insect Biochem. Mol. Biol. 22*, 459–469.

Andersen, S.O., Peter, M.G., Roepstorff, P., **1996**. Cuticular sclerotization in insects. *Comp. Biochem. Physiol. B 113*, 689–705.

Andersen, S.O., Roepstorff, P., **1978**. Phenolic compounds released by mild acid hydrolysis from sclerotized cuticle: purification, structure, and possible origin from cross-links. *Insect Biochem. 8*, 99–104.

Andersen, S.O., Roepstorff, P., **1981**. Sclerotization of insect cuticle. 2. Isolation and identification of phenolic dimers from sclerotized insect cuticle. *Insect Biochem. 11*, 25–31.

Andersen, S.O., Roepstorff, P., **1982**. Sclerotization of insect cuticle. 3. An unsaturated derivative of *N*-acetyldopamine and its role in sclerotization. *Insect Biochem. 12*, 269–276.

Andersen, S.O., Thompson, P.R., Hepburn, H.R., **1981**. Cuticular sclerotization in the honeybee (*Apis mellifera adansonii*). *J. Comp. Physiol. 145*, 17–20.

Asano, T., Ashida, M., **2001a**. Cuticular pro-phenoloxidase of the silkworm, *Bombyx mori*: purification and demonstration of its transport from hemolymph. *J. Biol. Chem. 276*, 11100–11112.

Asano, T., Ashida, M., **2001b**. Transepithelially transported pro-phenoloxidase in the cuticle of the silkworm, *Bombyx mori*: identification of its methionyl residues oxidized to methionine sulfoxides. *J. Biol. Chem. 276*, 11113–11125.

Ashida, M., Brey, P.T., **1995**. Role of the integument in insect defense: pro-phenol oxidase cascade in the cuticular matrix. *Proc. Natl Acad. Sci. USA 92*, 10698–10702.

Ashida, M., Dohke, K., **1980**. Activation of pro-phenoloxidase by the activating enzyme of the silk worm, *Bombyx mori*. *Insect Biochem. 10*, 37–47.

Ashida, M., Yamazaki, H.I., **1990**. Biochemistry of the phenoloxidase system in insects: with special reference to its activation. In: Ohnishi, E., Ishizaki, H. (Eds.), Molting and Metamorphosis. Japan Scientific Societies Press, Tokyo, pp. 239–265.

Aso, Y., Kramer, K.J., Hopkins, T.L., Lookhart, G.L., **1985**. Characterization of haemolymph protyrosinase and a cuticular activator from *Manduca sexta* (L.). *Insect Biochem. 15*, 9–17.

Aso, Y., Kramer, K.J., Hopkins, T.L., Whetzel, S.Z., **1984**. Properties of tyrosinase and DOPA quinone imine conversion factor from pharate pupal cuticle of *Manduca sexta* L. *Insect Biochem. 14*, 463–472.

Atkinson, P.W., Brown, W.V., Gilby, A.R., **1973a**. Autoxidation of insect cuticular lipids: stabilization of alkyl dienes by 3,4-dihydricphenols. *Insect Biochem. 3*, 103–112.

Atkinson, P.W., Brown, W.V., Gilby, A.R., **1973b**. Phenolic compounds from insect cuticle: identification of some lipid antioxidants. *Insect Biochem. 3*, 309–315.

Bailey, W.D., Kimbrough, T.D., Mills, R.R., **1999**. Catechol conjugation with hemolymph proteins and their incorporation into the cuticle of the American cockroach, *Perplaneta americana*. *Comp. Biochem. Physiol. C 122*, 139–145.

Barrett, F.M., **1977**. Recovery of ketocatechols from exuviae of last instar larvae of the cicada, *Tibicen pruinosa*. *Insect Biochem. 7*, 209–214.

Barrett, F.M., **1980**. Recovery of phenolic compounds from exuviae of the spruce budworm, *Choristoneura fumiferana* (Lepidoptera: Tortricidae). *Can. Entomol. 112*, 151–157.

Barrett, F.M., **1984**. Metabolic origin of 3,4-dihydroxyphenylacetic acid in *Tenebrio molitor* and differences in its content in larval, pupal, and adult cuticle. *Can. J. Zool. 62*, 1005–1010.

Barrett, F.M., **1987a**. Characterization of phenoloxidases from larval cuticle of *Sarcophaga bullata* and a comparison with cuticular enzymes from other species. *Can. J. Zool. 65*, 1158–1166.

Barrett, F.M., **1987b**. Phenoloxidases from larval cuticle of the sheep blowfly, *Lucilia cuprina:* characterization, developmental changes, and inhibition by antiphenoloxidase antibodies. *Arch. Insect Biochem. Physiol. 5*, 99–118.

Barrett, F.M., **1990**. Incorporation of various putative precursors into cuticular 3,4-dihydroxyphenylacetic acid of adult *Tenebrio molitor*. *Insect Biochem. 20*, 645–652.

Barrett, F.M., **1991**. Phenoloxidases and the integument. In: Binnington, K., Retnakaran, A. (Eds.), Physiology of the Insect Epidermis. CSIRO, Melbourne, pp. 195–212.

Barrett, F.M., Andersen, S.O., **1981**. Phenoloxidases in larval cuticle of the blowfly, *Calliphora vicina*. *Insect Biochem. 11*, 17–23.

Bernays, E.A., Chamberlain, D.J., McCarthy, P., **1980**. The differential effects of ingested tannic acid on different species of Acridoidea. *Entomol. Experiment. Applic. 28*, 158–166.

Bernays, E.A., Woodhead, S., **1982**. Incorporation of dietary phenols into the cuticle in the tree locust *Anacridium melanorhodon*. *J. Insect Physiol. 28*, 601–606.

Binnington, K.C., Barrett, F.M., **1988**. Ultrastructural localization of phenoloxidases in cuticle and haemopoietic tissue of the blowfly *Lucilia cuprina*. *Tissue Cell 20*, 405–419.

Bodnaryk, R.P., Brunet, P.C.J., **1974**. 3-*O*-Hydrosulphato-4-hydroxyphenethylamine (dopamine 3-*O*-sulphate), a metabolite involved in the sclerotization of insect cuticle. *Biochem. J. 138*, 463–469.

Brunet, P.C.J., **1980**. The metabolism of the aromatic amino acids concerned in the cross-linking of insect cuticle. *Insect Biochem. 10*, 467–500.

Brunet, P.C.J., Coles, B.C., **1974**. Tanned silks. *Proc. R. Soc. B 187*, 133–170.

Candy, D.J., **1979**. Glucose oxidase and other enzymes of hydrogen peroxide metabolism from cuticle of *Schistocerca gregaria*. *Insect Biochem. 9*, 661–665.

Chase, M.R., Raina, K., Bruno, J., Sugumaran, M., **2000**. Purification, characterization and molecular cloning of prophenoloxidases from *Sarcophaga bullata*. *Insect Biochem. Mol. Biol. 30*, 953–967.

Christensen, A.M., Schaefer, J., Kramer, K.J., Morgan, T.D., Hopkins, T.L., **1991**. Detection of cross-links in insect cuticle by REDOR NMR spectroscopy. *J. Am. Chem. Soc. 113*, 6799–6802.

Coles, G.C., **1966**. Studies on resilin biosynthesis. *J. Insect Physiol. 12*, 679–691.

Cottrell, C.B., **1964**. Insect ecdysis with particular emphasis on cuticular hardening and darkening. *Adv. Insect Physiol. 2*, 175–218.

Curtis, A.T., Hori, M., Green, J.M., Wolfgang, W.J., Hiruma, K., *et al.*, **1984**. Ecdysteroid regulation of the onset of cuticular melanization in allatectomized and *black* mutant *Manduca sexta* larvae. *J. Insect Physiol. 30*, 597–606.

Czapla, T.H., Hopkins, T.L., Kramer, K.J., **1989**. Catecholamines and related *o*-diphenols in the hemolymph and cuticle of the cockroach *Leucophaea maderae* (F.) during sclerotization and pigmentation. *Insect Biochem. 19*, 509–515.

Czapla, T.H., Hopkins, T.L., Kramer, K.J., **1990**. Cuticular strength and pigmentation of five strains of adult *Blattella germanica* (L.) during sclerotization: correlations with catecholamines, beta-alanine and food deprivation. *J. Insect Physiol. 36*, 647–654.

Czapla, T.H., Hopkins, T.L., Kramer, K.J., Morgan, T.D., **1988**. Diphenols in hemolymph and cuticle during development and cuticle tanning of *Periplaneta americana* (L.) and other cockroach species. *Arch. Insect Biochem. Physiol. 7*, 13–28.

Darlington, M.V., Meyer, H.J., Graf, G., Freeman, T.P., **1983**. The calcified puparium of the face fly, *Musca autumnalis* (Diptera: Muscidae). *J. Insect Physiol. 29*, 157–162.

Dennell, R., **1958**. The amino acid metabolism of a developing insect cuticle: the larval cuticle and puparium of *Calliphora vomitoria*. 3. The formation of the puparium. *Proc. R. Soc. B 148*, 176–183.

Fogal, W., Fraenkel, G., **1969**. The role of bursicon in melanization and endocuticle formation in the adult fleshfly, *Sarcophaga bullata*. *J. Insect Physiol. 15*, 1235–1247.

Fraenkel, G., Hsiao, C., **1965**. Bursicon, a hormone which mediates tanning of the cuticle in the adult fly and other insects. *J. Insect Physiol. 11*, 513–556.

Fraenkel, G., Rudall, K.M., **1940**. A study of the physical and chemical properties of insect cuticle. *Proc. R. Soc. B 129*, 1–35.

Fujimoto, K., Okino, N., Kawabata, S., Iwagana, S., Ohnishi, E., **1995**. Nucleotide sequence of the cDNA encoding the proenzyme of phenoloxidase A_1 of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA 92*, 7769–7773.

Gilby, A.R., McKellar, J.W., **1976**. The calcified puparium of a fly. *J. Insect Physiol. 22*, 1465–1468.

Grossmüller, M., Messner, B., **1978**. Zur quantitativen Verteilung der Peroxidase- und Phenoloxidase-Aktivität in der Kutikula von *Galleria mellonella*-Larven und -Puppen: Ein Beitrag zur Sklerotisierung der Insektenkutikula. *Zool. Jb. Physiol. 82*, 16–22.

Hall, M., Scott, T., Sugumaran, M., Söderhall, K., Law, J.H., **1995**. Proenzyme of *Manduca sexta* phenoloxidase: purification, activation, substrate specificity of the active enzyme and molecular cloning. *Proc. Natl Acad. Sci. USA 92*, 7764–7768.

Han, Q., Fang, J., Ding, H., Johnson, J.K., Christensen, B.M., *et al.*, **2002**. Identification of *Drosophila melanogaster* yellow-f and yellow-f2 proteins as dopachrome-conversion enzymes. *Biochem. J. 368*, 333–340.

Han, Q., Li, G., Li, J., **2000a**. Chorion peroxidase-mediated NADH/O_2 oxidoreduction cooperated by chorion malate dehydrogenase-catalyzed NADH production: a feasible pathway leading to H_2O_2 formation during chorion hardening in *Aedes aegypti* mosquitoes. *Biochim. Biophys. Acta 1523*, 246–253.

Han, Q., Li, G., Li, J., **2000b**. Purification and characterization of chorion peroxidase from *Aedes aegypti* eggs. *Arch. Biochem. Biophys. 378*, 107–115.

Hasson, C., Sugumaran, M., **1987**. Protein cross-linking by peroxidase: possible mechanism for sclerotization of insect cuticle. *Arch. Insect Biochem. Physiol. 5*, 13–28.

Hillerton, J.E., Vincent, J.F.V., **1979**. The stabilization of insect cuticles. *J. Insect Physiol. 25*, 957–963.

Hiruma, K., Riddiford, L.M., **1988**. Granular phenoloxidase involved in cuticular melanization in the tobacco hornworm: regulation of its synthesis in the epidermis by juvenile hormone. *Devel. Biol. 130*, 87–97.

Hopkins, T.L., Kramer, K.J., **1992**. Insect cuticle sclerotization. *Annu Rev. Entomol. 37*, 273–302.

Hopkins, T.L., Morgan, T.D., Aso, Y., Kramer, K.J., **1982**. *N*-β-alanyldopamine: major role in insect cuticle tanning. *Science 217*, 364–366.

Hopkins, T.L., Morgan, T.D., Kramer, K.J., **1984**. Catecholamines in haemolymph and cuticle during larval, pupal and adult development of *Manduca sexta* (L.). *Insect Biochem. 14*, 533–540.

Hopkins, T.L., Morgan, T.D., Mueller, D.D., Tomer, K.B., Kramer, K.J., **1995**. Identification of catecholamine β-glucosides in the hemolymph of the tobacco hornworm, *Manduca sexta* (L.), during development. *Insect Biochem. Mol. Biol. 25*, 29–37.

Hopkins, T.L., Starkey, S.R., Xu, R., Merritt, M.E., Schaefer, J., *et al.*, **1999**. Catechols involved in sclerotization of the grasshopper, *Melanoplus sanguinipes,* and their interactions with cuticular proteins. *Arch. Insect Biochem. Physiol. 40*, 119–128.

Huang, X., Xu, R., Hawley, M.D., Hopkins, T.L., Kramer, K.J., **1998**. Electrochemical oxidation of *N*-acyldopamines and regioselective reactions of their quinones with *N*-acetylcysteine and thiourea. *Arch. Biochem. Biophys. 352*, 19–30.

Jacobs, M.E., **1978**. β-Alanine tanning of *Drosophila* cuticles and chitin. *Insect Biochem. 8*, 37–41.

Jacobs, M.E., **1980**. Influence of β-alanine on ultrastructure, tanning, and melanization of *Drosophila melanogaster* cuticles. *Biochem. Genet. 18*, 65–76.

Jacobs, M.E., **1985**. Role of β-alanine in cuticular tanning, sclerotization, and temperature regulation in *Drosophila melanogaster. J. Insect Physiol. 31*, 509–515.

Karlson, P., Schlossberger-Raecke, I., **1962**. Zum Tyrosinstoffwechsel der Insekten. 8. Die Sklerotisierung der Cuticula bei der Wildform und der Albinomutante von *Schistocerca gregaria* Forsk. *J. Insect Physiol. 8*, 441–452.

Karlson, P., Sekeris, C.E., **1976**. Control of tyrosine metabolism and cuticle sclerotization by ecdysone. In: Hepburn, H.P. (Ed.), The Insect Integument. Elsevier, London, pp. 145–156.

Kawabata, T., Yasuhara, Y., Ochiai, M., Matsuura, S., Ashida, M., **1995**. Molecular cloning of insect pro-phenoloxidase: a copper-containing protein homologous to arthropod hemocyanin. *Proc. Natl Acad. Sci. USA 92*, 7774–7778.

Kawasaki, H., Sato, H., **1985**. The tanning agent in the silk of the Japanese giant silkmoth, *Dictyoploca japonica* Butler. *Insect Biochem. 15*, 681–684.

Kawasaki, H., Yago, M., **1983**. The identification of two *N*-acyldopamine glucosides in the left colleterial gland of the praying mantid, *Tenodera aridifolia sinensis* Saussure, and their role in the oothecal sclerotization. *Insect Biochem 13*, 267–271.

Kayser, H., **1985**. Pigments. In: Kerkut, G.A., Gilbert, L.I. (Eds.), Comprehensive Insect Physiology, Biochemistry, and Pharmacology, vol. 10. Pergamon, Oxford, pp. 367–415.

Kayser-Wegmann, I., **1976**. Differences in black pigmentation in lepidopteran cuticles as revealed by light and electron microscopy. *Cell Tissue Res. 171*, 513–521.

Kayser-Wegmann, I., Kayser, H., **1983**. Black pigmentation of insect cuticles: a view based on microscopic and autoradiographic studies. In: Scheller, K. (Ed.), The Larval Serum Proteins of Insects: Function, Biosynthesis, Genetics. Georg Thieme Verlag, Stuttgart, pp. 151–167.

Kerwin, J.L., Turecek, F., Xu, R., Kramer, K.J., Hopkins, T.L., *et al.*, **1999**. Mass spectrometric analysis of catechol-histidine adducts from insect cuticle. *Anal. Biochem. 268*, 229–237.

Koeppe, J.K., Gilbert, L.I., **1974**. Metabolism and protein transport of a possible pupal tanning agent in *Manduca sexta. J. Insect Physiol. 20*, 981–992.

Koeppe, J.K., Mills, R.R., **1972**. Hormonal control of tanning by the American cockroach: probable bursicon mediated translocation of protein-bound phenols. *J. Insect Physiol. 18*, 465–469.

König, M., Agrawal, O.P., Schenkel, H., Scheller, K., **1986**. Incorporation of calliphorin into the cuticle of the developing blowfly, *Calliphora vicina. Roux's Arch. Devel. Biol. 195*, 296–301.

Kramer, K.J., Hopkins, T.L., **1987**. Tyrosine metabolism for insect cuticle tanning. *Arch. Insect Biochem. Physiol. 6*, 279–301.

Kramer, K.J., Morgan, T.D., Hopkins, T.L., Roseland, C.R., Aso, Y., *et al.*, **1984**. Catecholamines and β-alanine in the red flour beetle, *Tribolium castaneum*: roles in cuticle sclerotization and melanization. *Insect Biochem. 14*, 293–298.

Krueger, R.A., Kramer, K.J., Hopkins, T.L., Speirs, R.D., **1989**. *N*-β-alanyldopamine levels and synthesis in integument and other tissues of *Manduca sexta* (L.) during the larval–pupal transformation. *Insect Biochem. 19*, 169–175.

Li, J., Hodgeman, B.A., Christensen, B.M., **1996**. Involvement of peroxidase in chorion hardening in *Aedes aegypti. Insect Biochem. Mol. Biol. 26*, 309–317.

Locke, M., **1969**. The localization of a peroxidase associated with hard cuticle formation in an insect, *Calpodes ethlius* Stoll, Lepidoptera, Hesperiidae. *Tissue and Cell 1*, 555–574.

Malek, S.R.A., **1957**. Sclerotization and melanization: two independent processes in the cuticle of the desert locust. *Nature 180*, 237.

Manthey, M.K., Pyne, S.G., Truscott, R.J.W., **1992**. Involvement of tyrosine residues in the tanning of proteins by 3-hydroxyanthranilic acid. *Proc. Natl Acad. Sci. USA 89*, 1954–1957.

Messner, B., Janda, V., **1991**. Der Peroxidase-Nachweis in der Kutikula von *Galleria mellonella* (Insecta,

Lepidoptera) und *Protophormia terraenovae* (Insecta, Diptera): Ein Beitrag zur Sklerotisierung der Insektenkutikula. *Zool. Jb. Physiol. 95*, 31–37.

Messner, B., Kerstan, U., **1991**. Die Peroxidase im Sklerotisierungsprozess von Hartteilen verschiedener wirbelloser Tiere. *Zool. Jb. Physiol. 95*, 23–29.

Mindrinos, M.N., Petri, W.H., Galanopoulos, V.K., Lombard, M.F., Margaritis, L.H., **1980**. Crosslinking of the *Drosophila* chorion involves a peroxidase. *Roux's Arch. Devel. Biol. 189*, 187–196.

Morgan, T.D., Hopkins, T.L., Kramer, K.J., Roseland, C.R., Czapla, T.H., *et al.*, **1987**. *N*-β-Alanylnorepinephrine: biosynthesis in insect cuticle and possible role in sclerotization. *Insect Biochem. 17*, 255–263.

Morgan, T.D., Thomas, B.R., Yonekura, M., Czapla, T.H., Kramer, K.J., *et al.*, **1990**. Soluble tyrosinases from pharate pupal integument of the tobacco hornworm, *Manduca sexta* (L.). *Insect Biochem. 20*, 251–260.

Okot-Kotber, B.M., Morgan, T.D., Hopkins, T.L., Kramer, K.J., **1996**. Catecholamine-containing proteins from the pharate pupal cuticle of the tobacco hornworm, *Manduca sexta*. *Insect Biochem. Mol. Biol. 26*, 475–484.

Pau, R.N., Acheson, R.M., **1968**. The identification of 3-hydroxy-4-*O*-β-D-glucosidobenzyl alcohol in the left colleterial gland of *Blaberus discoidalis*. *Biochim. Biophys. Acta 158*, 206–211.

Pérez, M., Wappner, P., Quesada-Allué, L.A., **2002**. Catecholamine-β-alanyl ligase in the medfly *Ceratitis capitata*. *Insect Biochem. Mol. Biol. 32*, 617–625.

Peter, M.G., Andersen, S.O., Hartmann, R., Miessner, M., Roepstorff, P., **1992**. Catechol–protein conjugates: isolation of 4-phenylphenoxazin-2-ones from oxidative coupling of *N*-acetyldopamine with aliphatic amino acids. *Tetrahedron 48*, 8927–8934.

Peter, M.G., Scheller, K., **1991**. Arylphorins and the integument. In: Binnington, K., Retnakaran, A. (Eds.), Physiology of the Insect Epidermis. CSIRO, Melbourne, pp. 113–122.

Peter, M.G., Vaupel, W., **1985**. Lack of stereoselectivity in the enzymatic conversion of *N*-acetyldopamine into *N*-acetylnoradrenaline in insect cuticle. *J. Chem. Soc., Chem. Commun.* 848–850.

Petri, W.H., Wyman, A.R., Kafatos, F.C., **1976**. Specific proteins synthesis in cellular differentiation. 3. The egg shell proteins in *Drosophila melanogaster* and their program of synthesis. *Devel. Biol. 49*, 185–199.

Pryor, M.G.M., **1940a**. On the hardening of the ootheca of *Blatta orientalis*. *Proc. R. Soc. B 128*, 378–392.

Pryor, M.G.M., **1940b**. On the hardening of cuticle of insects. *Proc. R. Soc. B 128*, 393–407.

Ricketts, D., Sugumaran, M., **1994**. 1,2-Dehydro-*N*-β-alanyldopamine as a new intermediate in insect cuticular sclerotization. *J. Biol. Chem. 269*, 22217–22221.

Roepstorff, P., Andersen, S.O., **1981**. Electron impact, chemical ionization and field desorption mass spectrometry of substituted dihydroxyphenyl-benzodioxins isolated from insect cuticle: occurrence of thermal decomposition and oligomerization reactions in the mass spectrometer. *Biomed. Mass Spectrom. 8*, 174–178.

Roseland, C.R., Kramer, K.J., Hopkins, T.L., **1987**. Cuticular strength and pigmentation of rust-red and black strains of *Tribolium castaneum*: correlation with catecholamine and β-alanine content. *Insect Biochem. 17*, 21–28.

Saul, S.J., Sugumaran, M., **1988**. A novel quinone: quinone methide isomerase generates quinone methides in insect cuticle. *FEBS Lett. 237*, 155–158.

Saul, S.J., Sugumaran, M., **1989a**. *o*-Quinone/quinone methide isomerase: a novel enzyme preventing the destruction of self-matter by phenoloxidase-generated quinones during immune response in insects. *FEBS Lett. 249*, 155–158.

Saul, S.J., Sugumaran, M., **1989b**. Characterization of a new enzyme system that desaturates the side chain of *N*-acetyldopamine. *FEBS Lett. 251*, 69–73.

Saul, S.J., Sugumaran, M., **1989c**. *N*-acetyldopamine quinone methide/1,2-dehydro-*N*-acetyldopamine tautomerase: a new enzyme involved in sclerotization of insect cuticle. *FEBS Lett. 255*, 340–344.

Saul, S.J., Sugumaran, M., **1990**. 4-Alkyl-*o*-quinone/2-hydroxy-*p*-quinone methide isomerase from the larval hemolymph of *Sarcophaga bullata*. 1. Purification and characterization of enzyme-catalyzed reaction. *J. Biol. Chem. 265*, 16992–16999.

Schaefer, J., Kramer, K.J., Garbow, J.R., Jacob, G.S., Stejskal, E.O., *et al.*, **1987**. Aromatic cross-links in insect cuticle: detection by solid-state ^{13}C and ^{15}N NMR. *Science 235*, 1200–1204.

Scheller, K., Zimmermann, H.-P., Sekeris, C.E., **1980**. Calliphorin, a protein involved in the cuticle formation of the blowfly, *Calliphora vicina*. *Z. Naturforsch. C 35*, 387–389.

Schenkel, H., Myllek, C., König, M., Hausberg, P., Scheller, K., **1983**. Calliphorin: studies on its biosynthesis and function. In: Scheller, K. (Ed.), The Larval Serum Proteins of Insects: Function, Biosynthesis, Genetics. Georg Thieme Verlag, Stuttgart, pp. 18–39.

Schlossberger-Raecke, I., Karlson, P., **1964**. Zum Tyrosinstoffwechsel der Insekten. 13. Radioautographische Lokalisation von Tyrosinmetaboliten in der Cuticula von *Schistocerca gregaria* Forsk. *J. Insect Physiol. 10*, 261–266.

Sekeris, C.E., **1991**. The role of molting hormone in sclerotization in insects. In: Gupta, A.P. (Ed.), Morphogenetic Hormones of Arthropods: Roles in Histogenesis, Organogenesis and Morphogenesis, vol. 1, part 3. Rutgers University Press, New Brunswick, NJ, pp. 150–212.

Seligman, M., Friedman, S., Fraenkel, G., **1969**. Bursicon mediation of tyrosine hydroxylation during tanning of the adult cuticle of the fly, *Sarcophaga bullata*. *J. Insect Physiol. 15*, 553–561.

Sugumaran, M., **1987**. Quinone methide sclerotization: a revised mechanism for β-sclerotization of insect cuticle. *Bioorgan. Chem. 15*, 194–211.

Sugumaran, M., **1988**. Molecular mechanisms for cuticular sclerotization. *Adv. Insect Physiol. 21*, 179–231.

Sugumaran, M., **1998**. Unified mechanism for sclerotization of insect cuticle. *Adv. Insect Physiol. 27*, 229–334.

Sugumaran, M., Dali, H., Semensi, V., **1989a**. Chemical and cuticular phenoloxidase-mediated synthesis of cysteinyl-catechol adducts. *Arch. Insect Biochem. Physiol. 11*, 127–137.

Sugumaran, M., Giglio, L., Kundzicz, H., Saul, S.P., Semensi, V., **1992**. Studies on the enzymes involved in puparial cuticle sclerotization in *Drosophila melanogaster. Arch. Insect Biochem. Physiol. 19*, 271–283.

Sugumaran, M., Hennigan, B., Semensi, V., Dali, H., **1988**. On the nature of nonenzymatic and enzymatic oxidation of the putative sclerotizing precursor, 1,2-dehydro-*N*-acetyldopamine. *Arch. Insect Biochem. Physiol. 8*, 89–100.

Sugumaran, M., Saul, S.P., Semensi, V., **1989b**. Trapping of transiently formed quinone methide during enzymatic conversion of *N*-acetyldopamine to *N*-acetylnorepinephrine. *FEBS Lett. 252*, 135–138.

Thomas, B.R., Yonekura, M., Morgan, T.D., Czapla, T.H., Hopkins, T.L., *et al.*, **1989**. A trypsin-solubilized laccase from pharate pupal integument of the tobacco hornworm, *Manduca sexta*. *Insect Biochem. 19*, 611–622.

Umebachi, Y., **1993**. The third way of dopamine. *Trends Comp. Biochem. Physiol. 1*, 709–720.

Vincent, J.F.V., **1980**. Insect cuticle: a paradigm for natural composites. *Symp. Soc. Exp. Biol. 34*, 183–209.

Vincent, J.F.V., Hillerton, J.E., **1979**. The tanning of insect cuticle: a critical review and a revised mechanism. *J. Insect Physiol. 25*, 653–658.

Wappner, P., Kramer, K.J., Hopkins, T.L., Merritt, M., Schaefer, J., *et al.*, **1995**. *White pupa:* a *Ceratitis capitata* mutant lacking catecholamines for tanning the puparium. *Insect Biochem. Mol. Biol. 25*, 365–373.

Welinder, B.S., Roepstorff, P., Andersen, S.O., **1976**. The crustacean cuticle. 4. Isolation and identification of cross-links from *Cancer pagurus* cuticle. *Comp. Biochem. Physiol. B 53*, 529–533.

Wigglesworth, V.B., **1985**. Sclerotin and lipid in the waterproofing of the insect cuticle. *Tissue and Cell 17*, 227–248.

Wigglesworth, V.B., **1988**. The source of lipids and polyphenols for the insect cuticle: the role of fat body, oenocytes and oenocytoids. *Tissue and Cell 20*, 919–932.

Wright, T.R.F., **1987**. The genetics of biogenic amine metabolism, sclerotization, and melanization in *Drosophila melanogaster. Adv. Genet. 24*, 127–222.

Xu, R., Huang, X., Hopkins, T.L., Kramer, K.J., **1997**. Catecholamine and histidyl protein cross-linked structures in sclerotized insect cuticle. *Insect Biochem. Mol. Biol. 27*, 101–108.

Xu, R., Huang, X., Kramer, K.J., Hawley, M.D., **1996a**. Characterization of products from the reactions of dopamine quinone with *N*-acetylcysteine. *Bioorgan. Chem. 24*, 110–126.

Xu, R., Huang, X., Morgan, T.D., Prakash, O., Kramer, K.J., *et al.*, **1996b**. Characterization of products from the reactions of *N*-acetyldopamine quinone with *N*-acetylhistidine. *Arch. Biochem. Biophys. 329*, 56–64.

Yago, M., Kawasaki, H., **1984**. The identification of five *N*-acyldopamine glucosides in the left colleterial gland of the praying mantid, *Hierodula patellifera* Serville. *Insect Biochem. 14*, 487–489.

Yago, M., Sato, H., Kawasaki, H., **1984**. The identification of *N*-acyldopamine glucosides in the left colleterial gland of the praying mantids, *Mantis religiosa* L., *Statilia maculata* Thunberg and *Tenodera angustipennis* Saussure. *Insect Biochem. 14*, 7–9.

Yago, M., Sato, H., Oshima, S., Kawasaki, H., **1990**. Enzymic activities involved in the oothecal sclerotization of the praying mantid, *Tenodera sinensis* Saussure. *Insect Biochem. 20*, 745–750.

Yamazaki, H.I., **1969**. The cuticular phenoloxidase in *Drosophila virilis*. *J. Insect Physiol. 15*, 2203–2211.

Yamazaki, H.I., **1972**. Cuticular phenoloxidase from the silkworm *Bombyx mori*: properties, solubilization, and purification. *Insect Biochem. 2*, 431–444.

4.5 Biochemistry of Digestion

W R Terra and C Ferreira, University of São Paulo, São Paulo, Brazil

© 2005, Elsevier BV. All Rights Reserved.

4.5.1. Introduction

Most reviews start with the statement that the field under study has undergone remarkable progress over the last decade and the same can be said about the biochemistry of insect digestion. This growth is a characteristic of science as a whole that on average doubles in size every 15 years (Price, 1963). The growth of knowledge in the biochemistry of insect digestion had a bright start during the first decades of the last century, but slowed down after the development of synthetic chemical insecticides in the 1940s. Later on, with the environmental problems caused by chemical insecticides, new approaches for insect control were investigated. Midgut studies were particularly stimulated after the realization that the gut is a very large and relatively unprotected interface between the insect and its environment. Hence, an understanding of gut function was thought to be essential when developing methods of control that act through the gut,

such as the use of transgenic plants to control phytophagous insects.

Applebaum (1985) in his review on the biochemistry of digestion for the first edition of this series recognized the beginning of the renewed growth of the field. He discussed contemporary research showing that most insect digestive enzymes are similar to their mammalian counterparts, but that insect exotic diets require specific enzymes. In the next decade it became apparent that even enzymes similar to those of mammals have distinct characteristics because each insect taxon deals with food in a special way (Terra and Ferreira, 1994). Since then, the field of digestive physiology and biochemistry has progressed dramatically at the molecular level.

The aim of this chapter is to review the recent and spectacular progress in the study of insect digestive biochemistry. To provide a broad coverage while keeping the chapter within reasonable size limits, only a brief account with key references is given for work done prior to 1994. Papers after 1994 have been selected from those richer in molecular details, and, when they were too numerous, representative papers were chosen, especially when rich in references to other papers. Throughout, the focus is on providing a coherent picture of phenomena and highlighting further research areas. Amino acid residues are denoted by the one-letter code, if in peptides, for the sake of brevity. When mentioned in text with a position number, amino acid residues are denoted by the three-letter code to avoid ambiguity. For consistency, traditional abbreviations, like BAPA for benzoyl-arginine *p*-nitroanilide, have been changed, in the example to B-R-pNA, because the one-letter code for arginine is R.

The chapter is organized into four parts. The first part (Sections 4.5.2 and 4.5.3) tries to establish uniform parameters for studying insect digestive enzymes, providing an overview of the biochemistry of insect digestion, and discusses factors affecting digestive enzymes *in vivo*. The second part (Sections 4.5.4–4.5.6) reviews digestive enzymes with emphasis on molecular aspects, whereas the third part (Sections 4.5.7 and 4.5.8) describes the details of the digestive biochemical process along insect evolution. Finally, the fourth part (Section 4.5.9) discusses data on digestive enzyme secretion mechanisms and control.

4.5.2. Overview of the Digestive Process

4.5.2.1. Initial Considerations

Digestion is the process by which food molecules are broken down into smaller molecules that are absorbed by cells in the gut tissue. This process is controlled by digestive enzymes and is dependent on their localization in the insect gut.

4.5.2.2. Characterization of Digestive Enzymes

Enzyme kinetic parameters are meaningless unless assays are performed in conditions in which enzymes are stable. If researchers adopt uniform parameters and methods, comparisons among similar and different insect species will be more meaningful. A rectilinear plot of product formation (or substrate disappearance) versus time will ensure that enzymes are stable in a given condition. Activities (velocities) calculated from this plot are reliable parameters. According to the International Union of Biochemistry and Molecular Biology, the assay temperature should be 30 °C, except when the enzyme is unstable at this temperature or altered for specific purposes. Owing to partial inactivation, the optimum temperature is not a true property of enzymes and therefore should not be included in the characterization. Enzyme pH optimum should be determined using different buffers to discount the effects of chemical constituents of the buffers and their ionic strength on enzyme activity. The number of molecular forms of a given enzyme should be evaluated by submitting the enzyme preparation to a separation process (gel permeation, ion-exchange chromatography, electrophoresis, gradient ultracentrifugation, etc.), followed by assays of the resulting fractions. Substrate specificity of each molecular form of a given enzyme should be evaluated and substrate preference quantified by determining Vm/Km ratios for each substrate, keeping the amount of each enzyme form constant. Substrate preference expressed as the percentage activity towards a given substrate in relation to the activity upon a reference substrate may be misleading because, in this condition, enzyme activities are determined at different substrate saturations. The isoelectric points of many enzymes can be determined after staining with specific substrates following the separation of the native enzymes on isoelectrofocusing gels.

If enzyme characterization is performed as part of a digestive physiology study, emphasis should be given to enzyme compartmentalization, substrate specificity, and substrate preference, in order to discover the sequential action of enzymes during the digestive process. Knowledge of the effect of pH on enzyme activity is useful in evaluating enzyme action in gut compartments (**Figure 1**) with different pH values. Finally, the determination of the molecular masses of digestive enzymes, associated with the ability of enzymes to pass through the peritrophic membrane, allows estimation of the pore sizes of the peritrophic membrane. Molecular

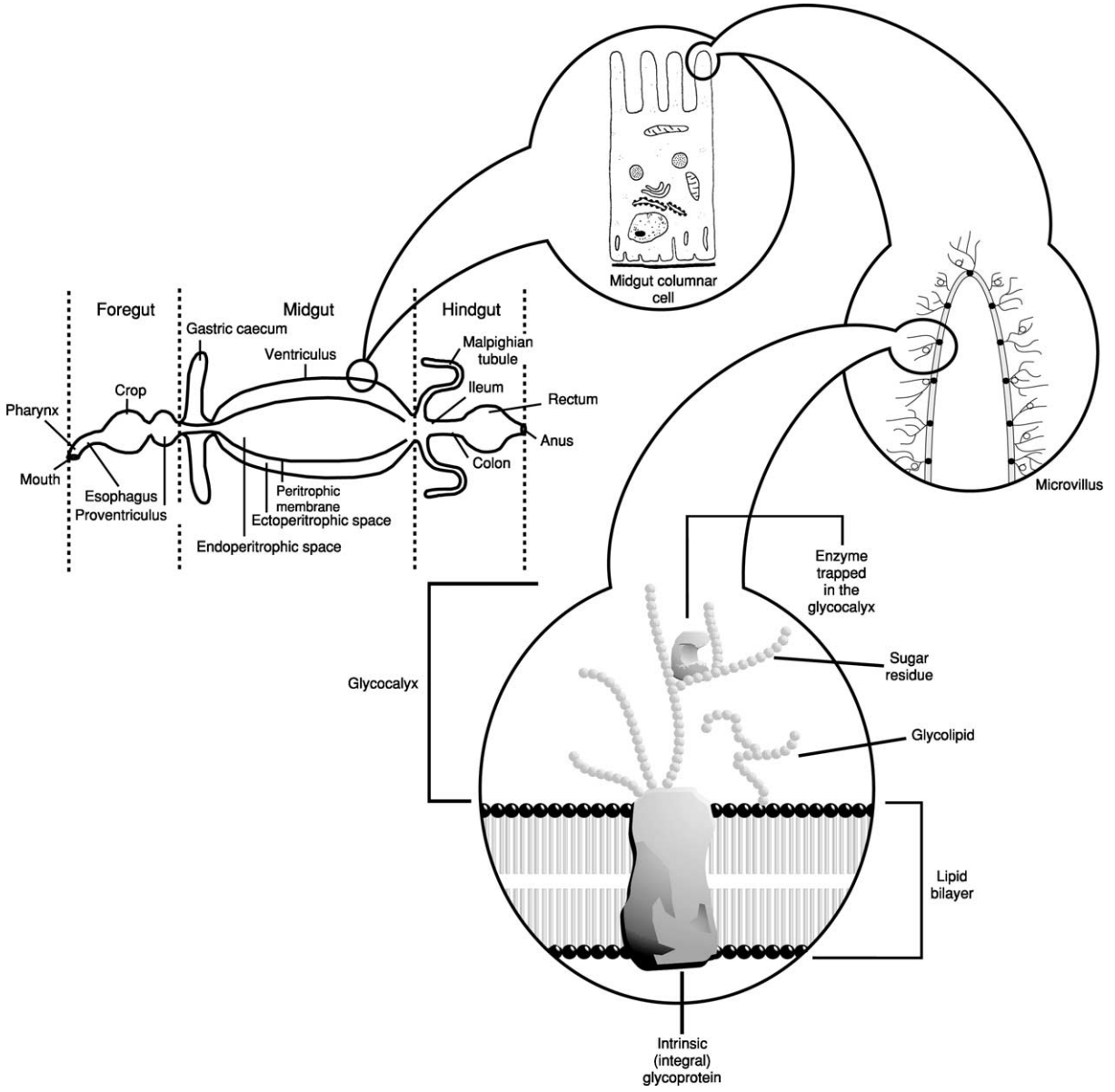


Figure 1 Diagrammatic representation of insect gut compartments. Glycocalyx: the carbohydrate moiety of intrinsic proteins and glycolipids occurring in the luminal face of microvillar membranes.

masses determined in nondenaturing conditions are preferred, since in these conditions the enzymes should maintain their *in vivo* aggregation states (not only their quaternary structures if present). The method of choice in this case is gradient ultracentrifugation.

Complete enzymological characterization requires purification to homogeneity and sequencing. Furthermore, details of the catalytic mechanisms, including involvement of amino acid residues in catalysis and substrate specificity, should be determined. This permits the classification of insect digestive enzymes into catalytic families, and discloses the structural basis of substrate specificities; it will also enable us to establish evolutionary relationships with enzymes from other organisms.

Cloning cDNA sequences encoding digestive enzymes enables the expression of large amounts of recombinant enzymes that may be crystallized or used for the production of antibodies. Antibodies are used in Western blots to identify a specific enzyme in protein mixtures or to localize the enzyme in tissue sections in a light or electron microscope. Enzyme crystals used for resolving three-dimensional (3D) structures (via X-ray diffraction or nuclear magnetic resonance (NMR)) need amounts

of purified enzymes that frequently are difficult to isolate from insects by conventional separation procedures. However, detailed 3D structures are necessary to understand enzyme mechanisms and the binding of inhibitors to enzyme molecules. Alternatively, cDNA may be amenable to site-directed mutagenesis for structure/function studies. Site-directed mutagenesis tests the role of individual amino acid residues in enzyme function or structure. Such knowledge is a prerequisite in developing new biotechnological approaches to control insects via the gut.

4.5.2.3. Classification of Digestive Enzymes

Digestive enzymes are hydrolases. The enzyme classification and numbering system used here is that recommended by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (Enzyme Commission).

Peptidases (peptide hydrolases, EC 3.4) are enzymes that act on peptide bonds and include the proteinases (endopeptidases, EC 3.4.21–24) and the exopeptidases (EC 3.2.4.11–19). Proteinases are divided into subclasses on the basis of catalytic mechanism, as shown with specific reagents or effect of pH. Specificity is only used to identify individual enzymes within subclasses. Serine proteinases (EC 3.4.21) have a serine and a histidine in the active site. Cysteine proteinases (EC 3.4.22) possess a cysteine in the active site and are inhibited by mercurial compounds. Aspartic proteinases (EC 3.4.23) have a pH optimum below 5, due to the involvement of a carboxyl residue in catalysis. Metalloproteinases (EC 2.3.24) need a metal ion in the catalytic process. Exopeptidases include enzymes that hydrolyze single amino acids from the N-terminus (aminopeptidases, EC 3.4.11) or from the C-terminus (carboxypeptidases, EC 3.4.16–18) of the peptide chain and those enzymes specific for dipeptides (dipeptide hydrolases, EC 3.4.13) (**Figure 2**).

Glycosidases (EC 3.2) are classified according to their substrate specificities. They include the enzymes that cleave internal bonds in polysaccharides and are usually named from their substrates, e.g., amylase, cellulase, pectinase, and chitinase. They also include the enzymes that hydrolyze oligosaccharides and disaccharides. Oligosaccharidases and disaccharidases are usually named based on the monosaccharide that gives its reducing group to the glycosidic bond and on the configuration (α or β) of this bond (**Figure 2**).

Lipids are a large and heterogeneous group of substances that are relatively insoluble in water but readily soluble in apolar solvents. Some contain fatty acids (fats, phospholipids, glycolipids, and waxes) and others lack them (terpenes, steroids, and carotenoids). Ester bonds are hydrolyzed in lipids containing fatty acids before they are absorbed. The enzymes that hydrolyze ester bonds comprise: (1) carboxylic ester hydrolases (EC 3.1.1), e.g., lipases, esterases, and phospholipases A and B; (2) phosphoric monoester hydrolases (EC 3.1.3), which are the phosphatases; and (3) phosphoric diester hydrolases (EC 3.1.4), including phospholipases C and D (**Figure 2**).

4.5.2.4. Phases of Digestion and Their Compartmentalization in the Insect Gut

Most food molecules to be digested are polymers such as proteins and starch and are digested sequentially in three phases. Primary digestion is the dispersion and reduction in molecular size of the polymers and results in oligomers. During intermediate digestion, these undergo a further reduction in molecular size to dimers; in final digestion, they become monomers. Digestion usually occurs under the action of digestive enzymes from the midgut, with little or no participation of salivary enzymes.

Any description of the spatial organization of digestion in an insect must relate the midgut compartments (cell, ecto-, and endoperitrophic spaces) to each phase of digestion and, hence, to the corresponding enzymes. To accomplish this, enzyme determinations must be performed in each midgut luminal compartment and in the corresponding tissue. Techniques of sampling enzymes from midgut luminal compartments and for identifying microvillar enzymes, and enzymes trapped in cell glycocalyx have been reviewed elsewhere (Terra and Ferreira, 1994).

Frequently, initial digestion occurs inside the peritrophic membrane (see Sections 4.5.7.1 and 4.5.7.2), intermediate digestion in the ectoperitrophic space, and final digestion at the surface of midgut cells by integral microvillar enzymes or by enzymes trapped into the glycocalyx (**Figure 1**). Exceptions to this rule, and the procedures for studying the organization of the digestive process, will be detailed below.

4.5.2.5. Role of Microorganisms in Digestion

Most insects harbor a substantial microbiota including bacteria, yeast, and protozoa. Microorganisms might be symbiotic or fortuitous contaminants from the external environment. They are found in the lumen, adhering to the peritrophic membrane, attached to the midgut surface, or within cells. Intracellular bacteria are usually found in special cells, the mycetocytes, which may be organized in groups, the

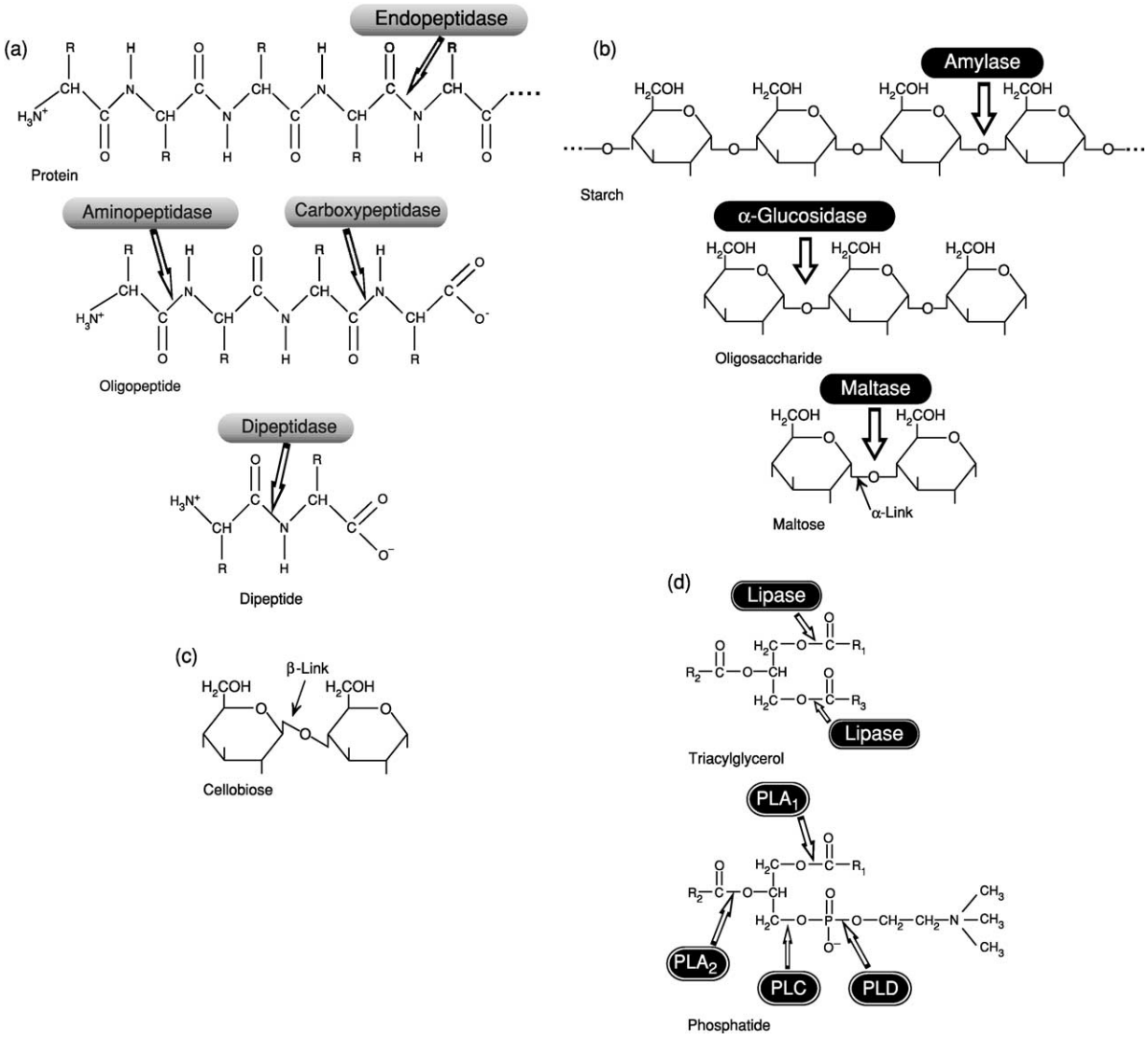


Figure 2 Digestion of important nutrient classes. Arrows point to bonds cleaved by enzymes. (a) Protein digestion; R, different amino acid moieties; (b) starch digestion; (c) β-linked glucoside; (d) lipid digestion; PL, phospholipase; R, fatty acyl moieties. (Reprinted with permission from Terra, W.R., **2003**. Digestion. In: Resh, V.H., Cardé, R.T. (Eds.), Encyclopedia of Insects. Academic Press, San Diego, CA, pp. 310–313; © Elsevier.)

mycetomes. Microorganisms produce and secrete their own hydrolases and cell death will result in the release of enzymes into the intestinal milieu. Any consideration of the spectrum of hydrolase activity in the midgut must include the possibility that some of the activity may derive from microorganisms. Despite the fact that digestive enzymes of some insects are thought to be derived from the microbiota, there are relatively few studies that show an unambiguous contribution of microbial hydrolases. Best examples are found among wood- and humus-feeding insects like termites, tipulid fly larvae, and scarabid beetle larvae. Although these insects may have their own cellulases (see Section 4.5.4.3.1), only fungi and certain filamentous bacteria developed a strategy for the chemical breakdown of lignin. Lignin is a phenolic polymer that forms an amorphous resin in which the polysaccharides of the secondary plant cell wall are embedded, thus becoming protected from enzymatic attack (Terra *et al.*, 1996; Brune, 1998; Dillon and Dillon, 2004).

Microorganisms play a limited role in digestion, but they may enable phytophagous insects to overcome biochemical barriers to herbivory (e.g., detoxifying flavonoids and alkaloids). They may also provide complex-B vitamins for blood-feeders and essential amino acids for phloem feeders, produce pheromone components, or withstand the colonization of the gut by nonindigenous species (including pathogens) (Dillon and Dillon, 2004).

4.5.3. Midgut Conditions Affecting Enzyme Activity

The pH of the contents of the midgut is one of the important internal environmental properties that affect digestive enzymes. Although midgut pH is hypothesized to result from adaptation of an ancestral insect to a particular diet, its descendants may diverge, feeding on different diets, while still retaining the ancestral midgut pH condition. Thus there is not necessarily a correlation between midgut pH and diet. In fact midgut pH correlates well with insect phylogeny (Terra and Ferreira, 1994; Clark, 1999).

The pH of insect midgut contents is usually in the 6–7.5 range. Major exceptions are the very alkaline midgut contents (pH 9–12) of Lepidoptera, scarab beetles and nematoceran Diptera larvae, the very acid (pH 3.1–3.4) middle region of the midgut of cyclorrhaphous Diptera, and the acid posterior region of the midgut of heteropteran Hemiptera (Terra and Ferreira, 1994; Clark, 1999). pH values may not be equally buffered along the midgut. Thus, midgut contents are acidic in the anterior midgut and nearly neutral or alkaline in the posterior midgut in Dictyoptera, Orthoptera, and most families of Coleoptera. Cyclorrhaphan Diptera midguts have nearly neutral contents in the anterior and posterior regions, whereas in middle midgut the contents are very acid (Terra and Ferreira, 1994).

The high alkanity of lepidopteran midgut contents is thought to allow these insects to feed on plant material rich in tannins, which bind to proteins at lower pH, reducing the efficiency of digestion (Berenbaum, 1980). This explanation may also hold for scarab beetles and for detritus-feeding nematoceran Diptera larvae that usually feed on refractory materials such as humus. Nevertheless, mechanisms other than high gut pH must account for the resistance to tannin displayed by some locusts (Bernays *et al.*, 1981) and beetles (Fox and Macauley, 1977). One possibility is the effect of surfactants, such as lysolecithin which is formed in insect fluids due to the action of phospholipase A on cell membranes (**Figure 2**), and which occurs widely in insect digestive fluids (De Veau and Schultz, 1992). Surfactants are known to prevent the precipitation of proteins by tannins even at pH as low as 6.5 (Martin and Martin, 1984). Present knowledge is not sufficient to relate midgut detergency to diet or phylogeny or to both.

Tannins may have deleterious effects other than precipitating proteins. Tannic acid is frequently oxidized in the midgut lumen, generating peroxides, including hydrogen peroxide, which readily diffuses across cell membranes and is a powerful cytotoxin. In some insects, e.g., *Orgyia leucostigma,* tannic acid oxidation and the generation of peroxides are suppressed by the presence of high concentrations of ascorbate and glutathione in the midgut lumen (Barbehenn *et al.*, 2003). Dihydroxy phenolics in an alkaline medium are converted to quinones that react with lysine ε-amino groups. This leads to protein aggregation and a decrease in lysine availability for the insect. Other compounds, e.g., oleuropein, alkylate lysine residues in proteins, causing the same problems as dihydroxy phenolics. These phenomena are inhibited in larvae of several lepidopteran species by secreting glycine into the midgut lumen. Glycine competes with lysine residues in the denaturating reaction (Konno *et al.*, 2001).

A high midgut pH may also be of importance, in addition to its role in preventing tannin binding to proteins, in freeing hemicelluloses from plant cell walls ingested by insects. Hemicelluloses are usually extracted in alkaline solutions for analytical purposes (Blake *et al.*, 1971) and insects, such as the caterpillar *Erinnyis ello*, are able to digest hemicelluloses efficiently without affecting the cellulose from the leaves they ingest to any degree (Terra, 1988). This explanation is better than the previous one in accounting for the very high pH observed in several insects, since a pH of about 8 is sufficient to prevent tannin binding to proteins (Terra, 1988).

The acid region in the cyclorrhaphous Diptera midgut is assumed to be involved in the process of killing and digesting bacteria, which may be an important food for maggots. This region is retained in Muscidae that have not diverged from the putative ancestral bacteria-feeding habit, as well as in the flesh-feeding Calliphoridae and in the fruit-feeding Tephritidae (Terra and Ferreira, 1994). The acid posterior midgut of Hemiptera may be related to their lysosome-like digestive enzymes (cysteine and aspartic proteinase) (see Sections 4.5.5.3 and 4.5.5.4).

Few papers have dealt with midgut pH buffering mechanisms. The early unsuccessful attempts to relate midgut buffering activity to the large amounts of phosphate frequently found in insect midguts, as well as other unsuccessful attempts to describe buffering mechanisms, are reviewed by House (1974). The results of more recent research on midgut buffering mechanisms are more encouraging. Dow (1992) showed that the lepidopteran larval midgut transports equal amounts of K^+ and alkali from blood to the midgut lumen. Based on this and other data he described a carbonate secretion system, which may be responsible for the high pH found in Lepidoptera midguts (**Figure 3**). Phosphorus NMR

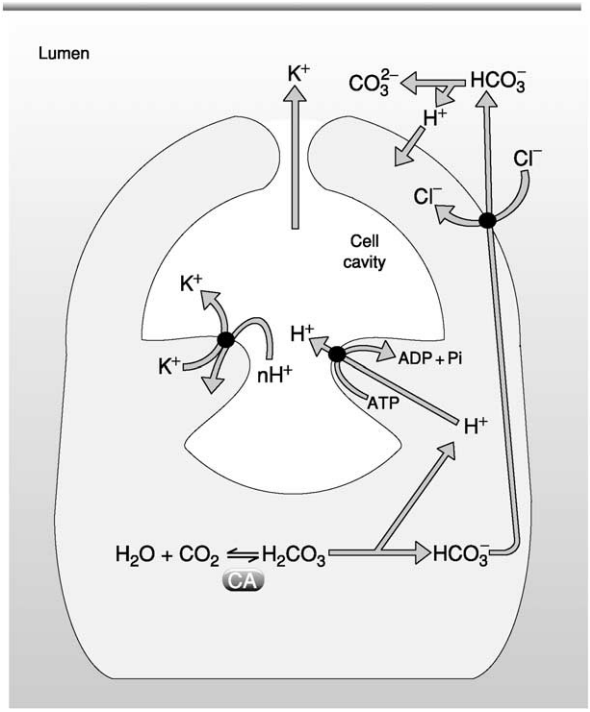


Figure 3 A model for generation of high gut pH by the goblet cells of lepidopteran larvae. Carbonic anhydrase (CA) produces carbonic acid that dissociates into bicarbonate and a proton. The proton is pumped by a V-ATPase into the goblet cell cavity, from where it is removed in exchange with K^+ that eventually diffuses into lumen. Bicarbonate is secreted in exchange with chloride and loses a proton due to the intense field near the membrane, forming carbonate and raising the gut pH. (Data from Dow, J.A.T., **1992**. pH gradients in lepidopteran midgut. *J. Exp. Biol. 172*, 355–375.)

microscopy has been used to show that valinomycin leads to a loss of alkalinization in the midgut of *Spodoptera litura* (Skibbe *et al.*, 1996). As valinomycin is known to transport K^+ down its concentration gradient, this result gives further support to the model described in **Figure 3**. It is not known whether midgut alkalinization in scarab beetles and nematoceran Diptera occurs by mechanisms similar to those of lepitopteran larvae. Terra and Regel (1995) determined pH values and concentrations of ammonia, chloride, and phosphate in the presence or absence of ouabain and vanadate in *Musca domestica* midguts. From the results they proposed that middle midgut acidification is accomplished by a proton pump of mammalian-like oxyntic cells, whereas the neutralization of posterior midgut contents depends on ammonia secretion (**Figure 4**).

Redox conditions in the midgut are regulated and may be the result of phylogeny, although data are scarce. Reducing conditions are observed in clothes moth, sphinx moths, owlet moths, and dermestid beetles (Appel and Martin, 1990) and in Hemiptera (Silva and Terra, 1994). Reducing conditions are important to open disulfide bonds in keratin ingested by some insects (clothes moths, dermestid beetles) (Appel and Martin, 1990), to maintain the activity of the major proteinase in Hemiptera (see Section 4.5.5.3), and to reduce the impact of some plant allelochemicals, such as phenol, in some herbivores (Appel and Martin, 1990). In spite of this, the artificial lowering of *in vivo* redox potentials did not

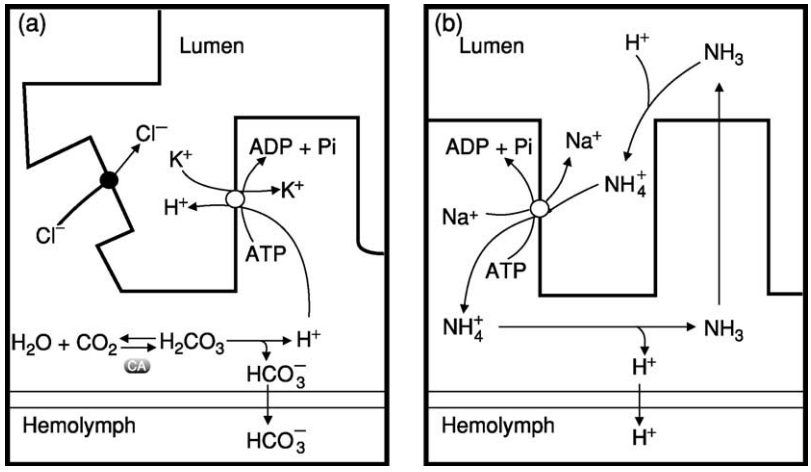


Figure 4 Diagrammatic representation of ion movements, proposed as being responsible for maintenance of pH in the larval midgut contents of *Musca domestica*. Carbonic anhydrase (CA) in cup-shaped oxyntic cells in the middle of the midgut (a) produces carbonic acid which dissociates into bicarbonate and a proton. Bicarbonate is transported into the hemolymph, whereas the proton is actively translocated into the midgut lumen acidifying its contents to pH 3.2. Chloride ions follow the movement of protons. NH_3 diffuses from anterior and posterior midgut cells (b) into the midgut lumen, becoming protonated and neutralizing their contents to pH 6.1–6.8. NH_4^+ is then exchanged for Na^+ by a microvillar Na^+/K^+-ATPase. Inside the cells, NH_4^+ forms NH_3, which diffuses into midgut lumen, and proton that is transferred into the hemolymph. (Reprinted with permission from Terra, W.R., Regel, R., **1995**. pH buffering in *Musca domestica* midguts. *Comp. Biochem. Physiol. A 112*, 559–564; © Elsevier.)

significantly impact digestive efficiency of the herbivore *Helicoverpa zea*, although the reducing agent used (dithiothreitol) inhibited some proteinases *in vitro* (Johnson and Felton, 2000).

Although several allelochemicals other than phenols may be present in the insect gut lumen, including alkaloids, terpene aldehydes, saponins, and hydroxamic acids (Appel, 1994), there are no sufficient data on their effect on digestion.

4.5.4. Digestion of Carbohydrates

4.5.4.1. Initial Considerations

Polysaccharides are major constituents of cell walls and energy reserves such as starch granules within plant cells and glycogen within animal cells. For phytophagous insects, disruption of plant cell walls is necessary in order to expose storage polymers in cell contents to polymer hydrolases. Cell wall breakdown may be achieved by mastication, but more frequently it is the result of the action of digestive enzymes. Thus, even insects unable to obtain nourishment from the cellulosic and noncellulosic cell wall biochemicals would profit from having enzymes active against these structural components. Cell walls are disrupted by β-glucanases, xylanases, and pectinases (plant cells), lysozyme (bacterial cells), or chitinase and β-1,3-glucanase (fungal cells). The carbohydrates associated with cellulose are frequently named "hemicelluloses" and the enzymes that attack them "hemicellulases." Thus, the hemicellulases include β-glucanases other than cellulases, xylanases, and pectinases. Following the loss of cell wall integrity, starch digestion is initiated by amylases. A complex of carbohydrases converts the oligomers resulting from the action of the polymer hydrolyzing enzymes into dimers (such as sucrose, cellobiose, and maltose that also occur free in some cells) and finally into monosaccharides like glucose and fructose.

4.5.4.2. Amylases

α-Amylases (EC 3.2.1.1) catalyze the endohydrolysis of long α-1,4-glucan chains such as starch and glycogen. Amylases are usually purified by glycogen–amylase complex formation followed by precipitation in cold ethanol or, alternatively, by affinity chromatography in a gel matrix linked to a protein amylase inhibitor. In sequence, isoamylases can be resolved by anion exchange chromatography (Terra and Ferreira, 1994).

Most insect amylases have molecular weights in the range 48–60 kDa, pI values of 3.5–4.0, and Km values with soluble starch around 0.1%. pH optima generally correspond to the pH prevailing in midguts from which the amylases were isolated. Insect amylases are calcium-dependent enzymes and are activated by chloride with displacement of the pH optimum. Activation also occurs with anions other than chloride, such as bromide and nitrate, and it seems to depend upon the ionic size (Terra and Ferreira, 1994).

The best-known insect α-amylase, and the only one whose 3D structure has been resolved, is the midgut α-amylase of *Tenebrio molitor* larvae. The enzyme has three domains. The central domain (domain A) is an $(\beta/\alpha)_8$-barrel that comprises the core of the molecule and includes the catalytic amino acid residues (Asp 185, Glu 222, and Asp 287) (*T. molitor* α-amylase numbering throughout). Domains B and C are almost opposite to each other, on each side of domain A. The substrate-binding site is located in a long V-shaped cleft between domains A and B. There, six saccharide units can be accommodated, with the sugar chain being cleaved between the third and fourth bound glucose residues. A calcium ion is placed in domain B and is coordinated by Asn 98, Arg 146, and Asp 155. This ion is important for the structural integrity of the enzyme and seems also to be relevant due its contact with His 189. This residue interacts with the fourth sugar of the substrate bound in the active site, forming a hinge between the catalytic site and the Ca^{2+}-binding site. A chloride ion is coordinated by Arg 183, Asn 285, and Arg 321 in domain A (Strobl *et al.*, 1998a). Domain C is placed in the C-terminal part of the enzyme, contains the so-called "Greek key" motif and has no clear function (**Figure 5**). These structural features are shared by all known α-amylases (Nielsen and Borchert, 2000), although not all α-amylases have a chloride-binding site (Strobl *et al.*, 1998a). The most striking difference between mammalian and insect α-amylases is the presence of additional loops in the vicinity of the active site of the mammalian enzymes (Strobl *et al.*, 1998a).

Asp 287 is conserved in all α-amylases. Comparative studies have shown that Glu 222 is the proton donor and Asp 185 the nucleophile, and that Asp 287 is important but not a direct participant in catalysis. It is proposed that its role is to elevate the pK_a of the proton donor (Nielsen and Borchert, 2000).

Chloride ion is an allosteric activator of α-amylases, leading to a conformation change in the enzyme that changes the environment of the proton donor. This change causes an increase in the pK_a of the proton donor, thus displacing the pH optimum of the enzyme and increasing its V_{max} (Levitzki and Steer, 1974). According to Strobl *et al.* (1998a), the increase in V_{max} is a consequence of chloride ion being close to the water molecule that has been

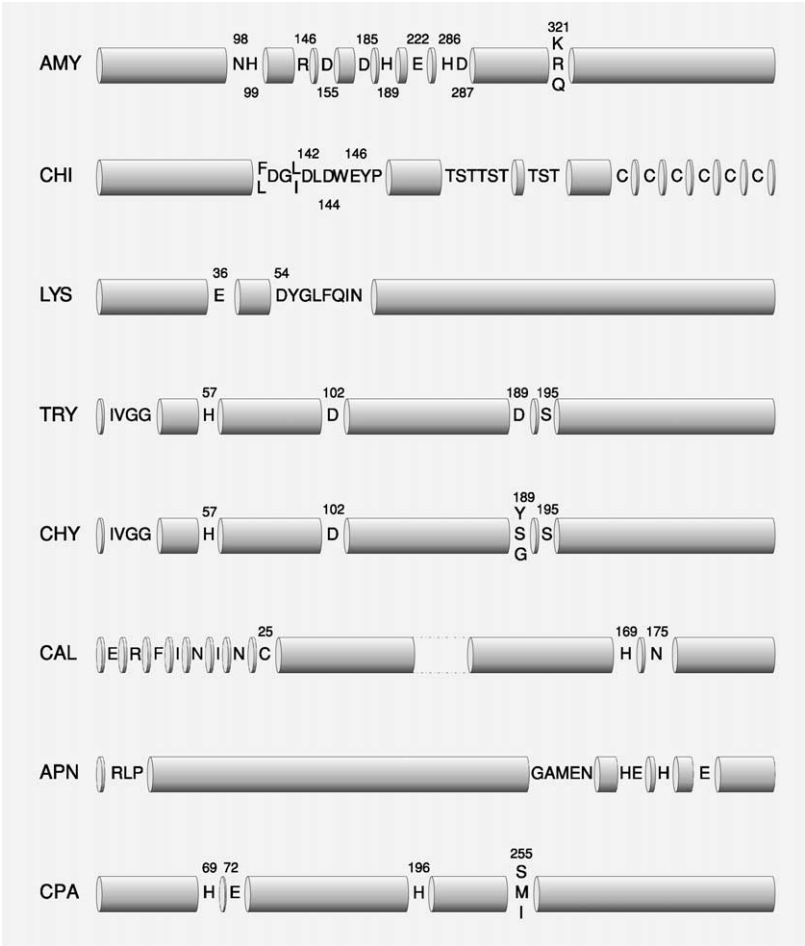


Figure 5 Conserved residues in the primary structures of major insect digestive enzymes. AMY (amylase) follows *Tenebrio molitor* amylase numbering; CHI (chitinase), molting-fluid *Manduca sexta* chitinase numbering. TRY (trypsin) and CHY (chymotrypsin) follow the bovine chymotrypsin numbering; CAL (cathepsin L), papain numbering: APN (aminopeptidase N) does not have a consensual numbering; CPA (carboxypeptidase A), mammalian CPA numbering.

suggested to initiate the cleavage of the substrate chain. The nucleophilicity of this water molecule might be enhanced by the negative charge of the ion.

There are 21 complete sequences of insect α-amylases registered in the GenBank (as of May 2003) excluding all of the sequences of *Drosophila*. The sequences correspond to 18 species found in four orders. Examples may be found among Hymenoptera (Ohashi *et al.*, 1999; Da Lage *et al.*, 2002), Coleoptera (Strobl *et al.*, 1997; Titarenko and Chrispeels, 2000), Diptera (Grossmann *et al.*, 1997; Charlab *et al.*, 1999), and Lepidoptera (Da Lage *et al.*, 2002). All the sequences have the catalytic triad (Asp 185, Glu 222, and Asp 287), the substrate-binding histidine residues (His 99, His 189, and His 286), and the Ca^{2+}-coordinating residues (Asn 98, Arg 146, and Asp 155) (**Figure 5**). From the residues found to be involved in chloride binding, Arg 183 and Asn 285 are conserved, whereas position 321 varies. According to D'Amico *et al.* (2000), all known chloride-activated α-amylases have an arginine or lysine residue at position 321. Insect α-amylase sequences have arginine at position 321, except those of *Zabrotes subfasciatus* and *Anthonomus grandis*, which have lysine, and the lepidopteran α-amylases which have glutamine. This agrees with the observation that most insect α-amylases are activated by chloride with the remarkable exception of lepidopteran amylases (Terra and Ferreira, 1994). The few coleopteran and hymenopteran α-amylases reported to be not affected by chloride (Terra and Ferreira, 1994) deserve reinvestigation. It is possible that another anion is replacing chloride as an activator, as shown for some hemipteran amylases (Hori, 1972).

Action pattern refers to the number of bonds hydrolyzed during the lifetime of a particular enzyme–substrate complex. If more than one bond

is hydrolyzed after the first hydrolytic step, the action pattern is said to be processive. The degree of multiple attack is the average number of hydrolyzed bonds after the first bond is broken. *Rhynchosciara americana* amylase has a degree of multiple attack between that of the amylase of *Bacillus subtilis* (1.7) and porcine pancreas (6). Amylases from larvae and adults of *Sitophilus zeamais*, *S. granarius*, and *S. oryzae* and larvae of *Bombyx mori* have action patterns similar to that of porcine pancreas amylase (Terra and Ferreira, 1994). These studies need to be taken further, including the determination of the affinities corresponding to each subsite in the active center. Such studies, specially if combined with crystallographic data, may describe in molecular detail the reasons why amylases act differently toward starches of distinct origins.

There is a variety of natural compounds that affect amylases, including many plant protein inhibitors (Franco *et al*., 2002). Crystallographic data have shown that these protein inhibitors always occupy the amylase active site (Strobl *et al*., 1998b; Nahoum *et al*., 1999). In the case of the *Amaranth* α-amylase inhibitor, a comparison of *T. molitor* amylase–inhibitor complex with a modeled complex between porcine pancreatic α-amylase and the inhibitor identified six hydrogen bonds that can be formed only in the *T. molitor* amylase–inhibitor complex (Pereira *et al*., 1999). This was the first successful explanation of how a protein inhibitor specifically inhibits α-amylases from insects, but not from mammalian sources. As will be discussed with details for trypsins (see Section 4.5.5.2.1), specific amylases are induced when insect larvae are fed with α-amylase inhibitor-containing diets (Silva *et al*., 2001). The mechanisms underlying this induction are unknown and are presently under study (C.P. Silva, personal communication).

4.5.4.3. β-Glucanases

β-Glucanases are enzymes acting on β-glucans. These are major polysaccharide components of plant cell walls and include β-1,4-glucans (cellulose), β-1,3-glucans (callose), and β-1,3;1,4-glucans (cereal β-glucans) (Bacic *et al*., 1988). The cell walls of certain groups of fungi have β-1,3;1,6-glucans (Bacic *et al*., 1988).

4.5.4.3.1. Cellulases Cellulose is by far the most important β-glucan. It occurs in the form of β-1,4-glucan chains packed in an ordered manner to form compact aggregates which are stabilized by hydrogen bonds. The resulting structure is crystalline and not soluble. According to work done with microbial systems, cellulose is digested by a combined action of two enzymes. An *endo*-β-1,4-glucanase (EC 3.2.1.4), with an open substrate-binding cleft, cleaves bonds located within chains in the amorphous regions of cellulose, creating new chain ends. An *exo*-β-1,4-glucanase (EC 3.2.1.91) processively releases cellobiose from the ends of cellulose chains in a tunnel-like active site. Surface loops in cellobiohydrolase prevent the dislodged cellulose chains from readhering to the crystal surface, as the enzyme progresses into crystalline cellulose (Rouvinen *et al*., 1990; Kleywegt *et al*., 1997). Cellobiohydrolase structure is modular, comprising a catalytic domain linked to a distinct cellulose-binding domain, which enhances the activity of the enzyme towards insoluble cellulose (Linder and Teeri, 1997).

Cellulose digestion in insects is rare, presumably because the dietary factor that usually limits growth in plant feeders is protein quality. Thus, cellulose digestion is unlikely to be advantageous to an insect that can meet its dietary requirements using more easily digested constituents. Nevertheless, cellulose digestion occurs in several insects that have, as a rule, nutritionally poor diets (Terra and Ferreira, 1994). The role of symbiotic organisms in insect cellulose digestion is less important than initially believed (Slaytor, 1992), although symbiotic nitrogen-fixing organisms are certainly involved in increasing the nutritive value of diets of many insects (Terra, 1990).

Few insect cellulases have been purified and characterized. Two *endo*-β-1,4-glucanases (41 and 42 kDa) were isolated from the lower termite *Reticulitermes speratus* (Watanabe *et al*., 1997). The cDNA that encodes this protein was cloned (Watanabe *et al*., 1998) and the protein was shown to be secreted from the salivary glands (Tokuda *et al*., 1999). The cDNAs that encode the *endo*-β-1,4-glucanase secreted by the midgut of the higher termite *Nasutitermes takasagoensis* (Tokuda *et al*., 1999) and the two *endo*-β-1,4-glucanases from the woodroach *Panestria cribrata* of 47 kDa (Tokuda *et al*., 1997) and 54 and 49 kDa (Scrivener and Slaytor, 1994), respectively, were also cloned (Tokuda *et al*., 1999; Lo *et al*., 2000).

Alignments of the sequences of termite and woodroach endoglucanases from data banks showed that they belong to family 9 of glycoside hydrolases (Coutinho and Henrissat, 1999). The paradigm of this family is the endo/exocellulase from the bacteria *Thermomonospora fusca*, whose catalytic center binds a cellopentaose residue and cleaves it into cellotetraose plus glucose or cellotriose plus cellobiose, and has Asp 55 as a nucleophile and Glu 424 as a proton donor (Sakon *et al*., 1997).

The active site groups are conserved in the termite and woodroach endoglucanases although these enzymes lack the cellulose-binding domains that improve the binding and facilitate the activity of the catalytic domain on crystalline cellulose (Linder and Teeri, 1997). The conclusions drawn from sequence alignments were confirmed by the 3D structure resolution of the *N. takasagoensis* endoglucanase, which also revealed a Ca^{2+}-binding site near its substrate binding cleft (Khadeni *et al.*, 2002). According to Slaytor (1992), the large production of endoglucanases in termites and woodroaches would compensate their low efficiency on crystalline cellulose.

The phytophagous beetle *Phaedon cochleariae* has cellulase activity in its midgut (Girard and Jouanin, 1999a). A cDNA that encodes one cellulase was cloned and shown to belong to the glycoside hydrolase family 45 and to consist only of a catalytic domain. The purification of an *exo*-β-1,4-glucanase (52 kDa) from the termite *Macrotermes mulleri* (Rouland *et al.*, 1988) and the partial resolution from *Ergates faber* (Coleoptera) midgut extracts of another *exo*-β-1,4-glucanase (Chararas *et al.*, 1983) suggest that these insects, in contrast to the others, have cellulases with cellulose-binding domains.

4.5.4.3.2. Laminarinases and licheninases Licheninases (EC 3.2.1.73) digest only β-1,3;1,4-glucans whereas laminarinases may hydrolyze β-1,3;1,4-glucans and also β-1,3-glucans (EC 3.2.1.6) or only the last polymer (EC 3.2.1.39).

In spite of laminarinase activities being widespread among insects (Terra and Ferreira, 1994), only the laminarinases of *Periplaneta americana* (LAM, LIC 1, and LIC 2) have been purified and studied in detail (Genta *et al.*, 2003). The enzymes are secreted by salivary glands, stabilized by calcium ions, and have pH optima around 6. LAM (46.2 kDa) is an *endo*-β-1,3-glucanase that processively releases mainly glucose from laminarin and shows lytic activity on fungal cells. LIC 1 (24.6 kDa) is an *endo*-β-1,3-(4)-glucanase highly active on sequences of β-1,3-linked cellotetraoses, releasing cellotetraose from lichenin and is also able to lyse fungal cells. The specificities of *P. americana* β-glucanases agree with the omnivorous detritus-feeding habit of this insect, as they are able to help the cellulases in attacking plant cell walls (mainly those from cereals) and also in opening the cell walls of fungi, usually present in detritus.

No insect licheninase has been characterized in detail to date.

4.5.4.4. Xylanases and Pectinases

Xylans constitute the major noncellulosic polysaccharides (hemicelluloses) of primary walls of grasses and secondary walls of all angiosperms, accounting for one-third of all renewable organic carbon available on earth (Bacic *et al.*, 1988). Chemically, xylans are a family of linear β-1,4-xylans with a few branches. *Endo*-β-1,4-xylanase activities (EC 3.2.1.8) were found in several insects (Terra and Ferreira, 1994). One of these enzymes was cloned from a beetle and shown to correspond to a protein of 22 kDa, with high sequence identity to fungal xylanases, and conserving the usual two catalytic regions (Girard and Jouanin, 1999a). An *exo*-β-1,4-xylanase (EC 3.2.1.37) was partially purified from termites (Matoub and Rouland, 1995) and thought to act synergistically with an *endo*-β-1,4-xylanase originating from fungus ingested by the termites. Much more work is needed on this class of enzymes that may be important mainly for detritivorous insects.

Pectin is a linear chain of a D-galacturonic acid units with α-1,4-linkages in which varying proportions of the acid groups are present as methyl esters. It is the main component of the rhamnogalacturonan backbone of the structure formed by the pectin polysaccharides. Pectin is hydrolyzed by pectinases (polygalacturonases, EC 3.2.1.15) described in many insects (Terra and Ferreira, 1994).

Pectinases are thought to be important for hemipterans, as they would facilitate penetration of their stylets through plant tissues into sap-conducting structures and for insects boring plant parts. Accordingly, pectinases have been found in hemipteran saliva (Vonk and Western, 1984) and have been isolated and characterized from two weevils (Shen *et al.*, 1996; Dootsdar *et al.*, 1997) and cloned from a phytophagous beetle (Girard and Jouanin, 1999a).

The pectinases from the weevils *S. oryzae* (Shen *et al.*, 1996) and *Diaprepes abbreviatus* (Dootsdar *et al.*, 1997) were purified to electrophoretical homogeneity from whole-body extracts and gut homogenates, respectively. Purification of the pectinases was achieved by affinity chromatography through cross-linked pectate in addition to two ion exchange chromatographic steps. The enzymes have molecular masses of 35–45 kDa, pH optimum 5.5 and Km values of 1–4 mg ml^{-1} for pectic acid. The *D. abbreviatus* pectinase is inhibited by a polygalacturonase-inhibitor protein that may be associated with plant resistance to insects (Dootsdar *et al.*, 1997). Although the weevil pectinases may originate from endosymbiotic bacteria (Campbell *et al.*, 1992), the finding that the cDNA-coding pectinase

of the beetle *P. cochleariae* has a poly(A) tail (Girard and Jouanin, 1999a) argues against this hypothesis. The beetle pectinase belongs to family 28 of the glycoside hydrolases and is most related to fungal endopolygalacturonases, conserving the active site signature centered on the His 223 catalytic residue (Girard and Jouanin, 1999a; Markovic and Janecek, 2001).

4.5.4.5. Chitinases and Lysozymes

Chitin, the simplest of the glycosaminoglycans, is a β-1,4-homopolymer of *N*-acetylglucosamine (see **Chapter 4.3**). Chitinolytic enzymes (Kramer and Koga, 1986) include: chitinase (EC 3.2.1.14), which catalyzes the random hydrolysis of internal bonds in chitin forming smaller oligosaccharides, and β-*N*-acetyl-D-glucosaminidase (EC 3.2.1.52), which liberates *N*-acetylglucosamine from the nonreducing end of oligosaccharides (Kramer and Koga, 1986). Lysozyme, as described below, also has some chitinase activity, whereas chitinase has no lysozyme activity. Chitinolytic enzymes associated with the ecdysial cycle have been studied and demonstrated to act synergistically in cuticular chitin degradation (Kramer and Koga, 1986). Nevertheless, these enzymes may also have a digestive role. Chitinase assays with midguts of several insects showed that there is a correlation between the presence of chitinase and a diet rich in chitin (Terra and Ferreira, 1994).

The best-known insect chitinase is the molting fluid chitinase from the lepidopteran *Manduca sexta* (**Figure 5**). The enzyme has a multidomain architecture that includes a signal peptide, an N-terminal catalytic domain, with the consensus sequence (F/L)DG(L/I)DLDWEYP, and a C-terminal cysteine-rich chitin-binding domain, which are connected by the interdomain serine/threonine-rich *O*-glycosylated linker. The residues Asp 142, Asp 144, Trp 145, and Glu 146 of the consensus sequence have been shown by site-directed mutagenesis to be involved in catalysis. Glu 146 functions as a proton donor in the hydrolysis like homologous residues in other glycoside hydrolases. Asp 144 apparently functions as an electrostatic stabilizer of the positively charged transition state, whereas Asp 142, and perhaps also Trp 145, influences the pK_a values of Asp 144 and Glu 146. The chitin-binding domains have 6 cysteines (with the consensus sequence CXnCXnCXnCXnCXnC, where Xn stands for a variable number of residues) and include several highly conserved aromatic residues (Tellam *et al.*, 1999). The three disulfide bonds in the domain may constrain the polypeptide to present the aromatic amino acids on the protein surface for interactions with the ring structures of sugars. Thus, the chitin-binding domains enhance activity toward the insoluble polymer, whereas the linker region facilitates secretion from the cell and helps to stabilize the enzyme in the presence of proteolytic enzymes (Kramer *et al.*, 1993; Lu *et al.*, 2002; Zhang *et al.*, 2002; Arakane *et al.*, 2003).

Chitinase in molting fluid of the beetle *T. molitor* is a large (about 320 kDa) multidomain protein with five catalytic domains, five serine/threonine-rich linker domains, four chitin-binding domains, and two mucin-like domains. There is evidence that the enzyme is secreted as a zymogen activated by trypsin (Royer *et al.*, 2002).

The *Anopheles gambiae* gut chitinase is secreted upon blood-feeding as an inactive proenzyme that is later activated by trypsin. Sequencing a cDNA coding the gut chitinase showed that the enzyme comprises a putative catalytic domain at the N-terminus, a putative chitin-binding domain at the C-terminus, and a serine/threonine/proline-rich amino acid stretch between them (Shen and Jacobs-Lorena, 1997). The mosquito chitinase seems to modulate the thickness and permeability of the chitin-containing peritrophic membrane (see Section 4.5.7.1). Supporting this conjecture the authors found that the peritrophic membrane is stronger and persisted longer when the mosquitoes were fed diets containing chitinase inhibitor.

The beetle *P. cochleariae* has one group of chitinases of 40–70 kDa active at pH 5.0 and detected in guts and another group of 40–70 kDa that are more active at pH 7.0 and that are associated with molting. A cDNA encoding a gut chitinase showed this enzyme has an active site centered on the catalytic residues Asp 146 and Glu 150 (*M. sexta* chitinase numbering), but lacks the C-terminal chitin-binding domain and the serine/threonine-rich interdomain (Girard and Jouanin, 1999b). The *T. molitor* midgut chitinase (GenBank accession no. AY325895) is similar to that of *P. cochleariae*. The putative role of *P. cochleariae* chitinase is in turnover of the peritrophic membrane (Girard and Jouanin, 1999b), as proposed above for *A. gambiae*. It is not clear, however, why the chitinase from *A. gambiae* in contrast to that of *P. cochleariae* has a chitin-binding domain if their putative role is to affect the same type (type I) (see Section 4.5.7.1) of peritrophic membrane. The *T. molitor* midgut chitinase may have the same proposed function, although it may also digest the cell walls of fungi usually present in its food. More research is necessary to clarify this point.

Lysozyme (EC 3.2.1.17) catalyzes the hydrolysis of the 1,4-β-glycosidic linkage between *N*-acetylmuramic acid and *N*-acetylglucosamine of the

peptidoglycan present in the cell wall of many bacteria, causing cell lysis. Lysozyme is part of an immune defense mechanism against bacteria and has been described in most animals, including insects (Hultmark, 1996). Lysozyme has also been implicated in the midgut digestion of bacteria by organisms which ingest large amounts of them, such as marine bivalves (McHenery *et al*., 1979), or that harbor a bacterial culture in their guts, as exemplified by ruminants (Stewart *et al*., 1987).

Among insects, the capacity of digesting bacteria in the midgut seems to be an ancestral trait of Diptera Cyclorrhapha (Lemos and Terra, 1991a; Regel *et al*., 1998), which agrees with the fact that most Diptera Cyclorrhapha larvae are saprophagous, feeding largely on bacteria (Terra, 1990). These insects have midgut lysozymes (Lemos *et al*., 1993; Regel *et al*., 1998) similar to those of vertebrate fermenters. Thus, these enzymes have low pI values, are more active at pH values 3–4, when present in media with physiological ionic strengths, and are resistant to the cathepsin D-like aspartic proteinase present in midguts (Lemos *et al*., 1993; Regel *et al*., 1998).

Sequence analyses (Kylsten *et al*., 1992; Daffre *et al*., 1994; Ito *et al*., 1995) showed that cyclorrhaphan (*Drosophila melanogaster* and *Musca domestica)* digestive lysozymes have the same conserved residues as vertebrate lysozymes (Imoto *et al*., 1972) (numbering according to Regel *et al*., 1998): positions 55–61, Glu 36, and Asp 54. Glu 36 is believed to act as a general acid in catalysis, whereas Asp 54 is postulated to stabilize the resulting metastable oxocarbonium intermediate (Imoto *et al*., 1972). More recently, Asp 54 has been implicated more strongly in catalysis of the hydrolysis of chitin-derived substrates (Matsumura and Kirsch, 1996). The ability of *D. melanogaster* and *M. domestica* purified lysozymes in hydrolyzing chitosan favors this view. The most remarkable sequence convergence of cyclorrhaphan digestive lysozymes with that of vertebrate foregut fermenters are Ser 104 and a decrease in the number of basic amino acids, suggesting that these features are necessary for a digestive role in an acid environment (Regel *et al*., 1998). These hypotheses must be tested by site-directed mutagenesis and by the resolution of the 3D structure of at least one of these insect digestive lysozymes.

Lysozyme is also found in the salivary glands of *R. speratus*. This insect is a termite that feeds mainly on dead wood, which tends to lack nitrogen. Fujita *et al*. (2001) suggested, on the basis of the distribution and activity of lysozyme in this termite, that wood-feeding termites use lysozyme secreted from the salivary gland to digest their hindgut bacteria, which are transferred by proctodeal trophallaxis. The termite lysozyme is active in neutral pH and has no serine in position 104 (Fujita *et al*., 2002), thus differing from the digestive cyclorrhaphan lysozymes.

Chitin and the bacterial peptidoglycan resemble one another chemically and structurally. Because of this one might anticipate that chitinases and lysozymes would be structurally related enzymes. Indeed, some lysozymes, exemplified by cyclorrhaphan digestive lysozymes (Regel *et al*., 1998) are very good chitinases. Structurally well-known chitinases and lysozymes share no significant amino acid sequence similarities, but have a structurally invariant core consisting of two helices and a three-stranded β-sheet, which form the substrate-binding catalytic cleft. These enzymes are considered to represent a superfamily of hydrolases which are likely to have arisen by divergent evolution (Monzingo *et al*., 1996). However, this picture does not take into account insect enzymes, for which 3D structural data are lacking.

4.5.4.6. α-Glucosidases

α-Glucosidases (EC 3.2.1.20) catalyze the hydrolysis of terminal, nonreducing α-1,4-linked glucose residues from aryl (or alkyl)-glucosides, disaccharides, or oligosaccharides. α-glucosidases are frequently named maltases, although some of them may have weak activity on maltose. Insect α-glucosidases occur as soluble forms in the midgut lumen or are trapped in the midgut cell glycocalyx. They are also bound to microvillar membranes (Terra and Ferreira, 1994), to perimicrovillar membranes (lipoprotein membranes ensheathing the midgut cell microvillar membranes in most hemipterans) (Silva and Terra, 1995), or to the modified perimicrovillar membranes of aphid midgut cells (Cristofoletti *et al*., 2003). The last two membrane-bound α-glucosidases, as well as the soluble enzyme from bee midguts (Nishimoto *et al*., 2001), were purified to electrophoretic homogeneity. *Culex pipiens* microvillar α-glucosidase is the primary target of the binary toxin of *Bacillus sphericus* and, although not purified, cDNA sequencing data suggest it is bound by a glycosyl phosphatidyl inositol anchor (Darboux *et al*., 2001). α-Glucosidase is a major protein in dipteran midgut microvillae (Terra and Ferreira, 1994) and probably because of that it is the receptor of endotoxins, similar to that observed with aminopeptidase N in lepidopteran midgut cells (see Section 4.5.5.5).

Although biochemical properties of many crude, partially or completely purified gut α-glucosidases

are known, including molecular masses (range 60–80 kDa or a multiple of these values), pH optima (range 5–6.5, irrespective of the corresponding midgut pH value), pI values (range 5.0–7.2), and inhibition by tris(hydroxylmethyl)aminomethane (Tris), only few studies report on α-glucosidases specificities. These studies showed that insect α-glucosidases hydrolyze oligosaccharides up to at least maltopentaose (Terra and Ferreira, 1994), although there are exceptions. The perimicrovillar α-glucosidase from *Dysdercus peruvianus* prefers oligosaccharides up to maltotetraose (Silva and Terra, 1995) and the midgut bee α-glucosidase, up to maltotriose (Nishimoto *et al.*, 2001). The purified midgut α-glucosidase of the pea aphid, *Acyrthosiphon pisum* catalyzes transglycosylation reactions in the presence of excess sucrose, thus freeing glucose from sucrose without increasing the osmolarity of the medium (Cristofoletti *et al.*, 2003). This phenomenon associated with a quick fructose absorption (Ashford *et al.*, 2000) explains why the midgut luminal osmolarity decreases as the ingested sucrose-containing phloem sap passes along the aphid midgut.

Three digestive α-glucosidases from dipterans have been sequenced: one salivary (James *et al.*, 1989) and two midgut (Zheng *et al.*, 1995; Darboux *et al.*, 2001) enzymes. All the sequences have the invariant residues: Asp 123, His 128, Asp 206, Arg 221, Glu 271, His 296, and Asp 297 (numbers are relative to the positions in the sequence of *C. pipiens* α-glucosidases) (Darboux *et al.*, 2001) that are involved in the active site of α-amylase family of enzymes, and the three residues Gly 69, Pro 77, and Gly 323, that have a structural role for some α-glucosidases (Janecek, 1997).

4.5.4.7. β-Glucosidases and β-Galactosidases

β-Glycosidases (EC 3.2.1) catalyze the hydrolysis of terminal, nonreducing β-linked monosaccharide residues from the corresponding glycoside. Depending on the monosaccharide that is removed, the β-glycosidase is named β-glucosidase (glucose), β-galactosidase (galactose), β-xylosidase (xylose), and so on. Frequently, the same β-glycosidase is able to hydrolyze several different monosaccharide residues from glycosides. In this case, β-glucosidase (EC 3.2.1.21) is used to name all enzymes that remove glucose efficiently. The active site of these enzymes is formed by subsites numbered from the glycosidic linkage to be broken with negative (towards the nonreducing end of the substrate) or positive (towards the reducing end of the substrate) integers (Davies *et al.*, 1997). The nonreducing monosaccharide residue binds at the glycone (−1) subsite, whereas the rest of the molecule accommodates at the aglycone subsite, which actually may correspond to several monosaccharide residue-binding subsites.

Some insects have three or four digestive β-glycosidases with different substrate specificity. In others only two of these enzymes are found, that are able to hydrolyze as many different β-glycosides as the other three or four enzymes together (Ferreira *et al.*, 1998; Azevedo *et al.*, 2003).

Insect β-glycosidases best characterized have molecular masses of 30–150 kDa, pH optima of 4.5–6.5, and pI values of 3.7–6.8, whereas Km values with cellobiose or *p*-nitrophenyl β-glucoside (NpβGlu) as substrates, are found in the range of 0.2–2 mM. Although hydrolyzing several similar substances, insect digestive β-glycosidases have different specificities, preferring β-glucosides or β-galactosides as substrates, with hydrophobic or hydrophilic moieties in the aglycone part of the substrate (Terra and Ferreira, 1994; Azevedo *et al.*, 2003).

A few insect digestive β-glycosidases are more active on galactosides than on glucosides. Such enzymes are found in *Locusta migratoria* adults (Morgan, 1975), *Abracris flavolineata* adults (Marana, Terra, and Ferreira, unpublished data). In *Rhynchosciara americana* (Ferreira and Terra, 1983) and *T. molitor* (Ferreira *et al.*, 2003) larvae, there is a β-glycosidase that hydrolyzes galactosides but not glucosides.

Based on relative catalytical efficiency on several substrates, insect β-glycosidases can be divided into two classes. Class A includes the enzymes that efficiently hydrolyze substrates with hydrophilic aglycones, such as disaccharides and oligosaccharides. Class B comprises enzymes that have high activity only on substrates with hydrophobic aglycones, such as alkyl-, *p*-nitrophenyl-, and methylumbelliferyl-glycosides. Based on the same properties, β-glycosidases were previously divided into three classes (Terra and Ferreira, 1994), but the characterization of more enzymes showed that the previously proposed class 2 of β-glycosidases does not exist. Enzymes from the former class 2 comprised β-glycosidases supposed to have high activity only on di- and oligosaccharides. Former classes 2 plus 1 are now grouped in class A. Former class 3 are now named class B.

Enzymes from class A are more abundant than β-glycosidases from class B. Class A β-glycosidases hydrolyze di- and oligosaccharides and have four subsites for glucose binding in the active site: one in the glycone (−1) and three in the aglycone (+1, +2, +3) position (Ferreira *et al.*, 2001, 2003;

Marana *et al*., 2001; Azevedo *et al*., 2003). Some enzymes seem to be adapted to use disaccharides besides oligosaccharides as substrates. Optimal hydrolysis of disaccharides relies on high affinities to glucose moieties in −1 and primarily in the +1 subsite (Ferreira *et al*., 2003). The enzymes highly active against oligosaccharides have subsites −1, +1, and +2 with similar affinities to glucose moieties (Ferreira *et al*., 2001, 2003).

The affinities of some β-glycosidases for alkyl-glucosides were determined and binding energies for each methylene group to the active site were calculated. Values obtained for class A enzymes from *Spodoptera frugiperda* βGly50 (Marana *et al*., 2001) and *T. molitor* βGly1 (Ferreira *et al*., 2001) are $1.3\,kJ\,mol^{-1}$ and $-0.97\,kJ\,mol^{-1}$, respectively. Surprisingly, the binding energy in the case of *A. flavolineata* alkyl-glycosidase, which is an enzyme from class B, is only $0.47\,kJ\,mol^{-1}$ (calculated from Marana *et al*., 1995). Since the enzymes from *T. molitor* and *S. frugiperda* can hydrolyze di- and oligosaccharides and the *A. flavolineata* alkyl β-glycosidase can only use synthetic and more hydrophobic compounds as substrates, it seems that the hydrophobicity of the aglycone region is not directly related to the type of substrate that is hydrolyzed by the enzyme. Probably the extension of the aglycone site plays a major role.

Class A β-glycosidases are able to hydrolyze β-1,3, β-1,4, and β-1,6 glycoside bonds from di- and oligosaccharides. These enzymes are likely to be involved in the intermediate and terminal digestion of cellulose, hemicellulose, and glycoproteins present in food.

Class B β-glycosidases such as *D. saccharalis* βGly2 (Azevedo *et al*., 2003) and *T. molitor* βGly4 seem to have two active sites (Ferreira *et al*., 2003). These were detected by using alternative substrates as inhibitors. Two active sites are also found in βGly47 from *S. frugiperda* (Marana *et al*., 2000) which is a class A enzyme, but one of its active site has properties of class B β-glycosidase. The three enzymes with two putative active sites are the main enzymes responsible for β-galactoside hydrolysis in the insect gut. The enzymes from the lepidopteran *S. frugiperda* and *D. saccharalis* hydrolyze mainly galactosides in one active site and glucosides in the other. The presence of two active sites in mammalian lactase-phlorizin hydrolase (the digestive mammalian β-glycosidase) has been known for a long time. The difference between the enzymes is that insect β-glycosidases with two active sites have the size of only one of the two mammalian β-glycosidase domains. Unfortunately, none of these types of enzymes has yet been cloned in insects. It is interesting to note that *T. molitor* βGly4 is activated by the detergent Triton X-100.

Class B β-glycosidases (or active site) with high activity against hydrophobic substrates may have the physiological role of hydrolyzing glycolipids, mainly galactolipidis that are found in high amount in vegetal tissues. The main galactolipids in plants are 2,3-diacyl β-galactoside D-glycerol (mono galactosyl diglyceride) and 2,3-diacyl 1-(α-galactosyl 1,6 β-galactosyl)-D-glycerol (digalactosyl diglyceride) (Harwood, 1980). These enzymes may act directly against the monogalactosyl diglyceride or on digalactosyl diglyceride after the removal of one of the galactose residues by α-galactosidase. The activation by amphipatic substances (as Triton X-100; see above) may be a mechanism to maintain high enzyme activity only in the neighborhood of plant cell membranes undergoing digestion in the insect midgut. Those membranes are the source of glycolipid substrates and activating detergent-like molecules. Distant from membranes, the β-glycosidase would be less active, thus hydrolyzing plant glucosides (see below) ingested by the insect with decreased efficiency.

In agreement with the hypothesis presented above, βGly47 from *S. frugiperda* can hydrolyze glycosylceramide, although with low activity (Marana *et al*., 2000). In mammals, sphingosine and ceramide hydrolysis are dependent on enzyme activation by proteins called saposins (Harzer *et al*., 2001). Given that genome sequences similar to saposins were found in *Drosophila melanogaster*, insect β-glycosidases active on glycolipids may also need the same kind of activators, and their absence in the assay reaction may explain why the activity against ceramides is low or not detected at all.

Free energy relationships (Withers and Rupitz, 1990) were used to compare the specificity of insect β-glycosidase subsites (Azevedo *et al*., 2003). The enzymes with more similar active sites are *Diatraea saccharalis* βGly3 and *T. molitor* βGly1, followed by *D. saccharalis* βGly1 and *T. molitor* βGly3. *Diatraea saccharalis* βGly1 and βGly3 have active sites somewhat similar to the *S. frugiperda* βGly50K and *T. molitor* βGly1 with *T. molitor* βGly3. *D. saccharalis* βGly2, *T. molitor* βGly4 (Ferreira *et al*., 2001), and *S. frugiperda* βGly47K (Marana *et al*., 2000) are similar in some aspects. The latter three enzymes seem to have two active sites and are, in each insect, primarily responsible for β-galactoside digestion. Since each of the three *D. saccharalis* β-glycosidases has a counterpart in *T. molitor*, Azevedo *et al*. (2003) speculated that insects with the same number of β-glycosidases could have similar enzymes.

The only insect β-glycosidases sequenced to date have been those from the moth *S. frugiperda* midgut (βGly50) (Marana *et al.*, 2001), the beetle *T. molitor midgut* (βGly1) (Ferreira *et al.*, 2001), and termite *Neotermes koshunensis* salivary glands (Tokuda *et al.*, 2002). The three enzymes share about 50% identity with one another and to other enzymes from glycosyl hydrolases from family 1, which have two glutamic acid residues responsible for catalysis and a structure of $(\beta/\alpha)_8$-barrel. In the βGly50 from *S. frugiperda* the pK_a of the nucleophile (Glu 399) is 4.9 and of the proton donor (Glu 187) is 7.5. In this enzyme, residue Glu 451 seems to be a key residue in determining the enzyme preference for glucosides versus galactosides. This is due to its interaction, in the glycone site, with substrate equatorial or axial hydroxyl 4, which is the only position where glucose differs from galactose. The steric hindrance of the same residue with hydroxyl 6 probably also explains why fucosides are the best substrate for many β-glycosidases (Marana *et al.*, 2002b).

Besides having a role in digestion, β-glycosidases are important in insect–plant relationships. To avoid herbivory plants synthesize a large number of toxic glucosides (**Figure 6**) which may be present in concentrations from 0.5% to 1% (Spencer, 1988). The presence of these glucosides in some insect diets may explain the variable number of β-glycosidases with different specificities present in their guts. Most plant glucosides have a hydrophobic aglycone and are β-linked *O*-glycosyl compounds. Since aglycones are usually more toxic than the glycosides themselves, intoxication may be avoided by decreasing the activity of the enzyme most active on toxic glucosides, without affecting the final digestion of hemicellulose and cellulose carried out by the other enzymes. This is exemplified by *D. saccharalis* larvae, which have three β-glycosidases in their midgut, feeding on diets containing the cyanogenic glucoside amygdalin. In this condition, the activity of the enzyme responsible for the hydrolysis of prunasin is decreased (Ferreira *et al.*, 1997). Prunasin is the saccharide resulting after the removal from amygdalin, and that forms the cyanogenic mandelonitrille upon hydrolysis (**Figure 6**). Resistance to toxic glucosides may also be achieved by the lack of enzymes able to hydrolyze toxic β-glucosides,

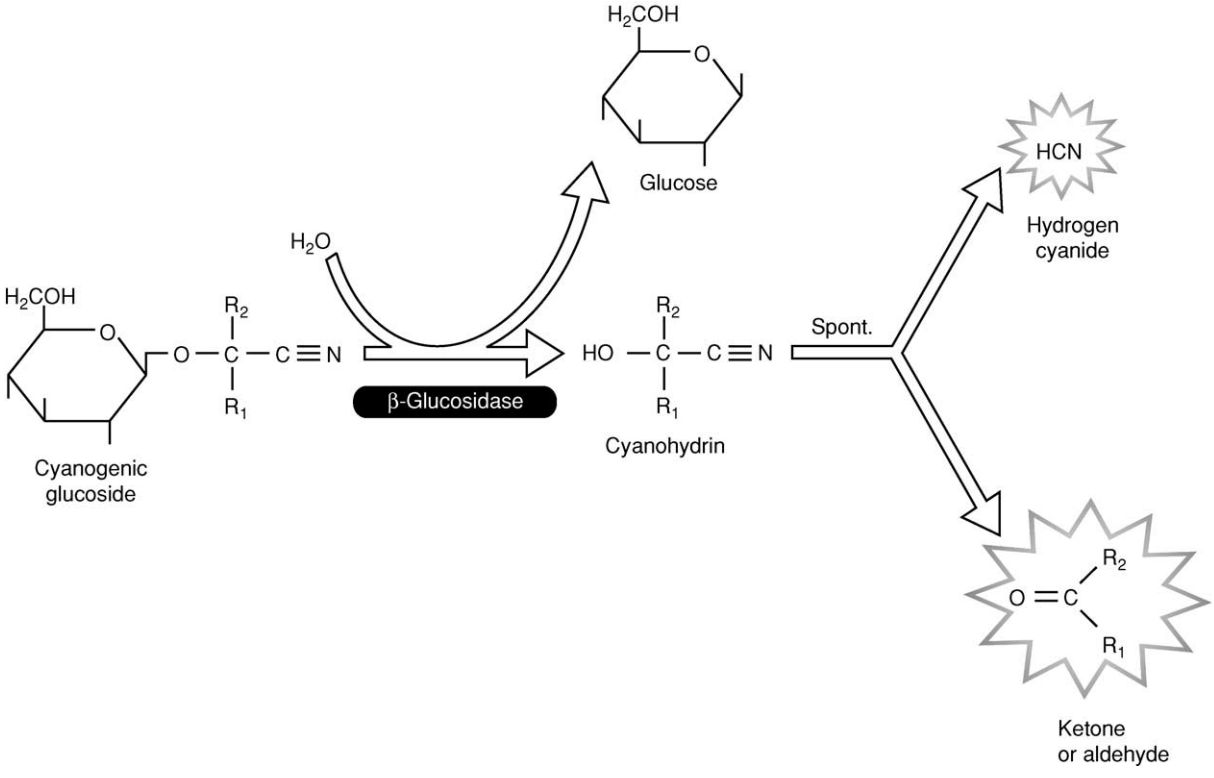


Figure 6 β-Glucosidase acting on a cyanogenic glucoside releases glucose and cyanohydrin that spontaneously decomposes, producing a ketone (or an aldehyde) and hydrogen cyanide. If $R_1 = R_2 = CH_3$, the glucoside is linamarin and the resulting ketone is acetone. If $R_1 = H$ and R_2 = phenyl, the glucoside is prunasin and the resulting aldehyde is benzaldehyde (see more examples in Vetter (2000)).

as observed with *S. frugiperda* larvae, which have two β-glycosidases unable to efficiently hydrolyze prunasin (Marana *et al.*, 2001; S.R. Marana, personal communication).

Progress in this field will require disclosing the mechanisms by which the presence of toxic β-glucosides differentially affects the midgut β-glycosidases and knowing the details of the active site architecture responsible for the specificity of these enzymes.

4.5.4.8. Trehalases

Trehalase (EC 3.2.1.28) hydrolyzes α,α′-trehalose into two glucose molecules and is one of the most widespread carbohydrases in insects, occurring in most tissues. Trehalase is very important for insect metabolism, since trehalose is the main circulating sugar in these organisms. Apical and basal trehalases can be distinguished in insect midguts. The apical trehalase may be soluble (glycocalyx-associated or secreted into the midgut lumen) or microvillar, whereas the midgut basal trehalase is an integral protein of the basal plasma membrane. The apical midgut trehalase is a true digestive enzyme. The midgut basal trehalase probably plays a role in the midgut utilization of hemolymph trehalose (Terra and Ferreira, 1994). In spite of the importance of trehalase, it is poorly studied in insects and also in other sources. There is not a single trehalase with known catalytical groups or with its 3D structure resolved.

Trehalases partially or completely purified from insect guts have pH optima from 4.8 to 6.0, Km from 0.33 to 1.1 mM, pI around 4.6, and molecular masses from 60 to 138 kDa (Terra and Ferreira, 1994).

Tris is usually a competitive inhibitor of trehalase. Inhibition of midgut trehalases at their optimum pH was reported in *Apis mellifera*, *R. americana*, and *B. mori*, with Kis of 50, 74, and 47 mM, respectively, (Terra and Ferreira, 1994). Tris inhibited *R. americana* trehalase at pH 9.0 with a higher Ki (182 mM) than at pH 6.0 (74 mM), suggesting the presence of a negative charge at the active site to which the protonated Tris binds (Terra *et al.*, 1978). In the Lepidoptera *Lymantria dispar* (Valaitis and Bowers, 1993) and *S. frugiperda* (Silva, Terra, and Ferreira, unpublished data) Tris up to 100 mM does not inhibit the enzymes near the pH optimum. Contrasting to what was found in trehalase from *R. americana*, in *S. frugiperda* Tris is a competitive inhibitor with a small Ki (0.55 mM) at pH 9.0. The results may be related to differences in the active site of midgut trehalases from Lepidoptera.

There have been a few attempts to identify important groups in the midgut trehalase active site. Terra *et al.* (1978, 1979, 1983) determined the pK_a values of the catalytical groups of the *R. americana* midgut trehalase. The pK_a value of the nucleophile was 5.0 (kinetic data) or 5.3 (carbodiimide modification results), whereas the pK_a value of the proton donor was 8.3 (kinetic data) or 7.7 (carbodiimide modification). Since there was a disagreement between the pK_a values determined for the proton donor, and taking into account that carbodiimide modification is only partially protected by trehalose, the authors suggested that the proton donor is near, but not at the active site, and that it participates in the reaction through another amino acid residue, like histidine (Terra *et al.*, 1979).

Lee *et al.* (2001) with the same approach as Terra *et al.* (1979) found pK_a values of 5.3 and 8.5 for the *A. mellifera* trehalase. These authors and Valaitis and Bowers (1993), who worked with *L. dispar* trehalase, showed that the trehalases were inactivated by diethyl pyrocarbonate (DPC), but the work did not progress further. Purified midgut *S. frugiperda* trehalase is modified by DPC only in the presence of 2 × Ki of α-methyl glucoside, a linear competitive inhibitor of the enzyme, indicating that there is a conformational change in the enzyme following inhibitor binding. The modification achieved with DPC affects an imidazole group and only inhibits 60% of the enzyme activity. The pK_a values obtained from kinetic data are 4.8 and 7.6, and the pK_a values calculated from carbodiimide and phenyl glyoxal modification are 4.9 and 7.8, respectively. Whereas 10 Km trehalose competitively protect the enzyme from phenyl glyoxal modification, only partial protection is seen when the modifier is carbodiimide. These and other results suggest that a carboxyl group is the nucleophile and an Arg residue is the proton donor, which pK_a is affected by a nearby imidazole group (Silva, Terra, and Ferreira, unpublished data).

The only insect midgut trehalase sequenced up till now is that from *B. mori* pupae (Su *et al.*, 1993). The corresponding mRNA is actively transcribed in midgut and Malpighian tubules, but not in the other tissues (Su *et al.*, 1994). It is interesting to note that, in mammals, the same trehalases are present in the microvillar membranes of small intestine and kidney. The midgut trehalase from *B. mori* has identity of only 46% with trehalase from the male accessory gland from *T. molitor*. Since similar low identities are found for *B. mori* trehalase (known to be from midgut) and the other insect trehalases deposited in the GenBank, these probably do not have a midgut origin.

Plant toxic β-glucosides and their aglycones can inhibit, with varied efficiency, some or all trehalases

from Malpighian tubules, fat body, midgut, and body wall of *P. americana*, *M. domestica*, *S. frugiperda*, and *D. saccharalis* (Silva, Terra, and Ferreira, unpublished data). Toxic β-glucosides are produced by many plant species and are present in high concentrations (see Section 4.5.4.7). It is not known whether those glucosides or aglycones are absorbed by the insect gut and interact with trehalases in tissues other than the midgut. It would be interesting to know if this happens and how insects resistant to toxic β-glucosides avoid the damage they can cause.

4.5.4.9. Acetylhexosaminidases, β-Fructosidases, and α-Galactosidases

An enzyme related to chitinolytic enzymes is β-*N*-acetyl-D-hexosaminidase (EC 3.2.1.52), which differs from β-*N*-acetyl-D-glucosaminidase in having a rather wide substrate specificity. The enzyme is found in many insects and its presumed physiological role is the hydrolysis of *N*-acetylglucosamine β-linked compounds such as glycoproteins (Terra and Ferreira, 1994). Detailed studies of this enzyme are lacking.

Sucrose hydrolysis is catalyzed by enzymes that are specific for the α-glucosyl (α-glucosidase, EC 3.2.1.20; see above) or for the β-fructosyl residue (β-fructosidase, EC 3.2.1.26) of the substrate. β-Fructosidase is characterized by its activity toward sucrose and raffinose and lack of activity upon maltose and melibiose. In insect midguts, sucrose hydrolysis generally occurs by action of the conspicuous α-glucosidase rather than by β-fructosidase. Only a small number of reports verify the presence of β-fructosidase in insects (Terra and Ferreira, 1994; Scrivener *et al.*, 1997) and there has been only one attempt (Santos and Terra, 1986a) at characterization. Larvae and adults of the moth *Erinnyis ello* have a midgut β-fructosidase with pH optimum 6.0, Km 30 mM (sucrose), pI 5.2, and molecular mass of 78 kDa. The physiological role of this enzyme is to hydrolyze sucrose, the major leaf (larvae) or nectar (adults) carbohydrate, which is not efficiently digested by *E. ello* midgut α-glucosidase (Santos and Terra, 1986b).

α-Galactosidase (α-D-galactoside galactohydrolase, EC 3.2.1.22) catalyzes the hydrolysis of α-D-galactosidic linkages in the nonreducing end of oligosaccharides, galactomannans, and galactolipids and is widely distributed in nature (Dey and Pridham, 1972). Galactooligosaccharides, such as melibiose, raffinose, and stachiose are common in plants, mainly in lipid-rich seeds (Shiroya, 1963), whereas galactolipids are widespread among leaves. The major lipids in chloroplast membranes are monogalactosyldiglyceride and digalactosyldiglyceride (Harwood, 1980).

There have been few attempts to resolve insect midgut α-galactosidases. Gel filtration and heat inactivation suggested that there is a single α-galactosidase activity (30 kDa, pH optimum 5.0) in *Dysdercus peruvianus* midgut that is more efficient on raffinose than on melibiose and NPαGal (Silva and Terra, 1997). There are two α-galactosidases in *Abracris flavolineata* midguts: the major (112 kDa, pH optimum 5.4) is active on melibiose and raffinose in addition to NPαGal, whereas the minor (70 kDa, pH optimum 5.7) hydrolyzes only NPαGal (Ferreira *et al.*, 1999). In the case of *Psacothea hilaris*, gel filtration gave evidence of the presence of multiple overlapping α-galactosidases more active on NPαGal than on melibiose (Scrivener *et al.*, 1997). There are three midgut luminal α-galactosidases (TG1, TG2, and TG3) from *T. molitor* larvae that are partially resolved by ion-exchange chromatography (Grossmann and Terra, 2001). The enzymes have approximately the same pH optimum (5.0), pI value (4.6), and molecular mass (46–49 kDa). TG2 hydrolyzes α-1,6-galactosaccharides, exemplified by raffinose, whereas TG3 acts on melibiose and apparently also on digalactosyldiglyceride, the most important compound in thylacoid membranes of chloroplast, converting it into monogalactosyldiglyceride. *Spodoptera frugiperda* larvae have three midgut α-galactosidases (SG1, SG2, and SG3) partially resolved by ion-exchange chromatography (Grossmann and Terra, 2001). The enzymes have similar pH optimum (5.8), pI value (7.2), and molecular mass (46–52 kDa). SG1 and SG3 hydrolyze melibiose and SG3 digests raffinose and, perhaps, also digalactosyldiglyceride.

4.5.5. Digestion of Proteins

4.5.5.1. Initial Considerations

The initial digestion of proteins is carried out by proteinases (endopeptidases) that break internal bonds in proteins. Different proteinases are necessary to do this because the amino acid residues vary along the peptide chain. There are three subclasses of proteinases involved in digestion classified according to their active site group (and hence by their mechanism): serine, cysteine, and aspartic proteinases. In each of the mentioned subclasses, there are several proteinases differing in substrate specificities. The oligopeptides resulting from proteinase action are attacked from the N-terminal end by aminopeptidases and from the C-terminal end by carboxypeptidases, both enzymes liberating one amino acid residue at each catalytic step. Finally, the dipeptides formed are hydrolyzed by dipeptidases.

4.5.5.2. Serine Proteinases

Serine proteinases (EC 3.4.21) (MEROPS) have serine, histidine, and aspartic acid residues (called the catalytic triad) in the active site. The family of enzymes homologous to chymotrypsin (Barrett *et al.*, 1998) includes the major digestive enzymes trypsin, chymotrypsin, and elastase. These enzymes differ in structural features that are associated with their different substrate specificities, as detailed below. The numbering of residues in enzyme polypeptide chains is referred to that of bovine chymotrypsin.

4.5.5.2.1. Trypsins Trypsins (EC 3.4.21.4) preferentially cleave protein chains on the carboxyl side of basic L-amino acids such as arginine and lysine. Most insect trypsins have molecular masses in the range 20–35 kDa, pI values 4–5, and pH optima 8–10. These enzymes occur in the majority of the insects, with the remarkable exception of hemipteran species and some taxa belonging to the series Cucujiformia of Coleoptera like Curculionidae (Terra and Ferreira, 1994). Nevertheless, some heteropteran Hemiptera have trypsin in the salivary glands (Zeng *et al.*, 2002).

Trypsin is usually identified in insect midgut homogenates using benzoyl-arginine *p*-nitroanilide (B-R-pNA, often referred to as BApNA) or benzoyl-arginine 7-amino-4-methyl coumarin (B-R-MCA) as substrates and with *N*-α-tosyl-L-lysine chloromethyl ketone (TLCK), phenylmethylsulfonyl fluoride (PMSF), or diisopropylfluorophosphate (DFP) as inactivating compounds. The substrates of choice for assaying insect trypsins are those in Figure 7. Trypsins from Orthoptera, Dictyoptera, and Coleoptera are usually purified by ion-exchange chromatography, whereas those from Diptera and Lepidoptera, by affinity chromatography, either in benzamidine-agarose (elution with benzamidine or by change in pH) or in soybean trypsin inhibitor (SBTI)-Sepharose (elution by change in pH). Due to significant autolysis, lepidopteran trypsins are more frequently purified by chromatography on benzamidine-Agarose with elution with benzamidine.

A total of 109 insect trypsin sequences were registered in GenBank in March 2003, corresponding to 34 species pertaining to five orders. Examples may be found among Hemiptera (Zeng *et al.*, 2002), Coleoptera (Girard and Jouanin, 1999a; Zhu and Baker, 1999), Diptera (Davis *et al.*, 1985; Barillas-Mury *et al.*, 1993), Siphonaptera (Gaines *et al.*, 1999), and Lepidoptera (Peterson *et al.*, 1994; Zhu *et al.*, 2000). Several sequences were not complete and some insect species are represented by many sequences that probably include enzymes other than digestive enzymes. The complete sequences have signal and activation peptides and the features typical of trypsin-like enzymes, including the conserved N-terminal residues IVGG, the catalytic amino acid triad of serine proteinase active sites (His 57, Asp 102, and Ser 195), three pairs of conserved cysteine residues for disulfide bonds, and the residue Asp 189 that determines specificity in trypsin-like enzymes (see Figure 5).

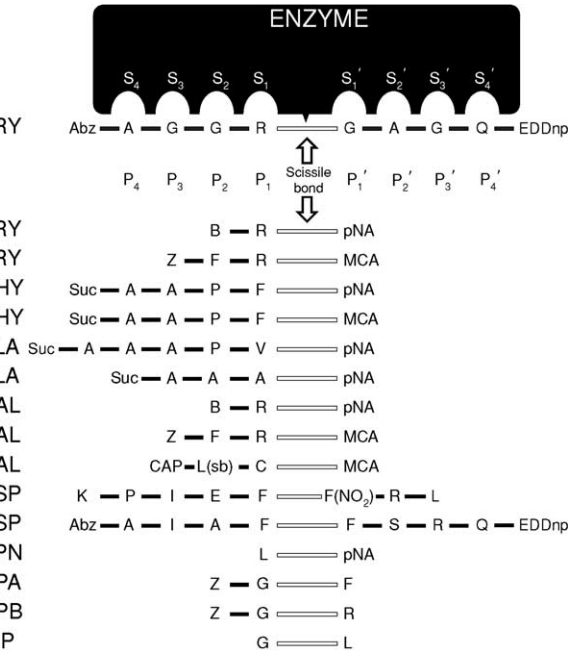


Figure 7 Substrates used in the assay of enzymes involved in protein digestion. S_n are subsites in the enzymes and P_n are amino acid residues in substrates. The arrows point to bonds cleaved by the different enzymes. Abz-Xn-EDDnp is a class of peptides with quenching (EDDnp) and fluorescent (Abz) groups at the C- and N-terminal ends, respectively, so that after hydrolysis the peptides become fluorescent. Substrates with C-terminal MCA are fluorescent and those with pNA are colorimetric. Contrary to GL, LG is also hydrolyzed by APN in addition to dipeptidase. For further details, see text. TRY, trypsin; CHY, chymotrypsin; ELA, elastase; CAL, cathepsin L-like enzyme; ASP, aspartic proteinase; APN, aminopeptidase N; CPA, carboxypeptidase A; CPB carboxypeptidase B; DIP, dipeptidase.

In spite of having structural features resembling vertebrate trypsins, insect trypsins differ from these because they are not activated or stabilized by calcium ions and frequently are unstable in acidic pH (Terra and Ferreira, 1994). Another contrasting aspect is the almost lack of reports on the isolation of inactive insect trypsinogens, that is, trypsin molecules still having the activation peptide (Reeck *et al.*, 1999). The activation of insect trypsinogen to trypsin deserves a closer look. Finally, other differences between vertebrate and insect trypsins include their substrate specificities and their interaction with protein inhibitors.

Amino acyl residues in proteinase substrates are numbered from the hydrolyzed peptide bonds as P_1, P_2, P_3, ..., P_n in the direction of the N-terminus and P'_1, P'_2, P'_3, ..., P'_n in the direction of the peptide C-terminus, whereas the corresponding enzyme subsites are numbered S_1, S_2, S_3, ..., S_n and S'_1, S'_2, S'_3, ..., S'_n (Schechter and Berger, 1967) (**Figure 7**). Mammalian trypsin preferably cleaves substrates having arginine rather than lysine at P_1 (primary specificity) (Craik *et al.*, 1985). The same primary specificity was found for insect trypsins, except those from lepidopterans, which prefer lysine at P_1 (Lopes *et al.*, 2004). This will be discussed below in relation to trypsin insensitivity to protein inhibitors. In order to characterize the trypsin specificity at subsites other than S_1, quenched fluorescent substrates like *o*-aminobenzoyl-GGRGAGQ-2,4-dinitrophenyl-ethylene diamine (where R stands for arginine at P_1 position) were synthesized with 15 amino acid replacements at each of the positions P'_1, P'_2, P_2, and P_3. The results suggested that trypsin subsites are more hydrophobic in trypsins from the more evolved insects (A.R. Lopes *et al.*, unpublished data). Trypsin from different insects also differ in the strength their subsites bind the substrate or the transition state (high-energy intermediate of the reaction). In other words, trypsin subsites differ in how they favor substrate binding or catalysis (Marana *et al.*, 2002a).

Plants have protein inhibitors (PIs) of insect midgut serine proteinases that affect insect development (Ryan, 1990). Insects may adapt to the presence of PIs in diet by overexpressing proteinases (Bown *et al.*, 1997; Broadway, 1997; Gatehouse *et al.*, 1997), by proteolytical inactivation of PIs mediated by the insect's own proteinases (Giri *et al.*, 1998), or by expressing new proteinases that are resistant to the inhibitor (Mazumbar-Leighton and Broadway, 2001a, 2001b). Current research is investigating the molecular basis of the difference between sensitive and inhibitor-insensitive trypsins, as well as the regulation of these enzymes.

PIs produced by plants have a region, named the reactive site, that interacts with the active site of their target enzymes. The reactive sites of many PIs are hydrophilic loops with a lysine residue at P_1 (Lopes *et al.*, 2004). As lepidopteran trypsins have hydrophobic subsites and prefer lysine rather than arginine at P_1 (see above), they are usually more resistant to PIs than the other insect trypsins. In this respect, it is interesting to note that PI-insensitive trypsins from *Heliothis virescens* bind more tightly to a hydrophobic chromatographic column than sensitive trypsins do (Brito *et al.*, 2001). These observations lead to the hypothesis that the molecular differences between sensitive and insensitive trypsins must rely on the interactions of PIs with residues in and around the enzyme active site.

An interesting approach to study insect–PI interactions was introduced by Volpicella *et al.* (2003). They compared the sequence of a sensitive trypsin from *Helicoverpa armigera* with the insensitive trypsin from the closely related species *H. zea*. The 57 different amino acids observed between the two enzymes were superimposed on the porcine trypsin crystal structure, where the residues known to be in contact to a Kunitz-type inhibitor (Song and Suh, 1998) were identified. The residues at positions (chymotrypsin numbering) 41, 57, 60, 95, 99, 151, 175, 213, 217, and 220 were considered by Volpicella *et al.* (2003) to be important in *H. zea* trypsin–PI interaction. However, some of the interacting residues may have been misidentified because trypsins from different species were compared. In a similar approach, Lopes *et al.* (2004) aligned all available trypsin sequences characterized as sensitive or insensitive to Kunitz-type inhibitor (Bown *et al.*, 1997; Mazumbar-Leighton and Broadway, 2001a) with porcine trypsin. After discounting conserved positions and positions not typical of sensitive or insensitive trypsins, the remaining positions that agree with those involved in porcine trypsin–PI (Bowman-Birk type, Lin *et al.*, 1993; Kunitz type, Song and Suh, 1998) or substrate (Koepke *et al.*, 2000) interactions were: 60, 94, 97, 98, 99, 188, 190, 213, 215, 217, 219, 228. These positions support the tree branches in a neighbor-joining analysis of sensitive (I, III) and insensitive (II) trypsin sequences (Lopes *et al.*, 2004) (**Figure 8a**). Site-directed mutagenesis of trypsin, followed by the determination of the binding constants of mutated trypsins with PIs, may help to resolve the discrepancy.

The mechanism by which PIs in diet induces the synthesis of insensitive trypsin in responsive insects remains completely unknown, as well as the regulatory element that may be involved.

4.5.5.2.2. Chymotrypsins Chymotrypsin (EC 3.4.21.1) preferentially cleaves protein chains at the carboxyl side of aromatic amino acids. Insect chymotrypsins usually have molecular masses of 20–30 kDa and pH optima of 8–11, and they differ from their vertebrate counterparts in their instability at acidic pH, inhibition pattern with SBTI (Terra and Ferreira, 1994) and, finally, in reacting with *N*-α-tosyl-L-phenylalanine chloromethyl ketone (TPCK) (see below). The distribution of chymotrypsin among insect taxa is similar to that of trypsin (Terra and Ferreira, 1994), including the occurrence in the salivary glands of some heteropteran bugs

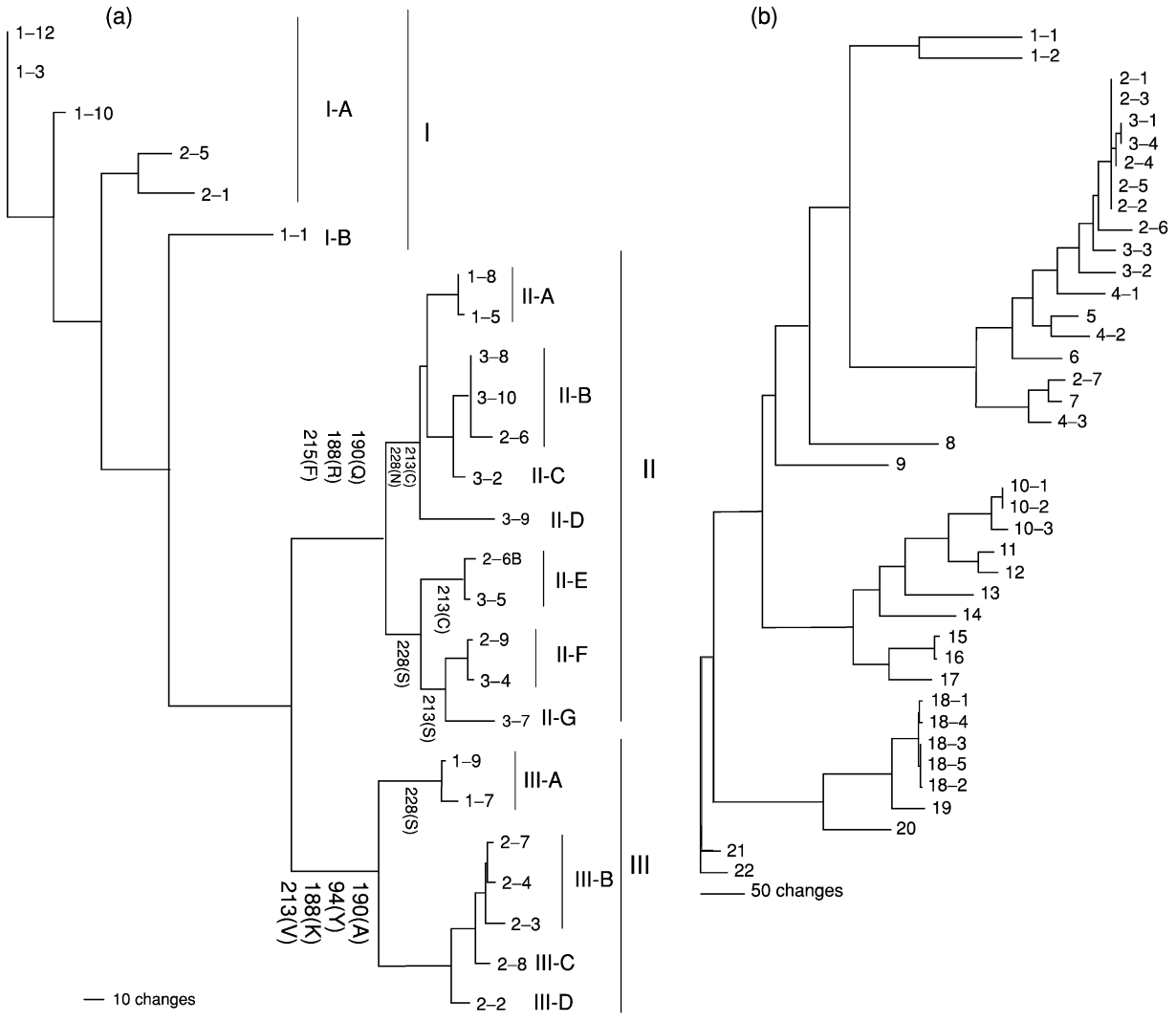


Figure 8 Neighbor-joining distance analysis tree of insect trypsin and chymotrypsin sequences. (a) Clusters formed by sequences of Lepidoptera trypsins sensitive (I, III) or insensitive (II) to protein inhibitors according to Bown *et al.* (1997) and Mazumbar-Leighton and Broadway (2001b). The amino acid positions important in defining the clusters are indicated. (b) Clusters formed by sequences of insect chymotrypsins. The sequences are identified by two digits: the first denotes the insect species (see below) and the second identify the different sequences of the same insect. Trypsin sequences: 1, *Agrotis ipsilon*; 2, *Helicoverpa zea*; 3, *H. armigera*. Chymotrypsin sequences: 1, *Plodia interpunctella*; 2, *H. armigera*; 3, *H. zea;* 4, *A. ipsilon*; 5, *Spodoptera frugiperda*; 6, *Manduca sexta*; 7, *Heliothis virescens*; 8, *Phaedon cochleariae*; 9, *Scirpophaga incertulas*; 10, *Anopheles gambiae*; 11, *A. aquasalis*; 12, *A. darlingi*; 13, *Rhyzopertha dominica*; 14, *Ctenocephalides felis*; 15, *Vespa crabro*; 16, *V. orientalis*; 17, *Solenopsis invicta*; 18, *Aedes aegypti*; 19, *Culex pipiens pallens*; 20, *Glossina morsitans morsitans*; 21, cow; 22, rat. (Courtesy of A.R. Lopes.)

(Colebatch *et al.*, 2002). The earlier failure to detect chymotrypsin activity in insect midguts was a consequence of using the mammalian chymotrypsin substrates, like benzoyl-tyrosine *p*-nitroanilide (B-Y-pNA) or benzoyl-tyrosine ethyl ester (B-Y-ee), in the assays. Insect chymotrypsins have an extended active site and larger substrates, like succinyl-AAPF-*p*-nitroanilide (Suc-AAPF-pNA), are necessary for their detection (Lee and Anstee, 1995; Lopes *et al.*, 2004) (**Figure 7**). Insect chymotrypsins are usually purified by affinity chromatography in phenyl butylamina-Sepharose (elution with phenyl butylamina) or in SBTI-Sepharose (elution with benzamidine) for enymes from lepidopterans, and by ion-exchange chromatography for those from dictyopterans, orthopterans, hymenopterans, and dipterans.

There are 63 complete sequences of insect chymotrypsins registered in the GenBank (as of March 2003) corresponding to 22 species pertaining to six different orders. Examples are found among Hemiptera (Colebatch *et al.*, 2002), Coleoptera (Girard and Jouanin, 1999a; Zhu and Baker, 2000), Hymenoptera (Jany *et al.*, 1983; Whitworth *et al.*, 1998),

Siphonaptera (Gaines *et al.*, 1999), Diptera (Casu *et al.*, 1994; de Almeida *et al.*, 2003), and Lepidoptera (Peterson *et al.*, 1995; Bown *et al.*, 1997). All the sequences have a signal peptide, an activation peptide (ending with an arginine residue), the catalytic triad (His 57, Asp 102, and Ser 195), three pairs of conserved cysteine residue, conserved N-terminal sequence IVGG and Ser/Gly/Tyr 189 which confers specificity to chymotrypsin-like enzymes (**Figure 5**).

The insect digestive chymotrypsin that has been most thoroughly studied is that of *Manduca sexta* (Lepidoptera: Sphingidae) (Peterson *et al.*, 1995). In this enzyme, the activation peptide is longer and has a net charge different from that of bovine chymotrypsinogen, leading the authors to suggest that the insect enzyme is activated by a peculiar mechanism. The mammalian chymotrypsin has a pH optimum around 8 with two catalytic important pK_as of 6.8 and 9.5, corresponding to the active-site histidine, and N-terminal leucine, respectively. In contrast, the *M. sexta* chymotrypsin has pH optimum 10.5–11 and a single kinetically significant pK_a at pH 9.2. This pK_a may represent the active-site histidine in an appropriate environment, although several other hypotheses were discussed (Peterson *et al.*, 1995). It is not clear whether the insect chymotrypsin active-site changes associated with TPCK resistance (see below) may also be the cause of the putative histidine pK_a displacement.

The resolution of the 3D structure of the fire ant digestive chymotrypsin led to the conclusion that it is strikingly similar to mammalian chymotrypsin, but has differences beyond those found among homologs from different mammalian systems (Botos *et al.*, 2000). The similarities include a conserved backbone scaffold and structural domains. Differences include the activation mechanism and substitutions in the subsite S_1 and mainly in the other subsites (S_4–S_4') that suggest different substrate specificities and interactions with PIs. In agreement with this, different insect chymotrypsins are sensitive to distinct PIs and like trypsins, PI-insensitive chymotrypsins may be induced in insects ingesting PI-containing diets (Bown *et al.*, 1997; Mazumbar-Leighton and Broadway, 2001b).

Chymotrypsin sequences form several branches in a neighbor-joining analysis that correspond to phylogenetic groups (**Figure 8b**). Lepidoptera species are clustered in the two branches supported by Trp 59 (all sequences have a conserved tryptophan at position 59), except for the oligophagous pyralid *Scirpophaga incertulas* (9 in **Figure 8b**). One of the major lepidopteran branches corresponds to sequences from the pyralid *Plodia interpunctella* (1 in **Figure 8b**) and the other two sequences from noctuids and sphingids, all of them polyphagous. The data suggest that Trp 59 is related to a polyphagous habit.

Chymotrypsins from insects that routinely ingest ketone-releasing compounds (like several plant glycosides) (see **Figures 6** and **9**) are not affected much by these compounds and others that react with His 57. Thus, in comparison with bovine chymotrypsin, the chymotrypsin from polyphagous lepidopteran insects reacts slowly with chloromethyl ketones, whereas those of oligophagous pyralid insects react rapidly (A.R. Lopes *et al.*, unpublished data). Modeling *Spodoptera frugiperda* (Noctuidae) chymotrypsin, based on its sequence and on crystallographic data of bovine chymotrypsin, showed that His 57 is probably protected from alkylation by a bulky (tryptophan) residue at position 59 (A.R. Lopes *et al.*, unpublished data). Bovine chymotrypsin has at position 59 Gly, thus rendering His 57 exposed to the medium. These adaptations are new examples of the interplay between insects and plants during their evolutionary arms race and deserve more attention through site-directed mutagenesis of recombinant chymotrypsins.

4.5.5.2.3. Elastases Since Christeller *et al.* (1990) described an elastase (EC 3.4.21.36)-like enzyme in

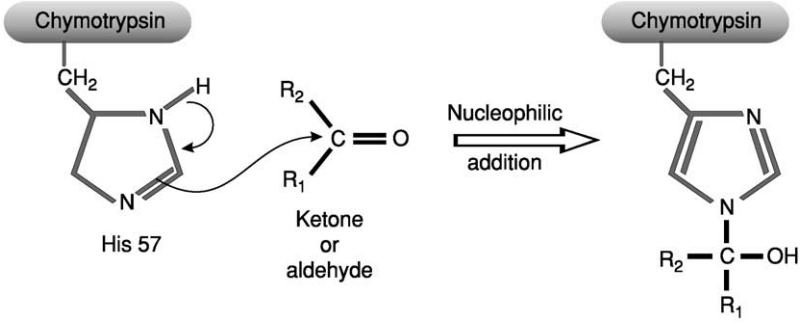


Figure 9 Ketones or aldehydes formed after the action of β-glucosidases on cyanogenic glucosides (**Figure 6**) may react with His 57 of chymotrypsin, inactivating it.

the cricket *Teleogryllus commodus*, this enzyme has been described in many other insects, including in homogeneous form (Terra and Ferreira, 1994; Whitworth *et al*., 1998). Usually elastase is identified with the substrate Suc-AAPL-pNA (**Figure 7**), combined with the observation of lack of activity on B-Y-pNA or B-Y-ee and resistance to TPCK. Since the mentioned substrate may also be hydrolyzed by chymotrypsin and lack of activity on B-Y-pNA and/or resistance to TPCK are usual among chymotrypsins (see Section 4.5.5.2.2), most described elastase may actually be chymotrypsins. True elastases were isolated from gypsy moth midguts (Valaitis, 1995) and from whole larvae of *Solenopsis invicta* (Whitworth *et al*., 1998). The last-mentioned enzymes hydrolyze Suc-APA-pNA, but not substrates with phenylalanine at P_1. Although the specific substrate for elastase (Suc-AAA-pNA) (Bieth *et al*., 1974) was not tested, the hydrolysis of Suc-AAAPV-pNA and the lack of hydrolysis of substrates with phenylalanine in P_1 discount a chymotrypsin. One of the *S. invicta* elastases (E2) was cloned, sequenced, and shown to be more similar to chymotrypsin than to elastase (Whitworth *et al*., 1999). This work confirms the occurrence of elastase in insect midgut. Further work is necessary to evaluate the extent of this enzyme in insect midguts and the importance in digestion.

4.5.5.3. Cysteine Proteinases

Cysteine proteinase is usually assayed in insect midgut contents or midgut homogenates at pH 5–6 with B-R-pNA, B-R-NA, casein, or hemoglobin as substrate. Activation by sulfhydryl agents (dithiothreitol (DTT) or cysteine) and inhibition by trans-epoxysuccinyl-L-leucyl-amido (4-guanidinobutane) (E-64) are usually indicative of the presence of the enzyme. The observation of inhibition of hydrolytic activity on any of the mentioned substrates by E-64 is insufficient for a positive identification of cysteine proteinase. Trypsin hydrolyzes the same substrates and may be reversibly inhibited by E-64 (Novillo *et al*., 1997). The identification of cysteine proteinase was made easier with the substrate ε-aminocaproyl-leucyl-(*S*-benzyl)-cysteinyl-MCA, which is hydrolyzed by cysteine proteinase but not by serine proteinases (Alves *et al*., 1996). Using such criteria, cysteine proteinases were described in Hemiptera Heteroptera and in species belonging to the series Cucujiformia of Coleoptera (Terra and Ferreira, 1994). Exceptions to this rule are the identification of cysteine proteinase in Hemiptera Auchenorrhyncha (aphids) (Cristofoletti *et al*., 2003) and the lack of this enzyme in cucujiform cerambycid beetles (Johnson and Rabosky, 2000).

Insect midgut cysteine proteinases were at first denoted as cathepsin B (EC 3.4.22.1)-like enzymes, because cathepsin B was the first animal cysteine proteinase described. Later on it became known that cathepsin B is more important as a peptidyl dipeptidase, rather than as an endopeptidase, because of the existence of an extended loop that forms a cap to the active-site cleft, and carries a pair of histidine residues that are thought to bind to the C-terminal carboxylate of the substrate (Barrett *et al*., 1998). Cathepsin L (EC 3.4.22.15) is a true endopeptidase that preferentially cleaves peptide bonds with hydrophobic amino acid residues in P_2 (cathepsin B prefers arginine at the same position) (Barrett *et al*., 1998). Thus, by using substrates like carbobenzoxy (Z)-FR-MCA and Z-RR-MCA it is possible to distinguish between the two enzymes (see **Figure 7**). Current research revealed that cathepsin L-like enzymes are the only insect cysteine proteinase quantitatively important. Much more difficult to ascertain is that a cathepsin L-like enzyme assayed in insect midguts has been secreted into midgut contents, and hence may be considered as a truly digestive enzyme. As in other animals (Barrett *et al*., 1998), cathepsin L-like enzymes in insects are expected to occur in lysosomes and never leave the cells. The same difficulties arise in trying to relate digestion to cathepsin L-like enzymes encoded by cDNAs cloned from midgut cells.

The problems that may arise during cathepsin L-like enzyme characterization are well illustrated in a study with *T. molitor* larvae. Three cathepsin L-like sequences were recognized in a cDNA library prepared from *T. molitor* midguts. One sequence after being expressed and used to raise antibodies was found to correspond to a lysosomal cathepsin L immunolocalized mainly at hemocytes and fat body cells. The second sequence was not related yet with any enzyme active in midgut. Finally, the third one corresponds to a cathepsin L-like enzyme purified from midgut contents (Cristofoletti *et al*., unpublished data).

Digestive cathepsin L-like enzymes have been purified to homogeneity only from *Diabrotica virgifera* (Coleoptera: Cucujiformia) (Koiwa *et al*., 2000), *Acyrthosiphon pisum* (Hemiptera: Auchenorrhyncha) (Cristofoletti *et al*., 2003), and *T. molitor* (Coleoptera: Cucujiformia) (Cristofoletti *et al*., unpublished data). The *A. pisum* enzyme is cell membrane-bound and faces the luminal contents, whereas *D. virgifera* and *T. molitor* ones are soluble enzymes secreted into midgut contents. The complete purification of those enzymes was achieved by affinity chromatography with soyacystatin as a ligand or by a combination of ion-exchange

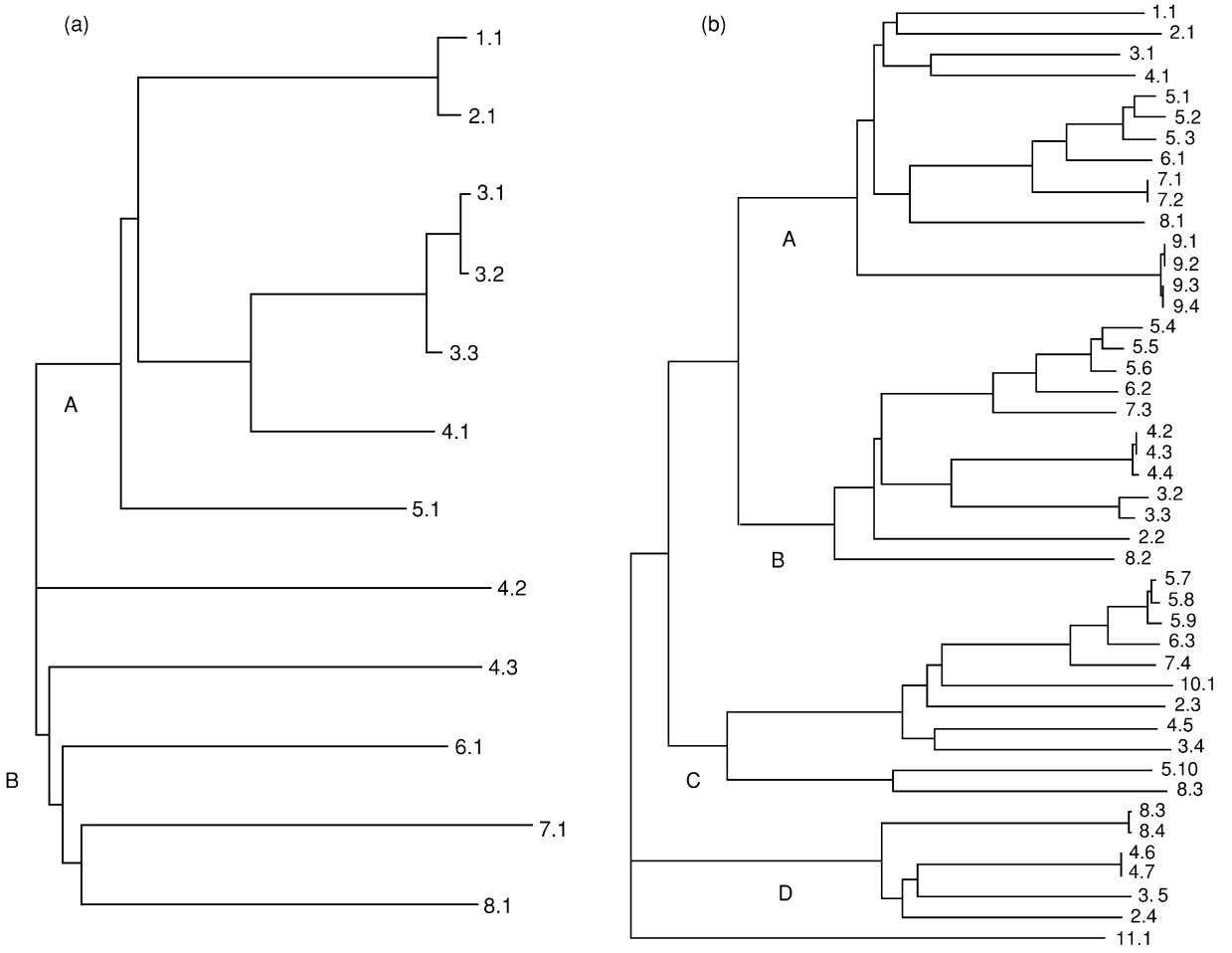


Figure 10 Neighbor-joining distance analysis tree of insect cathepsin L and aminopeptidase N sequences. (a) Clusters formed by cathepsin L-like sequences of insects known to have digestive cathepsins. (b) Clusters formed by coleopteran and lepidopteran aminopeptidases N. Sequences are identified as described in **Figure 8**. Cathepsin L-like sequences: 1, *Myzus persicae*; 2, *Aphis gossypii*; 3, *Sitophilus zeamais*; 4, *Tenebrio molitor; 5, Rhodnius prolixus*; 6, *Diabrotica virgifera*; 7, *Phaedon cochleariae*; 8, *Hypera postica*. Aminopeptidase N sequences: 1, *Epiphyas postvittana*; 2, *Lymantria dispar*; 3, *Bombyx mori*; 4, *Manduca sexta*; 5, *Helicoverpa armigera*; 6, *Heliothis punctigera*; 7, *H. virescens*; 8, *Plutella xylostella*; 9, *Plodia interpunctella*; 10, *Spodoptera littura*; 11, *Tenebrio molitor*. (Courtesy of P. T. Cristofoletti.)

chromatographies. The enzymes have pH optima of 5–6, molecular masses of 20–40 kDa, prefer Z-FR-MCA over Z-RR-MCA, are inhibited by E-64, and activated by cysteine or DTT. At least the *A. pisum* enzyme is also inhibited by chymostatin (completely) and elastatinal (partly) (Cristofoletti *et al.*, 2003).

Twelve cathepsin L-like sequences corresponding to eight species of coleopterans and hemipterans (those known to have digestive cathepsins) are registered in GenBank (as of March 2003). The sequences demonstrate the features characteristic of family 1 of cysteine proteinases (Barrett *et al.*, 1998): N-terminal propeptide that must be removed to activate the enzyme and the catalytic triad, Cys 25, His 169, and Asn 175 (papain numbering) and the ERFININ motif (**Figure 5**). The sequences form two major branches in a neighbor-joining analysis (**Figure 10a**) (P.T. Cristofoletti *et al.*, unpublished data). The data suggest that the Hemiptera digestive cathepsin is close to the lysosomal one from *T. molitor* (sequence 1), whereas those from the coleopterans (except from *Sitophilis oryzae*) are similar to the *T. molitor* digestive enzyme (sequence 3) (**Figure 10a**). It is probable that the *S. oryzae* enzymes are not true digestive but lysosomal, like *T. molitor* sequence 1 (P.T. Cristofoletti *et al.*, unpublished data). More work is needed to clarify the role of cathepsin L-like enzymes in insect digestion.

4.5.5.4. Aspartic Proteinases

Aspartic proteinases are active at acid pH, hydrolyze internal peptide bonds in proteins, and attack some synthetic substrates, either chromophoric (Dunn

et al., 1986) or internally quenched fluorescent substrates (Pimenta *et al*., 2001) (**Figure 7**). Mammalian cathepsin D has a substrate-binding cleft that can accommodate up to seven amino acids and prefers to cleave between two hydrophobic residues (Barrett *et al*., 1998).

The first report of aspartic proteinases in insects was made by Greenberg and Paretsky (1955), who found a strong proteolytic activity at pH 2.5–3.0 in homogenates of whole bodies of *Musca domestica*. Lemos and Terra (1991b) showed that the enzyme occurs in midguts and is cathepsin D-like. An aspartic proteinase similar to cathepsin D was found in families of Hemiptera, Heteroptera and in several families belonging to the cucujiform series of Coleoptera (Terra and Ferreira, 1994). Thus, it is possible that aspartic proteinases occur together with cysteine proteinase in Hemiptera and in most Coleoptera.

The aspartic proteinase isolated from *Callosobruchus maculatus* (pH optimum 3.3, 62 kDa) (Silva and Xavier-Filho, 1991) and *Tribolium castaneum* (pH optimum 3.0, 22 kDa) (Blanco-Labra *et al*., 1996) were partially purified and shown to be similar to cathepsin D. These studies need to be extended, so that the origin, specificity, and structure of insect cathepsin D-like enzymes be clarified.

4.5.5.5. Aminopeptidases

Aminopeptidases sequentially remove amino acids from the N-terminus of peptides and are classified on the basis of their dependence on metal ions (usually Zn^{2+} or Mn^{2+}) and substrate specificity. Aminopeptidase N (EC 3.4.11.2) has a broad specificity, although it removes preferentially alanine and leucine residues from peptides, whereas aminopeptidase A (EC 3.4.11.7) prefers aspartyl (or glutamyl)-peptides as substrates. Both are metalloenzymes (Norén *et al*., 1986).

In insect midguts, major amounts of soluble aminopeptidases are found in less evolved insects (e.g., Orthoptera, Hemiptera, Coleoptera Adephaga), whereas in more evolved insects (e.g., Coleoptera Polyphaga, Diptera, and Lepidoptera) aminopeptidase is found mainly bound to the microvillar membranes of midgut cells (Terra and Ferreira, 1994). Insect midgut aminopeptidases are metalloenzymes (ethylenediaminetetraacetic acid (EDTA) inhibition) and have pH optima of 7.2–9.0, irrespective of the pH of the midgut lumen from the different species, Km values (L-pNA) of 0.13–0.78 mM and molecular masses of 90–130 kDa. With a single exception (see below), all known insect aminopeptidases have a broad specificity, hydrolyzing a variety of amino acyl β-naphthylamides (except acidic amino acyl β-naphthylamides), indicating they are aminopeptidases N (Terra and Ferreira, 1994) (**Figure 7**). The exception is a soluble glycocalyx-associated midgut aminopeptidase from *R. americana*. This enzyme is an aminopeptidase removing N-terminal aspartic acid or glutamic acid residues from peptides that are not efficiently attacked by the other aminopeptidases (Klinkowstrom *et al*., 1994).

In addition to a midgut aminopeptidase A, the dipteran *R. americana* has three midgut aminopeptidases N (one soluble and two membrane-bound). The soluble aminopeptidase N (115.7 kDa) prefers tetrapeptides over tripeptides (Ferreira and Terra, 1984), like the minor 107 kDa membrane-bound enzyme, whereas the contrary is true for the major 169 kDa membrane-bound aminopeptidase (Ferreira and Terra, 1985, 1986a, 1986b). The single midgut aminopeptidase N of the coleopterans *Attagenus megatoma* (Baker and Woo, 1981) and *Tenebrio molitor* (Cristofoletti and Terra, 1999) resemble the 115.7 kDa and 107 kDa aminopeptidase of *R. americana*. Approximately the same substrate specificity was observed with the two midgut aminopeptidases of the lepidopteran *Tineola bisselliella* (Ward, 1975a, 1975b). The data suggest that panorpoid insects (Diptera and Lepidoptera) present multiple aminopeptidases with different substrate specificities, in contrast with the single aminopeptidase of coleopterans. However, much more data are needed to support this hypothesis.

There have been few attempts to characterize the active site of insect midgut aminopeptidases. Using multiple inhibition analysis and observing the protection against EDTA inactivation that different competitive inhibitors conferred to the enzyme, two subsites were proposed to occur in the active center of *R. americana* microvillar aminopeptidase: a hydrophobic subsite, to which isoamyl alcohol binds exposing the metal ion, and a polar subsite, to which hydroxylamine binds. Exposure of the metal ion after isoamyl alcohol binding may be analogous to the situation that results when part of the substrate occupies the hydrophobic subsite, causing conformational changes associated with the catalytic step (Ferreira and Terra, 1986b). The effect of pH at different temperatures on kinetic parameters of *T. molitor* midgut aminopeptidase and its inactivation by different compounds were studied (Cristofoletti and Terra, 2000). The data showed that *T. molitor* aminopeptidase catalysis depends on a metal ion, a carboxylate, and a protonated imidazole group and is, somehow, influenced by an arginine residue in the neighborhood of the active site. The catalytic metal binding depends on at least a deprotonated imidazole. In addition to the

above-mentioned groups involved in catalysis, at least one phenol group and one carboxylate are associated with substrate binding. Thus, *T. molitor* aminopeptidase shares common features with those of other zinc metallopeptidases, especially with mammalian aminopeptidase N, but it differs in some details. An imidazole group seems to be involved in catalysis in *T. molitor* aminopeptidase; this is not observed in mammalian aminopeptidase N, which has an imidazole group participating in substrate binding.

Aminopeptidase N sequences are available for the following lepidopterans: *Epiphyas postvittana, B. mori, Heliothis virescens, H. punctigera, Helicoverpa armigera, L. dispar, Manduca sexta, P. interpuctella*, and *Plutella xylostella* (Gill *et al.*, 1995; Knight *et al.*, 1995; Denolf *et al.*, 1997; Hua *et al.*, 1998; Chang *et al.*, 1999; Oltean *et al.*, 1999; Yaoi *et al.*, 1999; Garner *et al.*, 1999; Emmerling *et al.*, 2001; Rajagopal *et al.*, 2003). The sequences have a signal peptide, a conserved RLP motif near the N-terminal, a zinc binding/gluzincin motif $HEXXHX_{18}E$, a GAMEN conserved motif and a long hydrophobic C-terminal containing a glycosyl phosphatidyl inositol anchor (**Figure 5**). Based on the crystal structure of leukotriene A_4 hydrolase, the two histidine residues and the distant glutamic acid residue of the gluzincin motif are zinc ligands, the glutamic acid residue between the histidine residues is involved in catalysis (Hooper, 1994; Rawlings and Barrett, 1995), and the glutamic acid residue of the GAMEN motif binds the substrate N-terminal amino acid (Luciani *et al.*, 1998). In contrast to the situation in mammals, insect aminopeptidase N is membrane bound at the C-terminal. No soluble insect aminopeptidase N has been sequenced.

Dendrograms derived from alignments of coleopteran and lepidopteran midgut aminopeptidases suggest that there are at least four groups of lepidopteran aminopeptidases, with the isoforms of the same animal distributed among the groups (Chang *et al.*, 1999; Emmerling *et al.*, 2001; Nakanishi *et al.*, 2002; Rajagopal *et al.*, 2003; P. T. Cristofoletti *et al.*, unpublished data) (**Figure 10b**, clusters A, B, C, and D). The existence of a number of different aminopeptidases in lepidopterans could be explained by the need for enzymes with different substrate specificities (as shown above for *R. americana*) or different susceptibilities to inhibitors, similar to serine proteinases (see Section 4.5.5.2).

Probably associated with the fact that aminopeptidases are major proteins in some microvillar membranes (55% of *T. molitor* midgut microvillar proteins) (Cristofoletti and Terra, 1999), they are targets of insecticidal *Bacillus thuringiensis* crystal δ-endotoxins. These toxins, after binding to aminopeptidases and receptor molecules called cadherins, form channels through which cell contents leak leading to death of the insect (Gill *et al.*, 1995; Knight *et al.*, 1995; Denolf *et al.*, 1997). Although data on substrate specificity for lepidopteran aminopeptidase isoforms are lacking, there is evidence that the isoforms may have differences in toxin binding (Valaitis *et al.*, 1997; Zhu *et al.*, 2000; Nakanishi *et al.*, 2002; Rajagopal *et al.*, 2003).

Cloning and sequencing dipteran aminopeptidases, for which differences in substrate specificity are known, and a study of substrate specificities of lepidopteran aminopeptidases, may clarify the selective advantages of the evolution of aminopeptidase groups. Furthermore, this study may support the hypothesis that aminopeptidase gene duplications have occurred in the panorpoid ancestor, before differentiation between dipterans and lepidopterans.

4.5.5.6. Carboxypeptidases and Dipeptidases

Carboxypeptidases hydrolyze single amino acids from the C-terminus of the peptide chain and are divided into classes on the basis of their catalytic mechanism. There are two digestive metallocarboxypeptidases in mammals: carboxypeptidase A (EC 3.4.17.1), which hydrolyzes, in alkaline medium, C-terminal amino acids, except arginine, lysine, and proline, and carboxypeptidase B (EC 3.4.1.7.2), which releases, in alkaline conditions, C-terminal lysine and arginine preferentially. Insect digestive carboxypeptidases have been classified as carboxypeptidase A or B depending on activity in alkaline medium against Z-GF (or hippuryl β-phenyllactic acid) or Z-GR (or hippuryl-L-arginine), respectively (**Figure 7**). Digestive insect carboxypeptidase A-like enzymes are widespread among insects and most of them have pH otima of 7.5–9.0 and molecular masses of 20–50 kDa (Terra and Ferreira, 1994). They have been cloned and sequenced from Diptera (Ramos *et al.*, 1993; Edwards *et al.*, 1997) and Lepidoptera (Bown *et al.*, 1998) and the enzyme from the lepidopteran *H. armigera* was also submitted to crystallographic studies (Estébanez-Perpiñá *et al.*, 2001). The sequences have signal and activation peptides and the features typical of carboxypeptidases A, including the residues His 69, Glu 72, and His 196, which bind the catalytic zinc ion, and Arg 71, Asn 144, Arg 145, and Tyr 248 responsible for substrate binding and Arg 127 and Glu 270 for catalysis. In spite of the overall similarity of *H. armigera* procarboxypeptidase with human procarboxypeptidase A2, there are differences in the loops between the conserved secondary structures, including the loop where the activation processing occurs.

Another important difference is the residue 255 (bottom of the S'_1 pocket) that defines the enzyme specificity. In mammalian sequences Asp 255 is found in carboxypeptidase B and Ile 255 in carboxypeptidase A. In insect carboxypeptidases A, this residue varies (but never is an acid residue) (**Figure 5**).

Carboxypeptidases B-like enzymes have been detected in insect midguts (Terra and Ferreira, 1994) but none has been characterized in detail, because they are much less active than carboxypeptidases A.

Dipeptidases hydrolyze dipeptides and are classified according to their substrate specificities. Dipeptidases comprise the poorest known of the insect peptide hydrolases. There have been few studies in which dipeptidase assays were performed and even fewer attempts to characterize the enzymes (Terra and Ferreira, 1994). The larval midgut of *R. americana* has three dipeptidases (two soluble with 63 kDa and 73 kDa, respectively, and one membrane-bound) that hydrolyze Gly-Leu, resembling dipeptide hydrolase (dipeptidase, E.C. 3.4.13.18), although in contrast to the mammalian enzyme they are very active upon Pro-Gly (**Figure 7**). *Rhynchosciara americana* also seems to have an amino acyl-histidine dipeptidase (carnosinase, EC 3.4.13.3) (Klinkowstrom *et al.*, 1995). More work on insect digestive dipeptidases is urgently needed.

4.5.6. Digestion of Lipids and Phosphates

4.5.6.1. Overview

Lipids that contain fatty acids comprise storage lipids and membrane lipids. Storage lipids, such as oils present in seeds and fats in adipose tissue of animals, are triacylglycerols (triglycerides) and are hydrolyzed by lipases. Membrane lipids include phospholipids and glycolipids like mono- and digalactosyldiglycerides (see Section 4.5.4.7). Phospholipids are digested by phospholipases. A combination of α- and β-galactosidases may remove galactose residues from mono- and digalactosyldiglyceride to leave a diacylglycerol which may be hydrolyzed by a triacylglycerol lipase.

Phosphate moieties need to be removed from phosphorylated compounds prior to absorption. This is accomplished by nonspecific phosphatases. The phosphatases may be active in an alkaline (alkaline phosphatase, EC 3.1.3.1) or acid (acid phosphates, EC 3.1.3.2) medium.

4.5.6.2. Lipases

Triacylglycerol lipases (EC 3.1.1.3) are enzymes that preferentially hydrolyze the outer links of triacylglycerols and act only on the water–lipid interface. Activity of the lipase is increased as the interface becomes larger due to lipid emulsification caused by emulsifiers (surfactants). Insects lack emulsifiers comparable to the bile salts of vertebrates, but surfactant phospholipids, including lysolecithin, occur in their midguts in sufficient concentration to alter the surface tension of midgut contents (De Veau and Schultz, 1992). Lysolecithin, and other surfactants, may be formed by the action of phospholipase A on ingested phospholipids (see below and **Figure 1**).

Current research on insect lipases is focused on triacylglycerol lipase from the fat body and its interplay with flight muscles (Ryan and Van der Horst, 2000). Insect midgut triacylglycerol lipases have been studied in few insects and only in crude preparations. The data suggested that the enzyme preferentially releases fatty acids from the α-positions, prefers unsaturated fatty acids and is activated by calcium ions, thus resembling the action of mammalian pancreatic lipase. The resulting 2-monoacylglycerol may be absorbed or hydrolyzed (Terra *et al.*, 1996). Hydrolysis of 2-monoacylglycerol may be accomplished by the triacylglycerol lipase, following migration of the fatty acid to the 1-position, which seems to be favored by the alkaline midgut pH, at least in *M. sexta* (Tsuchida and Wells, 1988).

Esterases, which are usually named the carboxylesterases (ali-esterases, EC 3.1.1.1) catalyze the hydrolysis of carboxyl ester into alcohol and carboxylate. This enzyme, in contrast to lipases, attacks molecules that are completely dissolved in water. It also hydrolyzes water-insoluble long-chain fatty acid esters in the presence of surfactants, but at a rate much slower than that of triacylglycerol lipase. A role for esterases in digestion is unclear and because of this they are not reviewed in detail here.

In spite of the fact that a requirement for essential fatty acids is probably universal in insects, progress has been limited in the study of lipid digestion. Presumably, the lack of comparatively simple, sensitive assays and the complexities of digestion related to lipid solubilization have hindered work in this area. Another reason to study enzymes associated with lipid digestion is that they might be important in limiting the development of pathogens and parasites. Hydrolysis of membrane lipids might cause cellular lysis and fatty acid products of digestion may possess antibiotic effects.

4.5.6.3. Phospholipases

Phospholipase A_2 (EC 3.1.1.4) and phospholipase A_1 (EC 3.1.1.32) remove from phosphatides the fatty acid attached to the 2-position and 1-position,

respectively, resulting in a lysophosphatide (**Figure 2**). Lysophosphatide is more stable in micellar aggregates than on membranes. Thus, the action of phospholipase A on the membrane phosphatides causes the solubilization of cell membranes, rendering the cell contents free to be acted upon by the appropriate digestive enzymes. Phospholipase is widespread among insects (Terra *et al*., 1996; Nor Aliza *et al*., 1999). Phospholipase A_2 partially purified from the midgut of adult beetle *Cincindella circumpicta* has a molecular mass of 22 kDa and pH optimum 9.0, is calcium dependent, and is inhibited by the site-specific inhibitor oleyoxyethyl phosphorylcholine. Unfed beetles did not express the phospholipase in the midgut contents (Uscian *et al*., 1995). Although lysophosphatide may be further hydrolyzed by a lysophospholipase (phospholipase B, EC 3.1.1.5), evidence suggests it is absorbed intact by insects (Turunen and Kastari, 1979; Weiher and Komnick, 1997). Phosphatides may also be hydrolyzed by phospholipase C (EC 3.1.4.3) yielding the phosphoryl base moiety and diacylglycerol, or by phospholipase D (EC 3.1.4.4), resulting phosphatidate and the base (**Figure 2**). Both enzymes have been found in insect midgut (Turunen, 1993), but have not studied in detail.

4.5.6.4. Phosphatases

Alkaline phosphatase is usually a midgut microvillar membrane marker in dipteran and lepidopteran species, although it may also occur in midgut basolateral membranes and even as a secretory enzyme. Acid phosphatase is usually soluble in the cytosol of midgut cells in many insects and may also appear in midgut contents or be found membrane-bound in midgut cells (Terra and Ferreira, 1994).

The best-known alkaline phosphatases are those from *B. mori* (Lepidoptera: Bombycidae) larval midgut. The major membrane-bound and the minor soluble alkaline phosphatases were purified and shown to be monomeric enzymes with the following properties: (1) soluble enzyme, molecular mass of 61 kDa, pH optimum 9.8; (2) membrane-bound enzyme, molecular mass of 58 kDa, pH optimum 10.9. Both enzymes have wide substrate specificity and are inhibited by cysteine. The membrane-bound alkaline phosphatase occurs in the microvillar membranes of columnar cells, whereas the soluble enzyme is loosely attached to the goblet cell apical membrane facing the cell cavity (Eguchi, 1995). The determination of the complete sequence of the membrane-bound alkaline phosphatase led to the finding of putative regions for phosphatidylinositol anchoring, zinc-binding site but not for *N*-glycosylation, despite the fact that the enzyme contains *N*-linked oligosaccharides (Itoh *et al*., 1991). The sequence of the soluble alkaline phosphatase was also determined and has high identity with the membrane-bound enzyme (Itoh *et al*., 1999).

Acid phosphatases have been characterized in some detail only in *Rhodnius prolixus* (Hemiptera: Reduvidae). The major enzyme activity is soluble and has the following properties: wide specificity, a molecular mass of 82 kDa, Km for *p*-nitrophenyl phosphate 0.7 mM, and is inhibited by fluoride, tartrate, and molybdate. The minor enzyme activity is membrane-bound and is resolved into two enzymes (123 and 164 kDa) which are resistant to fluoride and tartrate (Terra *et al*., 1988).

4.5.7. The Peritrophic Membrane

4.5.7.1. The Origin, Structure, and Formation of the Peritrophic Membrane

There is a film surrounding the food bolus in most insects that occasionally is fluid (peritrophic gel) but more frequently is membranous (peritrophic membrane, PM). The PM is made up of a matrix of proteins (peritrophins) and chitin to which other components (e.g., enzymes, food molecules) may associate (see **Chapter 4.3**). This anatomical structure is sometimes called peritrophic matrix, but this term should be avoided because it does not convey the idea of a film and suggests it is the fundamental substance of some structure, usually filling a space as the mitochondrial matrix. The argument that membrane means a lipid bilayer is not valid because the PM is not a cell part, but an anatomical structure, like the nictitating membrane of birds and reptiles.

Peritrophins, the integral PM proteins, are made of several domains. The major domain (peritrophin A-domain), is a cysteine-rich domain with chitin-binding properties having the consensus sequence: $CX_{13–20}CX_{5–6}CX_{9–19}CX_{10–14}CX_{4–14}C$ (where X is any amino acid except cysteine) that includes several conserved aromatic amino acids (see Section 4.5.4.5). Variations of this chitin-binding domain are peritrophin-B and peritrophin-C domains with consensus sequences:

$CX_{12–13}CX_{20–21}CX_{10}CX_{12}CX_2CX_8CX_{7–12}C$ and $CX_{8–9}CX_{17–21}CX_{10–11}CX_{12–13}CX_{11}C$, respectively. Another kind of domain occurring in peritrophins are proline/threonine-rich domains that are heavily glycosylated and similar to mucins. Peritrophins may have one (e.g., Cb-peritrophin-15 from *Lucilia cuprina*) to several (e.g., peritrophin-44 from

L. cuprina) chitin-binding domains or chitin-binding domains with small (e.g., Ag-AperI from *Anopheles gambiae*) or very large mucin-like domains (e.g., IIM from *Trichoplusia ni*) (Wang and Granados, 1997; Shen and Jacobs-Lorena, 1998; Tellam *et al.*, 1999, 2003).

The 3D structure of PM is thought to result from chitin fibrils being interlocked with the chitin-binding domains of peritrophins. Mucin-like domains of peritrophins are thought to face the ectoperitrophic and endoperitrophic sides of the PM. As these domains are highly hydrated they lubricate the surface of the PM, easing the movement of food inside the PM and of the ectoperitrophic fluid outside the PM. Furthermore, the glucan chains associated with peritrophin mucin-like domains may assure high proteinase resistance to PM (see **Figure 9** in Schorderet *et al.*, 1998; **Figure 5** in Wang and Granados, 2001).

The structure of peritrophins prompted Terra (2001) to develop a speculative model of the origin and evolution of the PM. According to this model, ancestral insects had their midgut cells covered with a mucus similar to that found in most animals. This gastrointestinal mucus, at least in vertebrates, is a gel-like substance composed of mucins (Allen, 1983; Forstner and Forstner, 1986). It was proposed that evolutionary processes led to the development of the PM from the gastrointestinal mucus. According to this hypothesis, the peritrophins, the major PM proteins, evolved from mucins by acquiring chitin–binding domains. The concomitant evolution of chitin secretion by midgut cells permitted the formation of the chitin–protein network characteristic of PM structure described above. Later on in evolution, some peritrophins lost their mucin-like domains. If the hypothesis that the PM is derived from the gastrointestinal mucus is correct, it should have originally been synthesized by midgut cells along the whole midgut and should have had the properties of the mucus. Later in evolution, insect species would have appeared with a chitin–protein network resulting in PM formation. Therefore, the formation of the PM by the whole midgut epithelium is the ancestral condition, whereas the restriction of PM production to midgut sections, or the lack of a PM and its replacement by the peritrophic gel, are derived conditions.

PMs are classified into two types (Peters, 1992). Type I PM is found in cockroaches (Dictyoptera), grasshoppers (Orthoptera), beetles (Coleoptera), bees, wasps, and ants (Hymenoptera), moths and butterflies (Lepidoptera), and in hematophagous adult mosquitoes (Diptera). Type II PM occurs in larval and adult (except hematophagous ones) mosquitoes and flies (Diptera), and in a few adult Lepidoptera.

Type I PM is formed either by the whole midgut epithelium, or by part of it (anterior or posterior regions). PM produced by the whole or anterior midgut epithelium envelops the food along the whole midgut. When PM is produced only by the posterior part, the anterior midgut epithelium is usually covered with a viscous material, the peritrophic gel, as observed in carabid beetles (Ferreira and Terra, 1989) and bees (Jimenez and Gilliam, 1990). This gel is also observed in the anteriorly placed midgut caeca of some insects and along the whole midgut of others (Terra, 2001).

It has been shown with light microscopy, as well as transmission and scanning electron microscopy, that during the formation of type I PM chitin-containing fibrous material appears first at the tips of the microvilli of anterior midgut cells and then is rapidly included into a thin PM surrounding the food bolus (Harper and Hopkins, 1997). Chitin is synthesized outside the cells by a chitin synthase bound to microvillar membranes using precursors formed inside the cells (see **Chapter 4.3**). In agreement with this, antibodies raised against a translated product of a cDNA fragment coding chitin synthase immunolabeled only the apical ends of midgut microvilli (Zimoch and Merzendorfer, 2002). Chitin, after being self-organized into fibers, is interlocked by peritrophins. These are released by microapocrine secretion (Bolognesi *et al.*, 2001). The formation of these PMs is frequently induced by the distension of the gut caused by food ingestion.

Type II PM is secreted by a few rows of cells at the entrance of the midgut (cardia) and usually is found in insects irrespective of food ingestion. Peritrophins are secreted by exocytosis (Eisemann *et al.*, 2001).

Although a PM is found in most insects, it does not occur in Hemiptera and Thysanoptera, which have perimicrovillar membranes in their cells (see below). The other insects that apparently do not have a PM are adult Lepidoptera, Phthiraptera, Psocoptera, Zoraptera, Strepsiptera, Raphidioptera, Megaloptera, adult Siphonaptera, bruchid beetles, and some adult ants (Hymenoptera) (Peters, 1992). These insects may have a peritrophic gel instead of PM, one example being bruchid beetles (Terra, 2001), or may have had their PMs overlooked, because the insects were unfed.

Another possibility is that minute hematophagous insects (like Siphonaptera and Phthiraptera) have lost their PM because the blood clot assures countercurrent flows (see Section 4.5.7.2.2.3) and their

small size makes easy the efficient diffusion of digestion products up to the midgut surface.

The PM may have a large range of pore sizes: some small or very large and most of them in the middle range. The average pore sizes of PM may be determined by comparing molecular masses of enzymes restricted to the ectoperitrophic fluid (**Figure 1**) with those of enzymes present inside PM. This method of pore size estimation is probably the most accurate one since it reflects *in vivo* conditions. Pore sizes have also been determined by feeding insects with colloidal gold particles or fluorescent dextran molecules of known molecular masses and recording passage through the PM *in vivo* or using PM mounted as a sac and measuring diffusing rates. Determinations performed with these techniques by different authors (Zhuzhikov, 1964; Terra and Ferreira, 1983; Espinoza-Fuentes *et al*., 1984; Peters and Wiese, 1986; Santos and Terra, 1986; Wolfersberger *et al*., 1986; Miller and Lehane, 1990; Ferreira *et al*., 1994a) found pores in the range 7–9 nm for insects pertaining to different orders. Other authors described pores in the range 17–36 nm (Barbehenn and Martin, 1995; Edwards and Jacobs-Lorena, 2000). Pore sizes in the range 17–36 nm were obtained with fluorescent dextran molecules in conditions able to detect very small amounts of substances traversing the PM and, for this, they probably correspond to the large pores occurring at low frequency in PMs. Although these large pores are supposed to be of no importance regarding digestive events, they set the size limits that an infecting particle must have to successfully pass through the PM.

As a consequences of its small pores, the PM hinders the free movement of molecules, dividing the midgut lumen into two compartments (**Figure 1**) with different molecules. The functions of this structure include those of the mucus (protection against food abrasion and invasion by microorganisms) and several roles associated with the compartmentalization of the midgut. These roles result in improvements in digestive physiology efficiency thereby leading to decreased digestive enzyme excretion, and restrict the production of the final products of digestion close to their transporters, thus facilitating absorption. These roles will be detailed below (see Section 4.5.7.2).

In the light of current research, the distinction between PM and peritrophic gel may be more important than the traditional distinction between type I and type II PMs, in spite of the fact that the two types may have somewhat different structures (Tellam *et al*., 1999). Other major points needing clarification are how the chemical nature of peritrophic gel and PM define their strength, elasticity and porosity and how these structures are self-assembled in the midgut lumen.

4.5.7.2. The Physiological Role of the Peritrophic Membrane

4.5.7.2.1. Protection against food abrasion and invasion by microorganisms As mentioned before, gut cells in most animals are covered with a gel-like coating of mucus, which has been most thoroughly studied in mammals (Forstner and Forstner, 1986). In these animals, the mucus is supposed to lubricate the mucosa, protecting it from mechanical damage, and to trap bacteria and parasites. Since the insect midgut epithelium lacks a mucus coating, PM functions were supposed to be analogous to that of mucus. Thus, insects deprived of PM may have the midgut cells damaged by coarse food and may be liable to microorganism invasion in some reported cases (Peters, 1992; Tellam, 1996; Lehane, 1997).

The PM as a barrier against invasion by microorganisms has particular relevance in insects that transmit viruses and parasites to human beings, as these microorganisms may have specific developmental phases in insect tissues (Tellam, 1996; Lehane, 1997). Microorganisms invade the insect midgut cells after disrupting the PM with the use of chitinase (Shahabuddin, 1995) or by using a proteinase such as enhancin that affects specifically the peritrophins (Peng *et al*., 1999; Ivanova *et al*., 2003).

A barrier against microorganism invasion is probably less important for the majority of insects that feed on plants, as exemplified by observations carried out with the moth *T. ni*. Larvae of this insect deprived of PM by Calcofluor treatment show high mortality. Examination of dead larvae showed no signs of microbial infection or cell damage by Calcofluor, although these larvae were more susceptible to experimental infection (Wang and Granados, 2000). The results may be interpreted as Calcofluor killing larvae by affecting PM functions in digestion.

In the same direction goes the observation that some plants respond to herbivorous insect attack by producing a unique 33 kDa cysteine proteinase with chitin-binding activity. This proteinase damages the PM, resulting in significant reduction in caterpillar growth caused by impaired nutrient utilization (Pechan *et al*., 2002).

4.5.7.2.2. Enhancing digestive efficiency

4.5.7.2.2.1. Overview The proposal of roles for the PM in digestion has benefited from studies on the organization of the digestive process. These studies (reviews: Terra, 1990; Terra and Ferreira,

1994, 2003) revealed that in most insects initial digestion occurs in the endoperitrophic space (Figure 1), intermediate digestion in the ectoperitrophic space, and final digestion at the surface of midgut cells. Such studies led to the formulation of the hypothesis of the endo–ectoperitrophic circulation of digestive enzymes. It was suggested that there is a recycling mechanism (Figure 11), where food flows inside the PM from the anterior midgut to the posterior, whereas in the ectoperitrophic space water flows from the posterior midgut to the caeca. When the polymeric food molecules become sufficiently

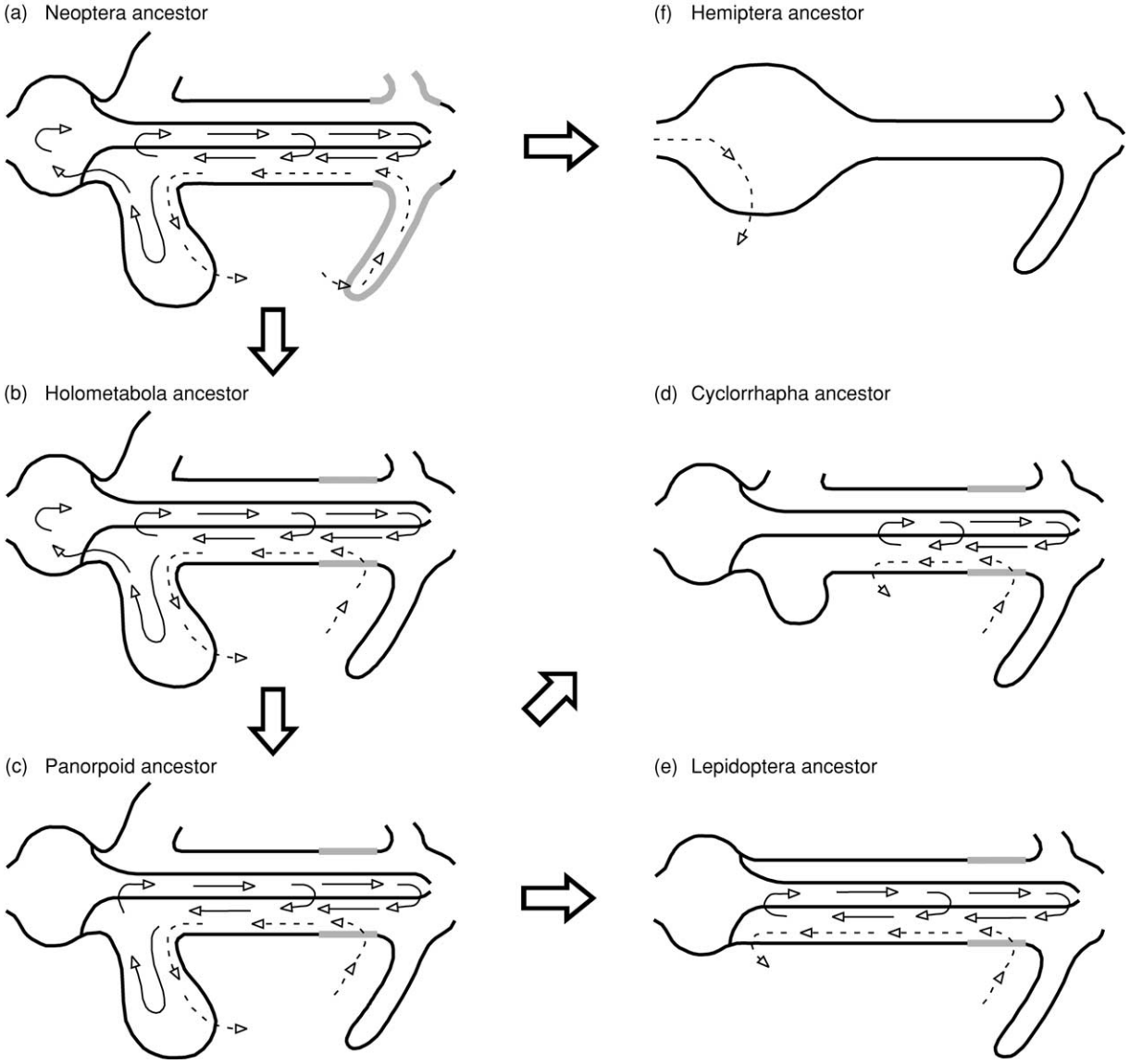


Figure 11 Diagrammatic representation of water fluxes (dotted arrows) and of the circulation of digestive enzymes (solid arrows) in putative insect ancestors that correspond to the major basic gut plans. In Neoptera ancestors (a), midgut digestive enzymes pass into the crop. Countercurrent fluxes depend on the secretion of fluid by the Malpighian tubules and its absorption by the caeca. Enzymes involved in initial, intermediate, and final digestion circulate freely among gut compartments. Holometabola ancestors (b) are similar except that secretion of fluid occurs in posterior ventriculus. The ancestors of hymenopteran and panorpoid (Lepidoptera and Diptera assemblage) insects (c) display countercurrent fluxes like Holometabola ancestors, midgut enzymes are not found in the crop, and only the enzymes involved in initial digestion pass through the peritrophic membrane. Enzymes involved in intermediate digestion are restricted to the ectoperitrophic space and those responsible for terminal digestion are immobilized at the surface of midgut cells. Cyclorrhapha ancestors (d) have a reduction in caeca, absorption of fluid in middle midgut, and anterior midgut playing a storage role. Lepidoptera ancestors (e) are similar to panorpoid ancestors, except that the anterior midgut replaces the caeca in fluid absorption. Hemiptera ancestors (f) have lost crop, caeca, and fluid-secreting regions. Fluid is absorbed in anterior midgut. (Reprinted with permission from Terra, W.R., Ferreira, C., **2003**. Digestive system. In: Resh, V.H., Cardé, R.T. (Eds.), Encyclopedia of Insects. Academic Press, San Diego, CA, pp. 313–323; © Elsevier.)

small to pass through the PM (with the accompanying polymer hydrolases) the flow patterns result in the carriage towards the caeca or the anterior midgut where intermediate and final digestion occurs.

Terra *et al*. (reviews: Terra and Ferreira, 1994; Terra, 2001) hypothesized that as a consequence of the compartmentalization of digestive events, there is an increase in the efficiency of digestion of polymeric food by allowing the removal of the oligomeric molecules from the endoperitrophic space, which is powered by the recycling mechanism associated with the midgut fluxes. Because oligomers may be substrates or inhibitors for some polymer hydrolases, their presence should decrease the rate of polymer degradation. Fast polymer degradation ensures that polymers are not excreted and hence increases their digestibility. Another possible consequence of compartmentalization is an increase in the efficiency of oligomeric food hydrolysis due to the transference of oligomeric molecules to the ectoperitrophic space and restriction of oligomer hydrolases to this compartment. In these conditions, oligomer hydrolysis occurs in the absence of probable partial inhibition (because of nonproductive binding) by polymer food and presumed nonspecific binding by nondispersed undigested food. This process should lead to the production of food monomers in the vicinity of midgut cell surface, causing an increase in the concentration of the final products of digestion close to their transporters, thus facilitating absorption. Experimental evidence supporting the adaptations for increasing digestive efficiency proposals are discussed in the following sections.

4.5.7.2.2.2. Prevention of nonspecific binding A model system was used to test the hypothesis that the PM prevents nonspecific binding of undigested material onto midgut cell surface (R. Bolognesi, W.R. Terra, and C. Ferreira, unpublished data). Aminopeptidase-containing microvillar membranes from *Spodoptera frugiperda* midgut cells were purified and assayed in the absence and presence of a concentration of midgut contents resembling *in vivo* conditions. The activity of the aminopeptidase in the presence of the midgut contents was about 50% of the activity in their absence, favoring the idea that undigested material in contact with microvillar enzymes negatively affects their activity.

4.5.7.2.2.3. Prevention of enzyme excretion This function was at first proposed based on results obtained with dipteran larvae (reviews: Terra, 1990; Terra and Ferreira, 1994). Both *R. americana* and *Musca domestica* present a decreasing trypsin gradient along midgut contents (putatively generated by the recycling mechanism) and excreted less than 15% of midgut luminal trypsin after each gut emptying. When the larvae were fed a diet with excess protein, the trypsin gradient along midgut contents becomes less discernible and trypsin excretion increases to 40%. This is exactly what would be expected if the recycling mechanism existed and an increase in undigested dietary protein prevents trypsin from diffusing into the ectoperitrophic space and moving into anterior midgut by the countercurrent flux of fluid. Subsequently, dye experiments showed the existence of the appropriate fluid fluxes.

More recently, experimental evidence that a recycling mechanism also occurs in Lepidoptera and Coleoptera was described. As predicted by the model, most trypsin activity is found in the lumen of *Manduca sexta* (Lepidoptera, type I PM) anterior midgut, whereas trypsin mRNA predominates in middle midgut (Peterson *et al*., 1994). Furthermore, immunocytochemical data showed the occurrence in the anterior midgut of significant amounts of a 41 kDa protein and a β-glycosidase secreted by the posterior midgut of *M. sexta* (Borhegyi *et al*., 1999) and middle midgut in *Tenebrio molitor* (Coleoptera, type I PM) (Ferreira *et al*., 2002), respectively. Finally, the decreasing trypsin and chymotrysin gradient along *S. frugiperda* midgut contents disappeared in Calcofluor-treated larvae lacking a peritrophic membrane (Bolognesi *et al*., 2001) and the excretory rate increased from 0.1% to 0.9% of midgut contents at each gut emptying (R. Bolognesi, W. R. Terra, and C. Ferreira, unpublished data).

4.5.7.2.2.4. Increase in the efficiency of digestion of polymeric food A model system was used to test this hypothesis (R. Bolognesi, W.R. Terra, and C. Ferreira, unpublished data). Midgut contents from *S. frugiperda* larvae were placed into dialysis bags suspended in stirred and unstirred media. Trypsin activities in stirred and unstirred bags were 210% and 160%, respectively, over the activities of similar samples maintained in a test tube. The results suggested that the diffusion of products from the trypsin reaction media favors enzyme action and that stirring (an *in vitro* model of the ectoperitrophic countercurrent flux) enhances the effect.

4.5.7.2.2.5. Increase in the efficiency of oligomeric food hydrolysis This putative function is supported by the experiments of Bolognesi *et al*. (R. Bolognesi, W.R. Terra, and C. Ferreira, unpublished data). They collected ectoperitrophic fluid from the large midgut caeca of *R. americana*. Aminopeptidase A, *N*-acetylglucosaminidase, and carboxypeptidase A are enzymes restricted to the

ectoperitrophic space. When those enzymes were put in the presence of PM contents their activities decreased in relation to controls as follows: aminopeptidase A, 46%; *N*-acetylglucosaminidase, 56%; carboxypeptidase A, 92%. These decreases in activity probably result from oligomer hydrolase competitive inhibition by luminal polymers.

4.5.7.2.2.6. Restriction of food monomer production at cell surface This is a consequence of restricting oligomer hydrolases to the ectoperitrophic space (see Section 4.5.7.2.2.5) and causes an increase in the concentration of the final products of digestion close to the carriers responsible for their absorption. A model system should be developed to test this hypothesis.

4.5.7.2.2.7. Enzyme immobilization Midgut luminal enzymes, in addition to occurring in the endoperitrophic and ectoperitrophic spaces, may be associated with the PM. For example, results obtained with *S. frugiperda* larvae showed that PM may contain up to 13% and 18% of the midgut luminal activity of amylase and trypsin, respectively (Ferreira *et al*., 1994b). Hence, enzyme immobilization may play a role in digestion, although a minor one. The attachment mechanism of enzymes in PM is not well known. Nevertheless, there is evidence, at least in *S. frugiperda*, that trypsin, amylase, and microvillar enzymes are incorporated into the jelly-like substance associated with PM when the enzymes, still bound to membranes, are released from midgut cells by a microaprocrine process (Jordão *et al*., 1999; Bolognesi *et al*., 2001).

4.5.7.2.2.8. Toxin binding Although potentially toxic dietary tannins are attached to and excreted with *Schistocerca gregaria* PM (Bernays and Chamberlain, 1980), toxin binding by the PM seems to be a less widespread phenomenon than previously suggested. Thus, tannins in *M. sexta* (Barbehenn and Martin, 1998) and lipophilic and amphiphilic noxious substances in *Melanoplus sanguinipes* (Barbehenn, 1999) are maintained in the endoperitrophic space because they form high molecular weight complexes, not because of PM binding.

4.5.7.2.2.9. Peritrophic membrane functions and insect phylogeny Current data detailed below suggest that PMs of all insects have functions (see Sections 4.5.7.2.1, 4.5.7.2.2.2–4.5.7.2.2.4), whereas functions (see Sections 4.5.7.2.2.5 and 4.5.7.2.2.6) are demonstrable only in PMs of Panorpodea (the taxon that includes Diptera and Lepidoptera) and of the hymenopteran sawflies. PM function (see Section 4.5.7.2.2.7) may occur in all insects but this needs further confirmation. Function (see Section 4.5.7.2.2.8), although it may be important for some insects, should be viewed as opportunistic. In other words, the PM probably evolved from a protective role (see Section 4.5.7.2.1) to more sophisticated functions (see Sections 4.5.7.2.2.2–4.5.7.2.2.7) under selective pressures and, due to the chemical properties of their constituents, the PM also developed the ability to bind different compounds including toxins.

4.5.8. Organization of the Digestive Process

4.5.8.1. Evolutionary Trends of Insect Digestive Systems

After studying the spatial organization of the digestive events in insects of different taxa and diets, it was realized that insects may be grouped relative to their digestive physiology, assuming they have common ancestors.Those putative ancestors correspond to basic gut plans from which groups of insects may have evolved by adapting to different diets (Terra and Ferreira, 1994, 2003).

The basic plan of digestive physiology for most winged insects (Neoptera ancestors) is summarized in **Figure 11**. In these ancestors, the major part of digestion is carried out in the crop by digestive enzymes propelled by antiperistalsis forward from the midgut. Saliva plays a variable role in carbohydrate digestion. After a while following ingestion, the crop contracts transferring digestive enzymes and partly digested food into the ventriculus. The anterior ventriculus is acid and has high carbohydrase activity, whereas the posterior ventriculus is alkaline and has high proteinase activity. This differentiation along the midgut may be an adaptation to instability of ancestral carbohydrases in the presence of proteinases. The food bolus moves backward in the midgut of the insect by peristalsis. As soon as the polymeric food molecules are digested to become sufficiently small to pass through the peritrophic membrane, they diffuse with the digestive enzymes into the ectoperitrophic space (**Figure 1**). The enzymes and nutrients are then displaced toward the caeca with a countercurrent flux caused by secretion of fluid at the Malpighian tubules and its absorption back by cells at the caeca (**Figure 11**), where final digestion is completed and nutrient absorption occurs. When the insect starts a new meal, the caeca contents are moved into the crop. As a consequence of the countercurrent flux,

digestive enzymes occur as a decreasing gradient in the midgut and their excretion is lowered.

The Neoptera basic plan gave origin to that of the Polyneoptera orders, which include Blattodea, Isoptera, and Orthoptera, and evolved to the basic plans of Paraneoptera and Holometabola. The characteristics of the Paraneoptera ancestors cannot be inferred because midgut function data are available only for Hemiptera.

The basic gut plan of the Holometabola (**Figure 11b**) (which include Coleoptera, Megaloptera, Hymenoptera, Diptera, and Lepidoptera) is similar to that of Neoptera, except that fluid secretion occurs by the posterior ventriculus, instead of by the Malpighian tubules. Because the posterior midgut fluid does not contain wastes, as is the case for Malpighian tubular fluid, the accumulation of wastes in caeca is decreased. Caeca loss probably further decreases the accumulation of noxious substances in the midgut, which would be more serious in insects that have high relative food consumption rates, as is common among Holometabola.

The basic plan of Coleoptera did not evolve dramatically from the holometabolan ancestor, whereas the basic plan of Hymenoptera, Diptera, and Lepidoptera ancestor (hymenopteran–panorpoid ancestor, **Figure 11c**) presents important differences. Thus, hymenopteran–panorpoid ancestors have countercurrent fluxes like holometabolan ancestors, but differ from these in the lack of crop digestion, midgut differentiation in luminal pH, and in which compartment is responsible for each phase of digestion. In holometabolan ancestors, all phases of digestion occur in the endoperitrophic space (**Figure 1**), whereas in hymenopteran–panorpoid ancestors only initial digestion occurs in that region. In the latter ancestors, intermediate digestion is carried out by free enzymes in the ectoperitrophic space and final digestion occurs at the midgut cell surface by immobilized enzymes. The free digestive enzymes do not pass through the PM because they are larger than the PM's pores. As a consequence of the compartmentalization of digestive events in hymenopteran and panorpoid insects, there is an increase in the efficiency of digestion of polymeric food as discussed before.

The evolution of insect digestive systems summarized above and in **Figure 11** was proposed, as discussed before, from studies carried out in 12 species pertaining to four insect orders. To give further support for the hypothesis that the characteristics of gut function and morphology depend more on phylogeny than on diet, another approach was used. A total of 29 gut morphology and digestive physiology characteristics (e.g., luminal pH, ratio of gut

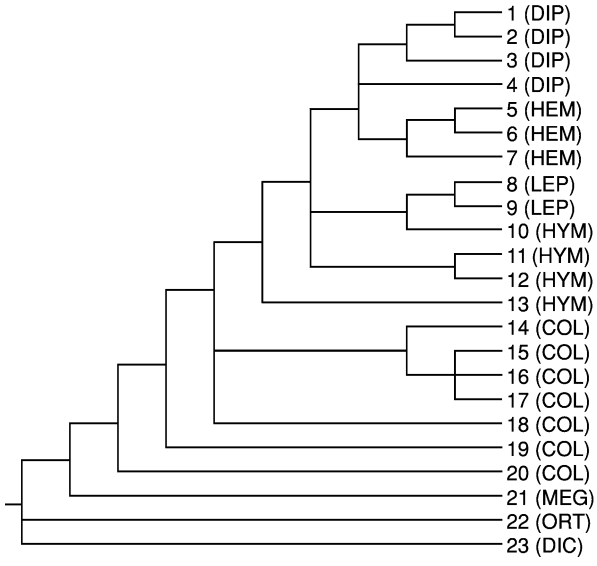


Figure 12 Cladogram of representative insects based on 29 gut morphology and digestive physiology characteristics. Insects: 1, *Trichosia pubescens*; 2, *Rhynchosciara americana*; 3, *Musca domestica*; 4, *Anopheles* spp.; 5, *Rhodnius prolixus*; 6, *Dysdercus peruvianus*; 7, *Acyrthosiphon pisum*; 8, *Erinnyis ello*; 9, *Spodoptera frugiperda*; 10, *Themos malaisei*; 11, *Camponotus rufipes*; 12, *Scaptotrigona bipunctata*; 13, *Bracon hebetor*; 14, *Tenebrio molitor*; 15, *Migdolus fryanus*; 16, *Sphenophorus levis*; 17, *Cyrtomon solana*; 18, *Dermestes maculatus*; 19, *Pyrearinus termitilluminans*; 20, *Pheropsophus aequinoctialis*; 21, *Corydalus* sp.; 22, *Abracris flavolineata*; 23, *Periplaneta americana*. (Courtesy of A.B. Dias.)

section volumes, type of peritrophic membrane, presence of special gut cells, distribution of digestive enzymes along the gut, major proteinase) were identified in 23 species from eight different insect orders. Making use of these characteristics, a cladogram was constructed putting together the data from studied species (**Figure 12**). The data confirmed that the morphological and functional traits associated with the digestive system are more dependent on taxon than on dietary habits of the different insects (Dias, Vanin, Marques, and Terra, unpublished data). There are two insect species that do not apparently fit the model: *Anopheles* spp. and *Themos malaisei*. *Anopheles* spp. is an adult, whereas the other Diptera is larval. *Themos malaisei* is an unexpected finding that will be discussed below (see Section 4.5.8.2.8).

4.5.8.2. Digestion in the Major Insect Orders

4.5.8.2.1. Initial comments Applebaum (1985) in his review for the first edition of this series divided the insects according to their diet when describing their digestive processes. Recent data (see Section 4.5.8.1) support the view that functional digestive traits of insects are linked with their phylogenetic

position. The organization of the digestive process in the different insect orders have been reviewed several times (Terra, 1988, 1990; Terra and Ferreira, 1994, 2003). The following section is therefore an abridged version of those texts, highlighting new findings and trying to identify points that deserve more research, especially in relation to molecular aspects. Only key references before 1994 are cited and the reader should find more references in the above-mentioned reviews.

4.5.8.2.2. Blattodea Cockroaches are usually omnivorous. It is thought that digestion in cockroaches occurs as described for the Neoptera ancestor (**Figure 11a**), except that part of the final digestion of proteins occurs on the surface of midgut cells (Terra and Ferreira, 1994). This was confirmed by the finding in *P. americana* that most trypsin, maltase, and amylase are found in the crop, whereas aminopeptidase predominates in the microvillar membranes of posterior midgut. There is a decreasing gradient of trypsin, maltase, and amylase along the midgut contents and less than 5% of trypsin and maltase (amylase, 27%) are excreted during each midgut emptying. This suggests the existence of midgut digestive enzyme recycling, with amylase excretion increased probably due to excess dietary starch. The recycling mechanism is thought to be powered by water fluxes as in the Neoptera ancestor, although there are no data supporting this. Major digestive proteinases are trypsin and chymotrypsin (Dias and Terra, unpublished data).

The differentiation of pH along the midgut (acid anterior midgut and alkaline posterior midgut) is not conserved among some cockroaches like *P. americana*, but it was maintained in others exemplified by the blaberid *Nauphoeta cinerea* (Elpidina *et al*., 2001a). The organization of digestion in this cockroach seems similar to that in *P. americana*, although data on enzyme excretion are lacking. At least blaberoid cockroaches possess proteinase inhibitory proteins active in the anterior midgut. These inhibitors are thought to be a primitive device to decrease the proteolytic inactivation of the animal's own carbohydrases, which are thus expected to be more active in the anterior midgut. The digestive carbohydrases from more evolved insects are stable in the presence of their own proteinases (Terra, 1988). Recently several proteinase inhibitors have been partially purified from *N. cinerea* (Elpidina *et al*., 2001b).

Another difference between cockroaches and the Neoptera ancestor is the enlargement of hindgut structures, noted mainly in wood-feeding cockroaches. These hindgut structures harbor bacteria producing acetate and butyrate from ingested wood or other cellulose-containing materials. Acetate and butyrate are absorbed by the hindgut of all cockroaches, but this is more remarkable with woodroaches (Terra and Ferreira, 1994). Cellulose digestion may be partly accomplished by bacteria in the hindgut of *P. americana* or protozoa in *Cryptocercus punctulatus* (Bignell, 1981). Nevertheless, now it is clear that *P. americana* saliva contains two cellulases and three laminarinases that may open plant cells and lyze fungal cells (Genta *et al*., 2003). This agrees with the omnivorous detritus feeding habit of the insect. The woodroach *Panestria cribrata* also has its own cellulase (Scrivener *et al*., 1989).

4.5.8.2.3. Isoptera Termites may be seen as insects derived from, and more adapted than, woodroaches in dealing with refractory material as wood and humus. Associated with this specialization, they lost the crop and midgut caeca, and enlarged their hindgut structures. Both lower and higher termites digest cellulose with their own cellulase, despite the occurrence of cellulose-producing protozoa in the paunch, an enlarged region of the anterior hindgut in lower termites. The products of cellulose digestion pass from the midgut into the hindgut, where they are converted into acetate and butyrate by hindgut bacteria as in woodroaches. Symbiotic bacteria are also responsible for nitrogen fixation in the hindgut, resulting in bacterial protein. This is incorporated into the termite body mass after being expelled in feces by one individual and being ingested and digested by another. This explains the ability of termites to develop successfully in diets very poor in protein. Both lower and higher feeding termites seem to have an endo–ectoperitrophic circulation of digestive enzymes (Terra and Ferreira, 1994; Nakashima *et al*., 2002; see also Section 4.5.4.3.1).

4.5.8.2.4. Orthoptera Grasshoppers feed mainly on grasses and their digestive physiology has clearly evolved from the Neoptera ancestor. Carbohydrate digestion occurs mainly in the crop, under the action of midgut enzymes, whereas protein digestion and final carbohydrate digestion take place at the anterior midgut caeca. The abundant saliva (devoid of significant enzymes) produced by grasshoppers saturate the absorbing sites in the midgut caeca, thus hindering the countercurrent flux of fluid. This probably avoids excessive accumulation of noxious wastes in the caeca, coming from Malpighian tubule secretion, and makes possible the high relative food consumption observed among locusts

in their migratory phases. Starving grasshoppers present midgut countercurrent fluxes. Cellulase found in some grasshoppers is believed to facilitate the access of digestive enzymes to the plant cells ingested by the insects by degrading the cellulose framework of cell walls (Dow, 1986; Terra and Ferreira, 1994; Marana *et al*., 1997).

4.5.8.2.5. Hemiptera The Hemiptera comprise insects of the major suborders Auchenorrhyncha (cicadas, spittlebugs, leafhoppers, and planthoppers) and Sternorrhyncha (aphids and white flies) that feed almost exclusively on plant sap, and Heteroptera (e.g., assassin bugs, plant bugs, stink bugs, and lygaeid bugs) that are adapted to different diets.

The ancestor of the entire order Hemiptera is supposed to have been a sap-sucker similar to present day Auchenorrhyncha. Sap-sucking Hemiptera may suck phloem or xylem sap. These food sources have very low contents of proteins (with the exception of few phloem saps; see below) and carbohydrate polymers and are relatively poor in free essential amino acids. In contrast to xylem sap, phloem sap is very rich in sucrose (Terra, 1990). Thus, except for dimer (sucrose) hydrolysis, no food digestion is usually necessary in sap-suckers. Upon adapting to dilute phloem and/or xylem sap, hemipteran ancestors would lose the enzymes involved in initial and intermediate digestion and lose the peritrophic membrane (**Figure 11f**). These changes are associated with the lack of luminal digestion.

The major problem facing a sap-sucking insect (specially on dilute phloem or xylem sap) is to absorb nutrients, such as essential amino acids, that are present in very low concentration in sap. Whichever mechanism is used, xylem feeders may absorb as much as 99% of dietary amino acids and carbohydrate (Andersen *et al*., 1989). Amino acids may be absorbed according to a hypothesized mechanism that depends on perimicrovillar membranes, which are membranes ensheathing the midgut microvilli with a dead end (**Figure 13**). A role in midgut amino acid absorption depends on the presence of amino acid–K^+ symports on the surface of the perimicrovillar membranes and of amino acid carriers and potassium pumps on the microvillar membranes. Although amino acid carriers have been found in the microvillar membranes of several insects (Wolfersberger, 2000), no attempts have been made to study the other postulated proteins. Thus, in spite of the model provided an explanation for the occurrence of these peculiar cell structures in Hemiptera, it is supported only by: (1) evidence that amino acids are absorbed with potassium ions in *Dysdercus peruvianus* (Silva and Terra, 1994); (2) occurrence of particles studying the cytoplasmic face of the midgut microvillar membranes of *D. peruvianus*. These might be ion pumps responsible for the putative potassium ion transport, like similar structures in several epithelia (Silva *et al*., 1995).

Another problem that deserves more attention regarding perimicrovillar membranes is their origin. Immunolocalization of the perimicrovillar enzyme marker, α-glucosidase, suggests that these membranes are formed when double membrane vesicles fuse their outer membranes with the microvillar membranes and their inner membranes with the perimicrovillar membranes. A double membrane Golgi cisterna (on budding) forms the double membrane vesicles (Silva *et al*., 1995).

Organic compounds in xylem sap need to be concentrated before they can be absorbed by the perimicrovillar system. This occurs in the filter chamber of Cicadoidea and Cercopoidea, which concentrates xylem sap tenfold. The filter chamber consists of a thin-walled, dilated anterior midgut in close contact with the posterior midgut and the proximal ends of the Malpighian tubules. This arrangement enables water to pass directly from the anterior midgut to the Malpighian tubules, concentrating food in midgut and eliminating excess water. The high permeability of the filter chamber membrane to water results from the occurrence of specific channels formed by proteins named aquaporins. These were characterized as membrane proteins with 15–26 kDa and were immunolocalized in the filter chamber of several xylem sap feeders (Le Cahérec *et al*., 1997).

Sternorrhyncha, as exemplified by aphids, may suck more or less continuously phloem sap of sucrose concentration up to 1.0 M and osmolarity up to three times that of the insect hemolymph. This results in a considerable hydrostatic pressure caused by the tendency of water to move from the hemolymph into midgut lumen. To withstand these high hydrostatic pressures, aphids have developed several adaptations. Midgut stretching resistance is helped by the existence of links between apical lamellae (replacing usual midgut cell microvilli) that become less conspicuous along the midgut. As a consequence of the links between the lamellae, the perimicrovillar membranes could no longer exist and were replaced by membranes seen associated with the tips of the lamellae, the modified perimicrovillar membranes (Ponsen, 1991; Cristofoletti *et al*., 2003). A modified perimicrovillar membrane-associated α-glucosidase frees fructose from sucrose without increasing the osmolarity by promoting transglycosylations. As the fructose is quickly absorbed, the osmolarity decreases, resulting in a

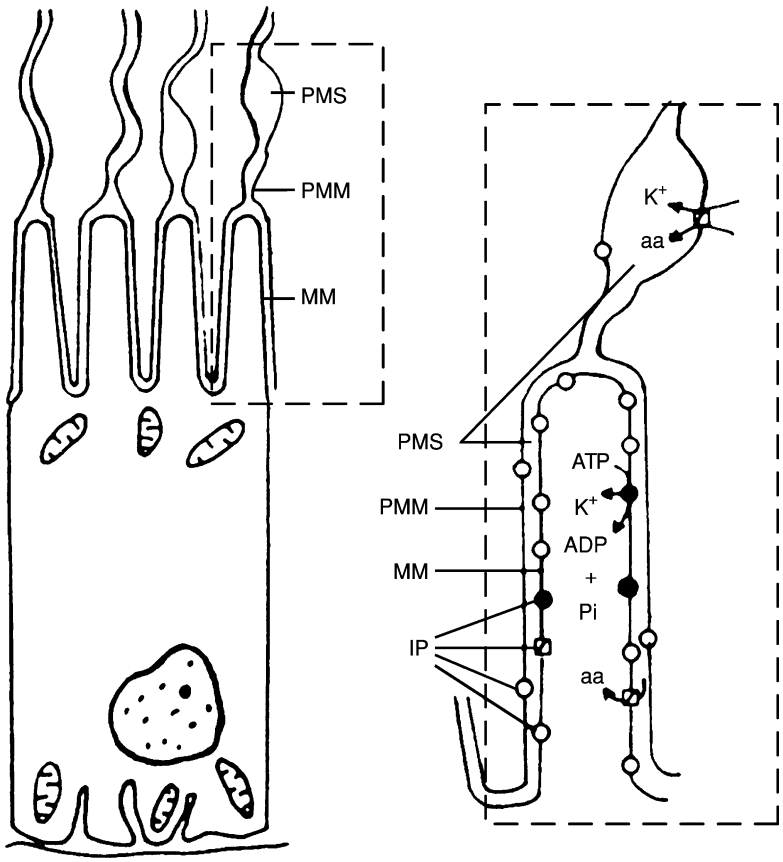


Figure 13 Model for the structure and physiological role of the microvillar border of midgut cells from Hemiptera. The left figure is a diagrammatic representation of a typical Hemiptera midgut cell and the right figure details its apex. The microvillar membrane (MM) is ensheathed by the perimicrovillar membrane (PMM), which extends toward the luminal compartment with a dead end. The microvillar and perimicrovillar membranes delimit a closed compartment, i.e., the perimicrovillar space (PMS). The microvillar membrane is rich and the perimicrovillar membrane is poor in integral proteins (IP). Microvillar membranes actively transport potassium ions (the most important ion in sap) from PMS into the midgut cells, generating a concentration gradient between the gut luminal sap and the PMS. This concentration gradient may be a driving force for the active absorption of organic compounds (amino acids, aa, for example) by appropriate carriers present in the PMM. Organic compounds, once in the PMS, may diffuse up to specific carriers on the microvillar surface. This movement is probably enhanced by a transfer of water from midgut lumen to midgut cells, following (as solvation water) the transmembrane transport of compounds and ions by the putative carriers. (Reprinted with permission from Terra, W.R., Ferreira, C., **1994**. Insect digestive enzymes: properties, compartmentalization and function. *Comp. Biochem. Physiol. B 109*, 1–62; © Elsevier.)

honeydew isoosmotic with hemolymph (Ashford *et al.*, 2000; Cristofoletti *et al.*, 2003). Another interesting adaptation is observed in whiteflies, where a trehalulose synthase forms trehalulose from sucrose, thus making available less substrate for an α-glucosidase that otherwise would increase the osmolarity of ingested fluid on hydrolyzing sucrose (Salvucci, 2003).

A cathepsin L (see Section 4.5.5.3) bound to the modified perimicrovillar membranes of *Acyrthosiphon pisum* (Cristofoletti *et al.*, 2003) may explain the capacity of some phloem sap feeders to rely on protein found in some phloem saps (Salvucci *et al.*, 1998) and the failure of other authors to find an active proteinase in sap feeders. They worked with homogenate supernatants or supernatants of Triton X-100-treated samples, under which conditions the cathepsin L would remain in the pellet.

Amino acid absorption in *A. pisum* midguts is influenced by the presence of the bacteria *Buchnera* in the mycetocytes of the mycetomes occurring in the aphid hemocoel (Prosser *et al.*, 1992). The molecular mechanisms underlying this phenomenon are not known, in spite of the fact that there is strong evidence showing that *Buchnera* uses the nonessential amino acids absorbed by the host in the synthesis of essential amino acids (Prosser and Douglas, 1992; Shigenobu *et al.*, 2000). It is likely that amino acid absorption through apical lamellar carriers depends on the amino acid concentration gradient between midgut lumen and hemolymph, whereas hemolymph titers vary widely according

to *Buchnera* metabolic activity (Liadouze *et al.*, 1995).

The evolution of Heteroptera was associated with regaining the ability to digest polymers. Because the appropriate digestive enzymes were lost, they instead used enzymes derived from lysosomes. Lysosomes are cell organelles involved in intracellular digestion carried out by special proteinases referred to as cathepsins. Compartmentalization of digestion was maintained by the perimicrovillar membranes, as a substitute for the absent peritrophic membrane. Digestion in the two major Heteroptera taxa, Cimicomorpha (exemplified by the blood-feeder *Rhodnius prolixus*), and Pentatomorpha (exemplified by the seed-sucker *Dysdercus peruvianus*), are similar. The dilated anterior midgut stores food and absorbs water and, at least in *D. peruvianus*, also absorbs glucose. Digestion of proteins and absorption of amino acids occurs in the posterior ventriculus. Most protein digestion occurs in lumen with the aid of a cysteine proteinase, and ends in the perimicrovillar space under the action of aminopeptidases and dipeptidases (Terra and Ferreira, 1994). Symbiont bacteria may occur in blood-feeders putatively to provide vitamins (see Section 4.5.2.5). At least in *R. prolixus*, the neuroendocrine system has factors important for maintaining the ultrastructural organization of the midgut epithelial cells (Gonzales *et al.*, 1998).

4.5.8.2.6. Megaloptera Megaloptera include alderflies and dobsonflies and are often considered to be the most primitive group of insects with complete metamorphosis. All their larvae are aquatic predators feeding on invertebrates (Theischinger, 1991). Megaloptera ancestors are like Holometabola ancestors except that for the anterior midgut caeca, which were lost and replaced in function by the anterior midgut. Thus, in *Corydalus* sp. larvae, most digestion occurs in the crop under the action of soluble amylase, maltase, aminopeptidase, and trypsin (major proteinase). Digestive enzyme recycling should occur, as less than 10% of midgut amylase, maltase, and aminopeptidase are lost at each midgut emptying. The higher excretory rate of trypsin (27%) probably results from excess dietary protein (Dias and Terra, unpublished data).

4.5.8.2.7. Coleoptera Coleoptera ancestors are like Megaloptera ones. Nevertheless, there are evolutionary trends leading to a great reduction or loss of the crop and, as in the panorpoid orders, occurrence of at least final digestion of proteins at the surface of midgut cells. Thus, in predatory Carabidae most of the digestive phases occur in the crop by means of midgut enzymes, whereas in predatory larvae of Elateridae initial digestion occurs extraorally by the action of enzymes regurgitated onto their prey. The preliquified material is then ingested by the larvae and its digestion is finished at the surface of midgut cells (Terra and Ferreira, 1994).

The entire digestive process occurs in the dermestid larval endoperitrophic space that is limited by a peritrophic gel in anterior midgut and a peritrophic membrane in posterior midgut. There is a decreasing gradient along the midgut of amylase, maltase, trypsin (major proteinase), and aminopeptidase suggesting the occurrence of digestive enzyme recycling (Terra and Ferreira, 1994; Caldeira, Dias, Terra, and Ribeiro, unpublished data).

Like dermestid beetles, the larvae of *Migdolus fryanus* (Cerambycidae) and *Sphenophorus levis* (Curculionidae) have a peritrophic gel and a peritrophic membrane in the anterior and posterior midgut, respectively, and a decreasing gradient of amylase, maltase, and proteinase along the midgut. In contrast to dermestids, aminopeptidase is a microvillar enzyme in both insects (Dias and Terra, unpublished data). These data do not confirm the earlier suggestion (Terra and Ferreira, 1994) that the final digestion of all nutrients occurs on the surface of midgut cells of Curculionidae.

Tenebrionid larvae also have aminopeptidase as a microvillar enzyme and the distribution of enzymes in gut regions of adults is similar to that in the larvae (Terra and Ferreira, 1994). This suggests that the overall pattern of digestion in larvae and adults of Coleoptera is similar, despite the fact that (in contrast to adults) beetle larvae usually lack a crop.

Insects of the series Cucujiformia (which includes Tenebrionidae, Chrysomelidae, Bruchidae, and Curculionidae) have cysteine proteinases (see Section 4.5.5.3) in addition to (or in place of) serine proteinases as digestive enzymes, suggesting that the ancestors of the whole taxon were insects adapted to feed on seeds rich in serine proteinase inhibitors. The occurrence of trypsin as the major proteinase in *M. fryanus* (Dias and Terra, unpublished data) confirmed the preliminary work (Murdock *et al.*, 1987) according to which cerambycid larvae reacquired serine proteinases.

Scarabaeidae and several related families are relatively isolated in the series Elateriformia and evolved considerably from the Coleoptera ancestor. Scarabid larvae, exemplified by dung beetles, usually feed on cellulose materials undergoing degradation by a fungus-rich flora. Digestion occurs in the midgut, which has three rows of caeca, with a ventral groove between the middle and posterior row. The alkalinity of gut contents increase to almost

pH 12 along the midgut ventral groove. This high pH probably enhances cellulose digestion, which occurs mainly in the hindgut fermentation chamber, through the probable action of bacterial cell-bound enzymes. The final product of cellulose degradation is mainly acetic acid, which is absorbed through the hindgut wall. There is controversy as to whether scarabid larvae ingest feces to obtain nitrogen compounds, as described above for termites (Terra and Ferreira, 1994; Biggs and McGregor, 1996).

4.5.8.2.8. Hymenoptera The organization of the digestive process is variable among hymenopterans and to understand its peculiarities it is necessary to review briefly their evolution. The hymenopteran basal lineages are phytophagous as larvae, feeding both ecto- and endophytically and include several superfamilies like Xyeloidea and Tenthredinoidea, all known as sawflies. Close to these are the Siricoidea (wood wasps) that are adapted to ingest fungus-infected wood. Wood wasp-like ancestors gave rise to the Apocrita (wasp-waisted Hymenoptera) that are parasitoids of insects. They use their ovipositor to injure or kill their host which represents a single meal for their complete development. A taxon sister of Ichneumonoidea in Apocrita gave rise to Aculeata (bees, ants, and wasps with thin waist) (Quicke, 2003).

The digestive systems of Hymenoptera ancestors are like the panorpoid ancestors (**Figure 11c**). However, there are evolutionary trends leading to the loss of midgut caeca (replaced in function by the anterior midgut) and changes in midgut enzyme compartmentalization. These trends appear to be associated with the development of parasitoid habits and were maintained in Aculeata, as described below.

The sawfly *T. malaisei* (Tenthredinoidea: Argidae) larva has a midgut with a ring of anterior caeca that forms a U at the ventral side. Luminal pH is above 9.5 in the first two-thirds of the midgut. Trypsin (major proteinase) and amylase have a decreasing activity along the endoperitrophic space, suggesting enzyme recycling. Maltase predominates in the anterior midgut tissue as a soluble glycocalyx-associated enzyme, whereas aminopeptidase is a microvillar enzyme in posterior midgut (Dias, Ribeiro, and Terra, unpublished data). These characteristics (except the presence of caeca) are similar to those of lepidopteran larvae (see Section 4.5.8.2.10) and explain the fact that this insect is close to the lepidopterans in **Figure 12**. Otherwise, Aculeata with their less sophisticated midgut (see below) branches closer to coleopterans (**Figure 12**).

Wood wasp larvae of the genus *Sirex* are believed to be able to digest and assimilate wood constituents by acquiring cellulase and xylanase, and possibly other enzymes, from fungi present in wood on which they feed (Martin, 1987).

The larvae of Apocrita present a midgut which is closed at its rear end, and which remains unconnected with the hindgut until the time of pupation. It is probable that this condition evolved as an adaptation of endoparasitoid Apocrita ancestors to avoid the release of toxic compounds into the host in which they lived (Terra, 1988).

In larval bees, most digestion occurs in the endoperitrophic space. Countercurrent fluxes seem to occur but there is no midgut luminal pH gradient. Adult bees ingest nectar and pollen. Sucrose from nectar is hydrolyzed in the crop by the action of a sucrase from the hypopharyngeal glands. After ingestion, pollen grains extrude their protoplasm in the ventriculus, where digestion occurs. As in larvae there is also evidence of an endo–ectoperitrophic circulation of digestive enzymes (Jimenez and Gilliam, 1990; Terra and Ferreira, 1994).

Although many authors favor the view that pollen grains are digested in bees after their extrusion by osmotic shock, this subject is controversial not only in bees but also among pollen-feeder beetles (Human and Nicholson, 2003).

Worker ants feed on nectar, honeydew, plant sap, or on partly digested food regurgitated by their larva. Thus, they frequently were said to lack digestive enzymes or display only those enzymes associated with intermediate and (or) final digestion (Terra and Ferreira, 1994). Although this seems true for leaf-cutting ants that appear to rely only on monosaccharides, produced by fungal enzymes acting on plant polysaccharides (Silva *et al.*, 2003), this is not widespread. Thus, adult *Camponotus rufipes* (Formicinae) have soluble amylase, trypsin (major proteinase), maltase, and aminopeptidase enclosed in a type I PM in their midguts. As only 14% of amylase and less than 7% of the other digestive enzymes are excreted during the midgut emptying, these insects may have a digestive enzyme recycling mechanism (Dias and Terra, unpublished data).

4.5.8.2.9. Diptera The Diptera evolved along two major lines: an assemblage (early Nematocera) of suborders corresponding to the mosquitoes, including the basal Diptera, and the suborder Brachycera that includes the most evolved flies (Cyclorrhapha). The Diptera ancestor is similar to the panorpoid ancestor (**Figure 11c**) in having the enzymes involved in intermediate digestion free in the ectoperitrophic fluid (mainly in the large caeca), whereas the enzymes of terminal digestion are

membrane bound at the midgut cell microvilli (Terra and Ferreira, 1994). Although these characteristics are observed in most nonbrachyceran larvae, the more evolved of these larvae may show reduction in size of midgut caeca (e.g., Culicidae). Nonhematophagous adults store liquid food (nectar or decay products) in their crops. Digestion occurs in their midgut as in larvae. Nectar ingested by mosquitoes (males and females) is stored in the crop, and is digested and absorbed at the anterior midgut. Blood, which is ingested only by females, passes to the posterior midgut, where it is digested and absorbed (Billingsley, 1990; Terra and Ferreira, 1994).

Adult *Aedes aegypti* midgut surface is covered in a large proportion by tubular bilayers with diameter fourfold smaller than microvilli. They fuse and branch forming bundles that seem to originate in the intercellular crypts and seem to be fused with the microvillar surface (Ziegler *et al.*, 2000). These structures are not related with the perimicrovillar membranes of Hemiptera. The latter envelop the microvilli before extending into the lumen in structures that may resemble the tubular membrane bilayers of *A. aegypti* (see Section 4.5.8.2.5 and **Figure 13**). The puzzling structures of *A. aegypti* should be further studied to discover their relationships with digestion.

The Cyclorrhapha ancestor (**Figure 11d**) evolved dramatically from the panorpoid ancestor (**Figure 11c**), apparently as a result of adaptations to a diet consisting mainly of bacteria. Digestive events in Cyclorrhapha larvae are exemplified by larvae of the housefly *Musca domestica*. These ingest food rich in bacteria. In the anterior midgut there is a decrease in the starch content of the food bolus, facilitating bacterial death. The bolus now passes into the middle midgut where bacteria are killed by the combined action of low pH, a special lysozyme (see Section 4.5.4.5) and an aspartic proteinase (see Section 4.5.5.4). Finally, the material released by bacteria is digested in the posterior midgut, as is observed in the whole midgut of insects of other taxa. Countercurrent fluxes occur in the posterior midgut powered by secretion of fluid in the distal part of the posterior midgut and its absorption back in middle midgut. The middle midgut has specialized cells for buffering the luminal contents in the acidic zone (**Figure 4**), in addition to those functioning in fluid absorption. Cyclorrhaphan adults, except for a few blood-suckers, feed mainly on liquids associated with decaying material (rich in bacteria) in a way similar to housefly *M. domestica* adults. These salivate (or regurgitate their crop contents) onto their food. After the dispersed material is ingested, starch digestion is accomplished primarily in the crop by the action of salivary amylase. Digestion is followed in the midgut, essentially as described for larvae (Terra and Ferreira, 1994).

The stable fly *Stomoxys calcitrans* stores and concentrates the blood meal in the anterior midgut and gradually passes it to the posterior midgut, where digestion takes place, resembling what occurs in larvae. These adults lack the characteristic cyclorrhaphan middle midgut and the associated luminal low pH. Stable flies occasionally take nectar (Jordão *et al.*, 1996a).

4.5.8.2.10. Lepidoptera Lepidopteran ancestors (**Figure 11e**) differ from panorpoid ancestors because they lack midgut caeca, have all their digestive enzymes (except those of initial digestion) immobilized at the midgut cell surface, and present long-neck goblet cells and stalked goblet-cells in the anterior and posterior larval midgut regions, respectively. Goblet cells excrete K^+ ions that are absorbed from leaves ingested by larvae. Goblet cells also seem to assist anterior columnar cells in water absorption and posterior columnar cells in water secretion (Terra and Ferreira, 1994; Ortego *et al.*, 1996).

Although most lepidopteran larvae have a common pattern of digestion, species that feed on unique diets generally display some adaptations. *Tineola bisselliella* (Tineidae) larvae feed on wool and display a highly reducing midgut for cleaving the disulfide bonds in keratin to facilitate proteolytic hydrolysis of this otherwise insoluble protein (Terra and Ferreira, 1994). Similar results were obtained with *Hofmannophila pseudospretella* (Christeller, 1996). Wax moths (*Galleria mellonella*) infest beehives and digest and absorb wax. The participation of symbiotic bacteria in this process is controversial. Another adaptation has apparently occurred in lepidopteran adults which feed solely on nectar. Digestion of nectar only requires the action of an α-glucosidase (or a β-fructosidase) to hydrolyze sucrose, the major component present. Many nectar-feeding lepidopteran adults have amylase in salivary glands and several glycosidases and peptidases in the midgut (Terra and Ferreira, 1994).

Woods and Kingsolver (1999) developed a chemical reactor model of the caterpillar midgut and used the model as a framework for generating hypotheses about the relationship between feeding responses to variable dietary protein and the physical and biochemical events in the midgut and body. They concluded that absorption (or postabsorptive processes) is limiting in a caterpillar maintained in artificial diets. Caterpillars eating leaves may not have

the same limiting step and this deserves a similar detailed study. Another interesting study would be the development of a model for the beetle midgut. This would determine whether beetles have digestion as the limiting step or consumption to compensate for their less sophisticated midguts.

4.5.9. Digestive Enzyme Secretion Mechanisms and Control

Digestive enzyme secretory mechanisms and control probably are the least understood areas in insect digestion. Studies of the secretory mechanisms have only described major differences, which seem to include unique aspects not seen in other animals.

Insects are continuous (e.g., Lepidoptera and Diptera larvae) or discontinuous (e.g., predators and many hematophagous insects) feeders. The synthesis and secretion of digestive enzymes in continuous feeders seem to be constitutive, that is, they occur continuously (at least between molts), whereas in discontinuous feeders they are regulated (Lehane *et al.*, 1996). Digestive enzymes, as with all animal proteins, are synthesized in the rough endoplasmic reticulum and processed in the Golgi complex, and are packed into secretory vesicles (Figure 14). There are several mechanisms by which the contents of the secretory vesicles are freed in the midgut lumen. In holocrine secretion, secretory vesicles are stored in the cytoplasm until they are released, at which time the whole secretory cell is lost to the extracellular space. During exocytic secretion, secretory vesicles fuse with the midgut cell apical membrane emptying their contents without any loss of cytoplasm (Figure 14a). In contrast, apocrine secretion involves the loss of at least 10% of the apical cytoplasm following the release of secretory vesicles (Figure 14b). These have previously undergone fusions originating larger vesicles that after release eventually free their contents by solubilization (Figure 14b). When the loss of cytoplasm is very small, the secretory mechanism is called microapocrine. Microapocrine secretion consists in releasing budding double-membrane vesicles (Figure 14c) or, at least in insect midguts, pinched-off vesicles that may contain a single or several secretory vesicles (Figure 14d). In both cases the secretory vesicle contents are released by membrane fusion and/or by membrane solubilization caused by high pH contents or by luminal detergents.

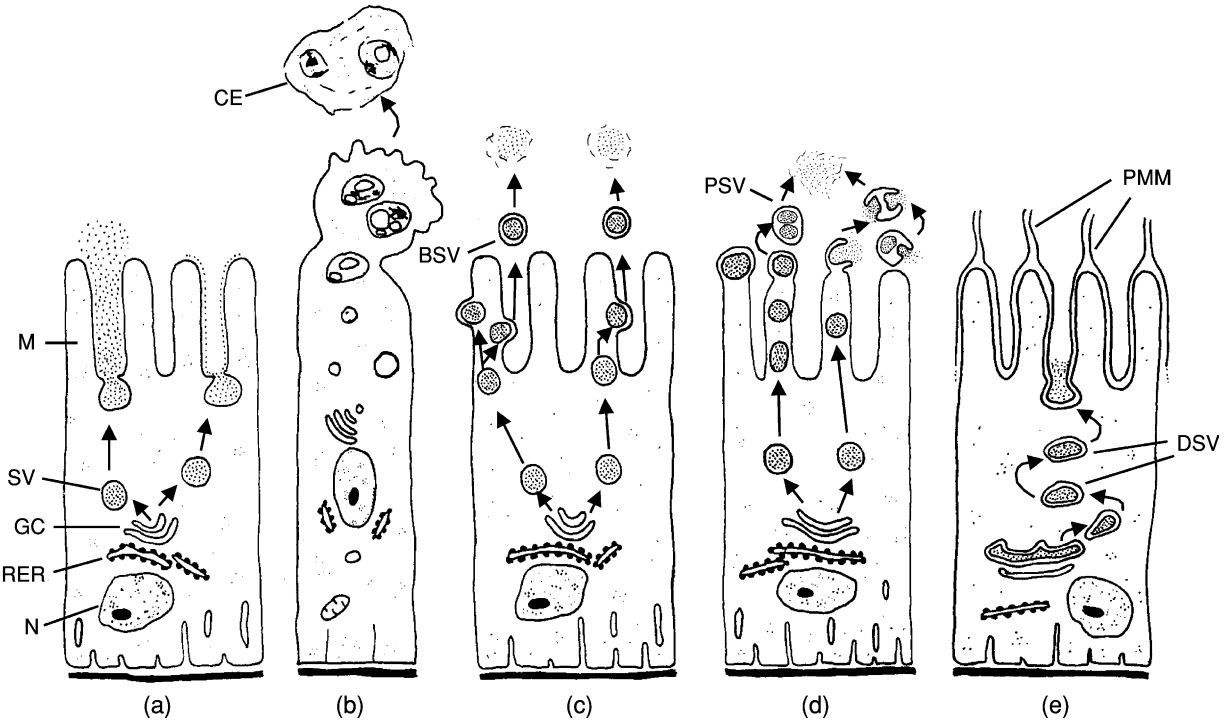


Figure 14 Models for secretory processes of insect digestive enzymes. (a) Exocytic secretion; (b) apocrine secretion; (c) microapocrine secretion with budding vesicles; (d) microapocrine secretion with pinched-off vesicles; (e) modified exocytic secretion in hemipteran midgut cell. BSV, budding secretory vesicle; CE, cellular extrusion; DSV, double-membrane secretory vesicle; GC, Golgi complex; M, microvilli; N, nucleus; PMM, perimicrovillar membrane; PSV, pinched-off secretory vesicle; RER, rough endoplasmatic reticulum; SV, secretory vesicle. (Reprinted with permission from Terra, W.R., Ferreira, C., **2003**. Digestive system. In: Resh, V.H., Cardé, R.T. (Eds.), Encyclopedia of Insects. Academic Press, San Diego, CA, pp. 313–323; © Elsevier.)

The secretory mechanisms of insect midgut cells reviewed below are based on immunocytolocalization data or on data combining biochemical procedures and electron micrographs. Studies based only on traditional cytology have been reviewed elsewhere (Terra and Ferreira, 1994; Lehane *et al*., 1996).

Holocrine secretion is usually described on histological grounds mainly in midgut of insects other than higher Holometabola. These insects have large number of regenerative cells in their midguts. Thus, it is probable that cell renewal in these insects is being misinterpreted as holocrine secretion (Terra and Ferreira, 1994). In spite of this, immunocytochemical data showed that trypsin-containing vesicles along with cell organelles are discharged by opaque zone cells of adult stable flies, suggesting holocrine secretion (Jordão *et al*., 1996a).

Exocytic, apocrine, and microaprocrine secretory mechanisms depend largely on midgut regions. Digestive enzymes are usually secreted by exocytosis in the posterior midgut, whereas alternate mechanisms may be observed in anterior midgut. Thus, trypsin is secreted by the posterior midgut of adult mosquitoes (Graf *et al*., 1986), larval flies (Jordão *et al*., 1996b), and caterpillars (Jordão *et al*., 1999) by exocytosis, as well as, β-glycosidase by *Tenebrio molitor* middle midguts (Ferreira *et al*., 2002). Trypsin is secreted by the anterior midgut of caterpillars using a microapocrine route (Santos *et al*., 1986; Jordão *et al*., 1999), whereas in the anterior midgut of *T. molitor* amylase secretion occurs by an apocrine mechanism (Cristofoletti *et al*., 2001). Based only on morphological evidence, one may say that, in addition to *E. ello* and *Spodoptera frugiperda*, microapocrine secretion occurs in other lepidopteran species, such as *Manduca sexta* (Cioffi, 1979), whereas apocrine secretion is observed in some Orthoptera (Heinrich and Zebe, 1973) and in many coleopteran species other than *T. molitor* (Bayon, 1981; Silva and Souza, 1981; Baker *et al*., 1984).

Immunocytolocalization data (Silva *et al*., 1995) showed that secretion by hemipteran midgut cells displays special features, as the cells have perimicrovillar membranes, in addition to microvillar ones (**Figure 14e**). In this case, double membrane vesicles bud from modified (double membrane) Golgi structures (**Figure 14e**). The double membrane vesicles move to the cell apex, their outer membranes fuse with the microvillar membrane, and their inner membranes fuse with the perimicrovillar membranes, emptying their contents (**Figure 14e**).

Control of digestive enzyme synthesis and secretion in insects has been extensively investigated but the results are often difficult to interpret. It is generally hypothesized that short-term variations in enzymatic activity are controlled by a secretagogue mechanism, whereas long-term changes are regulated hormonally (Lehane *et al*., 1996).

An example of the long-term effects of hormones is the transcription of the early trypsin gene which starts a few hours after mosquito emergence and is under the control of juvenile hormone. However, the early trypsin mRNA is stored in the midgut epithelium and remains untranslated until stimulated by a secretagogue mechanism (Noriega and Wells, 1999) as described below. Other examples are known in molecular detail. A decapeptide trypsin modulating oostatic factor (TMOF) was isolated from the ovaries of *A. aegypti*. It is an ovarian signal that terminates trypsin biosynthesis in the midgut cells after the blood has been digested and its amino acids have been utilized for egg yolk protein synthesis (Borovsky *et al*., 1994). A TMOF-like factor was found in *Heliothis virescens* hemolymph that seems to depress trypsin synthesis at the end of each larval instar (Naven *et al*., 2001).

The presence of food in the midgut is necessary to stimulate synthesis and secretion of digestive enzymes in some insects (Lehane *et al*., 1996). This secretagogue mechanism is known in molecular detail only in mosquitoes. On feeding, these insects express small amounts of early trypsin, using stored early trypsin mRNA (see above). This generates free amino acids and small peptides from blood proteins. These compounds are the initial signals that induce the synthesis and secretion of large amounts of late trypsins that complete digestion (**Figure 15**) (Noriega and Wells, 1999).

The involvement of putative endocrine cells in the control of synthesis and secretion of digestive enzymes has frequently been proposed (Lehane *et al*., 1996). In support of an endocrine role for these cells, the contents of their secretory granules have been shown immunocytochemically to share epitopes with vertebrate neuropeptides. Furthermore, feeding has been shown to cause quantitative changes in the levels of these putative peptide hormones in mosquitoes (Sehnal and Zitman, 1996). However, there is no clear evidence to show what role these putative endocrine cells play in control of midgut events.

4.5.10. Concluding Remarks

In spite of numerous gaps demanding further research, already indicated in this review, it is clear that insect digestive biochemistry is becoming a

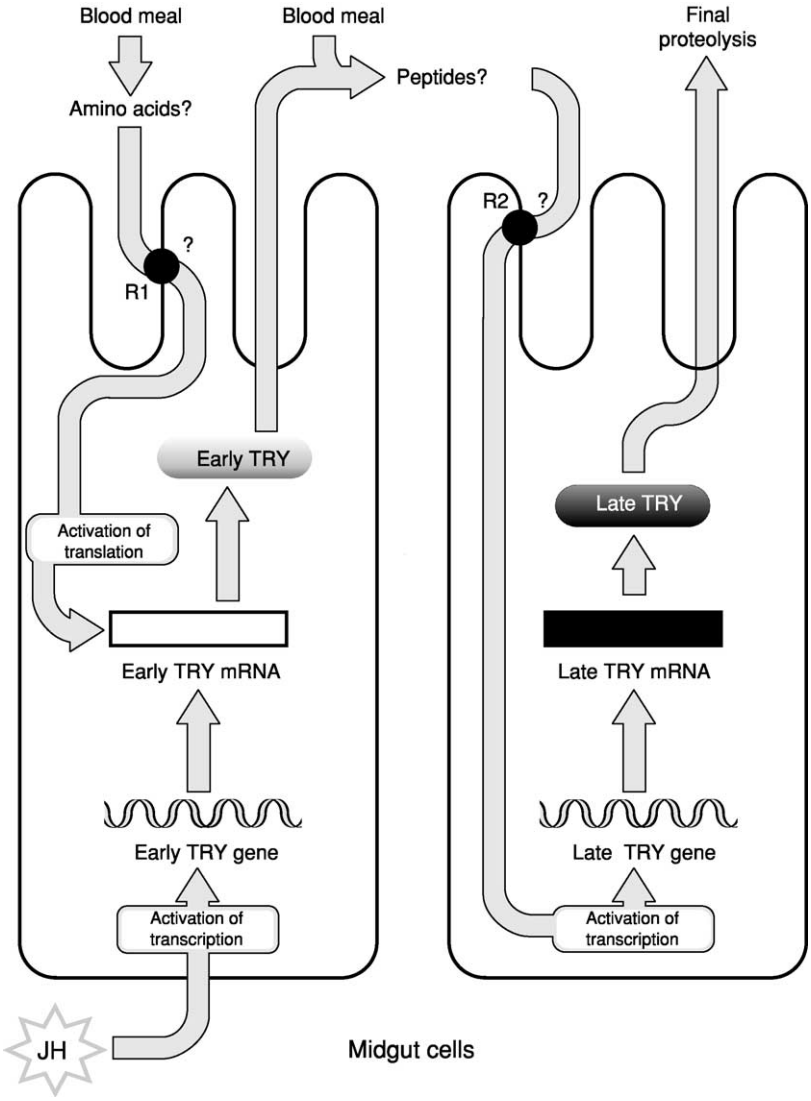


Figure 15 Model for regulation of trypsin (TRY) synthesis following a blood meal in the mosquito midgut. After emergence, juvenile hormone (JH) activates early trypsin gene transcription. Amino acids originating soon after the blood meal (unknown process) enter the cell (R1, amino acid–Na^+ symporter?) and activate early trypsin translation. Early trypsin causes limited proteolysis on blood meal proteins producing unidentified peptides. These probably bind at a receptor (R2) and somehow activate late trypsin transcription. Late trypsin carries out complete proteolysis of blood meal proteins. As digestion of the blood meal nears completion, the mRNA for late trypsin disappears from the midgut. The mechanisms underlying this phenomenon are unknown (Noriega and Wells, 1999).

developed science and that its methods are powerful enough to lead to steady progress. It is conceivable that, in the next few decades, knowledge of the structural biology and function of digestive enzymes and of the control of expression of alternate digestive enzymes and their secretory mechanisms, as well as on microvillar biochemistry, will support the development of more effective and specific methods of insect control. "Quem viver, verá"–Brazilian proverb which has a similar meaning to "Whoever is alive, will see".

Acknowledgments

The work on our laboratory greatly benefited from the long collaboration with our friend A.F. Ribeiro (Instituto de Biociencias, University of São Paulo) on cell biology. Also instrumental for our work were the collaboration with our friends J.R.P. Parra and M.C. Silva-Filho (Escola Superior de Agricultura Luiz de Queiroz, University of São Paulo) on insect nutrition and insect–plant interactions, respectively. We are indebted to Drs. J.E. Baker and B. Oppert

(US Department of Agriculture, Agricultural Research Station, Kansas, USA), R.J. Dillon (University of Bath, UK), and S.R. Marana (University of São Paulo) for critical review of the manuscript. We are also grateful to our former students and now colleagues and friends A.R. Lopes and P.T. Cristofoletti for help with the literature and enzymes sequence trees and to A.B. Dias for the tree of midgut characters and for preparing all drawings. Our work was supported by Brazilian research agencies Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) (Temático and SMOLBnet programs) and CNPq. The authors are staff members of the Biochemistry Department and research fellows of CNPq.

References

Allen, A., **1983**. Mucus: a protective secretion of complexity. *Trends Biochem. Sci. 8*, 169–173.

Alves, L.C., Almeida, P.C., Franzoni, L., Juliano, L., Juliano, M.A., **1996**. Synthesis of *N*-α-protected aminoacyl 7-amino-4-methyl-coumarin amide by phosphorus oxychloride and preparation of specific fluorogenic substrates for papain. *Peptide Res. 9*, 92–96.

Andersen, P.C., Brodbeck, B.V., Mizell, R.F., **1989**. Metabolism of aminoacids, organic acids and sugars extracted from the xylem fluid of four host plants by adult *Homalodisca coagulata*. *Entomol. Exp. Appl. 50*, 149–159.

Appel, H.M., **1994**. The chewing herbivore gut lumen: physicochemical conditions and their impact on plant nutrients, allelochemicals, and insect pathogens. In: Bernays, E.A. (Ed.), Insect–Plant Interactions, vol. 5. CRC Press, Boca Raton, FL, pp. 203–223.

Appel, H.M., Martin, M.M., **1990**. Gut redox conditions in herbivorous lepidopteran larvae. *J. Chem. Ecol. 16*, 3277–3290.

Applebaum, S.W., **1985**. Biochemistry of digestion. In: Kerkut, G.A., Gilbert, L.I. (Eds.), Comprehensive Insect Physiology, Biochemistry, and Pharmacology, vol. 4. Pergamon, Oxford, pp. 279–312.

Arakane, Y., Zhu, Q., Matsumiya, M., Muthukrishnan, S., Kramer, K.J., **2003**. Properties of catalytic, linker and chitin-binding domains of insect chitinase. *Insect Biochem. Mol. Biol. 33*, 631–648.

Ashford, D.A., Smith, W.A., Douglas, A.E., **2000**. Living on a high sugar diet: the fate of sucrose ingested by a phloem-feeding insect, the pea aphid *Acyrthosiphon pisum*. *J. Insect Physiol. 46*, 335–341.

Azevedo, T.M., Terra, W.R., Ferreira, C., **2003**. Purification and characterization of three β-glycosidases from midgut of the sugar cane borer, *Diathraea saccharalis*. *Insect Biochem. Mol. Biol. 33*, 81–92.

Bacic, A., Harris, P.J., Stone, B.A., **1988**. Structure and function of plant cell walls. In: Stumpf, P.K., Conn, E.E. (Eds.), The Biochemistry of Plants, vol. 14. Academic Press, New York, pp. 297–371.

Baker, J.E., Woo, S.M., **1981**. Properties and specificities of a digestive aminopeptidase from larvae of *Attagenus megatoma* (Coleoptera: Dermestidae). *Comp. Biochem. Physiol. B 69*, 189–193.

Baker, J.E., Woo, S.M., Byrd, R.V., **1984**. Ultrastructure of the gut of *Sitophilus granarius* (L.) (Coleoptera: Curculionidae) with notes on distribution of proteinases and amylases in crop and midgut. *Can. J. Zool. 62*, 1251–1259.

Barbehenn, R.V., **1999**. Non-absorption of ingested lipophilic and amphiphilic allelochemicals by generalist grasshoppers: the role of extractive ultrafiltration by the peritrophic envelope. *Arch. Insect Biochem. Physiol. 42*, 130–137.

Barbehenn, R.V., Martin, M.M., **1995**. Peritrophic envelope permeability in herbivorous insects. *J. Insect Physiol. 4*, 303–311.

Barbehenn, R.V., Martin, M.M., **1998**. Formation of insoluble and colloidally dispersed tannic acid complexes in the midgut fluid of *Manduca sexta* (Lepidoptera: Sphingidae): an explanation for the failure of tannic acid to cross the peritrophic envelopes of Lepidopteran larvae. *Arch. Insect Biochem. Physiol. 39*, 109–117.

Barbehenn, R.V., Poopat, U., Spencer, B., **2003**. Semiquinone and ascorbyl radicals in the gut fluids of caterpillars with EPR spectrometry. *Insect Biochem. Mol. Biol. 33*, 125–130.

Barillas-Mury, C., Graf, R., Hagedorn, H.H., Wells, M.A., **1993**. cDNA and deduced amino acid sequence of a blood meal-induced trypsin from the mosquito, *Aedes aegypti*. *Insect Biochem. 2*, 7–12.

Barrett, A.J., Rawlings, N.D., Wasner, J.F., **1998**. Handbook of Proteolytic Enzymes. Academic Press, London.

Bayon, C., **1981**. Ultrastrucuture de l'épithelium intestinal et flore pariétale chez la larve xylophage d' *Oryctes nasicornis* L. (Coleoptera: Scarabaeidae). *Int. J. Insect Morphol. Embryol. 10*, 359–371.

Berenbaum, M., **1980**. Adaptive significance of midgut pH in larval Lepidoptera. *Am. Natural 115*, 138–146.

Bernays, E.A., Chamberlain, D.J., **1980**. A study of tolerance of ingested tannin in *Schistocerca gregaria*. *J. Insect Physiol. 26*, 415–420.

Bernays, E.A., Chamberlain, D.J., Leather, E.M., **1981**. Tolerance of acridids to ingested condensed tannin. *J. Chem. Ecol. 7*, 247–256.

Bieth, J., Spiess, B., Wermuth, C.G., **1974**. The synthesis and analytical use of a highly sensitive and convenient substrate of elastase. *Biochem. Med. 11*, 350–357.

Biggs, D.R., McGregor, P.G., **1996**. Gut pH and amylase and protease activity PM larvae of the New Zealand grass grub (*Costelytra zealandica*; Coleoptera: Scarabaeidae) as a basis for selecting inhibitors. *Insect Biochem. Mol. Biol. 26*, 69–75.

Bignell, D.E., **1981**. Nutrition and digestion. In: Bell, W.J., Adiyodi, N.G. (Eds.), The American Cockroach. Chapman and Hall, London, pp. 57–86.

Billingsley, P.F., **1990**. The midgut ultrastructure of hematophagous insects. *Annu. Rev. Entomol. 35*, 219–248.

Blake, J.D., Murphy, P.T., Richards, G.N., **1971**. Isolation and A/B classification of hemicelluloses. *Carbohydr. Res. 16*, 49–57.

Blanco-Labra, A., Martinez-Gallardo, N.A., Sandoval-Cardoso, L., Delano-Frier, J., **1996**. Purification and characterization of a digestive cathepsin D proteinase isolated from *Tribolium castaneum* larvae (Herbst). *Insect Biochem. Mol. Biol. 26*, 95–100.

Bolognesi, R., Ribeiro, A.F., Terra, W.R., Ferreira, C., **2001**. The peritrophic membrane of *Spodoptera frugiperda*: secretion of peritrophins and role in immobilization and recycling digestive enzymes. *Arch. Insect Biochem. Physiol. 47*, 62–75.

Borhegyi, N.H., Molnár, K., Csikós, G., Sass, M., **1999**. Isolation and characterization of an apically sorted 41-kDa protein from the midgut of tobacco hornworm (*Manduca sexta*). *Cell Tissue Res. 297*, 513–525.

Borovsky, D., Powell, C.A., Nayar, J.K., Blalock, J.E., Hayes, T.K., **1994**. Characterization and localization of mosquito-gut receptors for trypsin modulating oostatic factor using a complementary peptide and immunocytochemistry. *FASEB J. 8*, 350–355.

Botos, I., Meyer, E., Nguyen, M., Swanson, S.M., Koomen, J.M., *et al.*, **2000**. The structure of an insect chymotrypsin. *J. Mol. Biol. 298*, 895–901.

Bown, D.P., Wilkinson, H.S., Gatehouse, J.A., **1997**. Differentially regulated inhibitor-sensitive and insensitive protease genes from the phytophagous insect pest, *Helicoverpa armigera*, are members of complex multigene families. *Insect Biochem. Mol. Biol. 27*, 625–638.

Bown, D.P., Wilkinson, H.S., Gatehouse, J.A., **1998**. Midgut carboxypeptidase from *Helicoverpa armigera* (Lepidopetra: Noctuidae) larvae: enzyme characterization, cDNA cloning and expression. *Insect Biochem. Mol. Biol. 28*, 739–749.

Brito, L.O., Lopes, A.R., Parra, J.R.P., Terra, W.R., Silva-Filho, M.C., **2001**. Adaptation of tobacco budworm *Heliothis virescens* to proteinase inhibitor may be mediated by the synthesis of new proteinases. *Comp. Biochem. Physiol. B 128*, 365–375.

Broadway, P., **1997**. Dietary regulation of serine proteinases that are resistant to serine proteinase inhibitors. *J. Insect Physiol. 43*, 855–874.

Brune, A., **1998**. Termite guts: the world's smallest bioreactors. *Trends Biotechnol. 16*, 16–21.

Campbell, B.C., Bragg, T.S., Turner, C.E., **1992**. Phylogeny of symbiotic bacteria of four weevil species (Coleoptera: Curculionidae) based on analysis of 16s ribosomal DNA. *Insect Biochem. Mol. Biol. 22*, 415–421.

Casu, R.E., Jarmey, J.M., Elvin, C.M., Eisemann, C.H., **1994**. Isolation of a trypsin-like serine protease gene family from the sheep blowfly *Lucilia cuprina*. *Insect Mol. Biol. 3*, 159–170.

Chang, W.X.Z., Gahan, L.J., Tabashnik, B.E., Heckel, D.G., **1999**. A new aminopeptidase from diamondback moth provides evidence for a gene duplication event in Lepidoptera. *Insect Mol. Biol. 8*, 171–177.

Chararas, C., Eberhard, R., Courtois, J.E., Petek, F., **1983**. Purification of three cellulases from the xylophagous larvae of *Ergates faber* (Coleoptera: Cerambycidae). *Insect Biochem. 13*, 213–218.

Charlab, R., Valenzuela, J.G., Rowton, E.D., Ribeiro, J.M., **1999**. Toward an understanding of the biochemical and pharmacological complexity of the saliva of a hematophagous sand fly *Lutzomyia longipalpis*. *Proc. Natl Acad. Sci. USA 96*, 15155–15160.

Christeller, J.T., **1996**. Degradation of wool by *Hofmannophila pseudospretella* (Lepidoptera: Oecophoridae) larval midgut extracts under condition simulating the midgut environment. *Arch. Insect Biochem. Physiol. 33*, 99–119.

Christeller, J.T., Laing, W.A., Shaw, B.D., Burgess, E.P.J., **1990**. Characterization and partial purification of the digestive proteases of the black field cricket, *Teleogryllus commodus* (Walker): elastase is a major component. *Insect Biochem. 20*, 157–164.

Cioffi, M., **1979**. The morphology and fine structure of the larval midgut of a moth (*Manduca sexta*) in relation to active ion transport. *Tissue Cell 11*, 467–479.

Clark, T.M., **1999**. Evolution and adaptive significance of larval midgut alkalinization in the insect superorder Mecopterida. *J. Chem. Ecol. 25*, 1945–1960.

Colebatch, G., Cooper, P., East, P., **2002**. cDNA cloning of a salivary chymotrypsin-like protease and the identification of a six additional cDNAs encoding putative digestive proteases from the green mirid, *Creontiades dilutus* (Hemiptera: Miridae). *Insect Biochem. Mol. Biol. 32*, 1065–1075.

Coutinho, P.M., Henrissat, B., **1999**. Carbohydrate-active enzymes server. http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html.

Craik, C.S., Largman, C., Fletcher, T., Roczmiak, S., Barr, P.J., *et al.*, **1985**. Redesigning trypsin: alteration of substrate specificity. *Science 228*, 291–297.

Cristofoletti, P.T., Ribeiro, A.F., Deraison, C., Rahbé, Y., Terra, W.R., **2003**. Midgut adaptation and digestive enzyme distribution in a phloem feeding insect, the pea aphid *Acyrthosiphon pisum*. *J. Insect Physiol. 49*, 11–24.

Cristofoletti, P.T., Ribeiro, A.F., Terra, W.R., **2001**. Apocrine secretion of amylase and exocytosis of trypsin along the midgut of *Tenebrio molitor*. *J. Insect Physiol. 47*, 143–155.

Cristofoletti, P.T., Terra, W.R., **1999**. Specificity, anchoring, and subsites in the active center of a microvillar aminopeptidase purified from *Tenebrio molitor* (Coleoptera) midgut cells. *Insect Biochem. Mol. Biol. 29*, 807–819.

Cristofoletti, P.T., Terra, W.R., **2000**. The role of amino acid residues in the active site of a midgut microvillar aminopeptidase from the beetle *Tenebrio molitor*. *Biochim. Biophys. Acta 1479*, 185–195.

Daffre, S., Kylsten, P., Samakovlis, C., Hultmark, D., **1994**. The lysozyme locus in *Drosophila melanogaster*:

an expanded gene family adapted for expression in the digestive tract. *Mol. Gen. Genet. 242*, 152–162.

Da Lage, J.L., Von Wormhoudt, A., Cariou, M.L., **2002**. Diversity and evolution of the alpha-amylase genes in animals. *Biologia 57*, 181–189.

D'Amico, S., Gerday, C., Feller, G., **2000**. Structural similarities and evolutionary relationships in chloride-dependent α-amylases. *Gene 253*, 95–105.

Darboux, I., Nielsen-LeRoux, C., Charles, J.F., Pauron, D., **2001**. The receptor of *Bacillus sphaericus* binary toxin in *Culex pipiens* (Diptera: Culicidae) midgut: molecular cloning and expression. *Insect Biochem. Mol. Biol. 31*, 981–990.

Davies, G.J., Wilson, K.S., Henrisat, B., **1997**. Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem. J. 321*, 557–559.

Davis, C.A., Riddell, D.C., Higgins, M.J., Holden, J.J., White, B.N., **1985**. A gene family in *Drosophila melanogaster* coding for trypsin-like enzymes. *Nucleic Acids Res. 13*, 6605–6619.

de Almeida, R.W., Tovar, F.J., Ferreira, I.I., Leoncini, O., **2003**. Chymotrypsin genes in the malaria mosquitoes, *Anopheles aquasalis* and *Anopheles darlingi*. *Insect Biochem. Mol. Biol. 33*, 307–315.

Denolf, P., Hendrickx, K., Van Damme, J., Jansens, S., Peferoen, M., *et al.*, **1997**. Cloning and characterization of *Manduca sexta* and *Plutella xylostella* midgut aminopeptidase N enzymes related to *Bacillus thuringiensis* toxin-binding proteins. *Eur. J. Biochem. 248*, 748–761.

De Veau, E.J.I., Schultz, J.C., **1992**. Reassessment of interaction between gut detergents and tannins in Lepidoptera and significance for gypsy moth larvae. *J. Chem. Ecol. 18*, 1437–1453.

Dey, P.M., Pridham, J.B., **1972**. Biochemistry of α-galactosidases. *Adv. Enzymol. 36*, 91–130.

Dillon, R.J., Dillon, V.M., **2004**. The insect gut bacteria: an overview. *Annu. Rev. Entomol. 49*, 71–92.

Dootsdar, H., McCollum, T.G., Mayer, R.T., **1997**. Purification and characterization of an endo-polygalacturonase from the gut of West Indies sugarcane rootstalk borer weevil (*Diaprepes abbreviatus* L.) larvae. *Comp. Biochem. Physiol. B 118*, 861–867.

Dow, J.A.T., **1986**. Insect midgut function. *Adv. Insect Physiol. 19*, 187–328.

Dow, J.A.T., **1992**. pH gradients in lepidopteran midgut. *J. Exp. Biol. 172*, 355–375.

Dunn, B.M., Jimenez, M., Parten, B.F., Valler, M.J., Rolph, C.E., *et al.*, **1986**. A systematic series of synthetic chromophoric substrates for aspartic proteinases. *Biochem. J. 237*, 899–906.

Edwards, M.J., Jacobs-Lorena, M., **2000**. Permeability and disruption of the peritrophic matrix and caecal membrane from *Aedes aegypti* and *Anopheles gambiae* mosquito larvae. *J. Insect Physiol. 46*, 1313–1320.

Edwards, M.J., Lemos, F.J.A., Donnelly-Doman, M., Jacobs-Lorena, M., **1997**. Rapid induction by a blood meal of a carboxypeptidase gene in the gut of the mosquito *Anopheles gambiae*. *Insect Biochem. Mol. Biol. 27*, 1063–1072.

Eguchi, M., **1995**. Alkaline phosphatase isozymes in insects and comparison with mammalian enzyme. *Comp. Biochem. Physiol. B 111*, 151–162.

Eisemann, C., Wijffels, G., Tellam, R.L., **2001**. Secretion of type 2 peritrophic matrix protein, peritrophin-15, from the cardia. *Arch. Insect Biochem. Physiol. 472*, 76–85.

Elpidina, E.N., Vinokurov, K.S., Gromenko, V.A., Rudenskoya, Y.A., Dunaevsky, Y.E., *et al.*, **2001a**. Compartmentalization of proteinases and amylases in *Nauphoeta cinerea* midgut. *Arch. Insect Biochem. Physiol. 48*, 206–216.

Elpidina, E.N., Vinokurov, K.S., Rudenskaya, Y.A., Dunoevsky, Y.E., Zhuzhikov, D.P., **2001b**. Proteinase inhibitors in *Nauphoeta cinerea* midgut. *Arch. Insect Biochem. Physiol. 48*, 217–222.

Emmerling, M., Chandler, D., Sandeman, M., **2001**. Molecular cloning of three cDNAs encoding aminopeptidases from the midgut of *Helicoverpa punctigera*, the Australian native budworm. *Insect Biochem. Mol. Biol. 31*, 899–907.

Espinoza-Fuentes, F.P., Ferreira, C., Terra, W.R., **1984**. Spatial organization of digestion in the larval and imaginal stages of the sciarid fly *Trichosia pubescens*. *Insect Biochem. 14*, 631–638.

Estébanez-Perpiñá, E., Bayés, A., Vendrell, J., Jongsma, M.A., Bown, D.P., *et al.*, **2001**. Crystal structure of a novel mid-gut procarboxypeptidase from the cotton pest *Helicoverpa armigera*. *J. Mol. Biol. 313*, 629–638.

Ferreira, A.H.P., Marana, S.R., Terra, W.R., Ferreira, C., **2001**. Purification, molecular cloning, and properties of a β-glycosidase isolated from midgut lumen of *Tenebrio molitor* (Coleoptera) larvae. *Insect Biochem. Mol. Biol. 31*, 1065–1076.

Ferreira, A.H.P., Ribeiro, A.F., Terra, W.R., Ferreira, C., **2002**. Secretion of β-glycosidase by middle midgut cells and its recycling in the midgut of *Tenebrio molitor* larvae. *J. Insect Physiol. 48*, 113–118.

Ferreira, A.H.P., Terra, W.R., Ferreira, C., **2003**. Characteization of a β-glycosidase highly active on disaccharides and of a β-galactosidase from *Tenebrio molitor* midgut lumen. *Insect Biochem. Mol. Biol. 33*, 253–265.

Ferreira, C., Capella, A.N., Sitnik, R., Terra, W.R., **1994a**. Properties of the digestive enzymes and the permeability of the peritrophic membrane of *Spodoptera frugiperda* (Lepidoptera) larvae. *Comp. Biochem. Physiol. A 107*, 631–640.

Ferreira, C., Capella, A.N., Sitnik, R., Terra, W.R., **1994b**. Digestive enzymes in midgut cells, endo- and ectoperitrophic contents and peritrophic membranes of *Spodoptera frugiperda* (Lepidoptera) larvae. *Arch. Insect Biochem. Physiol. 26*, 299–313.

Ferreira, C., Parra, J.R.P., Terra, W.R., **1997**. The effect of dietary plant glycosides on larval midgut β-glycosidases from *Spodoptera frugiperda* and *Diatraea saccharalis*. *Insect Biochem. Mol. Biol. 27*, 55–59.

Ferreira, C., Terra, W.R., **1983**. Physical and kinetic properties of a plasma membrane bound β-D-glucosidase (cellobiase) from midgut cells of an insect (*Rhynchosciara americana* larva). *Biochem. J. 213*, 43–51.

Ferreira, C., Terra, W.R., **1984**. Soluble aminopeptidases from cytosol and luminal contents of *Rhynchosciara americana* midgut caeca: properties and phenanthroline inhibition. *Insect Biochem. 14*, 145–150.

Ferreira, C., Terra, W.R., **1985**. Minor aminopeptidases purified from the plasma membrane of midgut caeca cells of an insect (*Rhynchosciara americana*) larva. *Insect Biochem. 15*, 619–625.

Ferreira, C., Terra, W.R., **1986a**. The detergent form of the major aminopeptidase from the plasma membrane of midgut caeca cells of *Rhynchosciara americana* (Diptera) larva. *Comp. Biochem. Physiol. B 84*, 373–376.

Ferreira, C., Terra, W.R., **1986b**. Substrate specificity and binding loci for inhibitors in an aminopeptidase purified from the plasma membrane of midgut cells of an insect (*Rhynchosciara americana*) larva. *Arch. Biochem. Biophys. 244*, 478–485.

Ferreira, C., Terra, W.R., **1989**. Spatial organization of digestion, secretory mechanisms and digestive enzyme properties in *Pheropsophus aequinoctialis* (Coleoptera: Carabidae). *Insect Biochem. 19*, 383–391.

Ferreira, C., Torres, B.B., Terra, W.R., **1998**. Substrate specificities of midgut β-glycosidases from insects of different orders. *Comp. Biochem. Physiol. B 119*, 219–225.

Ferreira, C., Marana, S.R., Silva, C., Terra, W.R., **1999**. Properties of digestive glycosidases and peptidases and the permeability of the peritrophic membranes of *Abracris flavolineata* (Orthoptera : Acrididae). *Comp. Biochem. Physiol. B 123*, 241–250.

Forstner, G.G., Forstner, J.F., **1986**. Structure and function of the gastrointestinal fluid. In: Desnuelle, P., Sjostrom, H., Noren, O. (Eds.), Molecular and Cellular Basis of Digestion. Elsevier, Amsterdam, pp. 125–143.

Fox, L.R., Macauley, B.J., **1977**. Insect grazing on *Eucalyptus* in response to variation in leaf tannins and nitrogen. *Oecologia 29*, 145–162.

Franco, O.L., Rigden, D.J., Melo, F.R., Grossi-de-Sá, M.F., **2002**. Plant α-amylase inhibitors and their interaction with insect α-amylases: structure, function and potential for crop protection. *Eur. J. Biochem. 269*, 397–412.

Fujita, A., Minamoto, T., Shimizu, I., Abe, T., **2002**. Molecular cloning of lysozyme-encoding cDNAs expressed in the salivary gland of a wood-feeding termite, *Reticulitermes speratus*. *Insect Biochem. Mol. Biol. 32*, 1615–1624.

Fujita, A., Shimizu, I., Abe, T., **2001**. Distribution of lysozyme and protease, and amino acid concentration in guts of a wood-feeding termite, *Reticulitermes speratus* (Kolbe): possible digestion of symbiont bacteria transferred by trophallaxis. *Physiol. Entomol. 26*, 116–123.

Gaines, P.J., Sampson, C.M., Rushlow, K.E., Stiegler, G.L., **1999**. Cloning of a family of serine protease genes from the cat flea *Ctenocephalides felis*. *Insect Mol. Biol. 81*, 11–22.

Garner, K.J., Hiremath, S., Lehtoma, K., Valaitis, A.P., **1999**. Cloning and complete sequence characterization of two gypsy moth aminopeptidase-N cDNAs, including the receptor for *Bacillus thuringiensis* Cry 1Ac toxin. *Insect Biochem. Mol. Biol. 29*, 527–535.

Gatehouse, L.N., Shannon, A.L., Burgess, E.P.J., Christeller, J.T., **1997**. Characterization of major midgut proteinase cDNAs from *Helicoverpa armigera* larvae and changes in gene expression in response to four proteinase inhibitors in the diet. *Insect Biochem. Mol. Biol. 27*, 929–944.

Genta, F.A., Terra, W.R., Ferreira, C., **2003**. Action pattern, specifity, lytic activities, and physiological role of five digestive β-glucanases isolated from *Periplaneta americana*. *Insect Biochem. Mol. Biol. 33*, 1085–1097.

Gill, S.S., Cowles, E., Francis, V., **1995**. Identification, isolation, and cloning of a *Bacillus thuringiensis* Cry 1ac-toxin binding protein from the midgut of the lepidopteran insect *Heliothis virescens*. *J. Biol. Chem. 270*, 27277–27282.

Girard, C., Jouanin, L., **1999a**. Molecular cloning of cDNAs encoding a range of digestive enzymes from a phytophagous beetle *Phaedon cochleariae*. *Insect Biochem. Mol. Biol. 29*, 1129–1142.

Girard, G., Jouanin, L., **1999b**. Molecular cloning of a gut-specific chitinase cDNA from the beetle *Phaedon cochleariae*. *Insect Biochem. Mol. Biol. 29*, 549–556.

Giri, A.P., Harsulkar, A.M., Deshpande, V.V., Sainani, M.N., Gupta, V.S., *et al.*, **1998**. Chickpea defensive proteinase inhibitors can be inactivated by podborer gut proteinases. *Plant Physiol. 116*, 393–401.

Gonzales, M.S., Nogueira, N.F.S., Feder, D., de Souza, W., Azambuja, P., *et al.*, **1998**. Role of the head in the ultrastructural midgut organization in *Rhodnius prolixus* larvae: evidence from head transplantation experiments and acdysone therapy. *J. Insect Physiol. 44*, 553–560.

Graf, R., Raikhel, A.S., Brown, M.R., Lea, A.O., Briegel, H., **1986**. Mosquito trypsin: immunocytochemical localization in the midgut of blood-fed *Aedes aegyti* (L.). *Cell Tissue Res. 245*, 19–27.

Greenberg, B., Paretsky, D., **1955**. Proteolytic enzymes in the housefly, *Musca domestica* (L.). *Ann. Entomol. Soc. America 48*, 46–50.

Grossmann, G.A., Terra, W.R., **2001**. Alpha-galactosidases from the larval midgut of *Tenebrio molitor* (Coleoptera) and *Spodoptera frugiperda* (Lepidoptera). *Comp. Biochem. Physiol. B 128*, 109–122.

Grossmann, G.L., Campos, Y., Senerson, D.W., James, A.A., **1997**. Evidence for two distinct members of the amylase gene family in the yellow fever mosquito, *Aedes aegypti*. *Insect Biochem. Mol. Biol. 27*, 769–781.

Harper, M.S., Hopkins, T.L., **1997**. Peritrophic membrane structure and secretion in European corn borer larvae (*Ostrinia nubilalis*). *Tissue Cell 29*, 461–475.

Harwood, J.L., **1980**. Plant acyl lipids: structure, distribution and analysis. In: Stumpf, P.K. (Ed.), The Biochemistry of Plants: A Comprehensive Treatise, vol. 4, Lipids: Structure and Function. Academic Press, New York, pp. 1–56.

Harzer, K., Hiraiwa, M., Paton, B.C., **2001**. Saposins (sap) A and C activate the degradation of galactosylsphingosine. *FEBS Lett. 508*, 107–110.

Heinrich, D., Zebe, E., **1973**. Zur Feinstruktur der Mitteldarmzellen von *Locusta migratoria* in verschiedenen Phasen der Verdauung. *Cytobiologie 7*, 315–326.

Hooper, N.M., **1994**. Families of zinc metalloproteases. *FEBS Lett. 354*, 1–6.

Hori, K., **1972**. Comparative study of a property of salivary amylase among various heteropterous insects. *Comp. Biochem. Physiol. B 42*, 501–508.

House, H.L., **1974**. Digestion. In: Rockstein, M. (Ed.), The Physiology of Insecta, 2nd edn., vol. 5. Academic Press, New York, pp. 63–117.

Hua, G., Tsukamoto, K., Ikezawa, H., **1998**. Cloning and sequence analysis of the aminopeptidase N isozyme (APN2) from *Bombyx mori* midgut. *Comp. Biochem. Physiol. B 121*, 213–232.

Hultmark, D., **1996**. Insect lysozymes. In: Jollés, P. (Ed.), Lysozymes: Model Enzymes in Biochemistry and Biology. Birkhäuser Verlag, Basel, pp. 87–102.

Human, H., Nicolson, S.W., **2003**. Digestion of maize and sunflower pollen by the spotted maize *Astylus atromaculatus* (Melyridae): is there a role for osmotic shock? *J. Insect Physiol. 49*, 633–643.

Imoto, T., Johnson, L.N., North, A.C.T., Phillips, D.C., Ruppley, J.A., **1972**. Vertebrate lysozymes. In: Boyer, P.D. (Ed.), The Enzymes, vol. 7. Academic Press, New York, pp. 665–868.

Ito, Y., Nakamura, M., Hotani, T., Imoto, T., **1995**. Insect lysozyme from house fly (*Musca domestica*) larvae: possible digestive function based on sequence and enzymatic properties. *J. Biochem. Tokyo 118*, 546–551.

Itoh, M., Kanamori, Y., Takao, M., Eguchi, M., **1999**. Cloning of soluble alkaline phosphatase cDNA and molecular basis of the polymorphic nature in alkaline phosphatases isozymes of *Bombyx mori* midgut. *Insect Biochem. Mol. Biol. 29*, 121–129.

Itoh, M., Takeda, S., Yamamoto, H., Izumi, S., Tomino, S., *et al.*, **1991**. Cloning and sequence analysis of membrane-bound alkaline phosphatase cDNA of the silkworm, *Bombyx mori*. *Biochim. Biophys. Acta 1129*, 135–138.

Ivanova, N., Sorokin, A., Anderson, L., Galleron, N., Candellon, B., *et al.*, **2003**. Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. *Nature 423*, 87–91.

James, A.A., Blackmer, K., Racioppi, J.V., **1989**. A salivary gland-specific, maltase-like gene of the vector mosquito, *Aedes aegypti*. *Gene 75*, 73–83.

Janecek, S., **1997**. Alpha-amylase family: molecular biology and evolution. *Progr. Biophys. Mol. Biol. 67*, 67–97.

Jany, K.D., Bekeler, K., Pfleiderer, G., Ishay, J., **1983**. Amino acid sequence of an insect chymotrypsin from the larvae of the hornet, *Vespa orientalis*. *Biochem. Biophys. Res. Commun. 110*, 1–7.

Jimenez, D.R., Gilliam, M., **1990**. Ultrastructure of ventriculus of the honey bee *Apis mellifera* (L): cytochemical localization of the acid phosphatase, alkaline phosphatase, and non-specific esterase. *Cell Tissue Res. 261*, 431–443.

Johnson, K.S., Felton, G.W., **2000**. Digestive proteinase activity in corn earworm (*Helicoverpa zea*) after molting and in response to lowered redox potential. *Arch. Insect Biochem. Physiol. 44*, 151–161.

Johnson, K.S., Rabosky, D., **2000**. Phylogenetic distribution of cysteine proteinases in beetles: evidence for an evolutionary shift to an alkaline digestive strategy in Cerambycidae. *Comp. Biochem. Physiol. B 126*, 609–619.

Jordão, B.P., Capella, A.N., Terra, W.R., Ribeiro, A.F., Ferreira, C., **1999**. Nature of the anchors of membrane-bound aminopeptidase, amylase, and trypsin and secretory mechanisms in *Spodoptera frugiperda* (Lepidoptera) midgut cells. *J. Insect Physiol. 45*, 29–37.

Jordão, B.P., Lehane, M.J., Terra, W.R., Ribeiro, A.F., Ferreira, C., **1996a**. An immunocytochemical investigation of trypsin secretion in the midgut of the stablefly *Stomoxys calcitrans*. *Insect Biochem. Mol. Biol. 26*, 445–453.

Jordão, B.P., Terra, W.R., Ribeiro, A.F., Lehane, M.J., Ferreira, C., **1996b**. Trypsin secretion in *Musca domestica* larval midguts: a biochemical and immunocytochemical study. *Insect Biochem. Mol. Biol. 26*, 337–346.

Khadeni, S., Guarino, L.A., Watenabe, H., Tokuda, G., Meyer, E.F., **2002**. Structure of an endoglucanase from termite, *Nasutitermes takasagoensis*. *Acta Cryst. D 58*, 653–658.

Kleywegt, G.J., Zou, J.Y., Divne, C., Davies, G.J., Sinning, I., *et al.*, **1997**. The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 angstrom resolution, and a comparison with related enzymes. *J. Mol. Biol. 272*, 383–397.

Klinkowstrom, A.M., Terra, W.R., Ferreira, C., **1994**. Aminopeptidase A from *Rhynchosciara americana* (Diptera) larval midguts: properties and midgut distribution. *Arch. Insect Biochem. Physiol. 27*, 301–305.

Klinkowstrom, A.M., Terra, W.R., Ferreira, C., **1995**. Midgut dipeptidases from *Rhynchosciara americana* (Diptera) larvae: properties of soluble and membrane bound forms. *Insect Biochem. Mol. Biol. 25*, 303–310.

Knight, P.J.K., Knowles, B.H., Eller, D.J., **1995**. Molecular cloning of an insect aminopeptidase N that serves as a receptor of *Bacillus thuringiensis* Cry 1a (c) toxin. *J. Biol. Chem. 270*, 17765–17770.

Koepke, J., Ermler, V., Warkentin, E., Wenzi, G., Flecker, P., **2000**. Crystal structure of cancer chemopreventive Bowman–Birk inhibition in ternary complex with bovine trypsin at 2–3 Å resolution: structural basis of

Janus-faced serine protease inhibitor specificity. *J. Mol. Biol. 298*, 477–491.

Koiwa, H., Shade, R.E., Zhu-Salzman, K., D'Urzo, M.P., Murdock, L.L., *et al.*, **2000**. A plant defensive cystatin (soyacystatin) targets cathepsin L-like digestive cysteine proteinases (DuCALs) in the larval midgut of the western corn rootworm (*Diabrotica virgifera virgifera*). *FEBS Lett. 471*, 67–70.

Konno, K., Okada, S., Hirayama, C., **2001**. Selective secretion of free glycine, a neutralizer against a plant defense chemical, in the digestive juice of the privet moth larvae. *J. Insect Physiol. 47*, 1451–1457.

Kramer, K.J., Corpuz, L., Choi, H., Muthukrishnan, S., **1993**. Sequence of a cDNA and expression of the genes encoding epidermal and gut chitinases of *Manduca sexta*. *Insect Biochem. Mol. Biol. 23*, 691–701.

Kramer, K.J., Koga, D., **1986**. Insect chitin: physical state, synthesis, degradation and metabolic regulation. *Insect Biochem. 16*, 851–877.

Kylsten, P., Kimbrell, D.A., Daffre, S., Samakovlis, C., Hultmark, D., **1992**. The lysozyme locus in *Drosophila melanogaster*: different genes are expressed in midgut and salivary glands. *Mol. Gen. Genet. 232*, 335–343.

Le Cahérec, F., Guillam, M.T., Beuron, F., Cavalier, A., Thomas, D., *et al.*, **1997**. Aquaporin-related proteins in the filter chamber of homopteran insects. *Cell Tissue Res. 290*, 143–151.

Lee, J.H., Tsuji, M., Nakamura, M., Nishimoto, M., Okuyama, M., *et al.*, **2001**. Purification and identification of the essential ionizable groups of honeybee, *Apis mellifera* L., trehalase. *Biosci. Biotechnol. Biochem. 65*, 2657–2665.

Lee, M.J., Anstee, J.H., **1995**. Endoproteases from the midgut of *Spodoptera littoralis* include a chymotrypsin-like enzyme with an extended binding site. *Insect Biochem. Mol. Biol. 25*, 49–61.

Lehane, M.J., **1997**. Peritrophic matrix structure and function. *Annu. Rev. Entomol. 42*, 525–550.

Lehane, M.J., Müller, H.M., Crisanti, A., **1996**. Mechanisms controlling the synthesis and secretion of digestive enzymes in insects. In: Lehane, M.J., Billingsley, P.F. (Eds.), Biology of the Insect Midgut. Chapman and Hall, London, pp. 195–205.

Lemos, F.J.A., Ribeiro, A.F., Terra, W.R., **1993**. A bacteria-digesting midgut lysozyme from *Musca domestica* (Diptera) larvae: purification, properties and secretory mechanism. *Insect Biochem. Mol. Biol. 23*, 533–541.

Lemos, F.J.A., Terra, W.R., **1991a**. Digestion of bacteria and the role midgut lysozyme in some insect larvae. *Comp. Biochem. Physiol. B 100*, 265–268.

Lemos, F.J.A., Terra, W.R., **1991b**. Properties and intracellular distribution of a cathepsin D-like proteinase active at the acid region of *Musca domestica* midgut. *Insect Biochem. 21*, 457–465.

Levitzki, A., Steer, M.L., **1974**. The allosteric activation of mammalian α-amylase by chloride. *Eur. J. Biochem. 41*, 171–180.

Liadouze, I., Febvay, G., Guillaud, J., Bonnot, G., **1995**. Effect of diet on the free amino acid pools of symbiotic and aposymbiotic pea aphids, *Acyrthosiphon pisum*. *J. Insect Physiol. 41*, 33–40.

Lin, G., Bode, W., Huber, R., Chi, C., Engh, R.A., **1993**. The 0.25-nm X-ray structure of Bowman–Birk-type inhibitor from mung bean in ternary complex with porcine trypsin. *Eur. J. Biochem. 212*, 549–555.

Linder, M., Teeri, T.T., **1997**. The roles and function of cellulose-binding domains. *J. Biotechnol. 57*, 15–28.

Lo, N., Tokuda, G., Watanabe, H., Rose, H., Slaytor, M., *et al.*, **2000**. Evidence from multiple gene sequences indicates that termites evolved from wood-feeding cockroaches. *Current Biol. 10*, 801–804.

Lopes, A.R., Juliano, M.A., Juliano, L., Terra, W.R., **2004**. Coevolution of insect trypsins and inhibitors. *Arch. Insect Biochem. Physiol. 55*, 140–152.

Lu, Y., Zen, K.-C., Muthukrishnan, S., Kramer, K.J., **2002**. Site-directed mutagenesis and functional analysis of active site acidic amino acid residues D142, D144 and E146 in *Manduca sexta* (tobacco hornworm) chitinase. *Insect Biochem. Mol. Biol. 32*, 1369–1382.

Luciani, V., Marie-Claire, C., Ruffet, E., Beaumont, A., Roques, B.P., *et al.*, **1998**. Characterization of Glu^{350} as a critical residue involved in the N-terminal amine binding site of aminopeptidase N (EC 3.4.11.2): insight into its mechanism of action. *Biochemistry 37*, 686–692.

Marana, S.R., Jacobs-Lorena, M., Terra, W.R., Ferreira, C., **2001**. Amino acid residues involved in substrate binding and catalysis in an insect digestive β-glycosidase. *Biochim. Biophys. Acta 1545*, 41–52.

Marana, S.R., Lopes, A.R., Juliano, L., Juliano, M.A., Ferreira, C., *et al.*, **2002a**. Subsites of trypsin active site favor catalysis or substrate binding. *Biochem. Biophys. Res. Commun. 290*, 494–497.

Marana, S.R., Ribeiro, A.F., Terra, W.R., Ferreira, C., **1997**. Ultrastructure and secretory activity of *Abracris flavolineata* (Orthoptera: Acrididae) midguts. *J. Insect Physiol. 43*, 465–473.

Marana, S.R., Terra, W.R., Ferreira, C., **1995**. Midgut β-D-glucosidases from *Abracris flavolineata* (Orthoptera: Acrididae): physical properties, substrate specificities and function. *Insect Biochem. Mol. Biol. 25*, 835–843.

Marana, S.R., Terra, W.R., Ferreira, C., **2000**. Purification and properties of a β-glycosidase purified from midgut cells of *Spodoptera frugiperda* (Lepidoptera) larvae. *Insect Biochem. Mol. Biol. 30*, 1139–1146.

Marana, S.R., Terra, W.R., Ferreira, C., **2002b**. The role of amino acid residues Q39 and E451 in the determination of substrate specificity of the *Spodoptera frugiperda* β-glycosidase. *Eur. J. Biochem. 269*, 3705–3714.

Markovic, O., Janecek, S., **2001**. Pectin degrading glycoside hydrolases of family 28: sequence-structural features, specificities and evolution. *Protein Eng. 14*, 615–631.

Martin, M.M., **1987**. Invertebrate–Microbial Interactions: Ingested Fungal Enzymes in Anthropod Biology. Cornell University Press, Ithaca, NY.

Martin, M.M., Martin, J.S., **1984**. Surfactants: their role in preventing the precipitation of proteins by tannins in insect guts. *Oecologia 61*, 342–345.

Matoub, M., Rouland, C., **1995**. Purification and properties of the xylanases from the termite *Macrotermes bellicosus* and its symbiotic fungus *Termitomyces* sp. *Comp. Biochem. Physiol. B 112*, 629–635.

Matsumura, I., Kirsch, J.F., **1996**. Is aspartate 52 essential for catalysis by chicken egg white lysozyme? The role of natural substrate-assisted hydrolysis. *Biochemistry 35*, 1881–1889.

Mazumbar-Leighton, S., Broadway, R.M., **2001a**. Transcriptional induction of diverse midgut trypsins in larval *Agrotis ipsilon* and *Helicoverpa zea* feeding on the soybean trypsin inhibitor. *Insect Biochem. Mol. Biol. 31*, 645–657.

Mazumbar-Leighton, S., Broadway, R.M., **2001b**. Identification of six chymotrypsin cDNAs from larval midgut of *Helicoverpa zea* and *Agrotis ipsilon* feeding on the soybean (Kunitz) trypsin inhibitor. *Insect Biochem. Mol. Biol. 31*, 633–644.

McHenery, J.G., Birkbeck, T.H., Allen, S.A., **1979**. The occurrence of lysozyme in marine bivalves. *Comp. Biochem. Physiol. B 63*, 25–28.

Miller, N., Lehane, M.J., **1990**. *In vitro* perfusion studies on the peritrophic membrane of the tsetse fly *Glossina morsitans morsitans* (Diptera, Glossinidae). *J. Insect Physiol. 36*, 813–818.

Monzingo, A.F., Marcotte, E.M., Hart, P.J., Roberts, J.D., **1996**. Chitinases, chitosanases, and lysozymes can be divided into prokaryotic and eucaryotic families sharing a conserved core. *Nature Struct. Biol. 3*, 133–140.

Morgan, M.R.J., **1975**. Relationship between gut cellobiase, lactose, aryl β-glucosidase, and aryl β-galactosidase activities of *Locusta migratoria*. *Insect Biochem. 5*, 609–617.

Murdock, L.L., Brookhart, G., Dunn, P.E., Foard, D.E., Kelley, S., *et al.*, **1987**. Cysteine digestive proteinases in Coleoptera. *Comp. Biochem. Physiol. B 87*, 783–787.

Nahoum, V., Farisei, F., Lê-Berre-Anton, V., Egloff, M.P., Rougé, P., *et al.*, **1999**. A plant-seed inhibitor of two classes of α-amylases: X-ray analysis of *Tenebrio molitor* larval α-amylases in complex with the bean *Phaseolus vulgaris* inhibitor. *Acta Crystallogr. D 55*, 360–362.

Nakanishi, K., Yaoi, K., Nagino, Y., Hara, H., Kitami, M., *et al.*, **2002**. Aminopeptidase N isoforms from the midgut of *Bombyx mori* and *Plutella xylostella*: their classification and the factors that determine their binding specificity to *Bacillus thuringiensis* Cry 1A toxin. *FEBS Lett. 519*, 215–220.

Nakashima, K., Watanabe, H., Saitoh, H., Tokuda, G., Azuma, J.I., **2002**. Dual cellulose-digesting system of the wood-feeding termite, *Coptotermes formosanus* Shiraki. *Insect Biochem. Mol. Biol. 32*, 777–784.

Naven, R., Sorge, D., Sterner, A., Borovsky, D., **2001**. TMOF-like factor controls the biosynthesis of serine proteases in the larval gut of *Heliothis virescens*. *Arch. Insect Biochem. Physiol. 47*, 169–180.

Nielsen, J.E., Borchert, T.U., **2000**. Protein engineering of bacterial α-amylases. *Biochim. Biophys. Acta 1543*, 253–274.

Nishimoto, M., Kubota, M., Tsuji, M., Mori, H., Kimura, A., *et al.*, **2001**. Purification and substrate specificity of honeybee, *Apis mellifera* L., alpha-glucosidase III. *Biosci. Biotech. Biochem. 65*, 1610–1616.

Nor Aliza, A.R., Rana, R.L., Skoda, S.R., Berkebile, D.R., Stanley, D.W., **1999**. Tissue polyunsaturated fatty acids and a digestive phospholipase A_2 in the primary screwworm, *Cochliomya hominivorax*. *Insect Biochem. Mol. Biol. 29*, 1029–1038.

Norén, O., Sjostrom, H., Danielsen, E.M., Cowell, G.M., Skovbjerg, H., **1986**. The enzymes of the enterocyte plasma membrane. In: Desnuelle, P., Sjostrom, H., Norén, O. (Eds.), Molecular and Cellular Basis of Digestion. Elsevier, Amsterdam, pp. 355–365.

Noriega, F.G., Wells, M.A., **1999**. A molecular view of trypsin synthesis in the midgut of *Aedes aegypti*. *J. Insect Physiol. 45*, 613–620.

Novillo, C., Castañera, P., Ortego, F., **1997**. Inhibition of digestive trypsin-like proteases from larvae of several lepidopteran species by the diagnostic cysteine protease inhibitor E-64. *Insect Biochem. Mol. Biol. 27*, 247–254.

Ohashi, K., Natori, S., Kubo, T., **1999**. Expression of amylase and glucose oxidase in the hypopharyngeal gland with an age-dependent role change of the worker honeybee (*Apis mellifera* L.). *Eur. J. Biochem. 265*, 127–133.

Oltean, D.I., Pullikuth, A.K., Lee, H.K., Gill, S.S., **1999**. Partial purification and characterization of *Bacillus thuringiensis* Cry 1A toxin receptor A from *Heliothis virescens* and cloning of the corresponding cDNA. *Appl. Environ. Microbiol. 65*, 4760–4766.

Ortego, F., Novillo, C., Catañera, P., **1996**. Characterization and distribution of digestive proteases of the stalk corn borer, *Sesamia nonagrioides* Lef. (Lepidoptera: Noctuidae). *Arch. Insect Biochem. Physiol. 33*, 136–180.

Pechan, T., Cohen, A., Williams, W.P., Luthe, D.S., **2002**. Insect feeding mobilizes a unique defense protease that disrupts the peritrophic matrix of caterpillars. *Proc. Natl Acad. Sci. USA 99*, 13319–13323.

Peng, J., Zhong, J., Granados, R.R., **1999**. A baculovirus enhancin alters the permeability of a mucosal midgut peritrophic matrix from lepidopteran larvae. *J. Insect Physiol. 45*, 159–166.

Pereira, P.J.B., Lozanov, V., Patthy, A., Huber, R., Bode, W., *et al.*, **1999**. Specific inhibition of insect α-amylases: yellow mealworm α-amylase in complex with the Amaranth at 2.0 Å resolution. *Structure 7*, 1079–1088.

Peters, W., **1992**. Peritrophic Membranes. Springer, Berlin.

Peters, W., Wiese, B., **1986**. Permeability of the peritrophic membranes of some Diptera to labelled dextrans. *J. Insect Physiol. 32*, 43–49.

Peterson, A.M., Barillas-Mury, C.V., Wells, M.A., **1994**. Sequence of three cDNAs encoding an alkaline midgut trypsin from *Manduca sexta*. *Insect Biochem. Mol. Biol. 24*, 463–471.

Peterson, A.M., Fernando, G.J.P., Wells, M., **1995**. Purification, characterization and cDNA sequence of an alkaline chymotrypsin from the midgut of *Manduca sexta*. *Insect Biochem. Mol. Biol. 25*, 765–774.

Pimenta, D.C., Oliveira, A., Juliano, M.A., Juliano, L., **2001**. Substrate specificity of human cathepsin D using internally quenched fluorescent peptides derived from reactive site loop of kallistatin. *Biochim. Biophys. Acta 1544*, 113–122.

Ponsen, M.B., **1991**. Structure of the digestive system of aphids, in particular *Hyalopterus* and *Coloradoa*, and its bearing on the evolution of filter chambers in Aphidoidea. Wageningen Agricultural University Press, Wageningen 91–95, pp. 3–61.

Price, D.J.S., **1963**. Little Science, Big Science. Columbia University Press, New York.

Prosser, W.A., Douglas, A.E., **1992**. A test of the hypotheses that nitrogen is upgraded and recycled in aphid (*Acyrthosiphon pisum*) symbiosis. *J. Insect Physiol. 38*, 93–99.

Prosser, W.A., Simpson, S.J., Douglas, A.E., **1992**. How an aphid (*Acyrthosiphon pisum*) symbiosis responds to variation in dietary nitrogen. *J. Insect Physiol. 38*, 301–307.

Quicke, D.L.J., **2003**. Hymenoptera (ants, bees, wasps). In: Resh, V.H., Cardé, R.T. (Eds.), Encyclopedia of Insects. Academic Press, San Diego, CA, pp. 534–546.

Rajagopal, R., Agrawal, N., Selvapandiyan, A., Sivakumar, S., Ahmad, S., *et al.*, **2003**. Recombinantly expressed isoenzymic aminopeptidases from *Helicoverpa armigera* (American cotton bollworm) midgut display differential interaction with closely related *Bacillus thuringiensis* insecticidal proteins. *Biochem. J. 379*, 971–978.

Ramos, A., Mahowald, A., Jacobs-Lorena, M., **1993**. Gut-specific genes from the black-fly *Simulium vittatum* encoding trypsin-like and carboxypeptidase-like proteins. *Insect Mol. Biol. 1*, 149–163.

Rawlings, N.D., Barrett, A.J., **1995**. Evolutionary families of metallopeptidases. *Methods Enzymol. 248*, 183–228.

Reeck, G., Oppert, B., Denton, M., Kanost, M., Baker, J., *et al.*, **1999**. Insect proteinases. In: Turk, V. (Ed.), Proteases. Birkhäuser Verlag, Basel, pp. 125–148.

Regel, R., Matioli, S.R., Terra, W.R., **1998**. Molecular adaptation of *Drosophila melanogaster* lysozymes to a digestive function. *Insect Biochem. Mol. Biol. 28*, 309–319.

Rouland, C., Civas, A., Renoux, J., Petek, F., **1988**. Purification and properties of cellulases from the termite *Macrotermes mulleri* (Termitidae: Macrotermitinae) and its symbiotic fungus *Termitomyces* sp. *Comp. Biochem. Physiol. B 91*, 449–458.

Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J.K.C., Jones, T.A., **1990**. Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science 249*, 380–386.

Royer, V., Fraichard, S., Bouhin, H., **2002**. A novel putative insect chitinase with multiple catalytic domains: hormonal regulation during metamorphosis. *Biochem. J. 366*, 921–928.

Ryan, C.A., **1990**. Proteinase inhibitors in plants: genes for improving defense against insects and pathogens. *Annu. Rev. Phytopathol. 28*, 425–449.

Ryan, R.O., Van der Horst, D.J., **2000**. Lipid transport biochemistry and its role in energy production. *Annu. Rev. Entomol. 45*, 233–260.

Sakon, J., Irwin, D., Wilson, D.B., Karplus, P.A., **1997**. Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nature Struct. Biol. 4*, 810–818.

Salvucci, M.E., **2003**. Distinct sucrose isomerases catalyze trehalulose synthesis in whiteflies, *Bemisia argentifolii*, and *Erwinia rhapontici*. *Comp. Biochem. Physiol. B 135*, 385–395.

Salvucci, M.E., Rosell, R.C., Brown, J.K., **1998**. Uptake and metabolism of leaf proteins by the silverleaf whitefly. *Arch. Insect Biochem. Physiol. 39*, 155–165.

Santos, C.D., Ribeiro, A.F., Terra, W.R., **1986**. Differential centrifugation, calcium precipitation and ultrasonic disruption of midgut cells of *Erinnyis ello* caterpillars: purification of cell microvilli and inferences concerning secretory mechanisms. *Can. J. Zool. 64*, 490–500.

Santos, C.D., Terra, W.R., **1986a**. Distribution and characterization of oligomeric digestive enzymes from *Erinnyis ello* caterpillars and inferences concerning secretory mechanisms and the permeability of the peritrophic membrane. *Insect Biochem. 16*, 691–700.

Santos, C.D., Terra, W.R., **1986b**. Midgut alpha-glucosidase and beta-fructosidase from *Erinnyis ello* larvae and imagoes: physical and kinetic-properties. *Insect Biochem. 16*, 819–824.

Schechter, I., Berger, A., **1967**. On the size of active site in proteases. *Biochem. Biophys. Res. Commun. 27*, 157–162.

Schorderet, S., Pearson, R.D., Vuocolo, T., Eisemann, C., Riding, G.A., *et al.*, **1998**. cDNA and deduced amino acid sequences of a peritrophic membrane glycoprotein, "Peritrophin-48," from the larvae of *Lucilia cuprina*. *Insect Biochem. Mol. Biol. 28*, 99–111.

Scrivener, A.M., Slaytor, M., **1994**. Properties of the endogenous cellulase from *Panestia cribrata* Saussure and purification of major endo-β-1,4-glucanase components. *Insect Biochem. Mol. Biol. 24*, 223–231.

Scrivener, A.M., Slaytor, M., Rose, H.A., **1989**. Symbiont-independent digestion of cellulose and starch in *Panesthia cribrata* Saussure, an Australian wood-eating cockroach. *J. Insect Physiol. 35*, 935–941.

Scrivener, A.M., Watanabe, H., Noda, H., **1997**. Diet and carbohydrate digestion in the yellow-spotted longicorn beetle *Psacothea hilaris*. *J. Insect Physiol. 43*, 1039–1052.

Sehnal, F., Zitman, D., **1996**. Midgut endocrine cells. In: Lehane, M.J., Billingsley, P.F. (Eds.), Biology of the Insect Midgut. Chapman and Hall, London, pp. 55–85.

Shahabuddin, M., **1995**. Chitinase as a vaccine. *Parasitol. Today 11*, 46–47.

Shen, Z., Jacobs-Lorena, M., **1997**. Characterization of a novel gut-specific chitinase gene from the human malaria vector *Anopheles gambiae*. *J. Biol. Chem. 272*, 28895–28900.

Shen, Z., Jacobs-Lorena, M., **1998**. A type I peritrophic matrix protein from the malaria vector *Anopheles gambiae* binds to chitin: cloning, expression and characterization. *J. Biol. Chem. 273*, 17665–17670.

Shen, Z., Reese, J.C., Reeck, G.R., **1996**. Purification and characterization of polygalacturonase from the rice weevil, *Sitophilus oryzae* (Coleoptera: Curculionidae). *Insect Biochem. Mol. Biol. 26*, 427–433.

Shigenobu, S., Watanabe, H., Hattori, M., Sasaki, Y., Ishikawa, H., **2000**. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature 407*, 81–86.

Shiroya, T., **1963**. Metabolism of raffinose in cotton seeds. *Phytochemistry 2*, 23–46.

Silva, A., Bacci, M., Jr., Siqueira, C.Q., Bueno, O.L., Pagnoca, F.C., *et al.*, **2003**. Survival of *Atta sexdens* workers on different food sources. *J. Insect Physiol. 49*, 307–313.

Silva, C.P., Ribeiro, A.F., Gulbenkian, S., Terra, W.R., **1995**. Organization, origin and function of the outer microvillar (perimicrovillar) membranes of *Dysdercus peruvianus* (Hemiptera) midgut cells. *J. Insect Physiol. 41*, 1093–1103.

Silva, C.P., Terra, W.R., **1994**. Digestive and absortive sites along the midgut of the cotton seed sucker bug *Dysdercus peruvianus* (Hemiptera: Pyrrhocoridae). *Insect Biochem. Mol. Biol. 24*, 493–505.

Silva, C.P., Terra, W.R., **1995**. An α-glucosidase from perimicrovillar membranes of *Dysdercus peruvianus* (Hemiptera: Pyrrhocoridae) midgut cells: purification and properties. *Insect Biochem. Mol. Biol. 25*, 487–494.

Silva, C.P., Terra, W.R., **1997**. Alpha-galactosidase activity in ingested seeds and in the midgut of *Dysdercus peruvianus* (Hemiptera: Pyrrhocoridae). *Arch. Insect Biochem. Physiol. 34*, 443–460.

Silva, C.P., Terra, W.R., de Sá, M.F.G., Samuels, R.I., Isejima, E.M., *et al.*, **2001**. Induction of digestive α-amylases in larvae of *Zabrotes subfasciatus* (Coleoptera: Bruchidae) in response to ingestion of common bean α-amylase inhibitor 1. *J. Insect Physiol. 47*, 1283–1290.

Silva, C.P., Xavier-Filho, J., **1991**. Comparison between the levels of aspartic and cysteine proteinase of the larval midguts of *Callosobruchus maculatus* (F.) and *Zabrotes subfasciatus* (Boh.) (Coleoptera: Bruchidae). *Comp. Biochem. Physiol. B 99*, 529–533.

Silva, I., Souza, V.B.V., **1981**. Ultrastructural aspects of the anterior midgut of *Oncideres s. saga* (Dalman, 1823) larva (Coleoptera: Cerambycidae). *Rev. Brasil. Entomol. 25*, 103–112.

Skibbe, U., Christeller, J.T., Callaghan, P.T., Eccles, C.D., Laing, W.A., **1996**. Visualization of pH gradients in the larval midgut of *Spodoptera litura* using ^{31}P-NMR microscopy. *J. Insect Physiol. 42*, 777–790.

Slaytor, M., **1992**. Cellulose digestion in termites and cockroaches: what role do symbionts play. *Comp. Biochem. Physiol. B 103*, 775–784.

Song, H.K., Suh, S.W., **1998**. Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from *Erythrina caffra* and tissue-type plasminogen activator. *J. Mol. Biol. 275*, 347–363.

Spencer, K.C., **1988**. Chemical mediation of coevolution in the *Passiflora–Heliconius* interaction. In: Spencer, K.C. (Ed.), Chemical Mediation of Coevolution. Academic Press, London, pp. 167–240.

Stewart, C.B., Schilling, J.W., Wilson, A.C., **1987**. Adaptive evolution in the stomach lysozymes of foregut fermenters. *Nature 330*, 401–404.

Strobl, S., Gomis-Ruth, F.X., Maskos, K., Frank, G., Huber, R., *et al.*, **1997**. The alpha-amylase from the yellow mealworm: complete primary structure, crystallization and preliminary X-ray analysis. *FEBS Lett. 409*, 109–114.

Strobl, S., Maskos, K., Betz, M., Wiegand, G., Huber, R., *et al.*, **1998a**. Crystal structure of yellow meal worm α-amylase at 1.64 Å resolution. *J. Mol. Biol. 278*, 617–628.

Strobl, S., Maskos, K., Wiegand, G., Huber, R., Gomis-Rüth, F.X., *et al.*, **1998b**. A novel strategy for inhibition of α-amylases: yellow mealworm α-amylases in complex with the *Ragi* bifunctional inhibitor at 2.5 Å resolution. *Structure 6*, 911–921.

Su, Z.H., Ikeda, M., Sato, Y., Saito, H., Imai, K., *et al.*, **1994**. Molecular characterization of ovary trehalase of the silkworm, *Bombyx mori*, and its transcriptional activation by diapause hormone. *Biochim. Biophys. Acta 1218*, 366–374.

Su, Z.H., Sato, Y., Yamashita, O., **1993**. Purification, cDNA cloning and Northern blot analysis of trehalase of pupal midgut of the silkworm, *Bombyx mori*. *Biochim. Biophys. Acta 1173*, 217–224.

Tellam, R.L., **1996**. The peritrophic matrix. In: Lehane, M.J., Billingsley, P.F. (Eds.), Biology of the Insect Midgut. Chapman and Hall, London, pp. 86–114.

Tellam, R.L., Vuocolo, T., Eisemann, C., Briscoe, S., Riding, G., *et al.*, **2003**. Identification of an immunoprotective mucin-like protein, peritrophin-55, from the peritrophic matrix of *Lucilia cuprina* larvae. *Insect Biochem. Mol. Biol. 33*, 239–252.

Tellam, R.L., Wijffels, G., Willadsen, P., **1999**. Peritrophic matrix proteins. *Insect Biochem. Mol. Biol. 29*, 87–101.

Terra, W.R., **1988**. Physiology and biochemistry of insect digestion: an evolutionary perspective. *Braz. J. Med. Biol. Res. 21*, 675–734.

Terra, W.R., **1990**. Evolution of digestive systems of insects. *Annu. Rev. Entomol. 35*, 181–200.

Terra, W.R., **2001**. The origin and functions of the insect peritrophic membrane and peritrophic gel. *Arch. Insect Biochem. Physiol. 47*, 47–61.

Terra, W.R., **2003**. Digestion. In: Resh, V.H., Cardé, R.T. (Eds.), Encyclopedia of Insects. Academic Press, San Diego, CA, pp. 310–315.

Terra, W.R., Ferreira, C., **1994**. Insect digestive enzymes: properties, compartmentalization and function. *Comp. Biochem. Physiol. B 109*, 1–62.

Terra, W.R., Ferreira, C., **1983**. Further evidence that enzymes involved in the final stages of digestion by *Rhynchosciara americana* do not enter the endoperitrophic space. *Insect Biochem. 13*, 143–150.

Terra, W.R., Ferreira, C., **2003**. Digestive system. In: Resh, V.H., Cardé, R.T. (Eds.), Encyclopedia of Insects. Academic Press, San Diego, CA, pp. 313–323.

Terra, W.R., Ferreira, C., De Bianchi, A.G., **1978**. Physical properties and Tris inhibition of an insect trehalase and a thermodynamic approach to the nature of its active site. *Biochim. Biophys. Acta 524*, 131–141.

Terra, W.R., Ferreira, C., Garcia, E.S., **1988**. Origin, distribution, properties and functions of the major *Rhodnius prolixus* midgut hydrolases. *Insect Biochem. 18*, 423–434.

Terra, W.R., Ferreira, C., Jordão, B.P., Dillon, R.J., **1996**. Digestive enzmes. In: Lehane, M.J., Billingsley, P.F. (Eds.), Biology of the Insect Midgut. Chapman and Hall, London, pp. 153–194.

Terra, W.R., Regel, R., **1995**. pH buffering in *Musca domestica* midguts. *Comp. Biochem. Physiol. A 112*, 559–564.

Terra, W.R., Terra, I.C.M., Ferreira, C., **1983**. Inhibition of an insect midgut trehalase by dioxane and δ-gluconolactone: enzyme pK_a values and geometric relationships at the active site. *Int. J. Biochem. 15*, 143–146.

Terra, W.R., Terra, I.C.M., Ferreira, C., de Bianchi, A.G., **1979**. Carbodiimide-reactive carboxyl groups at the active site of an insect midgut trehalase. *Biochim. Biophys. Acta 571*, 79–85.

Theischinger, G., **1991**. Megaloptera. In: Naumann, I.D., Carne, P.B., Lawrence, J.F., Nielsen, E.S., Spradbery, J.P., Taylor, R.W., Whitten, M.J., Littlejohn, M.J. (Eds.), The Insects of Australia. 2nd edn. Melbourne University Press, Melbourne, pp. 516–520.

Titarenko, E., Chrispeels, M.J., **2000**. cDNA cloning, biochemical characterization and inhibition by plant inhibitors of the alpha-amylases of the Western corn rootworm, *Diabrotica virgifera virgifera*. *Insect Biochem. Mol. Biol. 30*, 979–990.

Tokuda, G., Lo, N., Watanabe, H., Slaytor, M., Matsumoto, T., *et al.*, **1999**. Metazoan cellulase genes from termites: intron/exon structures and sites of expression. *Biochim. Biophys. Acta 1447*, 146–159.

Tokuda, G., Saito, H., Watanabe, H., **2002**. A digestive β-glycosidase from the salivary glands of the termite, *Neotermes koshunensis* (Shiraki): distribution, characterization and isolation of its precursor cDNA by 5′- and 3′-RACE amplifications with degenerate primers. *Insect Biochem. Mol. Biol. 32*, 1681–1689.

Tokuda, G., Watanabe, H., Matsumoto, T., Noda, H., **1997**. Cellulose digestion in the wood-eating higher termite, *Nasutitermes takasagoensis* (Shiraki): distribution of cellulases and properties of endo-β-1,4-glucanase. *Zool. Sci. 14*, 83–93.

Tsuchida, K., Wells, M.A., **1988**. Digestion, absorption, transport and storage of fat during the last stadium of *Manduca sexta*: changes in the role lipophorin in the delivery of dietary lipids to the fat body. *Insect Biochem. 18*, 263–268.

Turunen, S., **1993**. Metabolic pathways in the midgut epithelium of *Pieris brassicae* during carbohydrate and lipid assimilation. *Insect Biochem. Mol. Biol. 23*, 681–689.

Turunen, S., Kastari, T., **1979**. Digestion and absorption of lecithin in larvae of the cabbage butterfly, *Pieris brassicae*. *Comp. Biochem. Physiol. A 62*, 933–937.

Uscian, J.M., Miller, J.S., Sarath, G., Stanley-Samuelson, D.W., **1995**. A digestive phospholipase A_2 in the tiger beetle *Cincindella circumpicta*. *J. Insect Physiol. 41*, 135–141.

Valaitis, A.P., **1995**. Gypsy moth midgut proteinases: purification and characterization of luminal trypsin, elastase and the brush-border membrane leucine aminopeptidase. *Insect Biochem. Mol. Biol. 25*, 139–149.

Valaitis, A.P., Bowers, D.F., **1993**. Purification and properties of the soluble midgut trehalase from the gypsy moth, *Lymantria dispar*. *Insect Biochem. Mol. Biol. 23*, 599–606.

Valaitis, A.P., Mazza, A., Brousseau, R., Masson, L., **1997**. Interaction analyses of *Bacillus thuringiensis* Cry 1A toxins with two aminopeptidases from gipsy moth midgut brush border membranes. *Insect Biochem. Mol. Biol. 27*, 529–539.

Vetter, J., **2000**. Plant cyanogenic glycosides. *Toxicon 38*, 11–36.

Volpicella, M., Ceci, L.R., Cordewenwe, J., America, T., Gallerani, R., *et al.*, **2003**. Properties of purified gut trypsin from *Helicoverpa zea*, adapted to proteinase inhibitors. *Eur. J. Biochem. 270*, 10–19.

Vonk, H.J., Western, J.R.H., **1984**. Comparative Biochemistry and Physiology of Enzymatic Digestion. Academic Press, New York.

Wang, P., Granados, R.R., **1997**. Molecular cloning and sequencing of a novel invertebrate intestinal mucin. *J. Biol. Chem. 272*, 16663–16669.

Wang, P., Granados, R.R., **2000**. Calcofluor disrupts the midgut defense system in insects. *Insect Biochem. Mol. Biol. 30*, 135–143.

Wang, P., Granados, R.R., **2001**. Molecular structure of the peritrophic membrane (PM): identification of

potential PM target sites for insect control. *Arch. Insect Biochem. Physiol. 47*, 110–118.

Ward, C.W., **1975a**. Aminopeptidases in webbing clothes moth larvae: properties and specificities of the enzymes of intermediate electrophoretic mobility. *Biochim. Biophys. Acta 410*, 361–369.

Ward, C.W., **1975b**. Aminopeptidases in webbing clothes moth larvae: properties and specificities of enzymes of highest electrophoretic mobility. *Austral. J. Biol. Sci. 28*, 447–455.

Watanabe, H., Nakamura, M., Tokuda, G., Yamaoka, I., Scrivener, A.M., *et al.*, **1997**. Site of secretion and properties of endogenous endo-β-1,4-glucanase components from *Reticulitermes speratus* (Kolbe), a Japanese subterranean termite. *Insect Biochem. Mol. Biol. 27*, 305–313.

Watanabe, H., Noda, H., Tokuda, G., Lo, N., **1998**. A cellulase gene of termite origin. *Nature 394*, 330–331.

Weiher, B., Komnick, H., **1997**. Digestion of phosphatidylcholines, absorption, and esterification of lipolytic products by *Aeshna cyanea* larvae as studied *in vivo* and *in vitro*. *Arch Insect Biochem. Physiol. 36*, 273–293.

Whitworth, S.T., Blum, M.S., Travis, J., **1998**. Proteolytic enzymes from larvae of the fire ant, *Solenopsis invicta*: isolation and characterization of four serine endopeptidases. *J. Biol. Chem. 273*, 14430–14434.

Whitworth, S.T., Kordula, T., Travis, J., **1999**. Molecular cloning of Soli EC: an elastase-like serine proteinase from the imported red fire ant (*Solenopsis invicta*). *Insect Biochem. Mol. Biol. 29*, 249–254.

Withers, S.G., Rupitz, K., **1990**. Measurement of active-site homology between potato and rabbit muscle α-glucan phosphorylases through use of a free energy relationship. *Biochemistry 29*, 6405–6409.

Wolfersberger, M.G., **2000**. Amino acid transport in insects. *Annu. Rev. Entomol. 45*, 111–120.

Wolfersberger, M.G., Spaeth, D.D., Dow, J.A.T., **1986**. Permeability of the peritrophic membrane of tobacco hornworm larval midgut. *Am. Zool. 26*, 356.

Woods, H.A., Kingsolver, J.G., **1999**. Feeding rate and the structure of protein digestion and absorption in Lepidoptera midguts. *Arch. Insect Biochem. Physiol. 42*, 74–87.

Yaoi, K., Nakanishi, K., Kadotani, T., Imamura, M., Koizumi, N., *et al.*, **1999**. cDNA cloning and expression of *Bacillus thuringiensis* Cry 1 Aa toxin binding 120 kDa aminopeptidase N from *Bombyx mori*. *Biochim. Biophys. Acta 1444*, 131–137.

Zeng, F., Zhu, Y.C., Cohen, A.C., **2002**. Molecular cloning and partial characterization of a trypsin-like protein in salivary glands of *Lygus hesperus* (Hemiptera: Miridae). *Insect Biochem. Mol. Biol. 32*, 455–464.

Zhang, H., Huang, X., Fukamizo, T., Muthukrishnan, Kramer, K.J., **2002**. Site-directed mutagenesis and functional analysis of an active site tryptophan of insect chitinase. *Insect Biochem. Mol. Biol. 32*, 1477–1488.

Zheng, L., Whang, L.H., Kumar, V., Kafatos, F.C., **1995**. Two genes encoding midgut-specific maltase-like polypeptides from *Anopheles gambiae*. *Exp. Parasitol. 81*, 272–283.

Zhu, Y.C., Baker, J.E., **1999**. Characterization of midgut trypsin-like enzymes and three trypsinogen cDNAs from the lesser grain borer, *Rhyzopertha dominica* (Coleoptera: Bostrichidae). *Insect Biochem. Mol. Biol. 29*, 1053–1063.

Zhu, Y.C., Baker, J.E., **2000**. Molecular cloning and characterization of a midgut chymotrypsin-like enzyme from the lesser grain borer, *Rhyzopertha dominica*. *Arch. Insect Biochem. Physiol. 43*, 173–184.

Zhu, Y.C., Kramer, K.K., Oppert, B., Dowdy, A.K., **2000**. cDNAs of aminopeptidase-like protein genes from *Plodia interpunctella* strains with different susceptibilities to *Bacillus thuringiensis* toxins. *Insect Biochem. Mol. Biol. 30*, 215–224.

Zhuzhikov, D.P., **1964**. Function of the peritrophic membrane in *Musca domestica* L. and *Calliphora erythrocephala* Meig. *J. Insect Physiol. 10*, 273–278.

Ziegler, H., Garon, C.F., Fischer, E.R., Shahabuddin, M., **2000**. A tubular network associated with the brush-border surface of the *Aedes aegypti* midgut: implications for pathogen transmission by mosquitoes. *J. Exp. Biol. 203*, 1599–1611.

Zimoch, L., Merzendorfer, H., **2002**. Immunolocalization of chitin synthase in the tabacco hornworm. *Cell Tissue Res. 308*, 287–297.

Relevant Websites

http://www.chem.gmw.ac.uk/iubmb – International Union of Biochemistry and Molecular Biology, Enzyme Commission.

http://www.merops.ac.uk – MEROPS.

4.6 Lipid Transport

D J Van der Horst, Utrecht University, Utrecht, The Netherlands
R O Ryan, Children's Hospital Oakland Research Institute, CA, USA

© 2005, Elsevier BV. All Rights Reserved.

4.6.1. Historical Perspective

4.6.1.1. Lipophorin Structure and Morphology

Lipophorin was discovered 40 years ago as a major hemolymph component and key transport vehicle for water insoluble metabolites (Beenakkers *et al.*, 1985; Chino, 1985). Lipophorin is generally regarded as a multifunctional carrier because it displays a broad ability to accommodate hydrophobic biomolecules. In essence, lipophorin can be described as a noncovalent assembly of lipids and proteins, organized as a largely spherical particle. The core of the particle is made up of hydrophobic lipid molecules, such as diacylglycerol (DAG), hydrocarbons, and carotenoids. DAG, which serves as the transport form of neutral glycerolipid in hemolymph, provides an energy source for various tissues through oxidative metabolism of its fatty acid constituents. Hydrocarbons, in the form of long-chain aliphatic alkanes and alkenes, are extremely hydrophobic lipid molecules that are deposited on the cuticle where they serve to prevent desiccation and may function as semiochemicals. Carotenoids are plant-derived pigments used for coloration and as a precursor to visual pigments (Canavoso *et al.*, 2001). Another important lipid component of lipophorin is phospholipid. In general, the major glycerophospholipids present are phosphatidylcholine and phosphatidylethanolamine (Wang *et al.*, 1992). These amphiphilic lipids exist as a monolayer at the lipophorin particle surface, positioned in such a way that their fatty acyl chains interact with the hydrophobic core of the particle while their polar head groups are presented to the aqueous milieu. In this manner, the phospholipid moieties of lipophorin serve a key structural role. The other major structural component of lipophorin is protein. All lipophorin particles possess two apolipoproteins, termed apolipophorin I (apoLp-I) and apolipophorin II (apoLp-II). ApoLp-I and apoLp-II are integral components of the lipophorin particle and cannot be removed without destruction of lipophorin particle integrity. It is recognized that apoLp-I and apoLp-II are the product of the same gene and that the two proteins arise from a post-translational proteolytic processing event (Weers *et al.*, 1993). This finding is consistent with the fact that apoLp-I and apoLp-II are found in a 1 : 1

molar ratio in lipophorin particles. At present it is not known if one or the other apoLp possesses additional functions aside from its primary role in stabilizing lipophorin particle integrity although a role in receptor interactions is implied (see below). The structural role is fulfilled by the capacity of apoLp-I and apoLp-II to interact with lipid and create an interface between the nonpolar core of the particle and the external environment. In this capacity, apoLp-I and apoLp-II function in a manner similar to that proposed for apolipoprotein B in vertebrate plasma. Indeed, it is now recognized that the genes encoding these proteins are derived from a common ancestor (Babin *et al.*, 1999).

4.6.1.2. Lipophorin Subspecies

One of the hallmark features of lipophorin-mediated lipid transport relates to the dynamic nature of the particle. Lipophorin isolated from various life stages is generally of a unique density and lipid composition. For example, lipophorin from *Manduca sexta* fifth instar larvae displays a density of $1.15\,g\,ml^{-1}$ with a particle diameter in the range of 16 nm. By contrast, lipophorin isolated from adult hemolymph is of lower density and larger diameter. Indeed, in *M. sexta* a broad array of unique lipophorin subspecies has been identified, each with characteristic properties (Prasad *et al.*, 1986). On the basis of this diversity, a nomenclature system has been adopted that distinguishes various lipophorin subspecies based on their density. Since most particles fall within the density limits 1.21 and $1.07\,g\,ml^{-1}$, the term high-density lipophorin (HDLp) is commonly used. Because many lipophorin subspecies are present at well-defined developmental stages, a suffix may be added to denote this. Hence, HDLp-P and HDLp-A may be used to distinguish HDLp from pupal and adult hemolymphs, respectively.

One of the features of HDLp-A is its ability to associate with a third apolipophorin, apoLp-III. In insect species that use lipid as a fuel for flight (such as *Locusta migratoria* and *M. sexta*), apoLp-III is present in abundance in adult hemolymph as a lipid free protein. Whereas a small amount of apoLp-III may be associated with HDLp under resting conditions, flight activity induces association of large amounts of apoLp-III with the lipophorin particle surface (Van der Horst *et al.*, 1979). This process, which is dependent upon the uptake of DAG by the lipophorin particle, leads to the conversion of HDLp into low-density lipophorin (LDLp). LDLp has a larger diameter, a significantly increased DAG content, and a lower density. In studies of this conversion, it has been shown that apoLp-III associates with the surface of the expanding lipophorin particle as a function of DAG enrichment (Soulages and Wells, 1994b; Ryan and Van der Horst, 2000). Thus, it has been hypothesized that apoLp-III serves to stabilize the DAG-enriched particle, providing an interface between surface localized hydrophobic DAG molecules and the external aqueous medium. It is envisioned that continued DAG accumulation by HDLp results in partitioning of DAG between the hydrophobic core of the particle and the surface monolayer (Wang *et al.*, 1995). The presence of DAG in the surface monolayer exerts a destabilizing effect on the particle structure and, if allowed to persist, would result in deleterious particle fusion and aggregation. By "sensing" the presence of DAG in the lipophorin surface monolayer, apoLp-III is attracted to the particle surface and forms a stable binding interaction. This event is fully reversible and, upon removal of DAG from the particle, apoLp-III dissociates, leading to regeneration of HDLp. Importantly, it is recognized that lipophorin particles can then bind additional DAG, forming a cycle of transport. It is noteworthy that these concepts about apoLp-III association/dissociation from lipophorin emerged from physiological studies of flight activity in *L. migratoria* conducted in the late 1970s and early 1980s in The Netherlands and England (Mwangi and Goldsworthy, 1977, 1981; Van der Horst *et al.*, 1979, 1981). A cartoon depicting metabolic and biochemical processes related to induction of flight-related lipophorin conversions and the accompanying increase in neutral lipid transport capacity is presented in **Figure 1**. Elaboration of various aspects of this central scheme will occur in the next sections. At this point, however, it should be noted that this generalized mechanism differs fundamentally from metabolic processes in vertebrates, where lipoproteins do not have a function in the transport of energy substrates during exercise (Van der Horst *et al.*, 2002). That being said, it is evident that novel insight into structural and functional aspects of vertebrate lipid transport processes can be gained from the study of insect lipid transport. A vivid example of this emerged recently with the identification and functional characterization of the *Drosophila* homolog of the vertebrate microsomal lipid transfer protein (MTP) (Sellers *et al.*, 2003). Analysis of the *Drosophila* genome revealed the existence of an expressed sequence tag with 23% sequence identity to vertebrate MTP. Coexpression of this gene with that encoding a truncated version of apolipoprotein B in African green monkey kidney (COS) cells afforded the ability to assemble and secrete lipoproteins. Compared to mammalian MTP, the *Drosophila* MTP homolog possesses unique structural and catalytic properties that appear

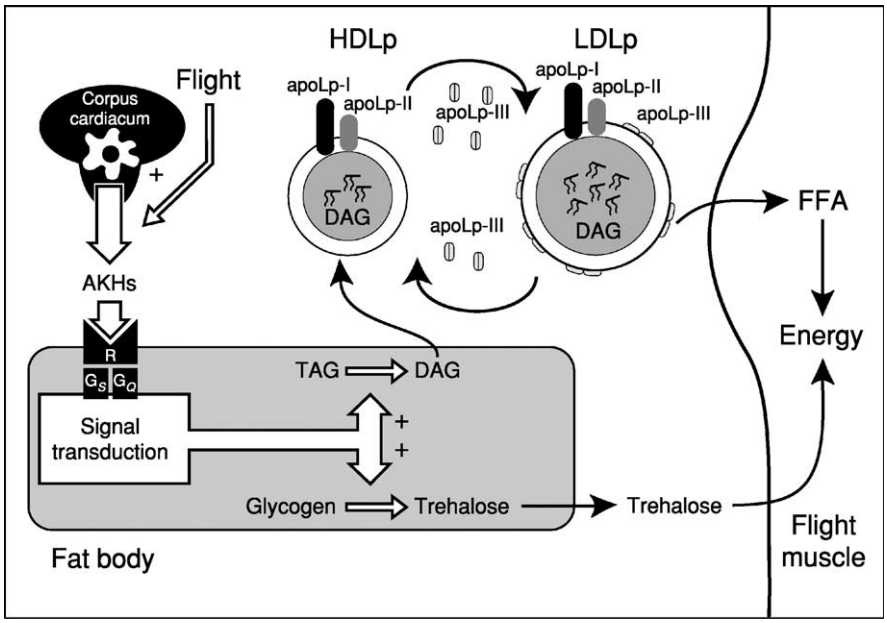


Figure 1 Molecular basis of the lipophorin lipid shuttle: AKH-controlled DAG mobilization from insect fat body during flight activity results in the reversible alternation of lipophorin from a relatively lipid-poor (HDLp) to a lipid-rich (LDLp) state, and apoLp-III from a lipid-free in a lipid-bound state. The reversible conformational change in apoLp-III induced by DAG loading of lipophorin is schematically visualized. AKHs, adipokinetic hormones; R, receptor; G, G protein; HDLp, high-density lipophorin; LDLp, low-density lipophorin; apoLp-I, -II, and -III, apolipophorin I, II, and III; TAG, triacylglycerol; DAG, diacylglycerol; FFA, free fatty acids. (Based on data from several insect species, particularly *Locusta migratoria* and *Manduca sexta*, reviewed in Ryan and Van der Horst (2000) and Van der Horst *et al.* (2001); and this review.)

to be related to distinct strategies of lipid transport operative in insects.

4.6.2. Flight-Related Processes

4.6.2.1. Adipokinetic Hormone

Insect flight involves the mobilization, transport, and utilization of endogenous energy reserves at extremely high rates. In insects that engage in long-distance flight, the demand for fuel, particularly lipids, by the flight muscles can remain elevated for extended periods of time. Peptide adipokinetic hormones (AKHs), synthesized and stored in neuroendocrine cells, play a crucial role in this process as they integrate flight energy metabolism. Insect AKHs are short peptides consisting of 8–11 amino acid residues. To date the structures of over 35 different AKHs are known from representatives of most insect orders; in spite of considerable variation in their structures they are clearly related (reviews: Gäde, 1997; Van der Horst *et al.*, 2001; Oudejans and Van der Horst, 2003). All AKHs are N-terminally blocked by a pyroglutamate (pGlu) residue and all but one (Köllisch *et al.*, 2000) are C-terminally amidated. Initiation of flight activity induces the release of AKHs from the intrinsic AKH-producing cells (adipokinetic cells) in the glandular lobes of the corpus cardiacum, a neuroendocrine gland located caudal to the insect brain and physiologically equivalent to the pituitary of mammals. The fat body plays a fundamental role in lipid storage, as well as in the process of lipolysis controlled by the AKHs. Binding of these hormones to their G protein-coupled receptors at the fat body target cells triggers a number of coordinated signal transduction processes that ultimately result in the mobilization of carbohydrate and lipid reserves as fuels for flight activity (see **Figure 1**). Energy-yielding metabolites are transported via the hemolymph to the contracting flight muscles. Carbohydrate (trehalose) in the circulation provides energy for the initial period of flight and is replenished from glycogen reserves. However, similar to sustained activity in many other animal species, flight activity of insects covering vast distances nonstop is powered principally by mobilization of endogenous reserves of triacylglycerol (TAG), the most concentrated form of energy available to biological tissues. As a result of TAG mobilization, the concentration of *sn*-1,2-DAG in the hemolymph increases progressively and gradually constitutes the principal fuel for flight. The mechanism for hormonal activation of glycogen phosphorylase, the enzyme determining

the rate of glycogen breakdown and trehalose biosynthesis, has been well established. In contrast, little is known of the mechanism by which the pivotal enzyme TAG lipase catalyzes AKH-controlled production of the DAG on which long-distance flight is dependent.

For a considerable part, the success of insects in long-distance flights is attributable to their system of neuropeptide AKHs integrating flight energy metabolism, involving the transfer of energy substrates, particularly lipids, to the flight muscles as discussed above. Therefore, recent advances in the strategy of adipokinetic cells in hormone storage and release will be discussed along with the effects of the AKHs on lipid mobilization.

4.6.2.2. Strategy of the Adipokinetic Cells

In view of their involvement in the regulation and integration of extremely intense metabolic processes, the AKH-producing cells (adipokinetic cells) of the corpus cardiacum constitute an appropriate model system for studying neuropeptide biosynthesis and processing, as well as the coherence between biosynthesis, storage, and release of these neurohormones (reviews: Ryan and Van der Horst, 2000; Van der Horst *et al*., 2001; Diederen *et al*., 2002). These processes have been particularly studied in two locust species notorious for their long-range flight capacity, *L. migratoria* and *Schistocerca gregaria*, which similarly to several other insect species mobilize more than one AKH. The three AKHs synthesized in the adipokinetic cells of *L. migratoria* consist of a decapeptide AKH-I and two octapeptides (AKH-II and -III). AKH-I is by far the most abundant peptide; the ratio of AKH-I : -II : -III in the corpus cardiacum is approximately 14 : 2 : 1 (Oudejans *et al*., 1993). All three AKHs are involved in the mobilization of both lipids and carbohydrates, although their action is differential (reviews: Vroemen *et al*., 1998; Van der Horst and Oudejans, 2003). In addition, several other effects of AKH are known, such as inhibition of the synthesis of proteins, fatty acids, and RNA (reviews: Gäde, 1996; Gäde *et al*., 1997).

The transport of these hydrophobic peptides in the circulation occurs independently of a carrier (Oudejans *et al*., 1996). The AKHs of *L. migratoria* appear to be catabolized differentially after their release; turnover half-times of AKH-I and -II during flight are relatively slow (35 and 37 min, respectively), whereas the hemolymph half-time of AKH-III is very rapid (3 min) (Oudejans *et al*., 1996). Degradation of the (single) AKH in the hemolymph of adult females of the cricket *Gryllus bimaculatus*, which do not fly well, was estimated to be remarkably short (half-life approximately 3 min) in the resting state (Woodring *et al*., 2002). A recent study in which AKH concentrations were measured by radioimmunoassay shows that the hemolymph concentration of two AKHs from *S. gregaria* (AKH-I and -II) increases within 5 min of initiation of flight and are maintained at approximately 15-fold (AKH-I) and 6-fold (AKH-II) the resting levels over flights of at least 60 min (Candy, 2002). The increase in hormone level preceded an increase in hemolymph lipid content. Furthermore, a rapid release of the AKHs over the first few minutes was followed by a slower release, maintaining the elevated hormone levels.

The AKH peptides are derived from preprohormones that are translated from separate mRNAs and subsequently enzymatically processed. Cotranslational cleavage of the signal sequences generates the AKH-I, -II, and -III prohormones, consisting of a single copy of AKH, a GKR or GRR processing site, and an AKH-associated peptide (AAP). AKH-I and -II prohormones are structurally very similar whereas AKH-III is remarkably different (Bogerd *et al*., 1995) (**Figure 2**). Prior to further processing, the AKH-I and -II prohormones dimerize at random by oxidation of their (single) cysteine residues in the AAP, giving rise to two homodimers and one heterodimer. Proteolytic processing of these dimeric products at their processing sites, involving removal of the two basic amino acid residues and amidation, using glycine as the donor, yields the bioactive hormones as well as three (two homodimeric and one heterodimeric) AKH-precursor related peptides (APRPs) with as yet unknown functions (reviews: Van der Horst *et al*., 2001; Diederen *et al*., 2002; Oudejans and Van der Horst, 2003). Recent data from capillary liquid chromatography-tandem mass spectrometry analysis indicate that these APRPs are further processed to form smaller peptides, designed AKH joining peptide 1 (AKH-JP I) and 2 (AKH-JP II), respectively (Baggerman *et al*., 2002) (**Figure 2**). The biosynthesis of AKH-III from its prohormone has only very recently been disclosed (Huybrechts *et al*., 2002). By the use of sophisticated techniques including capillary high-performance liquid chromatography (HPLC) and nanoflow electrospray ionization quantitative time-of-flight (Q-TOF) mass spectrometry, another (fourth) APRP was identified, a homodimer resulting from the cross-linking of two AKH-III prohormone molecules (in a parallel and/or antiparallel fashion) by two disulfide bridges formed between their (two) cysteine residues and subsequent proteolytic cleavage of the AKH-III molecules. This finding indicates that

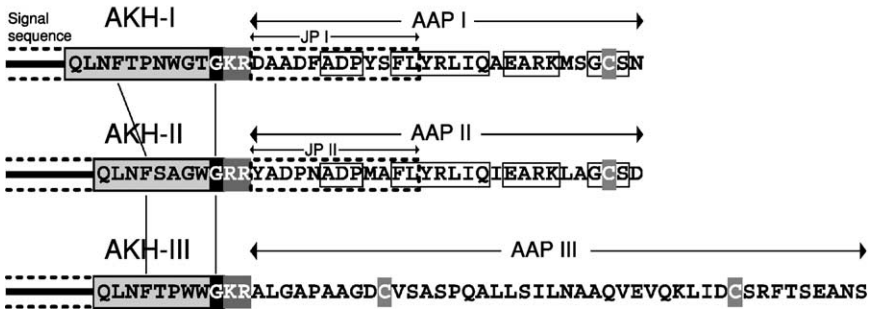


Figure 2 Sequence and proteolytic processing of *Locusta migratoria* AKH prohormones. The AKH sequence is followed by a processing site (GKR or GRR); identical residues in the AKH-associated peptides (AAPs) I and II are boxed. The cysteine residues forming disulfide bridges prior to proteolytic processing of all AKH prohormones are shown in white. JP, joining peptide. (Based on data from Bogerd, J., Kooiman, F.P., Pijnenburg, M.A.P., Hekking, L.H.P., Oudejans, R.C.H.M., *et al.*, **1995**. Molecular cloning of three distinct cDNAs, each encoding a different adipokinetic hormone precursor, of the migratory locust, *Locusta migratoria*: differential expression of the distinct adipokinetic hormone precursor genes during flight activity. *J. Biol. Chem. 270*, 23038–23043; Baggerman, G., Huybrechts, J., Clynen, E., Hens, K., Harthoorn, L., *et al.*, **2002**. New insights in adipokinetic hormone (AKH) precursor processing in *Locusta migratoria* obtained by capillary liquid chromatography-tandem mass spectrometry. *Peptides 23*, 635–644; and Huybrechts, J., Clynen, E., Baggerman, G., De Loof, A., Schoofs, L., **2002**. Isolation and identification of the AKH III precursor-related peptide from *Locusta migratoria*. *Biochem. Biophys. Res. Commun. 296*, 1112–1117.)

processing of AKH-III prohormone occurs similar to that of the AKH-I and -II prohormones. In contrast to the APRPs derived from AKH-I and -II prohormones, however, no evidence was found for further processing of the APRP generated along with AKH-III production.

In situ hybridization showed that the mRNA signals encoding the three different AKH preprohormones are colocalized in the cell bodies of the glandular lobes of the corpus cardiacum (Bogerd *et al.*, 1995). Following their synthesis in the rough endoplasmic reticulum in the cell bodies, the AKH prohormones are transported to the Golgi complex and packaged into secretory granules at the *trans*-Golgi network, whereas proteolytic processing of the prohormones to bioactive AKHs is presumed to take place in the secretory granules (reviews: Van der Horst *et al.*, 2001; Diederen *et al.*, 2002; Oudejans and Van der Horst, 2003). The intracellular location of the AKHs was probed with antibodies specific for the corresponding associated peptides (AAP I, II, and III), the amino acid sequences of which differ to a larger degree from each other than those of the AKHs. All three (dimeric) AAPs were shown to be colocalized in the same secretory granules, which implies that these three AKHs colocalize in these granules and are released simultaneously during flight (Harthoorn *et al.*, 1999). Since the membranes of exocytosed secretory granules fuse with the plasma membrane, the total content of the granules is released into the hemolymph. Consequently, in addition to bioactive AKHs, the APRPs, and possibly other products, are released. Whether the AKH-JPs are released is not yet clear (Baggerman *et al.*, 2002; Huybrechts *et al.*, 2002); data on AKH-JP I and II indicate that these peptides do not stimulate lipid release from the fat body nor activate fat body glycogen phosphorylase, both key functions of the AKHs (Baggerman *et al.*, 2002).

The AKH cells continuously synthesize AKHs, resulting in a steady increase in the amounts of the three hormones in the corpus cardiacum with age. Concurrently, the number of the AKH-containing secretory granules (diameter ~300 nm) also increases. In addition, particularly in older adults, intracisternal granules (ICGs) are produced. ICGs are present in both exocrine and endocrine cells, and originate from premature condensation of peptidergic products within cisternae of the rough endoplasmic reticulum (review: Diederen *et al.*, 2002). In the locust adipokinetic cells, these granules, which may attain diameters up to 5 µm and even more, appear to function as a store for AKH-I and -II prohormones, as shown immunocytochemically with specific anti-AAPs (Harthoorn *et al.*, 1999, 2000). The prohormone for AKH-III is absent, which points to differences in physiological function between AKH-III and the other two AKHs.

The secretory activity of the adipokinetic cells, which has been investigated *in vitro* primarily for AKH-I, is subject to many regulatory substances including neurogenic locustatachykinins and humoral crustacean cardioactive peptide (CCAP) as initiating factors, trehalose as an inhibitor, and several positive and negative modulators (reviews: Van der Horst *et al.*, 1999; Vullings *et al.*, 1999). Recent data on the release of AKH from the corpora cardiaca *in vitro* show that regulatory substances

(including CCAP) affect the release of all three AKHs in proportion to their concentration in the corpus cardiacum (Harthoorn *et al*., 2001). However, the only natural stimulus for the release of the AKHs is flight activity, and the relative contributions of all known substances effective in the process of release of these neurohormones remain to be established *in vivo*.

The amount of AKHs released during flight represents only a few percent of the huge stores harbored in the adipokinetic cells. On the other hand, only a limited part of these AKH stores appear to be actually releaseable. In studies in which young secretory granules were specifically labeled, these newly formed secretory granules were preferentially released (last in, first out) (reviews: Van der Horst *et al*., 2001; Diederen *et al*., 2002; Oudejans and Van der Horst, 2003). Following the biosynthesis of new AKH prohormones, their packaging into secretory granules and their processing to bioactive AKHs, which takes less than 1 h, granules containing newly synthesized AKHs appeared to be available for release during a restricted period of approximately 8 h before they are supposed to enter a pool of older secretory granules that appear to be unable to release their content upon secretory stimulation. This indicates that only a relatively small readily releasable pool of new secretory granules exists. Therefore, an important question is whether the secretory output of AKHs during flight would induce a stimulation of the rate of AKH biosynthesis. The mRNA levels of all three AKH preprohormones, however, did not appear to be affected by flight activity, while the rate of synthesis of AKH prohormones and AKHs was not affected either (Harthoorn *et al*., 2001). Apparently, a coupling between release and biosynthesis of AKHs is absent. Inhibition of AKH biosynthesis *in vitro* by Brefeldin A, a specific blocker of the transport of newly synthesized secretory proteins from the endoplasmic reticulum to the Golgi complex, resulted in a considerable decrease in the release of AKHs induced by CCAP, and highlighted once more that the regulated secretion of AKHs is completely dependent on the existence of a readily releasable pool of newly formed secretion granules (Harthoorn *et al*., 2002). Therefore, we conclude that the strategy of the adipokinetic cells to cope with variations in secretory output of AKHs apparently is to rely on the continuous biosynthesis of AKHs, which produces a readily releasable pool that is sufficiently large and constantly replenished.

An important question remaining unanswered is, what might be the rationale for the storage of such large quantities of hormones that are not accessible for secretory release? In addition, the possible function of the prohormones for AKH-I and -II in the ICGs in providing an additional source of AKH prohormones when called upon remains to be established.

4.6.2.3. Effect of Adipokinetic Hormones on Lipid Mobilization

Binding of the AKHs to their plasma membrane receptor(s) at the fat body cells is the primary step to the induction of signal transduction events that ultimately lead to the activation of target key enzymes and the mobilization of lipids as a fuel for flight. Although the AKHs constitute extensively studied neurohormones and their actions have been shown to occur via G protein-coupled receptors (reviews: Van Marrewijk and Van der Horst, 1998; Vroemen *et al*., 1998), the general properties of which are remarkably well conserved during evolution (review: Vanden Broeck, 2001), the identification of these receptors has not been successful. Very recently, however, the first insect AKH receptors have been identified at the molecular level, namely those of the fruitfly *Drosophila melanogaster* and the silkworm *Bombyx mori* (Staubli *et al*., 2002). They appear to be structurally related to mammalian gonadotropin-releasing hormone (GnRH) receptors. These data promise to elucidate the nature of AKH receptors from other insects; it is envisaged that insects such as the locust, that produce two or more different types of AKH, may have two or more different AKH receptors.

The signal transduction mechanism of the three locust AKHs has been studied extensively, and involves stimulation of cAMP production, which is dependent on the presence of extracellular Ca^{2+}. Additionally, the AKHs enhance the production of inositol phosphates including inositol 1,4,5-trisphosphate (IP_3), which may mediate the mobilization of Ca^{2+} from intracellular stores. This depletion of Ca^{2+} from intracellular stores stimulates the influx of extracellular Ca^{2+}, indicative of the operation of a capacitative (store-operated) calcium entry mechanism. The interactions between the AKH signaling pathways ultimately result in mobilization of stored reserves as fuel for flight (reviews: Van Marrewijk and Van der Horst, 1998; Vroemen *et al*., 1998; Ryan and Van der Horst, 2000; Van der Horst *et al*., 2001; Van Marrewijk, 2003). The concentration of DAG in the hemolymph increases progressively at the expense of stored TAG reserves in the fat body, which implies hormonal activation of the key enzyme, fat body TAG lipase. In a bioassay, all three AKHs are able to stimulate lipid mobilization, although their

relative potencies are different. This recalls the concept of a hormonally redundant system involving multiple regulatory molecules with overlapping actions (reviews: Goldsworthy *et al*., 1997; Vroemen *et al*., 1998). Results obtained with combinations of two or three AKHs, which are likely to occur together in locust hemolymph under physiological conditions *in vivo*, revealed that the maximal responses for the lipid-mobilizing effects were much lower than the theoretically calculated responses based on dose-response curves for the individual hormones. In the lower (probably physiological) range, however, combinations of the AKHs were more effective than the theoretical values calculated from the responses elicited by the individual hormones (review: Van Marrewijk and Van der Horst, 1998).

The mechanism by which TAG lipase catalyzes AKH-controlled production of the DAG on which long-distance flight depends is only poorly understood, mainly due to technical problems in isolating or activating the lipase. In vertebrates, hormone-sensitive lipase (HSL) controls mobilization of TAG stores in adipose tissue, and although contrary to insects, free fatty acids (FFA) are released into the blood for uptake and oxidation in muscle, there is a clear functional similarity between vertebrate adipose tissue HSL and insect fat body TAG lipase (reviews: Ryan and Van der Horst, 2000; Van der Horst *et al*., 2001; Van der Horst and Oudejans, 2003).

4.6.3. Apolipophorin III

4.6.3.1. Lipid Free Helix Bundle Structure

ApoLp-III was discovered in the late 1970s and early 1980s by research groups in Europe and North America (reviews: Blacklock and Ryan, 1994; Ryan and Van der Horst, 2000). ApoLp-III was first isolated from hemolymph of *L. migratoria* (Van der Horst *et al*., 1984) and the tobacco hawkmoth, *M. sexta* (Kawooya *et al*., 1984). *Manduca sexta* apoLp-III is a 166 amino acid protein that lacks tryptophan and cysteine (Cole *et al*., 1987). However, the well-characterized apoLp-III from *L. migratoria* is 164 residues long and lacks cysteine, methionine, and tyrosine (Kanost *et al*., 1988; Smith *et al*., 1994). *Manduca sexta* apoLp-III is nonglycosylated while *L. migratoria* apoLp-III contains two complex carbohydrate chains (Hård *et al*., 1993). Sequence analysis predicts that all apoLp-IIIs are composed of predominantly amphipathic α-helix secondary structure, consistent with far ultraviolet circular dichroism (CD) studies (Ryan *et al*., 1993; Weers *et al*., 1998). An important breakthrough in our understanding of the structure of apoLp-III occurred with determination of the X-ray crystal structure of *L. migratoria* apoLp-III (Breiter *et al*., 1991). These authors showed that apoLp-III exists as a globular, up-and-down amphipathic α-helix bundle in the absence of lipid. The molecule is composed of five discrete α-helix segments that orient their hydrophobic faces toward the center of the bundle. Using a convenient method for bacterial overexpression, recombinant *M. sexta* apoLp-III was enriched with stable isotopes (Ryan *et al*., 1995; Wang *et al*., 1997b). Application of heteronuclear multidimensional nuclear magnetic resonance (NMR) techniques to isotopically enriched *M. sexta* apoLp-III yielded a complete assignment of this protein (Wang *et al*., 1997a). Structure calculations revealed a five-helix bundle molecular architecture, representing the first full-length apolipoprotein whose high resolution solution structure has been determined in the absence of detergent (Wang *et al*., 2002) (**Figure 3**). In keeping with the X-ray structure of *L. migratoria* apoLp-III, this structure also reveals an up-and-down bundle of five amphipathic α-helices. Interesting, however, Wang and coworkers identified a distinct short segment of α-helix that connects helix 3 and helix 4 in the bundle (termed helix 3′). This sequence segment (P_{95}DVEKE$_{100}$) aligns perpendicular to the long axis of the bundle and, as discussed below, has been shown to play a role in the initiation of apoLp-III lipid interaction. More recently, Fan *et al*. (2001, 2003) employed multidimensional NMR techniques to obtain a complete assignment and solution structure determination for *L. migratoria* apoLp-III. This work is significant in that it permits direct comparison between the X-ray crystal structure and the NMR structure. Interestingly, Fan *et al*. provide previously unreported structural evidence for a solvent exposed short helix that is positioned perpendicular to the long axis of the helix bundle. These authors propose that this short helix can serve as a recognition helix for initiation of apoLp-III lipid interaction, leading to conformational opening of the helix bundle. Contrary to the model presented by Breiter *et al*. (1991), Fan *et al*. (2003) suggest an alternate opening mechanism. Further studies will be required to elucidate the precise mechanism whereby apoLp-III recognizes and binds to available lipid surfaces (see below).

4.6.3.2. Lipid Induced Conformation Change

The up-and-down antiparallel organization of helical segments in apoLp-III allows for a simple opening of the bundle about putative "hinge" loops that connect the helices as originally proposed

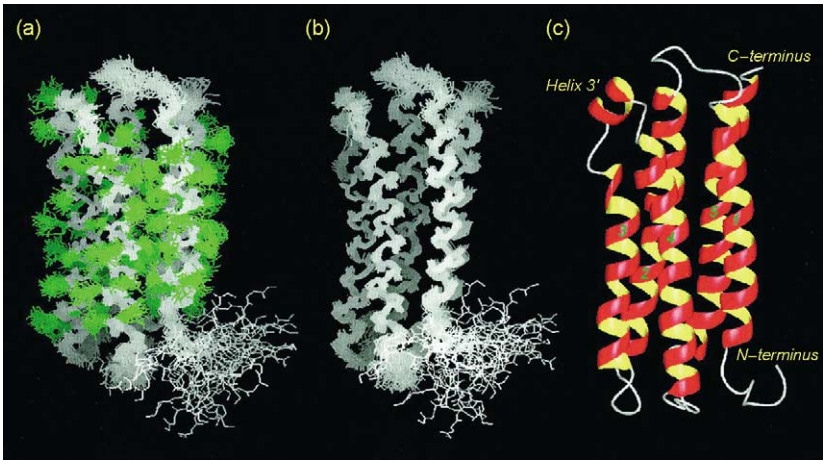


Figure 3 Nuclear magnetic resonance (NMR) visualizations of structure of lipid-free *Manduca sexta* apoLp-III. (a, b) Superposition of 40 NMR-derived structures of apoLp-III, with backbone atoms displayed in white and side chain heavy atoms displayed in green. (c) Ribbon representation of an energy-minimized, average structure of apoLp-III (PDB code 1EQ1). (Reproduced with permission from Wang, J., Sykes, B.D., Ryan, R.O., **2002**. Structural basis for the conformational adaptability of apolipophorin III, a helix bundle exchangeable apolipoprotein. *Proc. Natl Acad. Sci. USA 99*, 1188–1193; © by the National Academy of Sciences of the United States of America.)

by Breiter *et al.* (1991). The model suggests that apoLp-III initiates contact with lipid surfaces via one end of the helix bundle. Conformational opening could then occur with retention of helix boundaries present in the bundle configuration. Such an event would result in substitution of helix–helix interactions in the bundle conformation for helix–lipid interactions. Current evidence suggests that this conformational transition is triggered by availability of a suitable lipid surface and is reversible (Singh *et al.*, 1992; Liu *et al.*, 1993; Soulages and Wells, 1994a; Soulages *et al.*, 1995, 1996). Thus, it is conceivable that helix 3 and helix 4 move away from helices 1, 2, and 5 in concert as the bundle opens about the loop segments connecting helix 2 and helix 3 and helix 4 and helix 5 (Breiter *et al.*, 1991; Narayanaswami *et al.*, 1996b).

A well-known property of amphipathic exchangeable apolipoproteins in general is an ability to disrupt phospholipid bilayer vesicles and transform them into apolipoprotein–phospholipid disk complexes (Pownall *et al.*, 1978). This property represents a useful method to investigate aspects of the proposed lipid-induced helix bundle molecular switch process. The disk-shaped complexes formed between apoLp-III and dimyristoylphosphatidylcholine (DMPC) are of uniform size and composition, permitting detailed analysis of their structural organization (Wientzek *et al.*, 1994). Attenuated total reflectance Fourier transformed infrared spectroscopy has been employed to characterize helix orientation in apoLp-III–DMPC disk complexes (Raussens *et al.*, 1995, 1996). This analysis, and more recent studies (Soulages and Arrese, 2001) reveal that apoLp-III helical segments interact with phospholipid fatty acyl chains around the perimeter of the disk complex.

Several independent studies have provided convincing evidence that apoLp-III undergoes a significant conformational change upon association with lipid. Kawooya *et al.* (1986) used a monolayer balance to investigate apoLp-III behavior at the air–water interface, while Narayanaswami *et al.* (1996a) studied the unique fluorescence properties of the lone tyrosine in *M. sexta* apoLp-III. Near-ultraviolet CD analysis of *L. migratoria* apoLp-III indicates that helix realignment and reorientation occurs upon interaction with phospholipid vesicles (Weers *et al.*, 1994). Sahoo *et al.* (2000) used pyrene excimer fluorescence spectroscopy to investigate lipid binding induced realignment of helix 2 and helix 3 in *M. sexta* apoLp-III. In this study cysteine residues were introduced into the protein by site-directed mutagenesis (N40C and L90C). These sites were selected for introduction of cysteine residues based on the fact that they reside in close proximity in the helix bundle conformation. Covalent modification of the cysteine thiol groups with pyrene maleimide yielded a double pyrene labeled apoLp-III. In the absence of lipid, pyrene labeled apoLp-III adopts a helix bundle conformation. Fluorescence spectroscopy experiments revealed normal pyrene emission at 375 and 395 nm (excitation 345 nm) as well as excimer (excited state dimer) fluorescence at longer wavelengths (~460 nm). Control experiments verified that the excimer peak arose from intramolecular pyrene–pyrene interactions in the

labeled protein and was not due to intermolecular interactions. Because it is known that excimer fluorescence is manifest only when pyrene moieties are within 10 Å of one another, this property was used to assess the effect of lipid binding. The observation that excimer fluorescence was greatly reduced when apoLp-III was complexed with DMPC was taken as evidence for a conformational change in the protein upon lipid binding that results in relocation of helix 2 away from helix 3.

In fluorescence studies of apoLp-III, carried out by Soulages and Arrese (2000a, 2000b), site-directed mutagenesis was used to create various mutant apoLp-IIIs with a single tryptophen residue in each of the five helical segments of the protein. Data obtained in this study suggests that apoLp-III undergoes a conformational change that brings helices 1, 4, and 5 into contact with the lipid surface, while others (helices 2 and 3) appear to behave differently. In other studies Soulages *et al.* (2001) used disulfide bond engineering to show that conformational flexibility of helices 1 and 5 of *L. migratoria* apoLp-III play an important role in the lipid binding process. In other studies Dettloff *et al.* (2001b) reported that a C-terminal truncated apoLp-III from the wax moth *Galleria mellonella*, comprising the first three helical segments of the protein, retains structural integrity and an ability to interact with lipid surfaces. More recently, Dettloff *et al.* (2002) expanded this work to encompass two additional three helix mutants derived from *G. mellonella* apoLp-III, a C-terminal fragment comprising helices 3–5 and a core fragment comprising helices 2–4. All three truncation mutants retained their ability to solubilize bilayer vesicles of DMPC, an event that led to large increases in their α-helix content. The N-terminal and core fragment, but not the C-terminal fragment, were able to interact with phospholipase C modified human low-density lipoprotein, thereby preventing its aggregation. This result suggests that impairment of the lipid interaction properties of the C-terminal fragment has occurred as a result of removal of N-terminal helix segments. Taken together, it appears that the minimal essential elements required for apoLp-III lipid binding function is less than the intact five-helix bundle. Recent experiments have provided evidence that opening of the helix bundle is even more dramatic than originally postulated. It is now proposed that the protein adopts a fully extended belt-like conformation (Garda *et al.*, 2002; Sahoo *et al.*, 2002) (**Figure 4**). Garda *et al.* (2002) employed fluorescence resonance energy transfer methods while Sahoo *et al.* (2002) used pyrene excimer fluorescence to probe aspects of helix repositioning upon interaction with DMPC. In both of these studies, knowledge of the three-dimensional structure of the apoLp-III in the absence of lipid (i.e., the helix bundle conformation) allowed for structure guided site-directed mutagenesis to introduce strategically placed cysteine residues to which fluorescent reporter groups could be covalently attached. Subsequent characterization studies yielded a unifying model of apoLp-III conformation

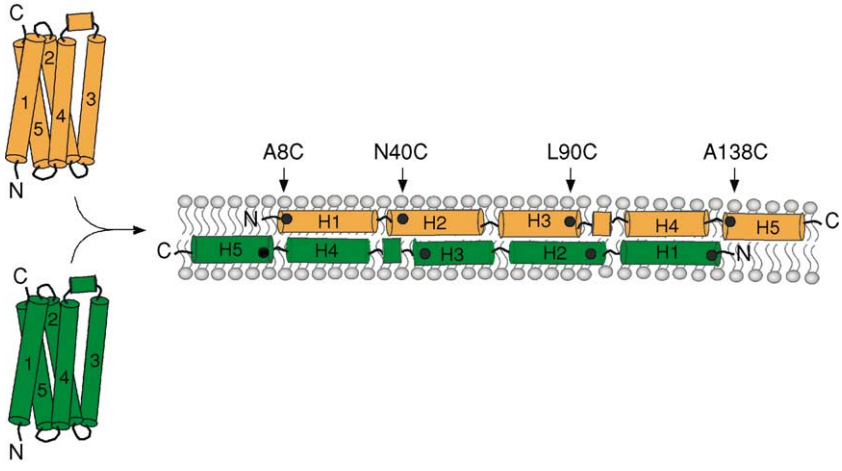


Figure 4 Model of apoLp-III bound to phospholipid discoidal complexes. ApoLp-III complexes with phospholipids on a discoidal particle adopting an extended α-helical conformation. Lipid-triggered association involves extension of H1 away from H5, helix bundle opening and repositioning of H2 and H3. The positions of cysteine substitution mutations employed in this and previous analyses are indicated: A8C, N40C, L90C, and A138C; H1, H2, H3, and H5. Apolp-III adopts an extended helical conformation around the periphery of discoidal phospholipid bilayer complexes, with neighboring molecules aligned antiparallel with respect to each other, and shifted by one helix. (Reprinted with permission from Sahoo, D., Weers, P.M.M., Ryan, R.O., Narayanaswami, V., **2002**. Lipid-triggered conformational switch of apolipophorin III helix bundle to an extended helix organization. *J. Mol. Biol. 321*, 201–214; © Elsevier.)

on disk complexes wherein the resulting structure resembles concepts and models that describe the organization of human apolipoprotein A-I on nascent high-density lipoproteins (Klon *et al.*, 2002).

4.6.3.3. Initiation of ApoLp-III Lipid Binding

Analysis of the structure of *L. migratoria* apoLp-III indicates the presence of solvent exposed leucine residues at one end of the protein (Breiter *et al.*, 1991). These authors proposed that this region of the molecule functions as a "hydrophobic sensor", which recognizes potential lipid surface binding sites. Surface plasmon resonance spectroscopy studies revealed that small amounts of DAG induce binding of apoLp-III to a phospholipid bilayer with its long molecular axis normal to the lipid surface (Soulages *et al.*, 1995). This interaction is proposed to be the first step in formation of a stable binding interaction. Site-directed mutagenesis was performed to determine whether alteration in the hydrophobicity of the putative sensor region of *L. migratoria* apoLp-III affects its ability to initiate contact with lipid surfaces (Weers *et al.*, 1999). In this study three partially exposed leucine residues, located at the end of the protein containing the loop segments that connect helix 1 and helix 2 and helix 3 and helix 4, were mutated to arginine. Three single arginine to leucine substitution mutants and a triple mutant were expressed in *Escherichia coli* and characterized in terms of their structural and stability properties. The effect of these mutations on phospholipid bilayer vesicle transformation into disk complexes versus lipoprotein binding suggests that the former binding interaction has an electrostatic component. Taken together, the data support the view that the end of the molecule bearing Leu 32, 34, and 95 is responsible for initiating contact with potential lipid surface binding sites.

The solution structure of *M. sexta* apoLp-III revealed the presence of helix 3′ at one end of the protein globule (Wang *et al.*, 1997a, 2002). One possibility is that helix 3′ reorientation facilitates contact with a lipid surface by exposing the hydrophobic interior of the helix bundle. The lipid surface could then trigger a molecular switch to induce conformational opening of the helix bundle and formation of a stable binding interaction. To investigate this, protein engineering was employed to remove helix 3′ and replace it with a sequence that has a high probability of forming a β-turn (Narayanaswami *et al.*, 1999). Characterization of the lipid binding properties of this "helix-to-turn" mutant apoLp-III revealed defective lipid binding properties. In more refined site-directed mutagenesis studies it was determined that Val 97, located in the center of helix 3′, is a critical residue for initiation of apoLp-III lipid binding. As described above, a similar short helix was identified in *L. migratoria* apoLp-III based on its NMR determined solution structure (Fan *et al.*, 2003). This helix, however, is present as the opposite end of the apoLp-III helix bundle suggesting that, if it is a recognition helix, bundle opening is different from that proposed for *M. sexta* apoLp-III by Narayanaswami *et al.* (1999) and Wang *et al.* (2002). It is conceivable that an experimental approach similar to that employed by Narayanaswami *et al.* (1999) will permit direct experimental assessment of the role of this short helix in *L. migratoria* apoLp-III lipid interaction.

Studies of the effect of the glycosyl moieties of *L. migratoria* apoLp-III on its lipid binding properties have also been investigated. Soulages *et al.* (1998) showed that recombinant apoLp-III, which lacks covalently bound carbohydrate, displayed a much stronger interaction with phospholipid vesicles than natural insect-derived apoLp-III. From the X-ray structure of *L. migratoria* apoLp-III in the absence of lipid, it is known that both glycosylation sites (at residues 18 and 85) are localized in the central region of the long axis of the bundle. Further study of this phenomenon revealed that apoLp-III sugar moieties interfere with helix bundle penetration into the bilayer surface during disruption and transformation into disk complexes (Weers *et al.*, 2000). Thus, it is apparent that structural aspects of the helix bundle as well as the composition of the lipid surface influence the ability of apoLp-III to initiate and form a stable lipid-binding interaction.

4.6.3.4. ApoLp-III Alternate Functions

Based on the observed developmentally timed upregulation of its mRNA, apoLp-III has been implicated in muscle and neuron programmed cell death (Sun *et al.*, 1995). When considered in light of its known lipid interaction properties, it is conceivable that it serves a function in membrane dissolution and/or lipid reabsorption during metamorphosis. Others have reported apoLp-III functions in insect immunity (Wiesner *et al.*, 1997). Indeed, recent reports suggest that it is lipid associated apoLp-III that manifests this biological activity (Dettloff *et al.*, 2001a, 2001c). These authors hypothesize that LDLp, formed *in vivo*, serves as an endogenous signal for immune activation, specifically mediated by lipid-associated apoLp-III interaction with hemocyte membrane receptors. From a structural standpoint, the truncated variants of *G. mellonella* apoLp-III (see above) that retain functional ability,

represent useful tools to probe the structural and physiological role of apoLp-III in innate immunity. Support for this general concept has emerged from studies of *G. mellonella* apoLp-III variants wherein point mutations were introduced at residues 66 and 68 (Niere *et al.*, 2001). The observation that mutation-induced decreases in apoLp-III lipid interaction properties correlate with decreased immune inducing activity is consistent with the hypothesis that apoLp-III immune activation is related to the conformational change that accompanies lipid interaction of this protein. On a broader scale it is important to understand the molecular details of this emerging group of proteins (Narayanaswami and Ryan, 2000) because the property of reversible interconversion between water-soluble and lipid bound states could have applications beyond their natural biological settings. Indeed, as work on this system continues, it is evident that apoLp-III and analogous helix bundle apolipoproteins represent novel biosurfactants with potentially useful properties, including biodegradability.

4.6.4. Lipophorin Receptor Interactions

4.6.4.1. The Low-Density Lipoprotein Receptor Family

In the concept of lipid transport during intense lipid utilization in insects, a major difference between the functioning of lipoproteins of mammals and insects is the selective mechanism by which insect lipoproteins transfer their hydrophobic cargo. Circulating HDLp particles may serve as a DAG donor or acceptor, dependent on the physiological situation, and function as a reusable lipid shuttle without additional synthesis or increased degradation of the apolipoprotein matrix, as discussed above. In apparent contrast to this concept, in fat body tissue of larval and young adult locusts, receptor-mediated uptake of HDLp was demonstrated (Dantuma *et al.*, 1997). A receptor has been cloned and sequenced from locust fat body cDNA, and identified as a novel member of the LDL receptor (LDLR) family, that is particularly expressed in fat body, oocytes, and the brain (Dantuma *et al.*, 1999). When stably transfected in an LDLR-deficient Chinese hamster ovary cell line, the locust receptor mediated endocytic uptake of fluorescently labeled HDLp that was absent in mock-transfected cells, suggesting that the receptor may function *in vivo* as an endocytic receptor for HDLp (Dantuma *et al.*, 1999). Domain organization of this insect lipophorin receptor (iLR; in this review LpR is adopted) shows a high similarity to mammalian LDLR. However, the ligand binding domain of LpR contains one additional cysteine-rich repeat compared to the seven repeats in LDLR, and is therefore identical to that of the human very low-density lipoprotein (VLDL) receptor (VLDLR), which also contains eight consecutive cysteine-rich repeats in this domain, as is schematically depicted in **Figure 5**. The amino acid sequence of the longer cytoplasmic tail of LpR is unique for insect lipophorin receptors: the 12 C-terminal amino acid residues of LDLR are completely different from those of LpR, whereas the C-terminal tail of LpR contains an additional 10 amino acid residues (Van Hoof *et al.*, 2002). A similar VLDLR homolog was identified in mosquito oocytes and shown to bind lipophorin (Cheon *et al.*, 2001). Three-dimensional models of the elements representing both the ligand binding

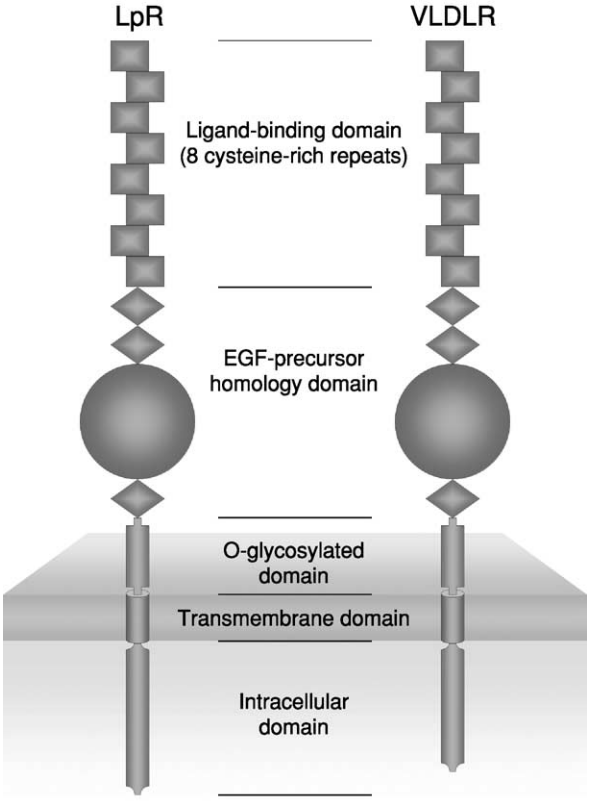


Figure 5 Schematic representation of the insect lipophorin receptor (LpR) and the mammalian VLDL receptor (VLDLR), indicating the identical domain organization. The mammalian LDL receptor has the same organization, but one ligand binding repeat less. EGF, epidermal growth factor. (Based on data from Dantuma, N.P., Potters, M., De Winther, M.P.J., Tensen, C.P., Kooiman, F.P., *et al.*, **1999**. An insect homolog of the vertebrate very low density lipoprotein receptor mediates endocytosis of lipophorins. *J. Lipid Res. 40*, 973–978.)

domain and the epidermal growth factor precursor homology domain of locust LpR bear a striking resemblance to those of mammalian LDLR (Van der Horst *et al*., 2002). Despite their pronounced structural similarity, however, the ligand specificity of LpR and LDLR for lipophorin and LDL, respectively, is mutually exclusive (Van Hoof *et al*., 2002). Additionally, the functioning of both receptors in lipid transport in the two animal groups seems to be intriguingly different, as discussed below.

Interaction of HDLp with a specific high-affinity binding site or receptor in the cell membrane of the fat body and other tissues of several insect species has been well documented (review: Ryan and Van der Horst, 2000; see also the recent data by Pontes *et al*., 2002; Van Hoof *et al*., 2003). In the binding of human LDL to its receptor, the most C-terminal 1000 amino acids in apoB are involved (Borén *et al*., 1998). Remarkably, although both the sequence and domain structure of the precursor protein of insect apoLp-I and apoLp-II (apoLp-II/I) resemble that of apoB100 (Babin *et al*., 1999; Van der Horst *et al*., 2002), apoLp-II/I does not show homology to this C-terminal part of apoB100, leaving the receptor-binding domain of apoLp-II/I to be disclosed in the future.

Immunocytochemical localization of HDLp has demonstrated the presence of the lipoprotein in endosomes of fat body of the larval dragonfly *Ashna cyanea* (Bauerfeind and Komnick, 1992) and in developing mosquito oocytes (Sun *et al*., 2000), suggesting endocytosis of circulating HDLp. Uptake of HDLp in the fat body of young adult locusts was shown to be receptor mediated (Dantuma *et al*., 1997). A recent study presents evidence for the involvement of LpR in the endocytic uptake mechanism for HDLp in the locust that is temporally present during specific periods of development (Van Hoof *et al*., 2003). Shortly after ecdysis, when lipid reserves are depleted, LpR is expressed in fat body tissue of young adult locusts as well as larvae, and fat body cells are able to endocytose the complete HDLp particle. On the fourth day after (larval or imaginal) ecdysis, however, expression of LpR is downregulated and drops below detectable levels; consequently, HDLp is no longer internalized. Downregulation of LpR was postponed by experimental starvation of adult locusts immediately after ecdysis. Moreover, by starving adult locusts after downregulation of LpR, expression of the receptor was induced. These data suggest that LpR expression is regulated by the demand of fat body tissue for lipid components (Van Hoof *et al*., 2003).

4.6.4.2. Ligand Recycling Hypothesis

Receptor-mediated uptake of HDLp in newly ecdysed adult and larval locusts may provide a mechanism for the uptake of specific lipid components separate from the mechanism of selective unloading of HDLp lipid cargo at the cell surface. At this time, however, the simultaneous existence of the two distinct mechanisms cannot be excluded. Downregulation of LpR in fat body cells suggests that this receptor is not involved in the lipophorin shuttle mechanism operative in the flying insect. Nevertheless, an endocytic uptake of HDLp seems to conflict with the selective process of lipid transport between HDLp and fat body cells without degradation of the lipophorin matrix. However, the pathway followed by the internalized HDLp may be different from the classical receptor-mediated lysosomal pathway typical of LDLR-internalized ligands. Therefore, an intriguing question is whether this novel LpR, in contrast to all other members of the LDLR family, is able to recycle its ligand after intracellular trafficking. In mammalian cells, LDL and diferric transferrin have been used extensively to study intracellular transport of ligands that are internalized by receptor-mediated endocytosis (Goldstein *et al*., 1985; Brown and Goldstein, 1986; Mellman, 1996; Mukherjee *et al*., 1997). Whereas LDL dissociates from its receptor and is completely degraded in lysosomes, transferrin remains attached to its receptor and is eventually resecreted from the cells (Ghosh *et al*., 1994). LDLR-expressing Chinese hamster ovary cells transfected with LpR cDNA were used to study the endocytic uptake and intracellular pathways of locust HDLp simultaneously with either human LDL or transferrin. Intracellular trafficking of multiple fluorescently labeled ligands was visualized by multicolor confocal laser scanning microscopy, and provided evidence for different intracellular routes followed by the mammalian and insect lipoproteins (Van Hoof *et al*., 2002) (**Figure 6**). In contrast to LDL, which is degraded in lysosomes after dissociating from its receptor, HDLp remained coupled to LpR and was transported to a nonlysosomal juxtanuclear compartment. Colocalization of HDLp with transferrin identified this organelle as the endocytic recycling compartment, from which internalized HDLp was eventually resecreted (half-time ~13 min), in a manner similar to that operative in the transferrin recycling pathway.

The above data indicate that, in mammalian cells, endocytosed insect HDLp, in contrast to human LDL, follows a recycling pathway mediated by LpR. This promises to elucidate new aspects of

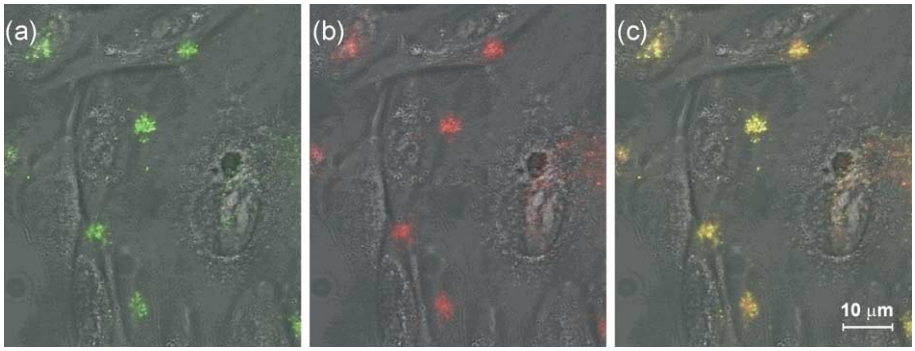


Figure 6 Confocal laser microscopic digital image of Chinese hamster ovary cells incubated with fluorescently labeled HDLp (a) and transferrin (b) after a chase period of 20 min. Colocalization of both ligands is visualized in yellow when images (a) and (b) are merged (c). (Reproduced with permission from Van der Horst, D.J., Van Hoof, D., Van Marrewijk, W.J.A., Rodenburg, K.W., **2002**. Alternative lipid mobilization: the insect shuttle system. *Mol. Cell. Biochem. 239*, 113–119; © Kluwer Academic Publishers.)

LDLR functions, although recycling of endocytosed HDLp in insect fat body cells remains to be shown. Additionally, although the acidic endosomal environment of endocytosed HDLp has been postulated to facilitate the transfer of lipid components other than DAG or cholesterol (Dantuma *et al*., 1997), both the function of the process of receptor-mediated endocytosis and the rationale for its occurrence during specific stages of insect development remain to be elucidated.

4.6.5. Other Lipid Binding Proteins

4.6.5.1. Lipid Transfer Particle

The concept that specialized proteins exist, which function in redistribution of hydrophobic lipid molecules, is well documented in the mammalian literature. A wide variety of distinct lipid transfer proteins have been characterized and their metabolic roles investigated. In 1986 a lipid transfer particle (LTP) was isolated from *M. sexta* larvae and shown to facilitate vectorial redistribution of lipids among plasma lipophorin subspecies (Ryan *et al*., 1986a, 1986b). In subsequent studies LTP was implicated in formation of LDLp from HDLp in response to AKH (Van Heusden and Law, 1989). The concept that LTP functions in flight-related lipophorin conversions correlates well with observed increases in LTP concentration in adult hemolymph compared with other developmental stages (Van Heusden *et al*., 1996; Tsuchida *et al*., 1998). When compared to other lipid transfer proteins, however, LTP displays unique structural characteristics. For example, it exists as a high molecular weight complex of three apoproteins (apoLTP-I, ~320 000 kDa; apoLTP-II, 85 000 kDa; and apoLTP-III, 55 000 kDa) and 14% noncovalently associated lipid (Ryan *et al*., 1988). LTPs exhibiting similar structural properties have been isolated from *L. migratoria, Periplaneta americana*, *B. mori*, and *Rhodnius prolixus* hemolymph (Hirayama and Chino, 1990; Takeuchi and Chino, 1993; Tsuchida *et al*., 1997; Golodne *et al*., 2001). The large size of LTP permitted examination of its structural properties by negative stain electron microscopy (Ryan *et al*., 1990a; Takeuchi and Chino, 1993). LTP from these two distinct species displays a highly asymmetric morphology with two major structural features, a quasispherical head region and an elongated cylindrical tail, which appears to possess a central hinge. The lipid component resembles that of lipophorin in that it contains predominantly phospholipid and DAG (Ryan, 1990). An important question arising from these physical characteristics relates to the requirement of the lipid component as a structural entity and/or its involvement in catalysis of lipid transfer. Studies employing lipoproteins containing radiolabeled lipids in incubations with LTP have revealed that the lipid component of the particle is in dynamic equilibrium with that of lipoprotein substrates (Ryan *et al*., 1988). Thus, it is evident that the lipid moiety is not merely a static structural component of LTP. Rather, it can be considered as a functional element in the mechanism of lipid transfer.

Very recently, studies of the *Drosophila* genome revealed the presence of a homolog of mammalian microsomal lipid transfer protein, MTP. Sellers *et al*. (2003) provided evidence that *Drosophila* MTP possesses the functional ability to transfer TAG onto nascent apolipoprotein B-containing lipoproteins in transfected COS cells, although it does not facilitate transfer of TAG among vesicles *in vitro*. Future studies, designed to elucidate the physiological role of this newly identified insect MTP and its participation in lipophorin assembly,

interconversions or some other aspect of lipid transport, represents a fruitful direction for future research. Given the size similarity between *Drosophila* MTP and apoLTP-II from other insect species, as well as their apparent functional similarities (see below), it is conceivable that apoLTP-II may be related to insect MTP.

4.6.5.1.1. Lipid substrate specificity Experiments have been conducted to examine the ability of LTP to utilize various substrate lipids. As reviewed earlier (Ryan and Van der Horst, 2000; Arrese *et al*., 2001) LTP catalyses the exchange and net transfer of DAG, in keeping with its proposed role in lipophorin interconversions *in vivo*. Extending this concept, Canavoso and Wells (2001) incubated radiolabeled midgut sacs with lipophorin containing medium. These authors found that transfer of DAG from the midgut sacs to lipophorin was blocked by preincubation with antibody against LTP, supporting the view that LTP functions in DAG export from the midgut to lipophorin. In a similar manner, LTP was shown by Jouni *et al*. (2003) to be required for DAG transfer from lipophorin to *B. mori* ovarioles.

In studies of other potential lipid substrates, Singh and Ryan (1991) used [$^{14}C_1$]acetate to label the DAG and hydrocarbon moiety of lipophorin *in vivo*. Subsequent lipid transfer experiments revealed that LTP is capable of facilitating transfer of hydrocarbon among lipoprotein substrates, suggesting LTP plays a role in movement of these extremely hydrophobic, specialized lipids from their site of synthesis to their site of deposition at the cuticle (Takeuchi and Chino, 1993). Interestingly, the rate of LTP-mediated hydrocarbon transfer was slower than DAG transfer. In other work, *B. mori* LTP was employed in studies of LTP-mediated carotene transfer among lipophorin particles (Tsuchida *et al*., 1998). Again, compared to DAG transfer, the rate of LTP-mediated carotene redistribution was much slower. Taken together, these results suggest that LTP may have a preference for DAG versus hydrocarbon or carotenes. Alternatively, the observed preference for DAG may be a function of the relative accessibility of the substrates within the donor lipoprotein. The ability of LTP to facilitate phospholipid transfer was studied by Golodne *et al*. (2001). These authors observed that LTP mediated a time-dependent transfer of phospholipid that was nonselective. In contrast to the requirement for LTP to mediate transfer of DAG, hydrocarbon, carotenoids, and phospholipids, Yun *et al*. (2002) provided evidence that LTP does not function in cholesterol transfer or redistribution in *M. sexta*. Rather, cholesterol is proposed to diffuse among tissues via mass action, freely transferring between lipophorin and tissues, depending on the physiological need. In keeping with this interpretation, Jouni *et al*. (2003) found that cholesterol transfer from lipophorin to *B. mori* ovarioles was unaffected by antibodies directed against LTP, whereas DAG transfer was inhibited.

4.6.5.1.2. Mechanism of facilitated lipid transfer In general lipid transfer catalysts may act as carriers of lipid between donor and acceptor lipoproteins or transfer may require formation of a ternary complex between donor, acceptor, and LTP. Based on the observed LTP-mediated net transfer of DAG from HDLp to human LDL (Ryan *et al*., 1990b), a strategy was developed to address this question experimentally (Blacklock *et al*., 1992). [^{3}H]-DAG–HDLp and unlabeled LDL were covalently bound to Sepharose matrices and packed into separate columns connected in series, followed by circulation of LTP or buffer. Circulation of LTP, but not buffer, resulted in a concentration-dependent increase in the amount of radiolabeled DAG recovered in the LDL fraction, revealing that LTP facilitates net lipid transfer via a carrier-mediated mechanism.

Blacklock and Ryan (1995) employed LTP apolipoprotein specific antibodies to probe the structure and catalytic properties of *M. sexta* LTP, obtaining evidence that apoLTP-II is a catalytically important apoprotein. In a similar manner Van Heusden *et al*. (1996) employed LTP antibody inhibition experiments to demonstrate that all three LTP apoproteins are important for lipid transfer activity. These authors found that, unlike apoLp-III, apoLTP-III is not found as a free protein in hemolymph (Van Heusden *et al*., 1996) despite the fact that it dissociates from the complex following exposure to nonionic detergent (Blacklock and Ryan, 1995).

4.6.5.2. Carotenoid Binding Proteins

In insects, the involvement of lipophorin in the hemolymph transport of dietary carotenoids is well documented (Tsuchida *et al*., 1998). Lipophorin may selectively deposit these isoprenoid lipid components at specific tissues, which is in keeping with the general function of the insect lipoprotein to selectively accept and deliver its hydrophobic cargo. The mechanism involved in this selective delivery is not fully understood, and may involve LTP and/or the lipophorin receptor, as discussed above. Carotenoids fulfill several important roles in insects. Certain carotenoids are the provitamins for vitamin A, which is required as the visual

pigment chromophore in the animal kingdom, including *D. melanogaster* (Giovannucci and Stephenson, 1999). Another example is the production of a yellow cocoon in the silkworm, *B. mori*, which is dependent on the availability of carotenoids in the silk gland (Tabunoki *et al.*, 2002). On both aspects, novel data illustrate both the presence and function of carotenoid binding proteins in insects.

Recently, a carotenoid binding protein (CBP) from the silk gland of *B. mori* larvae has been identified (Tabunoki *et al.*, 2002). The function of the 33 kDa protein was studied using several phenotypes of *B. mori* mutants; only in mutants carrying the dominant *Y* (yellow hemolymph) gene CBP is present in the villi of the midgut epithelium, suggesting that CBP may be involved in absorption of carotenoids. Similarly, using another mutant, it was inferred that CBP is involved in the uptake of carotenoids by the silk gland. The deduced amino acid sequence from CBP cDNA revealed the protein to be a novel member of the steroidogenic acute regulatory (StAR) protein family; CBP binds carotenoids rather than cholesterol.

Another protein involved in the cellular uptake of dietary carotenoids was discovered in *Drosophila*. Recent studies demonstrate that the molecular basis for blindness of the *ninaD* visual mutant fly is a defect in the cellular uptake of carotenoids (Kiefer *et al.*, 2002). The *ninaD* gene encodes a class B scavenger receptor with significant sequence identity to the mammalian class B scavenger receptors, SR-BI and CD36; the loss of this function abolishes carotenoid uptake and results in a vitamin A-deficient phenotype. In mammals, class B scavenger receptors, particularly SR-BI, are involved in cholesterol homeostasis and mediate the bidirectional flux of unesterified cholesterol between target cells and lipoproteins (Jian *et al.*, 1998; Yancey *et al.*, 2000). As lipophorin is structurally related to mammalian lipoproteins, it was speculated that the *ninaD* scavenger receptor mediates the transfer of carotenoid from lipophorin in a mechanistically similar manner (Kiefer *et al.*, 2002). *ninaD* mRNA levels were particularly high in pupae, which may indicate that NinaD has an important role in the redistribution of zeaxanthin, the larval storage form of carotenoids in the fat body, to the developing eyes during pupation (Giovannucci and Stephenson, 1999). Interestingly, in 1–2-day-old adults, *ninaD* mRNA levels were equally high, in contrast to older adults (>10 days) where *ninaD* mRNA was hardly visible. This suggests a similarity to the expression of the lipophorin receptor in the adult locust (Van Hoof *et al.*, 2003), and may indicate that both proteins function in specific lipid metabolic processes in the young adult stage that are no longer required in the older adult.

4.6.5.3. Fatty Acid Binding Proteins

Hydrolysis of LDLp-carried DAG by a lipophorin lipase at the flight muscles (review: Van der Horst *et al.*, 2001) results in the extracellular production of FFAs. After uptake by the flight muscle cells, these FFAs are oxidized for energy generation. The mechanism by which the extracellularly liberated FFAs are translocated across the plasma membrane is as yet unknown, but may involve membrane fatty acid transporter proteins similar to those identified in mammals, i.e., fatty acid translocase (FAT)/CD36, fatty acid transport protein (FATP), and plasma membrane fatty acid binding protein ($FABP_{pm}$) (reviews: Glatz and Van der Vusse, 1996; Brinkmann *et al.*, 2002). The intracellular transport of FFAs in insect flight muscle is mediated by a fatty acid binding protein (FABP) (review: Haunerland, 1997). This insect FABP belongs to the cytoplasmic FABPs that comprise a family of 14–15 kDa proteins that bind fatty acid ligands with high affinity and are involved in shuttling fatty acids to cellular compartments, modulating intracellular lipid metabolism, and regulating gene expression (review: Boord *et al.*, 2002). Intriguingly, in contrast to FABP in mammals, locust FABP is an adult-specific protein, the expression of which is directly linked to metamorphosis; to accomplish the extremely high metabolic rate encountered during migratory flight, the concentration of FABP in locust flight muscle cytosol is over three times that as in the mammalian heart (review: Haunerland, 1997). The high amino acid sequence similarity (82%) between the FABP of *L. migratoria* flight muscle and that of human skeletal muscles (Maatman *et al.*, 1994) is reflected in a strong similarity in their three-dimensional structures (Van der Horst *et al.*, 2002).

4.6.5.4. Lipases

Similar to those in mammals, the lipids that are oxidized during sustained physical exercise in insects are derived from stored TAG reserves by the action of a lipase. In mammals, however, FFAs are mobilized from adipose tissue and transported in the blood bound to serum albumin for uptake and oxidation in muscle, whereas in insects, DAG is released from the fat body and transported to the flight muscles by lipophorin as discussed above, requiring hydrolysis of DAG at the flight muscles to liberate the fatty acids that are eventually taken up. The lipase catalyzing the hydrolysis of TAG

stores in both mammals and insects plays a crucial role in energy metabolism by providing the source of energy for the working muscles, and thus between mammalian adipose tissue lipase (HSL) and insect fat body TAG lipase there is a clear functional similarity. However, there is an essential difference in the mode of action of both lipolytic enzymes, as TAG hydrolysis by HSL results in the release of FFAs, whereas that of the insect TAG lipase results in release of DAG.

HSL is regulated posttranscriptionally by phosphorylation; structural studies of the enzyme have identified several amino acids and regions of the molecule that are critical for enzymatic activity and regulation of HSL (reviews: Holm *et al*., 1997; Kraemer and Shen, 2002). In addition, other lipid-droplet associated proteins such as perilipins and lipoptransin may be involved in the interaction of HSL with the stored lipid (reviews: Ryan and Van der Horst, 2000; Van der Horst *et al*., 2001). Recent data in which the acute changes in HSL activity of human adipose tissue were monitored during physical exercise by assaying the phosphorylation state of the enzyme show that lipase activity increased over sixfold above baseline at 5 min of exercise and subsequently decreased gradually despite continuation of the exercise stimulus (Petridou and Mougios, 2002).

4.6.5.4.1. Regulation of lipolysis In spite of the importance of lipid mobilization for sustained insect flight, the regulation of lipolysis in insect fat body is largely unknown. The involvement of AKHs in lipolysis is without debate and was demonstrated both *in vivo* from enhanced levels of DAG in hemolymph of insects injected with the hormones and *in vitro* by the accumulation of DAG in isolated *L. migratoria* fat body tissue that was incubated in the presence of AKH; both cAMP and Ca^{2+} were shown to play an important role in the effect of AKH on lipolysis (reviews: Ryan and Van der Horst, 2000; Van der Horst *et al*., 2001; Van der Horst and Oudejans, 2003).

In two insect species that rely on lipid mobilization during flight activity, *L. migratoria* and *M. sexta*, it has been shown that the DAG, which is released from the fat body by the action of AKH, is stereospecific and has the *sn*-1,2 configuration (review: Ryan and Van der Horst, 2000). The pathway for the stereospecific synthesis of this *sn*-1,2-DAG is not well understood. Although a number of experimental data suggest that a stereospecific TAG lipase is involved (Arrese and Wells, 1997; Arrese *et al*., 2001) other data obtained by the same authors are conflicting with this view (review: Ryan and Van der Horst, 2000). For instance, direct stereospecific hydrolysis of TAG into *sn*-1,2-DAG would involve enzymatic removal of the *sn*-3 fatty acid. As the latter fatty acid was not accumulated in the fat body nor in the hemolymph (Arrese *et al*., 2001), conclusive data demonstrating unambiguously the occurrence of this mode of TAG lipase action is still lacking.

In summary, unlike the functional similarity between HSL in vertebrate adipose tissue and TAG lipase in insect fat body, there is a discrepancy in the mode of action of both lipolytic enzymes. The vertebrate HSL catalyzes nonstereospecific hydrolysis of TAG (as well as of DAG and monoacylglycerol) (review: Kraemer and Shen, 2002), leading to the release of fatty acids, whereas the action of the insect TAG lipase results in the formation and release of *sn*-1,2-DAG. So far, the reason for this discrepancy is unknown, but both structural and regulatory aspects may be involved. Elucidation of the mechanism of TAG lipase action and the regulation of its activity is essential for a better insight in the as yet poorly understood mechanism of flight-directed lipolysis in insects.

4.6.5.5. Vitellogenin

Oocyte development in adult females involves the accumulation of large amounts of lipid, most of which is extraovarian in origin and is delivered by lipophorin. Another lipid binding protein that serves a role in lipid transport to the oocyte is vitellogenin; although its overall contribution to the oocyte lipid accumulation is relatively modest (about 5%) (Sun *et al*., 2000). While the structural properties of insect vitellogenins are diverse, they generally possess ~10% lipid, primarily phospholipid and glycerolipid. Vitellogenin is synthesized and assembled in the fat body, secreted into hemolymph, and taken up by oocytes. Vitellogenin uptake is facilitated by members of a subfamily of the LDLR family that have been characterized in *Drosophila* (Schonbaum *et al*., 1995) and in the mosquito *Aedes aegypti* (Sappington *et al*., 1996). Recently, it was demonstrated that the ovarian vitellogenin receptor (VgR) is only distantly related to the lipophorin receptor (LpR), another ovarian LDLR homolog with a different ligand (Cheon *et al*. 2001). These data imply that the receptor-mediated mechanisms involved in the uptake of lipid and the accumulation of yolk protein precursors (which provide a key nutrient source for the developing oocyte), utilize two separate receptors (VgR and LpR) which are specific for their respective ligands, vitellogenin and lipophorin. Considerable early work was performed to characterize

the lipid transport properties of vitellogenin (review: Kunkel and Nordin, 1985), while more recent work has focused on molecular and evolutionary aspects of vitellogenin proteins and receptor-mediated endocytosis. These aspects, which are beyond the scope of the present treatise, have been comprehensively reviewed elsewhere (Sappington and Raikhel, 1998; Raikhel *et al.*, 2002) With respect to lipophorin internalization, it is important to note that immunocytochemical data in the mosquito (Sun *et al.*, 2000) revealed that only a small amount of lipophorin accumulates in developing oocytes as yolk protein, comprising ~3% of total ovarian proteins upon completion of protein internalization. Since lipid accounts for 35–40% of the insect egg dry weight (Kawooya and Law, 1988), Sun *et al.* (2000) proposed that internalization of lipophorin is unlikely to be the major route of lipid delivery to the developing oocyte. A dual mechanism for lipophorin-mediated lipid delivery to oocytes (a lipophorin shuttle mechanism involving LDLp and internalization of HDLp, with stripping of most of its lipid) has been demonstrated earlier (Kawooya and Law, 1988; Kawooya *et al.*, 1988; Liu and Ryan, 1991). However, considering that the LpR receptor involved in uptake of HDLp by mosquito oocytes bears a high structural similarity to the LpR discovered in locust fat body cell membranes (Dantuma *et al.*, 1999; Cheon *et al.*, 2001), the precise mechanism of LpR-mediated endocytosis in the oocyte, and the fate of HDLp, remain open questions. In light of the ligand recycling hypothesis discussed earlier for lipid delivery to the fat body, the possibility exists that LpR recycles its ligand after intracellular trafficking, providing another mechanism for the uptake of specific lipid components by the oocyte.

References

Arrese, E.L., Canavoso, L.E., Jouni, Z.E., Pennington, J.E., Tsuchida, K., *et al.*, **2001**. Lipid storage and mobilization in insects: current status and future directions. *Insect Biochem. Mol. Biol. 31*, 7–17.

Arrese, E.L., Wells, M.A., **1997**. Adipokinetic hormone-induced lipolysis in the fat body of an insect, *Manduca sexta*: synthesis of *sn*-1,2-diacylglycerols. *J. Lipid Res. 38*, 68–76.

Babin, P.J., Bogerd, J., Kooiman, F.P., Van Marrewijk, W.J.A., Van der Horst, D.J., **1999**. Apolipophorin II/I, apolipoprotein B, vitellogenin, and microsomal triglyceride transfer protein genes are derived from a common ancestor. *J. Mol. Evol. 49*, 150–160.

Baggerman, G., Huybrechts, J., Clynen, E., Hens, K., Harthoorn, L., *et al.*, **2002**. New insights in adipokinetic hormone (AKH) precursor processing in *Locusta migratoria* obtained by capillary liquid chromatography-tandem mass spectrometry. *Peptides 23*, 635–644.

Bauerfeind, R., Komnick, H., **1992**. Immunocytochemical localization of lipophorin in the fat body of dragonfly larvae (*Aeshna cyanea*). *J. Insect Physiol. 38*, 185–198.

Beenakkers, A.M.T., Van der Horst, D.J., Van Marrewijk, W.J.A., **1985**. Insect lipids and lipoproteins, and their role in physiological processes. *Prog. Lipid Res. 24*, 19–67.

Blacklock, B.J., Ryan, R.O., **1994**. Hemolymph lipid transport. *Insect Biochem. Mol. Biol. 24*, 855–873.

Blacklock, B.J., Ryan, R.O., **1995**. Structural studies of *Manduca sexta* lipid transfer particle with apolipoprotein-specific antibodies. *J. Lipid Res. 35*, 108–116.

Blacklock, B.J., Smillie, M., Ryan, R.O., **1992**. *Manduca sexta* lipid transfer particle can facilitate lipid transfer via a carrier-mediated mechanism. *J. Biol. Chem. 267*, 14033–14037.

Bogerd, J., Kooiman, F.P., Pijnenburg, M.A.P., Hekking, L.H.P., Oudejans, R.C.H.M., *et al.*, **1995**. Molecular cloning of three distinct cDNAs, each encoding a different adipokinetic hormone precursor, of the migratory locust, *Locusta migratoria*: differential expression of the distinct adipokinetic hormone precursor genes during flight activity. *J. Biol. Chem. 270*, 23038–23043.

Boord, J.B., Fazio, S., Linton, M.F., **2002**. Cytoplasmic fatty acid-binding proteins: emerging roles in metabolism and atherosclerosis. *Curr. Opin. Lipidol. 13*, 141–147.

Borén, J., Lee, I., Zhu, W., Arnold, K., Taylor, S., *et al.*, **1998**. Identification of the low density lipoprotein receptor-binding site in apolipoprotein B100 and the modulation of its binding activity by the carboxyl terminus in familial defective apo-B100. *J. Clin. Invest. 101*, 1084–1093.

Breiter, D.R., Kanost, M.R., Benning, M.M., Wesenberg, G., Law, J.H., *et al.*, **1991**. Molecular structure of an apolipoprotein determined at 2.5-Å resolution. *Biochemistry 30*, 603–608.

Brinkmann, J.F.F., Abumrad, N.A., Ibrahimi, A., Van der Vusse, G.J., Glatz, J.F.C., **2002**. New insights into long-chain fatty acid uptake by heart muscle: a crucial role for fatty acid translocase/CD36. *Biochem. J. 367*, 561–570.

Brown, M.S., Goldstein, J.L., **1986**. A receptor-mediated pathway for cholesterol homeostasis. *Science 232*, 34–47.

Canavoso, L.E., Jouni, Z.E., Karnas, K.J., Pennington, J.E., Wells, M.A., **2001**. Fat metabolism in insects. *Annu. Rev. Nutr. 21*, 23–46.

Canavoso, L.E., Wells, M.A., **2001**. Role of lipid transfer particle in delivery of diacylglycerol from midgut to lipophorin in larval *Manduca sexta*. *Insect Biochem. Mol. Biol. 31*, 783–790.

Candy, D.J., **2002**. Adipokinetic hormones concentrations in the haemolymph of *Schistocerca gregaria*, measured by radioimmunoassay. *Insect Biochem. Mol. Biol. 32*, 1361–1367.

Cheon, H.-M., Seo, S.-J., Sun, J., Sappington, T.W., Raikhel, A.S., **2001**. Molecular characterization of the

VLDL receptor homolog mediating uptake of lipophorin in oocytes of the mosquito *Aedes aegypti*. *Insect Biochem. Mol. Biol. 31*, 753–760.

Chino, H., **1985**. Lipid transport: biochemistry of hemolymph lipophorin. In: Kerkut, G.A., Gilbert, L.I. (Eds.), Comprehensive Insect Physiology, Biochemistry, and Pharmacology, vol. 10. Pergamon Press, Oxford, pp. 115–134.

Cole, K.D., Fernando-Warnakulasuriya, G.J.P., Boguski, M.S., Freeman, M., Gordon, J.I., *et al.*, **1987**. Primary structure and comparative sequence analysis of an insect apolipoprotein: apolipophorin III from *Manduca sexta*. *J. Biol. Chem. 262*, 11794–11800.

Dantuma, N.P., Pijnenburg, M.A.P., Diederen, J.H.B., Van der Horst, D.J., **1997**. Developmental down-regulation of receptor-mediated endocytosis of an insect lipoprotein. *J. Lipid Res. 38*, 254–265.

Dantuma, N.P., Potters, M., De Winther, M.P.J., Tensen, C.P., Kooiman, F.P., *et al.*, **1999**. An insect homolog of the vertebrate very low density lipoprotein receptor mediates endocytosis of lipophorins. *J. Lipid Res. 40*, 973–978.

Dettloff, M., Kaiser, B., Wiesner, A., **2001a**. Localization of injected apolipophorin III *in vivo*: new insights into the immune activation process directed by this protein. *J. Insect Physiol. 47*, 789–797.

Dettloff, M., Niere, M., Ryan, R.O., Kay, C.M., Wiesner, A., *et al.*, **2002**. Differential lipid binding of truncation mutants of *Galleria mellonella* apolipophorin III. *Biochemistry 41*, 9688–9695.

Dettloff, M., Weers, P.M.M., Niere, M., Kay, C.M., Ryan, R.O., *et al.*, **2001b**. An N-terminal three-helix fragment of an exchangeable insect apolipoprotein, apolipophorin III, conserves the lipid binding properties of the full length protein. *Biochemistry 40*, 3150–3157.

Dettloff, M., Wittwer, D., Weise, C., Wiesner, A., **2001c**. Lipophorin of a lower density is formed during immune responses in the lepidopteran insect *Galleria mellonella*. *Cell Tissue Res. 306*, 449–458.

Diederen, J.H.B., Oudejans, R.C.H.M., Harthoorn, L.F., Van der Horst, D.J., **2002**. Cell biology of the adipokinetic hormone-producing neurosecretory cells in the locust corpus cardiacum. *Microsc. Res. Tech. 56*, 227–236.

Fan, D., Reese, L., Ren, X., Weers, P.M.M., Ryan, R.O., *et al.*, **2001**. Complete ^{1}H, ^{15}N, and ^{13}C assignments of an exchangeable apolipoprotein, *Locusta migratoria* apolipophorin III. *J. Biomol. NMR 19*, 83–84.

Fan, D., Zheng, Y., Yang, D., Wang, J., **2003**. NMR solution structure and dynamics of an exchangeable apolipoprotein, *Locusta migratoria* apolipophorin III. *J. Biol. Chem. 278*, 21210–21220.

Gäde, G., **1996**. The revolution in insect neuropeptides illustrated by the adipokinetic hormone/red pigment-concentrating hormone family of peptides. *Z. Naturforsch. 51C*, 607–617.

Gäde, G., **1997**. The explosion of structural information on insect neuropeptides. *Prog. Chem. Organ. Nat. Prod. 71*, 1–128.

Gäde, G., Hoffmann, K.H., Spring, J.H., **1997**. Hormonal regulation in insects: facts, gaps and future directions. *Physiol. Rev. 77*, 963–1032.

Garda, H.A., Arrese, E.L., Soulages, J.L., **2002**. Structure of apolipophorin III in discoidal lipoproteins: interhelical distances in the lipid-bound state and conformational change upon binding to lipid. *J. Biol. Chem. 262*, 11794–11800.

Ghosh, R.N., Gelman, D.L., Maxfield, F.R., **1994**. Quantification of low density lipoprotein and transferrin endocytic sorting HEp2 cells using confocal microscopy. *J. Cell Sci. 107*, 2177–2189.

Giovannucci, D.R., Stephenson, R.S., **1999**. Identification and distribution of dietary precursors of the *Drosophila* visual pigment chromophore: analysis of carotenoids in wild type and *ninaD* mutants by HPLC. *Vision Res. 39*, 219–229.

Glatz, J.F.C., Van der Vusse, G.J., **1996**. Cellular fatty acid binding proteins: their function and physiological significance. *Prog. Lipid Res. 35*, 243–282.

Goldstein, J.L., Brown, M.S., Anderson, R.G.W., Russell, D.W., Schneider, W.J., **1985**. Receptor-mediated endocytosis: concepts emerging from the LDL receptor system. *Annu. Rev. Cell Biol. 1*, 1–39.

Goldsworthy, G.J., Lee, M.J., Luswata, R., Drake, A.F., Hyde, D., **1997**. Structures, assays and receptors for locust adipokinetic hormones. *Comp. Biochem. Physiol. B 117*, 483–496.

Golodne, D.M., Van Heusden, M.C., Gondim, K.C., Masuda, H., Atella, G.C., **2001**. Purification and characterization of a lipid transfer particle in *Rhodnius prolixus*: phospholipid transfer. *Insect Biochem. Mol. Biol. 31*, 563–571.

Hård, K., Van Doorn, J.M., Thomas-Oates, J.E., Kamerling, J.P., Van der Horst, D.J., **1993**. Structure of the Asn-linked oligosaccharides of apolipophorin III from the insect *Locusta migratoria*: carbohydrate-linked 2-amino-ethylphosphonate as a constituent of a glycoprotein. *Biochemistry 32*, 766–775.

Harthoorn, L.F., Diederen, J.H.B., Oudejans, R.C.H.M., Van der Horst, D.J., **1999**. Differential location of peptide hormones in the secretory pathway of insect adipokinetic cells. *Cell Tissue Res. 298*, 361–369.

Harthoorn, L.F., Diederen, J.H.B., Oudejans, R.C.H.M., Verstegen, M.M.A., Vullings, H.G.B., *et al.*, **2000**. Intracisternal granules in the adipokinetic cells of locusts are not degraded and apparently function as supplementary stores of secretory material. *Eur. J. Cell Biol. 79*, 27–34.

Harthoorn, L.F., Oudejans, R.C.H.M., Diederen, J.H.B., Van der Horst, D.J., **2002**. Coherence between biosynthesis and secretion of insect adipokinetic hormones. *Peptides 23*, 629–634.

Harthoorn, L.F., Oudejans, R.C.H.M., Diederen, J.H.B., Van de Wijngaart, D.J., Van der Horst, D.J., **2001**. Absence of coupling between release and biosynthesis of peptide hormones in insect neuroendocrine cells. *Eur. J. Cell Biol. 80*, 451–457.

Haunerland, N.H., **1997**. Transport and utilization of lipids in insect flight muscle. *Comp. Biochem. Physiol. B 117*, 475–482.

Hirayama, Y., Chino, H., **1990**. Lipid transfer particle in locust hemolymph: purification and characterization. *J. Lipid Res. 31*, 793–799.

Holm, C., Langin, D., Manganiello, V., Belfrage, P., Degerman, E., **1997**. Regulation of hormone-sensitive lipase activity in adipose tissue. *Methods Enzymol. 286*, 45–67.

Huybrechts, J., Clynen, E., Baggerman, G., De Loof, A., Schoofs, L., **2002**. Isolation and identification of the AKH III precursor-related peptide from *Locusta migratoria*. *Biochem. Biophys. Res. Commun. 296*, 1112–1117.

Jian, B., de la Llera-Moya, M., Ji, Y., Wang, N., Phillips, M.C., *et al.*, **1998**. Scavenger receptor class B type I as mediator of cellular cholesterol efflux to lipoproteins and phospholipid acceptors. *J. Biol. Chem. 273*, 5599–5606.

Jouni, Z.E., Takada, N., Gazard, J., Maekawa, H., Wells, M.A., *et al.*, **2003**. Transfer of cholesterol and diacylglycerol from lipophorin to *Bombyx mori* ovarioles *in vitro*: role of lipid transfer particle. *Insect Biochem. Mol. Biol. 33*, 145–153.

Kanost, M.R., Boguski, M.S., Freeman, M., Gordon, J.I., Wyatt, G.R., *et al.*, **1988**. Primary structure of apolipophorin-III from the migratory locust, *Locusta migratoria*. *J. Biol. Chem. 263*, 10568–10573.

Kawooya, J.K., Keim, P.S., Ryan, R.O., Shapiro, J.P., Samaraweera, P., *et al.*, **1984**. Insect apolipophorin III: purification and properties. *J. Biol. Chem. 259*, 10733–10737.

Kawooya, J.K., Law, J.H., **1988**. Role of lipophorin in lipid transport to the insect egg. *J. Biol. Chem. 263*, 8748–8753.

Kawooya, J.K., Meredith, S.C., Wells, M.A., Kézdy, F.J., Law, J.H., **1986**. Physical and surface properties of insect apolipophorin III. *J. Biol. Chem. 261*, 13588–13591.

Kawooya, J.K., Osir, E.O., Law, J.H., **1988**. Uptake of the major lipoprotein and its transformation in the insect egg. *J. Biol. Chem. 263*, 8740–8747.

Kiefer, C., Sumser, E., Wernet, M.F., von Lintig, J., **2002**. A class B scavenger receptor mediates the cullular uptake of carotenoids in *Drosophila*. *Proc. Natl Acad. Sci. USA 99*, 10581–10586.

Klon, A.E., Segrest, J.P., Harvey, S.C., **2002**. Comparative models for human apolipoprotein A-I bound to lipid in discoidal high-density lipoprotein particles. *Biochemistry 41*, 10895–10905.

Köllisch, G.V., Lorenz, M.W., Kellner, R., Verhaert, P.D., Hoffman, K.H., **2000**. Structure elucidation and biological activity of an unusual adipokinetic hormone from corpora cardiaca of the butterfly, *Vanessa cardui*. *Eur. J. Biochem. 267*, 5502–5508.

Kraemer, F.B., Shen, W.-J., **2002**. Hormone-sensitive lipase: control of intracellular tri- (di-)acylglycerol and cholesteryl ester hydrolysis. *J. Lipid Res. 43*, 1585–1594.

Kunkel, J.G., Nordin, J.H., **1985**. Yolk proteins. In: Kerkut, G.A., Gilbert, L.I. (Eds.), Comprehensive Insect Physiology, Biochemistry, and Pharmacology, vol. 1. Pergamon Press, Oxford, pp. 83–111.

Liu, H., Ryan, R.O., **1991**. Role of lipid transfer particle in transformation of lipophorin in insect oocytes. Biochim. Biophys. *Acta 1085*, 112–118.

Liu, H., Scraba, D.G., Ryan, R.O., **1993**. Prevention of phospholipase-C induced aggregation of low density lipoprotein by amphipathic apolipoproteins. *FEBS Lett. 316*, 27–33.

Maatman, R.G.H.J., Degano, M., Van Moerkerk, H.T.B., Van Marrewijk, W.J.A., Van der Horst, D.J., *et al.*, **1994**. Primary structure and binding characteristics of locust and human muscle fatty-acid-binding proteins. *Eur. J. Biochem. 221*, 801–810.

Mellman, I., **1996**. Endocytosis and molecular sorting. *Annu. Rev. Cell Devel. Biol. 12*, 575–625.

Mukherjee, S., Ghosh, R.N., Maxfield, F.R., **1997**. Endocytosis. *Physiol. Rev. 77*, 759–803.

Mwangi, R.W., Goldsworthy, G.J., **1977**. Diglyceride-transporting lipoproteins in *Locusta*. *J. Comp. Physiol. 114*, 177–190.

Mwangi, R.W., Goldsworthy, G.J., **1981**. Diacylglycerol-transporting lipoproteins and flight in *Locusta*. *J. Insect Physiol. 27*, 47–50.

Narayanaswami, V., Frolov, A., Schroeder, F., Oikawa, K., Kay, C.M., *et al.*, **1996a**. Fluorescence studies of lipid-association induced conformational adaptations of an exchangeable amphipathic apolipoprotein. *Arch. Biochem. Biophys. 334*, 143–150.

Narayanaswami, V., Ryan, R.O., **2000**. The molecular basis of exchangeable apolipoprotein function. *Biochim. Biophys. Acta 1483*, 15–36.

Narayanaswami, V., Wang, J., Kay, C.M., Scraba, D.G., Ryan, R.O., **1996b**. Disulfide bond engineering to monitor conformational opening of apolipophorin III during lipid binding. *J. Biol. Chem. 271*, 26855–26862.

Narayanaswami, V., Wang, J., Schieve, D., Kay, C.M., Ryan, R.O., **1999**. A molecular trigger of lipid-binding induced opening of a helix bundle exchangeable apolipoprotein. *Proc. Natl Acad. Sci. USA 96*, 4366–4371.

Niere, M., Dettloff, M., Maier, T., Ziegler, M., Wiesner, A., **2001**. Insect immune activation by apolipophorin III is correlated with the lipid-binding properties of this protein. *Biochemistry 40*, 11502–11508.

Oudejans, R.C.H.M., Mes, T.H.M., Kooiman, F.P., Van der Horst, D.J., **1993**. Adipokinetic peptide hormone content and biosynthesis during locust development. *Peptides 14*, 877–881.

Oudejans, R.C.H.M., Van der Horst, D.J., **2003**. Adipokinetic hormones: structure and biosynthesis. In: Henry, H.L., Norman, A.W. (Eds.), Encyclopedia of Hormones, vol. 1. Academic Press, London, pp. 38–42.

Oudejans, R.C.H.M., Vroemen, S.F., Jansen, R.F.R., Van der Horst, D.J., **1996**. Locust adipokinetic hormones: carrier-independent transport and differential inactivation at physiological concentrations during rest and flight. *Proc. Natl Acad. Sci. USA 93*, 8654–8659.

Petridou, A., Mougios, V., **2002**. Acute changes in triacylglycerol lipase of human adipose tissue during exercise. *J. Lipid Res. 43*, 1331–1334.

Pontes, E.G., Grillo, L.A.M., Gondim, K.C., **2002**. Characterization of lipophorin binding to the fat body of *Rhodnius prolixus*. *Insect Biochem. Mol. Biol. 32*, 1409–1417.

Pownall, H.J., Massey, J.B., Kusserow, S.K., Gotto, A.M., Jr., **1978**. Kinetics of lipid–protein interactions: interaction of apolipoprotein A-I from human plasma high density lipoproteins with phosphatidylcholines. *Biochemistry 17*, 1183–1188.

Prasad, S.V., Ryan, R.O., Wells, M.A., Law, J.H., **1986**. Changes in lipoprotein composition during larval–pupal metamorphosis of an insect, *Manduca sexta*. *J. Biol. Chem. 261*, 558–562.

Raikhel, A.S., Kokoza, V.A., Zhu, J., Martin, D., Wang, S.F., *et al.*, **2002**. Molecular biology of mosquito vitellogenesis: from basic studies to genetic engineering of antipathogen immunity. *Insect Biochem. Mol. Biol. 32*, 1275–1286.

Raussens, V., Goormaghtigh, E., Narayanaswami, V., Ryan, R.O., Ruysschaert, J.M., **1995**. Alignment of apolipophorin III α-helices in complex with dimyristoylphosphatidylcholine: a unique spatial orientation. *J. Biol. Chem. 270*, 12542–12547.

Raussens, V., Narayanaswami, V., Goormaghtigh, E., Ryan, R.O., Ruysschaert, J.M., **1996**. Hydrogen/deuterium exchange kinetics of apolipophorin III in lipid free and phospholipid bound states: an analysis by Fourier transform infrared spectroscopy. *J. Biol. Chem. 271*, 23089–23095.

Ryan, R.O., **1990**. The dynamics of insect lipophorin metabolism. *J. Lipid Res. 31*, 1725–1739.

Ryan, R.O., Howe, A., Scraba, D.G., **1990a**. Studies of the morphology and structure of the plasma lipid transfer particle from the tobacco hornworm, *Manduca sexta*. *J. Lipid Res. 31*, 871–879.

Ryan, R.O., Oikawa, K., Kay, C.M., **1993**. Conformational, thermodynamic and stability properties of *Manduca sexta* apolipophorin III. *J. Biol. Chem. 268*, 1525–1530.

Ryan, R.O., Prasad, S.V., Henriksen, E.J., Wells, M.A., Law, J.H., **1986a**. Lipoprotein interconversions in an insect, *Manduca sexta*: evidence for a lipid transfer factor in the hemolymph. *J. Biol. Chem. 261*, 563–568.

Ryan, R.O., Schieve, D.S., Wientzek, M., Narayanaswami, V., Oikawa, K., *et al.*, **1995**. Bacterial expression and site directed mutagenesis of a functional recombinant apolipoprotein. *J. Lipid Res. 36*, 1066–1072.

Ryan, R.O., Senthilathipan, K.R., Wells, M.A., Law, J.H., **1988**. Facilitated diacylglycerol exchange between insect hemolymph lipophorins: properties of *Manduca sexta* lipid transfer particle. *J. Biol. Chem. 263*, 14140–14145.

Ryan, R.O., Van der Horst, D.J., **2000**. Lipid transport biochemistry and its role in energy production. *Annu. Rev. Entomol. 45*, 233–260.

Ryan, R.O., Wells, M.A., Law, J.H., **1986b**. Lipid transfer protein from *Manduca sexta* hemolymph. *Biochem. Biophys. Res. Commun. 136*, 260–265.

Ryan, R.O., Wessler, A.N., Ando, S., Price, H.M., Yokoyama, S., **1990b**. Insect lipid transfer particle catalyzes bidirectional vectorial transfer of diacylglycerol from lipophorin to human low density lipoprotein. *J. Biol. Chem. 265*, 10551–10555.

Sahoo, D., Narayanaswami, V., Kay, C.M., Ryan, R.O., **2000**. Pyrene excimer fluorescence: a spatially sensitive probe to monitor lipid induced rearrangement of apolipophorin III. *Biochemistry 39*, 6594–6601.

Sahoo, D., Weers, P.M.M., Ryan, R.O., Narayanaswami, V., **2002**. Lipid-triggered conformational switch of apolipophorin III helix bundle to an extended helix organization. *J. Mol. Biol. 321*, 201–214.

Sappington, T.W., Kokoza, V.A., Cho, W.L., Raikhel, A.S., **1996**. Molecular characterization of the mosquito vitellogenin receptor reveals unexpected high homology to the *Drosophila* yolk protein receptor. *Proc. Natl Acad. Sci. USA 93*, 8934–8939.

Sappington, T.W., Raikhel, A.S., **1998**. Molecular characteristics of insect vitellogenins and vitellogenin receptors. *Insect Biochem. Mol. Biol. 28*, 277–300.

Schonbaum, C.P., Lee, S., Mahowald, A.P., **1995**. The *Drosophila* yolkless gene encodes a vitellogenin receptor belonging to the low-density lipoprotein receptor superfamily. *Proc. Natl Acad. Sci. USA 92*, 1485–1489.

Sellers, J.A., Hou, L., Athar, H., Hussain, M.M., Shelness, G.S., **2003**. A *Drosophila* microsomal triglyceride transfer protein homolog promotes the assembly and secretion of human apolipoprotein B: implications for human and insect lipid transport and metabolism. *J. Biol. Chem. 278*, 20367–20373.

Singh, T.K.A., Ryan, R.O., **1991**. Lipid transfer particle catalyzed transfer of lipoprotein-associated diacylglycerol and long chain aliphatic hydrocarbons. *Arch. Biochem. Biophys. 286*, 376–382.

Singh, T.K.A., Scraba, D.G., Ryan, R.O., **1992**. Conversion of human low density lipoprotein into a very low density lipoprotein like particle *in vitro*. *J. Biol. Chem. 267*, 9275–9280.

Smith, A.F., Owen, L.M., Strobel, L.M., Chen, H., Kanost, M.R., *et al.*, **1994**. Exchangeable apolipoproteins of insects share a common structural motif. *J. Lipid Res. 35*, 1976–1984.

Soulages, J.L., Arrese, E.L., **2000a**. Dynamics and hydration of the α-helices of apolipophorin III. *J. Biol. Chem. 275*, 17501–17509.

Soulages, J.L., Arrese, E.L., **2000b**. Fluorescence spectroscopy of single tryptophan mutants of apolipophorin-III in discoidal lipoproteins of dymyristoylphosphatidylcholine. *Biochemistry 39*, 10574–10580.

Soulages, J.L., Arrese, E.L., **2001**. Interaction of the α-helices of apolipophorin III with phospholipid acyl chains in discoidal lipoprotein particle: a fluorescence quenching study. *Biochemistry 40*, 14279–14290.

Soulages, J.L., Arrese, E.L., Chetty, P.S., Rodriguez, V., **2001**. Essential role of the conformational flexibility of helices 1 and 5 on the lipid binding activity of apolipophorin III. *J. Biol. Chem. 276*, 34162–34166.

Soulages, J.L., Pennington, J., Bendavid, O., Wells, M.A., **1998**. Role of glycosylation in the lipid-binding activity of the exchangeable apolipoprotein, apolipophorin-III. *Biochem. Biophys. Res. Commun. 243*, 372–376.

Soulages, J.L., Salamon, Z., Wells, M.A., Tollin, G., **1995**. Low concentrations of diacylglycerol promote the binding of apolipophorin III to a phospholipid bilayer: a surface plasmon resonance spectroscopy study. *Proc. Natl Acad. Sci. USA 92*, 5650–5654.

Soulages, J.L., van Antwerpen, R., Wells, M.A., **1996**. Role of diacylglycerol and apolipophorin-III in regulation of physicochemical properties of the lipophorin surface: metabolic implications. *Biochemistry 35*, 5191–5198.

Soulages, J.L., Wells, M.A., **1994a**. Effect of diacylglycerol content on some physicochemical properties of the insect lipoprotein, lipophorin: correlation with the binding of apolipophorin-III. *Biochemistry 33*, 2356–2362.

Soulages, J.L., Wells, M.A., **1994b**. Lipophorin: the structure of an insect lipoprotein and its role in lipid transport in insects. *Adv. Protein Chem. 45*, 371–415.

Staubli, F., Jørgensen, T.J.D., Cazzamali, G., Williamson, M., Lenz, C., *et al.*, **2002**. Molecular identification of the insect adipokinetic hormone receptors. *Proc. Natl Acad. Sci. USA 99*, 3446–3451.

Sun, D., Ziegler, R., Milligan, C.E., Fahrbach, S., Schwartz, L.M., **1995**. Apolipophorin III is dramatically up-regulated during the programmed cell death of insect skeletal muscle and neurons. *J. Neurobiol. 26*, 119–129.

Sun, J., Hiraoka, T., Dittmer, N.T., Cho, K.-H., Raikhel, A.S., **2000**. Lipophorin as a yolk protein in the mosquito, *Aedes aegypti*. *Insect Biochem. Mol. Biol. 30*, 1161–1171.

Tabunoki, H., Sugiyama, H., Tanaka, Y., Fujii, H., Banno, Y., *et al.*, **2002**. Isolation, characterization, and cDNA sequence of a carotenoid binding protein from the silk gland of *Bombyx mori* larvae. *J. Biol. Chem. 277*, 32133–32140.

Takeuchi, N., Chino, H., **1993**. Lipid transfer particle in the hemolymph of the American cockroach: evidence for its capacity to transfer hydrocarbons between lipophorin particles. *J. Lipid Res. 34*, 543–551.

Tsuchida, K., Arai, M., Tanaka, Y., Ishihara, R., Ryan, R.O., *et al.*, **1998**. Lipid transfer particle catalyzes transfer of carotenoids between lipophorins of *Bombyx mori*. *Insect Biochem. Mol. Biol. 28*, 927–934.

Tsuchida, K., Soulages, J.L., Moribayashi, A., Suzuki, K., Maekawa, H., *et al.*, **1997**. Purification and properties of a lipid transfer particle from *Bombyx mori*: comparison to the lipid transfer particle from *Manduca sexta*. *Biochim. Biophys. Acta 1337*, 57–65.

Vanden Broeck, J., **2001**. Insect G protein-coupled receptors and signal transduction. *Arch. Insect Biochem. Physiol. 48*, 1–12.

Van der Horst, D.J., Oudejans, R.C.H.M., **2003**. Adipokinetic hormones and lipid mobilization. In: Henry, H.L., Norman, A.W. (Eds.), Encyclopedia of Hormones, vol. 1. Academic Press, London, pp. 34–38.

Van der Horst, D.J., Van Doorn, J.M., Beenakkers, A.M.T., **1979**. Effects of the adipokinetic hormone on the release and turnover of hemolymph diglycerides and on the function of the diglyceride-transporting lipoprotein system during locust flight. *Insect Biochem. 9*, 627–635.

Van der Horst, D.J., Van Doorn, J.M., Beenakkers, A.M.T., **1984**. Hormone induced rearrangements of locust hemolymph lipopoproteins: the involvement of glycoprotein C_2. *Insect Biochem. 14*, 495–504.

Van der Horst, D.J., Van Doorn, J.M., De Keijzer, A.N., Beennakkers, A.M.T., **1981**. Interconversions of diacylglycerol-transporting lipoproteins in the haemolymph of *Locusta migratoria*. *Insect Biochem. 11*, 717–723.

Van der Horst, D.J., Van Hoof, D., Van Marrewijk, W.J.A., Rodenburg, K.W., **2002**. Alternative lipid mobilization: the insect shuttle system. *Mol. Cell. Biochem. 239*, 113–119.

Van der Horst, D.J., Van Marrewijk, W.J.A., Diederen, J.H.B., **2001**. Adipokinetic hormones of insects: release, signal transduction, and responses. *Int. Rev. Cytol. 211*, 179–240.

Van der Horst, D.J., Van Marrewijk, W.J.A., Vullings, H.G.B., Diederen, J.H.B., **1999**. Metabolic neurohormones: release, signal transduction and physiological responses of adipokinetic hormones in insects. *Eur. J. Entomol. 96*, 299–308.

Van Heusden, M.C., Law, J.H., **1989**. An insect lipid transfer particle promotes lipid loading from fat body to lipoprotein. *J. Biol. Chem. 264*, 17287–17292.

Van Heusden, M.C., Yepiz-Plascencia, G.M., Walker, A.M., Law, J.H., **1996**. *Manduca sexta* lipid transfer particle: synthesis by fat body and occurrence in hemolymph. *Arch. Insect Biochem. Physiol. 31*, 39–51.

Van Hoof, D., Rodenburg, K.W., Van der Horst, D.J., **2002**. Insect lipoprotein follows a transferrin-like recycling route that is mediated by the insect LDL receptor homologue. *J. Cell Sci. 115*, 4001–4012.

Van Hoof, D., Rodenburg, K.W., Van der Horst, D.J., **2003**. Lipophorin receptor-mediated lipoprotein endocytosis in insect fat body cells. *J. Lipid Res. 44*, 1431–1440.

Van Marrewijk, W.J.A., **2003**. Adipokinetic hormones and carbohydrate metabolism. In: Henry, H.L., Norman, A.W. (Eds.), Encyclopedia of Hormones, vol. 1. Academic Press, London, pp. 29–34.

Van Marrewijk, W.J.A., Van der Horst, D.J., **1998**. Signal transduction of adipokinetic hormone. In: Coast, G.M., Webster, S.G. (Eds.), Recent Advances in Arthropod Endocrinology. Cambridge University Press, Cambridge, pp. 172–188.

Vroemen, S.F., Van der Horst, D.J., Van Marrewijk, W.J.A., **1998**. New insights into adipokinetic hormone signaling. *Mol. Cell. Endocrinol. 141*, 7–12.

Vullings, H.G.B., Diederen, J.H.B., Veelaert, D., Van der Horst, D.J., **1999**. Multifactorial control of the release of hormones from the locust retrocerebral complex. *Microsc. Res. Tech. 45*, 142–153.

Wang, J., Gagne, S., Sykes, B.D., Ryan, R.O., **1997a**. Insight into lipid surface recognition and reversible conformational adaptations of an exchangeable apolipoprotein by multidimensional heteronuclear NMR techniques. *J. Biol. Chem. 272*, 17912–17920.

Wang, J., Liu, H., Sykes, B.D., Ryan, R.O., **1992**. ^{31}P-NMR study of the phospholipid moiety of lipophorin subspecies. *Biochemistry 31*, 8706–8712.

Wang, J., Liu, H., Sykes, B.D., Ryan, R.O., **1995**. Localization of two distinct microenvironments for the diacylglycerol component of lipophorin particles by 13C-NMR. *Biochemistry 34*, 6755–6761.

Wang, J., Sahoo, D., Schieve, D., Gagne, S., Sykes, B.D., *et al.*, **1997b**. Multidimensional NMR studies of an exchangeable apolipoprotein and its interaction with lipids. *Tech. Protein Chem. 8*, 427–438.

Wang, J., Sykes, B.D., Ryan, R.O., **2002**. Structural basis for the conformational adaptability of apolipophorin III, a helix bundle exchangeable apolipoprotein. *Proc. Natl Acad. Sci. USA 99*, 1188–1193.

Weers, P.M.M., Kay, C.M., Oikawa, K., Wientzek, M., Van der Horst, D.J., *et al.*, **1994**. Factors affecting the stability and conformation of *Locusta migratoria* apolipophorin III. *Biochemistry 33*, 3617–3624.

Weers, P.M.M., Narayanaswami, V., Kay, C.M., Ryan, R.O., **1999**. Interaction of an exchangeable apolipoprotein with phospholipid vesicles and lipoprotein particles: role of leucines 32, 34, and 95 in *Locusta migratoria* apolipophorin III. *J. Biol. Chem. 274*, 21804–21810.

Weers, P.M.M., Van der Horst, D.J., Ryan, R.O., **2000**. Interaction of locust apolipophorin III with lipoproteins and phospholipid vesicles: effect of glycosylation. *J. Lipid Res. 41*, 416–423.

Weers, P.M.M., Van Marrewijk, W.J.A., Beenakkers, A.M.T., Van der Horst, D.J., **1993**. Biosynthesis of locust lipophorin: apolipophorins I and II originate from a common precursor. *J. Biol. Chem. 268*, 4300–4303.

Weers, P.M.M., Wang, J., Van der Horst, D.J., Kay, C.M., Sykes, B.D., *et al.*, **1998**. Recombinant locust apolipophorin III: characterization and NMR spectroscopy. *Biochim. Biophys. Acta 1393*, 99–107.

Wientzek, M., Kay, C.M., Oikawa, K., Ryan, R.O., **1994**. Binding of insect apolipophorin III to dimyristoylphosphatidylcholine vesicles: evidence for a confirmational change. *J. Biol. Chem. 269*, 4605–4612.

Wiesner, A., Losen, S., Kopacek, P., Weise, C., Gotz, P., **1997**. Isolated apolipophorin III from *Galleria mellonella* stimulates the immune reactions of this insect. *J. Insect Physiol. 43*, 383–391.

Woodring, J., Lorenz, M.W., Hoffmann, K.H., **2002**. Sensitivity of larval and adult crickets (*Gryllus bimaculatus*) to adipokinetic hormone. *Comp. Biochem. Physiol. A 133*, 637–644.

Yancey, P.G., de la Llera-Moya, M., Swarnakar, S., Monzo, P., Klein, S.M., *et al.*, **2000**. High density lipoprotein phospholipid composition is a major determinant of the bi-directional flux and net movement of cellular free cholesterol mediated by scavenger receptor BI. *J. Biol. Chem. 275*, 36596–36604.

Yun, H.K., Jouni, Z.E., Wells, M.A., **2002**. Characterization of cholesterol transport from midgut to fat body in *Manduca sexta* larvae. *Insect Biochem. Mol. Biol. 32*, 1151–1158.

4.7 Proteases

M R Kanost and T E Clarke, Kansas State University, Manhattan, KS, USA

© 2005, Elsevier BV. All Rights Reserved.

4.7.1. Introduction and History

Proteases (peptidases) are enzymes that hydrolyze peptide bonds in proteins. Exopeptidases cleave a terminal amino acid residue at the end of a polypeptide; endopeptidases cleave internal peptide bonds. Hooper (2002) provides a useful introduction to the general properties of proteases. Proteases can be classified based on the chemical groups that function in catalysis. In serine proteases the hydroxyl group in the side chain of a serine residue in the active site acts as a nucleophile in the reaction that hydrolyzes a peptide bond, whereas in cysteine proteases the sulfhydryl group of a cysteine side chain performs this function. In aspartic acid proteases and metalloproteases, a water molecule in the active site (positioned by interacting with an aspartyl group or a metal ion, respectively) functions as the nucleophile that attacks the peptide bond. Proteases are classified on this basis of catalytic mechanism in a system developed by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (http://www.chem.qmul.ac.uk/iubmb/enzyme/EC3/4/). However, proteases can have the same catalytic mechanism but will be unrelated in amino acid sequence, as products of convergent evolution. The MEROPS classification system groups proteases into families based on sequence similarity (Rawlings and Barrett, 1999) (http://merops.sanger.ac.uk).

A protease cleaves a peptide bond, called the scissile bond, between two amino acid residues named P1 and P1′ (Schechter and Berger, 1967). Residues on the N-terminal side of the scissile bond are numbered in the C to N direction, whereas residues on the C-terminal side of the scissile bond (the "prime" side) are numbered in the N to C direction (**Figure 1**). The substrate specificity of most endopeptidases is highly dependent on the nature of the side chain of the P1 residue, but the sequence of other residues near the scissile bond can also affect binding of the substrate to the active site and thus influence substrate specificity.

Insects produce abundant proteases that function in digestion of dietary proteins in the gut. Such proteases are thoroughly discussed in 00053. This chapter focuses on nondigestive proteases, which have many diverse roles in insect biology. These proteases often function in cascade pathways, in which one protease activates the zymogen form of another protease, leading to amplification of an initial signal that may involve a few molecules and finally generating a very large number of effector molecules at the end of the pathway. The complement and blood coagulation pathways in mammals are well-understood examples of this type of protease cascade, which also occur in insect embryonic development and insect immune responses (Jiang and Kanost, 2000; Krem and DiCerra, 2002; see **Chapter 4.5**). Details of the organization and regulation of such pathways in insect biology are beginning to be understood. This chapter will include an

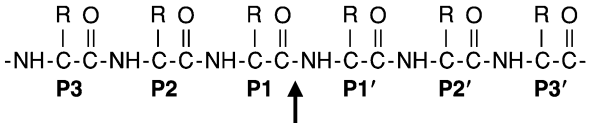


Figure 1 The Schechter and Berger (1967) notation for protease cleavage sites. The arrow designates the scissile peptide bond between amino acid residues P1 and P1′.

emphasis on the current state of knowledge in this rapidly developing area.

Insect proteases have previously been reviewed by Law *et al.* (1977), Applebaum (1985), Terra *et al.* (1996), and Reeck *et al.* (1999). These reviews deal primarily with proteases as they function in the digestion of food. Only quite recently has much detailed information appeared about proteases with other functions in insect biology. An exception is cocoonase, the first insect protease that was purified and well-characterized biochemically. Cocoonase is a serine protease from silk moths that functions to hydrolyze silk proteins in the cocoon, enabling the adult moth to emerge (Kafatos *et al.*, 1967a, 1967b). It digests sericin, the silk protein that cements fibroin threads together (see **Chapter 2.11**). A specialized tissue called the galea, derived from modified mouthparts, synthesizes and secretes the zymogen form, prococoonase (Kafatos, 1972). On the surface of the galea, prococoonase is activated by cleavage at a specific site by an unknown protease in the molting fluid (Berger *et al.*, 1971). Sequencing of an amino terminal fragment and the peptide containing the active site Ser residue indicated that the activation and catalytic mechanisms of coccoonase were quite similar to those of mammalian trypsin (Felsted *et al.*, 1973; Kramer *et al.*, 1973). It is surprising that no molecular cloning has apparently yet been carried out for this historically important insect protease.

4.7.2. Proteases in Eggs and Embryos

4.7.2.1. Proteases That Digest Egg Yolk Proteins

Vitellin and a few other egg-specific proteins stored in yolk granules of insect eggs are digested by proteases to release amino acids for use in embryonic development (Raikhel and Dhadialla, 1992; **Chapter 3.9**). Such enzymes in eggs represent several different protease families and mechanistic classes. In *Bombyx mori*, a serine protease that degrades vitellin was purified from *B. mori* eggs (Indrasith *et al.*, 1988), and its cDNA was cloned (Ikeda *et al.*, 1991). This protease cleaves after Arg or Lys P1 residues and is a member of the S1 (chymotrypsin-like) family of serine proteases. It is synthesized in ovaries as a zymogen and is activated during embryogenesis. A second serine protease from the S1 family specifically degrades the 30 kDa yolk proteins present in *B. mori* eggs (Maki and Yamashita, 1997, 2001). This protease, which is synthesized at the end of embryogenesis, has elastase-like specificity, cleaving after P1 residues with small side chains. A serine carboxypeptidase is synthesized in the fat body of a mosquito, *Aedes aegypti*, transported through the hemolymph, and taken up by oocytes (Cho *et al.*, 1991). This protease is synthesized as a zymogen and activated within eggs during embryogenesis.

Cysteine proteases have been characterized from eggs of several insect species. Those that have been sequenced are from the C1 (papain-like) family of cysteine proteases. They typically have acidic pH optima and have biochemical properties similar to mammalian cysteine proteases known as cathepsins (although not all proteases called cathepsins are cysteine proteases). A 47 kDa cysteine protease that can digest vitelllin has been purified from *B. mori* eggs (Kageyama and Takahashi, 1990; Yamamoto and Takahashi, 1993), and its cDNA has been cloned (Yamamoto *et al.*, 1994). It has sequence similarity to mammalian cathepsin L and a preference for cleaving at sites that have hydrophobic residues at the P2 and P3 positions. It is synthesized as a zymogen in ovary and fat body as a maternal product and taken up into oocytes (Yamamoto *et al.*, 2000). This cysteine protease is self-activated at low pH by proteolytic processing, apparently by a weak activity of the proenzyme under acidic conditions (Takahashi *et al.*, 1993).

Cysteine proteases with sequence similarity to mammalian cathepsin B have also been identified as enzymes that digest insect egg yolk proteins. In *Drosophila melanogaster*, a cysteine protease is associated with yolk granules (Medina *et al.*, 1988). Its zymogen is apparently activated by a serine protease during embryonic development, and the active cathepsin-B then digests yolk proteins. A cysteine protease that digests yolk proteins has also been identified in another higher dipteran, *Musca domestica* (Ribolla and De Bianchi, 1995). In *A. aegypti* a "vitellogenic cathepsin B" is synthesized in adult female fat body after the insect has taken a blood meal, and the zymogen is transported through the hemolymph and taken up by oocytes (Cho *et al.*, 1999). The enzyme is activated by proteolytic processing when embryonic development begins and then probably functions to digest vitellin. A cathepsin B-like protease that can digest vitellin is also synthesized in the fat body and ovaries of a

lepidopteran insect, *Helicoverpa armigera* (Zhao *et al.*, 2002). However, its gene is expressed in fat body of males and females and in larvae and pupae, and thus is not coordinated with vitellogenesis as is the mosquito cathepsin-B.

4.7.2.2. The Dorsal Pathway in Embryonic Development

A signal transduction system that regulates dorsal/ventral patterning in *D. melanogaster* embryonic development is activated by an extracellular serine protease cascade (Morisato and Anderson, 1995; LeMosy *et al.*, 1999). The members of this cascade are produced maternally and deposited in the space between the vitellin membrane and the embryo. The pathway was elucidated by genetic analysis, and recently the recombinant forms of the proteases have been studied. This pathway involves a serine protease cascade (**Figure 2**) that eventually cleaves an inactive protein called spätzle, making it competent to bind to a transmembrane receptor named Toll (Belvin and Anderson, 1996). Toll is homologous to the mammalian interleukin-1 receptor. Binding of spätzle to Toll initiates a signal transduction pathway that leads to activation of a transcription factor from the rel family named Dorsal.

A large (350 kDa) multi-domain protein called nudel, containing a serine protease domain, regions of LDL receptor repeats (see **Chapter 4.6**), and an N-terminal glycosaminoglycan modification, is secreted by the ovarian follicle cells into the perivitelline space (Hong and Hashimoto, 1995; Turcotte and Hashimoto, 2002). The nudel protease is autoactivated by a mechanism not yet understood (LeMosy *et al.*, 1998, 2000) and is thought to activate the second protease in the pathway, "gastrulation defective" by specific proteolysis. Mutations in nudel's protease domain produce a dorsalizing phenotype and can also result in fragile eggshells, suggesting an additional function for the protease activity (Hong and Hashimoto, 1996; LeMosy *et al.*, 1998, 2000; LeMosy and Hashimoto, 2000).

Gastrulation defective is a serine protease (Konrad *et al.*, 1998; Han *et al.*, 2000), with sequence similarity to mammalian complement factors C2 and B (DeLotto, 2001). Experiments with recombinant proteins have demonstrated that the gastrulation defective zymogen can be autoactivated when it interacts with the protease snake zymogen and that active gastrulation defective can in turn proteolytically activate the snake zymogen (Dissing *et al.*, 2001; LeMosy *et al.*, 2001). Computer modeling studies indicate that the zymogen activation site of gastrulation defective is a good fit in the active site of nudel and that the snake zymogen activation site can dock in the active site of gastrulation defective, consistent with the proposed functions of these enzymes in the cascade pathway (Rose *et al.*, 2003). A potential lower affinity interaction of the gastrulation defective active site with its own zymogen activation sequence may explain gastrulation defective's autoactivation in the absence of nudel when it is overexpressed in embryos or at high concentration *in vitro*.

The final two proteases in this cascade, snake and easter, contain C-terminal serine protease domains and N-terminal clip domains similar to horseshoe crab proclotting enzyme (DeLotto and Spierer, 1986; Chasan and Anderson, 1989; Gay and Keith, 1992; Smith and DeLotto, 1992, 1994). Clip domains, thought to function in protein–protein interactions, are also present in some hemolymph proteases of insects (Jiang and Kanost, 2000) (see Section 4.7.3.1 below). Mutations that eliminate the protease activity of snake (Smith *et al.*, 1994) or easter (Jin and Anderson, 1990) have abnormal dorsoventral phenotypes, indicating that a functional protease domain is essential for their roles in embryonic development. *In vitro* experiments with recombinant snake and easter zymogens confirm their order in the cascade indicated by genetic analysis: snake cleaves and activates easter, which cleaves prospätzle (Smith *et al.*, 1995; DeLotto and DeLotto, 1998; Dissing *et al.*, 2001; LeMosy *et al.*, 2001). These results are consistent with predictions

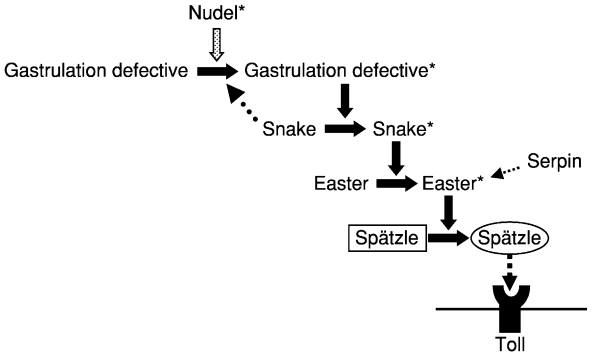


Figure 2 A model of the protease cascade that activates the Dorsal signal transduction pathway in *D. melanogaster* embryonic development. Nudel, gastrulation defective, snake, and easter are serine proteases that are synthesized as zymogens. The active forms of the proteases are indicated with an asterisk. Solid arrows indicate proteolytic activation steps that have been demonstrated by biochemical studies. A dotted arrow indicates that interaction between the snake and gastrulation defective zymogens can leads to activation of gastrulation defective. Easter* cleaves spätzle to produce an active ligand that binds to Toll, a transmembrane receptor. Easter is negatively regulated by interaction with an inhibitor from the serpin family.

based on computer modeling of the snake and easter three-dimensional structures and substrate interactions sites (Rose *et al*., 2003). Active easter is converted *in vivo* to a high molecular mass form that is probably a complex with a protease inhibitor that regulates its activity (Misra *et al*., 1998; Chang and Morisato, 2002). Female flies with a mutation in the gene for a serine protease inhibitor, serpin 27A, produce embryos that show defects in dorsal–ventral polarity, suggesting that this inhibitor is a maternal product that regulates at least one of the proteases in the pathway (Hashimoto *et al*., 2003; Ligoxygakis *et al*., 2003).

4.7.3. Hemolymph Plasma Proteases

4.7.3.1. Serine Proteases

Serine proteases in hemolymph have several types of physiological functions in defense against infection or wounding. An unusual phenomenon, perhaps related to protection against predation, involves serine proteases in the hemolymph of South American saturniid caterpillars of the genus *Lonomia* that are toxic to mammals. Contact with these caterpillars can result in acquired bleeding disorders due to potent fibrinolytic activity of these hemolymph proteases (Amarant *et al*., 1991; Arocha-Pinango and Guerrero, 2000).

Extracellular serine protease cascades mediate rapid responses to infection and wounding in vertebrate and invertebrate animals. Biochemical and genetic evidence indicates that activation of serine proteases in arthropod hemolymph is a component of several immune responses, including coagulation, melanotic encapsulation, activation of antimicrobial peptide synthesis, and modulation of hemocyte function (Yoshida and Ashida, 1986; Katsumi *et al*., 1995; Kawabata *et al*., 1996; Ashida and Brey, 1998; Levashina *et al*., 1999; Jiang and Kanost, 2000; Gorman and Paskewitz, 2001; Kanost *et al*., 2001). Microbial challenge induces expression of many nondigestive serine proteases (Dimopoulos and Della Torre, 1996; Oduol *et al*., 2000; Irving *et al*., 2001; Dimopoulos *et al*., 2002; Zhu *et al*., 2003a). In *Anopheles gamibiae*, a large, multidomain serine protease, Sp22D, is expressed in hemocytes, fat body, and midgut and is secreted into the hemolymph, with a low level of induced expression in response to infection (Danielli *et al*., 2000; Gorman *et al*., 2000a). Sp22D contains chitin-binding domains, lipoprotein receptor class A domains, scavenger receptor domains, and a C-terminal serine protease domain. This complex domain architecture with multiple domains that might function in binding to proteins or polysaccharides suggests that Sp22D may participate in the formation of protein complexes in defense responses or perhaps in tissue remodeling at metamorphosis.

Most of the hemolymph proteases are expressed in fat body or hemocytes, but a bacteria-induced protease, scolexin, from *Manduca sexta* is expressed in epidermis (Kyriakides *et al*., 1995; Finnerty *et al*., 1999). Another novel serine protease expressed in pupal yellow body of *Sarcophaga peregrina* has antibacterial activity distinct from its protease activity (Nakajima *et al*., 1997; Tsuji *et al*., 1998).

Serine proteases that contain a C-terminal protease domain and an N-terminal clip domain are known to act in cascade pathways in arthropod hemolymph (Jiang and Kanost, 2000). Among clip domain proteases with known function are horseshoe crab proclotting enzyme and clotting factor B (Kawabata *et al*., 1996), *D. melanogaster* snake and easter, and phenoloxidase-activating proteases described below. Clip domains are 35–55 amino acid residue sequences that contain three conserved disulfide bonds. They may function to mediate interaction of members of protease cascade pathways. Proteases may contain one or more N-terminal clip domains, followed by a 20–100 residue linking sequences connecting them to the catalytic protease domain.

Insects that have been investigated in some detail are known to contain a large number of genes for clip domain proteases. Among the 204 genes with homology to the S1 serine protease family in the *D. melanogaster* genome, 24 are clip domain proteases, most of whose functions are unknown (Ross *et al*., 2003). The *A. gambiae* genome contains 41 clip domain proteases (Christophides *et al*., 2002) but only a few have been studied in detail (Han *et al*., 1997; Paskewitz *et al*., 1999; Gorman *et al*., 2000b; review: Gorman and Paskewitz, 2001). In *M. sexta*, more than 20 clip domain proteases expressed in fat body or hemocytes have been identified (Jiang *et al*., 1999; Jiang, Y. Wang and M.R. Kanost, unpublished results; Kanost *et al*., 2001).

Melanization, a response to wounding and infection in insects and crustaceans, involves activation of a cascade of serine protease zymogens (**Figure 3**). This pathway leads to rapid activation of a protease that in turn activates a phenoloxidase zymogen (prophenoloxidase; proPO) (Söderhäll *et al*., 1996; Ashida and Brey, 1998). Oxidation of phenols by phenoloxidase leads to production of quinones that polymerize to form melanin (see **Chapters 4.3** and **4.4**). Melanization of encapsulated parasites is believed to be an important defensive response in insects, including insect vectors of human diseases

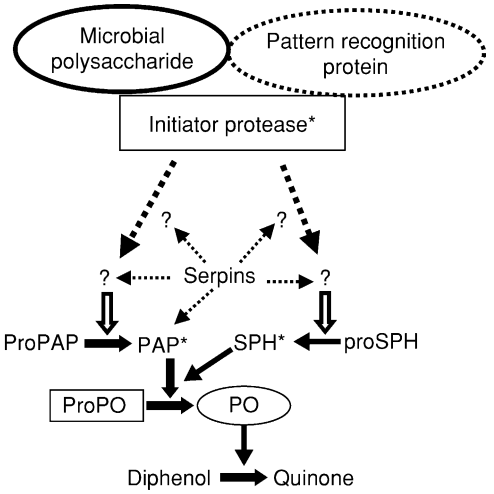


Figure 3 A model of the protease cascade that activates prophenoloxidase in response to infection. Hemolymph plasma proteins known as pattern recognition proteins bind to polysaccharides on the surface of microorganisms. This interaction leads to activation of initiator protease(s) by a mechanism not yet understood, which triggers a protease cascade. The number of proteases in the pathway is not yet known (indicated by dashed arrows). Activation of uncharacterized proteases leads to activation of activator of prophenoloxidase activating protease (PAP) and serine protease homologs (SPH) that function together to form a prophenoloxidase activating enzyme, which cleaves prophenoloxidase (ProPO) to form active phenoloxidase (PO). PO catalyzes the oxidation of hemolymph catecholic phenols to corresponding quinones, which can undergo further reactions to form melanin. Proteases in the pathway are regulated by serine protease inhibitors known as serpins.

(Gillespie *et al.*, 1997; Paskewitz and Gorman, 1999). Serine proteases that activate prophenoloxidase have been purified, and corresponding cDNAs have been cloned from two lepidopteran insects: *M. sexta* (Jiang *et al.*, 1998, 2003a, 2003b) and *B. mori* (Satoh *et al.*, 1999); a beetle, *Holotrichia diomphalia* (Lee *et al.*, 1998a, 1998b); and a crayfish (Wang *et al.*, 2001). All of these enzymes contain a C-terminal serine protease catalytic domain and one or two N-terminal clip domains, similar to horseshoe crab clotting enzyme and *D. melanogaster* easter (Jiang and Kanost, 2000). These proteases are synthesized as zymogens, which must be activated by a protease upstream in the pathway. However, it is not known how many steps occur between recognition of a microorganism by a hemolymph protein and activation of the terminal protease in the pathway. Furthermore, it is not yet understood how the first protease in the cascade is activated. Proteins that bind to microbial surfaces (pattern recognition proteins) are probably involved in this step, through interaction with microbial polysaccharides and an initiating protease. Plasma proteins that bind to bacterial lipopolysaccharide, peptidoglycan, or fungal glucans can stimulate activation of the prophenoloxidase cascade (Ochiai and Ashida, 1988, 1999, 2000; Söderhäll *et al.*, 1988; Yu *et al.*, 1999, 2002; Ma and Kanost 2000; Yu and Kanost, 2000).

Initial characterization of a proPO-activating proteinase in *M. sexta* indicated that the purified protease could not efficiently activate proPO, but required participation of a nonproteolytic protein fraction (Jiang *et al.*, 1998). This protein cofactor was identified in *H. diomphalia* (Kwon *et al.*, 2000) and *M. sexta* (Yu *et al.*, 2003) as a protein with a clip domain and a serine protease domain, in which the active site serine residue is changed to glycine. There are 13 such clip domain serine protease homolog genes in the *D. melanogaster* genome (Ross *et al.*, 2003). These serine protease homologs lack protease activity due to the incomplete catalytic triad and must therefore have other functions. One *D. melanogaster* serine protease homolog, masquerade, functions in nerve and muscle development (Morugasu-Oei *et al.*, 1995, 1996). A serine protease homolog in crayfish hemolymph has a role in immune responses, indicating evolutionary conservation of function in these unusual proteins (Lee and Söderhäll, 2001). The active form of the serine protease homologs that function as cofactors for proPO activation are themselves activated through specific cleavage by a serine protease in hemolymph (Kim *et al.*, 2002; Lee *et al.*, 2002b; Yu *et al.*, 2003) adding additional complexity to this pathway. The serine protease homologs from *M. sexta* that stimulate proPO activation bind to a hemolymph lectin, which is a recognition protein for lipopolysaccharides from Gram-negative bacteria, and to proPO and prophenoloxidase activating protease (Yu *et al.*, 2003). The interaction between the lectin and a proPO activation complex may serve to localize melanin synthesis to the surface of invading bacteria.

Serine proteases are also involved in other insect immune responses. The signal transduction system that regulates dorsal/ventral development in *D. melanogaster* embryos also regulates expression of the gene for an antifungal peptide in larvae and adults (review: Hoffmann, 2003). In embryonic development, this pathway involves an extracellular serine protease cascade that eventually cleaves an inactive protein, spätzle, making it competent to bind to a transmembrane receptor named Toll. This same receptor–ligand interaction activates a signal pathway that leads to activation of rel family transcription factors that stimulate expression of drosomycin, an antifungal peptide normally synthesized by the fat body after microbial challenge.

Mutants deficient in Toll or spätzle exhibit decreased induction of drosomycin (Lemaitre *et al*., 1996). These results suggest that infection leads to proteolytic processing of spätzle in the hemolymph. The activated spätzle binds to the Toll receptor in fat body membranes, leading to activation of the Dif pathway and expression of drosomycin. Mutants of gastrulation defective, snake, or easter do not have an impaired antifungal response, indicating that a different set of proteases functions in the immune response protease cascade. Like prophenoloxidase activation, this pathway is initiated by interactions of pattern recognition proteins with microbial surface polysaccharides (Kim *et al*., 2000; Leulier *et al*., 2003), presumably stimulating activation of the first protease in the cascade. A clip domain protease known as persephone has been shown to participate in this pathway (Ligoxygakis *et al*., 2002a), but its position in the cascade is not yet known.

4.7.3.2. Protease Inhibitors

Insect hemolymph contains high concentrations of serine protease inhibitors from several different gene families (reviews: Kanost and Jiang, 1996; Polanowski and Wilusz, 1996; Kanost, 1999). Protease cascade pathways in mammalian blood are regulated by ~45 kDa protease inhibitors known as serpins (Silverman *et al*., 2001; Gettins, 2002). Serpins in arthropod hemolymph also function to regulate protease cascades, preventing detrimental effects of uncontrolled immune responses. For example, each of the proteases in the horseshoe crab coagulation pathway is regulated by serpins produced by hemocytes (Kawabata *et al*., 1996). Fourteen serpin genes have been identified in the *A. gambiae* genome (Christophides *et al*., 2002) and 26 serpin genes have been annotated in the *D. melanogaster* genome. In *M. sexta*, 7 serpin genes have been identified (Jiang *et al*., 1996; Gan *et al*., 2001; Zhu *et al*., 2003b). However, the physiological target proteases of only a few insect serpins have been determined (Ashida and Sasaki, 1994; Jiang *et al*., 2003b; Zhu *et al*., 2003a).

The reactive site in a serpin protein that interacts with the target protease is part of an exposed loop near the C-terminal end of the serpin sequence. Some insect serpin genes have a unique structure in which mutually exclusive alternate splicing of an exon that encodes the reactive site loop results in production of several inhibitors with different specificity. This was first observed in the gene for *M. sexta* serpin-1, which contains 12 copies of its 9th exon. Each version of exon 9 encodes a different reactive site loop sequence and inhibits a different spectrum of proteases (Jiang *et al*., 1996; Jiang and Kanost, 1997). Structures of two of the *M. sexta* serpin-1 variants have been determined by X-ray crystallography (Li *et al*., 1999; Ye *et al*., 2001). Serpin genes with alternate exons in the same position as in *M. sexta* serpin-1 have been identified in other insect species, including the lepidopterans *B. mori* (Sasaki, 1991; Narumi *et al*., 1993) and *Mamestra configurata* (Chamankhah *et al*., 2003), in the dipterans *D. melanogaster* (Kruger *et al*., 2002) and *A. gambiae* (Danielli *et al*., 2003), and in the cat flea, *Ctenocephalides felis* (Brandt *et al*., 2004).

Like blood clotting, phenoloxidase activation is normally regulated *in vivo* as a local reaction of brief duration. This regulation involves serine protease inhibitors in plasma (Kanost and Jiang, 1996; Kanost, 1999). Two serpins from the hemolymph of *M. sexta* (serpin-1J and serpin-3) can inhibit proPO activating proteases (Jiang *et al*., 2003a, 2003b; Zhu *et al*., 2003a, 2003b). In *D. melanogaster*, serpin 27A (an apparent ortholog of *Manduca* serpin-3) is involved in regulation of melanization (DeGregorio *et al*., 2002; Ligoxygakis *et al*., 2002b). Similarly, a mutation in *D. melanogaster* serpin 43Ac (Necrotic) leads to constitutive expression of drosomycin, indicating that this serpin regulates a protease in the cascade that processes spätzle (**Figure 4**) (Levashina *et al*., 1999; Green *et al*., 2000). It is not yet known whether serpin 43Ac inhibits persephone or some other protease in the pathway.

In addition to serpins, lower molecular weight inhibitors from the Kunitz family (Sugumaran *et al*., 1985; Saul and Sugumaran, 1986; Aso *et al*., 1994) and a family of 4 kDa inhibitors from locusts (Boigegrain *et al*., 1992) can interfere with proPO activation (review: Kanost, 1999), although it is not yet known which proteases in the pathway they can inhibit.

4.7.4. Cellular Proteases

4.7.4.1. Serine Proteases

Serine proteases appear to function in tissue remodeling and extracellular matrix degradation required for cell movements in metamorphosis. Evagination of *D. melanogaster* imaginal discs is blocked by serine protease inhibitors, and the discs release serine proteases (Pino-Heiss and Schubiger, 1989). The *Stubble-stubbloid* gene encodes an integral membrane protein with an extracellular serine protease domain. It is expressed in imaginal discs under control of 20-hydroxyecdysone (see

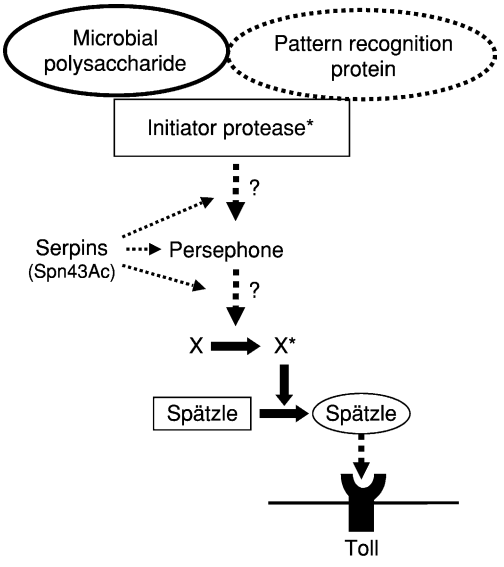


Figure 4 A model of the protease cascade that activates the *D. melanogaster* Toll pathway for synthesis of antimicrobial peptides. Hemolymph plasma proteins known as pattern recognition proteins bind to polysaccharides on the surface of microorganisms. This interaction leads to activation of initiator protease(s) by a mechanism not yet understood, which triggers a protease cascade. The number of proteases in the pathway is not yet known (indicated by dashed arrows). A clip domain protease called Persephone is known from genetic data to be a component in the pathway. The cascade leads to activation of a protease (designated as X in this figure) that cleaves spätzle to produce an active ligand that binds to Toll, a transmembrane receptor. A serpin, Spn43Ac negatively regulates at least one protease in the pathway.

Chapter 3.3) and may function to detach imaginal discs from extracellular matrices (Appel *et al*., 1993).

4.7.4.2. Cathepsin-Type Cysteine Proteases

Cysteine proteases related to cathepsin B and cathepsin L have been identified as proteins produced by hemocytes that participate in tissue remodeling in metamorphosis of several insects. In *S. peregrina*, a 26/29 kDa protease synthesized in hemocytes was identified as a cathepsin B (Saito *et al*., 1992; Kurata *et al*., 1992a; Takahashi *et al*., 1993; Fujimoto *et al*., 1999). This protease may be released from pupal hemocytes to cause dissociation of fat body at metamorphosis (Kurata *et al*., 1992b). A cathepsin B from hemocytes of *B. mori* may also function in tissue degradation during metamorphosis, including histolysis of silk glands (Shiba *et al*., 2001; Xu and Kawasaki, 2001; **Chapter 2.11**).

Cysteine proteases classified as cathepsin L have also been identified as participants in tissue remodeling at metamorphosis. A cathepsin L from *S. peregrina* appears to function in the differentiation of imaginal discs (Homma *et al*., 1994; Homma and Natori, 1996). A similar cathepsin L from another dipteran, *Delia radicum*, is expressed highly in midgut beginning in late third instar and may function in metamorphosis of the midgut (Hegedus *et al*., 2002). A cathepsin L-like protease expressed in a *D. melanogaster* haemocyte cell line is present in granules and may be a lysosomal enzyme functioning to degrade phagocytosed material (Tryselius and Hultmark, 1997).

4.7.4.3. Caspases

Programmed cell death, known as apoptosis, is an essential process in development (Richardson and Kumar, 2002; **Chapter 2.5**) and is also a response by insect cells to viral infection (Clarke and Clem, 2003). An intracellular cascade of cysteine proteases called caspases (MEROPS family C14) is a key pathway in initiating apoptosis. Caspases are conserved in structure and function in animal species. They cleave target protein substrates after Asp or Glu P1 residues. Caspases are synthesized as inactive zymogens. They are activated by specific proteolytic cleavage, yielding a small and a large subunit that together form an active site. These heterodimers further dimerize to form a hetrotetramer with two active sites (Earnshaw *et al*., 1999).

Caspases are divided into two groups, initiator caspases that interact with apoptosis-initiating proteins, and effector caspases, which are activated by the initiators. Initiator caspases contain a C-terminal cysteine protease domain and a long N-terminal region that contains death effector domains (DEDs) or caspase recruitment domains (CARDs), which interact with proteins that stimulate apoptosis. Dimerization of initiator caspases occurs when they associate with apoptosis-promoting proteins, causing caspase activation through conformational change in the absence of proteolytic cleavage (Boatright *et al*., 2003). The initiator caspases then activate the effector caspases through specific proteolytic cleavage. Effector caspases, which lack a long N-terminal domain, cleave structural components of the cytoskeleton and nucleus and proteins involved in signaling pathways, resulting in death of the cell. Cleavage by effector caspases activivates death-promoting enzymes, including certain kinases and nucleases (Earnshaw *et al*., 1999).

Apoptosis and caspases have been investigated in detail in *D. melanogaster* (**Figure 5**). The *D. melanogaster* genome contains seven caspase genes. Three genes encode caspases with long prodomains (*dronc*, *dcp-2*/*dredd*, and *strica*/*dream*) and four genes encode caspases with short prodomains (*dcp-1*, *drICE*, *decay*, and *damm*/*daydream*) (Vernooy *et al*., 2000; Kumar and Doumanis, 2000).

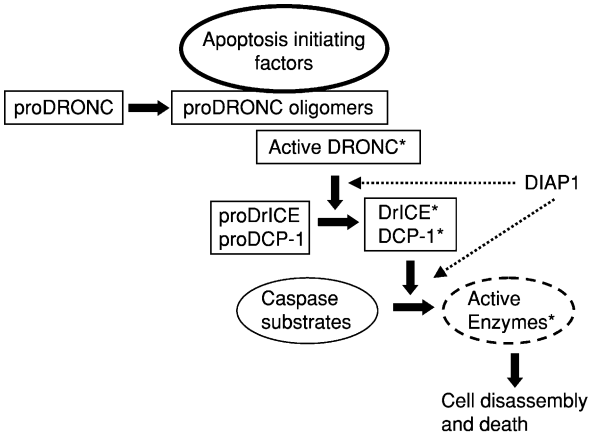


Figure 5 A model for a caspase cascade pathway in apoptosis in *D. melanogaster*. Cellular proapoptotic protein factors initiate oligomerization of the caspase zymogen proDRONC, which results in its self-processing and activation. Active DRONC cleaves the effector caspases DrICE and DcP-1, which then become active and cleave protein substrates including inactive forms of nucleases and kinases. The resulting active enzymes then catalyze reactions that result in progressive disassembly of the nucleus and cytoskeleton and cell death. A caspase inhibitor DIAP1 inhibits DRONC, DCP-1, and DrICE to regulate their activities.

Dronc contains an N-terminal CARD, and its overexpression in flies or in cultured cells stimulates apoptosis (Dorstyn *et al.*, 1999a; Quinn *et al.*, 2000). Dronc expression is stimulated by ecdysteroid, and it appears to function in apoptosis in tissue remodeling at metamorphosis (Dorstyn *et al.*, 1999a; Cakouros *et al.*, 2002; Lee *et al.*, 2002b; **Chapter 2.5**). Depletion of Dronc by RNA interference results in a decrease in normal programmed cell death in embryogenesis (Quinn *et al.*, 2000). Dronc can cleave substrates with P1 Glu or Asp residues. It can autoactivate by cleaving following a glutamate residue, but it activates drICE by cutting at an aspartate residue (Hawkins *et al.*, 2000). Dredd is an initiator caspase that participates in apoptosis (Chen *et al.*, 1998). Dredd also functions in a signal transduction pathway that leads to activation of immune responses (Elrod-Erickson *et al.*, 2000; Leulier *et al.*, 2000; Georgel *et al.*, 2001) through its cleavage and activation of a rel family transcription factor called relish (Stoven *et al.*, 2000, 2003). The third initiator caspase in *D. melanogaster*, STRICA, lacks either a CARD or a DED and instead contains a serine/threonine rich prodomain, suggesting that STRICA functions in an as yet unidentified pathway (Doumanis *et al.*, 2001).

The effector caspases Dcp-1 and DrICE are processed by active DRONC and then cleave target proteins that result in the disassembly of the cell. Both caspases induce apoptosis when overexpressed, have preferred target cleavage sites very similar to those of mammalian caspase-3, and can be inhibited by the baculovirus protein P35, a powerful inhibitor of effector caspases (Fraser and Evan, 1997; Fraser *et al.*, 1997; Song *et al.*, 1997; Meier *et al.*, 2000). Dcp-1 and DrICE have slight differences in specificity. Active DCP-1 can activate other proDCP-1 proteins and proDrICE, whereas active DrICE can not cleave and activate proDrICE (Song *et al.*, 2000). A third effector caspase, DECAY, shares similarity with DrICE and DCP-1 in sequence and preferred cleavage site, and can induce apoptosis when ectopically expressed in cultured cells (Dorstyn *et al.*, 1999b). In contrast, the short-prodomain caspase DAMM appears to have an accessory role in apoptosis, sensitizing the cell to prodeath signals rather then directly killing the cell. DAMM shares strong sequence similarity with STRICA, has a preferred cleavage site similar to mammalian caspases -1, -4, and -5, and is not processed *in vitro* by DAMM, DRONC, DECAY, DCP-1, or DrICE. DAMM mRNA is upregulated in apoptotic tissues, and DAMM over-expression in eye tissue sensitizes the cells to apoptosis from either developmental signals or radiation (Harvey *et al.*, 2001).

The caspase pathways in other insects may be even more complex. For example, the *A. gambiae* genome contains 12 caspase genes, compared to 7 in *D. melanogaster*. An effector caspase has been identified in the lepidopteran *Spodptera frugiperda* (Ahmad *et al.*, 1997; Seshagiri and Miller, 1997). This enzyme, Sf-caspase-1, is activated by an initiator caspase equivalent to DRONC (Manji and Friesen, 2001). The three-dimensional structure of Sf-caspase-1 has been determined (the first structure of a nonhuman caspase) by X-ray crystallography (Forsyth *et al.*, 2004) in the presence of a tetrapeptide inhibitor. The Sf-caspase-1 fold is very similar to that of human effector caspases, but it has unique features in its N-terminal prodomain that may be important in regulating its activation.

Caspases are regulated by specific inhibitor proteins that are present in the cytoplasm. In *D. melanogaster*, cells appear to be constitutively primed for caspase activation, and removal of caspase inhibition is sufficient to cause cell death. Three inhibitors of apoptosis (IAP) proteins, which inhibit caspases, have been identified in *D. melanogaster* (Hay *et al.*, 1995; Jones *et al.*, 2000). DIAP1 inhibits Dronc, DrICE, and DCP-1 (Kaiser *et al.*, 1998; Hawkins *et al.*, 1999; Meier *et al.*, 2000; Wilson *et al.*, 2002). *D. melanogaster* exhibits a constitutive proapoptotic predisposition, resulting in rapid apoptosis of cultured cells or embryos in the event of

the decreased expression of DIAP1 (Igaki *et al.*, 2002; Muro *et al.*, 2002; Rodriguez *et al.*, 2002).

Iap genes, similar to *D. melanogaster* diap1 and sharing the same capability of inhibiting initiator caspases, have been identified from three lepidopterans, *Trichoplusia ni* (Liao *et al.*, 2002; Seshagiri *et al.*, 1999), *B. mori* (Huang *et al.*, 2001), and *S. frugiperda* (Huang *et al.*, 2000). RNA interference depletion of Sf-IAP from *S. frugiperda* cell lines results in apoptosis, suggesting that, like *Drosophila* cells, *Spodoptera* cells are maintained in a state predisposed to cell death (Muro *et al.*, 2002).

Baculoviruses have evolved two types of caspase inhibitors that can block the lepidopteran apoptotic program. AcMNPV expresses its gene for the anti-apoptotic protein P35 at both early and late times during an infection, thus preventing infected cells from prematurely terminating viral replication through apoptosis (Clem *et al.*, 1991). P35 is a globular protein with an extended loop (Fisher *et al.*, 1999; Zoog *et al.*, 1999) that contains a caspase cleavage site recognized by a wide range of effector caspases, including human caspases-1, -3, -6, -7, -8, -10 (Zhou *et al.*, 1998), Sf-caspase-1 (Ahmad *et al.*, 1997), and DrICE (Fraser *et al.*, 1997), but not by initiator caspases such as human caspase-9 (Vier *et al.*, 2000), DRONC (Meier *et al.*, 2000), or Sf-caspase-X (LaCount *et al.*, 2000). When an effector caspase cleaves in the reactive site loop of P35, a thioester bond forms between the active site cysteine of the caspase and the P1 aspartic acid residue of P35, causing a conformational change in P35 that blocks access to water molecules, thus preventing completion of the peptide bond hydrolysis (Bertin *et al.*, 1996; dela Cruz *et al.*, 2001; Riedl *et al.*, 2001; Xu *et al.*, 2001).

Baculoviruses also contain genes for IAP proteins that block apoptosis by inhibiting initiator caspases (Huang *et al.*, 2000b; Birnbaum *et al.*, 1994; LaCount *et al.*, 2000; Means *et al.*, 2003) suggesting that, at some point during their evolution, baculoviruses co-opted the cell's own mechanism for regulating caspase activity. In fact, the first IAP protein to be discovered was Cp-IAP from the baculovirus *Cydia pomonella* granulovirus (Crook *et al.*, 1993).

4.7.4.4. Metalloproteases

Three families of metalloproteases that are not involved in digestion of food have been targets of fairly limited investigation in insects. They appear to function in remodeling of the extracellular matrix or in degradation of peptide hormones.

Matrix metalloproteases (MMP) are integral membrane proteins, present on the outer surface of cells. They are multidomain proteins that include a catalytic domain that incorporates a zinc ion in the active site. In mammals, these enzymes regulate processes involving morphogenesis in development and tissue remodeling by digesting protein components of the extracellular matrix (Nagase and Woessner, 1999). In *D. melanogaster* there are two MPP genes. Dm1-MMP is expressed strongly in embryos and may have a role in remodeling of the extracellular matrix in development of the central nervous system (Llano *et al.*, 2000). Dm2-MMP is expressed at a low level at all developmental stages, but with strong expression in regions of the brain and the eye imaginal discs (Llano *et al.*, 2002). Mutants in Dm1-MMP have defects in larval tracheal development and pupal head eversion, whereas mutants in Dm2-MMP do not undergo proper tissue remodeling during metamorphosis. However, normal embryonic development was observed even in double mutants lacking both MMPs (Page-McCaw *et al.*, 2003). It appears that these proteases are required primarily for remodeling of the extracellular matrix in metamorphosis. In mammals, the MMPs are regulated by proteins called tissue inhibitors of metalloproteases (TIMP). A homolog of these proteins in *D. melanogaster* has been shown to inhibit its MMPs (Wei *et al.*, 2003).

A different class of metalloproteases known as ADAMs (because they contain a disintegrin and metalloprotease domain) are also integral membrane proteins, with the zinc-containing protease domain on the extracellular surface. In mammals, they participate in growth factor processing, cell adhesion, cell fusion, and tissue remodeling processes, although their physiological functions, particularly of the protease domain, are still not well understood (Seals and Courtneidge, 2003). A *D. melanogaster* ADAM called kuzbanian has been shown to function in axon extension in nervous system development (Fambrough *et al.*, 1996; Rooke *et al.*, 1996). Kuzbanian is involved in initiation of a signal transduction pathway by a transmembrane receptor called Notch (Sotillos *et al.*, 1997). The metalloprotease domain of Kuzbanian may exert its physiological effect through cleavage of Notch (Lieber *et al.*, 2002) or the Notch ligand, Delta (Bland *et al.*, 2003).

A third family of zinc metalloproteases identified in insects contains members that may function in the degradation of peptide hormones. These transmembrane proteins are similar to mammalian neprilysins (also called neutral endopeptidases), which cleave oligopeptides on the amino side of hydrophobic residues. They cleave physiologically active signaling peptides with functions in nervous, cardiovascular,

inflammatory, and immune systems and have a wide tissue distribution (Turner *et al*., 2001). Metalloprotease activities with properties similar to neprylins have been identified as enzymes that can degrade tachykinin-related peptides in a cockroach, *Leucophaea maderae*, a locust, *Locusta migratoria,* a dipteran, *D. melanogaster,* and a lepidopteran, *Lacanobia oleracea* (Isaac *et al*., 2002; Isaac and Nassel, 2003). Similar activities that degrade adipokinetic hormone have been identified in the lepidopterans *Lymantria dispar* (Masler *et al*., 1996) and *M. sexta* (Fox and Reynolds, 1991), and a dipteran, *M. domestica* (Lamango and Isaac, 1993). cDNA clones for proteases with sequence similarity to neprilysin have been described in *B. mori* (Zhao *et al*., 2001), *M. sexta* (Zhu *et al*., 2003a), and *L. migratoria* (Macours *et al*., 2003). The *D. melanogaster* genome contains 24 neprilysin-like genes (Coates *et al*., 2000). Further study is needed to determine the substrates and physiological roles of these proteases, but it seems likely that they function as negative regulators of peptide hormones.

4.7.4.5. Aspartic Acid Proteases

The aspartic acid proteases from the MEROPS families A1 and A2 (similar to human pepsin) are a group that has received little study in insects. The *D. melanogaster* genome contains 34 genes that encode proteins similar to these enzymes, but none appear to have been investigated at the molecular level. A cathepsin D-like aspartic acid protease from *A. aegypti* has been identified as a lysosomal enzyme, which accumulates in fat body during vitellogenin synthesis (Cho and Raikhel, 1992; Dittmer and Raikhel, 1997).

An enzymatically inactive protein from the A1 family is an important allergen from a cockroach (*Blattella germanica*) that triggers asthmatic responses in humans (Pomés *et al*., 2002). It is present at highest concentration in the gut (Arruda *et al*., 1995), but since it apparently lacks proteolytic activity, its function in the insect is unknown.

4.7.4.6. Proteasomes

Proteasomes are complex intracellular proteases that function in regulated degradation of cellular proteins. Turnover of proteins by the proteasome regulates many processes including the cell cycle, circadian cycles (see **Chapter 4.11**), transcription, growth, development, as well as removal of abnormal proteins. Proteins are targeted for degradation by the proteasome by attachment of polyubiquitin chains to an amino group on a lysine side chain. Eukaryotic proteasomes are composed of four stacked heptameric rings that form a cylinder with multiple protease catalytic sites in its interior. This structure, the 20S proteasome, is composed of 28 subunits from multiple homologous gene products and has a mass of ~700 kDa. The 20S proteasome has little activity unless it is activated by another 700 kDa, 20 subunit protein, PA700, that can bind to one or both ends of the cylinder. When both ends of the 20S proteasome are capped by a PA700, the resulting complex is the 26S proteasome, which is active in degrading ubiquitinated proteins, an ATP-dependent process (DeMartino and Slaughter, 1999).

Proteasomes in insects have been studied primarily in *D. melanogaster* and *M. sexta* (Mykles, 1997, 1999). They have physical and catalytic properties similar to those of proteasomes from other eukaryotic species (Uvardy, 1993; Haire *et al*., 1995; Walz *et al*., 1998). Mutations in genes for subunits of *D. melanogaster* proteasomes or proteins, that regulate proteasome activity, result in lethal or otherwise complex phenotypes involving the disruption of multiple aspects of physiology and development (Schweisguth, 1999; Ma *et al*., 2002; Watts *et al*., 2003). Some mutations alter subunit composition or disrupt proteasome assembly (Covi *et al*., 1999; Smyth and Belote, 1999; Szlanka *et al*., 2003). Recent studies have used double-stranded RNA interference to study disruptions caused by reduced expression of individual proteasome components (Wojcik and DeMartino, 2002; Lundgren *et al*., 2003). In *M. sexta*, programmed cell death of intersegmental muscles at metamorphosis is accompanied by marked changes in proteasome activity and subunit composition (Dawson *et al*., 1995; Jones *et al*., 1995; Takayanagi *et al*., 1996; Low *et al*., 1997, 2000, 2001; Hastings *et al*., 1999).

4.7.5. Conclusions and Future Prospects

Tremendous advances have been made in the last 10 years in our knowledge of the existence, structure, and function of insect proteases that have biological roles unrelated to digestion of food. Much more remain to be discovered. Through examination of the genome sequences of *D. melanogaster* and *A. gambiae*, we can see that these insects have an enormous number of genes encoding proteases and that most of them are unstudied and have unknown functions. Nearly all of the research in this area has focused on a few dipteran and lepidopteran species. It is to be expected that detailed investigation of a broader range of species will reveal complex and diverse functions for proteases

in regulating intracellular and extracellular processes. A common feature of proteases is that they are synthesized as inactive zymogens, activated by proteolysis when the time is right, and then rapidly inhibited by specific inhibitors. Better understanding of molecular mechanisms of this tight regulation of multiple and varied protease cascade pathways will impact many areas of insect biology.

References

Ahmad, M., Srinivasula, S.M., Wang, L., Litwack, G., Fernandes-Alnemri, T., *et al.*, **1997**. *Spodoptera frugiperda* caspase-1, a novel insect death protease that cleaves the nuclear immunophilin FKBP46, is the target of the baculovirus antiapoptotic protein p35. *J. Biol. Chem. 272*, 1421–1424.

Amarant, T., Burkhart, W., LeVine, H., 3rd, Arocha-Pinango, C.L., Parikh, I., **1991**. Isolation and complete amino acid sequence of two fibrinolytic proteinases from the toxic Saturnid caterpillar *Lonomia achelous*. *Biochem. Biophys. Acta 1079*, 214–221.

Appel, L.F., Prout, M., Abu-Shumays, R., Hammonds, A., Garbe, J.C., *et al.*, **1993**. The *Drosophila Stubble-stubbloid* gene encodes an apparent transmembrane serine protease required for epithelial morphogenesis. *Proc. Natl Acad. Sci. USA 90*, 4937–4941.

Applebaum, S.W., **1985**. Biochemistry of digestion. In: Kerkut, G.A., Gilbert, L.I. (Eds.), Comprehensive Insect Physiology Biochemistry, and Pharmacology, vol. 4. Pergamon, Oxford, pp. 279–311.

Arocha-Pinango, C.L., Guerrero, B., **2000**. *Lonomia* genus caterpillar toxins: biochemical aspects. *Biochemie 82*, 937–942.

Arruda, L.K., Vailes, L.D., Mann, B.J., Shannon, J., Fox, J.W., *et al.*, **1995**. Molecular cloning of a major cockroach (*Blattella germanica*) allergen, Bla g 2. Sequence homology to the aspartic proteases. *J. Biol. Chem. 270*, 19563–19568.

Ashida, M., Brey, P.T., **1998**. Recent advances on the research of the insect prophenoloxidase cascade. In: Brey, P.T., Hultmark, D. (Eds.), Molecular Mechanisms of Immune Responses in Insects. Chapman and Hall, London, pp. 135–172.

Ashida, M., Sasaki, T., **1994**. A target protease activity of serpins in insect hemolymph. *Insect Biochem. Molec. Biol. 24*, 1037–1041.

Aso, Y., Yamashita, T., Meno, K., Murakami, M., **1994**. Inhibition of prophenoloxidase-activating enzyme from *Bombyx mori* by endogenous chymotrypsin inhibitors. *Biochem. Mol. Biol. International 33*, 751–758.

Belvin, M.P., Anderson, K.V., **1996**. A conserved signaling pathway: the *Drosophila* toll-dorsal pathway. *Annu. Rev. Cell Devel. Biol. 12*, 393–416.

Berger, E., Kafatos, F.C., Felsted, R.L., Law, J.H., **1971**. Cocoonase III. Purification, preliminary characterization, and activation of the zymogen of an insect protease. *J. Biol. Chem. 246*, 4131–4137.

Bertin, J., Mendrysa, S.M., LaCount, D.J., Gaur, S., Krebs, J.F., *et al.*, **1996**. Apoptotic suppression by baculovirus P35 involves cleavage by and inhibition of a virus-induced CED-3/ICE-like protease. *J. Virol. 70*, 6251–6259.

Birnbaum, M.J., Clem, R.J., Miller, L.K., **1994**. An apoptosis-inhibiting gene from a nuclear polyhedrosis virus encoding a polypeptide with Cys/His sequence motifs. *J. Virol. 68*, 2521–2528.

Bland, C.E., Kimberly, P., Rand, M.D., **2003**. Notch-induced proteolysis and nuclear localization of the Delta ligand. *J. Biol. Chem. 278*, 13607–13610.

Boatright, K.M., Renatus, M., Scott, F.L., Sperandio, S., Shin, H., *et al.*, **2003**. A unified model for apical caspase activation. *Mol. Cell. 11*, 529–541.

Boigegrain, R., Mattras, H., Brehélin, M., Paroutaud, P., Coletti-Previero, M., **1992**. Insect immunity: two proteinase inhibitors from hemolymph of *Locusta migratoria*. *Biochem. Biophys. Res. Commun. 189*, 790–793.

Brandt, K.S., Silver, G.M., Becher, A.M., Gaines, P.J., Maddux, J.D., *et al.*, **2004**. Isolation, characterization, and recombinant expression of multiple serpins from the cat flea, *Ctenocephalides felis*. *Arch. Insect Biochem. Physiol. 55*, 200–214.

Cakouros, D., Daish, T., Martin, D., Baehrecke, E.H., Kumar, S.J., **2002**. Ecdysone-induced expression of the caspase DRONC during hormone-dependent programmed cell death in *Drosophila* is regulated by Broad-Complex. *Cell Biol. 157*, 985–995.

Chamankhah, M., Braun, L., Visal-Shah, S., O'Grady, M., Baldwin, D., *et al.*, **2003**. *Mamestra configurata* serpin-1 homologues: cloning, localization and developmental regulation. *Insect Biochem. Mol. Biol. 33*, 355–369.

Chang, A.J., Morisato, D., **2002**. Regulation of Easter activity is required for shaping the Dorsal gradient in the *Drosophila* embryo. *Development 129*, 5635–5645.

Chasan, R., Anderson, K., **1989**. The role of easter, an apparent serine protease, in organizing the dorsal–ventral pattern of the *Drosophila* embryo. *Cell 56*, 391–400.

Chen, P., Rodriguez, A., Erskine, R., Thach, T., Abrams, J.M., **1998**. Dredd, a novel effector of the apoptosis activators REAPER, Grim, and Hid in *Drosophila*. *Devel. Biol. 201*, 202–216.

Cho, W.-L., Deitsch, K.W., Raikhel, A.S., **1991**. An extraovarian protein accumulated in mosquito oocytes is a carboxypeptidase activated in embryos. *Proc. Natl Acad. Sci. USA 88*, 10821–10824.

Cho, W.-L., Raikhel, A., **1992**. Cloning of cDNA for mosquito lysosomal aspartic protease. *J. Biol. Chem. 267*, 21823–21829.

Cho, W.-L., Tsao, S.-M., Hays, A.R., Walter, R., Chen, J.-S., *et al.*, **1999**. Mosquito cathepsin B-like protease involved in embryonic degradation of vitellin is produced as a latent extraovarian precursor. *J. Biol. Chem. 274*, 13311–13321.

Christophides, G.K, Evgeny, Z., Barillas-Mury, C., Birney, E., Blandin, S., *et al.*, **2002**. Immunity-related genes

and gene families in *Anopheles gambiae*. *Science 298*, 159–165.

Clarke, T.E., Clem, R.J., **2003**. Insect defenses against virus infection: the role of apoptosis. *Int. Rev. Immunol. 22*, 401–424.

Clem, R.J., Fechheimer, M., Miller, L.K., **1991**. Prevention of apoptosis by a baculovirus gene during infection of insect cells. *Science 254*, 1388–1390.

Coates, D., Siviter, R., Isaac, R.E., **2000**. Exploring the *Caenorhabditis elegans* and *Drosophila melanogaster* genomes to understand neuropeptide and peptidase function. *Biochem. Soc. Trans. 28*, 464–469.

Covi, J.A., Belote, J.M., Mykles, D.L., **1999**. Subunit compositions and catalytic properties of proteasomes from developmental temperature-sensitive mutants of *Drosophila melanogaster*. *Arch. Biochem. Biophys. 368*, 85–97.

Crook, N.E., Clem, R.J., Miller, L.K., **1993**. An apoptosis-inhibiting baculovirus gene with a zinc finger-like motif. *J. Virol. 67*, 2168–2174.

Danielli, A., Kafatos, F.C., Loukeris, T.G., **2003**. Cloning and characterization of four *Anopheles gambiae* serpin isoforms, differentially induced in the midgut by *Plasmodium berghei* invasion. *J. Biol. Chem. 278*, 4184–4193.

Danielli, A., Loukeris, T.G., Lagueux, M., Muller, H.-M., Richman, A., *et al.*, **2000**. A modular chitin-binding protease associated with hemocytes and hemolymph in the mosquito *Anopheles gambiae*. *Proc. Natl Acad. Sci. USA 97*, 7136–7141.

Dawson, S.P., Arnold, J.E., Mayer, N.J., Reynolds, S.E., Billett, M.A., *et al.*, **1995**. Developmental changes of the 26 S proteasome in abdominal intersegmental muscles of *Manduca sexta* during programmed cell death. *J. Biol. Chem. 270*, 1850–1858.

dela Cruz, W.P., Friesen, P.D., Fisher, A.J., **2001**. Crystal structure of baculovirus P35 reveals a novel conformational change in the reactive site loop after caspase cleavage. *J. Biol. Chem. 276*, 32933–32939.

De Gregorio, E., Han, S.J., Lee, W.J., Baek, M.J., Osaki, T., *et al.*, **2002**. An immune-responsive Serpin regulates the melanization cascade in *Drosophila*. *Devel. Cell 3*, 581–592.

DeLotto, R., **2001**. Gastrulation defective, a complement factor C2/B-like protease interprets a ventral prepattern in *Drosophila*. *EMBO Rep. 2*, 721–726.

DeLotto, Y., DeLotto, R., **1998**. Proteolytic processing of the *Drosophila* Spätzle protein by easter generates a dimeric NGF-like molecule with ventralising activity. *Mech. Devel. 72*, 141–148.

DeLotto, R., Spierer, P., **1986**. A gene required for the specification of dorsal–ventral pattern in *Drosophila* appears to encode a serine protease. *Nature 323*, 688–692.

DeMartino, G.N., Slaughter, C.A., **1999**. The proteasome, a novel protease regulated by multiple mechanisms. *J. Biol. Chem. 274*, 22123–22126.

Dimopoulos, G., Christophides, G.K., Meister, S., Schultz, J., White, K.P., *et al.*, **2002**. Genome expression analysis of *Anopheles gambiae*: responses to injury, bacterial challenge, and malaria infection. *Proc. Natl Acad. Sci. USA 99*, 8814–8819.

Dimopoulos, G., Richman, A., della Torre, A., Kafatos, F.C., Louisn, C., **1996**. Identification and characterization of differentially expressed cDNAs of the vector mosquito, *Anopheles gambiae*. *Proc. Natl Acad. Sci. USA 93*, 13066–13071.

Dissing, M., Giordano, H., DeLotto, R., **2001**. Autoproteolysis and feedback in a protease cascade directing *Drosophila* dorsal–ventral cell fate. *EMBO J. 20*, 2387–2393.

Dittmer, N.T., Raikhel, A.S., **1997**. Analysis of the mosquito lysosomal aspartic protease gene: an insect housekeeping gene with fat body-enhanced expression. *Insect Biochem. Mol. Biol. 27*, 323–335.

Dorstyn, L., Colussi, P.A., Quinn, L.M., Richardson, H., Kumar, S., **1999a**. DRONC, an ecdysone-inducible *Drosophila* caspase. *Proc. Natl Acad. Sci. USA 96*, 4307–4312.

Dorstyn, L., Read, S.H., Quinn, L.M., Richardson, H., Kumar, S., **1999b**. DECAY, a novel *Drosophila* caspase related to mammalian caspase-3 and caspase-7. *J. Biol. Chem. 274*, 30778–30783.

Doumanis, J., Quinn, L., Richardson, H., Kumar, S., **2001**. STRICA, a novel *Drosophila melanogaster* caspase with an unusual serine/threonine-rich prodomain, interacts with DIAP1 and DIAP2. *Cell. Death Differ. 8*, 387–394.

Earnshaw, W.C., Martins, L.M., Kaufmann, S.H., **1999**. Mammalian caspases: structure, activation, substrates, and functions during apoptosis. *Annu. Rev. Biochem. 68*, 383–424.

Elrod-Erickson, M., Mishra, S., Schneider, D., **2000**. Interactions between the cellular and humoral immune responses in *Drosophila*. *Curr. Biol. 10*, 781–784.

Fambrough, D., Pan, D., Rubin, G.M., Goodman, C.S., **1996**. The cell surface metalloprotease/disintegrin Kuzbanian is required for axonal extension in *Drosophila*. *Proc. Natl Acad. Sci. USA 93*, 13233–13238.

Felsted, R.L., Kramer, K.J., Law, J.H., Berger, E., Kafatos, F.C., **1973**. Cocoonase IV. Mechanism of activation of prococoonase from *Antheraea polyphemus*. *J. Biol. Chem. 248*, 3021–3028.

Finnerty, C.M., Karplus, P.A., Granados, R.R., **1999**. The insect immune protein scolexin is a novel serine proteinase homolog. *Protein Sci. 8*, 242–248.

Fisher, A.J., Cruz, W., Zoog, S.J., Schneider, C.L., Friesen, P.D., **1999**. Crystal structure of baculovirus P35: role of a novel reactive site loop in apoptotic caspase inhibition. *EMBO J. 18*, 2031–2039.

Forsyth, C.M., Lemongello, D., LaCount, D.J., Friesen, P.D., Fisher, A.J., **2004**. Crystal structure of an invertebrate caspase. *J. Biol. Chem. 279*, 7001–7008.

Fox, A.M., Reynolds, S.E., **1991**. Degradation of adipokinetic hormone family peptides by a circulating endopeptidase in the insect *Manduca sexta*. *Peptides 12*, 937–944.

Fraser, A.G., Evan, G.I., **1997**. Identification of a *Drosophila melanogaster* ICE/CED-3-related protease, drICE. *EMBO J. 16*, 2805–2813.

Fraser, A.G., McCarthy, N.J., Evan, G.I., **1997**. drICE is an essential caspase required for apoptotic activity in *Drosophila* cells. *EMBO J. 16*, 6192–6199.

Fujimoto, Y., Kobayashi, A., Kurata, S., Natori, S., **1999**. Two subunits of the insect 26/29-kDa proteinase are probably derived from a common precursor protein. *J. Biochem. (Tokyo) 125*, 566–573.

Gan, H., Wang, Y., Jiang, M., Mita, K., Kanost, M.R., **2001**. A bacteria-induced, intracellular serpin in granular hemocytes of *Manduca sexta*. *Insect Biochem. Molec. Biol. 31*, 887–898.

Gay, N.J., Keith, F.J., **1992**. Regulation of translation and proteolysis during the development of embryonic dorso–ventral polarity in *Drosophila*. Homology of easter proteinase with Limulus proclotting enzyme and translational activation of Toll receptor synthesis. *Biochim. Biophys. Acta 1132*, 290–296.

Georgel, P., Naitza, S., Kappler, C., Ferrandon, D., Zachary, D., *et al.*, **2001**. *Drosophila* immune deficiency (IMD) is a death domain protein that activates antibacterial defense and can promote apoptosis. *Devel. Cell. 1*, 503–514.

Gettins, P.G., **2002**. Serpin structure, mechanism, and function. *Chem. Rev. 102*, 4751–4804.

Gillespie, J.P., Kanost, M.R., Trenczek, T., **1997**. Biological mediators of insect immunity. *Ann. Rev. Entomol. 42*, 611–643.

Gorman, M.J., Andreeva, O.V., Paskewitz, S.M., **2000a**. Sp22D: a multidomain serine protease with a putative role in insect immunity. *Gene 251*, 9–17.

Gorman, M.J., Andreeva, O.V., Paskewitz, S.M., **2000b**. Molecular characterization of five serine protease genes cloned from *Anopheles gambiae* hemolymph. *Insect Biochem. Mol. Biol. 30*, 35–46.

Gorman, M., Paskewitz, S.M., **2001**. Serine proteases as mediators of mosquito immune responses. *Insect Biochem. Molec. Biol. 31*, 257–262.

Green, C., Levashina, E., McKimmie, C., Dafforn, T., Reichhart, J.-M., *et al.*, **2000**. The necrotic gene in *Drosophila* corresponds to one of a cluster of three serpin transcripts mapping at 43A1.2. *Genetics 156*, 1117–1127.

Haire, M.F., Clark, J.J., Jones, M.E., Hendil, K.B., Schwartz, L.M., *et al.*, **1995**. The multicatalytic protease (proteasome) of the hawkmoth, *Manduca sexta*: catalytic properties and immunological comparison with the lobster enzyme complex. *Arch. Biochem. Biophys. 318*, 15–24.

Han, J.-H., Lee, S.H., Tan, Y.-Q., Lemosy, E.K., Hashimoto, C., **2000**. Gastrulation defective is a serine protease involved in activating the receptor Toll to polarize the *Drosophila* embryo. *Proc. Natl Acad. Sci. 97*, 9093–9097.

Han, Y., Salazar, C., Reese-Stardy, S., Cornel, A., Gorman, M., *et al.*, **1997**. Cloning and characterization of a serine protease from the human malaria vector, *Anopheles gambiae*. *Insect Molec. Biol. 6*, 385–395.

Harvey, N.L., Daish, T., Mills, K., Dorstyn, L., Quinn, L.M., *et al.*, **2001**. Characterization of the *Drosophila* caspase, DAMM. *J. Biol. Chem. 276*, 25342–25350.

Hashimoto, C., Kim, D.R., Weiss, L.A., Miller, J.W., Morisato, D., **2003**. Spatial regulation of developmental signaling by a serpin. *Devel. Cell. 5*, 945–950.

Hastings, R.A., Eyheralde, I., Dawson, S.P., Walker, G., Reynolds, S.E., *et al.*, **1999**. A 220-kDa activator complex of the 26 S proteasome in insects and humans. A role in type II programmed insect muscle cell death and cross-activation of proteasomes from different species. *J. Biol. Chem. 274*, 25691–25700.

Hawkins, C.J., Wang, S.L., Hay, B.A., **1999**. A cloning method to identify caspases and their regulators in yeast: identification of *Drosophila* IAP1 as an inhibitor of the *Drosophila* caspase DCP-1. *Proc. Natl Acad. Sci. USA 96*, 2885–2890.

Hawkins, C.J., Yoo, S.J., Peterson, E.P., Wang, S.L., Vernooy, S.Y., *et al.*, **2000**. The *Drosophila* caspase DRONC cleaves following glutamate or aspartate and is regulated by DIAP1, HID, and GRIM. *J. Biol. Chem. 275*, 27084–27093.

Hay, B.A., Wassarman, D.A., Rubin, G.M., **1995**. *Drosophila* homologs of baculovirus inhibitor of apoptosis proteins function to block cell death. *Cell 83*, 1253–1262.

Hegedus, D., O'Grady, M., Chamankhah, M., Baldwin, D., Gleddie, S., *et al.*, **2002**. Changes in cysteine protease activity and localization during midgut metamorphosis in the crucifer root maggot (*Delia radicum*). *Insect Biochem. Molec. Biol. 32*, 1585–1596.

Hoffmann, J.A., **2003**. The immune response of *Drosophila*. *Nature 426*, 33–38.

Homma, K., Kurata, S., Natori, S., **1994**. Purification, characterization, and cDNA cloning of procathepsin L from the culture medium of NIH-SAPE-4, an embryonic cell line of *Sarcophaga peregrina* (flesh fly), and its involvement in the differentiation of imaginal disks. *J. Biol. Chem. 269*, 15258–15264.

Homma, H., Natori, N., **1996**. Identification of substrate proteins for cathepsin L that are selectively hydrolyzed during the differentiation of imaginal disks of *Sarcophaga peregrina*. *Eur. J. Biochem. 240*, 443–447.

Hong, C., Hashimoto, C., **1995**. An unusual mosaic protein with a protease domain, encoded by the nudel gene, is involved in defining embryonic dorsoventral polarity in *Drosophila*. *Cell 82*, 785–794.

Hong, C.C., Hashimoto, C., **1996**. The maternal nudel protein of *Drosophila* has two distinct roles important for embryogenesis. *Genetics 143*, 1653–1661.

Hooper, N.M., **2002**. Proteases: a primer. In: Hooper, N.M. (Ed.), Essays in Biochemistry: Proteases in Biology and Medicine. Portland Press, London, pp. 1–8.

Huang, Q., Deveraux, Q.L., Maeda, S., Salvesen, G.S., Stennicke, H.R., *et al.*, **2000**. Evolutionary conservation of apoptosis mechanisms: lepidopteran and baculoviral inhibitor of apoptosis proteins are inhibitors of mammalian caspase-9. *Proc. Natl Acad. Sci. USA 97*, 1427–1432.

Huang, Q., Deveraux, Q.L., Maeda, S., Stennicke, H.R., Hammock, B.D., *et al.*, **2001**. Cloning and characterization of an inhibitor of apoptosis protein (IAP) from *Bombyx mori*. *Biochim. Biophys. Acta 1499*, 191–198.

Igaki, T., Yamamoto-Goto, Y., Tokushige, N., Kanda, H., Miura, M., **2002**. Down-regulation of DIAP1 triggers a novel *Drosophila* cell death pathway mediated by Dark and DRONC. *J. Biol. Chem. 277*, 23103–23106.

Ikeda, M., Yaginuma, T., Kobayashi, M., Yamashita, O., **1991**. cDNA cloning, sequencing and temporal expression of the protease responsible for vitellin degradation in the silkworm *Bombyx mori*. *Comp. Biochem. Physiol. B 99*, 405–411.

Indrasith, L.S., Sasaki, T., Yamashita, O., **1988**. A unique protease responsible for selective degradation of a yolk protein in *Bombyx mori*. Purification, characterization, and cleavage profile. *J. Biol. Chem. 263*, 1045–1051.

Irving, P., Troxler, L., Heuer, T.S., Belvin, M., Kopczynski, C., *et al.*, **2001**. A genome-wide analysis of immune responses in *Drosophila*. *Proc. Natl Acad. Sci. USA 98*, 15119–15124.

Isaac, R.E., Nassel, D.R., **2003**. Identification and localization of a neprilysin-like activity that degrades tachykinin-related peptides in the brain of the cockroach, *Leucophaea maderae*, and locust, *Locusta migratoria*. *J. Comp. Neurol. 457*, 57–66.

Isaac, R.E., Parkin, E.T., Keen, J.N., Nassel, D.R., Siviter, R.J., *et al.*, **2002**. Inactivation of a tachykinin-related peptide: identification of four neuropeptide-degrading enzymes in neuronal membranes of insects from four different orders. *Peptides 23*, 725–733.

Jiang, H., Kanost, M.R., **1997**. Characterization and functional analysis of twelve naturally occurring reactive site variants of serpin-1 from *Manduca sexta*. *J. Biol. Chem. 272*, 1082–1087.

Jiang, H., Kanost, M.R., **2000**. The clip-domain family of serine proteinases in arthropods. *Insect Biochem. Molec. Biol. 30*, 95–105.

Jiang, H., Wang, Y., Huang, Y., Mulnix, A.B., Kadel, J., *et al.*, **1996**. Organization of serpin gene-1 from *Manduca sexta*: evolution of a family of alternate exons encoding the reactive site loop. *J. Biol. Chem. 271*, 28017–28023.

Jiang, H., Wang, Y., Kanost, M.R., **1998**. Pro-phenol oxidase activating proteinase from an insect, *Manduca sexta*: a bacteria-inducible protein similar to *Drosophila* easter. *Proc. Natl Acad. Sci. USA 95*, 12220–12225.

Jiang, H., Wang, Y., Kanost, M.R., **1999**. Four serine proteinases expressed in *Manduca sexta* hemocytes. *Insect Mol. Biol. 8*, 39–53.

Jiang, H., Wang, Y., Yu, X.-Q., Kanost, M.R., **2003a**. Prophenoloxidase-activating proteinase-2 from hemolymph of *Manduca sexta*. A bacteria-inducible serine proteinase containing two clip domains. *J. Biol. Chem. 278*, 3552–3561.

Jiang, H., Wang, Y., Yu, X.-Q., Zhu, Y., Kanost, M.R., **2003b**. Prophenoloxidase-activating proteinase-3 (PAP-3) from *Manduca sexta* hemolymph: a clip-domain serine proteinase regulated by serpin-1J and serine proteinase homologs. *Insect Biochem. Mol. Biol. 33*, 1049–1060.

Jin, Y., Anderson, K.V., **1990**. Dominant and recessive alleles of the *Drosophila* easter gene are point mutations at conserved sites in the serine protease catalytic domain. *Cell 60*, 873–881.

Jones, M.E., Haire, M.F., Kloetzel, P.M., Mykles, D.L., Schwartz, L.M., **1995**. Changes in the structure and function of the multicatalytic proteinase (proteasome) during programmed cell death in the intersegmental muscles of the hawkmoth, *Manduca sexta*. *Devel. Biol. 169*, 436–447.

Jones, G., Jones, D., Zhou, L., Steller, H., Chu, Y., **2000**. Deterin, a new inhibitor of apoptosis from *Drosophila melanogaster*. *J. Biol. Chem. 275*, 22157–22165.

Kafatos, F.C., **1972**. The cocoonase zymogen cells of silk moths: a model of terminal cell differentiation for specific protein synthesis. *Curr. Top. Devel. Biol. 7*, 125–191.

Kafatos, F.C., Law, J.H., Tartakoff, A.M., **1967b**. Cocoonase II. Substrate specificity, inhibitors, and classification of the enzyme. *J. Biol. Chem. 424*, 1488–1494.

Kafatos, F.C., Tartakoff, A.M., Law, J.H., **1967a**. Cocoonase I. Preliminary characterization of a proteolytic enzyme from silk moths. *J. Biol. Chem. 242*, 1477–1487.

Kageyama, T., Takahashi, S.Y., **1990**. Purification and characterization of a cysteine proteinase from silkworm eggs. *Eur. J. Biochem. 193*, 203–210.

Kaiser, W.J., Vucic, D., Miller, L.K., **1998**. The *Drosophila* inhibitor of apoptosis D-IAP1 suppresses cell death induced by the caspase drICE. *FEBS Lett. 440*, 243–248.

Kanost, M.R., **1999**. Serine proteinase inhibitors in arthropod immunity. *Devel. Comp. Immunol. 23*, 291–301.

Kanost, M.R., Jiang, H., **1996**. Proteinase inhibitors in invertebrate immunity. In: Söderhäll, K., Iwanga, S., Vasta, G. (Eds.), New Directions in Invertebrate Immunology. SOS Publications, Fair Haven, CT, pp. 155–173.

Kanost, M.R., Jiang, H., Wang, Y., Yu, X.-Q., Ma, C.C., *et al.*, **2001**. Hemolymph proteinases in immune responses of *Manduca sexta*. *Adv. Exp. Med. Biol. 484*, 319–328.

Katsumi, Y., Kihara, H., Ochiai, M., Ashida, M., **1995**. A serine protease zymogen in insect plasma. Purification and activation by microbial cell components. *Eur. J. Biochem. 228*, 870–877.

Kawabata, S., Muta, T., Iwanaga, S., **1996**. The clotting cascade and defense molecules found in the hemolymph of the horseshoe crab. In: Söderhäll, K., Iwanaga, S., Vasta, G. (Eds.), New Directions in Invertebrate Immunology. SOS Publications, Fair Haven, CT, pp. 255–283.

Kim, M.S., Baek, M.J., Lee, M.H., Park, J.W., Lee, S.Y., *et al.*, **2002**. A new easter-type serine protease cleaves a masquerade-like protein during prophenoloxidase activation in *Holotrichia diomphalia* larvae. *J. Biol. Chem. 277*, 39999–40004.

Kim, Y.S., Ryu, J.H., Han, S.J., Choi, K.H., Nam, K.B., *et al.*, **2000**. Gram-negative bacteria-binding protein, a pattern recognition receptor for lipopolysaccharide and beta-1,3-glucan that mediates the signaling for the

induction of innate immune genes in *Drosophila melanogaster* cells. *J. Biol. Chem. 275*, 32721–32727.

Konrad, K.D., Goralski, T.J., Mahowald, A.P., Marsh, J.L., **1998**. The gastrulation defective gene of *Drosophila melanogaster* is a member of the serine protease superfamily. *Proc. Natl Acad. Sci. USA 95*, 6819–6824.

Kramer, K.J., Felsted, R.L., Law, J.H., **1973**. Cocoonase V. Structural studies on an insect serine protease. *J. Biol. Chem. 248*, 3021–3028.

Krem, M.M., Di Cerra, E., **2002**. Evolution of enzyme cascades from embryonic development to blood coagulation. *Trends Biochem. Sci. 27*, 67–74.

Kruger, O., Ladewig, J., Koster, K., Ragg, H., **2002**. Widespread occurrence of serpin genes with multiple reactive centre-containing exon cassettes in insects and nematodes. *Gene 293*, 97–105.

Kryiakides, T.R., McKillip, J.L., Spence, K.D., **1995**. *Arch. Insect Biochem. Physiol. 29*, 269–280.

Kumar, S., Doumanis, J., **2000**. The fly caspases. *Cell Death Differ. 7*, 1039–1044.

Kurata, S., Saito, H., Natori, S., **1992a**. Purification of a 29 kDa proteinase of *Sarcophaga peregrina*. *Eur. J. Biochem. 204*, 911–914.

Kurata, S., Saito, H., Natori, S., **1992b**. The 29-kDa hemocyte proteinase dissociates fat body at metamorphosis of *Sarcophaga*. *Devel. Biol. 153*, 115–121.

Kwon, T.H., Kim, M.S., Choi, H.W., Loo, C.H., Cho, M.Y., *et al.*, **2000**. A masquerade-like serine proteinase homologue is necessary for phenoloxidase activity in the coleopteran insect, *Holotrichia diomphalia* larvae. *Eur. J. Biochem. 267*, 6188–6196.

LaCount, D.J., Hanson, S.F., Schneider, C.L., Friesen, P.D., **2000**. Caspase inhibitor P35 and inhibitor of apoptosis Op-IAP block *in vivo* proteolytic activation of an effector caspase at different steps. *J. Biol. Chem. 275*, 15657–15664.

Lamango, N.S., Isaac, R.E., **1993**. Metabolism of insect neuropeptides: properties of a membrane-bound endopeptidase form heads of *Musca domestica*. *Insect Biochem. Mol. Biol. 23*, 801–808.

Law, J.H., Dunn, P.E., Kramer, K.J., **1977**. Insect proteases and peptidases. *Adv. Enzymol. Relat. Areas Mol. Biol. 45*, 389–425.

Lee, C.Y., Cooksey, B.A., Baehrecke, E.H., **2002a**. Steroid regulation of midgut cell death during *Drosophila* development. *Devel. Biol. 250*, 101–111.

Lee, K.Y., Zhang, R., Kim, M.S., Park, J.W., Park, H.Y., *et al.*, **2002b**. A zymogen form of masquerade-like serine proteinase homologue is cleaved during prophenoloxidase activation by Ca^{2+} in coleopteran and *Tenebrio molitor* larvae. *Eur. J. Biochem. 269*, 4375–4383.

Lee, S.Y., Cho, M.Y., Hyun, J.H., Lee, K.M., Homma, K., *et al.*, **1998b**. Molecular cloning of cDNA for pro-phenol-oxidase-activating factor I, a serine proteinase is induced by lipopolysaccharide or 1,3-β-glucan in a coleopteran insect, *Holotrichia diomphalia* larvae. *Eur. J. Biochem. 257*, 615–621.

Lee, S.Y., Kwon, T.H., Hyun, J.H., Choi, J.S., Kawabata, S.I., *et al.*, **1998a**. *In vitro* activation of pro-phenol-oxidase by two kinds of pro-phenol-oxidase activating factors isolated from hemolymph of coleopteran *Holotrichia diomphalia* larvae. *Eur. J. Biochem. 15*, 50–57.

Lee, S.Y., Söderhäll, K., **2001**. Characterization of a pattern recognition protein, a masquerade-like protein, in the freshwater crayfish *Pacifastacus leniusculus*. *J. Immunol. 166*, 7319–7326.

Lemaitre, B., Nocolas, E., Michaut, L., Reichhart, J.M., Hoffmann, J.A., **1996**. The dorsoventral regulatory gene cassette spätzle/Toll/cactus controls the potent antifungal response in *Drosophila* adults. *Cell 86*, 973–983.

LeMosy, E.K., Hashimoto, C., **2000**. The nudel protease of *Drosophila* is required for eggshell biogenesis in addition to embryonic patterning. *Devel. Biol. 217*, 352–361.

LeMosy, E.K., Hong, C.C., Hashimoto, C., **1999**. Signal transduction by a protease cascade. *Trends Cell Biol. 9*, 102–107.

LeMosy, E.K., Kemler, D., Hashimoto, C., **1998**. Role of Nudel protease activation in triggering dorsoventral polarization of the *Drosophila* embryo. *Development 125*, 4045–4053.

LeMosy, E.K., Leclerc, C.L., Hashimoto, C., **2000**. Biochemical defects of mutant *nudel* alleles causing early developmental arrest or dorsalization of the *Drosophila* embryo. *Genetics 154*, 247–257.

LeMosy, E.K., Tan, Y.-Q., Hashimoto, C., **2001**. Activation of a protease cascade involved in patterning the *Drosophila* embryo. *Proc. Natl Acad. Sci. 98*, 5055–5060.

Leulier, F., Parquet, C., Pili-Floury, S., Ryu, J.H., Caroff, M., *et al.*, **2003**. The *Drosophila* immune system detects bacteria through specific peptidoglycan recognition. *Nat. Immunol. 4*, 478–484.

Leulier, F., Rodriguez, A., Khush, R.S., Abrams, J.M., Lemaitre, B., **2000**. The *Drosophila* caspase Dredd is required to resist gram-negative bacterial infection. *EMBO Rep 1*, 353–358.

Levashina, E.A., Langley, E., Green, C., Gubb, D., Ashburner, M., *et al.*, **1999**. Constitutive activation of Toll-mediated antifungal defense in serpin-deficient *Drosophila*. *Science. 285*, 1917–1919.

Li, J., Wang, Z., Canagarajah, B., Jiang, H., Kanost, M.R., *et al.*, **1999**. The structure of active serpin K from *Manduca sexta* and a model for serpin-protease complex formation. *Structure 7*, 103–109.

Liao, W.T., Yang, Y., Wu, X.F., **2002**. Expression and functional analysis of an inhibitor of apoptosis protein from *Trichoplusia ni*. *Biochem. Biophys. Res. Commun. 293*, 675–679.

Lieber, T., Kidd, S., Young, M.W., **2002**. Kuzbanian-mediated cleavage of *Drosophila* Notch. *Genes Devel. 16*, 209–221.

Ligoxygakis, P., Pelte, N., Hoffmann, J.A., Reichhart, J.M., **2002a**. Activation of *Drosophila* Toll during fungal infection by a blood serine protease. *Science 297*, 114–116.

Ligoxygakis, P., Pelte, N., Ji, C., Leclerc, V., Duvic, B., *et al.*, **2002b**. A serpin mutant links Toll activation to melanization in the host defence of *Drosophila*. *EMBO J. 21*, 6330–6337.

Ligoxygakis, P., Roth, S., Rechhart, J.M., **2003**. A serpin regulates dorsal–ventral axis formation in the *Drosophila* embryo. *Curr. Biol. 13*, 2097–2102.

Llano, E., Adam, G., Pendas, A.M., Quesada, V., Sanchez, L.M., *et al.*, **2002**. Structural and enzymatic characterization of *Drosophila* Dm2-MMP, a membrane-bound matrix metalloproteinase with tissue-specific expression. *J. Biol. Chem. 277*, 23321–23329.

Llano, E., Pendas, A.M., Aza-Blanc, P., Kornberg, T.B., Lopez-Otin, C., **2000**. Dm1-MMP, a matrix metalloproteinase from *Drosophila* with a potential role in extracellular matriz remodeling during neural development. *J. Biol. Chem. 275*, 35978–35985.

Low, P., Bussell, K., Dawson, S.P., Billett, M.A., Mayer, R.J., *et al.*, **1997**. Expression of a 26S proteasome ATPase subunit, M73, in muscles that undergo developmentally programmed cell death, and its control by ecdysteroid hormones in the insect *Manduca sexta*. *FEBS Lett. 400*, 345–349.

Low, P., Hastings, R.A., Dawson, S.P., Sass, M., Billett, M.A., *et al.*, **2000**. Localisation of 26S proteasomes with different subunit composition in insect muscles undergoing programmed cell death. *Cell Death Differ. 7*, 1210–1217.

Low, P., Reynolds, S.E., Sass, M., **2001**. Proteolytic activity of 26s proteasomes isolated from muscles of the tobacco hornworm, *Manduca sexta*: differences between surviving muscles and those undergoing developmentally programmed cell death. *Acta Biol. Hung. 52*, 435–442.

Lundgren, J., Masson, P., Realini, C.A., Young, P., **2003**. Use of RNA interference and complementation to study the function of the *Drosophila* and human 26S proteasome subunit S13. *Mol. Cell. Biol. 23*, 5320–5330.

Ma, C., Kanost, M.R., **2000**. A β-1,3-glucan recognition protein from an insect, *Manduca sexta*, agglutinates microorganisms and activates the phenoloxidase cascade. *J. Biol. Chem. 275*, 7505–7514.

Ma, J., Katz, E., Belote, J.M., **2002**. Expression of proteasome subunit isoforms during spermatogenesis in *Drosophila melanogaster. Insect Mol. Biol. 11*, 627–639.

Macours, N., Poels, J., Hens, K., Luciani, N., De Loof, A., *et al.*, **2003**. An endothelin-converting enzyme homologue in the locust, *Locusta migratoria*: functional activity, molecular cloning and tissue distribution. *Insect Mol. Biol. 12*, 233–240.

Maki, N., Yamashita, O., **1997**. Purification and characterization of a protease degrading 30 kDa yolk proteins of the silkworm, *Bombyx mori*. *Insect Biochem. Mol. Biol. 27*, 721–728.

Maki, N., Yamashita, O., **2001**. The 30 k protease A responsible for 30-kDa yolk protein degradation of the silkworm, *Bombyx mori*: cDNA structure, developmental change and regulation by feeding. *Insect Biochem. Mol. Biol. 31*, 407–413.

Manji, G.A., Friesen, P.D., **2001**. Apoptosis in motion. An apical, P35-insensitive caspase mediates programmed cell death in insect cells. *J. Biol. Chem. 276*, 16704–16710.

Masler, E.P., Wagner, R.M., Kovaleva, E.S., **1996**. *In vitro* metabolism of an insect neuropeptide by neural membrane preparations from *Lymantria dispar. Peptides 17*, 321–326.

Means, J.C., Muro, I., Clem, R.J., **2003**. Silencing of the baculovirus Op-iap3 gene by RNA interference reveals that it is required for prevention of apoptosis during *Orgyia pseudotsugata* M nucleopolyhedrovirus infection of Ld652Y cells. *J. Virol. 77*, 4481–4488.

Medina, M., Leon, P., Vallejo, C.G., **1988**. *Drosophila* cathepsin B-like proteinase: a suggested role in yolk degradation. *Arch. Biochem. Biophys. 263*, 355–363.

Meier, P., Silke, J., Leevers, S.J., Evan, G.I., **2000**. The *Drosophila* caspase DRONC is regulated by DIAP1. *EMBO J. 19*, 598–611.

Misra, S., Hecht, P., Maeda, R., Anderson, K.V., **1998**. Positive and negative regulation of Easter, a member of the serine protease family that controls dorsal-ventral patterning in the *Drosophila* embryo. *Development 125*, 1261–1267.

Morisato, D., Anderson, K.V., **1995**. Signaling pathways that establish the dorsal–ventral pattern of the *Drosophila* embryo. *Annu. Rev. Genet. 29*, 371–399.

Morugasu-Oei, B., Balakrishnan, R., Yang, X., Chia, W., Rodrigues, V., **1996**. Mutations in masquerade, a novel serine-protease-like molecule, affect axonal guidance and taste behavior in *Drosophila*. *Mech. Devel. 57*, 91–101.

Morugasu-Oei, B., Rodrigues, V., Yang, X., Chia, W., **1995**. Masquerade: a novel secreted serine protease-like molecule is required for somatic muscle attachment in the *Drosophila* embryo. *Genes Devel. 9*, 139–154.

Muro, I., Hay, B.A., Clem, R.J., **2002**. The *Drosophila* DIAP1 protein is required to prevent accumulation of a continuously generated, processed form of the apical caspase DRONC. *J. Biol. Chem. 277*, 49644–49650.

Mykles, D.L., **1997**. Biochemical properties of insect and crustacean proteasomes. *Mol. Biol. Rep. 24*, 133–138.

Mykles, D.L., **1999**. Structure and functions of arthropod proteasomes. *Mol. Biol. Rep. 26*, 103–111.

Nagase, H., Woessner, F.J., **1999**. Matrix metalloproteinases. *J. Biol. Chem., 274*, 21491–21494.

Nakajima, Y., Tsuji, Y., Homma, K., Natori, S., **1997**. A novel protease in the pupal yellow body of *Sarcophaga peregrina* (flesh/fly). Its purification and cDNA cloning. *J. Biol. Chem. 272*, 23805–23810.

Narumi, H., Hishida, T., Sasaki, T., Feng, D.F., Doolittle, R.F., **1993**. Molecular cloning of silkworm (*Bombyx mori*) antichymotrypsin. A new member of the serpin superfamily of proteins from insects. *Eur. J. Biochem. 214*, 181–187.

Ochiai, M., Ashida, M., **1988**. Purification of a β-1,3-glucan recognition protein in the prophenoloxidase activating system from hemolymph of the silkworm, *Bombyx mori*. *J. Biol. Chem. 263*, 12056–12062.

Ochiai, M., Ashida, M., **1999**. A pattern recognition protein for peptidoglycan. Cloning the cDNA and the gene of the silkworm, *Bombyx mori*. *J. Biol. Chem. 274*, 11854–11858.

Ochiai, M., Ashida, M., **2000**. A pattern recognition protein for β-1,3-glucan. *J. Biol. Chem. 275*, 4995–5002.

Oduol, F., Xu, J., Niare, O., Natarajan, R., Vernick, K.D., **2000**. Genes identified by an expression screen of the vector mosquito *Anopheles gambiae* display differential molecular immune response to malaria parasites and bacteria. *Proc. Natl Acad. Sci. USA 97*, 11394–11402.

Page-McCaw, A., Serano, J., Sante, J.M., Rubin, G.M., **2003**. *Drosophila* matrix metalloproteinases are required for tissue remodeling, but not embryonic development. *Devel. Cell 4*, 95–106.

Paskewitz, S.M., Gorman, M.J., **1999**. Mosquito immunity and malaria parasites. *M. Entomol. 45*, 80–94.

Paskewitz, S.M., Reese-Stardy, S., Gorman, M.J., **1999**. An easter-like serine protease from *Anopheles gambiae* exhibits changes in transcript abundance following immune challenge. *Insect Mol. Biol. 8*, 329–337.

Pino-Heiss, S., Schubiger, G., **1989**. Extracellular protease production by *Drosophila* imaginal disks. *Devel. Biol. 132*, 282–291.

Polanowski, A., Wilusz, T., **1996**. Serine proteinase inhibitors from insect hemolymph. *Acta Biochimica. Polonica. 43*, 445–454.

Pomés, A., Chapman, M.D., Vailes, L.D., Blundell, T.L., Dhanaraj, V., **2002**. Cockroach allergen Bla g2: structure, function, and implications for allergic sensitization. *Am. J. Respir. Crit. Care Med. 165*, 391–397.

Quinn, L.M., Dorstyn, L., Mills, K., Colussi, P.A., Chen, P., *et al.*, **2000**. An essential role for the caspase dronc in developmentally programmed cell death in *Drosophila*. *J. Biol. Chem. 275*, 40416–40424.

Raikhel, A.S., Dhadialla, T.S., **1992**. Accumulation of yolk proteins in insect oocytes. *Annu. Rev. Entomol. 37*, 217–251.

Rawlings, N.D., Barrett, A.J., **1999**. MEROPS: the peptidase database. *Nucleic Acids Res. 27*, 325–331.

Reeck, G., Oppert, B., Denton, M., Kanost, M., Baker, J., *et al.*, **1999**. Insect proteinases. In: Turk, V. (Ed.), Proteases: New Perspectives. Birkhauser, Basel, pp. 125–148.

Ribolla, P.E., De Bianchi, A.G., **1995**. Processing of procathepsin from *Musca domestica* eggs. *Insect Biochem. Mol. Biol. 25*, 1011–1017.

Richardson, H., Kumar, S., **2002**. Death to flies: *Drosophila* as a model system to study programmed cell death. *J. Immunol. Methods 265*, 21–38.

Riedl, S.J., Renatus, M., Snipas, S.J., Salvesen, G.S., **2001**. Mechanism-based inactivation of caspases by the apoptotic suppressor p35. *Biochemistry 40*, 13274–13280.

Rodriguez, A., Chen, P., Oliver, H., Abrams, J.M., **2002**. Unrestrained caspase-dependent cell death caused by loss of Diap1 function requires the *Drosophila* Apaf-1 homolog, Dark. *EMBO J. 21*, 2189–2197.

Rooke, J., Pan, D., Xu, T., Rubin, G.M., **1996**. KUZ, a conserved metalloprotease-disintegrin protein with two roles in *Drosophila* neurogenesis. *Science 273*, 1227–1231.

Rose, T., LeMosy, E.K., Cantwell, A.M., Banerjee-Roy, D., Skeath, J.B., *et al.*, **2003**. Three-dimensional models of proteases involved in patterning of the *Drosophila* embryo. *Crucial role of predicted cation binding sites. J. Biol. Chem. 278*, 11320–11330.

Ross, J., Jiang, H., Kanost, M.R., Wang, Y., **2003**. Serine proteases and their homologs in the *Drosophila melanogaster* genome: an initial analysis of sequence conservation and phylogenetic relationships. *Gene 304*, 117–131.

Saito, H., Kurata, S., Natori, S., **1992**. Purification and characterization of a hemocyte proteinase of *Sarcophaga*, possibly participating in elimination of foreign substances. *Eur. J. Biochem. 209*, 939–944.

Sasaki, T., **1991**. Patchwork-structure serpins from silkworm (*Bombyx mori*) larval hemolymph. *Eur. J. Biochem. 202*, 255–161.

Satoh, D., Horii, A., Ochiai, M., Ashida, M., **1999**. Prophenoloxidase-activating enzyme of the silkworm, *Bombyx mori*: purification, characterization and cDNA cloning. *J. Biol. Chem. 274*, 7441–7453.

Saul, S., Sugumaran, M., **1986**. Protease inhibitor controls prophenoloxidase activation in *Manduca sexta*. *FEBS Lett. 208*, 113–116.

Schechter, I., Berger, A., **1967**. On the size of the active site in proteases. *I. Papain. Biochem. Biophys. Res. Commun. 27*, 157–162.

Schweisguth, F., **1999**. Dominant-negative mutation in the β2 and β6 proteasome subunit genes affect alternate cell fate decisions in the Drosophila sense organ lineage. *Proc. Natl Acad. Sci. USA 96*, 11382–11386.

Seals, D.F., Courtneidge, S.A., **2003**. The ADAMs family of metalloproteases: multidomain proteins with multiple functions. *Genes Devel. 17*, 7–30.

Seshagiri, S., Miller, L.K., **1997**. Baculovirus inhibitors of apoptosis (IAPs) block activation of Sf-caspase-1. *Proc. Natl Acad. Sci. USA 94*, 13606–13611.

Seshagiri, S., Vucic, D., Lee, J., Dixit, V.M., **1999**. Baculovirus-based genetic screen for antiapoptotic genes identifies a novel IAP. *J. Biol. Chem. 274*, 36769–36773.

Shiba, H., Uchida, D., Kobayashi, H., Natori, M., **2001**. Involvement of cathepsin B- and L-like proteinases in silk gland histolysis during metamorphosis of *Bombyx mori*. *Arch. Biochem. Biophys. 390*, 28–34.

Silverman, G.A., Bird, P.I., Carrell, R.W., Church, F.C., Coughlin, P.B., *et al.*, **2001**. The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature. *J. Biol. Chem. 276*, 33293–33296.

Smith, C., DeLotto, R., **1992**. A common domain within the proenzyme regions of the *Drosophila* snake and easter proteins and *Tachypleus* proclotting enzyme

defines a new subfamily of serine proteases. *Protein Sci. 11*, 1225–1226.

Smith, C., DeLotto, R., **1994**. Ventralizing signal determined by protease activation in *Drosophila* embryogenesis. *Nature 368*, 548–551.

Smith, C., Giordano, H., DeLotto, R., **1994**. Mutational analysis of the *Drosophila* snake protease: an essential rule for domains within the proenzyme polypeptide chain. *Genetics 136*, 1355–1365.

Smith, C.L., Giordano, H., Schwartz, M., DeLotto, R., **1995**. Spatial regulation of Drosophila snake protease activity in the generation of dorsal–ventral polarity. *Development 121*, 4127–4135.

Smyth, K.A., Belote, J.M., **1999**. The dominant temperature-sensitive lethal DTS7 of *Drosophila melanogaster* encodes an altered 20S proteasome β-type subunit. *Genetics 151*, 211–220.

Söderhäll, K., Derenius, L., Johansson, M.W., **1996**. The prophenoloxidase activating system in invertebrates. In: Söderhäll, K., Iwanaga, S., Vasta, G. (Eds.), New Directions in Invertebrate Immunology. SOS Publications, Fair Haven, CT, pp. 229–253.

Söderhäll, K., Rogener, W., Söderhäll, I., Newton, R.P., Ratcliffe, N.A., **1988**. The properties and purification of a *Blaberus craniifer* plasma protein which enhances the activation of haemocyte prophenoloxidase by a β-1,3-glucan. *Insect Biochem. 18*, 323–330.

Song, Z., Guan, B., Bergman, A., Nicholson, D.W., Thornberry, N.A., *et al.*, **2000**. Biochemical and genetic interactions between *Drosophila* caspases and the proapoptotic genes rpr, hid, and grim. *Mol. Cell. Biol. 20*, 2907–2914.

Song, Z., McCall, K., Steller, H., **1997**. DCP-1, a *Drosophila* cell death protease essential for development. *Science 275*, 536–540.

Sotillos, S., Roch, F., Campuzano, S., **1997**. The metalloprotease-disintintegrin Kuzbanian participates in Notch activation during growth and patterning of Drosophila imaginal disks. *Development 124*, 4769–4779.

Stoven, S., Ando, I., Kadalayil, L., Engström, Y., Hultmark, D., **2000**. Activation of the *Drosophila* NF-kappaB factor Relish by rapid endoproteolytic cleavage. *EMBO Rep. 1*, 347–352.

Stoven, S., Silverman, N., Junell, A., Hedengren-Olcott, M., Erturk, D., *et al.*, **2003**. Caspase-mediated processing of the *Drosophila* NF-kappaB factor Relish. *Proc. Natl Acad. Sci. USA 100*, 5991–5996.

Sugumaran, M., Saul, S., Ramesh, N., **1985**. Endogenous protease inhibitors prevent undesired activation of prophenolase in insect hemolymph. *Biochem. Biophys. Res. Commun. 132*, 1124–1129.

Szlanka, T., Haracska, L., Kiss, I., Deak, P., Kurucz, E., *et al.*, **2003**. Deletion of proteasomal subunit S5a/Rpn10/p54 causes lethality, multiple mitotic defects and overexpression of proteasomal genes in *Drosophila melanogaster. J. Cell Sci. 116*, 1023–1033.

Takahashi, N., Kurata, S., Natori, S., **1993**. Molecular cloning of cDNA for the 29 kDa proteinase participating on decomposition of the larval fat body during metamorphosis of *Sarcophaga peregrina* (flesh fly). *FEBS Lett. 334*, 153–157.

Takahashi, S.Y., Yamamoto, Y., Shionoya, Y., Kageyama, T., **1993**. Cysteine proteinase from the eggs of the silkmoth, *Bombyx mori*: identification of a latent enzyme and characterization of activation and proteolytic processing *in vivo* and *in vitro*. *J. Biochem.* (*Tokyo*) *114*, 267–272.

Takayanagi, K., Dawson, S., Reynolds, S.E., Mayer, R.J., **1996**. Specific developmental changes in the regulatory subunits of the 26 S proteasome in intersegmental muscles preceding eclosion in *Manduca sexta*. *Biochem. Biophys. Res. Commun. 228*, 517–523.

Terra, W.R., Ferreira, C., Jordão, B.P., Dillon, R.J., **1996**. Digestive enzymes. In: Lehane, M.J., Billingsley, P.F. (Eds.), Biology of the Insect Midgut. Chapman and Hall, London, pp. 153–194.

Tryselius, Y., Hultmark, D., **1997**. Cysteine proteinase 1 (CP1), a cathespin L-like enzyme expressed in the *Drosophila melanogaster* haemocyte cell line mbn-2. *Insect Mol. Biol. 6*, 173–181.

Tsuji, Y., Nakajima, Y., Homma, K.-I., Natori, S., **1998**. Antibacterial activity of a novel 26-kDa serine protease in the yellow body of *Sarcophaga peregrina* (flesh fly) pupae. *FEBS Lett. 425*, 131–133.

Turcotte, C.L., Hashimoto, C., **2002**. Evidence for a glycosaminoglycan on the nudel protein important for dorsoventral patterning of the Drosophila embryo. *Devel. Dyn. 224*, 51–57.

Turner, A.J., Isaac, R.E., Coates, D., **2001**. The neprilysin (NEP) family of zinc metalloendopeptidases, genomics and function. *BioEssays 23*, 261–269.

Uvardy, A., **1993**. Purification and characterization of a multiprotein component of the *Drosophila* 26S (1500 kDa) proteolytic complex. *J. Biol. Chem. 268*, 9055–9062.

Vernooy, S.Y., Copeland, J., Ghaboosi, N., Griffin, E.E., Yoo, S.J., *et al.*, **2000**. Cell death regulation in *Drosophila*: conservation of mechanism and unique insights. *J. Cell Biol. 150*, F69–F75.

Vier, J., Furmann, C., Hacker, G., **2000**. Baculovirus P35 protein does not inhibit caspase-9 in a cell-free system of apoptosis. *Biochem. Biophys. Res. Commun. 276*, 855–861.

Walz, J., Erdmann, A., Kania, M., Typke, D., Koster, A.J., *et al.*, **1998**. 26S proteasome structure revealed by three-dimensional electron microscopy. *J. Struct. Biol. 121*, 19–29.

Wang, R., Lee, S.Y., Cerenius, L., Söderhall, K., **2001**. Properties of the prophenoloxidase activating enzyme of the freshwater crayfish, *Pacifastacus leniusculus*. *Eur. J. Biochem. 268*, 895–902.

Watts, R.J., Hoopfer, E.D., Luo, L., **2003**. Axon pruning during Drosophila metamorphosis: evidence for local degeneration and requirement of the ubiquitin-proteasome system. *Neuron. 38*, 871–887.

Wei, S., Xie, Z., Filenova, E., Brew, K., **2003**. *Drosophila* TIMP is a potent inhibitor of MMPs and TACE:

similarities in structure and function to TIMP-3. *Biochemistry 42*, 12200–12207.

Wilson, R., Goyal, L., Ditzel, M., Zachariou, A., Baker, D.A., *et al.*, **2002**. The DIAP1 RING finger mediates ubiquitination of Dronc and is indispensable for regulating apoptosis. *Nature Cell Biol. 4*, 445–450.

Wojcik, C., DeMartino, G.N., **2002**. Analysis of *Drosophila* 26S proteasome using RNA interference. *J. Biol. Chem. 277*, 6188–6197.

Xu, G., Cirilli, M., Huang, Y., Rich, R.L., Myszka, D.G., *et al.*, **2001**. Covalent inhibition revealed by the crystal structure of the caspase-8/p35 complex. *Nature 410*, 494–497.

Xu, Y., Kawasaki, H., **2001**. Isolation and expression of cathepsin B cDNA in hemocytes during metamorphosis of *Bombyx mori*. *Comp. Biochem. Physiol. B 130*, 393–399.

Yamamoto, Y., Takahashi, S.Y., **1993**. Cysteine proteinase from *Bombyx* eggs: role in programmed degradation of yolk proteins during embryogenesis. *Comp. Biochem. Physiol. B 106*, 35–45.

Yamamoto, Y., Takimoto, K., Izumi, S., Toriyama-Sakurai, M., Kageyama, T., *et al.*, **1994**. Molecular cloning and sequencing of cDNA that encodes cysteine proteinase in the eggs of the silkmoth, *Bombyx mori*. *J. Biochem.* (*Tokyo*) *116*, 1330–1335.

Yamamoto, Y., Yamahama, Y., Katou, K., Watabe, S., Takahashi, S.Y., **2000**. *Bombyx* acid cysteine protease (BCP): hormonal regulation of biosynthesis and accumulation in the ovary. *J. Insect Physiol. 46*, 783–791.

Ye, S., Cech, A.L., Belmares, R., Bergstrom, R.C., Tong, Y., *et al.*, **2001**. The structure of a Michaelis serpin-protease complex. *Nature Struct. Biol. 8*, 979–983.

Yoshida, H., Ashida, M., **1986**. Microbial activation of two serine enzymes and prophenoloxidase in the plasma fraction of hemolymph of the silkworm, *Bombyx mori*. *Insect Biochem. 20*, 539–545.

Yu, X.-Q., Gan, H., Kanost, M.R., **1999**. Immulectin, an inducible C-type lectin from an insect, *Manduca sexta*, stimulates activation of plasma prophenol oxidase. *Insect Biochem. Molec. Biol. 29*, 585–597.

Yu, X.-Q., Jiang, H., Wang, Y., Kanost, M.R., **2003**. Nonproteolytic serine proteinase homologs are involved in prophenoloxidase activation in the tobacco hornworm, *Manduca sexta*. *Insect Biochem. Mol. Biol. 33*, 197–208.

Yu, X.-Q., Kanost, M.R., **2000**. Immulectin-2, a lipopolysaccharide-specific lectin from an insect, *Manduca sexta*, is induced in response to Gram-negative bacteria. *J. Biol. Chem. 275*, 37373–37381.

Yu, X.-Q., Zhu, Y.-F., Ma, C., Fabrick, J.A., Kanost, M.R., **2002**. Pattern recognition proteins in *Manduca sexta* plasma. *Insect Biochem. Molec. Biol. 32*, 1287–1293.

Zhao, X., Mita, K., Shimada, T., Okano, K., Quan, G.X., *et al.*, **2001**. Isolation and expression of an ecdysteroid-inducible neutral endopeptidase 24.11-like gene in wing disks of *Bombyx mori*. *Insect Biochem. Mol. Biol. 31*, 1213–1219.

Zhao, X.F., Wang, J.X., Xu, X.L., Schid, R., Wieczorek, H., **2002**. Molecular cloning and characterization of the cathepsin B-like proteinase from the cotton boll worm, *Helicoverpa armigera*. *Insect Mol. Biol. 11*, 567–575.

Zhou, Q., Krebs, J.F., Snipas, S.J., Price, A., Alnemri, E.S., *et al.*, **1998**. Interaction of the baculovirus anti-apoptotic protein P35 with caspases. Specificity, kinetics, and characterization of the caspase/p35 complex. *Biochemistry 37*, 10757–10765.

Zhu, Y., Johnson, T.J., Myers, A.A., Kanost, M.R., **2003a**. Identification by subtractive suppression hybridization of bacteria-induced genes expressed in *Manduca sexta* fat body. *Insect Biochem. Mol. Biol. 33*, 541–559.

Zhu, Y., Wang, Y., Gorman, M.J., Jiang, H., Kanost, M.R., **2003b**. Manduca sexta serpin-3 regulates prophenoloxidase activation in response to infection by inhibiting prophenoloxidase-activating proteinases. *J. Biol. Chem. 278*, 46556–46564.

Zoog, S.J., Bertin, J., Friesen, P.D., **1999**. Caspase inhibition by baculovirus P35 requires interaction between the reactive site loop and the beta-sheet core. *J. Biol. Chem. 274*, 25995–26002.

4.8 Lipocalins and Structurally Related Ligand-Binding Proteins

H Kayser, University of Ulm, Germany

© 2005, Elsevier BV. All Rights Reserved.

4.8.1. Introduction

Studies of insects have significantly contributed to the establishment of the lipocalins as a new family of proteins. Members of the lipocalin family are defined by a specific three-dimensional motif of polypeptide folding and by the binding of small lipophilic molecules in their interior. Starting about 20 years ago, the number of lipocalins has increased enormously through the structural analysis of known proteins as well as the discovery of new proteins showing the lipocalin structural motif. Today, the lipocalins represent a superfamily of proteins including members from vertebrates, insects, mites, ticks, crustaceans, other invertebrates, and, more recently, from plants and even prokaryotes. This wide occurrence of lipocalins is thought to reflect a common ancestry and a long evolutionary history suggesting also the involvement of these proteins in many vital processes (Ganfornina *et al*., 2000).

The relatively late discovery of the lipocalins as a structural family of proteins is likely due to a low level of homology of their amino acid sequences. In pairwise comparisons of sequences, the identity between lipocalins may be below 20%, which is regarded as the lowest level for reliable alignments. Alternatively, lipocalins share a few sequence motifs in structurally conserved regions besides their overall highly conserved three-dimensional structure. Since sequence homology is of only limited value in identifying a new protein as a lipocalin, crystal structures have to be established for a definite assignment to this family, if modeling of its sequence into a known lipocalin structure is not successful. Hence, progress in the molecular characterization of

lipocalins depends on the availability of the protein of interest in sufficient quantity obtained either from a biological source or by heterologous expression of its cDNA. It depends further on the feasibility of obtaining well-diffracting crystals and, of course, on technical advancement in the crystallography field. The rapid growth of knowledge of the lipocalins with respect to structure and function in the past decade is accounted for by a recent special edition of *Biochimica Biophysica Acta* (vol. 1482, nos. 1–2, 2000) devoted entirely to this family of proteins. This issue offers a collection of relatively recent overviews of the lipocalin field dealing with general aspects of structures and functions as well as individual lipocalins including some from insects. The progress and growing importance of the lipocalins is further demonstrated by the first international Lipocalin Conference, which was held in 2003 in Copenhagen (Denmark). Those readers interested in the world of lipocalins are also referred to the lipocalin website.

The significant contributions from insect studies to the discovery and understanding of the lipocalins have not yet been comprehensively documented. The intention, therefore, of this chapter is to present an overview of the lipocalins identified in insects to date, including the fatty acid-binding proteins as structurally closely related ligand-binding proteins. Since all these diverse proteins represent variations of a common core structural theme, called an antiparallel β-barrel, this chapter describes the group-specific polypeptide folding and the underlying primary structures followed by the biological context, which comprises the occurrence of these proteins and the expression patterns of the corresponding genes, as well as the established and putative functions of each type of protein. The few lipocalins and structurally related proteins from ticks, mites, and crustaceans are included to cover all arthropods. Two well-known families of insect ligand-binding proteins that turned out not to be constructed as β-barrels, however, are also described to demonstrate their alternate folding. The lipocalins and fatty acid-binding proteins referred to in this chapter are listed in **Table 1**, including their accession codes. All multiple alignments of protein sequences shown in this chapter were performed using Clustal W, version 1.82, accessed via ExPASy. Graphical representations of crystal structures were prepared with Programme O (Jones *et al.*, 1991).

4.8.2. The β-Barrel Structural Motif

The characteristic lipocalin fold was first described for the human retinol-binding protein 20 years ago (Newcomer *et al.*, 1984), which thus became a kind of reference or prototypic lipocalin. Retinol-binding protein was soon followed by another vertebrate protein, β-lactoglobulin, the crystal structure of which showed high similarity to that of retinol-binding protein, though their sequences were not closely related (Papiz *et al.*, 1986). At this very early stage of the analysis of the lipocalin family, two insect proteins were studied independently for their crystal structures and, again, the close similarities in the folding motifs to that of retinol-binding protein were recognized instantly (Holden *et al.*, 1987; Huber *et al.*, 1987a, 1987b). These insect proteins were two lepidopteran biliproteins, which are discussed in detail below (see Section 4.8.3.1). The discovery of the architectural principle of retinol-binding protein in two evolutionarily distant insect proteins immediately suggested this new folding type to be very common across the phyla, though obviously not restricted to a specific functional class of proteins.

Typically, lipocalins are small extracellular, i.e., secreted, proteins of a size in the 20 kDa range, corresponding to about 180 amino acid residues, with affinity for specific small lipophilic molecules. Therefore, lipocalins are often referred to as extracellular lipid-binding proteins (eLBPs) to differentiate them from the related intracellular type (iLBPs) (see Section 4.8.6). The remarkably conserved lipocalin tertiary protein structure is called a β-barrel with a repeated +1 topology. This means it is made up of eight sequentially arranged antiparallel β-strands, assigned A to H, which are hydrogen-bonded to form a β-sheet that is folded back on itself to form an orthogonal stacking of two layers. The result is a barrel-like structure of a spherical shape. The eight β-strands are linked by seven loops, L1 to L7, which are all very short β-hairpins, except L1 that forms a large Ω loop covering partially the barrel opening with the internal ligand binding site. The other end of the barrel is closed by an N-terminal short 3_{10}-helical structure, which is followed immediately by strand A. One side of the barrel is lined with the C-terminal sequence, which includes a conserved α-helix preceding the short extra β-strand I, which does not contribute to the barrel structure. There are typically three conserved motifs, called structurally conserved regions (SCRs), each of them being characterized by a more or less extended invariant sequence motif. Two of the conserved regions, SCR1 and SCR3, which are present in all lipocalins, are located within the N- and C-terminal sequences comprising the 3_{10}-helix + strand A and strand H + the loop to the C-terminal α-helix, respectively. Another region, SCR2, that may be less

Table 1 Arthropod lipocalins and fatty acid-/retinoic acid-binding proteins referred to in the text, and their accession codes

Name of protein	*Source*	*Sequence (SwissProt/ TrEMBL)*	*Crystal structure (PDB)*
Lipocalins			
Bilin-binding protein	*Pieris brassicae*	P09464	1BBP
Insecticyanin a, mature form	*Manduca sexta*	P00305	coord. not in data bank
Insecticyanin a	*Manduca sexta*	Q00629	
Insecticyanin b	*Manduca sexta*	Q00630	
Biliverdin-binding protein I	*Samia cynthia ricini*	Q8T118	
Biliverdin-binding protein II	*Samia cynthia ricini*	Q8T119	
Nitrophorin 1	*Rhodnius prolixus*	Q26239	1NP1; 2NP1; 3NP1; 4NP1
Nitrophorin 2	*Rhodnius prolixus*	Q26241	1EUO
Nitrophorin 3	*Rhodnius prolixus*	Q94733	
Nitrophorin 4	*Rhodnius prolixus*	Q94734	1NP4; 1ERX; 1IKE; 1KOI; 1IKJ
Triabin	*Triatoma pallidipennis*	Q27049	1AVG
Pallidipin 2	*Triatoma pallidipennis*	Q27042	
Salivary platelet aggregation inhibitor 1	*Rhodnius prolixus*	Q94731	
Salivary platelet aggregation inhibitor 2	*Rhodnius prolixus*	Q94732	
Biogenic amine-binding protein	*Rhodnius prolixus*	Q86PT9	
Histamine-binding protein 1, female specific	*Rhipicephalus appendiculatus*	O77420	
Histamine-binding protein 2, female specific	*Rhipicephalus appendiculatus*	O77421	1QFT, 1QFV
Histamine-binding protein 3, male specific	*Rhipicephalus appendiculatus*	O77422	
Serotonin-histamine-binding protein	*Dermacentor reticulatus*	Q8WSK7	
Procalin	*Triatoma protracta*	Q9U6R6	
Bla g 4	*Blattella germanica*	P54962	
Lazarillo	*Schistocerca americana*	P49291	
glial Lazarillo	*Drosophila melanogaster*	Q9NAZ3 (Q9V6K5)	
neural Lazarillo	*Drosophila melanogaster*	Q9NAZ4 (Q8SXR1)	
Gallerin	*Galleria mellonella*	Q24996	
Bombyrin	*Bombyx mori*	Q95P97	
Hyphantrin	*Hyphantria cunea*	Q8T5Q9	
Tergal gland protein Lma-P22	*Leucophaea maderae*	O46130	
Tergal calycin p18	*Leucophaea maderae*	Q95VP9	
Crustacyanin A_1 subunit	*Homarus gammarus*	P58989	1GKA (holoA_1/A_2-dimer)
Crustacyanin A_1 subunit	*Homarus gammarus*		1H91 (apoA_1/A_1-dimer)
Crustacyanin A_2 subunit	*Homarus gammarus*	P80007	
Crustacyanin C_1 subunit	*Homarus gammarus*	P80029	1I4U (apoC_1/C_1-dimer)
Fatty acid-/retinoic acid-binding proteins			
Fatty acid-binding protein MFB1	*Manduca sexta*	P31416	
Fatty acid-binding protein MFB2	*Manduca sexta*	P31417	1MDC
Fatty acid-binding protein	*Schistocerca gregaria*	P41496	1FTP
Fatty acid-binding protein	*Locusta migratoria*	P41509	
Fatty acid-binding protein	*Heliothis zea*	O76515	
Fatty acid-binding protein	*Drosophila melanogaster*	Q9VGM2	
Fatty acid-binding protein	*Anopheles gambiae*	Q17017	
Fatty acid-binding protein	*Apis mellifera*	Q9Y1C6	
Fatty acid-binding protein Blo t 13	*Blomia tropicalis*	Q17284	
Fatty acid-binding protein Lep d 13	*Lepidoglyphus destructor*	Q9U5P1	
Fatty acid-binding protein Aca s 13	*Acarus siro*	O76821	
Cellular retinoic acid-binding protein	*Manduca sexta*	O61236	

PDB, protein data bank.

well conserved, covers parts of the strands F and G and the connecting loop L6 at the bottom of the barrel. All SCRs are located at the closed end of the β-barrel which may provide a docking site for other proteins. Furthermore, lipocalin sequences display typically four cysteine residues that form two disulfide bridges at comparable positions. The typical arrangement of secondary structures and structurally

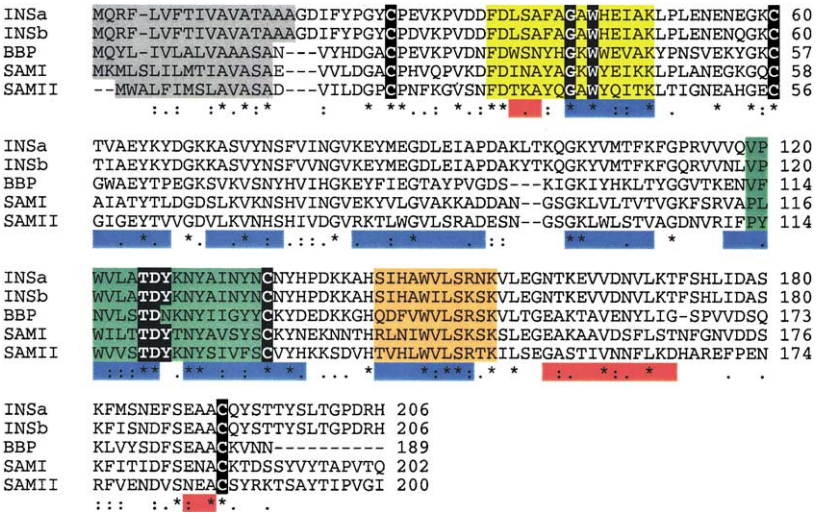


Figure 1 Alignment of the biliproteins with known full sequences. The proteins are insecticyanin a (INSa) and insecticyanin b (INSb) from *Manduca sexta*, the bilin-binding protein (BBP) from *Pieris brassicae*, and the biliverdin-binding proteins I and II (SAMI and SAMII) from *Samia cynthia ricini*. Black boxes show residues typically conserved in lipocalins. N-terminal secretion signals are boxed in gray and the structurally conserved regions SCR1, SCR2, and SCR3 are in yellow, green, and orange, respectively. Below the sequences, the approximate positions of helices and β-strands are labeled as red and blue bars, respectively. Asterisks indicate identical residues, double and single points more or less conserved substitutions.

conserved regions is illustrated in Figure 1 with sequences from biliproteins representing prototypic lipocalins (see Section 4.8.3).

The interior of the β-barrel harbors the binding site for a (usually single) molecule of ligand, which can be very different in shape, size, and physicochemical properties depending on the structure of the loop scaffold and the cavity of the individual lipocalin. Ligands can be retinoids, bilins, heme, carotenoids, and odorants (not in insects), depending on the physiological context of the carrier protein. However, the physiological ligands are not known from all lipocalins; some may also function without any bound ligand (see Section 4.8.5). The collective name "lipocalin," proposed in 1987 (Pervais and Brew, 1987), denotes these proteins as carriers of lipophilic ligands that are harbored within a "calyx" (from Greek, meaning cup). Though lipocalins may be found as monomers *in vivo* and *in vitro*, they have a tendency to form homomeric oligomers, as frequently observed in the crystalline state. Lipocalins may also complex to other soluble proteins (e.g., retinol-binding protein combines with transthyretrin in vertebrate blood) or to specific cell surface receptors. The latter aspect is not yet well documented, however.

There are several other groups of proteins that share a repeated +1 topology β-barrel with the lipocalins but display significant differences in topology and the construction of the barrel. The lipocalin-related proteins are represented by the superfamily of fatty acid-binding proteins (see Section 4.8.6), which were structurally discovered nearly one decade later, and triabin, an unusual single lipocalin that was isolated from an insect even more recently (see Section 4.8.4.3). The specific modifications of the β-barrel structure, as described above for the lipocalins, will be described in the appropriate sections. Two other protein families, the biotin-binding avidins and the bacterial metalloproteinases, which are structurally related to the lipocalins, fatty acid-binding proteins, and triabin, are outside the scope of this chapter dealing with insect proteins. All these variants of β-barrel proteins are part of a structural superfamily, called "calycins," as they are all characterized by a cuplike cavity as a real or potential internal binding site for a ligand (reviews: Flower *et al.*, 1993, 2000; LaLonde *et al.*, 1994).

4.8.3. Biliproteins: Prototypic Lipocalins

4.8.3.1. Biliproteins with Known Crystal Structure

The green coloration of many insect larvae, thought to provide camouflage at their feeding sites on plants, has been subject to many biological and chemical studies in the past. This green color is known to result from a mixture of yellow carotenoids derived from the food and blue bile pigments synthesized by the insect. Both pigments are usually associated with different proteins, the chromatographic separation of which demonstrates nicely the two underlying coloring principles (review: Kayser, 1985). Insect larvae that are fed artificial diets, usually very poor in carotenoids,

Figure 2 Last instar larvae of *Manduca sexta* fed tobacco leaves (green form) and artificial diet (blue form), respectively. The difference in color results from the lack of yellow carotenoids in the artificial diet. The blue coloration is due to the presence of insecticyanin, a biliprotein synthesized by the larvae.

consequently exhibit not a green but a blue coloration. Insect scientists working with a laboratory strain of the tobacco hornworm, *Manduca sexta*, for example, may have rarely seen the "wild" green coloration of the insect in nature (**Figure 2**). Blue chromoproteins, all most likely representing biliproteins, are present in many insect orders (Kayser, 1985). The eye-catching colors of biliproteins have certainly contributed to the fact they were studied early at the biochemical level and thus had a good chance to contribute to the discovery of the lipocalins. As mentioned above, the third and fourth lipocalin, the structures of which were independently solved by X-ray crystallography, were not only two proteins from insects but also two rather similar ones: the biliproteins from the butterfly *Pieris brassicae* and from the moth *M. sexta* (see below). Long before, both were known as proteins with an attractive sky-blue color, which is due to the presence of a bilatrien bile pigment, biliverdin IXγ. Both proteins were studied for their crystal structures in the expectation of finding a predominantly α-helical folding like that established for the light-harvesting cyanobacterial C-phycocyanins that represent biliproteins with a covalently bound derivative of biliverdin IXα (Schirmer *et al.*, 1987).

4.8.3.1.1. Bilin-binding protein from *Pieris brassicae* The biliprotein from *P. brassicae*, which was given the specific name bilin-binding protein (BBP) (Huber *et al.*, 1987a), was isolated from whole butterflies, where it is located predominantly in the wings, and characterized in the author's laboratory as a monomeric, nonglycosylated protein that is noncovalently associated with biliverdin IXγ at a 1 : 1 stoichiometry (Zipfel, 1982; Kayser, unpublished data). A molecular mass of ~20 kDa was obtained for purified BBP by both gel filtration and sodium dodecylsulfate (SDS)-gel electrophoresis consistent with its monomeric status. Alignment of the BBP sequence revealed closest homology to the biliprotein from *M. sexta*, known as insecticyanin (see Section 4.8.3.1.2), and to human apolipoprotein D with identities of 43% and 31%, respectively. The mature protein, as isolated from the butterflies, comprises 174 amino acid residues corresponding to a molecular mass of 19 790 Da in agreement with the chromatographic results (Suter *et al.*, 1988). An N-terminal signal sequence of 15 amino acids is found in the precursor, which has been cloned from larvae (Schmidt and Skerra, 1994). In the crystal state BBP was found as a tetramer consisting of two dimers (Huber *et al.*, 1987a, 1987b). The folding of the monomers was that of the typical eight-stranded β-barrel with orthogonal arrangement in the two sheets (**Figure 3**), as was first described for the retinol-binding protein. A typical short 3_{10}-helix is located in the N-terminal sequence just before the first β-strand. A long α-helix is attached along one side of the barrel in the C-terminal region that terminates in a short helical sequence. BBP contains two cysteine bridges that are believed to support this folding structure. Though the sequence homology of these two lipocalins is only ~10%, their tertiary structures match almost perfectly in the central and lower segments, while the loop region around the open end of the barrel is more varied. The calyx-like cavity of BBP harbors

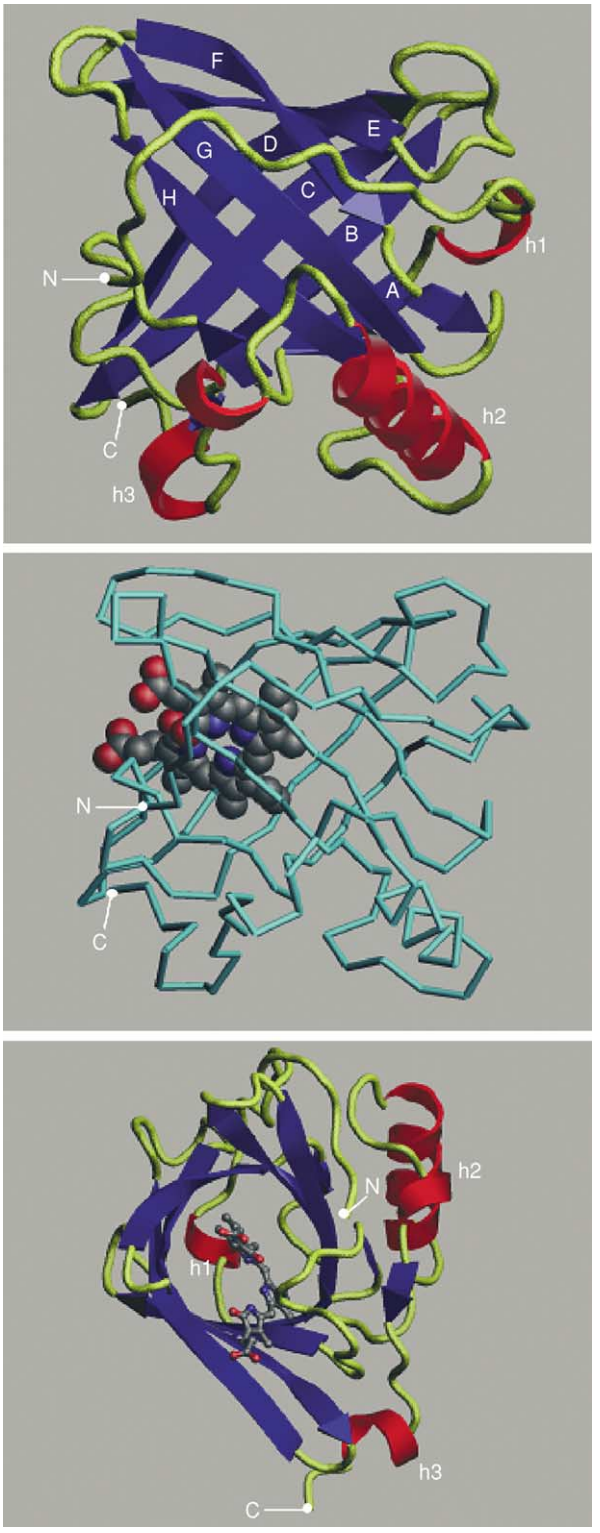


Figure 3 Crystal structure of the bilin-binding protein (BBP) from *Pieris brassicae*. Upper and lower panels: ribbon drawing with eight β-strands (blue arrows, labeled A–H) and three helices (red, labeled h1–h3). Middle panel: Cα-backbone structure (light blue) of the protein (same orientation as in the upper panel) with bound biliverdin-IXγ represented as stick-and-ball model. Oxygens are in red and nitrogens in blue. The N- and C-termini are marked with white dots.

the chromophore ligand, which is not deeply buried in the cavity but located close to the open end of the barrel with the two carboxyl groups of the heme propionate side chains pointing to the solvent. This implies only little contribution of these polar groups to the binding of the tetrapyrrole. The pyrrole nitrogens of the bilin are complexed to water molecules and in contact with the protein via a few hydrogen bonds.

The identity of the ligand of BBP was confirmed by its crystal structure as the γ-isomer of biliverdin IX. The bilin shows a cyclic helical structure characterized by all-*Z* configuration and all-*syn* conformation (**Figure 3**). This geometry is adopted also by free bilatrienes in solution (Kayser, 1985) and in agreement with the visible absorption spectrum of the bound chromophore (see below). The cyclic helical geometry of a bilatrien gives rise to two enantiomers with opposite helical sense, which are rapidly interconverted in solution. The crystallographic analysis revealed that the protein binds only one enantiomer of the ligand (**Figure 3**), thus inducing chirality in a compound that is achiral by structure (Huber *et al.*, 1987b; Scheer and Kayser, 1988). The induced optical activity of the *P. brassicae* BBP is nearly as strong as that in the cyanobacterial phycocyanins, which has a different basis, however. The circular dichroism (CD) spectrum of BBP shows a positive Cotton effect in the visible range, which could be correlated with the right-handed helix of the bilin for the first time (Huber *et al.*, 1987b). Treatment of the holoprotein with urea completely abolishes the optical activity of the holoprotein due to the unfolding of the protein that destroys the specific ligand binding site (Scheer and Kayser, 1988). As shown with the recombinant BBP apoprotein derived from larval cDNA, the holoprotein can be reconstituted by incubation with free biliverdin IXγ to yield an apparently native product as judged from its characteristic absorption spectrum (Schmidt and Skerra, 1994). However, no CD spectrum was recorded, hence the correct binding of the chromophore as the right-handed helical enantiomer has not been demonstrated. As indicated above, the visible spectrum of the bound chromophore is similar to that of the free form, as both are in a cyclic helical conformation. Characteristic differences, however, are seen at the long-wave absorption peak with a bathochromic shift of ~25 nm and a significant hyperchromic effect to produce an absorbance plateau around 670 nm; the sharp short-wave peak (the typical Soret band of tetrapyrroles) is at 383 nm (Zipfel, 1982; Scheer and Kayser, 1988). The tight fitting of ligand binding to its carrier lipocalin, as is obvious in case of BBP, may

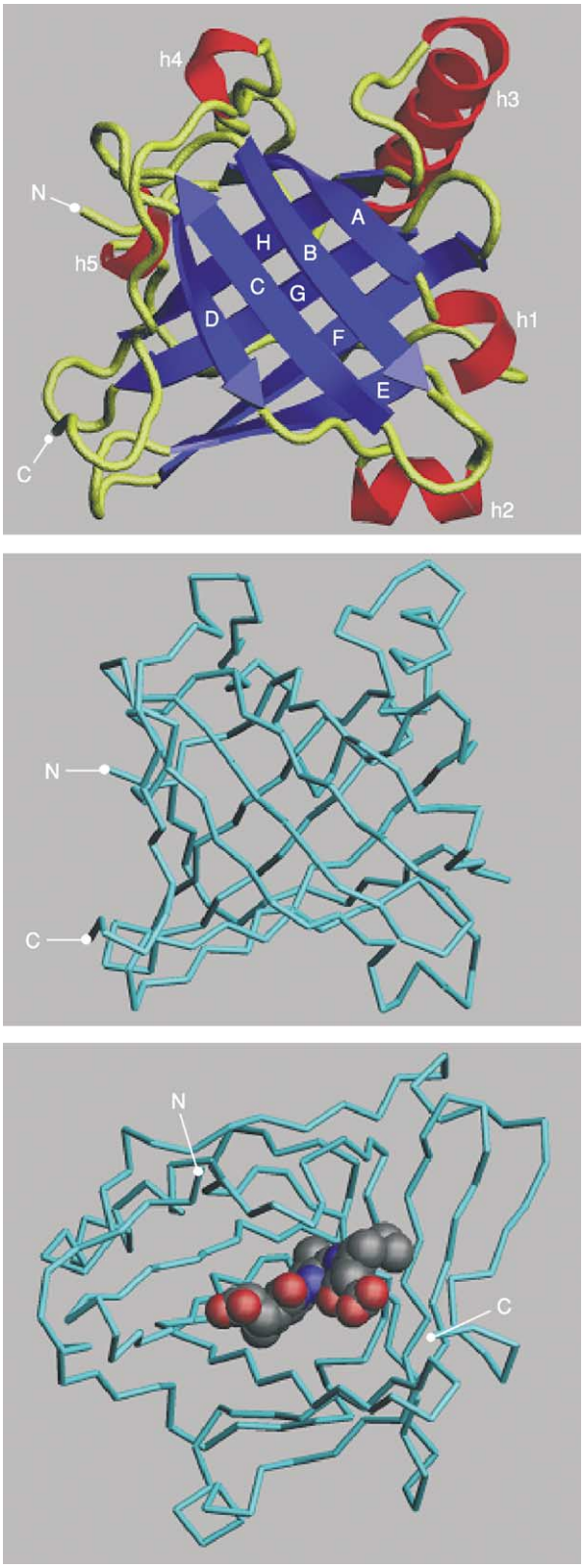


Figure 4 Crystal structure of insecticyanin from *Manduca sexta*. Upper panel: ribbon drawing with eight β-strands (blue arrows, labeled A–H) and several helices (red, labeled h1–h5). Middle and lower panels: Cα-backbone structure (light blue) of the protein with bound biliverdin-IXγ in the lower panel. The ligand is represented as stick-and-ball model with oxygens in red and nitrogens in blue. The upper and middle panels show the protein in the same orientation. The N- and C-termini are marked with white dots.

generally result in some electronic interaction between the bound chromophore and the protein. Conversely, a biliprotein with a fairly unchanged absorption spectrum versus the free bilin is unlikely to represent a lipocalin.

4.8.3.1.2. Insecticyanin from *Manduca sexta* Insecticyanin from larvae of the tobacco hornworm, *M. sexta*, was the first insect biliprotein that was purified and studied at the biochemical level (Cherbas, 1973). It can be easily obtained from larval hemolymph, where it may account for up to 5% of the total protein. Insecticyanin is very similar to BBP from *P. brassicae* with respect to its molecular size (189 amino acids for the protein as isolated; molecular mass of 21 378 Da), the presence of two disulfide bridges, the absorbance spectrum, the noncovalent binding of one molecule of biliverdin IXγ per monomer as well as the absence of bound carbohydrate and lipid (Riley *et al.*, 1984; Goodman *et al.*, 1985). The crystal structure of insecticyanin, solved independently and published only weeks after that of BBP, revealed very close similarity to that of retinol-binding protein (Holden *et al.*, 1987). This finding extends the overall similarity between the biliproteins from *M. sexta* and *P. brassicae* to the level of their crystal structures, though the sequence identity is only 40%. Moreover, both biliproteins crystallized as tetramers.

As shown in **Figure 4**, insecticyanin is folded as a barrel made up of eight antiparallel β-strands with an orthogonal arrangement of the two sheets, comparable to BBP. The N- and C-terminal sequences with several conserved helical structures are wrapped near the opening of the barrel. There is an extra helical structure inserted between the fourth and fifth β-strand (labeled h2 in **Figure 4**). The biliverdin ligand is located near the opening with the two propionate side chains directed to the solvent. The ligand of insecticyanin is in contact with hydrophobic residues on both sides, one of which is more polar than the other one in contrast to the situation in the related BBP. Like in the latter, biliverdin IXγ in insecticyanin adopts a porphyrin-like cyclic helical conformation, which is also right-handed like that in BBP (cf. **Figures 3 and 4**). There is no mention of the helical sense in the publication by Holden *et al.* (1987). As expected from the study of the *P. brassicae* protein, the chromophore of

insecticyanin is optically active, producing a CD spectrum that is consistent with a right-handed helical sense (Kayser, unpublished data).

4.8.3.2. Biliproteins with Unknown Crystal Structure

Due to their widespread occurrence in insects and easy visibility, a number of biliproteins have been isolated, mostly from lepidopteran species, and biochemically characterized in more or less detail (Table 2). The biliproteins from *P. brassicae* and *M. sexta* have been studied most thoroughly including their crystal structures, as described above. The possible classification of the other biliproteins as lipocalins depends on the results of future crystallographic analysis or successful modeling of amino acid sequences into the crystal structures of one of the known biliproteins. The full-length amino acid sequences are also known for the two biliproteins from the silk moth *Samia cynthia ricini* and from two other proteins from the butterfly *P. rapae*. The latter two proteins, which differ from each other by

Table 2 Purified insect biliproteins

Insects	*Name of protein*[a]	*Subunit size (kDa)*	*Oligomeric state*	*Bilin type (bilin/subunit)*	*Glyco/ Lipo protein*	*Reference*
Lepidoptera						
Pieris brassicae	Bilin-binding protein BBP[b,c]	19.8	Monomer	BV IXγ (1 : 1)	−/−	Suter *et al.* (1998)
Pieris rapae	Bilin-binding proteins BBP-1, BBP-2	~18	?	BV IXγ ? (?)	?/?	Yun (personal communication)
Manduca sexta	Insecticyanin[b,c]	21.4	Tetramer	BV IXγ (1 : 1)	−/−	Riley *et al.* (1984)
Agrius convolvuli	*A. convolvuli* insecticyanin[d]	21.2	Trimer (?)	BV IXγ? (1 : 1)	?/?	Saito and Shimoda (1997)
Attacus atlas	Biliverdin-binding protein (BBP)[d]	22.4	Dimer	BV IXγ? (1 : 1)	?/?	Saito (1997)
Samia cynthia ricini	Biliverdin-binding proteins BBP-I,[c] BBP-II[c]	I: 20.5 II: 22.7	Monomer Dimer	BV IXγ (1 : 1) BV IXγ (1 : 1)	?/? ?/?	Saito (1998a)
Antheraea yamamai	Biliprotein BP = P-II[d]	21.5	Monomer	Phorcabilin (?)	?/?	Saito *et al.* (1998)
Rhodinia fugax	Biliproteins BP-I,[d] BP-II[d]	I: 22.6 II: 22.9	I: Monomer II: Dimer	I: phorcbilin (?) II: BV IXγ (?)	?/? ?/?	Saito (1998b)
Cerura vinula	*Cerura (vinula)* biliprotein	~80	Tetramer ?	unknown (?)	?/?	Scheer and Kayser (1988)
Heliothis zea	Blue chromoprotein	~150	Tetramer	BV (?)	+/+	Haunerland and Bowers (1986)
Trichoplusia ni	Chromoprotein	~150	Dimer	BV (?)	+/+	G. Jones *et al.* (1988)
Spodoptera litura	Biliverdin-binding proteins BP-1–BP-4	~165	Dimers	BV (?)	+/+	Yoshiga and Tojo (1995)
Spodoptera litura	Biliverdin-binding vitellogenin	~188	Dimer	BV (?)	?/+	Maruta *et al.* (2002)
Orthoptera						
Locusta migratoria	Cyanoprotein	~83	Tetramer	BV (2 : 1)	+/−	Chino *et al.* (1983)
Heteroptera						
Podisus maculiventris	Blue lipophorin	~250/~80/~20	>700 kDa	BV IXγ (?)	?/+	Haunerland *et al.* (1992b)
Riptortus clavatus	Cyanoproteins CP1–CP4	~76	Hexamers	BV IXγ (4 : 1)	+/−	Chinzei *et al.* (1990)

[a]As referred to in the reference.
[b]Crystal structure available; see **Table 1**.
[c]Full sequence available; see **Table 1**.
[d]N-terminal sequence available.
BV, biliverdin.

a few residues only, are of the same size as BBP from *P. brassicae* to which they show ~94% identity (Yun, personal communication). Hence, it is most likely that the biliproteins from *P. rapae* have crystal structures almost identical to that of BBP from the closely related *P. brassicae*. N-terminal sequences are known from several moth biliproteins (see **Table 2**).

The chromophores of most insect biliproteins have been described on the basis of their visible absorption spectra as either biliverdin IXγ or "biliverdin(-like)" without further specification (**Table 2**). A more rigorous approach by, for example, microchemical degradation has not been performed with any of these chromoproteins. On the other hand, the identification of biliverdin IXγ, though tentative so far, in most of the lepidopteran proteins is not surprising since this biliverdin isomer is typical for this insect order (review: Kayser, 1985).

Phorcabilin, a derivative of biliverdin IXγ with extended chromophore geometry, seems to be associated with proteins in two saturniid moths only, in accordance with the restrictive occurrence of this bilin (Kayser, 1985). The ligand of the biliprotein from the moth *Cerura vinula* has not yet been identified. Preliminary results suggest a new chromophore, possibly a derivative of biliverdin IXγ (Scheer and Kayser, 1988; Kayser, unpublished data).

In many cases, the names of the insect biliproteins that have been described to date may give rise to confusion, because they are used for different proteins, or they simply refer to the blue color, as documented in **Table 2**. Though this is no attempt to propose a nomenclature for biliproteins, it should be kept in mind that the name "bilin-binding protein" or "BBP" has been given specifically to the biliprotein(s) from *P. brassicae* (Huber *et al.*, 1987a, 1987b). Similarly, "insecticyanin" has been coined for the biliprotein from *M. sexta* (Cherbas, 1973). These names should not be used for biliproteins from other species. Furthermore, a protein may be denoted as "biliverdin-binding" only if the chromophore has been definitely identified as one of the isomers of biliverdin IX. If the chemical class of the (blue) ligand is unknown, the protein may be described just as, for example, a "blue chromoprotein," "blue lipophorin," or "cyanoprotein," as in the past. To avoid confusion in future work it is proposed to use the generic term "biliprotein," abbreviated as "BP," linked to the name of the species from which it was isolated.

At least two groups of biliproteins may be discriminated on the basis of their overall characteristics (**Table 2**). One group is characterized by small proteins of ~20 kDa, lack of sugar and lipid constituents, binding of a single ligand per subunit, and occurrence as monomers or as homo-oligomers. This group may represent lipocalins, as typified by BBP from *P. brassicae* and insecticyanin from *M. sexta*. The second group comprises proteins characterized by much larger subunits in the 80 kDa or even 150 kDa range, the presence of sugar and/or lipid and the aggregation to complexes with sizes of up to 600 kDa and more. Some proteins of this type, such as the biliverdin-binding vitellogenin from *Spodoptera litura* (Maruta *et al.*, 2002), the blue lipophorin from *Podisus maculiventris* (Haunerland *et al.*, 1992b), and the cyanoprotein from *Riptortus clavatus* (Chinzei *et al.*, 1990) show properties that relate them to the storage proteins which are used during the embryo and metamorphosis stages (see Haunerland, 1996). In most of the biliproteins listed in **Table 2**, the absorption spectrum of the bound chromophore differs more or less significantly from that of the free ligand obtained after extraction with an organic solvent. As in the *Pieris* BBP, the long-wave absorption maximum of the bilin is usually redshifted (bathochromic effect) and increased (hyperchromic effect), which implies some electronic interaction between the ligand and the protein. Whether such spectral changes indicate that the ligand is specifically bound with respect to its structure and the binding site is not clear. It may well be that less specific binding of a ligand to a site on the protein surface or to the lipid moiety of a protein also results in a shift in the absorption spectrum. In any case, the binding cavity of the known lipocalins would not be large enough to harbor more than one bilin per monomer. So, the nature of the ligand-binding site of the large-size biliprotein remains conjecture. However, the crystal structure of the ~80 kDa subunit biliprotein from *Cerura vinula*, isolated in the author's laboratory, may be solved in the near future if current attempts to obtain high-quality crystals are successful.

4.8.3.3. Molecular Variants of Biliproteins

In a number of species, multiple biliproteins have been isolated and characterized by their amino acid composition, N-terminal sequence, and molecular mass (**Table 2**). Examples are found among the small-size (~20 kDa) proteins from two saturniid moths. In *Samia cynthia ricini*, the two biliverdin-binding proteins (BBP-I and BBP-II) from larval hemolymph share only 48% identity in their N-terminal 50 amino acid residues. BBP-II is likely identical to the biliprotein isolated from the molting fluid at pupation from the same insect (Saito, 1993, 1998a). The identity of the predominant BBP-I from this moth with BBP from *P. brassicae*

and insecticyanin from *M. sexta* is 46% and 52%, respectively. The corresponding value for BBP-II is 42% for both comparisons. In *Rhodinia fugax*, the biliproteins BP-I from larval hemolymph and cuticle and BP-II from epidermis are similar to one another and to those from *P. brassicae* and *M. sexta*, respectively (Saito, 1998b).

An example for multiple large-size (~165 kDa) proteins has been found in *Spodoptera litura*, from which four dimeric hemolymph biliverdin-binding proteins, differing in their isoelectric point, have been isolated indicating the presence of several non-identical subunits of about equal size (Yoshiga and Tojo, 1995). Unfortunately, partial sequences are not available to support this conclusion, but immunological differences have been recognized. As mentioned above, another biliverdin-binding protein from the same moth has been identified as a vitellogenin as it is female-specific and shows significant sequence homology with lepidopteran vitellogenins (Maruta *et al.*, 2002).

The oligomeric situation with the cyanoproteins from the bug *Riptortus clavatus* is different (Chinzei *et al.*, 1990). Here, two distinct subunits of about equal size combine to form four biliverdin-associated hexameric complexes which have been isolated as CP1 to CP4. According to their cDNA-deduced amino acid sequences, these proteins are closely related to arthropod hemocyanins and phenol oxidases, which are members of the superfamily of hexamerins (Miura *et al.*, 1998). Their hexameric state and the high ratio of chromophore per subunit of 4 : 1, as calculated for CP1, support the conclusion that these bug cyanoproteins are not lipocalins.

Full data at the gene level on molecular variants of biliproteins have been obtained for the established lipocalin from *M. sexta* (Kiely and Riddiford, 1985; Riddiford *et al.*, 1990; Li and Riddiford, 1992, 1994, 1996). In this insect, two major isoelectric forms of insecticyanin, described as a more acidic INS-a (pI 5.5) and a more basic INS-b (pI 5.7), have been identified. Both forms are present in the larval epidermis and cuticle, whereas only INS-b is found in the hemolymph. It is this form which has been studied as the first insect biliprotein by Cherbas (1973). The protein sequences encoded by the cDNAs for INS-a and INS-b, respectively, differ in 13 of the 189 amino acid residues of the mature protein; the N-terminal signal sequences are identical. The 3′ noncoding regions of the two cDNAs contain a sequence stretch that is unique for each gene and led to specific probes for expression studies of the two duplicated insecticyanin genes.

Comparable to the results on insecticyanin in *M. sexta*, two isoelectric forms of the BBP were purified from *P. brassicae* with pI values of 6.4 (BBP-I) and 6.2 (BBP-II), respectively (Kayser, unpublished data). The two proteins differ only in the N-terminal residue, which is asparagine in the predominant form BBP-I and aspartate in the minor variant BBP-II (Huber *et al.*, 1987b). This single difference in sequence fully accounts for the observed difference in the isoelectric point of the two forms. As expected, the crystal structures of the two BBP variants are identical, but BBP-II does not dimerize. In contrast to the two isoforms obtained as proteins from adult insects, only one cDNA sequence coding for BBP (174 amino acid residues of the mature protein plus an N-terminal signal peptide of 15 amino acids) has been found in last instar larvae, which are actively expressing this gene (Schmidt and Skerra, 1994), and accumulate most of the BBP holoprotein during the insects' life (Kayser, 1984; Kayser and Krull-Savage, 1984). Hence, BBP-II may arise artifactually by deamidation of BBP-I during the process of purification. The presence of the minor form BBP-II *in vivo* has yet to be confirmed.

Two isoelectric forms of biliproteins are also known from the related butterfly, *P. rapae*, which differ in 12 nucleotides in their cDNA sequences (Yun, personal communication). Hence, there are obviously separate genes for the two biliproteins in *P. rapae*, in contrast to only a single gene in *P. brassicae*.

4.8.3.4. Developmental Expression and Biosynthesis of Biliproteins

Though biliproteins are widespread among insects and may share a number of physicochemical features, there seems to be no clear common pattern with respect to their location in the body, their site of synthesis, and their fate during development. Detailed studies of the transcription of genes coding for apobiliproteins and the corresponding synthesis of proteins have been performed for insecticyanin in *M. sexta* (Riddiford, 1982; Trost and Goodman, 1986; Riddiford *et al.*, 1990; Li and Riddiford, 1994). Insecticyanin is present in the larval epidermis and hemolymph up to the adult stage. The highest concentration is found in the hemolymph of early fourth instar larvae and in adults after eclosion (up to 0.8 mg ml^{-1}), while the levels decrease after each larval molt. The genes *ins*-a and *ins*-b coding for the two isoelectric forms of insecticyanin are expressed during all larval instars with significant cyclic variations, followed by corresponding increases and decreases of the protein titers. Since the expression of the *ins* genes is restricted to the larva, insecticyanin present in the adult insect must

be retained from the larval stage. Both insecticyanin variants are mainly synthesized in the larval epidermis and to a lesser extent in the fat body. While the epidermis stores both proteins in granular form and secretes them into the cuticle, it delivers only INS-b to the hemolymph. INS-b is also produced as the predominant isoform in the fat body, but obviously released quickly and not stored there, as this tissue lacks blue coloration. The epidermal expression of *ins*-a lasts until pupal commitment, while that of *ins*-b continues up to the wandering stage. This differential control of gene expression is due to a small peak in the ecdysteroid titer occurring after the disappearance of juvenile hormone (Riddiford *et al*., 1990; Li and Riddiford, 1994). Thus, the developmental regulation of insecticyanin synthesis is apparently part of the overall endocrine program that controls the larval–pupal transition. In the black mutant of *M. sexta*, which is characterized by a lack of juvenile hormone (see **Chapter 3.7**), the expression of both *ins* genes is reduced in the epidermis, while it is enhanced in the fat body, suggesting a direct, tissue-specific effect of juvenile hormone on the expression of these genes (Li and Riddiford, 1996).

Insecticyanin is present in mature eggs of *M. sexta* at a concentration exceeding that in the hemolymph, though the *ins* genes are not expressed in any cell type in pupae and adults. The protein persisting from the larval stage is taken up from the hemolymph into the developing oocytes by a saturable and specific membrane-bound mechanism that requires the presence of calcium, as found by Kang *et al*. (1995, 1997). An apparently multimeric receptor with an estimated size of ~185 kDa that selectively binds insecticyanin with high affinity (K_d ~17 nM) has been characterized in detergent extracts from oocyte membranes. Thus, insecticyanin in the *Manduca* egg behaves like the storage proteins vitellogenin and lipophorin despite the differences in protein architecture, molecular size and mechanism of uptake into the egg (review: Haunerland, 1996) (see **Chapter 3.9**).

The expression of the BBP in *P. brassicae* is under developmental and tissue-specific control. Its mRNA can be demonstrated in penultimate and last instar larvae with a depression during the last larval and larval–pupal molts (Schmidt and Skerra, 1994). This kind of cyclic expression resembles that seen for insecticyanin in *M. sexta*. On the other hand, the BBP gene is not expressed in the larval epidermis but in the fat body, predominantly in the dorsolateral sheets and in the fat body tissue underlying the epidermis. This is not surprising since this tissue is of bluish coloration. Only little BBP mRNA is seen in the ventrolateral fat body, which is colorless indicating that the holoprotein is not stored there to a significant extent. Furthermore, the gene seems to be active also in the larval "gonads" (testes, according to this author). In marked contrast to the situation in *M. sexta*, the expression of the BBP gene in *P. brassicae* is again strong in the late pupa (pharate adult) and in young adults before it ceases almost completely (Schmidt and Skerra, 1994). This developmental pattern of BBP expression is in accordance with results from biochemical studies on the protein titer and the incorporation of specific precursors into the bilin and the apoprotein, respectively (Kayser, 1984; Kayser and Krull-Savage, 1984). The most active phases of holoprotein synthesis are in the last larval instar before the wandering stage and in late pupae following the increased ecdysteroid titer, which triggers adult development (see **Chapter 3.5**). The BBP synthesis ceases 2 days after adult emergence. The strong synthesis of BBP in developing adults is surprising as it is not sequestered into the eggs of *P. brassicae*, in contrast to the fate of insecticyanin in *M. sexta*. Most of the BBP in the butterfly is located in the wings, which are blue underneath the layer of white scales containing large amounts of pterin pigments in granular form (review: Kayser, 1985). Recent studies in the author's laboratory revealed that this wing BBP is synthesized *de novo* as holoprotein in the wings during adult development, while the BBP gene is switched off in the rest of the body (Sehringer, 1999; Sehringer and Kayser, unpublished data). These studies also demonstrated that the wings are capable of synthesizing the bilin ligand of BBP from the specific heme precursor 5-aminolevulinate (see below). Results from a developmental study of the activity of one of the key enzymes of tetrapyrrole synthesis, porphobilinogen synthase, are in accordance with the prominent phases of BBP synthesis and with the wings as the sites of BBP synthesis in the pupal stage (Kayser and Rilk-van Gessel, unpublished data).

There is an interesting difference between *P. brassicae* and *P. rapae* relating to the regulation of their biliproteins during development. In *P. rapae*, the isoform BBP-1 is abundant in the larval hemolymph, while there is more BBP-2 in the pupal stage (Yun, personal communication). The site of synthesis of the two *P. rapae* biliproteins remains to be studied. No isoforms with differential temporal expression are known from the closely related *P. brassicae*, as mentioned above.

In the larvae of *Spodoptera litura*, the four biliverdin-binding proteins that are composed of two ~165 kDa lipoprotein subunits (**Table 2**) also

show hemolymph levels that fluctuate cyclically in the course of larval development (Yoshiga *et al*., 1998), as described above for insecticyanin. The timing of BP-4 differs from that of the others, as it appears in earlier instars and has a low titer in the last instar, when the levels of the others are high. Synthesis of BP-4 seems to be triggered by juvenile hormone (Yoshiga and Tojo, 2001). At pupation, the biliproteins are taken up by the fat body, which turns blue in the pupa. A comparable sequestering of large-size biliproteins as glycolipoproteins by the fat body from the hemolymph at pupation is observed in the moth *Heliothis zea* (Haunerland and Bowers, 1986) and in the bug *Riptortus clavatus* (Chinzei *et al*., 1990). These biliproteins with subunits of ~150 kDa and ~76 kDa, respectively, represent high-density lipoproteins (see **Chapter 4.6**) and behave like storage proteins that are used up for adult development and egg formation. It is remarkable that the *Manduca* insecticyanin, a typical ~20 kDa lipocalin, also behaves like a storage protein in the embryo, as described above. This demonstrates that the presence of a bilin ligand does not denote a specific functional class of proteins.

As for the biosynthesis of the bilin chromophore of biliproteins, labeling studies in a number of insects have demonstrated that the heme precursor 5-aminolevulinate is efficiently incorporated into the bilin (review: Kayser, 1985). More detailed experiments, also in relation to the synthesis of the apoprotein, seem to be available only for the BBP from *P. brassicae* (Kayser, 1984; Kayser and Krull-Savage, 1984). These studies revealed that the syntheses of the bilin and the apoprotein are coordinated but independent from each other, since cycloheximide has a strong inhibitory effect on the synthesis of the apoprotein, while the incorporation of radiolabel from 5-aminolevulinate into protein-bound bilin was only weakly reduced. This suggests the presence of a relatively large pool of apoprotein available for binding of the heme derivative. A study of the time course of BBP synthesis in young adults of *P. brassicae* demonstrated further that label from [^{14}C]5-aminolevulinate appeared in the bilin almost immediately with no lag phase, comparable to the results for the apoprotein labeled with [^{3}H]leucine (Kayser and Krull-Savage, 1984). This result is consistent with a straightforward synthesis of the bilin, presumably via a heme precursor. Obviously, the bilin is not accumulated like a degradation product of some relatively stabile heme protein, for example like hemoglobin in vertebrates. The formation of bilins from heme has not yet been studied at the enzyme level in insects, though the action mechanism of heme oxygenase is fairly well understood and evolutionarily conserved (reviews: Maines, 1997; Wilks, 2002). It is furthermore unknown whether apobiliproteins or their precursors could serve as heme oxygenases, or whether the open chain tetrapyrrole is synthesized independently to associate with the apoprotein, as studies in *P. brassicae* suggested (Kayser and Krull-Savage, 1984). This possibility gains some support from the apparently successful reconstitution of BBP from its recombinant apoprotein (Schmidt and Skerra, 1994).

4.8.3.5. Putative Functions of Biliproteins

Usually, biliproteins are considered to contribute primarily to camouflage coloration of insects, in particular to reduce the risk for larvae feeding on (green) plants of being preyed upon. This is due to their strong blue color, which combines with the yellow color of carotenoids derived from the food plant to yield green coloration that may be modified by, e.g., dark pigmentation, to further enhance any camouflage effect (review: Kayser, 1985). This role is certainly true and evident in many insects, as in the example shown in **Figure 5**. However, there are also examples where biliproteins are present in epidermis, hemolymph, and/or fat body of the larva but are masked by other pigmentation of the cuticle. One obvious example is the butterfly *P. brassicae*. The larval integument displays a pattern of black and yellow pigments that mask the biliprotein underneath, and in the adults, where the bulk of

Figure 5 Fourth instar larva of *Cerura vinula* as an example for camouflage coloration based on yellow carotenoids and a blue biliprotein present in epidermis and hemolymph, combined with ''lytic'' dark pigmentation in the cuticle.

the biliprotein is located in the wings, the scales are bright white, except for a few black patches. Presumably, the camouflage potential is a secondary property of biliproteins that is successfully utilized in many insects.

A major, more general role of biliproteins is inferred from several considerations. The high cost of establishing the complex multienzyme pathway of heme synthesis followed by heme cleavage seems to be inadequate for biliproteins to be limited to just a role in camouflage, which might be achieved by other, "cheaper" means. Furthermore, the strong synthesis of heme products without any known need for a corresponding large quantity of a heme protein, such as hemoglobin or cytochrome P450 (see **Chapter 4.1**), suggests a genuine and vital role for the biliproteins as end (not waste) products of the heme pathway. In newly emerged adults of *P. brassicae*, for example, about 90% of the heme precursor 5-aminolevulinate is allocated directly to the synthesis of BBP, even at the time of peak formation of mitochondrial cytochromes during flight muscle development (Kayser, 1984).

The photochemical properties of the cyanobacterial biliproteins, which function as accessory light-harvesting complexes, led to a study of the photochemical behavior of an insect representative, BBP from *P. brassicae* (Scheer and Kayser, 1988; Schneider *et al*., 1988). The fluorescence intensity of BBP was very weak, compared to that of the cyanobacterial counterparts, and there was no indication of an intermolecular energy transfer in accordance with the monomeric nature of this protein. Overall, the photochemical properties of BBP are much too weak to support any speculation about a possible role based on the use of light for energy conversion or signaling. Most likely, this is true for insect (animal) biliproteins in general.

Concerning possible metabolic roles of biliproteins, it must be recalled that the cleavage of heme yields an open-chain tetrapyrrole, which is a biliverdin in all studied organisms, as well as carbon monoxide and iron in equimolar quantities. All these products are biologically active. A role for the released iron is difficult to reconcile since it has to be introduced into the porphyrin for its cleavage via its Fe-complex, heme. The production of carbon monoxide is possibly more relevant, as evidence is being accumulated for this gas to function as a second messenger comparable to the well-established role of nitric oxide which regulates a wide spectrum of biochemical and physiological processes by binding to the heme moiety of soluble guanylyl cyclases. Moreover, the pathways of nitric oxide and carbon monoxide signaling have been shown to interact (reviews: Barañano and Snyder, 2001; Hartsfield, 2002; Ryter *et al*., 2002). Thus, heme oxygenase is in a position to potentially control a number of cellular reactions. Isoforms of heme oxygenases with different functions and tissue specificity are known from vertebrates. Insects have been widely neglected in this field so far.

There are more speculative roles for bilins and their carrier proteins. The cleavage of the heme ring to an open-chain bilin is unlikely to be just a means to control the cellular levels of porphyrins because of the high investment to establish such a mechanism. Multistep biosynthetic pathways are usually regulated at the level of initial (key) enzymes, not at the terminal ones. Porphyrins are cytotoxic due to their photochemical properties and therefore have to be inactivated or eliminated. Their cleavage to a linear tetrapyrrole requires the insertion of ferrous iron for the activation of molecular oxygen and reduction equivalents, usually NADPH. However, bilins may also have regulatory roles according to recent studies in vertebrates. For example, biliverdin has been shown to inhibit soluble guanylyl cyclase *in vivo* and *in vitro* (Koglin and Behrends, 2002), to play a role in embryonic development (Falchuk *et al*., 2002), and, as a redox partner of bilirubin, to play a vital role in the prevention of cellular damage by reactive oxygen and nitrogen species (Barañano *et al*., 2002; Kaur *et al*., 2003). Apparently, a variety of actions of linear tetrapyrroles in biological systems are just emerging.

4.8.4. Nitrophorins and Related Proteins

4.8.4.1. The Problem of Blood Feeders

The basic lipocalin fold has evolved as a universal scaffold that can be modified to achieve affinity for chemically diverse ligands involved in a variety of biological functions. This universality becomes strikingly evident with the nitrophorins and functionally related proteins that have evolved independently in several arthropods specialized in taking blood meals. General aspects of this special feeding behavior have been reviewed very recently (Ribeiro and Francischetti, 2003). Blood-sucking arthropods, such as a number of bugs and ticks, have developed biochemical means to maintain a continuous blood flow at the feeding site and to prevent or delay any counteraction of the host's body that would reduce feeding success. This common requirement of blood feeders obviously guided the evolution of proteins specialized to interact at different sites and in different ways with the complex

process of blood coagulation in vertebrates. Obviously, the lipocalin fold provided a versatile basis to create binding sites for powerful ligands acting together to support the insects' needs. The basic studies on this group of functionally highly specialized lipocalins have been performed in *Rhodnius prolixus*, a South American bug known to transmit *Trypanosoma cruzi*, the vector of Chagas' disease. A general overview with a focus on the structures of the specialized proteins of blood-feeding insects has been given by Montfort *et al.* (2000). The chemical and physical properties of nitrophorins have been reviewed by Walker *et al.* (1999).

4.8.4.2. Crystal Structures of Nitrophorins from *Rhodnius prolixus*

The nitrophorins of the bug *R. prolixus* are stored in the salivary gland, which therefore looks red, and are injected, together with a number of other proteins of related functions, into the blood of the host. There are four nitrophorins, NP1 to NP4, that have been purified from the insect tissue, cloned, and produced as recombinant proteins for crystallization studies (Champagne *et al.*, 1995; Ribeiro *et al.*, 1995; Andersen *et al.*, 1997; Sun *et al.*, 1998). The numbering of the nitrophorins is in the order of their relative abundance with NP1 representing the predominant one. Each protein has a mass of ~20 kDa as is typical for members of the lipocalin family. In pairwise comparisons, NP1 and NP4 show ~90% sequence identity, while this is ~80% for NP2 and NP3. The identities between the two pairs are much lower, about 45%. This demonstrates different degrees of relatedness suggesting that the four nitrophorins originated from two gene duplications. The overall sequence identity of the nitrophorins is 38%.

The crystal structures of the nitrophorins, which are available from NP1, NP4, and NP2 in various complexes, are largely identical (Andersen *et al.*, 1998; Weichsel *et al.*, 1998, 2000). All of them show the typical conserved lipocalin folding of a β-barrel consisting of eight antiparallel β-strands with orthogonal orientation in the two sheets (**Figure 6**), as described in detail for the biliproteins from *P. brassicae* and *M. sexta* (see Section 4.8.3.1). Moreover, the nitrophorins share with the biliproteins the conserved long helix (actually two helical regions here, labeled h2 and h3 in **Figure 6**) at the extended C-terminal sequence that does not contribute to the barrel structure. Furthermore, two disulfide bridges are present at positions comparable to those in the biliproteins. The main difference between the two types of lipocalins refers to the ligand, which is heme in the ferric state in each of the nitrophorins (**Figure 7**). The heme is held in place

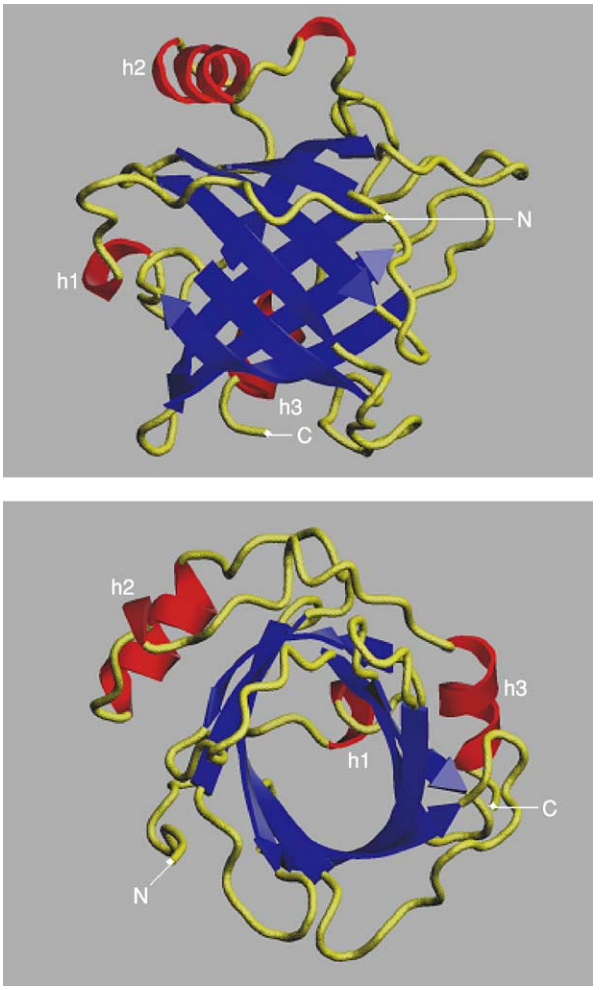


Figure 6 Crystal structure of nitrophorin 1 from *Rhodnius prolixus*. Ribbon drawings with eight β-strands (blue arrows) and several helices (red, labeled h1–h3). The N- and C-termini are marked with white dots. For the heme ligand, see **Figure 7**.

via 10 hydrophobic amino acid side chains and a histidine (His59) as the fifth iron ligand. This detail is reminiscent of the hemoglobins in which the heme is in the ferrous state, however. Even more, these two types of hemoproteins represent quite different protein structures: β-barrels in the nitrophorins and α-helical globular folds in the hemoglobins, demonstrating that they are evolutionarily unrelated. In the nitrophorins, the structure of the heme is very unusual as it is distorted and nonplanar. The carboxyl groups of the two propionate side chains of heme are hydrogen-bonded to the same lysine, and one of the propionates is also linked by an unusual carboxylate–carboxylate hydrogen bond to an aspartate side chain. Both heme propionate groups point to the solvent (**Figure 7**). The orientation of the heme in the nitrophorins is thus comparable to that of the cyclic but open-chain bilins in the biliproteins, as is

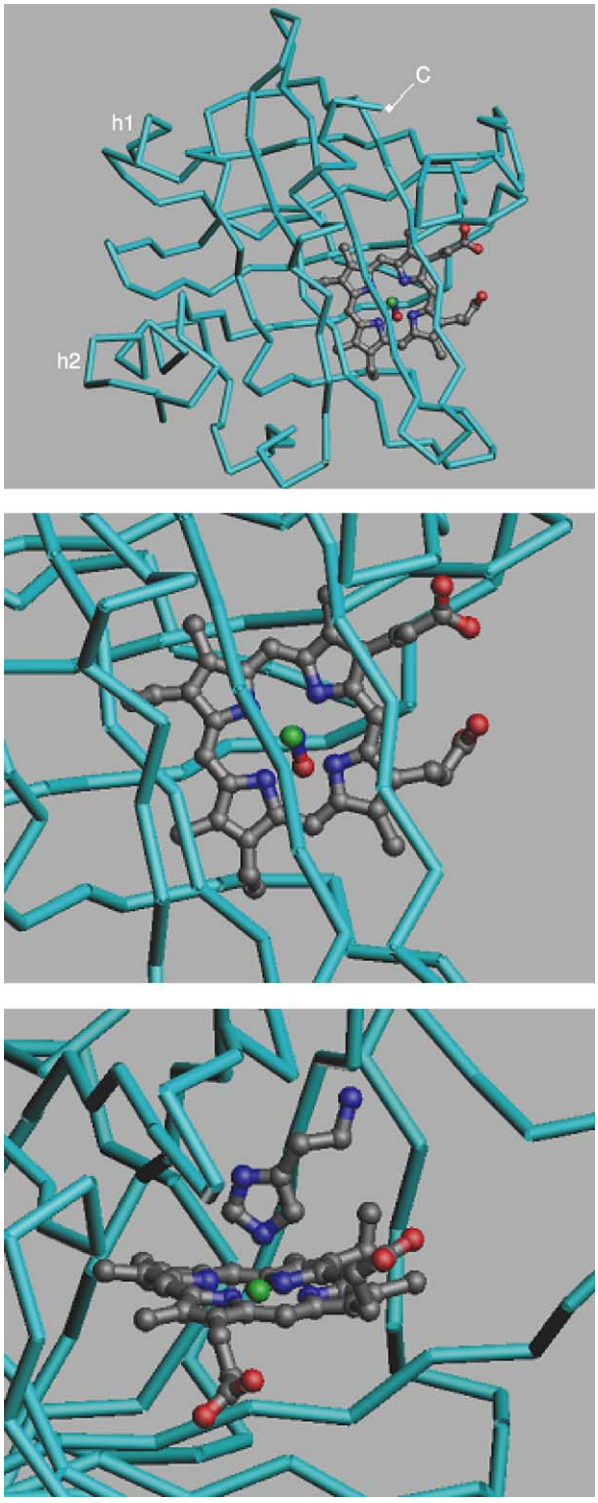


Figure 7 Crystal structure of nitrophorin 1 from *Rhodnius prolixus*. Cα-backbone structures (light blue) with the heme ligand in complex with nitric oxide (upper and middle panels) and histamine (lower panel), respectively. Heme is represented as stick-and-ball model with oxygen in red, nitrogen in blue, and iron in green. Two helices (h1 and h2) and the C-terminus are labeled. For further orientation, compare to **Figure 6**.

the location of the heme near the opening of the barrel. Though the nitrophorins show a higher degree of sequence identity with the biliproteins, their core lipocalin structures fit better to the vertebrate lipocalins. On the other hand, the positions of the disulfide bridges are identical in the two groups of insect lipocalins, while they are different from the mammalian lipocalins. Evolutionary relatedness in this case may hence be better reflected by the disulfide pattern than by identical positions of the core atoms of the β-barrel.

As can be seen from the crystal structure depicted in **Figure 7**, the nitrophorins are isolated as complexes of the heme iron with nitric oxide, which can be exchanged by histamine (Weichsel *et al.*, 1998). The heme is in the low-spin state in the complexes with nitric oxide and histamine (Soret maximum at 419 nm and 413 nm, respectively), and in the high-spin state when these ligands are replaced by water (Soret maximum at 404 nm). The dissociation constants (K_d values) of nitric oxide in complex with the various ferric-heme nitrophorins is in the 10×10^{-9} M to 1000×10^{-9} M range, depending on pH and the protein isoform. The NP2/NP3 pair shows higher affinities than the NP1/NP4 isoforms. In NP4, binding of nitric oxide changes the conformation of two loops at the β-barrel's opening so that nitric oxide gets enclosed by hydrophobic residues after several water molecules have been expelled from the cavity (Weichsel *et al.*, 2000). Keeping the heme iron in the oxidized (ferric) state is very important since the nitric oxide complex with ferrous iron is more stable by about six orders of magnitude, which means practically irreversible binding. The ferric state of heme is stabilized by the protein structure. The affinity for the heme ligands is modulated by pH in such a way that at the weakly acidic pH (~5) of the salivary gland the nitric oxide–heme complex is more stable than at the higher pH (~7.5) at the host's feeding site. This pH difference facilitates both the release of nitric oxide into the host's blood and the binding of histamine, due to its higher affinity, in place of nitric oxide. Histamine binds with a K_d of $\sim 20 \times 10^{-9}$ M at the host's pH, which is a ~100-fold higher affinity compared to that of nitric oxide in the same environment. NP2, also known as prolixin-S, also exerts anticoagulant activity in addition to its nitric oxide and histamine binding properties. Though details are not known, NP2 interferes at an early step with the multilevel cascade of blood coagulation.

Very recently, two additional proteins most likely representing new nitrophorins have been identified in the salivary gland of *R. prolixus*, and named NP5 and NP6 (Moreira *et al.*, 2003). One of these new

proteins, NP5, has been partially characterized as a heme protein, according to its absorption spectrum, with an N-terminal sequence revealing high similarity to NP4. Furthermore, NP5 binds nitric oxide and shows a spectral shift of the Soret peak from 404 nm to 422 nm upon binding similar to that of the other nitrophorins. The other protein, NP6, is only tentatively described as a nitrophorin so far. Interestingly, the relative abundance of the nitrophorins in the salivary gland depends on the life cycle stage of this bug (see Section 4.8.4.5).

It may be of interest to note that the bedbug *Cimex lectularius*, a blood feeder like *R. prolixus* but a member of a different family of Hemiptera, also has a heme-based donor protein for nitric oxide (Valenzuela and Ribeiro, 1998). By this property, it may also be called a nitrophorin. However, its amino acid sequence is unrelated to those of the *Rhodnius* nitrophorins and shows no lipocalin signature. It will therefore not be discussed further here. These nonhomologous but functionally equivalent proteins have likely evolved independently to meet the common constraints of blood feeding.

4.8.4.3. Other Lipocalins from Blood-Feeding Insects

The anticoagulant activity of the nitrophorin NP2 from *R. prolixus* has already been mentioned above. Another protein supporting the action of the nitrophorins as donors of nitric oxide to the host has been isolated and cloned from the salivary gland of this bug (Francischetti *et al.*, 2000). This protein, called *Rhodnius prolixus* aggregation inhibitor 1 (RPAI-1) or salivary platelet aggregation inhibitor 1, comprises 155 amino acids in its mature form and has a molecular mass of ~19 kDa. Though its crystal structure is not known, it obviously represents also a lipocalin as suggested by its sequence homology with some other insect proteins believed to have a lipocalin structure. These are pallidipin and triabin (see below) from the bug *Triatoma pallidipennis,* which both interfere with blood coagulation (Noeske-Jungblut *et al.*, 1994, 1995). RPAI-1 has a nucleotide binding site specific for ADP (and other adenine nucleotides) that is known to potentiate platelet aggregation induced by agonists such as collagen (Francischetti *et al.*, 2002). RPAI-1 is suggested to be most efficient in scavenging small concentrations of ADP that would otherwise induce the formation of large platelet aggregates. The sequence of a related salivary platelet aggregation inhibitor 2 has been reported also from *R. prolixus* (Champagne *et al.*, 1996).

Quite recently, another protein has been detected in the saliva of *R. prolixus* that apparently supports the bug in taking a blood meal. The new protein binds biogenic amines that are known to promote platelet aggregation in concert with ADP (Andersen *et al.*, 2003). This biogenic amine-binding protein (ABP) binds norepinephrine with highest affinity (K_d ~25 nM), followed by serotonin (K_d ~100 nM) and epinephrine (K_d ~350 nM). Alignment of the amino acid sequence of ABP with those of the four nitrophorins revealed significant similarity, a little lower than that between the nitrophorins. The lipocalin nature of ABP is further stressed by four cysteines at comparable positions (**Figure 8**). ABP is not associated with heme, which is explained by the absence of the proximal histidine that acts as fifth ligand of the heme iron in the nitrophorins. Hence, ABP lacks the structural basis for nitric oxide binding.

Triabin from *Triatoma pallidipennis*, a blood-sucking bug like *R. prolixus*, is a 16 kDa protein consisting of 142 amino acids in the mature form that inhibits thrombin by forming an equimolar noncovalent complex with this key compound for blood coagulation (Noeske-Jungblut *et al.*, 1995). Its structure is remarkable (**Figure 9**), as it turned out to be an unusual lipocalin with an eight-stranded β-barrel in which, however, the strands B and C of the first sheet are exchanged resulting in an "up–up–down–down" topology instead of the strict antiparallel "up–down–up–down" strand orientation (Fuentes-Prior *et al.*, 1997). Triabin lacks the long C-terminal α-helix that terminates in a short β-strand in typical lipocalins, which accounts for its lower size. Instead short helices are found at both termini. Like the retinol-binding proteins but unlike the biliproteins and nitrophorins, triabin contains three disulfide bridges. A sequence aligment of triabin and nitrophorins, together with related proteins, is shown in **Figure 8**.

4.8.4.4. Lipocalins from Blood-Feeding Ticks

The common feeding biology of blood-feeding insects and ticks has resulted in the convergent evolution of proteins serving the same goal, which is to support blood feeding by counteracting the hosts' responses to the attack. These functions are obviously best performed on the basis of a common protein structure, which turned out to be the conserved lipocalin fold. The protein structures as well as their ligands have been studied in two ticks, *Rhipicephalus appendiculatus*, which prefers cattle, and *Dermacentor reticulatus*, feeding on rodents.

Three histamine-binding proteins (HBP1 to HBP3) were identified in salivary glands of *R. appendiculatus* and cloned, and the crystal structure of one of them was solved (review: Paesen *et al.*, 2000).

Figure 8 Aligment of the nitrophorins 1 and 2 (NP1, NP2) and biogenic amine-binding protein (ABP) from *Rhodnius prolixus*, triabin (TRI) from *Triatoma pallidipennis* and Bla g 4 (Bg 4) from *Blattella germanica*. N-terminal secretion signals are boxed in gray. Residues typically conserved in lipocalins are boxed in black. In NP1 and NP2, the heme-binding His59 are also boxed in black. Asterisks indicate identical residues, double and single points more or less conserved substitutions.

Remarkably, these proteins are sex-specific, as two forms, HBP1 and HBP2, are found only in females, while HBP3 is restricted to males. Moreover, while the female-specific proteins are present only in adult ticks and produced only during the early phase of the single feeding period of females, the male-specific form is not restricted to the adult stage but occurs also in nymphs and larvae, and is synthesized over the entire feeding period involving several attacks of a host. These differences suggest that the three forms of HBPs are adapted to the sex-specific feeding biology of this tick species and may serve different needs. According to alignment studies of the amino acid sequences, the two female-specific isoforms, HBP1 and HBP2, are more closely related to each other (66% identity) than to the male protein, HBP3 (32% and 39% identity, respectively). The HBP sequences are unrelated to those of the nitrophorins and other lipocalins. The three structurally conserved regions of the lipocalins are also found in the HBPs, though as substantially modified motifs not readily recognized in the sequences. The molecular masses, based on the sequences without the N-terminal signals, are 19.5 kDa for both HBP1 and HBP2, and 21 kDa for HBP3. There are more sex-specific differences between these proteins: the male protein HBP3 is glycosylated and secreted as a dimer apparently linked by a disulfide bond, while HBP1 and HBP2 from females are monomeric and not posttranscriptionally modified. All HBPs, studied as recombinant proteins, bind histamine with high specificity but different affinities that are not related to sex, however. HBP1 binds histamine with a K_d of 18×10^{-9} M, while the values for HBP2 and HBP3 are in the $1–2 \times 10^{-9}$ M range.

The crystal structure of HBP2 has been reported (Paesen *et al.*, 1999). The overall structure matches the lipocalin fold with eight-stranded antiparallel sheets arranged to a fairly spherical β-barrel (**Figure 10**). This basic fold is modified by several structural details that are unique to this protein. Nevertheless, the structure of HBP2 is most similar to that of the BBP from the butterfly, *P. brassicae* (see Section 4.8.3.1.1). The N-terminal sequence of HBP2, comprising two helical structures, is extended and attached to the β-barrel via hydrogen bonds. The second helix, labeled h1 in **Figure 10**, occludes the open end of the barrel. The long C-terminal region comprises another helix, labeled h2, which is fixed to the barrel via two disulfide bridges. The most characteristic feature of HBP2 refers to the barrel cavity, which offers two distinct binding sites for two molecules of histamine (**Figure 10**), whereas a single ligand is usually present in classical lipocalins. The two sites are located at opposite ends of the cavity, which is separated into two parts due to the side chains of a tyrosine, a glutamate, and a tryptophane that are conserved in all three HBPs. The two histamine ligands are bound with different affinities and exhibit different orientations at their sites. At the high-affinity site, the histamine is perpendicular to the long axis of the β-barrel, while the other ligand at the low-affinity site is oriented parallel to that axis (**Figure 10**). The histamine at the high-affinity site of HBP2 is in a position comparable to that of heme in the

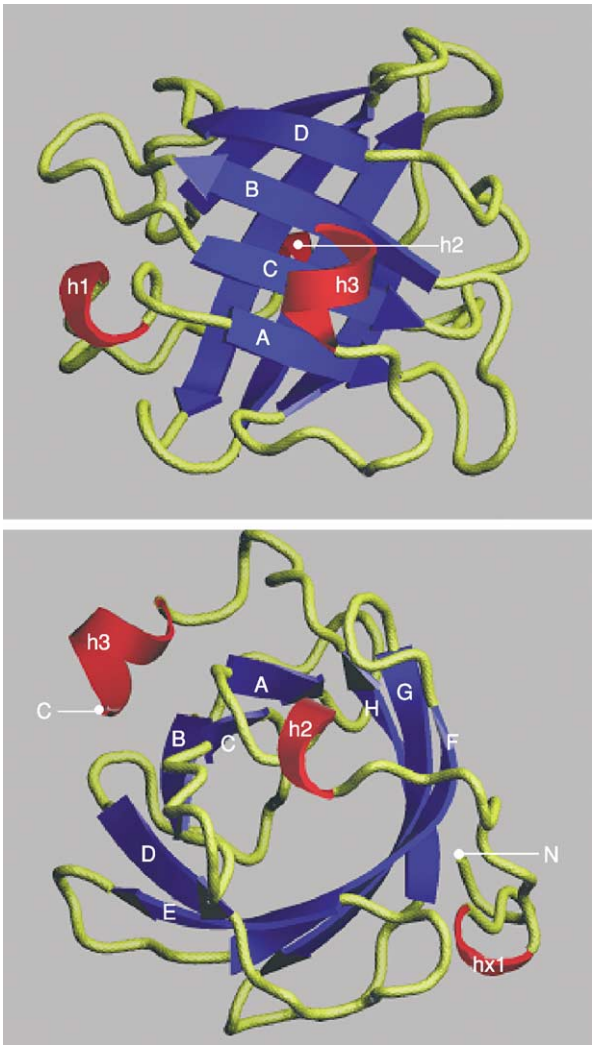


Figure 9 Crystal structure of triabin from *Triatoma pallidipennis*. Ribbon drawings with eight β-strands (blue arrows, labeled A–H) and several helices (red, labeled h1–h3). The N- and C-termini are marked with white dots. Note the unusual topology of the strands A–C–B–D.

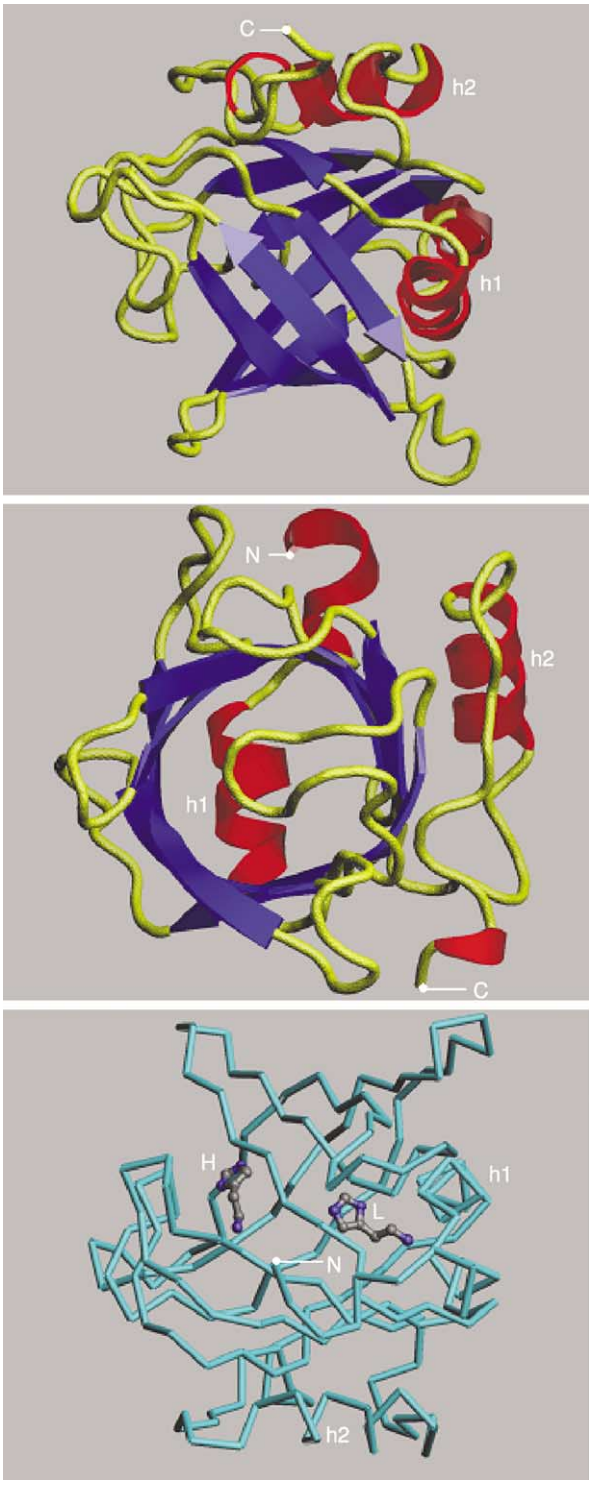


Figure 10 Crystal structure of the histamine-binding protein HBP2 from *Rhipicephalus appendiculatus*. Upper and middle panels: ribbon drawings with eight β-strands (blue arrows) and several helical regions (red, the major ones labeled h1 and h2). Lower panel: Cα-backbone structure (light blue) with two histamines bound at the high (H) and low (L) affinity sites. The N- and C-termini are marked with white dots.

nitrophorins; the low-affinity site is buried in the protein. How the histamines access the two different sites is not clear. To accommodate the hydrophilic histamine, which is positively charged in a physiological environment, both binding sites in the cavity are lined with a number of negatively charged and therefore polar amino acid residues. This is in contrast to typical lipocalins, which usually harbor hydrophobic ligands. In addition to the acidic residues, aromatic side chains contribute to the binding of histamine at the high-affinity site.

An interesting variant of a HBP has been recently found in the salivary gland of another tick, *Dermacentor reticulatus* (Sangamnatdej *et al.*, 2002). cDNA

cloning provided a sequence of 192 amino acids for the mature protein, corresponding to a molecular mass of ~22 kDa. The expressed protein exists as a dimer. It is glycosylated, and four putative glycosylation sites are derived from the sequence, which is 36–40% identical with those of the HBPs from *R. appendiculatus*, described above. Binding experiments with histamine revealed two specific binding sites in the recombinant protein from *D. reticulatus*. This is in accordance with the HBPs from the other tick, *R. appendiculatus*. The histamine at one of the sites can be competed by serotonin, which binds with high affinity (K_d of 6×10^{-10} M) to this single site. Thus, this salivary protein is a serotonin-histamine-binding protein (SHBP) with separate binding sites for the two distinct ligands. Though SHBP has not been crystallized, modeling of its sequence into the structure of the HBP2 revealed that the high-affinity site is well conserved, while the low-affinity site is enlarged mainly due to the substitution of an aspartate by glycine. This more spaceous site obviously serves as the binding site for serotonin that is larger than histamine.

With respect to the dependence of the synthesis of the HBPs on sex and sex-specific feeding behavior in *R. appendiculatus* (see above), it is of interest to note that the corresponding SHBP of *D. reticulatus* is expressed in both sexes. Both types of proteins are produced at increased rates over the feeding period though the two ticks differ in their feeding biology.

4.8.4.5. Developmental Expression of Nitrophorins in *Rhodnius prolixus*

The four well-studied nitrophorins have been isolated from the salivary glands of the adults of *Rhodnius prolixus*, where they are stored at different concentrations with NP1 as the dominant isoform, followed by NP2 and NP3, and NP4. This isoform pattern is not constant during the insect's life cycle, according to recent studies (Moreira *et al.*, 2003). In the first instar, NP2 represents the only major nitrophorin that is accompanied by two recently discovered nitrophorins, named NP5 and NP6, which have been only partially studied so far. At each instar, one additional nitrophorin is acquired, which is NP4 in the second instar and NP1 in the third instar. NP3 appears in the fifth instar and accumulates up to the adult stage. During larval development, the amounts of NP5 and NP6 decrease steadily; they are practically absent in the adult insect. In conclusion, the isoform pattern of the nitrophorins in the salivary gland of *R. prolixus* is development-specific. It is interesting to note that the nitrophorin form appearing earliest is NP2, the only protein of this group with anticoagulant activity in addition to its capability to reversibly bind nitric oxide and histamine. So, this nitrophorin variant provides the first instar nymphs with the full spectrum of activities that are relevant to blood feeding.

It should be mentioned that *R. prolixus* synthesizes in the fat body another heme protein that is released into the hemolymph and taken up by the developing oocytes. This protein with a molecular mass of ~15 kDa has a high content of α-helices. Therefore, this *Rhodnius* heme-binding protein is not a lipocalin. This is further stressed by the lack of any sequence similarity to the nitrophorins (Oliveira *et al.*, 1995; Paiva-Silva *et al.*, 2002).

4.8.4.6. Functions of Nitrophorins and Related Proteins

The main roles of the nitrophorins and functionally related proteins have been described in the context of their crystal structures, which allowed a molecular understanding of their function (see Section 4.8.4.2). So, this paragraph mainly summarizes the functional architectures of proteins adapted to a specific feeding biology. A study of these proteins has revealed a sophisticated molecular concept that enables insects and ticks to feed on vertebrate blood. The overall strategy is to prevent the host's response to the feeders' attack in several independent ways that act together in a concerted action.

In blood-sucking insects, the nitrophorins act as heme-based stable stores of nitric oxide, which is released into the blood after dilution and as a consequence of the higher pH at the feeding site (Nussenzveig *et al.*, 1995). The nitric oxide is distributed rapidly into the neighboring cells where it binds to the heme of soluble guanylyl cyclase for their activation that finally results in smooth muscle relaxation and vasodilatation. In exchange for nitric oxide, the nitrophorins bind histamine with high affinity that is produced by the host's mast cells to induce inflammation, immune response, and wound healing. All four nitrophorins studied in detail (NP1 to NP4) exert both activities as nitric oxide donors and as histamine scavengers. Nitrophorin NP2 exhibits anticoagulation activity in addition (Ribeiro *et al.*, 1995). Several other proteins of salivary glands reduce platelet aggregation by binding either ADP (RPAI-1), biogenic amines (ABP), or thrombin (triabin), which all promote blood coagulation. In conclusion, blood-sucking insects have developed a most remarkable battery of lipocalins targeted at diverse steps of the complex blood coagulation cascade in order to maintain the host's blood flow at the feeding site.

Ticks have developed a similar strategy to support their blood-feeding by using the same versatile protein fold, that of the lipocalins, to provide binding sites for histamine, but without employing heme, in contrast to the molecular concept of the nitrophorins of bugs. The tick proteins do not transport nitric oxide because they lack heme. The binding of serotonin in addition to histamine in the protein (SHBP) from *D. reticulatus* is not surprising since biogenic amines in general are established potentiators of blood coagulation and the anti-inflammation response, which are also targeted by the biogenic ABP from the bug *R. prolixus*.

4.8.5. Lipocalins with Unknown Ligands

4.8.5.1. Lipocalins Putatively Related to Development

As described in the foregoing, lipocalins are typically secreted, hence extracellular, proteins found in fluids of insects, like hemolymph (e.g., biliproteins) and saliva (e.g., nitrophorins), although they are also frequently stored as holoproteins within the cells, where they are synthesized (e.g., in the epidermis or fat body) or taken up (e.g., into the eggs). All these lipocalins are typified as carriers of specific ligands that are known, or at least presumed, to serve some metabolic roles in the insects. Well-studied examples of ligands are porphyrin products, nitric oxide, histamine, and biogenic amines. However, a number of proteins have been isolated or cloned from cDNA that were identified as lipocalins on the basis of sequence similarities and the presence of typical motifs in their secondary structures without knowing any real or potential ligands. It is possible that not all lipocalins carry a ligand but may function by, for example, docking to some other proteins as members of a signaling pathway. Under this view, it is interesting to note that several lipocalins without known ligands have been found to be associated with embryonic and postembryonic phases of development, frequently linked to that of the nervous system.

4.8.5.1.1. Lazarillo from *Schistocerca americana* Studies of the development of the nervous system uncovered a most unusual form of a lipocalin, as it is not free but covalently bound to the outer cell surface. There is no other eukaryotic lipocalin firmly anchored to a membrane (bacterial lipocalins are bound to the outer membrane). Playing a role as a guide to developing neurons, this protein was named Lazarillo after "Lazarillo de Tomes, a crafty boy who guided a blind man," as explained by Ganfornina *et al.* (1995), who discovered this lipocalin by the use of a monoclonal antibody in embryos of the grasshopper *Schistocerca americana*. A recent overview of Lazarillo has been published by Sánchez *et al.* (2000a).

Lazarillo is highly glycosylated, thus behaving like a ~45 kDa protein as affinity-isolated with the monoclonal antibody used to detect this lipocalin. From its cDNA, a molecular mass of ~20 kDa was predicted for the mature protein and seven potential *N*-glycosylation sites, however. The amino acid sequence of Lazarillo is further characterized by hydrophobic sequences at both ends. The N-terminal sequence likely represents a secretion signal, and the C-terminal sequence serves as an anchor for the cell surface attachment via a glycosylphosphatidylinositol (GPI) linkage. As can be seen from **Figure 11**, the Lazarillo sequence shows four cysteine residues which are most likely oxidized to form disulfide bridges corresponding to the two conserved disulfide bridges in other lipocalins. Moreover, the structurally conserved regions (SCRs) that characterize members of the lipocalin superfamily of proteins (Flower *et al.*, 1993) are well conserved in Lazarillo. At the amino acid sequence level, Lazarillo is about 30% identical with the other lipocalins, a value comparable to other members of this superfamily. A more detailed comparison places Lazarillo in the neighborhood of the biliproteins from *P. brassicae* and *M. sexta*, together with various forms of apolipoprotein D from mammals. Based on its relationship to the biliproteins, a homology model for Lazarillo was built using the atomic coordinates of insecticyanin from *M. sexta* (Sánchez *et al.*, 2000a). This model (**Figure 12**) shows a close match of the two structures. Remarkably, six of the seven potential glycosylation sites are located on the same side of the protein and close to the C-terminal GPI attachment site, thus resulting in a very polarized glycoprotein. The modeled cavity of Lazarillo also compares well with that of the biliprotein with respect to the distribution of hydrophobic and polar sites. It should be clearly stated at this point that no natural or artificial ligand of Lazarillo is known, though the binding cavity could accommodate a tetrapyrrole, for example.

4.8.5.1.2. Lazarillo-like lipocalins from *Drosophila melanogaster* The sequencing of the *Drosophila* genome provided two novel sequences with similarities to Lazarillo. These novel lipocalins were called *Drosophila* neural Lazarillo (DNLaz) and *Drosophila* glial Lazarillo (DGLaz), respectively (Sánchez *et al.*, 2000b). As deduced from their sequences, DNLaz is an acidic protein (pI 4.3), while DGLaz

```
GAL   ----------MLRMVLFAFVAAASASAVHEGKCPDFKPVDNFNLTAYQGVWYEISKTPND 50
BOM   ----------MLRLVLLTLLAAATAEVIHEGTCPELKPVNNFNLTAYQGIWYEISKFPNE 50
DNL   MNHHSSSHLLLLLISVVFGAVWVAHAQVPFPGKCPDVKLLDTFDAEAYMGVWYEYAAYPFA 60
DGL   MMSGQPLGSRVWLLSGVLLVTSAGTDAYGFGRCPNYPSMPKFNMSRVLGHWYEVERSFYL 60
HYP   ----------MWRLYLLKFLATASAQIPSLGWCPDFQAMANFNMDRFLGTWYEAERYFTV 50
LAZ   -----MIRRGLLSVTAALVLLSVSCSAQETMGCADRTAINDFNATLYMGKWYEYAKMGSM 55
                    :         :   .   .      *.:    :   *:      *  ***

GAL   -AEKNGKCGQAEYKLEG---------EVVKVKNSHV--VDGVQKYVEGTAKFAEDANKSA 98
BOM   -SEKNGKCSSAEYKLEG---------DVVKVKNVHI--IDGVKKYIEGTAKLTDDANKAA 98
DNL   -FEIGKKCIYANYSLIDN--------STVSVVNAAINRFTGQPSNVTGQAKVLGPG---- 107
DGL   -PEIASGCTTFQFEPYNKGEQSKFSNSKLAVAIKNINRITGNPNVNIGYATPENSRSSIM 119
HYP   -SELGTRCVTTHYTATPE---------GRILITNEITNSITGFKRLMEGHLQMVGREGEGR 101
LAZ   PYEEGGVCVTAEYSMSSN---------NITVVNSMKDNTTHEVNTTTGWAEFASELHTDG 106
        *    *   .:                : :              *

GAL   KLLVTLTYGAVNR-----ESPLNVIATDYQNYAIAYTCKYDEKSKSHNDSIWILSRAKKL 153
BOM   KLTVTFKFGEISR-----DGSVQVLATDYNNYAIAYNCKYDDKKKSHQVFVWILSRNKKL 153
DNL   QLAVAFYPTQPLT-----KANYLVLGTDYESYAVVYSCTS-VTPLANFKIVWILTRQREP 161
DGL   DFKFTTRFPDVIARLLPGSGKYQVLYTDYENFAILWSCGS-IGSLGHSDQIWILGRDRDF 178
HYP   VLVKYSSLPLPYD------FEYSILDTDYDTYAVMWACSG--IGPVHTQNTWLLSRERLP 153
LAZ   KLSVHFPNSPSVG-------NYWILSTDYDNYSIVWSCVKRPDSAASTEISWILLRSRNS 159
        :                      :: ***:.::: : *              *:* * :

GAL   EGDAKTAVDNYLKEH----AKEIDASK--LVQTDFSEEACKFT----------STSAVTE 197
BOM   EGDAKTAVDNFIKEH----SKEIDSSK--LVHTDFSEEACKFT----------ISSVITE 197
DNL   SAEAVDAARKILEDNDVSQAFLIDTVQ--KNCPRLDGNGTGLTGEDGLDVDDFVSTTVPN 219
DGL   EVDIRSKVYDVLKRLS------LDPER--LIISKNKQCPEAL------------------ 212
HYP   SMSVMQNAYAVLDKFKISRTFFQKTNQ--ADCQFYQHRLPQMN----------------- 194
LAZ   SNMTLERVEDELKNLQLDLNKYTKTEQSAKYCAGAEHVVGAMLS---------VAIASLF 210
      .      .   :.          .. :        .     :

GAL   PQTKKQ 203
BOM   HGKH-- 201
DNL   AIEKA- 224
DGL   ------
HYP   ------
LAZ   ALLH-- 214
```

Figure 11 Alignment of lipocalins putatively related to development. The proteins are gallerin (GAL) from *Galleria mellonella*, bombyrin (BOM) from *Bombyx mori*, neural and glial Lazarillo-like proteins (DNL and DGL, respectively) from *Drosophila melanogaster*, hyphantrin (HYP) from *Hyphantria cunea*, and Lazarillo (LAZ) from *Schistocerca americana*. N-terminal secretion signals are boxed in gray. Residues typically conserved in lipocalins are boxed in black. Stars indicate identical residues, double and single points more or less conserved substitutions.

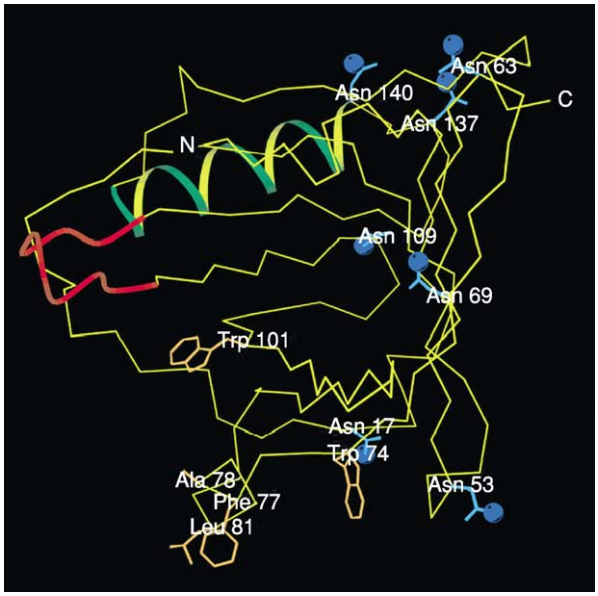


Figure 12 Homology model of Lazarillo from *Schistocerca americana* using the atomic coordinates of insecticyanin. The C-terminal helix (green) and the helical region with hydrophobic side chains (orange) might be involved in protein–protein interactions. The seven Asn glycosylation sites are labeled with blue balls. (Courtesy of Dr. Diego Sánchez.)

is basic (pI 8.6). Both proteins are ~21 kDa in size, exhibit the four conserved cysteines of lipocalins and represent secretory products according to their N-terminal signal sequences. Both may also be glycosylated; there are four potential sites in DNLaz, but only a single one in DGLaz. The two *Drosophila* proteins are different from other lipocalins and also from each other due to the presence of extra sequences (Figure 11). DGLaz has two of them of different lengths in the middle part of the coding region, which are located in two loops at the bottom of the lipocalin β-barrel. DNLaz shows an extra long C-terminal sequence, which, however, is hydrophilic in contrast to that of Lazarillo. No GPI tail is detected in the two Lazarillo-like proteins from *Drosophila*.

4.8.5.1.3. Gallerin from *Galleria mellonella* The sequence of gallerin was deduced from a brain cDNA library from the wax moth, *Galleria mellonella* (Filippov *et al.*, 1995). Gallerin comprises 203 amino acid residues including a signal sequence of 15 residues. According to homology searches, gallerin is related to BBP, the bilin-binding protein

from *P. brassicae*, and to the related insecticyanin from *M. sexta*. (see Section 4.8.3.1). Gallerin shows 40% identity with the two biliproteins, the four cysteines are at comparable positions, and also the three structurally conserved regions (SCRs) of typical lipocalins are present (**Figure 11**). Unfortunately, gallerin was not compared with the more known neuronal lipocalin Lazarillo, although mentioned in the respective publication (Filippov *et al*., 1995). The same publication also reports the sequence of another protein, called sericotropin, which was used to pick the gallerin sequence. According to a homology search performed by this author using BLAST, sericotropin is not a lipocalin as claimed by Filippov *et al*. (1995) since all the reference proteins are in fact of mainly helical structure. This correction is in agreement with the current view of the senior author of this publication (Sehnal, personal communication).

4.8.5.1.4. Bombyrin from *Bombyx mori* Bombyrin was identified in a brain cDNA library from silkworm, *Bombyx mori* (Sakai *et al*., 2001). The sequence of bombyrin was assigned to a lipocalin-type protein based on typical signatures like the four cysteines and the structurally conserved regions (**Figure 11**). A sequence alignment study, performed by this author, revealed that bombyrin is closely related to gallerin (68% identity), while no significant overall homologies (18–23% identity) were found to Lazarillo from the locust and the two Lazarillo-like proteins from the fruit fly.

4.8.5.1.5. Hyphantrin from *Hyphantria cunea* Another putatively developmental lipocalin, named hyphantrin, was cloned from the fall webworm *Hyphantria cunea* (Seo and Cheon, 2003). The sequence of hyphantrin was first assumed to represent a BBP though it is not associated with any blue pigment (Seo, personal communication). Only recently has hyphantrin been recognized as a protein with homology to the two Lazarillo-like proteins, to bombyrin, gallerin and Lazarillo (Seo, personal communication) (**Figure 11**). Hyphantrin is similarly related to various forms of apolipoprotein D and biliproteins (range of 25–30% identity, according to this author).

4.8.5.1.6. Expression and putative functions of developmental lipocalins By the use of the monoclonal antibody that led to the discovery of Lazarillo in *S. americana*, this protein was not ubiquitious in the nervous system of the embryo but only in subsets of neurons of the central nervous system (CNS), in the enteric nervous system, and in all sensory cells of the peripheral nervous system (Sánchez *et al*., 1995). Lazarillo is also expressed in a number of cells at the tips of the Malpighian tubules, in a group of nephrocytes, and in mesodermal cells. Lazarillo is expressed during the entire life cycle from the embryo up to the adult stage. In any case, the occurrence of this lipocalin is restricted to the cell surface, where it seems to be evenly distributed. What role Lazarillo could play in such diverse cell types is presently unclear. Studies of embryos of *S. americana* suggest that Lazarillo guides axon outgrowth during the development of the nervous system, as axon growth was no longer directed to make the correct contacts in the presence of an antibody against Lazarillo. Similar studies have been performed on the developing brain in the related species, *S. gregaria*, confirming the essential role of Lazarillo for directed axon pathfinding (Graf *et al*., 2000). Several hypotheses have been put forward on the mode of action of Lazarillo taking into account typical properties of lipocalins: to be designed as vehicles or receptors of mostly hydrophobic small ligands and to interact with other proteins as a means of intercellular communication (Sánchez *et al*., 2000a). In the latter speculative role, the large sugar moiety of Lazarillo could play an important role in cellular recognition.

In *Drosophila*, DNLaz and DGLaz are both expressed during embryogenesis in the CNS and in some nonneuronal tissues, but not in the peripheral and enteric nervous systems. The temporal patterns of expression of DNLaz and DGLaz are different during embryogenesis but similar thereafter with a low expression in the larvae and a high expression in pupae and adult flies (Sánchez *et al*., 2000b). This compares well with Lazarillo that is also mainly expressed in embryos and adults of the locust (Ganfornina *et al*., 1995). While a hypothesis on the role of Lazarillo in the locust has been formulated on experimental ground, the functions of the Lazarillo-like lipocalins in the fly are completely conjectural. The analysis of loss-of-function mutants, which is now under way, may provide some answer in the near future (Sánchez, personal communication).

According to Filippov *et al*. (1995), gallerin is expressed in the CNS of larvae, pupae, and adults of the wax moth. Gallerin expression was also found in the larval fat body, not in any other tissues. In contrast to the Lazarillo-type proteins and to gallerin, hyphantrin is found only in the pupal stage of *H. cunea* during a few days (days 4–6). Being mainly expressed in the epidermis, hyphantrin may not represent a typical neuronal lipocalin, although positive Northern plots were obtained with brain and fat body (Seo, personal communication).

4.8.5.2. Lipocalins Putatively Related to Reproduction

While the above developmental lipocalins, as they are tentatively grouped here, play only a speculative role in the (growing) insect, i.e., between cells, some other proteins, also belonging to the lipocalin superfamily, seem to be involved in the interactions between the (adult) insect and the environment. As in the developmental forms, ligands are also yet unknown for these lipocalins described below.

4.8.5.2.1. A lipocalin related to sexual behavior A lipocalin apparently involved in sexual behavior has been identified in the cockroach *Leucophaea maderae* (Korchi *et al.*, 1999). This protein, Lma-P22, is specific for adult males, where it is synthesized only in the epidermis of the tergites 2, 3, and 4 and becomes a constituent of the secretion of the dermal gland. The products of this gland are ingested by the female during courtship to act as an aphrodisiac. Partial sequences of the isolated Lma-P22 allowed the preparation of a cDNA encoding a protein with a molecular mass of 19.7 kDa. It consisted of 178 amino acids including an N-terminal signal sequence of 20 residues. Two putative sites for *N*-glycosylation were identified. Sequence alignments revealed an overall identity of 17–26% between Lma-P22 and several established and putative lipocalins including insecticyanin and gallerin. Moreover, the predicted secondary structure of Lma-P22 was in agreement with the lipocalin family. Whether Lma-P22 carries a ligand under physiological conditions is unknown, though a pheromone or another hydrophobic compound may qualify as candidates. Another protein with high homology to Lma-P22 has also been identified in the tergal secretion of male *L. maderae* (Cornette *et al.*, 2001).

4.8.5.2.2. A lipocalin related to oviposition A female-specific protein, referred to as Jf23, has been identified in the tarsi of the forelegs of the swallowtail butterfly, *Atrophaneura alcinous* (Tsuchihara *et al.*, 2000). Jf23 can be extracted from the tarsi as a ~23 kDa protein. The sequence of Jf23, as deduced from its cDNA, revealed a preprotein consisting of 203 amino acid residues, including a leader peptide of 15 amino acids. The calculated molecular mass of 20 296 Da of the Jf23 precursor is in good agreement with the estimated size of the isolated protein. Nearly half of the amino acids are hydrophobic. With a sequence identity of 38%, Jf23 is most closely related to BBP, the bilin-binding protein from *P. brassicae*, followed by insecticyanin, gallerin, and some mammalian lipocalins. A lipocalin structure of Jf23 is further supported by the presence of the conserved cysteines and of two structurally conserved regions at the corresponding locations. The presence of Jf23 in the female tarsi suggests that it may play a role in the transduction of plant chemical signals in the context of egg deposition. In fact, in electrophysiological experiments with the sensilla of female tarsi the response to compounds of a host plant was partially suppressed after pretreatment of the sensilla with antiserum against Jf23.

4.8.6. Fatty Acid-Binding Proteins

4.8.6.1. Variation of the Lipocalin Structure

Fatty acid-binding proteins (FABPs) are ubiquitous in animals. Members of this superfamily have been isolated from numerous vertebrates including humans, and from a variety of tissues such as intestine, liver, and heart muscle. While lipocalins are said to function typically as transporters of a variety of lipophilic molecules between cells, i.e., in the extracellular space (e.g., biliproteins in insect hemolymph or retinol-binding protein in vertebrate serum), FABPs are intracellular proteins and specialized as an almost single class of ligands, as their name indicates. FABPs are acidic proteins (pI ~5) and smaller than lipocalins, being composed of about 130 amino acid residues, corresponding to ~15 kDa proteins. On the other hand, FABPs share the overall structure of a β-barrel and a repeated +1 topology with the lipocalins. Because of some differences in the construction of the β-barrel, however, they are grouped as a separate superfamily within the calycin structural superfamily of proteins. Two other groups of intracellular proteins exhibit the same structural motif as the FABPs, which therefore all belong to the same superfamily. These proteins are the cellular retinol- and retinoic acid-binding proteins (CRBPs and CRABPs), respectively, reviewed by Banaszak *et al.* (1994) and Newcomer (1995). Only one representative has been identified in an insect to date (see Section 4.8.6.3). Because of their location in distinct compartments, the FABPs/CRBPs/CRABPs and the lipocalins are also referred to as intracellular and extracellular lipid-binding proteins, iLBPs and eLBPs, respectively. As a further difference between the two types of proteins, the orientation of the ligand is different. In the iLBPs, the polar group of the ligand (e.g., the carboxyl group of a fatty acid) is directed to the interior of the cavity, while it points to the barrel opening in the eLBPs (e.g., the carboxyl groups of

the tetrapyrroles). The crystal structure of the first representative of the FABPs was described in 1988 for the P2 myelin protein from the vertebrate peripheral nervous system (Jones *et al*., 1988). To date, the crystal structures of only two insects FABPs have been resolved, as described below. General overviews of the two major forms of β-barrel ligand-binding proteins, the lipocalins and the FABPs, have been written by Flower *et al*. (1993) and LaLonde *et al*. (1994), while the excellent review by Banaszak *et al*. (1994) is much more comprehensive and detailed.

Overall, the β-barrel of the FABPs is flatter and more clamlike in comparison to the more spherical lipocalin barrel. The major difference in the folding pattern is that the β-barrel of the FABPs is constructed not by eight, but by 10 antiparallel β-strands that are hydrogen-bonded to form two sheets of four and six strands, respectively. Like in the lipocalins, the strands forming the two sheets are in orthogonal orientation to one another. A characteristic feature of the general folding of the FABPs is the gap between the fourth and the fifth strand, assigned βD and βE. The first two strands are separated by a helix–turn–helix motif that is not present in the lipocalins. A scheme of the sequential arrangement of secondary structures of the FABPs is depicted in **Figure 13**. The helical parts of the N-terminal sequence occlude the β-barrel similarly to the first loop, the large Ω loop, in the lipocalins (see Section 4.8.2). FABPs lack the C-terminal helix of the lipocalins. The cavity of the β-barrel is mostly lined with hydrophobic side chains, but polar groups are also present. It is considerably more spacious than is required to bind a single molecule of fatty acid. In contrast to the situation in the lipocalins, it is unknown where or how the ligand enters the cavity of a FABP. It is speculated that the loading and unloading of the ligand may require dynamic changes in the protein folding that could be induced by the fatty acid itself. Nuclear magnetic resonance (NMR) studies at least indicate some flexibility in the region of the helices possibly serving as a portal for the ligand. The gap between the strands βD and βE is apparently not suitable for this role.

Despite the widespread occurrence of FABPs, only two representatives from insects have been purified, crystallized, and studied for their structure to date. These FABPs have been isolated from *M. sexta* and *S. gregaria*. Several other proteins have been identified as FABPs on the basis of sequence similarity with established FABPs, as described below (see Section 4.8.6.3). A number of FABPs from insects and mites are well-known allergens, as will be discussed separately (see Section 4.8.7). Those readers with particular interest in muscle-type FABPs from mammals and insects may consult the comparative article by Zanotti (1999). More comprehensive reviews of the FABPs from vertebrates are from Veerkamp and Maatman (1995) and Glatz and van der Vusse (1996). For more details and an overall view of lipid transport in insects, see **Chapter 4.6**.

4.8.6.2. Fatty Acid-Binding Proteins with Known Crystal Structure

4.8.6.2.1. Fatty acid-binding proteins from *Manduca sexta* Two FABPs, referred to as MFB1 and MFB2, have been isolated from the midgut cytosol of fifth instar larvae of *M. sexta*, as the most

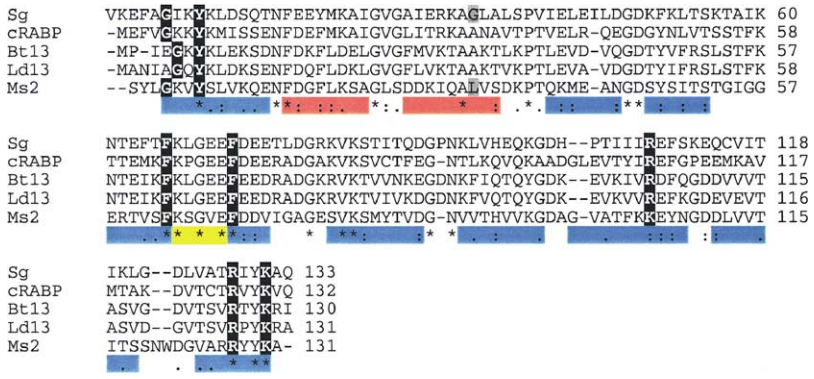


Figure 13 Aligment of fatty acid-/retinoic acid-binding proteins. The proteins are the fatty acid-binding proteins from *Schistocera gregaria* (Sg), *Manduca sexta* (Ms2: MFB2), Blo t 13 from *Blomia tropicalis* (Bt13), Lep d 13 (Ld13) from *Lepidoglyphus destructor*, and the cellular retinoic acid-binding protein (CRABP) from *M. sexta*. Black boxes show residues typically conserved in fatty acid-binding proteins. The residues boxed in gray (Gly in Sg and Leu in Ms2) are apparently responsible for the conformation of the bound fatty acid (see text for details). Below the sequences, the approximate positions of helices and β-strands are labeled as red and blue bars, respectively. The gap between the strands D and E is marked in yellow. Asterisks indicate identical residues, double and single points more or less conserved substitutions.

abundant proteins amounting to 2% and 12%, respectively, of the total soluble protein (Smith *et al.*, 1992). The predominant form, MFB2, has been crystallized and its structure determined as the first representative of an insect FABP (Benning *et al.*, 1992). The crystal structure of MFB2 (**Figure 14**) agrees with the folding motif of vertebrate FABPs. The β-barrel is made up of 10 up-and-down β-strands that are combined to form one sheet with four and one sheet with six strands. The N-terminal sequence is characterized by two helices that link strand A and strand B with a helix–turn–helix motif giving rise to the conserved clamlike overall structure of the barrel. The amino acid side chains defining the cavity for binding a fatty acid are hydrophobic, and most of them are uncharged. The aliphatic chain of the ligand is in contact with these hydrophobic sites, while the polar carboxyl group is hydrogen-bonded mainly to a conserved arginine (Arg127) and a tyrosine (Tyr129) located close to the C-termius that lacks the lipocalin-typical helix. The carboxyl group of the ligand is additionally bound to a sulfate which may be an artifact since the protein was crystallized from an ammonium sulfate solution. It is unknown whether a polar small molecule plays a role in the binding of a fatty acid in a physiological environment. While the position of the carboxyl group of the fatty acid in MFB2 is comparable to that in the vertebrate proteins, the conformation of the aliphatic chain is different. This is obviously due to a difference in the amino acid sequences which has an impact on size and shape of the cavity. In MFB2, a bulky leucine (Leu32) residue, corresponding to a glycine or alanine in the vertebrate proteins, forces the aliphatic tail of the fatty acid beyond C6 to an opposite orientation compared to the vertebrate proteins.

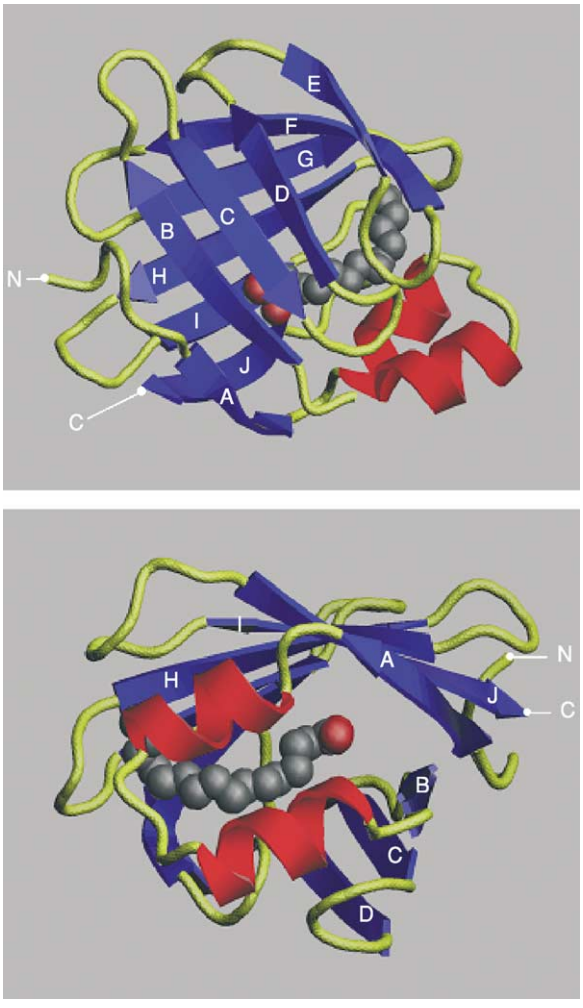


Figure 14 Crystal structure of the fatty acid-binding protein MFB2 from *Manduca sexta*. Ribbon drawings with ten β-strands (blue arrows, labeled A–J) and two helices (red) located between strands A and B (cf. **Figure 13**). The N- and C-termini are marked with white dots. The bound fatty acid is shown in space-filling representation with the carboxyl oxygens in red. Note the gap between strands D and E.

The minor form, MFB1, of the FABPs from the midgut of *M. sexta* has not been crystallized but its three-dimensional structure is thought to be very similar to that of MFB2. As derived from their cDNA sequences, MFB1 and MFB2 are of identical length comprising 131 amino acid residues, corresponding to molecular masses of 14 834 Da and 14 081 Da, respectively. As in most other FABPs, the N-termini are blocked also in the two *Manduca* proteins. The sequence identity between MFB1 and MFB2 is 56%, which is low for proteins of the same type and from the same tissue and species. The identities of the FABPs from the moth to those from various vertebrates are around 30% or lower. Remarkably, the two moth FABPs are not evenly distributed along the midgut. MFB1 is more concentrated in the anterior part but also present in the posterior part, while MFB2 is limited the posterior part of the midgut. Neither of the two proteins could be detected in extracts from fat body, muscles, and eggs. Like the lipocalins, MFB1 and MFB2 bind ligands in a 1 : 1 molar ratio. As isolated from the midgut, both proteins were loaded with a mixture of saturated and unsaturated fatty acids among which $C_{16:0}$ and $C_{18:2}$ fatty acids were predominant. The bound fatty acid can be easily exchanged in MFB1, but not in MFB2. Saturation binding experiments performed with MFB1 and radiolabeled oleic acid yielded an apparent dissociation constant of 14 μM for the exchange reaction. This is, of course, not a true measure for the affinity of the ligand to the apoprotein.

4.8.6.2.2. Fatty acid-binding protein from *Schistocerca gregaria* The first invertebrate FABP was isolated from *S. gregaria*. It was discovered in extracts from the flight muscles from adult locusts where it makes up about 18% of the total cytosolic protein (Haunerland and Chrisholm, 1990; Haunerland *et al*., 1993). As a soluble acidic protein (pI 5.2) with a molecular mass of ~15 kDa and fatty acids bound in a 1 : 1 molar ratio, this locust protein fits quite well the overall characteristics of other FABPs (see Section 4.8.6.1). By contrast, however, the N-terminus of the locust FABP was not blocked. The profile of the fatty acids, bound to the FABP from *S. gregaria* as isolated, was very similar to that of the two proteins from *M. sexta*. The locust FABP crystallized as a dimer in which the monomers are connected via the conserved helix–turn–helix motif that occludes the barrel like a lid (Haunerland *et al*., 1994). Overall, the protein shows the same conserved tertiary structure that is characteristic of members of this superfamily (**Figure 15**). These features are the highly conserved β-barrel with 10 antiparallel up-and-down β-strands with a gap between the βD and βE strands (**Figure 15**). The interior of the barrel contains 23 ordered water molecules, five of which are conserved between the proteins from *S. gregaria* and *M. sexta*. The locust FABP was obviously studied in the apoprotein form, as no fatty acid-ligand was found in the cavity suggesting its loss during the purification of the protein. However, the conserved C-terminal arginine and tyrosine residues that provide binding sites for the carboxyl group of the fatty acid in the moth MFB2, are also present in the locust FABP. While this suggests a similar mode of ligand binding, the conformation of the hydrocarbon tail of the ligand in the locust protein is expected to be opposite to that in MFB2, as the relevant bulky leucine (Leu32) in MFB2 corresponds to a small glycine (Gly34) in the locust FABP. This ligand conformation would be comparable to that in the vertebrate FABPs.

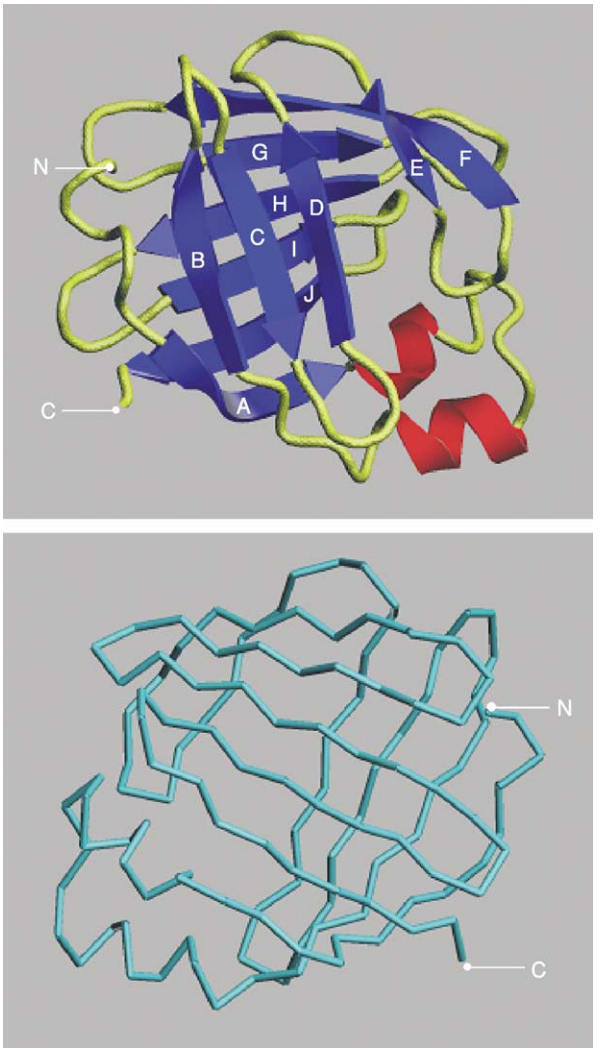


Figure 15 Crystal structure of the fatty acid-binding protein from *Schistocerca gregaria*. Upper panel: ribbon drawing showing ten β-strands (blue arrows, labeled A–J) and two helices (red) located between strands A and B (cf. **Figure 13**). The N- and C-termini are marked with white dots. Lower panel: Cα-backbone structure of the upper representation flipped around a vertical axis. The protein was isolated in the apo form.

The FABP from *S. gregaria*, which comprises 133 amino acids (**Figure 13**), shows unexpected low similarity to the two proteins from *M. sexta* and to those from vertebrates (about 30% sequence identity), while it is closer related (41% identity) to the FABP from human cardiac muscle (Price *et al*., 1992). Generally, FABPs from the same type of tissue but from different species turned out to be more similar to each other than those from different tissues in the same species. A comparison between the FABPs from insect flight muscles and midgut, for example, would be very interesting in this respect. Not surprisingly, however, the amino acid sequences of the FABPs from the flight muscles of the two locusts *S. gregaria* and *Locusta migratoria* are nearly identical (see Section 4.8.6.2.3).

The FABP from *S. gregaria* has also been studied at the gene level (Wu and Haunerland, 2001; Wu *et al*., 2001). In mammals, the FABP genes are characterized by three introns of varied length inserted at identical positions of the coding sequence. The locust gene shows only two introns, exactly corresponding to the first and third mammalian introns. This suggests a common evolutionary orgin of the FABPs. The promoter region contains a palindrome with a unique sequence of 19 bp that is essential for the induced expression of the gene by fatty acids. How this transcriptional effect is

mediated is unknown. It has been hypothesized that the conserved helix–turn–helix motif of a FABP could play a role in gene regulation, as it is reminiscent of a known transcription factor motif.

4.8.6.2.3. Fatty acid-binding protein from *Locusta migratoria* Another locust FABP is known from the flight muscles of *L. migratoria* (Maatman *et al.*, 1994). Cloning and heterologous expression of its cDNA has provided data on its sequence as well as binding of various ligands to the recombinant FABP. Although crystals of the protein were obtained, structural data have not yet been published. The *L. migratoria* FABP comprises 133 amino acids, corresponding to a molecular mass of 14 935 Da, and has a nonacetylated N-terminus, like in the other locust FABP. The sequence of the FABP from *L. migratoria* shows 98% identity with that of the *S. gregaria* homolog and 42% identity to that of the human muscle protein (Maatman *et al.*, 1994). According to the conserved features of its primary structure, the FABP from *L. migratoria* is expected to also posses the conserved tertiary structure known for this protein superfamily. Binding experiments, using the recombinant protein, provided a K_d for oleic acid of ~0.5 μM and confirmed the 1 : 1 stoichiometry of ligand binding. The FABP from *L. migratoria* is less acidic (pI 6.1) than that from *S. gregaria* (pI 5.2).

4.8.6.3. Fatty Acid-Binding Proteins with Unknown Crystal Structure

The sequence of a FABP from the moth *Heliothis zea*, the corn earworm, has been deposited in the databases (Heilmann, 1998). It has been identified on the basis of homology of its cDNA-derived sequence with that of established members of the FABP superfamily. Sequence alignment studies with insect FABPs, performed by this author using the Clustal W program, revealed a relatively high identity of the muscle protein from *H. zea* with the two midgut proteins from *M. sexta* (44% and 39% for MFB1 and MFB2, respectively), while the corresponding values for almost all other FABPs, including those from flight muscles of the two locusts, were in the 20% range. Another FABP has been reported (as a DNA sequence fragment) from the honeybee, *Apis mellifera*, where it is expressed only in larvae of workers, not of queens (Evans and Wheeler, 1999). A 14-kDa protein with a free N-terminus was purified from the flight muscles of the bug *Dipetalogaster maximus*, and identified as a heart-type FABP based on its N-terminal sequence (Cavagnari *et al.*, 2000). The sequencing of the genome of *Drosophila melanogaster* has revealed several genes and gene fragments obviously coding for FABPs. Similarly, one gene coding for a FABP of 141 amino acids and a calculated molecular mass of 15 071 Da has been cloned from the mosquito *Anopheles gambiae* (Favia *et al.*, 1996). There are other FABPs that are discussed subsequently as they have been identified as important allergens, a property also shared by a number of lipocalins (see Section 4.8.7).

The FABPs are not the only intracellular ligand- or lipid-binding proteins characterized by a β-barrel of ten antiparallel strands. The same folding-type is also found in a family of proteins called cellular retinol- and retinoic acid-binding proteins (CRBPs and CRABPs), respectively, as mentioned above (see Section 4.8.6.1). Their presence in vertebrates is well known, and the crystal structures of several forms have been described (reviews: Banaszak *et al.*, 1994; Newcomer, 1995). Only one insect member of the CRABP family has been identified to date. This protein isolated from *M. sexta* exhibits 71% identity with bovine and murine CRABP I and could therefore well be modeled into the crystal structures of the vertebrate homologs (Mansfield *et al.*, 1998). The sequence of this moth protein matches well the conserved FABP pattern (**Figure 13**). The natural ligand is unknown but the binding cavity could accommodate retinoic acid, for example.

4.8.6.4. Functions and Developmental Expression of Fatty Acid-Binding Proteins

As their name indicates, FABPs are intracellular transporters specialized for fatty acids as ligands. Though it seems evident that fatty acids require a carrier system that is mobile in the aqueous cytosol while offering a suitable site for the hydrophobic ligand, the implicated uptake and release of fatty acids by these binding proteins in a physiological environment has not yet been demonstrated clearly (Haunerland, 1997). As discussed above (see Section 4.8.6.1), the conserved folding structure of the FABPs does not suggest an opening of the cavity through which the fatty acid could access or exit the protein cavity. This is quite in contrast to the lipocalins with their calyx structure (see Section 4.8.2). While experiments *in vitro* with the *M. sexta* protein MFB1 demonstrated that a bound ligand can readily be exchanged, this was not possible in the case of the homolog MFB2 (Smith *et al.*, 1992). Hence, there may be significant differences between the various FABPs with respect to the ease and the rate of ligand exchange, as well as to the affinities for potential ligands. It is presumed that these different binding properties can be traced back to differences in the amino acid sequences translating

into differences in the ligand-binding pocket of the proteins. It would be of interest to compare the two FABPs from *M. sexta* in this respect and to relate their structural differences to the physiology of those parts of the midgut, where they are predominantly localized. The same considerations may be applied to the various tissue-specific isoforms of FABPs that are known from vertebrates.

In spite of the lack of understanding of the mechanisms of ligand uptake and release in the FABPs at the level of their crystal structure, these proteins are apparently tightly linked to lipid metabolism, specifically to the utilization of fatty acids for energy production (Van der Horst *et al.*, 1993). This seems to hold for insects as well as vertebrates. In the latter, the heart muscle is a typical tissue specialized on lipid oxidation to generate energy, and it contains a high concentration of tissue-specific FABP. In insects, migratory species similarly depend on the use of fatty acids to fuel long-distance flight. The flight muscles of locusts, for example, are an exceptionally rich source of FABP (Haunerland *et al.*, 1993). Under the conditions of extended flight, lipids are released from the fat body and transported *via* specialized hemolymph proteins to the flight muscles (see Section 4.8.8.2), where the fatty acids are taken up for β-oxidation (reviews: Haunerland, 1997; Ryan and Van der Horst, 2000) (see **Chapter 4.6**). The FABPs are supposed to be responsible for the intracellular transport of the fatty acids from the cell membrane to the mitochondria. It has been proposed that FABPs may also play a role in regulating the intracellular concentration of free fatty acids to prevent damage due to their detergent effect. In *S. gregaria*, FABP was localized in the cytosol as well as in the nucleus but not in mitochondria. The nuclear presence might suggest a genomic action, possibly on the expression of the FABP gene itself, as the transcription of the FABP gene can be enhanced by fatty acids. Thus, this transporter protein could be regulated according to physiological needs (Haunerland *et al.*, 1992a; Chen and Haunerland, 1994).

The FABP from *S. gregaria* flight muscle is present only in the adult. Only mature locusts are able to perform extended flight that requires the presence of this type of protein. The concentration of FABP starts to rise a few days after adult ecdysis reaching a plateau after about 10 days. This increase in FABP protein follows a transient but strong increase in its mRNA before it returns to a constant, low level (Haunerland *et al.*, 1992a; Haunerland, 1997). The high capacity to utilize fatty acids for energy production develops only several days after the molt to the adult stage, in parallel with the ability to perform extended flight. Overall, FABP synthesis appears to be part of the developmental program of an adult-specific muscle. This is also demonstrated by the application of inhibitors of metamorphosis that prevent the development of mature flight muscles and, in consequence, the appearance of the FABPs (Haunerland *et al.*, 1993).

4.8.7. Lipocalins and Fatty Acid-Binding Proteins as Allergens

Proteins with a β-barrel structure are of growing medical interest after several major respiratory allergens of mammalian and insect origin were identified as lipocalins and fatty acid-binding proteins, respectively (reviews: Mäntyjärvi *et al.*, 2000; Arlian, 2002). Though these calycins share the conserved protein architecture, conserved sequence regions, and some overall sequence homology, the structural determinants of their allergenicity have not yet been identified. It is evident, however, that it is not the β-barrel per se since by far not all lipocalins induce an allergic immune response. Well-studied examples of allergenic lipocalins from vertebrates are Bos d 2 from cow dander and Mus m 1 from mouse urine. The major sources for arthropod allergens from the superfamilies of lipocalins and fatty acid-binding proteins (FABPs), respectively, are species belonging to cockroaches, bugs, and mites.

Among the numerous allergens from cockroaches, Bla g 4 from *Blattella germanica* has been studied at the structural level and recognized as a typical lipocalin (Arruda *et al.*, 1995). The sequence of Bla g 4 comprises 182 amino acids with a calculated molecular mass of 20 904 Da, as deduced from its cDNA. The N-terminal 12 residues suggest that the encoded protein is secreted. The sequence contains one potential *N*-glycosylation site. The allergenic nature of the protein was confirmed using its recombinant form. A gene coding for Bla g 4 was also identified in *Periplaneta americana*, but its expression could not be demonstrated in this species, however. Searches for similarity to the Bla g 4 sequence revealed several established lipocalins from insects and vertebrates as related proteins. This cockroach allergen contains all three structurally conserved regions (SCRs) and the four conserved cysteines that define membership in the lipocalin family (**Figure 8**). Though the overall identity is low, as typically found among lipocalins, the sequence of Bla g 4 was successfully modeled into the crystal structure of the bilin-binding protein (BBP) from *P. brassicae*. The biological function of

Bla g 4 is unknown as is the binding of any ligand, although a role as a carrier of pheromones has been proposed.

Another group of insects with high allergenic potency are hematophagous bugs. The allergenic activity is stored in the salivary glands in many *Triatoma* species. In *T. protracta*, most of this activity is associated with a 19 kDa protein, named procalin, that comprises, as predicted from its cDNA, 169 amino acids with an N-terminal hydrophobic signal sequence (Paddock *et al.*, 2001). Unlike Bla g 4, procalin has no potential glycosylation site. An analysis of the procalin sequence predicts that procalin adopts a typical lipocalin fold with eight β-strands and a C-terminal helix. In fact, the closest homologs of procalin are the salivary platelet aggregation inhibitors from *R. prolixus* and triabin from *T. pallidipennis* with its unusual lipocalin fold (see Section 4.8.4.3). Moreover, the similarity of procalin to several established lipocalins, such as the moth biliproteins, retinol-binding protein, and β-lactoglobulin, has already been noted.

One of the allergens of the dust mite *Blomia tropicalis*, a most prevalent species in tropical areas, shows ~40% sequence identity with several FABPs from vertebrates and a flatworm. In fact, this allergen with the official name Blo t 13 represents a typical FABP with respect to size (130 amino acids; molecular mass of 14 800 Da) and binding specificity for fatty acids. Moreover, Blo t 13 is characterized by ten up-and-down β-strands, and its sequence could well be fitted into the conserved crystal structure of FABPs (Caraballo *et al.*, 1997; Puerta *et al.*, 1999). Further FABP allergens from mites are Lep d 13 from *Lepidoglyphus destructor* and Aca s 13 from *Acarus siro* (Eriksson *et al.*, 1999, 2001). The sequences of Blo t 13 and Lep d 13 are presented in **Figure 13** in comparison to those of other FABPs.

4.8.8. Ligand-Binding Proteins That Are Not β-Barrels

4.8.8.1. Odorant-Binding Proteins

Olfaction plays a major role in the lifes of insects and vertebrates (see **Chapter 3.15**). It is based on air-borne chemicals that could be pheromones (see **Chapter 3.14**), plant-derived compounds or other signaling substances. Volatile compounds are usually small and hydrophobic, requiring a carrier system that takes them from the cuticular surface of insects or from the nasal epithelium of vertebrates to the olfactory receptors. This carrier function is supposed to rely on a specialized set of proteins, called odorant-binding proteins (OBPs). They comprise the general odorant-binding proteins that occur in both sexes and have no clear ligand preferences and the pheromone(sex attractant)-binding proteins, which are male-specific and specialized in binding a defined ligand, the female pheromone. Many of these OBPs from insects and vertebrates have been purified, cloned, and studied for their functional properties (Pelosi and Maida, 1995; Field *et al.*, 2000). OBPs from insects and vertebrates share several features: they are small (15 to 20 kDa), acidic (pI ~5) and secreted in high concentrations. However, there is no sequence similarity between insect and vertebrate OBPs. This is reflected in their three-dimensional structures: the OBPs from vertebrates have been clearly identified as members of the lipocalin superfamily with all the features of β-barrel proteins, while those from insects are rich in α-helical structures and therefore not lipocalins. The protein folds of two OBPs from insects have recently been elucidated by NMR spectroscopy and X-ray diffraction, respectively: these are a hemolymph protein from the beetle *Tenebrio molitor* (Rothemund *et al.*, 1999) and a pheromone-binding protein in the complex with bombykol from *Bombyx mori* (Sandler *et al.*, 2000). Both studies revealed a new structural class of proteins defined by six α-helices forming a small hydrophobic core that provides a ligand-binding site. Three disulfide bridges are also conserved. Recently, an X-ray study of a chemosensory protein from the moth *Mamestra brassicae* demonstrated another novel fold consisting of six α-helices linked by α–α loops (Lartigue *et al.*, 2002). Evidence for a predominant helical structure has also been obtained for a chemosensory protein from the honeybee, studied by CD spectroscopy (Briand *et al.*, 2002). Furthermore, the genome of *Drosophila* has been searched for OBP genes; all identified 38 genes clearly featured a six-helical fold (Graham and Davies, 2002). It is remarkable from an evolutionary standpoint that insects and vertebrates developed completely different protein structures to achieve functionally equivalent tools for chemical communication.

4.8.8.2. Lipophorins

Hydrophobic molecules in general need amphiphilic transport systems operating in the extracellular space between the tissues as well as in the intracellular milieu (see **Chapter 4.6**). This aspect has already been discussed in the intracellular transport of fatty acids by a specialized set of binding proteins, the fatty acid-binding proteins (FABPs), which represent a variation of the β-barrel structure as it is found in the lipocalins (see Section 4.8.6.1). Insects performing extended flight like migratory

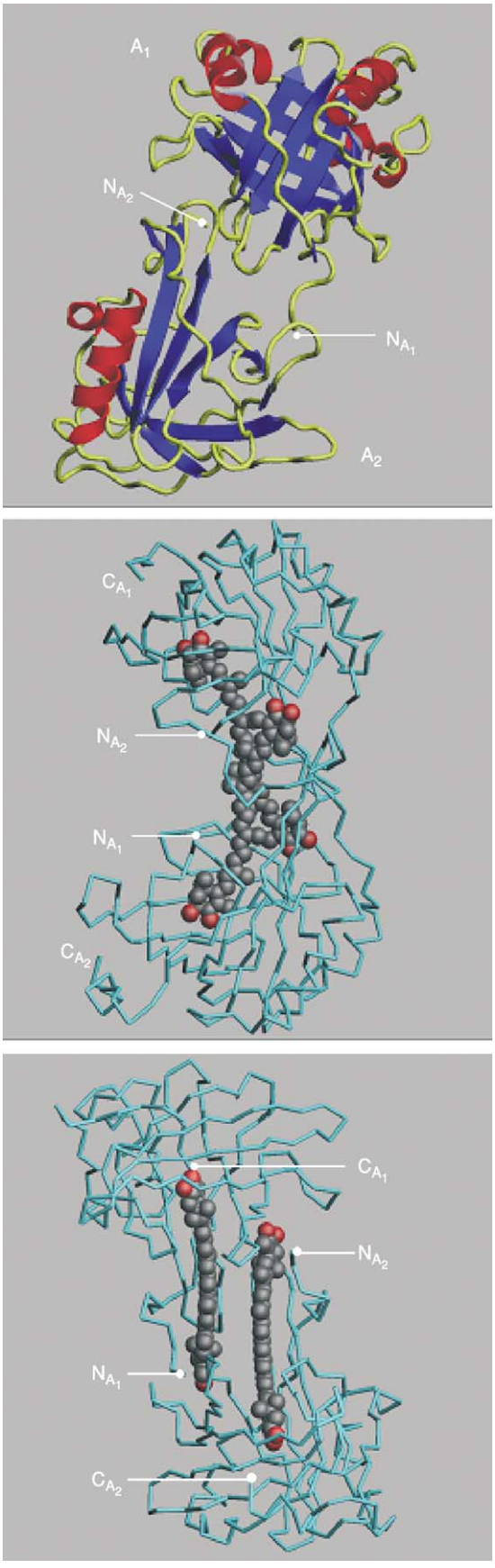
locusts and sphingid moths rely on the oxidation of lipids to produce energy. The lipids are stored in the fat body and activated to fuel flight by the conversion to diacylglycerols, representing their transport form. The transport of lipids to the cells, which finally β-oxidize the fatty acids, is performed by a special carrier system, called lipophorin. Lipophorin represents a large lipoprotein complex composed of two integral units of different size, apolipophorin I and apolipophorin II, with molecular masses of ~250 kDa and ~80 kDa, respectively. Lipophorin is charged with a variable amount of lipid depending on the intensity of lipid utilization by the insect and can also be associated with steroids, carotenoids, or other hydrophobic compounds. Another protein reversibly associates with the core lipophorin to stabilize this carrier complex and to allow more diacylglycerol to bind. This additional protein, apolipophorin III, with a molecular mass of ~19 kDa is highly abundant in the hemolymph as a lipid-free monomer, according to studies in *L. migratoria* and *M. sexta*. While the amino acid sequences of apolipophorin III are known from a number of insects, the only available crystal structure is that of the protein from *L. migratoria* (Breiter *et al.*, 1991). The locust apolipophorin III does not represent a lipocalin or a structural variant thereof like the fatty acid-binding proteins (FABPs). By contrast, it consists of five long α-helices connected by short loops. The helices show a clear amphiphilic arrangement of hydrophobic and hydrophilic amino acid residues, thus fitting the physicochemical requirements to interface the lipid and aqueous compartments. The sequence homology between the apolipophorins III from *L. migratoria* and *M. sexta* suggests a fairly conserved structure of this lipophorin partner. For more details on lipophorin in the overall context of lipid transport (see **Chapter 4.6**).

4.8.9. Crustacean Lipocalins with Established Crystal Structure

Lobster (*Homarus gammarus*) is of long-standing interest not only for gourmets but also for scientists interested in the secrets of the blue coloration of its shell turning into appetizing red upon cooking. The red color is due to astaxanthin (3,3′-dihydroxy-β,β-carotene-4,4′-dione), a carotenoid metabolically derived from β-carotene by oxidation at both cyclohexene end rings to produce an extended chromophore with a light absorption maximum in the range of 470–490 nm, depending on the solvent (review: Kayser, 1985). In lobster, astaxanthin is noncovalently bound to protein in a complex, called crustacyanin, with an absorption maximum of the carotenoid at ~630 nm, resulting in a blue protein. This huge bathochromic shift of more than 100 nm upon binding is due to a specific arrangement of the red carotenoid in the native protein complex (i.e., before cooking), as described below.

Crustacyanin or, more specifically, α-crustacyanin is a multimeric protein complex, which comprises eight heterodimeric units, called β-crustacyanins. Each of the β-crustacyanin dimers is composed of two types of monomers (the known monomers are $A_{1–3}$ and $C_{1–2}$) of similar size (~20 kDa) each with one bound molecule of astaxanthin. Thus, the holoprotein with a molecular mass of ~320 kDa contains 16 molecules of astaxanthin. The amino acid sequences of the 20 kDa subunit show homology to members of the lipocalin family (e.g., retinol-binding protein; biliproteins from *P. brassicae* and *M. sexta*; see Section 4.8.3.1), and the predicted tertiary structure is consistent with the antiparallel β-barrel fold. Moreover, two of the sequence regions and the two disulfide bridges that are conserved in the lipocalins are also present in the crustacyanin subunits. On the other side, the homology between subunits may be surprisingly low (for example, <40% between A_1 and A_2) for functionally identical proteins from the same species (Keen *et al.*, 1991).

The predicted lipocalin fold has been confirmed with crystals of the apo-C_1 subunit, which forms homodimers by packing the open ends of the two monomeric barrels face-to-face together (Gordon *et al.*, 2001). Comparable results have been obtained for the dimer of the apo A_1 subunit (Cianci *et al.*, 2001). The positions of the two astaxanthins were recently identified in a β-crustacyanin heterodimer that could be crystallized and its structure determined (Cianci *et al.*, 2002). As presented in **Figure 16**, the subunits A_1 and A_2 also dimerize in a face-to-face way and each subunit exhibits the typical lipocalin fold comparable to that in the previously studied apo-C_1 homodimer. Most surprisingly, however, the astaxanthin ligand of each subunit is not enclosed in the cavity formed by each barrel but extends into the other subunit by about half of its length. This unusual feature explains the tight binding of the chromophores in the dimeric β-crustacyanins in contrast to the labile complex in the monomeric state. In the dimer, the polyene chains of the two carotenoids are parallel to each other and close together. This position is fixed by the two central methyl groups (C20 and C20′) of the chain that are strongly bound to the protein by hydrophobic interaction. The cyclohexenone end rings are twisted to come into plane with the all-*trans* polyene chain, unlike in the free carotenoid.



This brings the double bonds of the end rings into full conjugation with the polyene chain resulting in an extension of the chromophore, and thus in a significant bathochromic shift of the absorption spectrum of the bound carotenoid. Further structural details, thought to enhance this effect, are the way of binding of the carbonyl groups of the end rings to the protein, the electronic interaction between the amino acids and the conjugated chain, and the formation of the finite α-crustacyanin aggregate (Cianci *et al.*, 2002).

Astaxanthin is typically encountered in crustaceans and other marine animals and provides the basis for similar camouflage colorations like in lobster. This red carotenoid is also found in insects from several orders where it most likely is also associated with protein (review: Kayser, 1985). However, an astaxanthin-based blue carotenoprotein has not yet been described from insects. The highly lipophilic carotenoids are presumed to be in general associated with proteins and, in fact, carotenoproteins have been reported from numerous insects, but none of these has been characterized with respect to the mode of ligand binding and protein architecture to date.

4.8.10. Lipocalin Engineering: An Insect Protein Takes the Lead

The preceding paragraphs have provided numerous examples of the evolutionary highly conserved core structure of the calycins as a structural superfamily, which comprises the lipocalins as the most diverse members (review: Flower *et al.*, 2000). The structural concept of a barrel constructed from an orthogonal arrangement of two sheets of antiparallel β-strands provides a remarkable rigid type of protein that is small, water-soluble, and resistant to heat and proteases. Its interior provides a binding site that is adapted in size, shape, and distribution of hydrophobic and polar amino acid residues to the specific binding of typically hydrophobic molecules. The lipocalin ligands enter the barrel at one end through a wide opening that is surrounded by four

Figure 16 Crystal structure of a β-crustacyanin dimer composed of the subunits A_1 and A_2 from lobster (*Homarus gammarus*). Upper panel: ribbon drawing with β-strands (blue arrows) and helices (red). Middle and lower panels: Cα-backbone structures in different orientations with one molecule of bound astaxanthin per monomer. The oxygens at the ionone end rings of the astaxanthin ligands are shown in red in the space-filling representations. The N- and C-termini are marked with white dots. The suffix A_1 and A_2, respectively, specifies the termini of the respective subunit.

loops connecting the strands. The amino acid residues in these loops play a major role in the recognition of the specific ligand. This is reflected by the variability in loop sequence and length in the lipocalins adapted to different ligands.

Based on these features, the lipocalins were early recognized as a potential scaffold to construct new versions of binding proteins for nonphysiological ligands. This concept for protein design has been successfully applied by Skerra and co-workers (reviews: Skerra, 2000; Weiss and Lowman, 2000) to the bilin-binding protein (BBP) from *P. brassicae*, which was one of the first lipocalins established by its crystal structure at high resolution (see Section 4.8.3.1.1). This insect protein was preferred to other members of this superfamily, such as the vertebrate retinol-binding protein, because of its more spacious binding cavity and the presence of only two, instead of three, disulfide cross-links. By mutating a number of amino acids in all four loops at the open end of the barrel, which were selected on the basis of the known crystal structure of BBP, variants of the wild-type lipocalin were created with affinity and selectivity for fluorescein and digoxigenin as new ligands, which bind with affinities in the nanomolar range. These engineered lipocalins were termed "anticalins," because they have similar molecular recognition properties as antibodies (Beste *et al.*, 1999; Skerra, 2001). The crystal structure of the fluorescein-binding anticalin, which was solved very recently (**Figure 17**), demonstrated that the loops and even the core of the β-barrel are tolerant to sequence changes and hence offer possibilities for the design of novel receptor proteins for various targets (Korndörfer *et al.*, 2003). Moreover, fusion proteins were generated between a BBP variant with a binding site for digoxigenin and a functional reporter enzyme, linked N- or C-terminally, or between two different engineered BBPs with affinity for fluorescein and digoxigenin, respectively. These fused lipocalins carried two functionalities and were hence called "duocalins" (Schlehuber and Skerra, 2001).

Furthermore, in earlier studies by the same group, a binding site for Zn(II) and other metal ions was introduced into the mammalian retinol-binding protein via three histidine side chains. The metal binding site was located at the outer surface of the β-barrel and bound Zn(II) in a 1 : 1 stoichiometry in the nanomolar range without affecting the binding of retinol in the interior of the engineered lipocalin (Skerra, 2000).

The potential of lipocalins to provide a scaffold for the design of numerous binding sites is comparable to that of antibodies while offering several

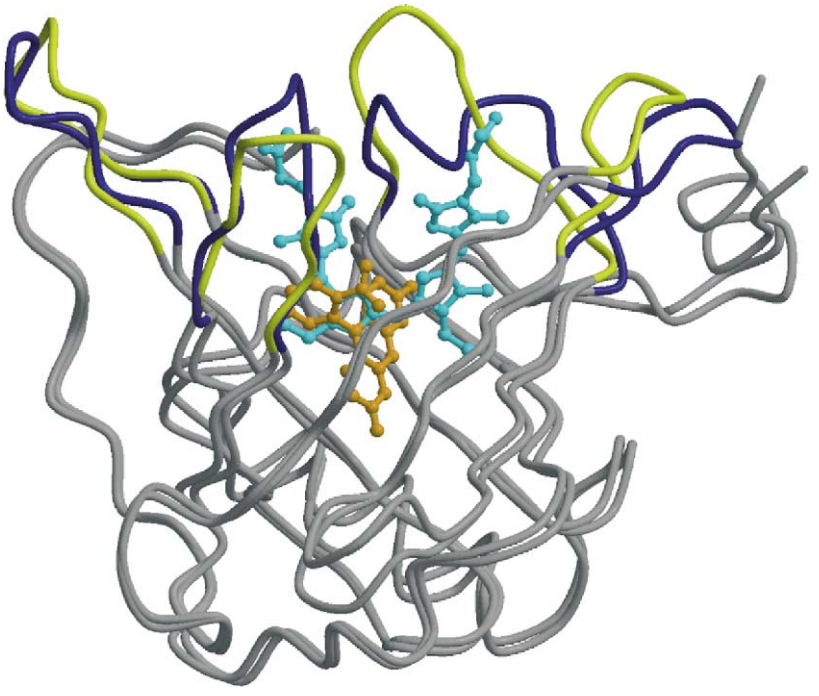


Figure 17 Crystal structure of engineered bilin-binding protein (BBP) from *Pieris brassicae* with high affinity for fluorescein. The structures of BBP (blue loop region) and the fluorescein-binding construct (yellow loop region) are superposed via the Cα-backbone structure of the β-barrel. The bound biliverdin IXγ (green) in BBP and fluorescein (brown) in the engineered protein are represented as stick-and-ball models. (Courtesy of Prof. A. Skerra.)

advantages (for details, see Skerra, 2001). Such "anticalins" may be used as new tools in medicine, biotechnology, bioanalytics, cosmetics, pest control, and other fields in the future. These novel approaches in the design of new proteins represent also a convincing example for the successful transformation of results from basic research on insects into a variety of potential applications that could not be foreseen at the beginning.

4.8.11. Lipocalins – One Fold, Many Roles: Summary and Outlook

The objective of the present overview of lipocalins and structurally related proteins, taken together as the calycins, was to demonstrate the functional versatility of an ancient motif of protein folding dating back to the roots of eukaryotes or even earlier periods of evolution (Ganfornina *et al*., 2000). Most remarkably, the core structure of the calycins, the β-barrel with antiparallel stands, has been highly conserved, while amino acid residues have been allowed to exchange to tailor the protein to the specific and, in most cases, high-affinity binding of small, hydrophobic ligands (Flower, 1995). The very low sequence homologies between members of this structural superfamily are a consequence of this evolutionary strategy.

Due to this wide variation of an ingenious theme, calycins got involved into many roles within cells and between cells. In most cases, however, the functions of the various types of calycins are far from being thoroughly understood, even in proteins characterized by a specific ligand. In lipocalins that lack a ligand, assigning a function may be even more speculative. This is not the case for the insect nitrophorins representing a group of highly specialized lipocalins with obviously well-defined tasks in the context of blood-feeding. The fatty acid-binding proteins are similarly specialized for a single chemical class of ligands; however, it is not clear yet whether their role is limited to the transfer of fatty acids from the cell membrane to the mitochondria, or whether they may also directly regulate gene expression and other vital cellular processes.

According to an increasing number of reports, lipocalins seem to serve not only specialized roles but are also involved in general cellular processes. While several lipocalins have long been implicated in various aspects of cell regulation (Flower, 1994), recent studies suggested that, for example, a lipocalin might trigger apoptosis in leukocytes (Yousefi and Simon, 2002). Results with the human tear lipocalin, which accepts a variety of ligands, indicate that this protein may be able to protect cells from oxidative damage by scavenging potentially harmful products (Lechner *et al*., 2001).

As mentioned at the beginning (see Section 4.8.1), a few lipocalins have also been identified in plants in the past few years. The first plant lipocalins were two enzymes of the xanthophyll cycle, violaxanthin de-epoxidase and zeaxanthin epoxidase (Bugos *et al*., 1998). These proteins are larger than the typical lipocalins with a β-barrel as a central core structure, as deduced from the sequences. Other proteins from *Arabidopsis thaliana* and wheat were found to show the typical lipocalin signature and homology to established members of this family, mainly to locust Lazarillo and vertebrate apolipoprotein D (Charron *et al*., 2002). These plant lipocalins are induced by heat-shock and cold acclimation. It may be of interest to look for similar effects in insects. Another protein from *A. thaliana* shows homology to insect biliproteins and vertebrate lipocalins (Hieber *et al*., 2000).

Before the identification of the two plant enzymes as lipocalins, the lipocalin-type of prostaglandin D synthase (the other type of prostaglandin D synthase, called the hematopoetic type, is unrelated to the lipocalin-type by sequence and protein structure), which is abundant in vertebrate cerebrospinal fluid, was the only lipocalin with enzymatic activity (Urade and Hayashi, 2000). *In vitro*, the lipocalin-type of prostaglandin D synthase also binds the nonsubstrate compounds biliverdin and bilirubin with high affinity; thyroid hormones and retinoids show markedly less affinity (Beuckmann *et al*., 1999). In conclusion, the lipocalin-type of prostaglandin D synthase may act like a typical, though nonspecific, lipocalin in addition to its role as an enzyme. The biological significance of this dual role is not known. Since the synthesis of prostaglandins has also been documented for insects (Büyükgüzel *et al*., 2002) (see **Chapter 4.9**), as study of the corresponding synthase(s) may be rewarding also from the lipocalin point of view.

Finally, it is worth mentioning apolipoprotein D, a classical mammalian lipocalin that is present in many tissues. This protein is unrelated to the apolipophorin family but shows close homology with the insect biliproteins insecticyanin and BBP. Moreover, the crystal structures of apolipoprotein D and the insect biliproteins are superposable and the distribution of hydrophobic and polar sites of the binding cavities is comparable among the three proteins (Peitsch and Boguski, 1990). Hence, apolipoprotein D has been also viewed as a mammalian bilin-binding protein. A detailed phylogenetic analysis of the lipocalin superfamily performed by Ganfornina *et al*. (2000) demonstrated a close

relationship between the arthropod lipocalins and the various forms of apolipoprotein D, clustering together in the same clade.

The insect biliproteins represent a kind of mystery due to their tight coupling to a strong heme pathway that is not understood from the relatively low requirement of heme products in insects compared to the hemoglobin-dependent vertebrates. Biliproteins are frequently referred to as the blue basis for camouflage coloration in insects (see Section 4.8.3.5). This role is certainly true for a number of insects, especially for larval stages. However, there is no clear correlation between the presence of biliproteins and their contribution to the visible coloration of a given insect. Moreover, the synthesis of open-chain bilins along the porphyrin pathway via heme is a costly undertaking that is unlikely to be repaid by a simple visible protection effect. Cryptic coloration might be achieved with much less effort by utilizing, for example, chlorophyll-derived products simply sequestered from the plant food. In conclusion, the primary roles of biliproteins in insects are likely still to be discovered.

Another most interesting property of lipocalins in general is binding to other proteins, in particular to receptors that may function in cellular signaling. Though cell surface receptors have been claimed for several lipocalins there is only little experimental proof for their presence and identity even in the case of the prototypic lipocalin retinol-binding protein (review: Flower, 2000). A well-studied example even at the crystal level, however, is the 2 : 1 complex of the retinol-binding protein with the tetrameric transthyretrin as it occurs in vertebrate blood. In insects, studies of the uptake of insecticyanin into developing oocytes of *M. sexta* provided preliminary data for a membrane receptor that is likely involved in uptake or internalization of this biliprotein (see Section 4.8.3.4). Another relation of lipocalins to membranes is seen in those proteins firmly associated with the cell surface like for Lazarillo that plays a still enigmatic role in developing neuronal networks (see Section 4.8.5.1.6). It is not known whether Lazarillo requires a specific ligand to be active. It is tempting to speculate about retinoic acid (or a related retinoid) as a potential ligand as this compound is known as a powerful mediator of cell differentiation and embryonic development. This would place Lazarillo near the cellular retinoic acid-binding proteins (CRABPs) that are members of the FABP superfamily, however (see Section 4.8.6.3).

Taken together, lipocalins seem to act everywhere and in a variety of ways in living systems. Research in insects has contributed significantly to the lipocalin field from the beginning and will hopefully do so in the future. There are many intriguing aspects of lipocalins insects may provide an ideal platform to focus on. This chapter may be helpful in this respect in attempting to provide an encouraging overview of this field.

Acknowledgments

The author is grateful to Lars Prade (Basel, Switzerland) for preparing all the basic figures of the crystal structures for this chapter. Hazel M. Holden (Madison, USA) kindly provided the atomic coordinates for insecticyamin. He further likes to thank Diego Sánchez (Valladolid, Spain), Arne Skerra (Munich, Germany), Chi-Young Yun (Deajeon, Korea), Sook Jae Seo (Gyeongnam, Korea), Hitoshi Saito (Tokyo, Japan), and Frantisek Sehnal (Ceské Budejovice, Czech Republic) for providing figures, reprints, accepted manuscripts, data of unpublished research, and other useful information.

References

Andersen, J.F., Champagne, D.E., Weichsel, A., Ribeiro, J.M.C., Balfour, C.A., *et al.*, **1997**. Nitric oxide binding and crystallization of recombinant nitrophorin I, a nitric oxide transport protein from the blood-sucking bug *Rhodnius prolixus*. *Biochemistry 36*, 4423–4428.

Andersen, J.F., Francischetti, I.M.B., Valenzuela, J.G., Schuck, P., Ribeiro, J.M.C., **2003**. Inhibition of hemostasis by a high affinity biogenic amine-binding protein from the saliva of a blood-feeding insect. *J. Biol. Chem. 278*, 4611–4617.

Andersen, J.F., Weichsel, A., Balfour, C.A., Champagne, D.E., Montfort, W.R., **1998**. The crystal structure of nitrophorin 4 at 1.5 Å resolution: transport of nitric oxide by a lipocalin-based heme protein. *Structure 15*, 1315–1327.

Arlian, L.G., **2002**. Arthropod allergens and human health. *Annu. Rev. Entomol. 47*, 395–433.

Arruda, L.K., Vailes, L.D., Hayden, M.L., Benjamin, D.C., Chapman, M.D., **1995**. Cloning of cockroach allergen, Bla g 4, identifies ligand binding proteins (or calycins) as a cause of IgE antibody responses. *J. Biol. Chem. 270*, 31196–31201.

Banaszak, L., Winter, N., Xu, Z., Bernlohr, D.A., Cowan, S., *et al.*, **1994**. Lipid-binding proteins: a family of fatty acid and retinoid transport proteins. *Adv. Protein Chem. 45*, 89–151.

Barañano, D.E., Rao, M., Ferris, C.D., Snyder, S.H., **2002**. Biliverdin reductase: a major physiologic cytoprotectant. *Proc. Natl Acad. Sci. USA 99*, 16093–16098.

Barañano, D.E., Snyder, S.H., **2001**. Neural roles for heme oxygenase: contrasts to nitric oxid synthase. *Proc. Natl Acad. Sci. USA 98*, 10996–11002.

Benning, M.M., Smith, A.F., Wells, M.A., Holden, H.M., **1992**. Crystallization, structure determination and least-squares refinement to 1.75 Å resolution of the fatty-acid-binding protein isolated from *Manduca sexta* L. *J. Mol. Biol. 228*, 208–219.

Beste, G., Schmidt, F.S., Stibora, T., Skerra, A., **1999**. Small antibody-like proteins with prescribed ligand specificities derived from the lipocalin fold. *Proc. Natl Acad. Sci. USA 96*, 1898–1903.

Beuckmann, C.T., Aoyagi, M., Okazaki, I., Hiroike, T., Toh, H., *et al.*, **1999**. Binding of biliverdin, bilirubin, and thyroid hormones to lipocalin-type prostaglandin D synthase. *Biochemistry 38*, 8006–8013.

Breiter, D.R., Kanost, M.R., Benning, M.M., Wesenberg, G., Law, J.H., *et al.*, **1991**. Molecular structure of an apolipoprotein determined at 2.5-Å resolution. *Biochemistry 30*, 603–608.

Briand, L., Swasdipan, N., Nespoulous, C., Bézirard, V., Blon, F., *et al.*, **2002**. Characterization of a chemosensory protein (ASP3c) from honeybee (*Apis mellifera* L.) as a brood pheromone carrier. *Eur. J. Biochem. 269*, 4586–4596.

Bugos, R.C., Hieber, A.D., Yamamoto, H.Y., **1998**. Xanthophyll cycle enzymes are members of the lipocalin family, the first identified in plants. *J. Biol. Chem. 273*, 15321–15324.

Büyükgüzel, K., Tunaz, H., Putnam, S.M., Stanley, D., **2002**. Prostaglandin biosynthesis by midgut tissue isolated from the tobacco hornworm, *Manduca sexta*. *Insect Biochem. Mol. Biol. 32*, 435–443.

Caraballo, L., Puerta, L., Jimenez, S., Martinez, B., Mercado, D., *et al.*, **1997**. Cloning and IgE binding of a recombinant allergen from the mite *Blomia tropicalis*, homologous with fatty acid-binding proteins. *Int. Arch. Allergy Immunol. 112*, 341–347.

Cavagnari, B.M., Scaraffia, P.Y., Haller, J.F., Gerez de Burgos, N.M., Santomé, J.A., **2000**. Presence of a fatty acid-binding protein and lipid stores in flight muscles of *Dipetalogaster maximus* (Hemiptera: Reduviidae). *J. Med. Entomol. 37*, 938–944.

Champagne, D.E., Andersen, J.F., Ribeiro, J.M.C., **1996**. Sequence of salivary platelet aggregation inhibitor 2 submitted to EMBL/GenBank/DDBJ databases (URL for: GenBank – www.ncbi.nlm.nib.gov; EMBL – www.ebi.ac.uk/embl; DDBJ –www.ddbj.nig.ac.jp).

Champagne, D.E., Nussenzveig, R.H., Ribeiro, J.M.C., **1995**. Purification, partial characterization, and cloning of nitric oxide-carrying heme proteins (nitrophorins) from salivary glands of the blood-sucking insect *Rhodnius prolixus*. *J. Biol. Chem. 270*, 8691–8695.

Charron, J.-B.F., Breton, G., Badawi, M., Sarhan, F., **2002**. Molecular and structural analysis of a novel temperature stress-induced lipocalin from wheat and *Arabidopsis*. *FEBS Lett. 517*, 129–132.

Chen, X., Haunerland, N.H., **1994**. Fatty acid binding protein expression in locust flight muscle: induction by flight, adipokinetic hormone, and low density lipoprotein. *Insect Biochem. Mol. Biol. 24*, 573–579.

Cherbas, P., **1973**. Biochemical studies of insecticyanin. Ph.D. thesis, Harvard University, Cambridge, MA.

Chino, H., Abe, Y., Takahashi, K., **1983**. Purification and characterization of a biliverdin-binding cyanoprotein from the locust hemolymph. *Biochim. Biophys. Acta 748*, 109–115.

Chinzei, Y., Haruna, T., Miura, K., Numata, H., Nakayama, S., **1990**. Purification and characterization of biliverdin-associated cyanoprotein from eggs and hemolymph of the bean bug, *Riptortus clavatus* (Heteroptera: Alydidae). *Insect Biochem. 20*, 545–555.

Cianci, M., Rizkallah, P.J., Olczak, A., Raftery, J., Chayen, N.E., *et al.*, **2001**. Structure of lobster apocrustacyanin A_1 using softer X-rays. *Acta Crystallog. D57*, 1219–1229.

Cianci, M., Rizkallah, P.J., Olczak, A., Raftery, J., Chayen, N.E., *et al.*, **2002**. The molecular basis of the coloration mechanism in lobster shell: β-crustacyanin at 3.2-Å resolution. *Proc. Natl Acad. Sci. USA 99*, 9795–9800.

Cornette, R., Farine, J.-P., Quennedey, B., Brossut, R., **2001**. Molecular characterization of a new adult male putative calycin specific to tergal gland aphrodisiac secretion in the cockroach *Leucophaea maderae*. Submitted to the EMBL/GenBank/DDBJ databases (URL for: GenBank – www.ncbi.nlm.nib.gov; EMBL – www.ebi.ac.uk/embl; DDBJ – www.ddbj.nig.ac.jp).

Eriksson, T.L.J., Rasool, O., Huecas, S., Whitley, P., Crameri, R., *et al.*, **2001**. Cloning of three new allergens from the dust mite *Lepidoglyphus destructur* using phage surface display technology. *Eur. J. Biochem. 268*, 287–294.

Eriksson, T.L.J., Whitley, P., Johansson, E., van Hage-Hamsten, M., Gafvelin, G., **1999**. Identification and characterization of two allergens from the dust mite *Acarus siro*, homologous with fatty acid-binding proteins. *Int. Arch. Allergy Immunol. 119*, 275–281.

Evans, J.D., Wheeler, D.E., **1999**. Differential gene expression between developing queens and workers in the honey bee, *Apis mellifera*. *Proc. Natl Acad. Sci. USA 96*, 5575–5580.

ExPASy, **2003**. http://us.expasy.org.

Falchuk, K.H., Contin, J.M., Dziedzik, T.S., Feng, Z., French, T.C., *et al.*, **2002**. A role for biliverdin IXα in dorsal axis development of *Xenopus laevis* embryo. *Proc. Natl Acad. Sci. USA 99*, 251–256.

Favia, G., Mariotti, G., Coluzzi, M., Mathiopoulos, K.D., **1996**. Sequence submitted to EMBL/GenBank/DDBJ databases (URL for: GenBank – www.ncbi.nlm.nib.gov; EMBL – www.ebi.ac.uk/embl; DDBJ – www.ddbj.nig.ac.jp).

Field, L.M., Pickett, J.A., Wadhams, L.J., **2000**. Molecular studies in insect olfaction. *Insect Mol. Biol. 9*, 545–551.

Filippov, V.A., Filippova, M.A., Kodrík, D., Sehnal, F., **1995**. Two lipocalin-like peptides of insect brain. In: Suzuki, A., Kataoka, H., Matsumoto, S. (Eds.), Molecular Mechanisms of Insect Metamorphosis and

Diapause. Industrial Publishing and Consulting, Tokyo, pp. 35–43.
Flower, D.R., **1994**. The lipocalin protein family: a role in cell regulation. *FEBS Lett. 354*, 7–11.
Flower, D.R., **1995**. Multiple molecular recognition properties of the lipocalin protein family. *J. Mol. Recogn. 8*, 185–195.
Flower, D.R., **2000**. Beyond the superfamily: the lipocalin receptors. *Biochim. Biophys. Acta 1482*, 327–336.
Flower, D.R., North, A.C.T., Attwood, T.K., **1993**. Structure and sequence relationships in the lipocalins and related proteins. *Protein Sci. 2*, 753–761.
Flower, D.R., North, A.C.T., Sansom, C.E., **2000**. The lipocalin protein family: structural and sequence overview. *Biochim. Biophys. Acta 1482*, 9–24.
Francischetti, I.M.B., Andersen, J., Ribeiro, J.M.C., **2002**. Biochemical and functional characterization of recombinant *Rhodnius prolixus* platelet aggregation inhibitor 1 as a novel lipocalin with high affinity for adenosine diphosphate and other adenine nucleotides. *Biochemistry 41*, 3810–3818.
Francischetti, I.M.B., Ribeiro, J.M.C., Champagne, D., Andersen, J., **2000**. Purification, cloning, expression, and mechanism of action of a novel platelet aggregation inhibitor from the salivary gland of the blood-sucking bug, *Rhodnius prolixus*. *J. Biol. Chem. 275*, 12639–12650.
Fuentes-Prior, P., Noeske-Jungblut, C., Donner, P., Schleuning, W.-D., Huber, R., *et al.*, **1997**. Structure of the thrombin complex with triabin, a lipocalin-like exosite-binding inhibitor derived from a triatomine bug. *Proc. Natl Acad. Sci. USA 94*, 11845–11850.
Ganfornina, M.D., Gutiérrez, G., Bastiani, M., Sánchez, D., **2000**. A phylogenetic analysis of the lipocalin protein family. *Mol. Biol. Evolution 17*, 114–126.
Ganfornina, M.D., Sánchez, D., Bastiani, M.J., **1995**. Lazarillo, a new GPI-linked surface lipocalin, is restricted to a subset of neurons in the grasshopper embryo. *Development 121*, 123–134.
Glatz, J.F.C., van der Vusse, G.J., **1996**. Cellular fatty acid-binding proteins: their function and physiological significance. *Prog. Lipid Res. 35*, 243–282.
Goodman, W.G., Adams, B., Trost, J.T., **1985**. Purification and characterization of a biliverdin-associated protein from the hemolymph of *Manduca sexta*. *Biochemistry 24*, 1168–1175.
Gordon, E.J., Leonard, G.A., McSweeney, S., Zagalsky, P.F., **2001**. The C_1 subunit of α-crustacyanin: the *de novo* phasing of the crystal structure of a 40 kDa homomeric protein using the anomalous scattering from S atoms combined with the direct methods. *Acta Crystallog. D57*, 1230–1237.
Graf, S., Ludwig, P., Boyan, G., **2000**. Lazarillo expression reveals a subset of neurons contributing to the primary axon scaffold of the embryonic brain of the grasshopper *Schistocerca gregaria*. *J. Comp. Neurol. 419*, 394–405.
Graham, L.A., Davies, P.L., **2002**. The odorant-binding proteins of *Drosophila melanogaster*: annotation and characterization of a divergent gene family. *Gene 292*, 43–55.
Hartsfield, C.L., **2002**. Cross talk between carbon monoxide and nitric oxide. *Antiox. Redox Signal. 4*, 301–307.
Haunerland, N.H., **1996**. Insect storage proteins: gene families and receptors. *Insect Biochem. Mol. Biol. 26*, 755–765.
Haunerland, N.H., **1997**. Transport and utilization of lipids in insect flight muscles. *Comp. Biochem. Physiol. B 117*, 475–482.
Haunerland, N.H., Andolfatto, P., Chrisholm, J.M., Wang, Z., Chen, X., **1992a**. Fatty-acid-binding protein in locust flight muscle: developmental changes of expression, concentration and intracellular distribution. *Eur. J. Biochem. 210*, 1045–1051.
Haunerland, N.H., Bowers, W.S., **1986**. A larval specific lipoprotein: purification and characterization of a blue chromoprotein from *Heliothis zea*. *Biochem. Biophys. Res. Commun. 134*, 580–586.
Haunerland, N.H., Chen, X., Andolfatto, P., Chrisholm, J.M., Wang, Z., **1993**. Developmental changes of FABP concentration, expression, and intracellular distribution in locust flight muscle. *Mol. Cell. Biochem. 123*, 153–158.
Haunerland, N.H., Chrisholm, J.M., **1990**. Fatty acid binding protein in flight muscle of the locust, *Schistocera gregaria*. *Biochim. Biophys. Acta 1047*, 233–238.
Haunerland, N.H., Jacobson, B.L., Wesenberg, G., Rayment, I., Holden, H.M., **1994**. Three-dimensional structure of the muscle fatty-acid-binding protein isolated from the desert locust *Schistocera gregaria*. *Biochemistry 33*, 12378–12385.
Haunerland, N.H., Ortego, F., Strausfeld, C.M., Bowers, W.S., **1992b**. Blue lipophorin from *Podisus maculaventris*. *Arch. Insect Biochem. Physiol. 20*, 49–59.
Heilmann, L.J., **1998**. Fatty acid binding protein of *Helicoverpa zea*. Submitted to the EMBL/GenBank/DDBJ databases (URL for: GenBank – www.ncbi.nlm.nib.gov; EMBL – www.ebi.ac.uk/embl; DDBJ – www.ddbj.nig.ac.jp).
Hieber, A.D., Bugos, R.C., Yamamoto, H.Y., **2000**. Plant lipocalins: violaxanthin de-epoxidase and zeaxanthin epoxidase. *Biochim. Biophys. Acta 1482*, 84–91.
Holden, H.M., Rypniewski, W.R., Law, J.H., **1987**. The molecular structure of insecticyanin from the tobacco hornworm *Manduca sexta* L. at 2.6 Å resolution. *EMBO J. 6*, 1565–1570.
Huber, R., Schneider, M., Epp, O., Mayr, I., Messerschmidt, A., *et al.*, **1987a**. Crystallization, crystal structure analysis and preliminary molecular model of the bilin binding protein from the insect *Pieris brassicae*. *J. Mol. Biol. 195*, 423–434.
Huber, R., Schneider, M., Mayr, I., Müller, R., Deutzmann, R., *et al.*, **1987b**. Molecular structure of the bilin binding protein (BBP) from *Pieris brassicae* after refinement at 2.0 Å resolution. *J. Mol. Biol. 198*, 499–513.
Jones, G., Ryan, R.O., Haunerland, N.H., Law, J.H., **1988**. Purification and characterization of a very high

density chromolipoprotein from the hemolymph of *Trichoplusia ni* (Hübner). *Arch. Insect Biochem. Physiol. 7*, 1–11.

Jones, T.A., Bergfors, T., Sedzik, J., Unge, T., **1988**. The three-dimensional structure of P2 myelin protein. *EMBO J. 7*, 1597–1604.

Jones, T.A., Zou, J.-Y., Cowan, S.W., Kjeldgaard, M., **1991**. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. A47*, 110–119.

Kang, Y., Kulakosky, P.C., van Antwerpen, R., Law, J.H., **1995**. Sequestration of insecticyanin, a blue hemolymph protein, into the egg of the hawkmoth *Manduca sexta*. Evidence for receptor-mediated endocytosis. *Insect Biochem. Mol. Biol. 25*, 503–510.

Kang, Y., Ziegler, R., van Antwerpen, R., Law, J.H., **1997**. Characterization of the solubilized oocyte membrane receptor for insecticyanin, a biliprotein of the hawkmoth, *Manduca sexta*. *Biochem. Biophys. Acta 1324*, 285–295.

Kaur, H., Hughes, M.N., Green, C.J., Naughton, P., Foresti, R., *et al.*, **2003**. Interaction of bilirubin and biliverdin with reactive nitrogen species. *FEBS Lett. 543*, 113–119.

Kayser, H., **1984**. *De novo* synthesis and levels of cytochrome *c* and a biliprotein during pupal–adult development of a butterfly, *Pieris brassicae*. *Z. Naturforsch. 39c*, 938–947.

Kayser, H., **1985**. Pigments. In: Kerkut, G.A., Gilbert, L.I. (Eds.), Comprehensive Insect Biochemistry, Physiology and Pharmacology, vol. 10. Pergamon Press, Oxford, pp. 367–415.

Kayser, H., Krull-Savage, U., **1984**. Development-specific incorporation of [^{14}C]5-aminolevulinate and [^{3}H]leucine into cytochrome *c* and biliprotein in the butterfly, *Pieris brassicae*: correlation with the ecdysteroid titer in the pupa. *Z. Naturforsch. 39c*, 948–957.

Keen, J.N., Caceres, I., Eliopoulos, E.E., Zagalsky, P.F., Findlay, J.B.C., **1991**. Complete sequence and model for the C_1 subunit of the carotenoprotein, crustacyanin, and model for the dimer, β-crustacyanin, formed from the C_1 and A_2 subunits with astaxanthin. *Eur. J. Biochem. 202*, 31–40.

Kiely, M.L., Riddiford, L.M., **1985**. Temporal programming of epidermal cell protein synthesis during the larval–pupal transformation of *Manduca sexta*. *Roux's Arch. Devel. Biol. 194*, 325–335.

Koglin, M., Behrends, S., **2002**. Biliverdin IX is an endogenous inhibitor of soluble guanylyl cyclase. *Biochem. Pharmacol. 64*, 109–116.

Korchi, A., Brossut, R., Bouhin, H., Delachambre, J., **1999**. cDNA cloning of an adult male putative lipocalin specific to tergal gland aphrodisiac secretion in an insect (*Leucophaea maderae*). *FEBS Lett. 449*, 125–128.

Korndörfer, I.P., Beste, G., Skerra, A., **2003**. Crystallographic analysis of an "anticalin" with tailored specificity for fluorescein reveals high structural plasticity of the lipocalin loop region. *Proteins 53*, 121–129.

LaLonde, J.M., Bernlohr, D.A., Banaszak, L.J., **1994**. The up-and-down β-barrel proteins. *FASEB J. 8*, 1240–1247.

Lartigue, A., Campanacci, V., Roussel, A., Larsson, A.M., Jones, T.A., *et al.*, **2002**. X-ray structure and ligand binding study of a moth chemosensory protein. *J. Biol. Chem. 277*, 32094–32098.

Lechner, M., Wojnar, P., Redl, B., **2001**. Human tear lipocalin acts as an oxidative-stress-induced scavenger of potentially harmful lipid peroxidation products in a cell culture system. *Biochem. J. 356*, 129–135.

Li, W.-C., Riddiford, L.M., **1992**. Two distinct genes encode two major isoelectric forms of insecticyanin in the tobacco hornworm, *Manduca sexta*. *Eur. J. Biochem. 205*, 491–499.

Li, W.-C., Riddiford, L.M., **1994**. The two duplicated insecticyanin genes, *ins*-a and *ins*-b are differentially expressed in the tobacco hornworm, *Manduca sexta*. *Nucl. Acids Res. 22*, 2945–2950.

Li, W.-C., Riddiford, L.M., **1996**. Differential expression of the two duplicated insecticyanin genes, *ins*-a and *ins*-b, in the black mutant of *Manduca sexta*. *Arch. Biochem. Biophys. 330*, 65–70.

Lipocalin, **2004**. www.jenner.ac.uk/lipocalin.htm.

Maatman, R.G.H.J., Degano, M., Van Moerkerk, H.T.B., Van Marrewijk, W.J.A., Van der Horst, D.J., *et al.*, **1994**. Primary structure and binding characteristics of locust and human muscle fatty-acid-binding proteins. *Eur. J. Biochem. 221*, 801–810.

Maines, M.D., **1997**. The heme oxygenase system: a regulator of second messenger gases. *Annu. Rev. Pharmacol. Toxicol. 37*, 517–554.

Mansfield, S.G., Cammer, S., Alexander, S.C., Muehleisen, D.P., Gray, R.S., *et al.*, **1998**. Molecular cloning and characterization of an invertebrate cellular retinoic acid binding protein. *Proc. Natl Acad. Sci. USA 95*, 6825–6830.

Mäntyjärvi, R., Rautiainen, J., Virtanen, T., **2000**. Lipocalins as allergens. *Biochim. Biophys. Acta 1482*, 308–317.

Maruta, K., Yoshiga, T., Katagiri, C., Ochiai, M., Tojo, S., **2002**. Purification and characterization of biliverdin-binding vitellogenin from the hemolymph of the common cutworm, *Spodoptera litura*. *Arch. Insect Biochem. Physiol. 50*, 97–106.

Miura, K., Shinoda, T., Yura, M., Nomura, S., Kamiya, K., *et al.*, **1998**. Two hexameric cyanoprotein subunits from an insect, *Riptortus clavatus*: sequence, phylogeny and developmental and juvenile hormone regulation. *Eur. J. Biochem. 258*, 929–940.

Montfort, W.R., Weichsel, A., Andersen, J.F., **2000**. Nitrophorins and related antihemostatic lipocalins from *Rhodnius prolixus* and other blood-sucking arthropods. *Biochim. Biophys. Acta 1482*, 110–118.

Moreira, M.F., Coelho, H.S.L., Zingali, R.B., Oliveira, P.L., Masuda, H., **2003**. Changes in salivary nitrophorin profile during the life cycle of the blood-sucking bug *Rhodnius prolixus*. *Insect Biochem. Mol. Biol. 33*, 23–28.

Newcomer, M.E., **1995**. Retinoid-binding proteins: structural determinants important for function. *FASEB J. 9*, 229–239.

Newcomer, M.E., Jones, T.A., Aqvist, J., Sundelin, J., Eriksson, U., *et al.*, **1984**. The 3-dimensional structure of retinol-binding protein. *EMBO J. 3*, 1451–1454.

Noeske-Jungblut, C., Haendler, B., Donner, P., Alagon, A., Possani, L., *et al.*, **1995**. Triabin, a highly potent exosite inhibitor of thrombin. *J. Biol. Chem. 270*, 28629–28634.

Noeske-Jungblut, C., Kratzschmar, J., Haendler, B., Alagon, A., Possani, L., *et al.*, **1994**. An inhibitor of collagen-induced platelet aggregation from the saliva of *Triatoma pallidipennis*. *J. Biol. Chem. 269*, 5050–5053.

Nussenzveig, R.H., Bentley, D.L., Ribeiro, J.M.C., **1995**. Nitric oxide loading of the salivary nitric-oxide-carrying hemoproteins (nitrophorins) in the blood sucking bug *Rhodnius prolixus*. *J. Exp. Biol. 198*, 1093–1098.

Oliveira, P.L., Kawooya, J.K., Ribeiro, J.M.C., Meyer, T., Poorman, R., *et al.*, **1995**. A heme-binding protein from hemolymph and oocytes of the blood-sucking insect, *Rhodnius prolixus*. *J. Biol. Chem. 270*, 10897–10901.

Paddock, C.D., McKerrow, J.H., Hansell, E., Foreman, K.W., Hsieh, I., *et al.*, **2001**. Identification, cloning, and recombinant expression of procalin, a major Triatomine allergen. *J. Immunol. 167*, 2694–2699.

Paesen, G.C., Adams, P.L., Harlos, K., Nuttall, P.A., Stuart, D.I., **1999**. Tick histamine-binding proteins: isolation, cloning, and three-dimensional structure. *Mol. Cell 3*, 661–671.

Paesen, G.C., Adams, P.L., Nuttall, P.A., Stuart, D.I., **2000**. Tick histamine-binding proteins: lipocalins with a second binding site. *Biochim. Biophys. Acta 1482*, 92–101.

Paiva-Silva, G.O., Sorgine, M.H.F., Benedetti, C.E., Meneghini, R., Almeida, I.C., *et al.*, **2002**. On the biosynthesis of *Rhodnius prolixus* heme-binding protein. *Insect Biochem. Mol. Biol. 32*, 1533–1541.

Papis, M.Z., Sawyer, L., Eliopoulos, E.E., North, A.C.T., Findlay, J.B.C., *et al.*, **1986**. The structure of beta-lactoglobulin and its similarity to plasma retinol-binding protein. *Nature 324*, 383–385.

Peitsch, M.C., Boguski, M.S., **1990**. Is apolipoprotein D a mammalian bilin-binding protein? *New Biol. 2*, 197–206.

Pelosi, P., Maida, R., **1995**. Odorant-binding proteins in insects. *Comp. Biochem. Physiol. B 111*, 503–514.

Pervais, S., Brew, K., **1987**. Homology and structure-function correlations between alpha 1-acid glycoprotein and serum retinol-binding protein and its relatives. *FASEB J. 1*, 209–214.

Price, H.M., Ryan, R.O., Haunerland, N.H., **1992**. Primary structure of locust flight muscle fatty acid binding protein. *Arch. Biochem. Biophys. 297*, 285–290.

Puerta, L., Kennedy, M.W., Jimenez, S., Caraballo, L., **1999**. Structural and ligand binding analysis of recombinant blo t 1 allergen from *Blomia tropicalis* mite, a fatty acid binding protein. *Int. Arch. Allergy Immunol. 119*, 181–184.

Ribeiro, J.M.C., Francischetti, I.M.B., **2003**. Role of arthropod saliva in blood feeding: sialome and post-sialome perspectives. *Annu. Rev. Entomol. 48*, 73–88.

Ribeiro, J.M.C., Schneider, M., Guimarães, J.A., **1995**. Purification and characterization of prolixin S (nitrophorin 2), the salivary anticoagulant of the blood-sucking bug *Rhodnius prolixus*. *Biochem. J. 308*, 243–249.

Riddiford, L.M., **1982**. Changes in translatable mRNAs during the larval–pupal transformation of the epidermis of the tobacco hornworm. *Devel. Biol. 92*, 330–342.

Riddiford, L.M., Palli, S.R., Hiruma, K., Li, W.-C., Green, J., *et al.*, **1990**. Developmental expression, synthesis, and secretion of insecticyanin by the epidermis of the tobacco hornworm, *Manduca sexta*. *Arch. Insect Biochem. Physiol. 14*, 171–190.

Riley, C.T., Barbeau, B.K., Keim, P.S., Kézdy, F.J., Heinrikson, R.L., *et al.*, **1984**. The covalent protein structure of insecticyanin, a blue biliprotein from the hemolymph of the tobacco hornworm, *Manduca sexta* L. *J. Biol. Chem. 259*, 13159–13165.

Rothemund, S., Liou, Y.-C., Davies, P.L., Krause, E., Sönnichsen, F.D., **1999**. A new class of hexahelical insect proteins revealed as putative carriers of small hydrophobic ligands. *Structure 7*, 1325–1332.

Ryan, R.O., van der Horst, D.J., **2000**. Lipid transport biochemistry and its role in energy production. *Annu. Rev. Entomol. 45*, 233–260.

Ryter, S.W., Otterbein, L.E., Morse, D., Choi, A.M.K., **2002**. Heme oxygenase/carbon monoxide signalling pathways: regulation and functional significance. *Mol. Cell. Biochem. 234/235*, 249–263.

Saito, H., **1993**. Purification and characterization of a biliverdin-binding protein from the molting fluid of the eri silkworm, *Samia cynthia ricini* (Lepidoptera: Saturniidae). *Comp. Biochem. Physiol. B 105*, 473–479.

Saito, H., **1997**. Purification and characterization of a biliverdin-binding protein from the larval cuticle of the saturniid silkworn, *Attacus atlas*. *Int. J. Wild Silkmoth Silk 3*, 1–6.

Saito, H., **1998a**. Purification and characterization of two insecticyanin-type proteins from the larval hemolymph of the Eri-silkworm, *Samia cynthia ricini*. *Biochim. Biophys. Acta 1380*, 141–150.

Saito, H., **1998b**. Purification and properties of two blue biliproteins from the larval hemolymph and integument of *Rhodinia fugax* (Lepidoptera: Saturniidae). *Insect Biochem. Mol. Biol. 28*, 995–1005.

Saito, H., Shimoda, M., **1997**. Insecticyanin of *Agrius convolvuli*: purification and characterization of the biliverdin-binding protein from the larval hemolymph. *Zool. Sci. 14*, 777–783.

Saito, H., Yamada, H., Kato, Y., **1998**. Isolation and partial characterization of chromoprotein from the larval hemolymph of the Japanese oak silkworm (*Antheraea yamamai*). *Comp. Biochem. Physiol. B 119*, 625–630.

Sakai, M., Wu, C., Suzuki, K., **2001**. Nucleotide and deduced amino acid sequences of a cDNA encoding a lipocalin protein in the central nervous system of *Bombxy mori*. *Nippon Sanshi-gaku Zasshi 70*, 105–111.

Sánchez, D., Ganfornina, M.D., Bastiani, M.J., **1995**. Developmental expression of the lipocalin Lazarillo and its role in axonal pathfinding in the grasshopper embryo. *Development 121*, 135–147.

Sánchez, D., Ganfornina, M.D., Bastiani, M.J., **2000a**. Lazarillo, a neuronal lipocalin in grasshoppers with a role in axon guidance. *Biochim. Biophys. Acta 1482*, 102–109.

Sánchez, D., Ganfornina, M.D., Torres-Schumann, S., Speese, S.D., Lora, J.M., *et al.*, **2000b**. Characterization of two novel lipocalins expressed in the *Drosophila* embryonic nervous system. *Int. J. Devel. Biol. 44*, 349–359.

Sandler, B.H., Nikonowa, L., Leal, W.S., Clardy, J., **2000**. Sexual attraction in the silkworm moth: structure of the pheromone-binding-protein-bombykol complex. *Chem. Biol. 7*, 143–151.

Sangamnatdej, S., Paesen, G.C., Slovak, M., Nuttall, P.A., **2002**. A high affinity serotonin- and histamine-binding lipocalin from tick saliva. *Insect Mol. Biol. 11*, 79–86.

Scheer, H., Kayser, H., **1988**. Conformational studies of biliproteins from the insects *Pieris brassicae* and *Cerura vinula*. *Z. Naturforsch. 43c*, 84–90.

Schirmer, T., Bode, W., Huber, R., **1987**. Refined three-dimensional structures of two cyanobacterial C-phycocyanins at 2.1 and 2.5 Å resolution. *J. Mol. Biol. 196*, 677–695.

Schlehuber, S., Skerra, A., **2001**. Duocalins: engineered ligand-binding proteins with dual specificity derived from the lipocalin fold. *Biol. Chem. 382*, 1335–1342.

Schmidt, F.S., Skerra, A., **1994**. The bilin-binding protein of *Pieris brassicae*: cDNA sequence and regulation of expression reveal distinct features of this insect pigment protein. *Eur. J. Biochem. 219*, 855–863.

Schneider, S., Baumann, F., Geiselhart, P., Kayser, H., Scheer, H., **1988**. Biliproteins from the butterfly *Pieris brassicae* studied by time-resolved fluorescence and coherent anti-stokes Raman spectroscopy. *Photochem. Photobiol. 48*, 239–242.

Sehringer, B., **1999**. Bilin-bindendes protein von *Pieris brassicae*: untersuchungen zur regulation des mRNA-gehalts und der proteinbiosynthese dieses pigmentproteins während der puppenentwicklung. Diploma thesis, University of Freiburg, Germany.

Seo, S.-J., Cheon, H.-M., **2003**. Molecular characterization of hyphantrin from *Hyphantria cunea*. Submitted to the EMBL/GenBank/DDBJ databases (URL for: GenBank – www.ncbi.nlm.nib.gov; EMBL – www.ebi.ac.uk/embl; DDBJ – www.ddbj.nig.ac.jp).

Skerra, A., **2000**. Lipocalins as a scaffold. *Biochim. Biophys. Acta 1482*, 337–350.

Skerra, A., **2001**. "Anticalins": a new class of engineered ligand-binding proteins with antibody-like properties. *Rev. Mol. Biotechnol. 74*, 257–275.

Smith, A.F., Tsuchida, K., Hanneman, E., Suzuki, T.C., Wells, M.A., **1992**. Isolation, characterization, and cDNA sequence of two fatty acid-binding proteins from the midgut of *Manduca sexta* larvae. *J. Biol. Chem. 267*, 380–384.

Sun, J., Yuda, M., Miura, K., Chinzei, Y., **1998**. Characterization and cDNA cloning of a hemoprotein in the salivary glands of the blood-sucking insect, *Rhodnius prolixus*. *Insect Biochem. Mol. Biol. 28*, 191–200.

Suter, F., Kayser, H., Zuber, H., **1988**. The complete amino-acid sequence of the bilin-binding protein from *Pieris brassicae* and its similarity to a family of serum transport proteins like the retinol-binding proteins. *Biol. Chem. Hoppe-Seyler 369*, 497–505.

Tsuchihara, K., Ueno, K., Yamanaka, A., Isono, K., Endo, K., *et al.*, **2000**. A putative binding protein for lipophilic substances related to butterfly oviposition. *FEBS Lett. 478*, 299–303.

Trost, J.T., Goodman, W.G., **1986**. Hemolymph titers of the biliprotein, insecticyanin, during development of *Manduca sexta*. *Insect Biochem. 16*, 353–358.

Urade, Y., Hayashi, O., **2000**. Prostaglandin D synthase: structure and function. *Vitam. Hormones 58*, 89–120.

Valenzuela, J.G., Ribeiro, J.M.C., **1998**. Purification and cloning of the salivary nitrophorin from the hemipteran *Cimex lectularius*. *J. Exp. Biol. 201*, 2659–2664.

Van der Horst, D.J., van Doorn, J.M., Passier, P.C.C.M., Vork, M.M., Glatz, J.F.C., **1993**. Role of fatty acid-binding protein in lipid metabolism of insect flight muscle. *Mol. Cell. Biochem. 123*, 145–152.

Veerkamp, J.H., Maatman, R.G.H.J., **1995**. Cytoplasmic fatty acid-binding proteins: their structure and genes. *Prog. Lipid Res. 34*, 17–52.

Walker, F.A., Ribeiro, J.M.C., Montfort, W.R., **1999**. Novel nitric oxide-liberating heme proteins from the saliva of bloodsucking insects. *Metal Ions Biol. Syst. 36*, 621–663.

Weichsel, A., Andersen, J.F., Champagne, D.E., Walker, F.A., Montfort, W.R., **1998**. Crystal structures of a nitric oxide transport protein from a blood-sucking insect. *Nature Struct. Biol. 5*, 304–309.

Weichsel, A., Andersen, J.F., Roberts, S.A., Montfort, W.R., **2000**. Nitric oxide binding to nitrophorin 4 induces complete distal pocket burial. *Nature Struct. Biol. 7*, 551–554.

Weiss, G.A., Lowman, H.B., **2000**. Anticalins versus antibodies: made-to-order binding proteins from small molecules. *Chem. Biol. 7*, R177–R184.

Wilks, A., **2002**. Heme oxygenase: evolution, structure, and mechanism. *Antiox. Redox Signal. 4*, 603–614.

Wu, Q., Andolfatto, P., Haunerland, N.H., **2001**. Cloning and sequence of the gene encoding the muscle fatty acid binding protein from the desert locust, *Schistocerca gregaria*. *Insect Biochem. Mol. Biol. 31*, 553–562.

Wu, Q., Haunerland, N.H., **2001**. A novel fatty acid response element controls the expression of the flight muscle FABP gene of the desert locust, *Schistocerca gregaria*. *Eur. J. Biochem. 268*, 5894–5900.

Yoshiga, T., Maruta, K., Tojo, S., **1998**. Developmental changes of storage proteins and biliverdin-binding proteins in the haemolymph and fat body of the common cutworm, *Spodoptera litura*. *J. Insect Physiol. 44*, 67–76.

Yoshiga, T., Tojo, S., **1995**. Purification and characterization of four biliverdin-binding proteins from larval haemolymph of the common cutworm, *Spodoptera litura*. *Insect Biochem. Mol. Biol. 25*, 575–581.

Yoshiga, T., Tojo, S., **2001**. Effect of a juvenile hormone analog, methoprene, on the hemolymph titers of biliverdin-binding proteins in the common cutworm, *Spodoptera litura* (Lepidoptera: Noctuidae). *Appl. Entomol. Zool. 36*, 337–343.

Yousefi, S., Simon, H.U., **2002**. Granulocyte apoptosis: death by a secreted lipocalin? *Cell Death Different. 9*, 595–597.

Zanotti, G., **1999**. Muscle fatty acid-binding protein. *Biochim. Biophys. Acta 1441*, 94–105.

Zipfel, D., **1982**. Isolierung und Charakterisierung eines Biliproteins von *Pieris brassicae* L. Diploma thesis, University of Ulm, Germany.

Relevant Websites

http://us.expasy.org – ExPASy, site for various software including Clustal W.
http://www.jenner.ac.uk – database for lipocalins.
http://www.ncbi.nlm.nih.gov – database for GenBank.
http://www.ebi.ac.uk – database for EMBL.
http://www.ddbj.nig.ac.jp – database for DDBJ.
http://www.rcsb.org – Protein Data Bank.

4.9 Eicosanoids

D W Stanley, University of Nebraska, Lincoln, NE, USA

© 2005, Elsevier BV. All Rights Reserved.

4.9.1. Introduction

Dadd's chapter in the first edition of *Comprehensive Insect Biochemistry, Physiology, and Pharmacology* (Dadd, 1985) treated organismal nutrition, an inquiry into the requirements, metabolism, and cellular functions of the chemicals which make up the natural diets of animals. Among the many classes of nutrients, Dadd reviewed what were then recent revelations of the requirements, presence and biological significance of arachidonic acid (AA) in insect nutritional science. He drew on the biomedical background to speculate that one of the main functions of AA in insect biology would be its role as direct precursor for prostaglandin (PG) biosynthesis, although there was very little idea of meaning which could be attached to the presence of PGs in insect tissues. In the early 1980s, it was known that PGE_2 results in egg-laying behavior in two cricket species (Loher *et al.*, 1981) and that mosquito larvae required dietary AA to complete the developmental excursion through pupation and successful adult emergence (Dadd and Kleinjan, 1979). In the years since Dadd's seminal review of insect nutrition, considerablely more insight into the biological significance of PGs and other eicosanoids in insects has come to light. Thus, this chapter is regarded (in the second edition of the comprehensive series) as an update on progress in the wake of Dadd's speculation.

The term "eicosanoid" was originally coined by Corey *et al.* (1980), who drew upon the Greek word "eikosi" meaning twenty, to formulate a single, encompassing term for all biologically active, oxygenated metabolites of AA and two other C_{20} polyunsaturated fatty acids (PUFAs). Three groups of eicosanoids: PGs, epoxyeicosatrienoic acids, and various lipoxygenase (LOX) products are recognized. Most of the known eicosanoids were first discovered in research into the physiology of humans and other mammals. The first known eicosanoids, PGs, were discovered in the 1930s. PGs, so named because they were first associated with the human prostate gland, stimulated contractions in some, but not all, mammalian smooth-muscle preparations (Goldblatt, 1933; von Euler, 1936). Von Euler (whose father Hans von Euler shared the 1929 Nobel Prize in Chemistry for work on the enzymology of sugar metabolism) eventually took part in the 1970 Nobel Prize in Medicine or Physiology for his discoveries of hormone-like substances, including PGs and adrenalin. The early research revealed the presence of a stimulating substance in human semen, which had an acidic and lipoidal nature. Further research into the substance was stimulated by the isolation and determination of the chemical structures of three PGs (Bergstrom *et al.*, 1962a, 1962b). Inspection of these structures immediately revealed the close similarity between

AA and PGs, from which it became apparent that PGs were formed by enzymatic oxygenation of AA.

Biosynthesis of PGs and other eicosanoids from AA might not have attracted the enormous attention it received were it not for the work of Sir John Vane, who discovered that the analgesic properties of aspirin and indomethacin were based on inhibition of PG biosynthesis (Ferreira *et al*., 1971; Vane, 1971). Discovery that a range of discomforts associated with fever, inflammation, and dysmenorrhea (among other ailments) could be eased by inhibition of the enzymes involved in PG biosynthesis had immediate and obvious pharmacological implications. A very large research venture in government, university, and corporate laboratories, which continues unabated, ensued. Sune Bergstrom, his student Bengt Samuelsson, and John Vane were awarded the 1982 Nobel Prize in Medicine or Physiology for their work on the chemistry and pharmacology of PGs (Oates, 1982).

It would be difficult to overstate the biological and pathophysiological significance of PGs and other eicosanoids in human and veterinary medicine. These compounds are present and responsible for important biological actions in virtually all mammalian tissues. Among mammals, eicosanoids are responsible for physiological events registered at the organismal (fever), tissue (smooth-muscle contraction or relaxation), cellular (downregulation of mast cell function), and molecular (gene expression) levels of biological organization. Many eicosanoid-regulated events relate directly to pathophysiology and other aspects of human health and well-being. The slow-reacting substance of anaphylaxis, for example, is a mixture of eicosanoids. Virtually all over-the-counter (and some prescription) pain relievers act by inhibiting PG biosynthesis. Again, taken as a group of biologically active compounds, eicosanoids exert profound influences at several levels of mammalian biology.

The biological significance of PGs and other eicosanoids seems all the more interesting when viewed in the broader context of nonmammalian vertebrates and invertebrates. Stanley and Howard (1998) and Stanley (2000) expressed this as the biological paradigm of eicosanoids, in distinction to what might be regarded as a more traditional biomedical model. This view arises from the observation that PGs and other eicosanoids are present and biologically active in representatives of all major animal phyla, including single celled organisms (Stanley, 2000). Eicosanoids probably comprise one of the earliest signaling systems in cell biology. Throughout animal – particularly invertebrate – evolution, eicosanoids have been recruited into a large number of diverse biological roles. Some of these roles, such as modulation of ion transport physiology or cellular immune defense reactions, may be fundamental eicosanoid actions which occur in all animal cells, while others appear to be rather specific to certain evolutionary or ecological contexts. For example, PGE_2 releases egg-laying behavior in newly mated females of a few, but only a few, insect species (Stanley-Samuelson and Loher, 1986). The large diversity of eicosanoid actions in animal biology suggests considerable explanatory power. That is to say, curiosity about eicosanoids can reveal new information on the regulation and mediation of physiological events in animals, and for our purposes, in the Class Insecta.

The goal of this chapter is to provide an overview of eicosanoids as they relate to insect biology. As mentioned earlier, eicosanoids occur in virtually all invertebrates and this broader information has been reviewed elsewhere (Stanley-Samuelson, 1987, 1991; Stanley and Howard, 1998; Stanley, 2000). Discussion begins with a brief overview of the chemistry and biochemistry of eicosanoids, to which we now turn.

4.9.2. Definitions, Structures, and Biosynthesis of Eicosanoids

4.9.2.1. Introduction

Our understanding of eicosanoids was developed from research on the chemistry, biochemistry, and biology of these molecules in mammals. Due to its clinical significance, this research has produced a very large corpus of literature, which for convenience is called the "biomedical model." Probably because research on eicosanoid actions in mammals preceded inquiry into invertebrates by a couple of decades, a great deal of the research on eicosanoids in invertebrate systems followed the protocols as well as the language of the biomedical model.

Eicosanoids are biosynthesized by enzymatic oxygenation of AA or two other C_{20} PUFAs. Information on the presence, metabolism, and nutritional requirements of PUFAs in insects can be taken from from several reviews (Dadd, 1973, 1977, 1981, 1983, 1985; Stanley-Samuelson *et al*., 1988; Stanley, 2000).

Three PUFAs, $C_{20:3n-6}$, $C_{20:4n-6}$, and $C_{20:5n-3}$, are potential direct substrates for eicosanoid biosynthesis, although most of the available information on eicosanoid biosynthesis is focused on metabolism of AA. Indeed, eicosanoid biosynthesis is often taken to be synonymous with "arachidonate metabolism," sometimes incorrectly referred to as the

"arachidonate cascade." AA does not actually enter a cascade; instead, it enters a range of metabolic pathways. Perhaps the sheer number of potential metabolic fates gives the sense of a cascade, but, this is not consistent with the meaning of the word. Eicosanoid biosynthesis, or arachidonate metabolism, more accurately describe the processes, and are the appropriate terms.

4.9.2.2. The Biomedical Model of Eicosanoid Biosynthesis

The following paragraphs outline the main pathways of eicosanoid biosynthesis, as understood in mammals. This outline represents our understanding of eicosanoid biosynthesis in insects and other invertebrates as well. Nonetheless, we can expect insects will express considerable departures from this model, particularly in biochemical details.

4.9.2.2.1. Phospholipase A_2 (PLA_2) is the first step in eicosanoid biosynthesis Eicosanoid biosynthesis is thought to begin with hydrolysis of AA from the *sn*-2 position of cellular phospholipids (PLs) (Dennis, 1994, 1997; Balsinde *et al*., 1999). PUFAs are generally associated with the *sn*-2 position of PLs, while saturated or monounsaturated fatty acids are associated with the *sn*-1 position. This asymmetry in the molecular localization of fatty acids within PLs was partly responsible for suggesting that PLA_2 might be a regulatory step in eicosanoid biosynthesis.

The number of recognized types of PLA_2s seem to steadily increase (Dennis, 1994, 1997; Balsinde *et al*., 1999), but there are two major PLA_2 classes. The secretory PLA_2s include those found in snake and arthropod venoms, those found in synovial fluids in some cases of pathology, and those associated with digestion. Most of these enzymes are low molecular mass (about 14 kDa) proteins stabilized by five to seven disulfide bridges. They require the presence of mM concentrations of Ca^{2+} for full catalytic activity. The other major class consists of the intracellular, or cytosolic PLA_2s ($cPLA_2$s). These are high molecular mass (>85 kDa) proteins that achieve full catalytic activity in the presence of μM calcium concentrations. Some of the $cPLA_2$s are of particular interest because they show a marked preference for PL substrate with AA in the *sn*-2 position. These enzymes are thought to mediate the first step in eicosanoid biosynthesis.

$cPLA_2$s have been purified from several mammalian sources (Dennis, 1994, 1997; Balsinde *et al*., 1999), including rat mesangial cells, human monocytes, human platelets, leukocytes, Swiss mouse 3T3 cells, glomerular mesangial cells, mouse keratinocytes, and mouse peritoneal macrophages. These enzymes are variously stimulated by proinflammatory cytokines, tumor necrosis factor, lipopolysaccharides (LPSs), and mitogens. The stimulations result in the translocation of the enzyme from the cytosol to the cellular membranes, where AA is selectively released from PLs. While the concentration of free AA is maintained at submicromolar levels by a reacylation pathway, stimulation of PLA_2 activity produces a rapid increase in free AA, which can then be available for eicosanoid biosynthesis.

The $cPLA_2$s are upregulated by several mechanisms. One is a calcium-dependent translocation to the membrane fractions of cells, as just mentioned. Alternatively, the enzymes are activated and deactivated inphosphorylation/dephosphorylation cycles. In other cases, a receptor-associated G protein activates $cPLA_2$s. Finally, some of these enzymes are activated by transcriptional activation, resulting in increased levels of $cPLA_2$ protein. The key point is that $cPLA_2$s represent the first step in eicosanoid biosynthesis.

Figure 1 provides an overview of the major eicosanoid biosynthetic pathways. The cyclooxygenase (COX) pathways yield the PGs and thromboxanes and the LOX pathways convert AA into various hydroperoxyeicosatetranoic acids (HPETEs) and hydroxyeicosatetranoic acids (HETEs). These species are themselves biologically active; they are also potential substrates for further metabolism to leukotrienes and still other biologically active products. The epoxygenases are cytochromes P450 (see **Chapter 4.1**), which yield the various epoxyeicosatrienoic acids.

4.9.2.2.2. The cyclooxygenase pathways PGs are C_{20} carboxylic acids with a five-membered ring variously substituted at C9 and C11, and two aliphatic chains featuring a substitution at C15 and one, two, or three double bonds. Three PUFAs, $C_{20:3n-6}$, $C_{20:4n-6}$, and $C_{20:5n-3}$, are potential substrates for the COX pathways, although AA is the commonest substrate among mammals. The PGs always have two fewer double bonds than their parental PUFAs, giving rise to the 1-, 2-, and 3-series PGs (Figure 2). PGs are defined by the substitutions at C9 and C11. PGE, for example, features a keto function at C9 and a hydroxyl function at C11. Specific PGs are identified by combining the number and letter designations. PGE_1, to continue the example, features one double bond at C13. The parental fatty acid is $C_{20:3n-6}$.

Figure 3 indicates that PG biosynthesis requires three enzyme steps. First is the COX step, which catalyzes the bis-oxygenation of AA to form the endoperoxide PGG_2. Second, the endoperoxide

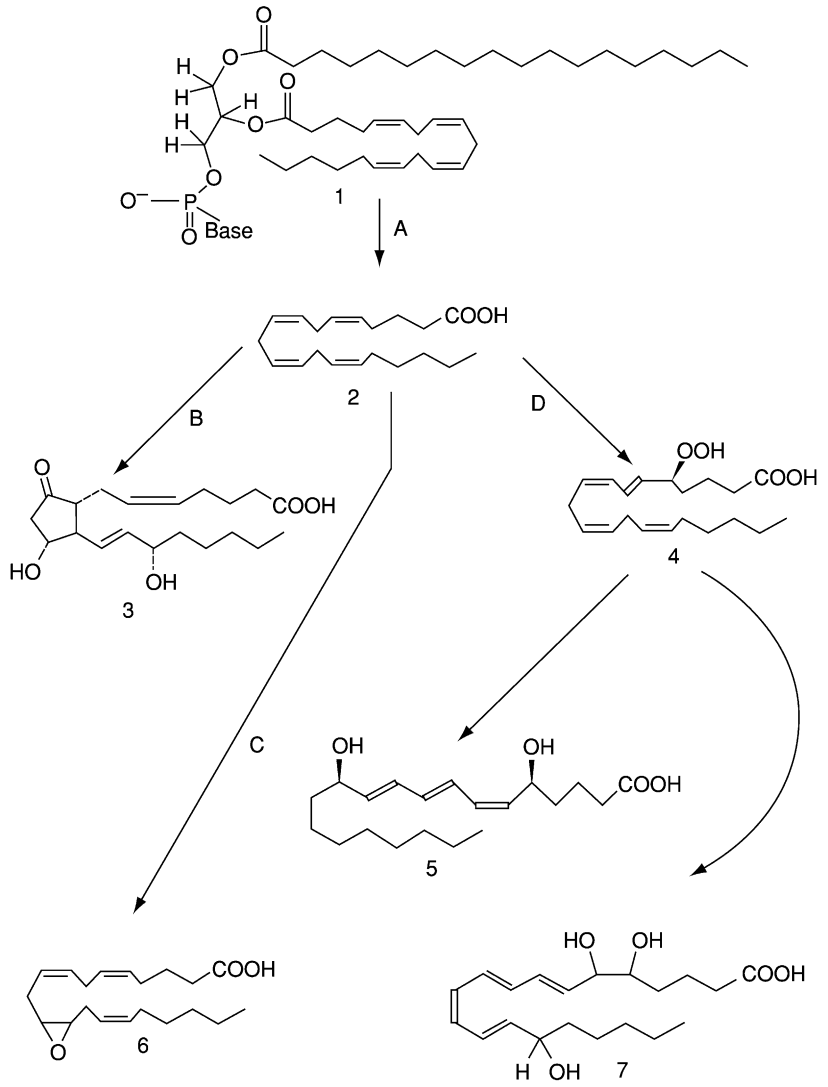


Figure 1 A sketch of arachidonic acid (AA) metabolism based on the biomedical background. Three fatty acids, $C_{20:3n-6}$, AA, and $C_{20:5n-3}$ are potential substrates for eicosanoid biosynthesis. Chemical structures are denoted by numerals: 1, cellular phospholipid; 2, hydrolyzed AA; 3, prostaglandin E_2; 4, 5-hydroperoxyeicosatetraenoic acid; 5, leukotriene B_4; 6, 11,12-epoxyeicosatrienoic acid; 7, lipoxin A. Capital letters indicate major enzyme systems responsible for eicosanoid biosynthesis: A, phospholipase A_2; B, cyclooxygenase (COX); C, cytochrome P450 epoxygenase; D, lipoxygenase (LOX). (Reproduced with permission from Stanley, D.W., **2000**. Eicosanoids in Invertebrate Signal Transduction Systems. Princeton University Press, Princeton, NJ; © Princeton University Press.)

undergoes a two-electron reduction at C15, catalyzed by a peroxidase activity. The peroxidase activity yields PGH_2, from which other biologically active PGs are produced. The COX and peroxidase activities are juxtaposed in a single protein, known for many years as PG endoperoxide synthase, or PGH synthase. It is now termed COX.

A couple of interesting features of COX help us understand PG biosynthesis. For one, the COX undergoes "suicide" inactivation, thought to be an intrinsic property of the enzyme. This was first recognized because COX is inactivated before all available substrate is converted into product, in mammalian cells after about 1300–1400 catalytic operations (Smith and Marnett, 1991; Smith *et al*., 1991). The suicide step is a feature of the COX component of the protein because the peroxidase activity remains after the COX has faded. Suicide inactivation may set an upper limit on cellular capacity for PG biosynthesis.

A second interesting feature of mammalian COXs is the intracellular localization of the protein. COX is a glycoprotein associated with membrane fractions, mainly endoplasmic reticulum and to some extent the nuclear membrane. The enzyme was originally thought to feature transmembrane domains,

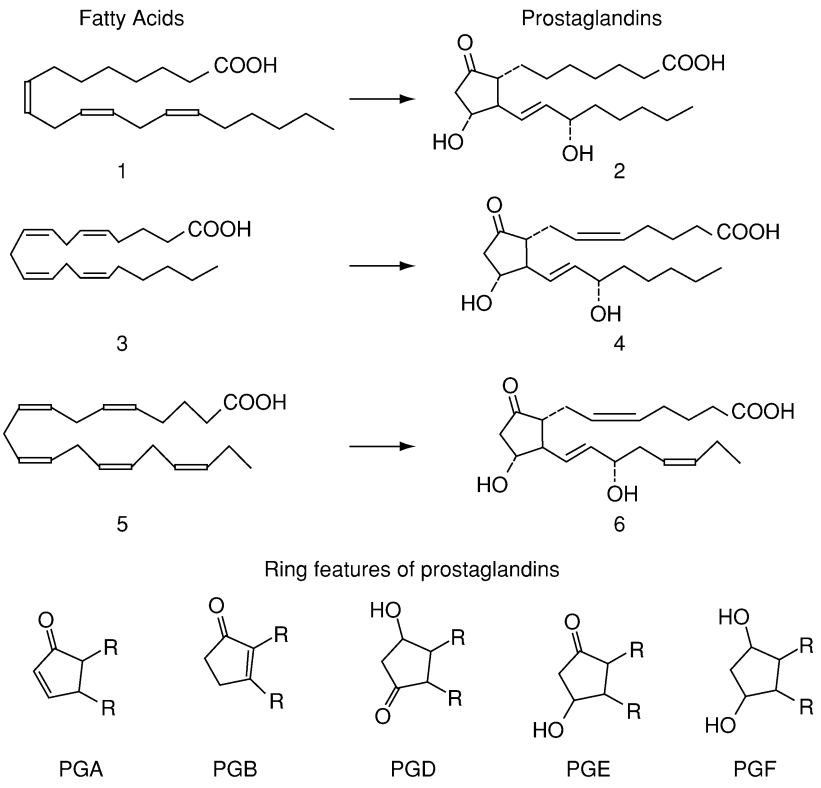


Figure 2 Cyclooxygenase (COX) converts certain C_{20} PUFAs into their respective 1-, 2-, and 3-series prostaglandin (PG) products. Each PG series features two fewer double bonds than its corresponding precursor PUFA. 1, $C_{20:3n-6}$; 2, PGE_1; 3, $C_{20:4n-6}$; 4, PGE_2; 5, $C_{20:5n-3}$; 6, PGE_3. The lower panel presents the ring features of five prostaglandins, where R represents the aliphatic chains that appear on the complete structures. (Reproduced with permission from Stanley, D.W., **2000**. Eicosanoids in Invertebrate Signal Transduction Systems. Princeton University Press, Princeton, NJ; © Princeton University Press.)

but current thinking places the enzyme entirely within the lumen of the endoplasmic reticulum and nuclear membranes (Otto and Smith, 1995).

This model held sway until early 1991 when it became clear that some cells express another form of COX. The two forms are now called COX-1 and COX-2. All the preceding remarks are based on our understanding of COX-1, which is thought to serve as a housekeeping isozyme. That is, it is responsible for biosynthesizing PGs active in physiological homeostasis. This enzyme is constitutively expressed in most mammalian tissues, although not in all cells within a tissue. COX-1 is thought to release PGH_2 into the cytosol where it is converted by other enzymes into biologically active PGs. The active PGs may act within the cell or may exit the cells, probably assisted by a PG transporter (Kanai *et al.*, 1995).

COX-2 mediates the biosynthesis of PGs for inflammatory processes, ovulation, and mitogenesis. In contrast to COX-1, COX-2 is not expressed in most mammalian cells (Smith *et al.*, 1996). However, it can be induced rapidly in many cells, including fibroblasts, endothelial cells, monocytes, and ovarian follicles. COX-2 is an inducible, rather than constitutive enzyme, and its expression is increased tremendously, from 10- to 80-fold by proinflammatory or mitogenetic factors, such as cytokines and tumor-promoting phorbol esters. While also associated with the membrane fractions of cells, COX-2 is mainly associated with the luminal surface of nuclear membranes. It may release PGH_2 into the nucleus, and the PGH_2 or another PG derived from it may interact with nuclear proteins to influence gene expression. Hence, COX-1 and COX-2 may represent two unrelated pools of active enzyme within the same cell, each with separate biological functions (Otto and Smith, 1995). While previously noted, solid information on the forms of the enzymes responsible for PG biosynthesis in invertebrates is not yet available.

Recently, the possibility of additional COX forms has arisen, surfaced by efforts to understand the pharmacology of acetaminophen (Tylenol in the USA and Paracetamol in the UK). While acetaminophen is an effective analgesic and antipyretic, it is a poor antiinflammatory drug and weak inhibitor of COX-1 and COX-2. The possibility of a COX-3 was first raised by Simmons and his colleagues

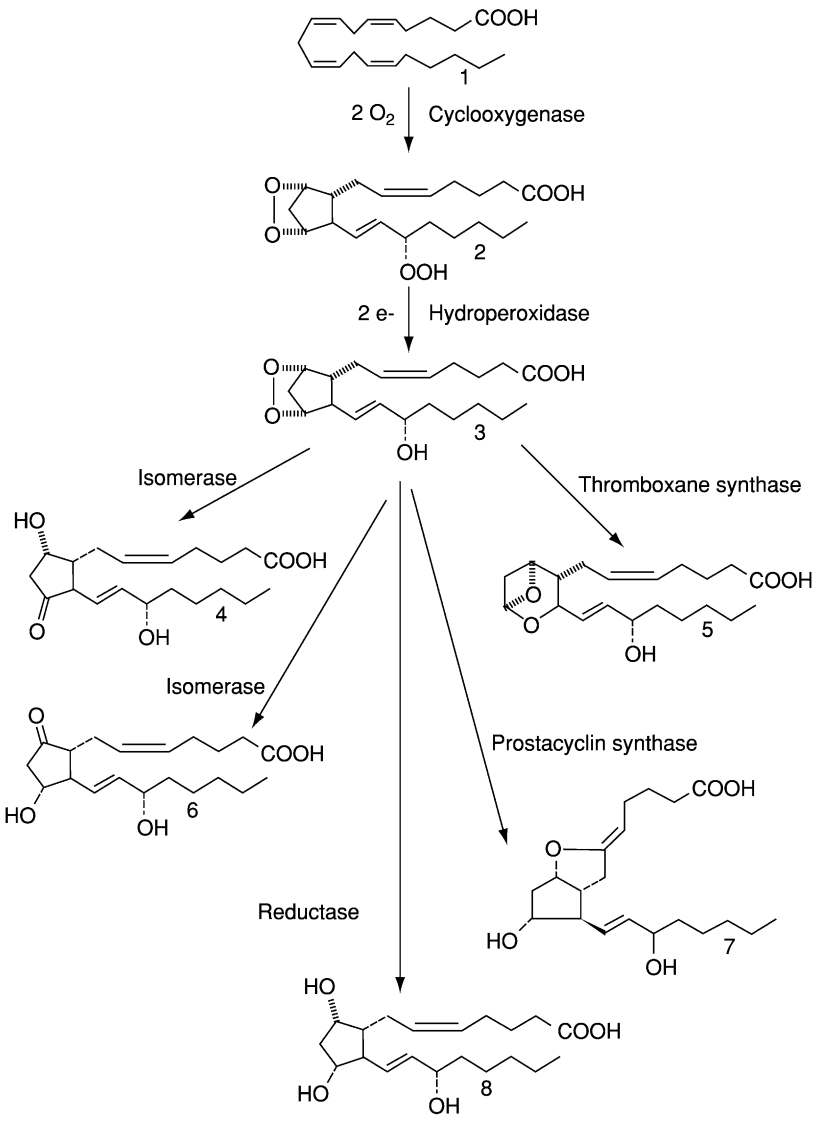


Figure 3 Prostaglandin (PG) biosynthesis requires three enzymatic steps. First, AA (structure 1) is oxygenated to the unstable endoperoxide, PGG_2 (structure 2). Second, a hydroperoxidase step reduces PGG_2 to the more stable PGH_2 (structure 3). Both of these enzymatic steps are catalyzed within a single protein, COX. PGH_2 is converted into the classical prostaglandins, including PGE_2 (4), thromboxane B_2 (5), PGD_2 (6), PGI_2 (prostacyclin, structure 7), and $PGF_{2\alpha}$ (8) by actions of cell-specific enzymes, as indicated. (Reproduced with permission from Stanley, D.W., **2000**. Eicosanoids in Invertebrate Signal Transduction Systems. Princeton University Press, Princeton, NJ; © Princeton University Press.)

(Simmons *et al.*, 1999), who found that treating a murine macrophage line (J774.2) with any of a variety of nonsteroidal antiinflammatory drugs induced expression of COX-2. The key finding was that, within the same cell line, the COX-2 induced by antiinflammatory drugs was more sensitive to inhibition with acetaminophen than COX-2 induced by LPSs. The authors postulated two enzyme activities with different properties. In another paper the authors suggested a COX-3, which is expressed as a variant of the COX-1 gene (Chandrasekharan *et al.*, 2002). The story is further complicated by the suggestion that COX-3 is a variant of COX-2 (Botting, 2002). However, in all cases, the new COX expressed increased sensitivity to acetaminophen. This work opens the possibility of multiple isoforms of COX, all derived from just two separate genes (Warner and Mitchell, 2002) and also for the eventual discovery of rather specific therapeutics which act on specific tissues. Again, we have no information on COX isoforms in insects, although studies in this area may contribute important new knowledge on the enzymes responsible for PG biosynthesis.

Figure 3 also shows the third enzymatic step in the biosynthesis of PGs. PGH_2 is a substrate for several

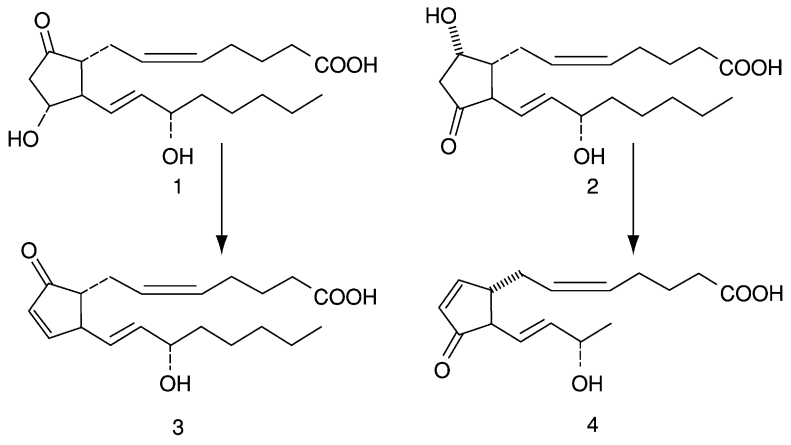


Figure 4 Two prostaglandins (PGs) are formed by nonenzymatic rearrangements of other PGs, rather than by enzymatic treatment of PGH_2. PGA_2 (3) is formed from PGE_2 (1). Δ^{12}-PGJ_2 (4) is formed by a nonenzymatic dehydration of PGD_2 (2). (Reproduced with permission from Stanley, D.W., **2000**. Eicosanoids in Invertebrate Signal Transduction Systems. Princeton University Press, Princeton, NJ; © Princeton University Press.)

enzymes responsible for converting PGH_2 into the active PGs. The PGs D, E, and F are formed by three groups of enzymes, respectively, the PGD synthases, the PGE synthases, and the PGF synthases, (Urade *et al.*, 1995). Two other PGs are formed directly from PGH_2, thromboxane A_2 and prostacyclin, also called PGI_2 (**Figure 3**). Both of these compounds act in platelet aggregation, among other things. Thromboxase strongly induces platelet aggregation while PGI_2 exerts the opposite action. PGI_2 synthase is responsible for converting PGH_2 into PGI_2, and this enzyme is also a member of the cytochrome P450 superfamily (Tanabe and Ullrich, 1995). The structures of thromboxane A_2 and PGI_2 are a little different from the other PGs, and for this reason we see the term "prostanoid" used to describe all COX products.

Two other PGs, Δ^{12}-PGJ_2 and PGA_2 (**Figure 4**), are not formed directly from PGH_2 (Negishi *et al.*, 1995a). PGA_2 is produced through a nonenzymatic rearrangement of PGE_2. PGJ_2 is a nonenzymatic dehydration product of PGD_2, which is converted to the Δ^{12}-PGJ_2 in the presence of serum albumin. Most PGs, including PGE_2, $PGF_{2\alpha}$, PGD_2, PGI_2, and thromboxane A_2, express their actions through specific receptors located on cell surfaces. PGA_2 and Δ^{12}-PGJ_2 operate through a different mechanism. These PGs are thought to be actively moved into cells via a transporter protein. An intracellular carrier molecule facilitates transport into the nucleus, where the PGs bind with thiol groups of nuclear proteins. The PG–protein complex then interacts with DNA, resulting in the expression of genes. Among their other biological actions, the A and J series PGs induce expression of genes for heatshock proteins in many normal and tumor cells.

4.9.2.2.3. The lipoxygenase (LOX) pathways Polymorphonuclear leukocytes are central to many mammalian inflammatory events. Samuelsson and his colleagues used these cells to investigate the possibility that AA could be metabolized into products other than PGs. They discovered that the main AA oxygenation pathways in leukocytes yielded 5-HETE and a series of products later called leukotrienes (Oates, 1982; Samuelsson, 1983).

The mammalian AA LOXs produce a series of six products (**Figure 5**). These are 5-, 8-, 9-, 11-, 12-, and 15-HPETEs. Each of the LOXs is named according to the carbon that is oxygenated. Thus, 5-LOX yields 5-HPETE, and so forth (Pace-Asciak and Asotra, 1989). The LOX products are quickly taken into various metabolic systems, which yield other biologically active products. In one of the most common fates, the hydroperoxide products are quickly reduced to the corresponding HETE by various glutathione peroxidases, which are abundant in most mammalian cells. The HETEs are biologically active (Spector *et al.*, 1988). For example, two products, 5- and 12-HETE, induce degranulation of human neutrophils. Similarly, 12-HETE is a potent chemoattractant for polymorphonuclear leukocytes. Several HETEs are active in various pathophysiological events, including proinflammatory processes.

One of the LOX products, 5-HPETE, serves as substrate for biosynthesis of the LTs (**Figure 6**). The root LT is LTA_4, an unstable epoxide of 5-HPETE. The enzyme responsible for this is called LTA_4 synthase. The 5-LOXs from human leukocytes and mouse mast cells also express LTA_4 synthase activity, and many workers believe both enzymatic steps are carried out by the same bifunctional enzyme.

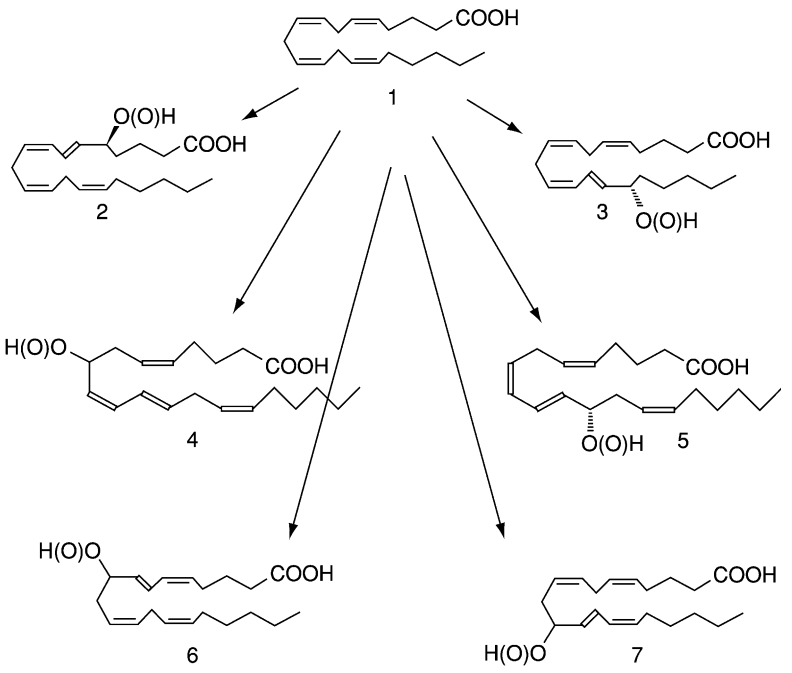


Figure 5 Lipoxygenase (LOX) is responsible for converting AA (1) into various hydroperoxyeicosatetraenoic acids (HPETEs) and hydroxyeicosatetraenoic acids (HETEs). Each of these compounds are formed by specific LOXs, identified by the positional specificity of the introduced oxygen. For example, 5-LOX yields 5-HPETE (2). The hydroperoxy acids may be reduced to their corresponding hydroxy acids by glutathione peroxidases. The remaining structures are identified by the position of the introduced oxygen: (3), 15-HPETE and 15-HETE; (4), 8-HPETE and 8-HETE; (5), 12-HPETE and 12-HETE; (6), 9-HPETE and 9-hydroxyeicosatetraenoic acid; (7), 11-HPETE and 11-HETE. (Reproduced with permission from Stanley, D.W., **2000**. Eicosanoids in Invertebrate Signal Transduction Systems. Princeton University Press, Princeton, NJ; © Princeton University Press.)

LTA_4 is an unstable substance whose main function is to serve as a substrate for the biosynthesis of other LTs. There are two main pathways (Pace-Asciak and Asotra, 1989). The first is catalyzed by LTA_4 hydrolase, which yields LTB_4 (**Figure 6**). This LT is a proinflammatory mediator of host defense reactions in mammals. It activates polymorphonuclear leukocytes, myeloid cells, and mast cells. LTB_4 also induces neutrophils to adhere to endothelial cell walls (Metters, 1995).

Alternatively, LTA_4 can be modified into the cysteinyl LTs or peptidoLTs (**Figure 6**). LTC_4 synthetase catalyzes addition of glutathione, in covalent linkage, to C6, yielding LTC_4. This LT can be converted to LTD_4 by a single transpeptidase step which catalyzes hydrolysis of the terminal amino acid residue from LTC_4. LTD_4 undergoes another transformation to LTE_4 by hydrolyzing the glycine residue from LTD_4. This step is catalyzed by a dipeptidase. LTF_4 can be formed by adding an amino acid residue.

The cysteinyl LTs make up the slow-reacting substance of anaphylaxis (Samuelsson, 1983). The major biological action of these compounds is contraction of smooth muscles associated with respiratory and vascular systems, and with the alimentary canals of mammals. These actions are mediated through specific receptors, of which two types are known (Metters, 1995). These are designated cys-LT_1 and cys-LT_2 receptors. Unlike the PG receptors, which exhibit marked specificity for each PG, the receptors for the cysteinyl LTs are much less fastidious. Both types of receptors have equal affinity for LTC_4 and LTD_4 and less affinity for LTE_4. We have no information on receptors for LTF_4. The LTs have not yet been considered in insect physiology.

Beside the HPETEs, HETEs, and LTs, the LOX pathways can yield another suite of compounds, none of which is known from invertebrates. Similarly, products of the epoxygenase pathway mentioned in **Figure 1** have not been discovered for invertebrates. These are considered in detail elsewhere (Stanley, 2000).

Gerwick (1993) noted that the term eicosanoid is limited to oxygenated metabolites of a limited group of fatty acids, specifically C_{20} PUFAs. He suggested that a new term, "oxylipin," was required to serve as a broader term for all oxygenated compounds formed from fatty acids of any chain length by reactions involving at least one step of a monooxygenase- or dioxygenase-dependent oxygenation. This broad word informs our appreciation of many fatty acid-derived products in various animal systems. Another

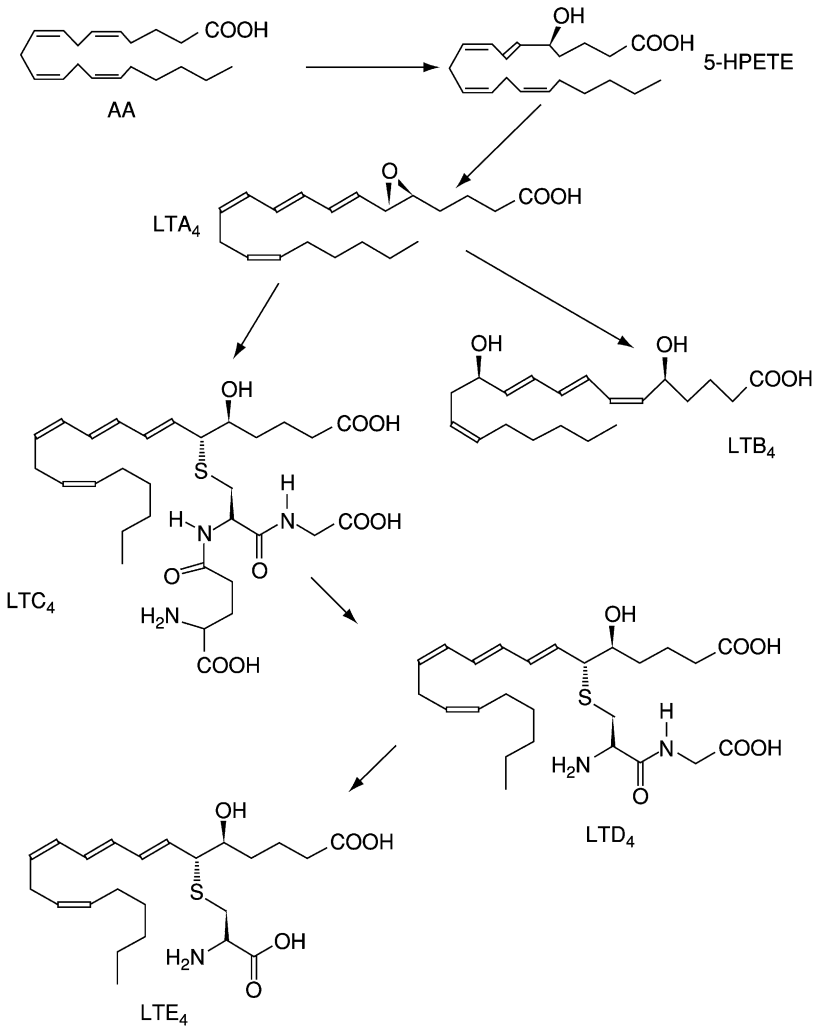


Figure 6 The biosynthesis and structures of leukotrienes. The pathways begin with conversion of AA to 5-HPETE by a 5-LOX. Leukotriene A_4 synthase yields the unstable epoxide leukotriene A_4 (LTA_4). LTA_4 may serve as substrate for leukotriene biosynthesis. In a one-step pathway, leukotriene hydrolyase yields leukotriene B_4 (LTB_4). The alternative pathway yields peptidoleukotrienes: leukotrienes C_4 (LTC_4), D_4 (LTD_4), and E_4 (LDE_4). (Reproduced with permission from Stanley, D.W., **2000**. Eicosanoids in Invertebrate Signal Transduction Systems. Princeton University Press, Princeton, NJ; © Princeton University Press.)

term, phytooxylipins, similarly describes a very wide array of oxygenated fatty acids that serve important roles in plant defense reactions (Blee, 1998). We may infer that the various forms of oxygenated fatty acids are crucial mediators in the life histories of most organisms. One area of most fundamental importance is reproduction, discussed in the next section.

4.9.3. Reproduction

Appreciation of the biological significance of eicosanoids began with the discovery that a substance in unfractionated human semen causes contractions of uterine smooth muscle (Kurzrok and Lieb, 1930). Although we now know eicosanoids occur and exert biological actions in virtually every mammalian tissue and body fluid, their influence on uterine muscle contraction, a function vital in reproductive biology, marks the discovery of the first known eicosanoid action.

Release of egg-laying behavior in newly mated house crickets, *Acheto domesticus*, another action in reproductive biology, also marks discovery of the first known biological action of PGs in insects and other invertebrates (Destephano and Brady, 1977). Although only coincidental, it is quite intriguing because at a superficial glance, it appears that PGs mediate events in reproduction of mammals and a very distantly related invertebrate species. The occurrence and actions of PGs in insect reproductive systems have been reviewed (Stanley-Samuelson and Loher, 1986; Stanley-Samuelson, 1994a; Miller and Stanley, 1998a; Stanley 2000). These reviews

remain useful because this is an area in which new information has been slow coming. This section presents an abbreviated synopsis of known information.

The most detailed studies on the role of PGs in releasing egg-laying behavior come from the work of Werner Loher on the Australian field cricket, *Teleogryllus commodus*. Loher and his colleagues (Loher *et al*., 1981; Stanley-Samuelson *et al*., 1987) developed an "enzyme-transfer" model to account for transfer of a still-unidentified form of COX (along with its substrate, AA) from males to females via spermatophores. This model seems to hold for some, but certainly not all, cricket species. PGs may release egg-laying behavior in newly mated silkmoths, *Bombyx mori*, although some of the data on this species are not entirely convincing (Yamaja Setty and Ramaiah, 1979, 1980).

PGs also may release egg-laying behavior in a few other insect species. Buprofezin is an insect growth regulator, which functions by inhibiting chitin biosynthesis and cuticle deposition in larvae of some insect species. Izawa *et al*. (1986) found that this compound inhibited egg-laying behavior in adults of the 28-spotted ladybird, *Henosepilachna vigintioctopunctata*. Similarly, Uchida *et al*. (1987) discovered that buprofezin inhibited egg-laying behavior and PG biosynthesis in adults of the brown rice planthopper, *Nilaparvata lugens*. The PG biosynthesis inhibiting action of the growth regulator suggested that PGs may release egg-laying behavior in these species.

The occurrence among insect species of PG-mediated egg-laying behaviors cannot be estimated in a reliable way. The model seems to apply to a few orthopterans, a lepidopteran, a coleopteran, and a hemipteran. Several points suggest that this is not a general model for insects. First, there are several insect species in which PG titers increase in female reproductive tracts after mating, but they do not release egg-laying behavior. Second, the role of PGs in releasing egg-laying behavior needs to be regarded in the context of individual mating systems (Thornhill and Alcock, 1983). In many species, eggs are formed, fertilized, and oviposited in separate time frames, and mating, *per se*, does not result in immediate egg-laying activities. Moreover, among those species in which mating does release egg-laying behavior, there are several mechanisms of releasing egg-laying behavior not involving PGs or other eicosanoids. Mechanical stimulation of egg-laying behavior is one example.

PGs or PG biosynthetic activity are transferred from males to females of several insect species in which PGs do not influence egg-laying behavior (Stanley, 2000). Our question is: what is the meaning of this information? In general terms, one could speculate that the PGs act in modulating and coordinating various details in reproductive physiology. The recent work by Medeiros and colleagues provides a specific example (Medeiros *et al*., 2002). As seen in other insect species, the ovaries of *Rhodnius prolixus* incorporate vitellogenin as well as several other specific proteins (see **Chapter 3.9**). Medeiros *et al*. (2002) tested the idea that eicosanoids influence the ovarian incorporation of *Rhodnius* heme-binding protein (RHBP). Ovaries of vitellogenic females were incubated in the presence of iodinated RHBP and either PGE_2 or indomethacin, a COX inhibitor. Their results indicated that incubations in the presence of PGE_2 downregulated RHBP uptake by up to 35%, while incubations in the presence of indomethacin upregulated RHBP uptake by up to 50%, both in comparison to control incubations. The authors also showed that the *Rhodnius* ovaries secreted PGE_2 into the culture medium and that the amount of secreted PGE_2 was reduced significantly in incubations conducted in the presence of indomethacin. Although the indomethacin did not influence the surface area nor the patency of follicle cells, indomethacin treatments did influence dephosphorylation of two ovarian proteins, an 18 kDa and a 25 kDa protein. Medeiros *et al*. (2002) concluded that eicosanoids influence the reproductive physiology of *Rhodnius* ovaries. Phosphorylation signal transduction pathways may be involved in PG modulation of RHBP uptake. The significance of this paper lies in its direct illumination of a previously unrecognized role of PGs in insect reproduction. Perhaps more important, however, the paper illustrates the point that many more-or-less subtle physiological steps in the overall process of producing progeny may be influenced by PGs and other eicosanoids.

Aside from the roles of PGs in releasing egg-laying behavior in a few insect species and in modulating RHBP endocytosis in *Rhodnius* ovaries, eicosanoids act in various other aspects in reproduction of many invertebrates and lower invertebrates. This work on animals outside of Insecta has been treated elsewhere (Stanley-Samuelson, 1994a; Miller and Stanley, 1998; Stanley, 2000).

It is worth noting that eicosanoids influence reproduction by acting at several levels of biological organization, as mentioned earlier for mammals. Actions such as modulating insect ovarian endocytosis of a protein (Medeiros *et al*., 2002) and many other eicosanoid actions in invertebrate reproduction (Stanley, 2000) are expressed at the cellular level, i.e., the eicosanoids influence cellular events. Other actions, including release of the egg-laying behavioral program are registered at the organismal level, even though they are probably driven by PG

actions at the level of a few select neurons (Loher *et al.*, 1981). Some eicosanoid actions take place at the population level. Sorensen *et al.* (1988), for example, reported that PGs of the F series act as the postovulatory pheromone in goldfish. There remains much to be discovered in the roles of eicosanoids in insect reproductive biology.

4.9.4. Eicosanoids in Ion Transport Physiology

It is generally thought that cellular life evolved in aqueous environments. Cells live in more or less aqueous environments, and virtually all cells have evolved physiological mechanisms to maintain homeostasis of ion and water balance. Terrestrial insects tend to lose water to their environments, and they have mechanisms to restrict and offset water losses (Kirschner, 1991). Freshwater animals tend to maintain their extracellular fluid compartments hyperosmotic to their environments. Still other animals live in water subject to rapid changes in osmotic concentration, such as estuaries and rocky pools in intertidal zones. The osmotic homeostasis of all animals is subject to frequent challenges at the cellular and organismal levels.

Many homeostatic mechanisms are expressed at the organismal level. The respiratory surfaces of all terrestrial animals, including insects, are internalized. The internalization of these surfaces is regarded as a major adaptation to terrestrial life. Terrestrial insects have a small layer of hydrocarbons and other lipids on their integument which helps reduce water loss. Some insects drink water to offset losses while other species can absorb water from humid atmospheres. Finally, some insects are able to sustain extreme dehydration. Larvae of the midge *Polypedilum vanderplanki*, which live in ephemeral pools on rocks in Africa, provide the best example of this. These pools evaporate in the dry season, and the larvae can tolerate nearly complete dehydration until the following rainy season (Hinton, 1960). Cellular mechanisms for maintaining water balance are integrated into organismal-level reactions to osmotic challenge. Vertebrate kidneys and the Malpighian tubules of insects and certain other invertebrates, for example, are influenced by diuretic and antidiuretic hormones which act to eliminate or conserve body water or solutes.

To appreciate eicosanoid actions in ion transport, a brief overview of some mechanisms of solute transport is given (Alberts *et al.*, 1994). The lipid bilayers of cellular membranes are fairly permeable to water and relatively impermeable to biologically important ions, such as chloride, sodium, potassium, magnesium, and calcium. These ions are transported across membranes by members of two classes of membrane-associated proteins, channels and carrier proteins. Channels are transmembrane proteins that form hydrophilic pores through membranes. The pores are generally selective, allowing specific ions, such as potassium or calcium, to pass through. Carrier proteins also bind specific solutes, which are transported due to conformational changes in the carriers. All channel-mediated and some carrier-mediated transport actions are driven solely by local electrochemical gradients known as passive transport or facilitated diffusion. Some carrier proteins are associated with an energy source, usually an ATPase. These carriers sometimes transport solutes against steep electrochemical gradients. Other carrier proteins act as coupled transporters in which a transport protein is responsible for movement of one solute with simultaneous movement of another. Symports move the two solutes in the same direction; antiports move the two solutes in opposite directions. For an example, the ubiquitous sodium–potassium antiport is responsible for actively pumping of sodium ions out of cells and potassium ions into cells. Passive and active transport mechanisms also are responsible for moving water across membranes. This brief glimpse allows us to regard the homeostasis of water and solute concentrations at the organismal and cellular levels as the outcome of regulated actions of specific intracellular, and often intramembrane, proteins. Eicosanoids are among the molecules which modulate homeostasis of water and solute concentrations.

As usual, most of our knowledge of the roles of eicosanoids in the physiology of water and solute homeostasis comes from the studies on vertebrate systems, particularly various kidney and toad preparations. Nephrons are the operative cells in kidney water and solute transport, and eicosanoid biosynthesis and actions vary along the nephrons, as detailed by Bonvalet *et al.* (1987).

Dalton (1977a, 1977b) provided the first recognition that eicosanoids act in the physiology of fluid secretion or ion transport in invertebrates. Fluid secretion by salivary glands isolated from the blow fly *Calliphora erythrocephala* can be stimulated by treating the glands with serotonin, a biogenic amine. In this system, the serotonin acts as an external hormone, and it is the natural ligand for salivary gland cell surface receptors. Serotonin-receptor interactions stimulate increased concentrations of intracellular cAMP in the salivary gland preparations (Berridge, 1970; Prince *et al.*, 1972) and these increased intracellular cAMP concentrations lead to increased secretion of an isosmotic potassium-rich

fluid. Dalton (1977a) investigated the influence of PGE_1 on salivary gland physiology, showing that treating isolated salivary glands with low doses (10^{-7}–10^{-9} M) of PGE_1 did not alter their basal fluid secretion rates. Alternatively, these low doses of PGE_1 attenuated the usual stimulatory influence of serotonin on fluid secretion rates. Dalton (1977b) suggested that PGE_1 reduces intracellular cAMP concentrations by downregulating adenylate cyclase, with no influence on phosphodiesterase. It appeared that at least one eicosanoid, PGE_1, plays an important physiological role in fluid secretion physiology in an invertebrate. In another early line of work with the locust rectum, Phillips and his colleagues suggested that eicosanoids may stimulate increased intracellular cAMP concentrations, which in turn lead to increased chloride transport (Phillips, 1980).

Petzel and Stanley-Samuelson (1992) suggested that PGs modulate basal fluid secretion rates in Malpighian tubules of female yellow fever mosquitoes, *Aedes aegypti*. This hypothesis was tested in a series of fluid secretion assays based on the Ramsey protocol (Petzel, 1993). In each experiment Malpighian tubules were incubated in the presence of physiological buffer. After an equilibration period, half of the buffer was exchanged with the same volume of buffer containing an inhibitor of eicosanoid biosynthesis. First, the influence of eicosatetraynoic acid (ETYA), an AA analog with triple bonds in place of the usual double bonds, was assessed. ETYA inhibits many enzymes which process AA, including PLA_2, LOX, and COX. In the presence of ETYA, basal fluid secretion decreased from about 1.0 nl min^{-1} in the pretreatment period to 0.55 nl min^{-1} in the experimental period. These findings supported the idea that PGs are involved in modulating basal fluid secretion rates in mosquito Malpighian tubules.

Results of similar experiments with selective LOX and expoxygenase inhibitors indicated that products of these two pathways do not influence basal fluid secretion rates in Malpighian tubule preparations. Experiments with the COX inhibitor indomethacin, however, showed that PGs modulate basal fluid secretion rates. In the presence of 100 µM indomethacin, fluid secretion decreased from 0.9 to 0.5 nl min^{-1} in a dose-dependent manner. It was inferred from this work that PGs, but not LOX or epoxygenase products, are involved in maintaining basal fluid secretion rates in mosquito Malpighian tubules (Petzel and Stanley-Samuelson, 1992).

The idea that PGs act in mosquito Malpighian tubules raised the issue of the occurrence and metabolism of AA in these tissues. Petzel *et al.* (1993) determined the presence of AA and PGE_2 in mosquito Malpighian tubules. They used immunohistochemistry to detect PGs (Howard *et al.*, 1992; Petzel *et al.*, 1993). The repeated dark-brown staining pattern showed the presence of PGE_2 in principal, but not stellate, cells of the tubule. The principal cells are responsible for secreting fluid into the lumen of the tubule. A similar staining pattern was observed for PGE_2 in Malpighian tubules of the yellow mealworm, *Tenebrio molitor* (Howard *et al.*, 1992). $PGF_{2\alpha}$ is also present in Malpighian tubules from the mosquito and the mealworm, shown by similar histological procedures. The $PGF_{2\alpha}$ staining patterns are quite different from the patterns obtained for PGE_2. For $PGF_{2\alpha}$, the staining is not restricted to the principal cells, but seems to be more or less evenly distributed among both major Malpighian tubule cell types.

Hypothesizing that PGs act through G protein-coupled receptors, Petzel also conducted a preliminary investigation of the possibility that PGs stimulate increased cAMP biosynthesis in mosquito Malpighian tubules (D.H. Petzel, personal communication). The tubules were isolated from adult female mosquitoes, then incubated in mosquito saline. PGE_2 was added to the saline and after selected incubation periods the tubules were frozen in liquid nitrogen and cAMP concentrations in the tubules were determined. This work showed that PGE_2 stimulated a four- to fivefold increase in intracellular cAMP concentrations. Petzel also recorded the influence of PGE_2 on fluid secretion rates. In these experiments, tubules were equilibrated as usual in mosquito saline, then PGE_2 was added to the saline. PGE_2 stimulated increased fluid secretion in mosquito Malpighian tubules, suggesting that PGs are among the regulatory elements in mosquito Malpighian tubule physiology. Van Kerkhove *et al.* (1995) conducted a similar line of work on Malpighian tubules from workers of the ant *Formica polyctena*. These findings are similar to the data with mosquito Malpighian tubules and support the notion that PGs act in insect Malpighian tubule physiology.

Phillips (1980) implicated PGs in the physiology of the locust rectum. Radallah *et al.* (1995) continued this work using an everted sac preparation of the locust rectum to assess the influence of AA and PGE_2 on water resorption. The everted sac is formed by tying one end of the rectum onto a catheter tube. Test compounds could then be injected into the sac, thereby exposing the hemolymph side of the preparation to the compounds. PGE_2 stimulated increased fluid resorption in a dose-dependent manner. Maximal stimulation, 59% greater than controls, was obtained at about 10^{-9} M PGE_2. They could also stimulate increased fluid resorption by adding AA, which resulted in a 79% increase in resorption at

10^{-6} M AA. Similar experiments with aspirin and indomethacin also stimulated dose-dependent increases in fluid resorption. These data suggest that PGs and inhibitors of PG biosynthesis exert similar influences on the rectal preparations. Several possibilities help resolve this apparent paradox. For one, indomethacin restricts biosynthesis of all COX products, some of which exert inhibitory influences on cellular processes. The possible inhibitory PG may exert greater influence on overall fluid transport dynamics than the stimulatory PG. If so, inhibiting the biosynthesis of a possible inhibitory PG may create an apparent stimulation of fluid transport. The idea of considering possible inhibitory and stimulatory PG actions in the same system has not been explored.

Radallah *et al.* (1995) determined the influence of AA and PGE_2 on selected intracellular signal transduction messengers. To record the influence of AA, PGE_2, and indomethacin on the influx of calcium into rectal cells they used a microfluorimetric technique. PGE_2 and AA stimulated large increases in calcium influx, which decreased to baseline within minutes. As in the fluid transport experiments, indomethacin similarly provoked increased calcium influx. Pretreating the tissues with nifedipine, which blocks L-type calcium channels, completely blocked the effects of AA and PGE_2 on calcium transport. This paper also reported that treating the locust rectum with AA or PGE_2 stimulated substantial increases in phospholipase C (PLC) activity, which indirectly influences calcium transport (Berridge, 1993). Once again, aspirin and indomethacin similarly induced increased PLC activity. This apparently depends on the influx of calcium because an L-type calcium channel blocker strongly attenuated the influences of AA and PGE_2 on PLC activity. They also showed that the PLC activity is associated with the release of inositol phosphates.

Radallah *et al.* (1995) present a convincing postulate that AA and at least one eicosanoid, PGE_2, are involved in modulating fluid transport in the locust rectum. The locust rectum is probably under primary regulation of the antidiuretic hormone, neuroparsin, which stimulates fluid resorption in the rectum. The hormone apparently acts through a G protein-coupled receptor, and it stimulates increased PLC activity. But it seems that the hormone also stimulates the release of AA from cellular PLs, which leads to increased PGE_2. The PGE_2 probably acts by an autocoidal mechanism to coordinate the cellular reactions to the central hormone. If so, PGE_2 acts similarly in the insect rectum as in segments of the mammalian kidney (Bonvalet *et al.*, 1987).

PGs also modulate fluid secretion rates in salivary glands of the lone star tick, *Amblyomma americanum* (Qian *et al.*, 1997). Tick salivary glands are osmoregulatory organs during host–parasite interactions. The salivary glands concentrate the nutrients associated with vertebrate blood and in the process they form a copious salt-rich saliva which is injected back into their hosts during feeding. The salivary glands are under nervous control and the neurotransmitter dopamine stimulates fluid secretion in isolated salivary glands (Sauer *et al.*, 1995). Qian *et al.* (1997) investigated the influence of PGs on dopamine-stimulated fluid secretion. Compared to controls, treatments with the PLA_2 inhibitor oleyloxyethyl phosphorylcholine (OOPC) and the COX inhibitor aspirin resulted in about 30–40% reductions in dopamine-stimulated fluid secretion rates. Longer incubations produced greater reductions and the influence of both inhibitors was expressed in a dose-dependent manner. The influence of OOPC was reversed by incubating the inhibitor-treated glands with 100 µM PGE_2 or its stable analog 17-phenyl trinor PGE_2. The PG treatments did not reverse the inhibitory influence of the COX inhibitors aspirin and diclofenac.

Dopamine stimulates fluid secretion through G protein-coupled receptors resulting in increased intracellular cAMP concentrations (Sauer *et al.*, 1995). Qian *et al.* (1997) determined the influence of PG biosynthesis inhibitors and of PGs on salivary gland cAMP concentrations. They found that the PLA_2 inhibitor OOPC and the COX inhibitor indomethacin inhibited dopamine-stimulated increases in intracellular cAMP concentrations by about 25%. In the presence of OOPC, PGE_2 and its analog stimulated 20–40% increases in the OOPC-treated cAMP concentrations.

The eicosanoid system influenced neither fluid secretion nor intracellular cAMP concentrations in the absence of dopamine stimulation. The authors' results provide strong support for their hypothesis that eicosanoids modulate dopamine-stimulated fluid secretion rates in tick salivary glands. Moreover, this work marks the first identification of physiologically functional PG receptors in an invertebrate system. Their data from binding studies support the view that the salivary gland receptor is functionally coupled to a stimulatory G protein. The authors concluded that the tick salivary gland expresses a functional PGE_2 receptor that does not directly regulate adenylate cyclase. They speculated that the PGE_2 receptor may influence calcium mobilization, which they confirmed in a subsequent paper (Qian *et al.*, 1998). The biological significance of the salivary gland PGE_2 receptor is linked

to secretion, or exocytosis, of anticoagulant proteins, which are thought to facilitate blood-feeding. Qian *et al.* (1998) demonstrated this by incubating dispersed salivary gland tissue in the presence of PGE_2, after which they recorded increased release of anticoagulant proteins.

Taken together, work on blowfly salivary glands, locust rectum, insect Malpighian tubules, and tick salivary glands provide quite convincing evidence for the biological actions of eicosanoids in water and solute transport in representatives of arthropods. Eicosanoids similarly act in other invertebrates, lower vertebrates, and mammals (Stanley, 2000). It is suggested that regulating the actions of specific proteins involved in ion transport is a fundamental biological action of PGs and other eicosanoids in animals.

4.9.5. Insect Immunity

4.9.5.1. Introduction

The ability of organisms to defend themselves from parasites and pathogens is a fundamental aspect of biology. Many microbes produce and secrete enzymes and toxins that provide a measure of protection from other microbes. Plants elaborate sophisticated defense reactions to wounding as well as invasions by bacteria and fungi. Metazoan animals thrive in constant contact with pathogenic and nonpathogenic microbes, fungi, and parasites. An animal's ability to defend itself from microbial invasions is known as immunity.

Immunologists recognize separable categories of immune reactions, innate immunity and acquired (or adaptive) immunity. Innate immunity is made up of several protective mechanisms that function in a nonspecific way. In mammals and other higher vertebrates, clonal biosynthesis of antibodies to specific invading antigens is a main line of acquired, or adaptive, immunity. Invertebrates lack lymphocytes and immunoglobulins, and therefore do not produce specific antibody reactions to infections. On this basis invertebrates are often said, incorrectly, to lack immune systems. Rowley (1996) registered this incongruity, noting that some of the very important early experiments in immunology were conducted on invertebrates.

Invertebrates and vertebrates express several forms of passive immunity, so named because they do not entail directed protective actions. These might include killing ingested organisms due to relatively harsh pH conditions (either strongly acid as in the mammalian stomach or very alkaline as in many insect midguts) in the alimentary canal. Foodborne microbes may be killed by hydrolytic actions of digestive enzymes. The integument serves as a very effective barrier to invading microbes.

Animals also express innate immunity to infections. Armstrong *et al.* (1996) recognized several expressions of innate immunity, some of which are seen in vertebrates and in invertebrates. Inducible antibacterial peptides of insects, defensins in insects and mammals, and the α_2-macroglobulins, known in mammals and some arthropods, are elements of innate immunity.

The innate immunity of insects is often divided into two major categories, cellular and humoral immunity. Cellular immunity involves direct interactions between circulating hemocytes and invading organisms (reviews: Brehelin, 1986; Dunn, 1986; Gupta, 1986, 1991; Boman and Hultmark, 1987; Lackie, 1988; Strand and Pech, 1995; Gillespie *et al.*, 1997; Levine and Strand, 2002). Cellular defense reactions include phagocytosis, a form of endocytosis; nodulation, entrapping bacterial cells into clusters of hemocytes; and encapsulation of organisms too large for phagocytosis, such as eggs of parasitoid insects. Following capture, the internalized microorganisms are destroyed by intracellular killing mechanisms. Formation of oxygen radicals and nitric oxide are killing mechanisms within invertebrate cells (Conte and Ottaviani, 1995). These immune reactions create a general inflammatory response to injury, infection, and parasitization.

Insect humoral immunity involves induced biosynthesis of antibacterial proteins (Tauszig *et al.*, 2000; Hoffmann and Reichart, 2002). This very important area of inquiry has revealed a large number of antibacterial proteins and has generated new understanding of fundamental aspects of biology, including gene regulation. The chapter by Hultmark is the best entry point to literature on humoral immunity.

Insects respond to large bacterial infections by nodule formation, the predominant response to bacterial infections (Dunn and Drake, 1983; Horohov and Dunn, 1983). Nodulation begins with entrapment of bacterial cells by granule-containing hemocytes. The granule-containing cells undergo degranulation, which releases proinflammatory chemicals. The nodulation process is completed by attaching layers of flattened phagocytes to the mature nodule. The last stage is a melanization process, leaving darkened, easily visible nodules attached to the inner sides of the body wall or various organs. Invading bacterial cells are topologically removed from circulation by effectively forming an impermeable wall between the bacterial mass and the remainder of the organism.

Nodulation can be regarded as a form of cellular encapsulation, seen in response to infection by organisms larger than bacterial cells, such as parasites and other foreign materials. Wounds to the integument also evoke encapsulation reactions. As in nodulation, encapsulation begins with release of proinflammatory chemicals by degranulation of granule-containing cells. Again, the ending stages involve attaching layers of phagocytes, which flatten and melanize the capsule.

4.9.5.2. Eicosanoids in Insect Immunity

The inflammatory defense reactions just described have been studied quite extensively in many invertebrate species, particularly arthropods (Levine and Strand, 2002), although we have relatively little information on the biochemical molecules that mediate inflammation in invertebrates. Some mediators known from mammalian host defense reactions operate in invertebrates. Proinflammatory cytokines, or cytokine-like molecules, exist in invertebrates, including a bivalve mollusc, *Mytilus edulis*, a sea star, *Asterias forbesi*, and a urochordate, *Styele clava*, all reviewed by Rowley (1996). Downer and his colleagues showed that biogenic amines stimulate hemocytic inflammatory reactions in cockroaches (Baines *et al*., 1992; Baines and Downer, 1994) and waxmoths (Dunphy and Downer, 1994; Diehl-Jones *et al*., 1996). Similarly, it appears that eicosanoids, to which we now turn attention, also serve as crucial mediators of insect inflammatory reactions to bacterial infections.

4.9.5.2.1. Immune reactions to bacterial infection
Eicosanoids exert many influences, in some cases stimulatory and in others inhibitory, on mammalian host defense systems. Based on this background, Stanley-Samuelson considered the possible actions of eicosanoids in insect immunity (Stanley-Samuelson *et al*., 1991). The results of these experiments, all using fifth instar *Manduca sexta* larvae, strongly supported the idea that eicosanoids mediate one or more cellular reactions to bacterial infections (Stanley-Samuelson *et al*., 1991).

This paper prompted the question: which of the hemocytic defense reactions to bacterial infections are mediated by eicosanoids? Stanley and his colleagues hypothesized that eicosanoids mediate insect nodulation responses to bacterial infections (Miller *et al*., 1994). Nodulation can be assessed by counting numbers of melanized nodules within insect hemocoels following infection (described in detail by Miller and Stanley, 1998b). It was inferred from the results of these experiments that eicosanoids mediate one or more of the early steps in the nodulation reaction to bacterial infection (Miller *et al*., 1994).

These two investigations with tobacco hornworms were the first experiments on eicosanoid actions in invertebrate immune systems. The assays for microaggregation and nodulation reactions are simple, and they facilitate investigation of a broader hypothesis. Specifically, do eicosanoids similarly act in nodulation reactions in other insect species? This question was tested in a series of similar exercises with other species. These include the tenebrionid beetle, *Zophobas atratus* (Miller *et al*., 1996), black cutworms, *Agrotis ipsilon*, and true armyworms, *Pseudaletia unipuncta* (Jurenka *et al*., 1997), the silkworm, *B. mori* (Stanley-Samuelson *et al*., 1997), and caterpillars of the butterfly, *Colias eurytheme* (Stanley *et al*., 1999). The results of all the experiments with these species supported the hypothesis. Similar findings emerged from work with adults of the cricket *Gryllus assimilis* (Miller *et al*., 1999), adult cockroaches *Periplaneta americana* (Tunaz and Stanley, 1999) and adult 17-year periodical cicadas, *Magicicada septendecim* and *M. cassini* (Tunaz *et al*., 1999). Although these species do not establish an exhaustive representation of the Class Insecta, they make up a sufficient sampling of insects to suggest that PGs and other eicosanoids are key mediators of insect cellular immunity.

Beyond the work described here, several other laboratories have investigated the influence of eicosanoids in insect immunity. Mandato *et al*. (1997) carried out a detailed investigation of eicosanoid actions in three discrete cellular processes within the overall nodulation reaction of the waxmoth larvae, *Galleria mellonella*. They found that eicosanoids mediate phagocytosis, cell spreading, and prophenyloxidase (PPO) activation in waxmoth larvae. Their work on PPO activation is quite interesting, because other groups found that eicosanoids do not influence PPO activation. Morishima *et al*. (1997) suggested a completely new role for eicosanoids in insect immunity. As mentioned earlier, the immune reactions of insects and other invertebrates to bacterial infections include humoral and cellular responses. The humoral responses include induced synthesis of antibacterial proteins, including cecropins and lysozymes. Cecropins are not found in the hemolymph in unchallenged insects, and the gene for this protein is somehow activated upon bacterial infection. Lysozyme occurs at low, constitutive levels in hemolymph, and expression of the lysozyme gene is upregulated following stimulation with bacterial cells or components of the bacterial cells. Morishima *et al*. (1997) suggested that eicosanoids mediate induction of the genes for cecropin

and lysozyme in fat body of silkworm, *B. mori*. This work marks the recognition of a newly discovered eicosanoid action in invertebrate immunity, made all the more interesting in light of the recent discovery of a functional coupling between eicosanoid biosynthetic pathways and the immune deficiency (imd) pathway in *Drosophila* (Yajima *et al.*, 2003). These authors reported that treatments with either of the two PLA_2 inhibitors, dexamethasone or *p*-bromophenacyl bromide, inhibited activation of the imd pathway. The inhibitory effects of these two inhibitors were attenuated by additional treatments with eicosanoid biosynthesis precursor fatty acids. Yajima *et al.* (2003) also found that AA alone did not activate the imd pathway, which indicated that eicosanoids participate in the activation of the imd pathway, but requires further LPS stimulation.

4.9.5.2.2. Immune reactions to pathogenic fungal infection All the work just mentioned is based on insect cellular defense reactions to bacterial infections. Dean *et al.* (2002) broadened the scope of this work with their hypothesis that eicosanoids mediate insect cellular reactions to the fungal pathogen *Metarhizium anisopliae*. They showed that treating tobacco hornworms with dexamethasone prior to challenge with fungal spores resulted in reduced number of nodules in reaction to the spores and that the influence of the drug could be reversed by treating experimental hornworms with AA. This was the first suggestion that eicosanoids mediate insect cellular reactions to fungal challenge. Lord *et al.* (2002) carried out a similar investigation using the fungal pathogen *Beauveria bassiana*. They reported that treating experimental tobacco hornworms with dexamethasone or the LOX inhibitors caffeic acid or esculetin or the COX inhibitor ibuprofen resulted in substantial reduction in number of nodules formed in reaction to fungal challenge. They also conducted a series of revealing rescue experiments. They found that the dexamethasone effect on nodulation could be reversed using the LOX product 5-HPETE, but not with the COX product PGH_2. Moreover, the influence of caffeic acid and esculetin was reversed with 5-HPETE but the ibuprofen effect was not changed by treating experimental hornworms with PGH_2. The authors inferred that products of the LOX pathways, but not the COX pathways, act in mediating hornworm immune reactions to *B. bassiana*. Lord *et al.* (2002) also noted that eicosanoids did not influence PPO activation, as seen in waxmoth larvae (Mandato *et al.*, 1997). These two findings with different insect fungal pathogens add important new information to understanding signaling mechanisms in insect cellular immunity.

4.9.5.2.3. Immune reactions to parasitoid eggs Carton *et al.* (2002) added a new insight into insect immunity with their report that dexamethasone treatments inhibited cellular reactions to parasitoid invasion. In these experiments, larvae of the fruitfly *Drosophila melanogaster* were injected with dexamethasone, then exposed to adults of the parasitoid wasp *Leptopilina boulardi*. Compared to control larvae, in which approximately 90% of the parasitized larvae encapsulated the wasp eggs, about one-third of the larvae treated with 5 mg of dexamethasone encapsulated eggs and about 14% of larvae treated with 8 mg of dexamethasone encapsulated eggs. These results demonstrated a substantial dose-dependent effect of dexamethasone on encapsulation of parasitoid eggs, from which the authors inferred that eicosanoids mediate cellular encapsulation reactions to parasitoid eggs.

The information reviewed here may suggest that eicosanoids mediate cellular immune reactions to bacterial, fungal, and parasitoid challenge. It is suggested that the roles of eicosanoids in immunity may be another fundamental eicosanoid action in animals. More to the point of understanding insect immunity, however, it has been suggested that the immune system of *D. melanogaster* discriminates between bacterial and fungal challenge, seen with respect to differential activation of signal transduction pathways, which lead to expression of genes for antimicrobial peptides (Lemaitre *et al.*, 1997). In light of this, it is noted that *M. sexta* cellular reactions to challenge by one fungal pathogen apparently depend on LOX, rather than COX, pathways, while reactions to bacterial challenge may be mediated by products of both pathways. As seen with the pathways for expression of genes for antimicrobial proteins, fungal and bacterial challenges seem to activate different pathways of eicosanoid biosynthesis.

4.9.5.2.4. Interactions between insect immune and neuroendocrine systems It was noted earlier that the signal transduction systems in insect cellular immunity include other important biomolecules, including peptides and biogenic amines. Beyond this, recent evidence points to interactions between the immune and endocrine systems in insect cellular immune reactions to infection. Wiesner *et al.* (1997) reported that apolipophorin-III stimulates immune reactions in *G. mellonella*. Halwani and Dunphy (1999) also noted that apolipophorin-III potentiates immune reactions in *G. mellonella*. Goldsworthy *et al.* (2003) considered another element of the neuroendocrine system in their study of adipokinetic hormone-I, apolipophorin-III, and eicosanoids in the locust,

Locusta migratoria. They found that adipokinetic hormone and eicosanoids are important in signaling nodulation reactions while adipokinetic hormone and apolipophoren-III are crucial for stimulating PPO activation. In line with the report by Lord *et al.* (2002), eicosanoids do not seem to influence PPO activation in locusts. As noted elsewhere, there is room for many mediators in insect cellular immunity.

Cellular reactions to immune challenge involve an unknown, albeit conceivably large, number of discrete cellular events (Miller *et al.*, 1994, 1996; Stanley, 2000). Nodulation, in particular, undoubtedly involves many discrete cellular activities, only a few of which have been identified. While not meant to be an exhaustive list, these may include recognition of microbial cell wall components, producing and secreting various signal moieties which attract other hemocytes toward the site of infection, cell migration, various cell adhesion actions (microbe–hemocyte and hemocyte–hemocyte), cell spreading, and activation of prophenoloxidase. One of the major gains of investigating the roles of eicosanoids in cellular immunity is the possibility of recognizing additional cell actions and identifying which of many possible eicosanoids are responsible for signaling particular cell actions.

4.9.5.3. Testing the Eicosanoid Hypothesis

The idea that eicosanoids mediate cellular immune reactions to microbial and parasitoid challenge is a robust notion, which requires testing from a number of approaches. By and large, the evidence supporting the hypothesis has been circumstantial in nature. On one hand, we see that treating insects with pharmaceutical inhibitors of eicosanoid biosynthesis impairs cellular defense reactions to immune challenge. On the other hand, the ability of insect tissues, including immunity-conferring tissues, to produce PGs and other eicosanoids has been documented. However, this biochemical work does not, in itself, forge a link between eicosanoids and cellular immune signaling. In the following paragraphs, a few experiments designed to more directly test the eicosanoid hypothesis are reviewed.

The outcomes of experiments with inhibitors of eicosanoid biosynthesis uniformly indicate that insects which had been treated with the inhibitors were impaired in their ability to clear bacterial cells from hemolymph circulation and in their ability to form microaggregates and nodules following infection. If the eicosanoid hypothesis makes sense, the impairments are due to the inability of hemocytes or other tissues to biosynthesize eicosanoids in response to microbial challenge. It should follow, then, that microbial infections stimulate production of various eicosanoids.

Jurenka *et al.* (1999) investigated this possibility in a series of experiments with black cutworms. Cutworms were artificially infected by injecting heat-killed bacteria, *Serratia marcescens*, into their hemocoels. After 30 min incubations, hemolymph was collected and processed for eicosanoid extraction. The eicosanoids were derivatized with the fluorescent compound ADAM, then analyzed on high-performance liquid chromatography (HPLC) equipped with a flow-through fluorescence detector. This work revealed large increases in the titer of one PG, $PGF_{2\alpha}$, in hemolymph of experimental, but not vehicle-injected control insects. Moreover, pretreating experimental cutworms with selected COX inhibitors blocked the infection-stimulated increases in $PGF_{2\alpha}$. This work strongly supports the eicosanoid hypothesis because it indicates that insects increase PG production following bacterial challenge.

Miller and Stanley (2001) considered the issue of signaling among immune tissues by investigating microaggregation reactions to bacterial challenge in isolated hemocyte populations. Hemocytes were prepared from tobacco hornworms, then challenged with lyophilized bacteria, *S. marcescens*. After selected incubation periods, generally 2 h, number of microaggregates were determined. These experiments documented the point that isolated hemocytes preparations are able to form microaggregates when challenged with bacteria. Results of similar experiments in which hemocytes were pretreated with various inhibitors of eicosanoid biosynthesis indicated that the microaggregation reactions also depend on eicosanoids.

This work led to the suggestion that isolated hemocytes are competent to biosynthesize and secrete eicosanoids in reaction to bacterial challenge. This idea was confirmed by experiments with hemocyte culture media that had been conditioned by challenging the hemocytes with bacteria (Miller and Stanley, 2001). Hemocyte preparations were challenged with bacteria and the hemocyte media were filtered through microfilters to produce "conditioned medium." Naive hemocyte preparations were then treated with the conditioned medium, and hemocyte microaggregation was determined. These experiments showed that exposing naive hemocyte preparations to conditioned medium was sufficient to generate microaggregation reactions in the naive cells. In another control experiment, it was observed that media, which was conditioned using hemocytes, that have been pretreated with dexamethasone or other inhibitors of eicosanoid biosynthesis, did not generate microaggregation reactions.

However, this attractive view suggesting rapidly expanding discoveries of eicosanoid actions in

invertebrate immunity is best viewed with skepticism (Stanley-Samuelson, 1994b). Several important issues should be addressed before the biological actions of eicosanoids in immunity are accepted. First, as mentioned elsewhere (Stanley-Samuelson and Dadd, 1983; Stanley-Samuelson *et al.*, 1988; Stanley, 2000), the presence and significance of eicosanoid-precursor PUFAs is not well appreciated in insect biochemistry. Second, although there are several reports on the topic, there is very little information on eicosanoid biosynthesis in insects and other invertebrates. Third, the pharmacological fates of COX and LOX inhibitors in invertebrates is not known with certainty. All of these issues relate to various aspects of the biochemistry of eicosanoids, to which we now turn.

4.9.5.4. The Biochemistry of Eicosanoid Systems in Immune Tissues

The fat body and hemocytes are thought to be the immunity-conferring tissues of insects, and the eicosanoid systems in these tissues were of prime interest. Soon after the hypothesis that eicosanoids mediate clearance of invading bacterial cells from hemolymph circulation was tested, an investigation into the main points of eicosanoid biosynthesis in these two tissues from the tobacco hornworm, *M. sexta* was launched.

4.9.5.4.1. Polyunsaturated fatty acids The fatty acid compositions of total lipids and selected PL fractions prepared from hemocytes isolated from fifth instar tobacco hornworms were determined (Ogg *et al.*, 1991). For these experiments, hemolymph was prepared by pericardial puncture, to prevent inadvertent activation of hemocytes. The hemocytes were pelleted by centrifugation, washed to eliminate hemolymph contamination, then processed for lipid analysis. As expected (Stanley-Samuelson *et al.*, 1988), only traces (less than 0.1% of the fatty acids) of $C_{20:3n-6}$, AA, or $C_{20:5n-3}$ were associated with total hemocyte lipids, or with total hemocyte PLs. This finding opened the possibility that AA or other eicosanoid-precursor components are sequestered into particular PL fractions. Two major glycerophospholipids, phosphatidylcholine and phosphatidylethanolamine, were isolated. Analysis of the fatty acids associated with these PLs again revealed only traces of the eicosanoid-precursor components. With the idea that these low levels might be related to an environmental constraint, the artificial culture medium was analyzed, which revealed higher proportions of AA in the medium than in the hemocytes. Hence, it was concluded the presence of low proportions of AA in hornworm hemocytes is the usual biological condition (Ogg *et al.*, 1991).

Certainly there are exceptions to this notion. One of the long-standing pillars of the field of animal lipids is that PUFAs are mostly associated with PLs, and less so with neutral lipids. This is generally true for insects, as well (Stanley-Samuelson and Dadd, 1983), but such ideas mut be tested. The triacylglycerol fractions prepared from tobacco hornworm heads yielded the highest proportions of AA ever recorded in a lepidopteran, i.e., AA constituted about 12% of the triacylglyerol fatty acids (Ogg and Stanley-Samuelson, 1992). A few other such pools of AA and $C_{20:5n-3}$ have been recorded in insect studies, one of which certainly relates to PG biosynthesis. Stanley-Samuelson and Loher (1983) found AA in a special pool in spermatophores of the cricket *T. commodus*. The AA was present at about 24% of phosphatidylcholine fatty acids, and virtually absent from phosphatidylethanolamine fatty acids. The authors suggested that the AA was specially sequestered for transfer from male to female crickets during mating. Perhaps the high AA proportions in hornworm heads represents a pool of storage AA that can be drawn upon as needed. Obviously, this is a speculation which requires more concrete examination. More recently, Nor Aliza *et al.* (2001) reported the presence of high proportions of AA and $C_{20:5n-3}$ in tissue PLs of adult fireflies, *Photinus pyralis*. They recorded AA at about 25% of PLs fatty acids in light organ and fat body from males, 13% for testes, and about 8% for midgut epithelia. Although such high proportions might indicate a very high potential for PG biosynthesis, preliminary experiments indicated that firefly tissues produced PGs at rates no different from other studied insect systems.

Recording very low proportions of AA in *M. sexta*, fat body and hemocytes is consistent with the general picture of PUFAs in terrestrial insects. These low proportions elicit questions about insect fatty acid biochemistry. Do hemocytes actively maintain low proportions of AA and other eicosanoid-precursor PUFAs? This was investigated by tracing the incorporation and remodeling of radioactive fatty acids into hemocyte lipids (Gadelhak and Stanley-Samuelson, 1994).

Four radioactive fatty acids, $C_{18:1n-9}$, $C_{18:2n-6}$, $C_{20:4n-6}$, and $C_{20:5n-3}$, were used in separate incorporation experiments. Consistent with the general background of animal lipid biochemistry, the *Manduca* hemocytes incorporated all four of the radioactive fatty acids into cellular complex lipids. Of the four, $C_{18:2n-6}$ was most efficiently incorporated into PLs, while $C_{18:1n-9}$ and $C_{18:2n-6}$ were efficiently incorporated into triacylglycerols. About 1–3% of the starting radioactivity in $C_{18:1n-9}$ was recovered

in diacylglycerols and monoacylglycerols. Very little of the other fatty acids was incorporated into these two fractions. The incorporation patterns generally agreed with the fatty acid composition of hemocytes.

The incorporated fatty acids were redistributed among lipid fractions in hornworm hemocytes during longer incubation periods (Gadelhak and Stanley-Samuelson, 1994). This redistribution is due to selected hydrolysis of some components from PLs, allowing incorporation of other components. Chilton and Murphy (1986) documented this remodeling process in human neutrophils, showing that after initial incorporation into ester-linked PLs, radioactive AA was selectively remodeled into ether- and plasmalogen-linked PL pools over time. These PL fractions were not considered in our studies because they have not yet been sufficiently detailed in insect systems.

Remodeling of incorporated fatty acids from PLs to triacylglyerols was recorded (Gadelhak and Stanley-Samuelson, 1994). After longer incubations, the radioactivity recovered in the PL fraction declined, with concomitant increases in radioactivity in the triacylglycerols. For AA, the radioactivity recovered in PLs declined to about 97% at 20 min, and to about 83% by 120 min. This directional shift was not seen with incorporated $C_{18:2n-6}$ and $C_{18:1n-9}$, although some of the radioactivity associated with these fatty acids was shifted between these two major cellular lipid fractions.

A similar picture emerged from the analysis of selected PL fractions. After 5 min incubation periods, most of the radioactivity associated with AA was recovered in phosphatidylcholine. With longer incubations, the proportions of radioactivity recovered in phosphatidylcholine declined, with attending increased radioactivity in phosphatidylethanolamine. After 120 min incubations, slightly more radioactivity associated with AA was detected in phosphatidylethanolamine, rather than phosphatidylcholine. Again, these results support the idea that PUFAs are selectively remodeled among the complex lipid fractions in hornworm hemocytes. Similar analyses of other insect tissues have not yet been carried out, but nonetheless, remodeling dynamics of this nature are to be expected. Again, these processes help understand the low proportions of AA and other eicosanoid-precursor PUFAs in terrestrial insects.

AA is present in fat body from larvae of the beetle *Z. atratus* (Howard and Stanley-Samuelson, 1996; Miller *et al.*, 1996). Similarly, AA is present in trace levels in black cutworms, *A. ipsilon*, true armyworms, *P. unipuncta* (Jurenka *et al.*, 1997), in silkworms, *B. mori* (Stanley-Samuelson *et al.*, 1997), and in adults of the cricket *G. assimilis* (Miller *et al.*, 1999). Relative to the work on waxmoth hemocytes (Mandato *et al.*, 1997), AA and $C_{20:5n-3}$ were confirmed in these insects during early studies of fatty acid nutritional metabolism (Stanley-Samuelson and Dadd, 1984; Stanley-Samuelson *et al.*, 1988). Hence, one element required for eicosanoid biosynthesis in invertebrate immune tissues is present, albeit at low levels, in all systems in which eicosanoids are present.

4.9.5.4.2. Phospholipase A_2 is the first step in eicosanoid biosynthesis Two of the eicosanoid biosynthesis inhibitors used in the immune studies, namely dexamethasone and *p*-bromophenacylbromide, are thought to act at the level of PLA_2. These two reagents do not directly inhibit eicosanoid-biosynthetic enzymes, but inhibit the overall process of eicosanoid biosynthesis by arresting the release of AA or other PUFA from cellular PLs. In the absence of free substrate, eicosanoid biosynthesis cannot proceed. Because these two inhibitors attenuate cellular and humoral immune reactions to bacterial infections and the influence of one of these, dexamethasone, can be reversed by treating experimental insects with free substrate, AA, it was proposed that an intervening PLA_2 step must be operative in the eicosanoid-mediated immune reactions to bacterial infections.

The PLA_2s are responsible for hydrolyzing the acyl moiety from the *sn*-2 position of glycerophospholipids and an intracellular PLA_2 in *Manduca* fat body has been characterized (Uscian and Stanley-Samuelson, 1993). In brief, phosphatidylcholine with radioactive AA esterified at the *sn*-2 position was prepared in the form of substrate vesicles to facilitate enzyme activity. *Manduca* fat body homogenates were centrifuged to produce microsomal-enriched fractions used as enzyme sources. The enzyme was incubated with substrate vesicles, then total lipids were extracted from the reaction mixtures. The lipid extracts were separated and fractions associated with unprocessed substrate, free fatty acids, and diacylglycerol were radioassayed. PLA_2 activity was calculated from the amount of radioactivity associated with the free fatty acid fraction.

The fat body PLA_2 was sensitive to the usual biophysical parameters and the calcium requirements for this enzyme were assessed. For this experiment, the fat body was homogenized and centrifuged in calcium-free buffer containing the calcium chelator ethylene glycol bis (B-aminoethyl/ether)-*N*,*N*,*n*′, *N*′-tetraacetic acid (EGTA). Reactions run in the presence or absence of calcium yielded similar results, from which we inferred that the *Manduca*

fat body PLA_2 is a calcium-independent enzyme, or possibly one with very low stringency calcium requirements. To test this hypothesis, enzyme preparations were dialyzed against 200 volumes of EGTA buffer and enzyme activity was assessed. Again, rigorous removal of calcium did not reduce enzyme activity.

A similar characterization of a PLA_2 in hemocytes from fifth instar larvae was reported by Schleusener and Stanley-Samuelson (1996). For these experiments, hemolymph was collected by pericardial puncture, diluted in buffer, the hemocytes rinsed in fresh buffer, and the cells were homogenized by sonication. Microsomal-enriched fractions were prepared from the sonicates and these were used as the enzyme source.

Again, the hemocyte PLA_2 was sensitive to substrate concentration, protein concentration, and pH. As seen in virtually all mammalian preparations, the hemocyte enzyme was inhibited in the presence of 5–50 μM OOPC. The calcium requirements of the hemocyte PLA_2 were determined, and as in the fat body preparations, all hemocyte preparation steps were carried out in buffer containing EGTA. Enzyme activity increased with increasing concentrations of EGTA, which suggested that the chelator removes from the reaction medium an ion that downregulates the hemocyte PLA_2 (Schleusener and Stanley-Samuelson, 1996).

Based on findings with some mammalian $cPLA_2$s, the idea that the hemocyte PLA_2 exhibits a selectivity for arachidonyl-associated PLs was considered. In these experiments, enzyme sources were prepared as usual, then incubated with either of the two substrates, one a palmitoyl-associated phosphatidylcholine and the other an arachidonyl-linked phosphatidylcholine. Enzyme activity with the palmitoyl-associated substrate was reduced by about 40% relative to activity with the arachidonyl-associated substrate. It is more appropriate to assess substrate specificity using purified, or at least highly enriched, enzyme preparations since most animal cells have several forms of cytosolic PLA_2, and experiments meant to detect the substrate preferences of a subset of these enzymes can yield ambiguous results. In this case, however, the 40% reduction in enzyme activity indicates that at least one PLA_2 in *Manduca* hemocytes has a fairly strong preference for arachidonyl-linked PL substrate. This enzyme may regulate eicosanoid biosynthesis in hemocytes.

These preliminary investigations serve to document both the presence of a cytosolic PLA_2 in hornworm fat body and the fact that hemocytes can hydrolyze AA from the *sn*-2 position of cellular PLs. These results are preliminary because documenting the presence of an enzyme in fat body and hemocytes does not demonstrate unequivocally that the enzyme is an integral part of the eicosanoid-mediated cellular reactions to bacterial infections, although recent work supports the idea that PLA_2 is a key step in eicosanoid biosynthesis following bacterial infection. The circumstantial evidence on the influence of PLA_2 inhibitors on cellular and humoral immune reactions has been bolstered with more direct experiments. Given the very low levels of AA in *Manduca* immune tissues, some of these experiments could be better designed using other invertebrate systems. For now, however, it can be said that the tobacco hornworm expresses $cPLA_2$s, and some forms of these enzymes may be involved in immune reactions to bacterial infections.

4.9.5.4.3. Eicosanoid biosynthesis in immune tissues While eicosanoid biosynthesis has been demonstrated in a number of invertebrate systems, the idea that eicosanoids mediate cellular and humoral immune reactions to bacterial infections in insects opens a window of opportunity for investigation. Eicosanoid biosynthesis in *Manduca* fat body and hemocyte preparations has been recorded, but before considering this work, an important issue should be considered. The methodology of tracing PG biosynthesis in invertebrates has not yet matured to a series of standard techniques. Although most methods in essence amount to reacting an enzyme source with radioactive AA, there are a few subtle points which deserve attention because they can influence the outcome of such experiments.

In work with the fat body (Stanley-Samuelson and Ogg, 1994), the eicosanoid biosynthesis reaction mixtures included radioactive AA emulsified in buffer containing a cofactor cocktail (2.4 mM reduced glutathione, 0.25 mM hydroquinone, and 25 μg hemoglobin in each reaction). The reactions were preceded by 3 min preincubations with all reaction components except the enzyme sources. The reactions were started by adding the microsomal-enriched enzyme source, and stopped by acidification and extraction of products.

The radioactive PGs were separated by thin-layer chromatography (TLC) (Hurst *et al.*, 1987). Chromatography with this system differs from usual procedures because the plates are placed into the solvent immediately after the solvent is transferred to the developing chamber. There is no equilibration period in this system, which requires about 2.5 h to develop. In some experiments, the reaction products were also separated by two-dimensional TLC. Bands corresponding to authentic standards were transferred to vials, and the radioactivity in each fraction

was determined. PG biosynthesis was calculated from the amount of radioactivity in each fraction.

PUFAs are fairly unstable molecules in oxygen atmospheres and experimental artifacts can result from spontaneous oxygenation reactions, or autoxidation, during reactions and subsequent work up steps. Autoxidation was determined by conducting "zero-time" reactions. In these control experiments, the substrate was preincubated in parallel with the experimental tubes. At time zero, when the enzyme source would ordinarily be added, the control preincubations were stopped and extraction and chromatographic separation followed. Zero-time reactions were done in every experiment, and values from these control experiments were used to correct for autoxidation in the biosynthesis reactions.

The fat body preparations yielded four PGs: $PGF_{2\alpha}$, PGE_2, PGD_2, and PGA_2. Under most experimental conditions, PGA_2 was the predominant fat body product. The enzyme preparation was sensitive to routine biophysical parameters although the literature on PG biosynthesis in invertebrates presents an ambiguous picture with respect to optimal reaction times. Drawing on the mammalian data (Smith *et al*., 1996), the first step in PG biosynthesis is catalyzed by COX, which undergoes "suicide" inactivation after about 1400 catalytic operations. The suicide inactivation is regarded as an autoregulatory mechanism, which imposes an upper limit on cellular potential to biosynthesize PGs. PG biosynthesis therefore occurs in short, very rapid bursts of enzyme activity. However, the overall dynamics of maintaining appropriate PG titers within cells includes PG biosynthesis and degradation. This could produce two phases of PG biosynthesis in *in vitro* reactions and possibly in intact cells. Early in the reactions, the rapid bursts of biosynthesis would exceed PG degradation, thereby favoring product accumulation. During the second phase of the reaction, the suicide inactivation of COX would lead to decreased PG biosynthesis and product degradation would exceed product formation. The overall result of this asymmetry in the reaction progress would be registered as higher product accumulation in shorter reaction periods which decreases during longer reaction periods.

In work with the *Manduca* fat body, an initial burst of PG biosynthesis which peaked at about 1 min was observed. Thereafter, product accumulation decreased over the next 9 min. These findings agree with mammalian data, and also with the time course of PG biosynthesis seen in housefly preparations. Wakayama *et al*. (1986) also observed rapid PG biosynthesis during 2 min incubations. After the first 2 min, there was a very gradual increase in product formation over the following 58 min. In contrast, Brenner and Bernasconi (1989) reported a linear increase in PGE_2 biosynthesis over a 60 min time course. In their characterization of PG biosynthesis by spermatophore content from males of the Australian field cricket, *T. commodus*, Tobe and Loher (1983) also found a linear increase in synthesis over 60 min reaction periods. The systems in which longer reaction times promote increased product formation may point to important differences between the well-established mammalian studies and the emerging information on invertebrate eicosanoid systems.

All of the biological investigations into the roles of eicosanoids in cellular and humoral immunity described in this section were based on the use of pharmacological inhibitors of eicosanoid biosynthesis. Some of these inhibitors designed for studies on mammalian tissues reduce PG biosynthesis in hornworm tissues. In these experiments, the fat body preparations were incubated in the presence of selected doses of the pharmacological reagents, and the reaction products were then extracted and analyzed. The fat body preparations were very sensitive to the COX inhibitors, indomethacin and naproxin. At the low dosage of 0.1 μM, indomethacin and naproxin reduced total PG biosynthesis by about 80%. Higher dosages virtually abolished PG biosynthesis. Along with inhibition of PG biosynthesis, reactions with naproxin yielded substantial levels of a radioactive product, tentatively identified as the LOX product 15-HETE. The influence of naproxin on increased LOX activity occurred in a dose-dependent manner. In the presence of increasing naproxin concentrations, the decreasing COX activity was accompanied by increasing LOX activity. While indomethacin effectively inhibited COX activity, there was no accompanying increase in LOX activity. It may be presumed that the indomethacin also inhibits the insect LOX, as it does in some mammalian preparations.

Tobacco hornworm hemocyte preparations also are capable of eicosanoid biosynthesis (Gadelhak *et al*., 1995). Pools of hemocytes were washed, then homogenized by sonification, microsomal-enriched preparations were prepared by centrifugation, and eicosanoid biosynthesis assayed. The hemocyte preparations yielded two major products, the COX product PGA_2 and the LOX product 15-HETE. The identification of 15-HETE is based on a single TLC step, and undoubtedly that fraction contains more than one LOX product. For this reason, this product is simply noted as total LOX activity.

Two inhibitors of eicosanoid biosynthesis influenced eicosanoid biosynthesis by the hemocyte

preparations. The COX inhibitor naproxin reduced COX activity by 88% at 0.1 mM and by 96% at 1.0 mM. Contrary to results with the fat body, the hemocyte LOX activity was not enhanced in the presence of naproxin. Reactions carried out in the presence of the LOX inhibitor esculetin also yielded reduced COX and LOX product formation. At 0.1 mM, esculetin nearly inhibited all eicosanoid biosynthesis. These studies support the view that eicosanoid biosynthesis observed on TLC is not an artifact of substrate autoxidation. However, the influence of these eicosanoid-biosynthesis inhibitors is quite instructive. While these are well-characterized pharmacological agents in mammalian systems, their actions in invertebrate systems may differ in important ways. For example, indomethacin, a COX inhibitor, potently inhibited cellular defense reactions and COX activity in *Manduca* fat body preparations. This compound did not inhibit PG biosynthesis in preparations of male reproductive tracts from the cricket *A. domesticus* (Destephano *et al*., 1976). More to the point, naproxin is accepted as a specific COX inhibitor. Results with the hornworm fat body preparation would suggest that this is so in invertebrates as well, because low doses of naproxin (0.1 μM) effectively inhibited COX, but not LOX, activity. These results suggest that the influence of eicosanoid-biosynthesis inhibitors should be investigated in each species, and each tissue within a species, to ensure that the inhibitors act as they do in vertebrate experiments.

This comment is quite relevant to work on immunity. On the basis of a single experiment with naproxin, one might conclude that COX activity is essential to cellular immune reactions to bacterial challenge. However, the results of our biochemical experiments with hemocyte preparations make it clear that naproxin may act by inhibiting LOX as well as COX activities. This is one of the reasons for conducting experiments with several separate inhibitors.

Eicosanoid biosynthesizing enzymes are rather uniformly distributed within mammalian cellular fractions, with COXs exclusively associated with endomembrane fractions of cells. The subcellular localization of COX and LOX activities in hemocytes can be assessed by determining eicosanoid biosynthesis in the mitochondrial, microsomal, and cytosolic fractions. The COX activity was unevenly distributed among the three cellular fractions, about 5% in the mitochondrial fraction, 58% in the microsomal fraction, and about 36% in the cytosolic fraction. Most of the LOX activity (87%) was recovered in the cytosolic fraction, and the remaining 13% in the microsomal fraction. These data indicate that insect COXs do not have the same subcellular distribution as one finds in mammals.

Although the broad themes are similar, there is substantial variation in eicosanoid systems among mammalian species and among tissues within a mammalian species. Relative to invertebrates, much more variation will emerge as research activities which continue to generate new information. Therefore, it is important to investigate at least some aspects of eicosanoid biosynthesis in several insect species in which eicosanoids play a role. AA and other C_{20} PUFAs as well as PG biosynthesis capacity is present in tissues from the tenebrionid beetle *Z. atratus* (Miller *et al*., 1996). Similarly, AA is present in fat body PLs from true armyworms and black cutworms (Jurenka *et al*., 1997). The fat body from these larvae also expressed an intracellular PLA_2 that can hydrolyze AA from cellular PLs. Fat body preparations from both species are able to convert radioactive AA into PGA_2, PGD_2, PGE_2, and $PGF_{2\alpha}$. A similar line of documentation experiments showed the presence of eicosanoid-biosynthesizing enzymes in fat body from silkworms, *B. mori* (Stanley-Samuelson *et al*., 1997). Miller *et al*. (1999) recorded traces of AA in the fat body of adult crickets, *G. assimilis*, as well as biosynthesis of three eicosanoids by fat body preparations, PGA_2, PGE_2, and a hydroxyeicosatetraenoic acid. As in the silkworm fat body preparations, the LOX product(s) was the major product produced by the cricket fat body preparations.

The point of these exercises is to document the presence of an eicosanoid-biosynthesizing system in insects thought to use eicosanoids in cellular immune signal transduction mechanisms. Even at this superficial level of analysis, however, substantial differences are evident in the eicosanoid systems among these few insect species. More detailed characterizations of the elements of eicosanoid biosynthesis in these species will undoubtedly reveal more differences. The biochemistry of these systems merits considerably more research attention.

4.9.5.4.4. The pharmacology of eicosanoid biosynthesis inhibitors There is very little knowledge concerning the movements or metabolic fates of eicosanoid biosynthesis inhibitors in invertebrates. Murtaugh and Denlinger (1982) maintained a group of house crickets on diets with indomethacin. After several days on these diets, they found reduced levels of PGE_2 and $PGF_{2\alpha}$ in testes from the experimental males and spermathecae from mated females. These findings suggest that indomethacin can move from the alimentary canal to at least two tissues in house crickets and also suggest the possibility of substantial biochemical differences between

the action of these compounds in whole animals and enzyme preparations.

These issues motivated the study of pharmacology of indomethacin in *Manduca*. Using fifth instar larvae, radioactive indomethacin was injected into the hemocoels of the animals and its movement, excretion, and metabolism examined. Over 99% of the radioactivity associated with injected indomethacin was cleared from the hemolymph circulation within the first 3 min after injection. The indomethacin was rapidly taken up by hornworm tissues. Radioactive indomethacin was recovered from all tissues analyzed, including integument, ventral nerve cord, salivary gland, fat body, Malpighian tubules, and gut epithelium. The greatest amount of radioactivity was recovered from fat body, the largest tissue. However, when normalized to wet tissue weight, most radioactivity was recovered from the salivary glands. These results indicate that indomethacin is probably distributed among all tissues in this insect, but not uniformly, as seen in mammals. Less radioactivity was recovered from integument and nerve cord, while more was recovered from fat body, silk gland, and Malpighian tubules. This is also consistent with the pharmacology of indomethacin in mammals.

The extraction system used to recover indomethacin and its metabolites from hornworm tissues provided information on the metabolic fate of indomethacin. Indomethacin is highly soluble in chloroform and other organic solvents, and is virtually insoluble in water. Rigorous multiple extraction procedures yielded about 90% of the recovered radioactivity in the collected organic phases. The remaining 10% of the radioactivity was detected in the aqueous phases. This 10% represents polar metabolites of indomethacin, taken to indicate that about 10% of the injected indomethacin was metabolized to water-soluble products.

There was no evidence for the metabolism of indomethacin in most tissues, including integument, nerve cord, fat body, Malpighian tubules, and gut epithelium. Virtually all of the radioactivity recovered from these tissues cochromatographed with authentic indomethacin on TLC. Contrary to results with other tissues, the salivary gland produced at least one polar product of indomethacin, which made up no more than 25% of the starting material. Hence, the material taken up into virtually all hornworm tissues is present as indomethacin.

The excretion of radioactive indomethacin was recorded by collecting frass. In this experiment, radioactive indomethacin was injected into the hemocoel of fifth instar larvae. The insects were held in individual cups, and frass pellets were collected every 2 h for the following 40 h. About 56% of the injected material was recovered over the 40 h incubation period. A trace of radioactivity appeared in the frass as early as 2 h after injection and almost half of the injected radioactivity was recovered between 4 and 12 h. These data reveal that substantial amounts of injected indomethacin remain in the hornworm tissues in its original form during the first 12 h after treatments. For purposes of experimental design, the inhibitor was present in the hornworm throughout the time course of our immunology experiments.

The products extracted from the frass were analyzed and three to five indomethacin metabolites were present in frass, accounting for about 30% of the recovered radioactivity. Since enteric microbes may have been responsible for the apparent metabolism, radioactive indomethacin was incubated with freshly collected frass pellets and the data showed that about 10% of the starting material was metabolized into more polar products after 40 h. This result suggests that one or more factors in hornworm frass are responsible for indomethacin metabolism prior to excretion.

The composite data indicate that indomethacin is an appropriate probe for assessing the roles of eicosanoid biosynthetic pathways in the cellular immunity of tobacco hornworms, rather than insects or invertebrates, because there are substantial differences in the pharmacology of indomethacin among mammalian species. For example, dogs and guinea pigs require about 20 min to clear 50% of injected indomethacin doses from the circulation (Yesair *et al.*, 1970), whereas rats require about 4 h (Hucker *et al.*, 1966). The distribution of indomethacin among tissues within mammals also differs among species. Rats maintain higher concentrations of indomethacin in blood circulation than in tissues at all times after injection. In guinea pigs, indomethacin is concentrated from blood into liver, kidney, and small intestine. The excretion of indomethacin also differs among mammals (Hucker *et al.*, 1966; Yesair *et al.*, 1970). Nearly all injected indomethacin is excreted in feces in dogs, while rabbits excrete most injected indomethacin in the urine. Guinea pigs and humans excrete about half of the injected indomethacin in urine, the remainder in feces. These findings emphasize differences in distribution, metabolism, and excretion of injected indomethacin among mammals. Similar differences are to be expected among invertebrates. Indeed, the different biochemical effects of indomethacin on eicosanoid biosynthesis in insect systems have been discussed. Overall, this work points to the importance of understanding the pharmacology of eicosanoid biosynthesis inhibitors in the work on assessing the roles of eicosanoids

in invertebrates, and the pharmacology of inhibitors in general.

4.9.6. Emerging Topics

During the time Dadd (1985) was developing his review of insect nutrition, appreciation of the biological significance of PGs and other eicosanoids in insects or other invertebrates was an emerging topic in insect biochemistry and physiology. There was virtually no background literature on the presence or meaning of AA in insect tissue lipids (Stanley-Samuelson and Dadd, 1983) and very little idea of PG action in insects. Understanding of eicosanoids in insect biology has grown in the last two decades, mostly due to studies that revealed new – or emerging – information. In this section several emerging areas have been highlighted.

4.9.6.1. An Insect Pathogen Inhibits Eicosanoid-Mediated Immunity

The significance of the hypothesis that eicosanoids mediate insect cellular immunity is the identification of a previously unrecognized and crucial signal transduction system in insect immunity. The biological actions of eicosanoids involve a suite of enzymes responsible for eicosanoid biosynthesis and, presumably, specific receptors and intracellular receptor interactions with effector systems. These enzymes, receptors, and intracellular proteins all represent potential novel targets that may ultimately be exploited in the ongoing development of new insecticides. Given the diversity of microbial organisms, it is not surprising to see that at least one insect pathogen has already exploited the inhibition of eicosanoid biosynthesis as a mechanism to downregulate insect immunity.

The nematode *Steinernema carpocapsae* lives in a symbiotic relationship with the bacterium *Xenorhabdus nematophilus* from which both partners gain advantage. The nematode serves as a productive environment for growth of the bacterium and *X. nematophilus* serves the nematode by killing newly invaded insect hosts. The biology of this and other bacterium–nematode–insect systems has been well documented (Kaya and Gaugler, 1993). Park and Kim (2000) recently illuminated a new aspect of these systems with their report that eicosanoids attenuate the lethality of *X. nematophilus* in the insect *Spodoptera exigua*. The authors inferred that the bacterium somehow inhibits eicosanoid biosynthesis in newly infected host insects, thereby impairing the insect cellular immune reactions to the presence of the bacterium.

It was hypothesized that the bacterium inhibited eicosanoid biosynthesis in insect immune tissues by restraining the action of PLA_2, the first step in eicosanoid biosynthesis. Testing this hypothesis began by showing that bacterial infections stimulate increased PLA_2 activity in hemocytes from tobacco hornworms (Tunaz *et al.*, 2003). The significance of this work is that it supports the view that bacterial infections lead to increased eicosanoid biosynthesis (as shown by Jurenka *et al.* (1999)) by stimulating hemocytic PLA_2 activity. It was subsequently found that, as Park and Kim (2000) originally suggested for *S. exigua*, *X. nematophilus* impairs nodulation reactions to bacterial infection in tobacco hornworms by inhibiting eicosanoid biosynthesis (Park *et al.*, 2003). This work showed that hornworms challenged with living bacteria produced far fewer nodules than did hornworms challenged with heat-killed bacteria. The immunity-impairing influence of challenge with living bacteria could be off-set by treating experimental hornworms with AA. It may be inferred that *X. nematophilus* inhibits eicosanoid biosynthesis in all of its host insect species.

It appears from this work (Park *et al.*, 2003) that the bacterium *X. nematophilus* secretes a specific product that inhibits eicosanoid biosynthesis. To investigate this, *X. nematophilus* cells were cultured, then separated from their culture medium by centrifugation; the medium was then fractionated into an aqueous and an organic fraction. After concentrating the extracts, each fraction was tested for its influence on nodulation. The organic fraction inhibited nodulation reactions to living bacteria while the aqueous fraction had no influence on nodulation. The organic fraction was fractionated into five discrete fractions by column chromatography. After testing each of the five, nodulation-inhibiting activity was present in only one. Moreover, results indicate *X. nematophilus* cells and its immunity-impairing product specifically inhibits PLA_2 in tobacco hornworm hemocytes (Y. Park *et al.*, unpublished data). The identity of the inhibitory factor should be determined in the near future.

This work revealed important new information on the biochemical mechanisms, which underlie the relationships among certain entomopathogenic nematodes, their symbiotic bacteria, and their insect hosts. It also strongly supported the hypothesis that eicosanoids mediate insect cellular defense reactions to microbial infection. Another aspect of the significance of these findings is that they indicate the potentials for discovery of mediating mechanisms in insect physiology and biochemistry.

4.9.6.2. Bioactive Oxylipids from Linoleic Acid

In addition to oxygenation of AA via the COX or LOX pathways, biomedical research on mammalian systems has revealed over a dozen of oxygenated

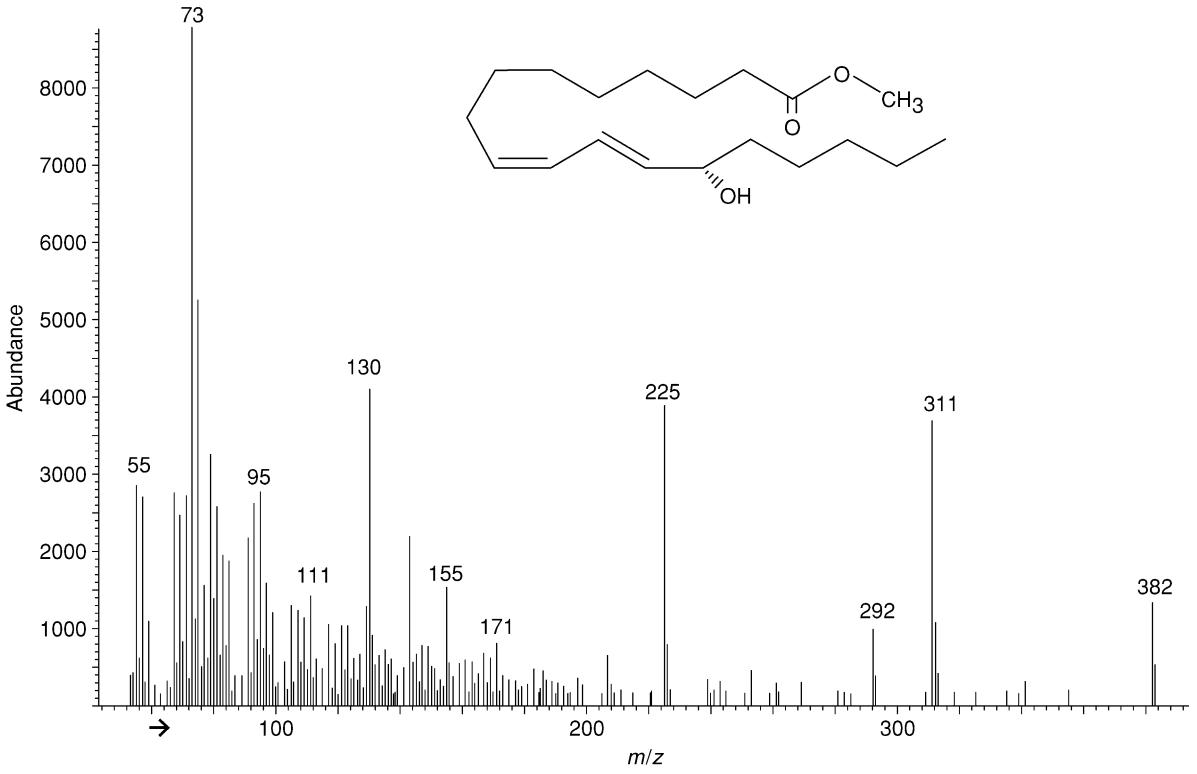


Figure 7 The structure of 13-hydroxyoctadecanoic acid methyl ester and an electron impact mass spectrum of the trimethylsilyl derivative of 13-hydroxyoctadecanoic acid methyl ester synthesized by a microsomal-enriched preparation of *Manduca sexta* fat body.

metabolites of linoleic acid, some of which exert important biological influence on cells. Products of linoleic acid metabolism include 9-hydroxyoctadecadienoic acid (9(S)-HODE) and 13(S)-HODE. Glasgow and Eling (1990) reported that epidermal growth factor stimulates linoleic acid metabolism in fibroblasts of BALB/c 3T3 mice. The same group later reported that 13(S)-HODE augments the epidermal growth factor receptor signaling pathway (Glasgow *et al.*, 2002).

Based on information in the biomedical literature, Putnam and Stanley (unpublished data) considered the hypothesis that insect tissues, also, are able to form potentially bioactive products of linoleic acid. Using fat body from tobacco hornworms, the biosynthesis of two bioactive lipids from radioactive $C_{18:2n-6}$, specifically 9- and 13-HODE, was recorded. The structures of these compounds were determined by gas chromatography-mass spectrometry (**Figure 7**). The possibility that metabolites of $C_{18:2n-6}$ may play important regulatory roles in insect cellular physiology is now emerging.

4.9.6.3. PG Receptors

Turning attention to another emerging topic, recall the broad claim of the eicosanoid hypothesis, i.e., eicosanoids somehow mediate insect cellular immune reactions to challenge. While the general theme of the hypothesis is strongly supported, information on the mechanisms of eicosanoid actions in insect biology remains the most prominent gap in our current knowledge. In contemporary models of eicosanoid action with respect to immunity (Stanley-Samuelson and Pedibhotla, 1996; Stanley *et al.*, 2002), circulating hemocytes recognize an immune challenge, most likely by recognition of specific cell components, such as LPS in the case of infection with Gram-negative bacteria. Upon stimulation by an invader, hemocytes produce and secrete eicosanoids which enter the hemolymph and influence the action and behavior of other hemocytes.

There are important implications in this model. First, the secreted eicosanoids are thought to act on other hemocytes, presumably through functional receptors which have been well characterized in mammalian systems. Second, hemocyte eicosanoid receptors are not uniformly expressed in all hemocytes. To the contrary, because eicosanoids mediate various cell defense actions – such as microaggregation, phagocytosis, cell spreading, nodulation, melanization, and encapsulation – it is possible

that subpopulations of hemocytes express receptors for particular eicosanoids.

The most important implication is that an understanding of eicosanoid receptors can lead to a more detailed understanding of hemocytic immunity and other areas of insect physiology because it will become possible to establish a direct linkage between particular eicosanoids and specific cell actions (Stanley, 2000).

4.9.7. Biochemical Mechanisms of Prostaglandin Actions

4.9.7.1. Background on G Protein-Coupled Receptors

The physiological actions of most PGs are mediated by G protein-coupled receptors (GPCRs) (Breyer *et al.*, 2001). The exceptions are PGA and other cyclopentenone PGs, which readily traverse cellular membranes and interact with intracellular receptors (Negishi *et al.*, 1995b). GPCRs are seven-transmembrane proteins responsible for transducing extracellular signals into intracellular compartments, where cellular responses to the signal are initiated. There are literally hundreds of GPCRs, which are specific for a wide range of ligands, including those involved in vision, olfaction, taste, neurotransmission, neuromodulation, and hormone actions. All multicellular animals have GPCRs, and the genes that encode these receptors make up substantial proportions of their entire genomes. GPRCs make up about 6% of the *Caenorhabditis elegans* genome, possibly 2% of mammalian genomes, and about 1% of the *D. melanogaster* genome (Brody and Cravchik, 2000; Vanden Broeck, 2001). The many known GPCRs are sorted into families, including the olfactory receptor family, a glutamate receptor family, a biogenic amine receptor family, and a diuretic hormone receptor family. PGs interact with GPCRs of the rhodopsin family. GPCRs transduce extracellular signals through selective coupling with intracellular G proteins, which in turn regulate the activity of cellular effector proteins, such as adenylyl cyclase, phospholipase C, or ion channels (Blenau and Baumann, 2001).

Receptors for most PGs are localized in the plasma membrane, although nuclear PG receptors have been reported (Bhattacharya *et al.*, 1999). PG receptors are classified on the basis of amino acid sequences and the effects of synthetic ligands which interact with the receptors (Coleman *et al.*, 1990). PGs D, E, F, I, and thromboxane A have corresponding receptors, denoted as DP, EP, FP, IP, and TP. So far, four subclasses of PGE receptors have been described, EP1, EP2, EP3, and EP4. These receptors affect changes in intracellular signal moieties by coupling with various G proteins. For example, EP1 receptors stimulate increased intracellular Ca^{2+}, and EP2 and EP4 receptors stimulate increased intracellular cAMP. EP3 receptors are unique as so far understood because they exist in variants which differ only in their intracellular C-termini, due to splice variations (Hatae *et al.*, 1989). This allows the different splice variants to couple with multiple G proteins, producing different physiological responses. For example, the EP3A can stimulate increased cAMP concentrations, while EP3B can exert the opposite effect.

4.9.7.2. PG Receptors in Insects

While many GPCRs have been identified and cloned from various insect sources (Roeder, 1999; Blenau and Baumann, 2001; Torfs *et al.*, 2001; Vanden Broeck, 2001), PG receptors have not been reported for insects. Indeed, information on the biochemical and molecular features of insect PG receptors remains the most prominent gap in understanding eicosanoid action in insect physiology (Stanley, 2000). Tunaz and Stanley (unpublished data) now have preliminary data on PGE_2 receptors in tobacco hornworm hemocytes. They used classical ligand binding studies, drawn from Negishi *et al.* (1993) and from Qian *et al.* (1997), who reported on PGE_2 receptors in salivary glands from the lone star tick, to register the presence of PGE_2 receptors in membrane preparations made from tobacco hornworm hemocytes. Data from this work indicate the presence of saturable, specific PGE_2 receptors associated with membrane preparations of hemocytes from tobacco hornworms. Analysis of the binding data indicted a single binding site model with a K_D of approximately 35 nM and B_{max} of approximately 7.5 fmol mg^{-1} protein.

It should be borne in mind that analysis of insect PG receptors by radioligand binding studies can be quite problematic. For example, insect tissues are small and they seem to express a small number of PG binding proteins relative to mammalian cells while binding studies require a great deal of tissue. Second, because PGs and other eicosanoids are themselves lipids, they present potentially serious problems with nonspecific or low-affinity binding. Finally, it is very difficult to isolate and study separate subpopulations of insect hemocytes or other subpopulations of cells in other tissues. The difficulties inherent in classical radioligand binding studies have been addressed in work on mammalian PG receptors. The contemporary approach is to clone the genes for the receptors, stably express the genes in established cell lines, and use the cell lines to characterize the receptors. This strategy

has yielded a very detailed picture of PG receptors in the biomedical context of mammalian systems.

Information from the *Drosophila* and the *Anopheles* genomes indicates that sequences are present in both insect genomes, which are quite similar to mammalian sequences for PG receptors. In their analysis of the *Anopheles* genome, Hill *et al.* (2002) identified a total of 276 GPCRs. Of these, the authors predicted the presence of five classes of PGCRs, including rhodopsin-like receptors. In mammals, PG receptors are members of the rhodopsin-like receptor family. Within the mosquito rhodopsin-like receptors, 22 were identified as "orphan receptors," that is, receptors for which the physiological ligand is unknown. The predicted number of orphan rhodopsin-like receptors in the *Drosophila* genome is also 22.

Haas and Stanley (unpublished data) used the orphan receptors 1 through 21 from *A. gambiae* (drawn from the supporting online material in Hill *et al.* (2002)) to conduct BLAST searches for similar receptor sequences in SWISSPROT. Two of these orphans, GPRorpha5 and GPRorpha11, are very similar to sequences representing receptors for various PGs, with over 100 hits with at least some similarity to mammalian PG receptors and over 40 hits with very high similarity (bit scores >44; E values <0.001).

Analysis of the two *Drosophila* orthologs to the *A. gambiae* orphans yielded similar findings. Again, there were over 100 hits with some similarity and over 40 hits with very high similarity, judged by the same statistical criteria. It may be inferred from these two series of *in silico* searches that at least some of the rhodopsin-like GPCRs, now recognized as orphans, may turn out to be specific PG receptors which mediate crucial physiological actions in insects.

Of course, it cannot be known whether these are, indeed, PG receptors of physiological significance until they have been analyzed in detail. Such analysis is now under way in the laboratory of Jozef Vanden Broeck with his student Vanessa Franssens in Leuven, Belgium. One of the *Drosophila* gene sequences (GC 4794) has been cloned from an expressed sequence tag and ligated into an expression vector. The expression vector was transvected into established *Drosophil*a cell lines for detailed biochemical studies. The results of this work should add a great deal of new information on the biological actions of PGs in insects.

4.9.8. Conclusions: The Potentials of Eicosanoids in Insect Biology

A great deal of new information on the presence and biological actions of PGs and other eicosanoids in insects, as well as in other invertebrates, has come to light in the nearly 20 years since publication of the first edition of *Comprehensive Insect Biochemistry, Physiology, and Pharmacology*. Work on the biological significance of eicosanoids in insects is rewarding because there is tremendous possibility for discovery. On one hand, research in this area has revealed biochemical signal mechanisms responsible for mediating crucial events in the lives of insects, such as cellular immune reactions to infection, ovarian development, and releasing egg-laying behavior in some insect species. Plainly, continued interest in this topic will yield new knowledge about how insects modulate a large variety of physiological actions.

Analysis of eicosanoid biosynthesis suggests that eicosanoids act in most insect tissues. For example, Büyükgüzel *et al.* (2002) characterized PG biosynthesis in midgut tissue from tobacco hornworms, showing that the alimentary canals of at least some insects are able to produce PGs. It is reasonable to now question the roles of these PGs in insect midguts.

Beyond gaining new appreciation of insect function, most invertebrate groups are far older than mammals. Although eicosanoids were first discovered in mammals, the long evolutionary history of eicosanoid actions in invertebrates has provided ample opportunity to recruit eicosanoids into many biological roles, most of which remain to be discovered. For example at least one PG serves as a postovulatory pheromone in some fish species. Do PGs serve as pheromones in any other animal group? Another example is this: it is thought that some corals produce very high levels of tissue PGs, which serve to reduce herbivory by inducing vomiting in fish which feed on the coral. In these cases it appears that PGs mediate some protective ecological interactions. There are other examples of PG actions in invertebrates and lower vertebrates which are not seen in mammals, and their study should lead to the elucidation of new roles for PGs in the future.

Work on invertebrate eicosanoid systems also will produce new information on fundamental issues. Efforts to clone a gene encoding COX from insect tissues (collaboration between David Stanley and Kulliki Varvas in Tallinn, Estonia) have been frustrated by the substantial differences among the mammalian genes that encode COX and their insect counterparts. Work on insects and other invertebrates will likely turn up a large number of COX isoforms, all of which are unknown at present. This work will enable scientists to better understand the organization and evolution of vertebrate and invertebrate COXs. Similarly, molecular studies of receptors for PGs and other eicosanoids

should enhance our understanding of intracellular signal transduction mechanisms substantially.

Research into insect eicosanoid systems may also have important practical applications. In the early days of work on PGs in insects, it was not at all clear how this information might be integrated into insect pest management schemes. The work by Park and Kim (Park and Kim, 2000; Park *et al*., 2003) opens possibilities for the discovery of natural products that impair insect immunity by inhibiting eicosanoid biosynthesis. Moreover, their work suggests mechanisms for the delivery of the products in a targeted way to pest populations. This may be the only first inkling of future opportunity to apply our understanding of eicosanoid systems.

Studies of PGs may teach us a great deal about how insects work, may generate information on fundamental aspects of eicosanoid biosynthesis and mechanisms of action, and may yield information of practical importance in agricultural and medical entomology. Young scientists are entering this emerging field, and these investigators will surely make truly exciting discoveries in the future. I look forward to discussing these discoveries in the future editions of this series.

References

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., *et al*., **1994**. Molecular Biology of the Cell, 3rd edn. Garland, New York.

Armstrong, P.B., Melchior, R., Quigley, J.P., **1996**. Humoral immunity in long-lived arthropods. *J. Insect Physiol. 42*, 53–64.

Baines, D., DeSantis, T., Downer, R.G.H., **1992**. Octopamine and 5-hydroxy-tryptamine enhance the phagocytic and nodule formation activities of cockroach (*Periplaneta americana*) haemocytes. *J. Insect Physiol. 11*, 905–914.

Baines, D., Downer, R.G.H., **1994**. Octopamine enhances phagocytosis in cockroach hemocytes: involvement of inositol triphosphate. *Arch. Insect Biochem. Physiol. 22*, 249–261.

Balsinde, J., Balboa, M.A., Insel, P.A., Dennis, E.A., **1999**. Regulation and inhibition of phospholipase A_2. *Annu. Rev. Pharmacol. Toxicol. 39*, 175–198.

Bergstrom, S., Dressler, F., Krabisch, C., Ryhage, R., Sjovall, J., **1962a**. The isolation and structure of a smooth muscle simulating factor in normal sheep and pig lungs. *Ark. Kem. 20*, 63–66.

Bergstrom, S., Ryhage, R., Samuelsson, B., Sjovall, J., **1962b**. The structure of prostaglandin E, F_1 and F_2. *Acta Chem. Scandin. 16*, 501–502.

Berridge, M.J., **1970**. The role of 5-hydroxytryptamine and cyclic AMP in the control of fluid secretion by isolated salivary glands. *J. Exp. Biol. 53*, 171–186.

Berridge, M.J., **1993**. Inositol triphosphate and calcium signalling. *Nature 361*, 315–325.

Bhattacharya, M., Varma, D.R., Chemtob, S., **1999**. Nuclear prostaglandin receptors. *Gene Therapy Mol. Biol. 4*, 323–338.

Blee, E., **1998**. Phytooxylipins and plant defense reactions. *Prog. Lipid Res. 37*, 33–72.

Blenau, W., Baumann, A., **2001**. Molecular and pharmacological properties of insect biogenic amine receptors: lessons from *Drosophila melanogaster* and *Apis mellifera*. *Arch. Insect Biochem. Physiol. 48*, 13–38.

Boman, H.G., Hultmark, D., **1987**. Cell-free immunity in insects. *Annu. Rev. Microbiol. 41*, 103–126.

Bonvalet, J.-P., Pradelles, P., Farman, N., **1987**. Segmental synthesis and actions of prostaglandins along the nephron. *Am. J. Physiol. 253*, F377–F387.

Botting, R.M., **2002**. Mechanism of action of acetaminophen: is there a COX 3? *Clin. Infect. Diseases 31*, S202–S210.

Brehelin, M. (Ed.) **1986**. Immunity in Invertebrates. Springer-Verlag, New York.

Brenner, X., Bernasconi, Z., **1989**. Prostaglandin biosynthesis in the gonads of the hematophagous insect *Triatoma infestans*. *Comp. Biochem. Physiol. 93B*, 1–4.

Breyer, R.M., Bagdassarian, C.K., Myers, S.A., Breyer, M.D., **2001**. Prostanoid receptors: subtypes and signaling. *Annu. Rev. Pharmacol. Toxicol. 41*, 661–690.

Brody, T., Cravchik, A., **2000**. *Drosophila melanogaster* G protein-coupled receptors. *J. Cell Biol. 150*, F83–F88.

Büyükgüzel, K., Tunaz, H., Putnam, S.M., Stanley, D.W., **2002**. Prostaglandin biosynthesis by midgut tissue isolated from the tobacco hornworm, *Manduca sexta*. *Insect Biochem. Mol. Biol. 32*, 435–443.

Carton, Y., Frey, F., Stanley, D.W., Vass, E., Nappi, A.J., **2002**. Dexamethasone inhibition of the cellular immune response of *Drosophila melanogaster* against a parasitoid. *J. Parasitol. 288*, 405–407.

Chandrasekharan, N.V., Dai, H., Roos, K.L.T., Evanson, N.K., Tomsik, J., *et al*., **2002**. COX-3, a COX-1 variant inhibited by acetaminophen and other analgesic/antipyretic drugs: cloning, structure and expression. *Proc. Natl Acad. Sci. USA 99*, 13926–13931.

Chilton, F.H., Murphy, R.C., **1986**. Remodeling of arachidonate-containing phosphoglycerides within the human neutrophil. *J. Biol. Chem. 261*, 7771–7777.

Coleman, R.A., Kennedy, L., Humphrey, P.P.A., Bunce, K., Lumley, P., **1990**. Prostanoids and their receptors. In: Transch, C., Sammes, F.G., Taylor, J. (Eds.), Comprehensive Medicinal Chemistry, vol. 3. Pergamon Press, Oxford, p. 643.

Conte, A., Ottaviani, E., **1995**. Nitric oxide synthase activity in molluscan hemocytes. *FEBS Lett. 365*, 120–124.

Corey, E.J., Albright, J.O., Barton, A.E., Hashimoto, S., **1980**. Chemical and enzymic syntheses of 5-HPETE, a key biological precursor of slow-reacting substance of anaphylaxis (SRS) and 5-HETE. *J. Am. Chem. Soc. 102*, 1435–1436.

Dadd, R.H., **1973**. Insect nutrition: current developments and metabolic implications. *Annu. Rev. Entomol. 18*, 331–420.

Dadd, R.H., **1977**. Quantitative requirements and utilization of nutrients: insects. In: Rechcigl, M. (Ed.), Handbook Series in Nutrition and Food, vol. 1. CRC Press, Boca Raton, FL, p. 305.

Dadd, R.H., **1981**. Essential fatty acids for mosquitoes, other insects and vertebrates. In: Bhaskaran, G., Friedman, S., Rodrigue, J.G. (Eds.), Current Topics in Insect Endocrinology and Nutrition. Plenum Press, New York, pp. 184–214.

Dadd, R.H., **1983**. Essential fatty acids: insects and vertebrates compared. In: Mittler, T.E., Dadd, R.H. (Eds.), Metabolic Aspects of Lipid Nutrition in Insects. Westview Press, Boulder, pp. 107–147.

Dadd, R.H., **1985**. Nutrition: organisms. In: Kerkut, G. A., Gilbert, L.I. (Eds.), Comprehensive Insect Physiology Biochemistry and Pharmacology, vol. 4. Pergamon Press, Oxford, pp. 313–390.

Dadd, R.H., Kleinjan, J.E., **1979**. Essential fatty acid for the mosquito *Culex pipiens*: arachidonic acid. *J. Insect Physiol. 25*, 495–502.

Dalton, T., **1977a**. Threshold and receptor reserve in the action of 5-hydroxytryptamine on the salivary glands of *Calliphora erythrocephala*. *J. Insect Physiol. 23*, 625–631.

Dalton, T., **1977b**. The effect of prostaglandin E_1 on cyclic AMP production in the salivary glands of *Calliphora erythrocephala*. *Experientia 33*, 1329–1330.

Dean, P., Gadsden, J.C., Richards, E.H., Edwards, J.P., Charnley, A.K., *et al.*, **2002**. Modulation by eicosanoids of immune responses by the insect *Manduca sexta* to the pathogenic fungus *Metarhizium anisopliae*. *J. Invert. Pathol. 79*, 93–101.

Dennis, E.A., **1994**. Diversity of group types, regulation, and function of phospholipase A_2. *J. Biol. Chem. 269*, 13057–13060.

Dennis, E.A., **1997**. The growing phospholipase A_2 superfamily of signal transduction enzymes. *Trends Biochem. Sci. 22*, 1–2.

Destephano, D.B., Brady, U.E., **1977**. Prostaglandin and prostaglandin synthetase in the cricket, *Acheta domesticus*. *J. Insect Physiol. 23*, 905–911.

Destephano, D.B., Brady, U.E., Woodall, L.B., **1976**. Partial characterization of prostaglandin synthetase in the reproductive tract of the male house cricket, *Acheta domesticus*. *Prostaglandins 11*, 261–273.

Diehl-Jones, W.L., Mandato, C.A., Whent, G., Downer, R.G.H., **1996**. Monoaminergic regulation of hemocyte activity. *J. Insect Physiol. 42*, 12–19.

Dunn, P.E., **1986**. Biochemical aspects of insect immunity. *Annu. Rev. Entomol. 31*, 321–339.

Dunn, P.E., Drake, D.R., **1983**. Fate of bacteria injected into naive and immunized larvae of the tobacco hornworm *Manduca sexta*. *J. Invert. Pathol. 41*, 77–85.

Dunphy, G.B., Downer, R.G.H., **1994**. Octopamine: a modulator of the haemocytic nodulation response of non-immune *Galleria mellonella* larvae. *J. Insect Physiol. 40*, 267–272.

Ferreira, S.H., Moncada, S., Vane, J.R., **1971**. Indomethacin and aspirin abolish prostaglandin release from the spleen. *Nature New Biol. 231*, 237–239.

Gadelhak, G.G., Pedibhotla, V.K., Stanley-Samuelson, D.W., **1995**. Eicosanoid biosynthesis by hemocytes from the tobacco hornworm, *Manduca sexta*. *Insect Biochem. Mol. Biol. 25*, 743–749.

Gadelhak, G.G., Stanley-Samuelson, D.W., **1994**. Incorporation of polyunsaturated fatty acids into phospholipids of hemocytes from the tobacco hornworm, *Manduca sexta*. *Insect Biochem. Mol. Biol. 24*, 775–785.

Gerwick, W.H., **1993**. Carbocyclic oxylipins of marine origin. *Chem. Rev. 93*, 1807–1823.

Gillespie, J.P., Kanost, M.R., Trenczek, T., **1997**. Biological mediators of insect immunity. *Annu. Rev. Entomol. 42*, 611–643.

Glasgow, W.C., Eling, T.E., **1990**. Epidermal growth factor stimulates linoleic acid metabolism in BALBL/c 3T3 fibroblasts. *Mol. Pharmacol. 38*, 503–510.

Glasgow, W.C., Hui, R., Kameda, H., Eling, T.E., **2002**. 13(S)-HpODE modulates mitogenic signal transduction through enhancing the phosphorylation and association of EGF receptor with the tyrosine phosphatase SHP-2. *Adv. Exp. Med. Biol. 507*, 463–467.

Goldblatt, M.W., **1933**. A depressor substance in seminal fluid. *Chem. Industry 52*, 1056–1057.

Goldsworthy, G., Mullen, L., Opoku-Ware, K., Chandrakant, S., **2003**. Interactions between the endocrine and immune systems in locusts. *Physiol. Entomol. 28*, 54–61.

Gupta, A.P. (Ed.) **1986**. Hemocytic and Humoral Immunity in Arthropods. John Wiley, New York.

Gupta, A.P. (Ed.) **1991**. Immunology of Insects and Other Arthropods. CRC Press, Boca Raton, FL.

Halwani, A.E., Dunphy, G.B., **1999**. Apolipophorin-III in *Galleria mellonella* potentiates haemolymph lytic activity. *Devel. Comp. Immunol. 23*, 563–570.

Hatae, N., Sugimoto, Y., Ichikawa, A., **1989**. Prostglandin receptors: advances in the study of EP3 receptor signaling. *J. Biochem. (Tokyo) 131*, 781–784.

Hill, C.A., Fox, A.N., Pitts, R.J., Kent, L.B., Tan, P.L., *et al.*, **2002**. G protein-coupled receptors in *Anopheles gambiae*. *Science 298*, 176–178.

Hinton, H.E., **1960**. Cryptobiosis in the larva of *Polypedilum vanderplanki* Hint. (Chironomidae). *J. Insect Physiol. 5*, 286–300.

Hoffmann, J.A., Reichart, J.M., **2002**. *Drosophila* innate immunity: an evolutionary perspective. *Nature Immunol. 3*, 121–126.

Horohov, D.W., Dunn, P.E., **1983**. Phagocytosis and nodule formation by hemocytes of *Manduca sexta* larvae following injection of *Pseudomonas aeruginosa*. *J. Invert. Pathol. 41*, 203–213.

Howard, R.W., Stanley-Samuelson, D.W., **1996**. Fatty acid composition of the fat body and Malpighian tubules of the tenebrionid beetle, *Zophobas atratus*:

significance in eicosanoid-mediated physiology. *Comp. Biochem. Phyiosol. B 115*, 429–437.

Howard, R.W., Witters, N.A., Stanley-Samuelson, D.W., **1992**. Immunohistochemical localization of PGE_2 and $PGF_{2\alpha}$ and fatty acid compositions of Malpighian tubule phospholipids from laboratory and feral populations of *Tenebrio molitor* L. (Coleoptera: Tenebrionidae). *Ann. Entomol. Soc. America 85*, 489–498.

Hucker, H.B., Zacchei, A.G., Cox, S.V., Brodie, D.A., Cantwell, N.H.R., **1966**. Studies on the absorption, distribution and excretion of indomethacin in various species. *J. Pharmacol. Exp. Therapy 153*, 237–249.

Hurst, J.S., Flatman, S., McDonald-Gibson, R.G., **1987**. Thin-layer chromatography (including radio thin-layer chromatography and autoradiography) of prostaglandins and related compounds. In: Benedetto, C., McDonald-Gibson, R.G., Nigam, S., Slater, T.F. (Eds.), Prostaglandins and Related Substances: A Practical Approach. IRL Press, Oxford, pp. 53–73.

Izawa, Y., Uchida, M., Yasui, M., **1986**. Mode of action of buprofezin on the twenty-eight spotted ladybird, *Henosepilachna vigintioctopunctata* Fabricius. *Agric. Biol. Chem. 50*, 1369–1371.

Jurenka, R.A., Miller, J.S., Pedibhotla, V.K., Rana, R.L., Stanley-Samuelson, D.W., **1997**. Eicosanoids mediate microaggregation and nodulation responses to bacterial infections in black cutworms, *Agrotis ipsilon*, and true armyworms, *Pseudaletia unipuncta*. *J. Insect Physiol. 43*, 125–133.

Jurenka, R.A., Pedibhotla, V.K., Stanley, D.W., **1999**. Prostaglandin production in response to a bacterial infection in true armyworm larvae. *Arch. Insect Biochem. Physiol. 41*, 225–232.

Kanai, N., Lu, R., Satrian, J.A., Bao, Y., Wolkoff, A.W., *et al.*, **1995**. Identification and characterization of a prostaglandin transporter. *Science 268*, 866–869.

Kaya, H.K., Gaugler, R., **1993**. Entomopathogenic nematodes. *Annu. Rev. Entomol. 38*, 181–206.

Kirschner, L.B., **1991**. Water and ions. In: Prosser, C. L. (Ed.), Comparative Animal Physiology, 4th edn: Environmental and Metabolic Animal Physiology. Wiley–Liss, New York, pp. 13–107.

Kurzrok, R., Lieb, C.C., **1930**. Biochemical studies of human semen. 2. The action of semen on the human uterus. *Proc. Soc. Exp. Biol. Med. 28*, 268–272.

Lackie, A.M., **1988**. Hemocyte behaviour. *Adv. Insect Physiol. 21*, 85–178.

Lemaitre, B., Reichhart, J.-M., Hoffmann, J.A., **1997**. *Drosophila* host defense: differential induction of antimicrobial peptide genes after infection by various classes of microorganisms. *Proc. Natl Acad. Sci. USA 94*, 14614–14619.

Levine, M.D., Strand, M.R., **2002**. Insect hemocytes and their role in immunity. *Insect Biochem. Mol. Biol. 32*, 1295–1309.

Loher, W., Ganjian, I., Kubo, I., Stanley-Samuelson, D.W., Tobe, S.S., **1981**. Prostaglandins: their role in egg-laying in the cricket *Teleogryllus commodus*. *Proc. Natl Acad. Sci. USA 78*, 7835–7838.

Lord, J.C., Anderson, S., Stanley, D.W., **2002**. Eicosanoids mediate *Manduca sexta* cellular response to the fungal pathogen *Beauveria bassiana*: a role for the lipoxygenase pathway. *Arch. Insect Biochem. Physiol. 51*, 46–54.

Mandato, C.A., Diehl-Jones, W.L., Moore, S.J., Downer, R.G.H., **1997**. The effects of eicosanoid biosynthesis inhibitors on prophenoloxidase activation, phagocytosis and cell spreading in *Galleria mellonella*. *J. Insect Physiol. 43*, 1–8.

Medeiros, M.N., Oliveira, M.P., Paiva-Silva, G.O., Silva-Neto, M.A.C., Romeiro, A., *et al.*, **2002**. The role of eicosanoids on *Rhodnius* heme-binding protein (RHBP) endocytosis by *Rhodnius prolixus* ovaries. *Insect Biochem. Mol. Biol. 32*, 537–545.

Metters, K.M., **1995**. Leukotriene receptors. *J. Lipid Mediators Cell Signal. 12*, 413–427.

Miller, J.S., Howard, R.W., Nguyen, T., Nguye, A., Rosario, R.M.T., *et al.*, **1996**. Eicosanoids mediate nodulation responses to bacterial infections in larvae of the tenebrionid beetle, *Zophobas atratus*. *J. Insect Physiol. 42*, 3–12.

Miller, J.S., Howard, R.W., Rana, R.L., Tunaz, H., Stanley, D.W., **1999**. Eicosanoids mediate nodulation reactions to bacterial infections in adults of the cricket, *Gryllus assimilis*. *J. Insect Physiol. 45*, 75–83.

Miller, J.S., Nguyen, T., Stanley-Samuelson, D.W., **1994**. Eicosanoids mediate insect nodulation responses to bacterial infections. *Proc. Natl Acad. Sci. USA 91*, 12418–12422.

Miller, J.S., Stanley, D.W., **1998a**. Eicosanoids in animal reproduction: what can we learn from invertebrates? In: Rowley, A.F., Kuhn, H., Schewe, T. (Eds.), Eicosanoids and Related Compounds in Plants and Animals. Portland Press, London, pp. 183–196.

Miller, J.S., Stanley, D.W., **1998b**. The nodule formation reaction to bacterial infection: assessing the role of eicosanoids. In: Wiesner, A., Dunphy, G.B., Marmaras, V.J., Morishima, I., Sugumaran, M., Yamakawa, M. (Eds.), Techniques in Insect Immunology. SOS Publications, Fair Haven, pp. 265–270.

Miller, J.S., Stanley, D.W., **2001**. Eicosanoids mediate microaggregation reactions to bacterial challenge in isolated insect hemocyte preparations. *J. Insect Physiol. 47*, 61–69.

Morishima, I., Yamano, Y., Inoue, K., Matsuo, N., **1997**. Eicosanoids mediate induction of immune genes in the fat body of the silkworm, *Bombyx mori*. *FEBS Lett. 419*, 83–86.

Murtaugh, M.P., Denlinger, D.L., **1982**. Prostaglandin E and $F_{2\alpha}$ in the house cricket and other insects. *Insect Biochem. 12*, 599–603.

Negishi, M., Koizumi, T., Ichikawa, A., **1995a**. Biological actions of Δ^{12}-prostaglandin J_2. *J. Lipid Mediators Cell Signal. 12*, 443–448.

Negishi, M., Sugimoto, Y., Ichikawa, A., **1993**. Prostanoid receptors and their biological actions. *Prog. Lipid Res. 32*, 414–434.

Negishi, M., Sugimoto, Y., Ichikawa, A., **1995b**. Molecular mechanisms of diverse actions of prostanoid receptors. *Biochim. Biophys. Acta 1259*, 109–120.

Nor Aliza, A.R., Bedick, J.C., Rana, R.L., Tunaz, H., Hoback, W.W., *et al.*, **2001**. Arachidonic and eicosapentaenoic acids in tissues of the firefly, *Photinus pyralis* (Insecta: Coleoptera). *Comp. Biochem. Physiol. A 128*, 251–257.

Oates, J.A., **1982**. The 1982 Nobel Prize in physiology or medicine. *Science 218*, 765–768.

Ogg, C.L., Howard, R.W., Stanley-Samuelson, D.W., **1991**. Fatty acid composition and incorporation of arachidonic acid into phospholipids of hemocytes from the tobacco hornworm, *Manduca sexta*. *Insect Biochem. 21*, 109–114.

Ogg, C.L., Stanley-Samuelson, D.W., **1992**. Phospholipid and triacylglycerol fatty acid compositions of the major life stages and selected tissues of the tobacco hornworm *Manduca sexta*. *Comp. Biochem. Physiol. B 101*, 345–351.

Otto, J.C., Smith, W.L., **1995**. Prostaglandin endoperoxide synthases-1 and -2. *J. Lipid Mediators Cell Signal. 12*, 139–156.

Pace-Asciak, C.R., Asotra, S., **1989**. Biosynthesis, catabolism, and biological properties of HPETEs, hydroperoxide derivatives of arachidonic acid. *Free Radical Biol. Med. 7*, 409–433.

Park, Y., Kim, Y., **2000**. Eicosanoids rescue *Spodoptera exigua* infected with *Xenorhabdus nematophilus*, the symbiotic bacteria to the entomopathogenic nematode *Steinernema carpocapsae*. *J. Insect Physiol. 46*, 1469–1476.

Park, Y., Kim, Y., Putnam, S.M., Stanley, D.W., **2003**. The bacterium *Xenorhabdus nematophilus* depresses nodulation reactions to infection by inhibiting eicosanoid biosynthesis in tobacco hornworms, *Manduca sexta*. *Arch. Insect Biochem. Physiol. 52*, 71–80.

Petzel, D.H., **1993**. Prostanoids and fluid balance in insects. In: Stanley-Samuelson, D.W., Nelson, D.R. (Eds.), Insect Lipids, Chemisry, Biochemistry and Biology. University of Nebraska Press, Lincoln, NE, pp. 139–178.

Petzel, D.H., Parrish, A.K., Ogg, C.L., Witters, N.A., Howard, R.W., *et al.*, **1993**. Arachidonic acid and prostaglandin E_2 in Malpighian tubules of female yellow fever mosquitoes. *Insect Biochem. Mol. Biol. 25*, 431–437.

Petzel, D.H., Stanley-Samuelson, D.W., **1992**. Inhibition of eicosanoid biosynthesis modulates basal fluid secretion in the Malpighian tubules of the yellow fever mosquito (*Aedes aegypti*). *J. Insect Physiol. 38*, 1–8.

Phillips, J.E., **1980**. Epithelial transport and control of recta of terrestrial insects. In: Locke, M.L., Smith, D.E. (Eds.), Insect Biology of the Future. Academic Press, New York, pp. 145–177.

Prince, W.T., Berridge, M.J., Ramussen, H., **1972**. Role of calcium and adenosine 3′-5′cyclic monophosphate in controlling fly salivary gland secretion. *Proc. Natl Acad. Sci. USA 69*, 553–557.

Qian, Y., Essenberg, R.C., Dillwith, J.W., Bowman, A.S., Sauer, J.R., **1997**. A specific prostaglandin E_2 receptor and its role in modulating salivary secretion in the female tick, *Amblyomma americanum* (L.). *Insect Biochem. Mol. Biol. 27*, 387–395.

Qian, Y., Yuan, J., Essenberg, R.C., Bowman, A.S., Shook, A.L., *et al.*, **1998**. Prostaglandin E_2 in the salivary glands of the female tick, *Amblyomma americanum* (L.): calcium mobilization and exocytosis. *Insect Biochem. Mol. Biol. 28*, 221–228.

Radallah, D., Nogaro, M., Fournier, B., **1995**. Arachidonic acid and prostaglandin E_2 stimulate phospholipase C activity in the rectum of the African locust, *Locusta migratoria migratorioides*. *J. Insect Physiol. 41*, 7–16.

Roeder, T., **1999**. Octopamaine in invertebrates. *Prog. Neurobiol. 59*, 533–561.

Rowley, A.F., **1996**. The evolution of inflammatory mediators. *Mediat. Inflamm. 5*, 3–13.

Samuelsson, B., **1983**. Leukotrienes: mediators of immediate hypersensitivity reactions and inflammation. *Science 220*, 568–575.

Sauer, J.R., McSwain, J.L., Bowman, A.S., Essenberg, R.C., **1995**. Tick salivary gland physiology. *Annu. Rev. Entomol. 40*, 245–267.

Schleusener, D.R., Stanley-Samuelson, D.W., **1996**. Phospholipase A_2 in hemocytes of the tobacco hornworm, *Manduca sexta*. *Arch. Insect Biochem. Physiol. 33*, 63–74.

Simmons, D.L., Botting, R.M., Robertson, P.M., Madsen, M.L., Vane, J.R., **1999**. Induction of an acetaminopohen-sensitive COX with reduced sensitivity to nonsteroid anti-inflammatory drugs. *Proc. Natl Acad. Sci. USA 96*, 3275–3280.

Smith, W.L., Garavito, R.M., DeWitt, D.L., **1996**. Prostaglandin endoperoxide H synthases (COXs)-1 and -2. *J. Biol. Chem. 271*, 33157–33160.

Smith, W.L., Marnett, L.J., **1991**. Prostaglandin endoperoxide synthase: structure and catalysis. *Biochim. Biophys. Acta. 1083*, 1–17.

Smith, W.L., Borgeat, P., Fitzpatrick, F.A., **1991**. The eicosanoids: cyclooxygenase, lipoxygenase, and epoxygenase pathways. In: Vance, D.E., Vance, J.E. (Eds.), Biochemistry of Lipids, Lipoproteins, and Membranes. Elsevier, Amsterdam, pp. 297–325.

Sorensen, P.W., Hara, T.J., Stacey, N.E., Goetz, F.W., **1988**. F prostaglandins function as potent olfactory stimulants that comprise the postovulatory female sex pheromone in goldfish. *Biol. Reprod. 39*, 1039–1050.

Spector, A.A., Gordon, J.A., Moore, S.A., **1988**. Hydroxyeicosatetraenoic acids (HETEs). *Prog. Lipid Res. 27*, 271–323.

Stanley, D.W., **2000**. Eicosanoids in Invertebrate Signal Transduction Systems. Princeton University Press, Princeton, NJ.

Stanley, D.W., Hoback, W.W., Bedick, J.C., Tunaz, H., Rana, R.L., *et al.*, **1999**. Eicosanoids mediate nodulation reactions to bacterial infections in larvae of the butterfly, *Colias eurytheme*. *Comp. Biochem. Physiol. C 123*, 217–223.

Stanley, D.W., Howard, R.W., **1998**. The biology of prostaglandins and related eicosanoids in invertebrates:

cellular, organismal and ecological actions. *Am. Zool. 38*, 369–381.

Stanley, D.W., Nor Aliza, A.R., Tunaz, H., Putnam, S.M., Park, Y., *et al.*, **2002**. Eicosanoids in insect biology. *Neotrop. Entomol. 31*, 341–350.

Stanley-Samuelson, D.W., **1987**. Physiological roles of prostaglandins and other eicosanoids in invertebrates. *Biol. Bull. 173*, 92–109.

Stanley-Samuelson, D.W., **1991**. Comparative eicosanoid physiology in invertebrate animals. *Am. J. Physiol. 260*, R849–R853.

Stanley-Samuelson, D.W., **1994a**. Prostaglandins and related eicosanoids in insects. *Adv. Insect Physiol. 24*, 115–212.

Stanley-Samuelson, D.W., **1994b**. Assessing the significance of prostaglandins and other eicosanoids in insect physiology. *J. Insect Physiol. 40*, 3–11.

Stanley-Samuelson, D.W., Dadd, R.H., **1983**. Long-chain polyunsaturated fatty acids: patterns of occurrence in insects. *Insect Biochem. 13*, 549–558.

Stanley-Samuelson, D.W., Dadd, R.H., **1984**. Polyunsaturated fatty acids in the lipids from adult *Galleria mellonella* reared on diets to which only one unsaturated fatty acid had beed added. *Insect Biochem. 14*, 321–327.

Stanley-Samuelson, D.W., Jensen, E., Nickerson, K.W., Tiebel, K., Ogg, C.L., *et al.*, **1991**. Insect immune response to bacterial infection is mediated by eicosanoids. *Proc. Natl Acad. Sci. USA 88*, 1064–1068.

Stanley-Samuelson, D.W., Jurenka, R.A., Blomquist, G.J., Loher, W., **1987**. Sexual transfer of prostaglandin precursor in the field cricket, *Teleogryllus commodus*. *Physiol. Entomol. 12*, 347–354.

Stanley-Samuelson, D.W., Jurenka, R.A., Cripps, C., Blomquist, G.J., de Renobales, M., **1988**. Fatty acids in insects: composition, metabolism, and biological significance. *Arch. Insect Biochem. Physiol. 9*, 1–33.

Stanley-Samuelson, D.W., Loher, W., **1983**. Arachidonic and other long-chain polyunsaturated fatty acids in spermatophores and spermathecae of *Teleogryllus commodus*: significance in prostaglandin-mediated reproductive behavior. *J. Insect Physiol. 29*, 41–45.

Stanley-Samuelson, D.W., Loher, W., **1986**. Prostaglandins in insect reproduction. *Ann. Entomol. Soc. America 79*, 841–853.

Stanley-Samuelson, D.W., Ogg, C.L., **1994**. Prostaglandin biosynthesis by fat body from the tobacco hornworm, *Manduca sexta*. *Insect Biochem. Mol. Biol. 24*, 481–491.

Stanley-Samuelson, D.W., Pedibhotla, V.K., **1996**. What can we learn from prostaglandins and related eicosanoids in insects? *Insect Biochem. Mol. Biol. 26*, 223–234.

Stanley-Samuelson, D.W., Pedibhotla, V.K., Rana, R.L., Nor Aliza, A.R., Hoback, W.W., *et al.*, **1997**. Eicosanoids mediate nodulation responses to bacterial infections in larvae of the silkmoth, *Bombyx mori*. *Comp. Biochem. Physiol. A 118*, 93–100.

Strand, M.R., Pech, L.L., **1995**. Immunological basis for compatibility in parasitoid–host relationships. *Annu. Rev. Entomol. 40*, 31–56.

Tanabe, T., Ullrich, V., **1995**. Prostacyclin and thromboxane synthases. *J. Lipid Mediators Cell Signal. 12*, 243–255.

Tauszig, S., Jauanguy, E., Hoffmann, J.A., Imler, J.L., **2000**. Toll-related receptors and the control of antimicrobial peptide expression in *Drosophila*. *Proc. Natl Acad. Sci. USA 97*, 10520–10525.

Thornhill, R., Alcock, R., **1983**. The Evolution of Insect Mating Systems. Harvard University Press, Cambridge, MA.

Tobe, S.S., Loher, W., **1983**. Properties of the prostaglandin synthetase complex in the cricket *Teleogryllus commodus*. *Insect Biochem. 13*, 137–141.

Torfs, H., Oonk, H.B., Vanden Broeck, J., Poels, J., Van Poyer, W., *et al.*, **2001**. Pharmacological characterization of STKR, an insect G protein-coupled receptors for tachykinin-like peptides. *Arch. Insect Biochem. Physiol. 48*, 39–49.

Tunaz, H., Bedick, J.C., Miller, J.S., Hoback, W.W., Rana, R.L., *et al.*, **1999**. Eicosanoids mediate nodulation reactions to bacterial infections in adults of two 17-year periodical cicadas, *Magicicada septendecim* and *M. cassini*. *J. Insect Physiol. 45*, 923–931.

Tunaz, H., Park, Y., Büyükgüzel, K., Bedick, J.C., Nor Aliza, A.R., *et al.*, **2003**. Eicosanoids in insect immunity: bacterial infection stimulates hemocytic phospholipase A_2 activity in tobacco hornworms. *Arch. Insect Biochem. Physiol. 52*, 1–6.

Tunaz, H., Stanley, D.W., **1999**. Eicosanoids mediate nodulation reactions to bacterial infections in adults of the American cockroach, *Periplaneta americana* (L.). *Proc. Entomol. Soc. Ontario 130*, 97–108.

Uchida, M., Izawa, Y., Sugimoto, T., **1987**. Inhibition of prostaglandin biosynthesis and oviposition by an insect growth regulator, Buprofezin, in *Nilaparvata lugens* Stal. *Pesticide Biochem. Physiol. 27*, 71–75.

Urade, Y., Watanabe, K., Hayaishi, O., **1995**. Prostaglandin D, E, and F synthases. *J. Lipid Mediators Cell Signal. 12*, 257–273.

Uscian, J.M., Stanley-Samuelson, D.W., **1993**. Phospholipase A_2 activity in the fat body of the tobacco hornworm *Manduca sexta*. *Arch. Insect Biochem. Physiol. 24*, 187–201.

Vanden Broeck, J., **2001**. Insect G protein-coupled receptors and signal transduction. *Arch. Insect Biochem. Physiol. 48*, 1–12.

Vane, J.R., **1971**. Inhibition of prostaglandin synthesis as a mechanism of action for aspirin-like drugs. *Nature New Biol. 231*, 232–235.

Van Kerkhove, E., Pirotte, P., Petzel, D.H., Stanley-Samuelson, D.W., **1995**. Eicosanoid biosynthesis inhibitors modulate basal fluid secretion rates in the Malpighian tubules of the ant, *Formica polyctena*. *J. Insect Physiol. 41*, 435–441.

von Euler, U.S., **1936**. On the specific vasodilating and plain muscle stimulating substances from accessory genital glands in men and certain animals (prostaglandin and vesiglandin). *J. Physiol. (London) 88*, 213–234.

Wakayama, E.J., Dillwith, J.W., Blomquist, G.J., **1986**. Characterization of prostaglandin synthesis in the housefly, *Musca domestica* (L.). *Insect Biochem. 16*, 903–909.

Warner, T.D., Mitchell, J.A., **2002**. COX-3 (COX-3): filling in the gaps toward a COX continuum? *Proc. Natl Acad. Sci. USA 99*, 13371–13373.

Wiesner, A., Losen, S., Kopacek, P., Weise, C., Gotz, P., **1997**. Isolated apolipophorin III from *Galleria mellonella* stimulates the immune reactions of this insect. *J. Insect Physiol. 43*, 383–391.

Yajima, M., Takada, M., Takahashi, N., Kikuchi, H., Natori, S., *et al.*, **2003**. A newly established *in vitro* ulture using transgenic *Drosophila* reveals functional coupling between the phospholipase A_2-generated fatty acid cascade and lipopolysaccharide-dependent activation of the immune deficiency (imd) pathway in insect immunity. *Biochem. J. 371*, 205–210.

Yamaja Setty, B.N., Ramaiah, T.R., **1979**. Isolation and identification of prostaglandins from the reproductive organs of male silkmoth, *Bombyx mori* (L.). *Insect Biochem. 9*, 613–617.

Yamaja Setty, B.N., Ramaiah, T.R., **1980**. Effect of prostaglandins and inhibitors of prostaglandin biosynthesis on oviposition in the silkmoth *Bombyx mori*. *Ind. J. Exp. Biol. 18*, 539–541.

Yesair, D.W., Callahan, M., Remington, L., Kensler, C.J., **1970**. Comparative effects of salicylic acid, phenylbutazone, probenecid and other anions on the metabolism, distribution and excretion of indomethacin by rats. *Biochem. Pharmacol. 19*, 1591–1600.

4.10 Ferritin

J J Winzerling, University of Arizona, Tucson, AZ, USA
D Q-D Pham, University of Wisconsin-Parkside, Kenosha, WI, USA

© 2005, Elsevier BV. All Rights Reserved.

4.10.1. Importance of Ferritin

Scientific interest in iron lies in the fact that it is both an essential nutrient and a potential toxin. Iron is required for many metabolic processes as a cofactor of numerous proteins. However, in the presence of oxygen, ferrous is readily oxidized to ferric rendering it insoluble at physiological pH. Ferrous oxidation also can result in formation of hydroxyl radical anions by the Haber–Weiss and Fenton reactions:

$$Fe^{2+} + O_2 \rightarrow Fe^{3+} + O_2^{-\bullet}$$
$$2O_2^{-\bullet} + 2H^+ \rightarrow H_2O_2 + O_2$$
$$Fe^{2+} + H_2O_2 \rightarrow OH^{\bullet} + OH^- + Fe^{3+}$$

Hydroxyl radical anions can damage many biological molecules and produce a variety of reactive oxygen species (ROS) (Kehrer, 2000; Aisen *et al*., 2001). Organisms must have adequate supplies of iron to meet nutritional requirements, and at the same time, prevent this potential cytotoxic effect. This balance is met, in part, by storing iron inside ferritin.

Ferritin is the primary iron storage protein of mammals. It is found predominately in the cell cytoplasm, and influences iron availability and response to oxidative stress (Aisen *et al*., 2001). In mammalian cells, the loss of ferritin expression is associated with susceptibility to iron-dependent toxicity and increased oxidative stress (DeRusso *et al*., 1995; Kakhlon *et al*., 2001). Ferritin overexpression is associated with increased iron deposition, changes in cell growth and improved response to oxidative challenge (Corsi *et al*., 1998; Epsztejn *et al*., 1999; Kakhlon *et al*., 2002; Kato and Niitsu, 2002).

The study of iron metabolism of insects was initiated by the work of Michael Locke and his colleagues (Locke and Leung, 1984; Nichol and Locke, 1989; Huebers *et al*. (1988); Bartfeld and Law, 1990). Since then much progress has been made and several reviews have been written on the subject of insect iron metabolism (Locke and Nichol, 1992; Winzerling and Law, 1997; Nichol *et al*., 2002). This chapter will focus on the composition, structure, expression, and putative roles of ferritin in insects.

4.10.2. Ferritin Structure and Iron Storage

Insect ferritins are usually compared with the well-characterized vertebrate cytoplasmic ferritins. Vertebrate cytoplasmic ferritins have a mass $\geq$450 kDa and consist of 24 subunits of $\sim$19 kDa (light chain, L) and $\sim$21 kDa (heavy chain, H) configured as a spherical polymer wherein iron is stored (Harrison and Arosio, 1996; Chasteen and Harrison, 1999). Vertebrate H subunits are characterized by a ferroxidase center which facilitates ferrous oxidation and uptake (Lawson *et al*., 1989, 1991; Levi *et al*., 1994; Santambrogio *et al*., 1996). The vertebrate L subunits have glutamic acid residues that constitute a nucleation site that allows iron deposition within the ferritin core, as well as other residues that form salt bridges that contribute to the structural stability of this large molecule (Trikha *et al*., 1995; Gallois *et al*., 1997; Hempstead *et al*., 1997).

Apoferritin becomes holoferritin when ferrous is oxidized, moved inside the molecule, and forms a crystalline mineral. *In vitro* studies show that Fe^{2+} binds to the glutamate and histidine residues of the ferroxidase center of the H subunits and is rapidly oxidized to Fe^{3+} (Aisen *et al.*, 2001). Fe^{3+} is then released from the oxidase site and transported to the nucleation site. Mineralization proceeds as the surface of the growing mineral core catalyzes autooxidation of joining iron molecules. Iron is stored in ferritin as hydroxide polymers:

$$n\mathrm{Fe}^{3+}(\mathrm{OH})_3 \xrightarrow{-n\mathrm{H_2O}} -(\mathrm{FeOOH})_{n-}$$

Although ferritin can accommodate 4500 hydroxyl-bridged iron atoms, isolated mammalian ferritins usually contain only 2000–2500 iron atoms (Aisen *et al.*, 2001). Iron is released from ferritin *in vitro* by strong chelators.

How iron is incorporated and released from ferritin *in vivo* has yet to be elucidated. *In vitro* iron loading by the ferroxidase site produces H_2O_2 and oxidative damage to the H subunit (Welch *et al.*, 2002). For this reason, it is suggested that *in vivo* iron loading occurs enzymatically and that the ferroxidase site is used only when this process is exceeded (Welch *et al.*, 2002). Treatment of ferritin with the 20S proteosome is accompanied by the release of iron (Aisen *et al.*, 2001), whether this serves to provide iron for cellular needs is not known.

Insect ferritins have a mass of 450–660 kDa with subunits that range in size from 21 to 36 kDa (Nichol and Locke, 1989; Dunkov *et al.*, 1995; Winzerling *et al.*, 1995; Charlesworth *et al.*, 1997; Du *et al.*, 2000). Although crystal structure studies of insect ferritins have not been done, several lines of evidence support the hypothesis that the structure of insect ferritins is similar to that of vertebrate cytoplasmic ferritin. Insect ferritins assume a rosette appearance on electron microscopic analysis that is characteristic of vertebrate ferritins (Nichol and Locke, 1989). Further, the mass of the protein predicted by a combination of 24 subunits agrees with the respective size of the native insect ferritin (Winzerling *et al.*, 1995; Nichol and Locke, 1999). In addition, the structure of the insect ferritin subunits is similar to that of the vertebrate ferritin subunits.

4.10.2.1. Comparative Structure of Ferritin Subunits

The terminology identifying the insect ferritin subunits is varied. The subunits with greatest structural similarity to the vertebrate H subunit are referred to in the literature as the heavy chain homolog (HCH) (Dunkov *et al.*, 1995; Zhang *et al.*, 2001a), S (Nichol and Locke, 1999), Fer-1 (Charlesworth *et al.*, 1997), and Fersub1 (Du *et al.*, 2000). The amino acid residues that constitute the ferroxidase center of vertebrate H subunits are conserved in all of these subunits (**Table 1**). Insect subunits that retain this feature will be referred to as an HCH.

A second type of insect ferritin subunit that lacks the residues of the ferroxidase site is referred to in the literature as the light chain homolog (LCH) (Pham *et al.*, 1996; Georgieva *et al.*, 2002b), G (Nichol and Locke, 1999), and Fersub2 (Du *et al.*, 2000). We will refer to subunits of this type as an LCH. Although the LCH nomenclature has been used, the LCH subunits show no greater similarity to vertebrate L subunits than to vertebrate H subunits (**Table 2**) and are actually larger than the insect HCHs.

Nucleotide sequences of HCH subunits are available for *Manduca sexta* (hawkmoth, Lepidoptera (Zhang *et al.*, 2001a)), *Calpodes ethlius* (skipper butterfly, Lepidoptera (Nichol and Locke, 1999)), *Galleria mellonella* (wax moth, Lepidoptera (Kim *et al.*, 2001)), *Nilaparvata lugens* (plant hopper, Hemiptera (Du *et al.*, 2000)), *Aedes aegypti* (yellow fever mosquito, Diptera (Dunkov *et al.*, 1995)), *Drosophila melanogaster* (fruit fly, Diptera (Charlesworth *et al.*, 1997)), *Bombyx mori* (silk moth, Lepidoptera (Nichol *et al.*, 2002)), and *Anopheles gambiae* (malaria mosquito, Diptera (Holt *et al.*, 2002)). LCH subunit sequences are reported for *M. sexta* (Pham *et al.*, 1996), *C. ethlius* (Nichol and Locke, 1999), *G. mellonella* (Kim *et al.*, 2002), *B. mori* (Nichol *et al.*, 2002), *N. lugens* (Du *et al.*, 2000), *D. melanogaster* (Dunkov and Georgieva, 1999; Georgieva *et al.*, 2002b), *A. gambiae* (Holt *et al.*, 2002) and *A. aegypti* (Geiser *et al.*, 2003). A comparison of the deduced amino acid sequences of the HCH and LCH insect subunits shows a much lower percentage of identity for subunits of the same species than for the same subunit among species (**Table 2**).

The structure of the insect ferritin subunits is similar to that of vertebrates (Pham, 2000). Vertebrate ferritin subunits consist of five α-helices (A–E) and an internal loop (Lawson *et al.*, 1991; Trikha *et al.*, 1994; Gallois *et al.*, 1997; Hempstead *et al.*, 1997). The modeling programs PHD (Rost and Sander, 1993, 1994a, 1994b; Rost *et al.*, 1995) and Swiss-PdbViewer (Peitsch, 1995, 1996; Guex and Peitsch, 1997; Guex *et al.*, 1999) detect the first four α-helices (A, B, C, and D) in the HCHs and LCHs from lepidopterans and dipterans (Pham,

Table 1 Characteristics of the insect ferritin subunits

Species (strain)	*Subunit name*	*Deduced mature peptide* (∼kDa)	*Matched N-terminal sequence* (∼kDa)	*Ferroxidase residues*	*Glycosylation site*[a] *(ConA)*	*Iron-responsive element*	*GenBank accession number*	*Reference*
Aedes aegypti (Bahaman)	HCH	21	24 26	+	−	+	L37082	Dunkov *et al.* (1995)
Aedes aegypti	LCH	23	28	−	+(+)	−	AY171561	Geiser *et al.* (2003)
Calpodes ethlius	HCH (S)	22	24	+	+(−)	+	AF161707	Nichol and Locke, (1999)
Calpodes ethlius	LCH (G)	24	31	−	+(+)	+	AF161709	Nichol and Locke, (1999)
Drosophila melanogaster	HCH	21	25 26	+	−	+	Y15629 U91524	Charlesworth *et al.* (1997)
Drosophila melanogaster	LCH	23	28	−	−	−	AF145124	Georgieva *et al.* (2002b)
Galleria mellonella	HCH	22	26	+	−	+	AF142340	Kim *et al.* (2001)
Galleria mellonella	LCH	27	32	−	+	+	AF329683	Kim *et al.* (2002)
Manduca sexta	HCH	22		+	−	+	AY032659	Zhang *et al.* (2001b)
Manduca sexta	LCH	25		−	+(+)	+	AF270492	Pham *et al.* (1996)
Nilaparvata lugens	HCH (Fersub1)	24		+	+(−)		AJ251148	Du *et al.* (2000)
Nilaparvata lugens	LCH (Fersub2)	24	26	−	+(+GNA)	+	AJ251147	Du *et al.* (2000)

[a]Putative glycosylation site by computer analysis; conA, Concanaolin A; GNA, *Galanthus nivalis*.

Reprinted, with permission from the Annual Review of Entomology, Volume 47, © 2002 by Annual Reviews www.annualreviews.org.

Table 2 Comparative identities of insect ferritin subunits

	Species													
	Subunit	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*
1	Human H														
2	Human L	55													
3	*Aa* HCH	35	23												
4	*Aa* LCH	17	17	17											
5	*Ce* HCH	36	23	41	16										
6	*Ce* LCH	19	19	16	23	16									
7	*Dm* HCH	35	31	50	18	49	17								
8	*Dm* LCH	27	24	17	28	27	32	23							
9	*Gm* HCH	27	18	39	18	74	19	50	19						
10	*Gm* LCH	21	21	18	26	18	69	35	22	19					
11	*Ms* HCH	43	42	42	20	70	19	52	18	83	22				
12	*Ms* LCH	26	23	24	27	17	67	40	23	20	78	25			
13	*Nl* HCH	30	29	40	16	48	23	42	22	46	14	45	15		
14	*Nl* LCH	21	21	15	29	14	34	34	16	14	34	35	18	14	

Human H, human heavy chain, human H chain (GenBank assession number L20941); Human L, human light chain, human L chain (GenBank assession number M10119); HCH, heavy chain homolog; LCH, light chain homolog; *Aa, Aedes aegypti*, *Ce, Calpodes ethlius; Dm, Drosophila melanogaster; Gm, Galleria mellonella; Ms, Manduca sexta; Nl, Nilaparvata lugens*. Assession numbers are listed in **Table 1**.
Reprinted, with permission from the Annual Review of Entomology, Volume 47, © 2002 by Annual Reviews www.annualreviews.org.

2000). However, only the Swiss model recognizes with certainty the fifth helix (E) in both orders of insects.

Although the dipteran and lepidopteran HCHs have diverged from their vertebrate counterparts, they still retain a significant structural similarity. Specifically, the tracing of the HCH carbon backbone is easily superimposed with the human H subunit backbone and the α-helices align as do the different types of amino acid residues (such as nonpolar, polar, acidic, and basic). Further, the seven amino acid residues (four glutamates, tyrosine, histidine, and glutamine) involved in the ferroxidase center maintain the same three-dimensional structure as the vertebrate ferroxidase center (Pham, 2000). Iron loading of ferritin *in vitro* using the ferroxidase site results in free radical-mediated damage of residue C90 of the H subunit and protein aggregation (Welch *et al*., 2002). Interestingly, this residue is not conserved in insect HCH subunits, suggesting that the ferroxidase site could be used for iron oxidation without damage to the subunit at this residue.

Although the α-helices in the insect LCHs are identified by modeling programs, these subunits exhibit less similarity to their vertebrate counterparts than the HCHs (**Table 2**) and lack important features associated with the vertebrate L subunits (Trikha *et al*., 1994; Gallois *et al*., 1997; Hempstead *et al*., 1997). Vertebrate L subunits have a porphyrin-binding pocket associated with ferrihydrite nucleation and residues that form salt bridges that contribute to structural integrity (Andrews *et al*., 1992; Trikha *et al*., 1994; Gallois *et al*., 1997; Hempstead *et al*., 1997). Sequence alignment indicates that the cluster of glutamic acids that functions as the porphyrin-binding pocket in the vertebrate L subunits is missing in insect LCHs (Pham, 2000). The loss of this cluster suggests that nucleation for iron storage occurs differently in insect ferritins. Further, although some amino acid residues that permit salt bridge formation appear to be semiconserved, the substitutions at these sites in insect LCHs make bridge formation unlikely because the salt interaction is lost. Since these bridges maintain the stability of vertebrate ferritins, the loss of these sites suggests that either the insect ferritins are less stable or they are stabilized by other forces. Current data do not support the former hypothesis because insect ferritins have been shown to be quite stable (Nichol and Locke, 1989; Dunkov *et al*., 1995; Winzerling *et al*., 1995). Finally, insect LCHs have a unique attribute: all contain tyrosine kinase phosphorylation sites (Pham, 2000).

In summary, the insect HCHs retain many structural features relevant to iron uptake and storage, whereas, the LCHs maintain neither the porphyrin ring nor the salt bridges that are found in vertebrate L subunits. In addition, significant differences exist in primary structure between the insect LCHs and the vertebrate L subunits, suggesting that species-specific activities of insect ferritins are a function of the LCH.

Most of the insect ferritin subunits identified to date have hydrophobic leader sequences that signal secretion, and secreted ferritin appears to serve important roles in insects. However, an insect ferritin subunit that lacks a secretion signal sequence was identified in the *Drosophila* genome (Adams *et al*., 2000). The deduced amino acid sequence indicates that this subunit retains the ferroxidase center and is most similar to a vertebrate H subunit (Nichol *et al*., 2002). A cytosolic ferritin also was described for *Philaenus spumarius* (spittle bug, Homoptera), which consists of a single subunit that is much smaller than other insect ferritin subunits (Collin *et al*., 1988). The sequence of this subunit has not been reported.

4.10.3. The Biology of Insect Ferritin

4.10.3.1. Roles of Ferritin in Insects

How insects absorb iron and maintain iron homeostasis is unknown. Plant-feeders are exposed to a variety of iron compounds and plant ferritins (Theil, 1987; Andrews *et al*., 1992; Briat, 1996). Alternatively, hematophagous insects are exposed to an iron load in the forms of hemoglobin and ferric-transferrin. Since various sources of iron are available to insects, the mechanisms of iron absorption and loss among insects are likely to be noticeably different.

Lepidopterans express high levels of ferritin messages in gut tissues (Pham *et al*., 1996; Nichol and Locke, 1999; Kim *et al*., 2001, 2002). Iron administration increases gut mRNA expression and, in *C. ethlius* larvae, provokes secretion of holoferritin into the posterior midgut lumen (Locke and Leung, 1984). Locke and colleagues have suggested that iron homeostasis is preserved in these insects by secreting holoferritin into the gut lumen for excretion in the feces (Locke and Leung, 1984; Nichol *et al*., 2002). Like the lepidopterans, ferritin mRNA and protein are increased in the gut tissues of the dipterans *D. melanogaster* and *Musca domestica* (housefly) in response to iron (Capurro *et al*., 1996; Georgieva *et al*., 2002b), and in *A. aegypti* females following a blood meal (Dunkov *et al*., 2002). However in contrast to lepidopterans, Dunkov *et al*. (2002) found no evidence of ferritin in the gut contents of *A. aegypti* after blood feeding. Others have shown that substantial amounts of heme are bound to the peritrophic matrix in *A. aegypti* and lost when the matrix is shed into lumen at the end of blood digestion (Pascoa *et al*., 2002). The absence of ferritin in the gut contents in these insects probably reflects this alternate mechanism of iron disposal. Interestingly, yet another mechanism limiting iron absorption is found in *Rhodnius prolixus* (kissing bug, Hemiptera), where heme from the blood meal is sequestered in the gut as hemozoin (Oliveira *et al*., 2000). Despite limiting heme absorption, blood-feeding insects would still receive substantial iron from ferric-transferrin. Iron from this source could account for the increase in gut ferritin expression observed in *A. aegypti*. In any case, insects appear to have developed a variety of mechanisms for maintaining iron homeostasis by limiting iron uptake or by enhancing iron excretion. The secretion of holoferritin into the hindgut, the upregulation of ferritin expression in gut tissues, and the sequestering of heme all represent mechanisms that would protect insects against iron overload and iron-mediated oxidative challenge.

Other roles of insect ferritin in iron metabolism are not clear. In the majority of insects, ferritin is found in the hemolymph suggesting that this protein could be involved in iron movement among body tissues. In *C. ethlius*, apoferritin secreted from fat body becomes holoferritin in the hemolymph (Nichol and Locke, 1999; Locke, 2003). Further, when *M. sexta* larvae are injected with iron, holoferritin in the hemolymph increases and constitutes about 1% of the total protein (Winzerling *et al*., 1995; Zhang *et al*., 2001b). Similarly, in *D. melanogaster* (Georgieva *et al*., 2002b) and *A. aegypti* (Dunkov *et al*., 2002), hemolymph ferritin is increased following exposure to iron. Together these findings indicate that iron exposure increases ferritin in the hemolymph, that ferritin can be loaded with iron in this tissue, and that ferritin serves as the major iron-sequestering protein in insects. By sequestering iron in hemolymph, ferritin also could serve to protect insects from iron overload and subsequent oxidative stress, as well as to limit the iron available to invading pathogens.

The tissue source of hemolymph ferritin could vary with the insect species. In lepidopterans, hemolymph ferritin appears to originate in the fat body (Locke, 1991; Locke and Nichol, 1992; Nichol and Locke, 1999). Ferritin expression is increased in fat body in response to iron (Huebers *et al*., 1988; Locke *et al*., 1991; Pham *et al*., 1996; Nichol and Locke, 1999) and apoferritin is secreted from the fat body into the hemolymph (Nichol and Locke, 1999; Locke, 2003). In *Apis mellifera* (honey bee, Hymenoptera (Keim *et al*., 2002)), as well as several other insects, holoferritin also is found in the secretory pathway of fat body tissue (Locke and Leung, 1984; Locke and Nichol, 1992; Locke, 2003). However, in *D. melanogaster*, ferritin mRNA and protein are very low in fat body, but are high in gut tissues (Georgieva *et al*., 2002b). Georgieva *et al*. (2002b)

have suggested that in these insects hemolymph ferritin could originate from the gut. Regardless of the source, the secretion of ferritin from insect cells into the hemolymph appears crucial to insect iron metabolism.

Locke has reviewed the secretory pathway in insects (Locke, 2003). In dictyopterans, dipterans, coleopterans, hemipterans, and lepidopterans, intracellular holoferritin is rarely seen in the cytosol or the nuclear compartment, but is consistently observed in the secretory pathway (Nichol and Locke, 1990; Locke and Nichol, 1992). Only in *P. spumarius* are similar levels of cytosolic and nuclear holoferritin observed (Nichol and Locke, 1990). Insect holoferritin has been found in the rough endoplasmic reticulum (RER), Golgi complex, and secretory vesicles (Locke and Leung, 1984; Locke and Nichol, 1992; Nichol *et al*., 2002). Apparently, ferritin assembles in the RER, but is not secreted immediately. How 24 subunits come together in an orderly manner in the lumen of the RER remains unknown. The subunits from several insects have cysteine residues in the N-terminus that could serve as acylation sites (Nichol and Locke, 1999; Nichol *et al*., 2002). Nichol and Locke (1999) showed that ferritin subunits from *C. ethlius* can be acylated *in vitro*. In addition, computational analysis predicts that some of the subunits have myristoylation sites, as well as tyrosine kinase phosphorylation sites that could be involved in ferritin assembly and secretion (Pham, 2000). Reversible associations of myristoylated proteins with plasma membranes have been observed (Trowler *et al*., 1988; Peitzsch and McLaughlin, 1993). Perhaps some subunits are reversibly associated with the inner membrane through a myristoylated site allowing ferritin assembly, and in some cases, iron loading. Once this process is finished, ferritin detaches and is secreted. Possibly, conformational changes occur with phosphorylation/dephosphorylation of the tyrosine kinase sites of the LCH subunits that results in detachment from the membrane. Protein detachment from membranes following conformational changes caused by phosphorylation/dephosphorylation at tyrosine kinase sites has been documented (Hubbard *et al*., 1998).

Another potential role for ferritin in insects could be the provision of iron for the developing embryo. In vertebrates, iron is required for growth and deficiency results in developmental defects. In dipterans, ferritin is expressed in the ovaries and is found in eggs (Dunkov *et al*., 2002; Georgieva *et al*., 2002b) and its presence suggests that iron stored in ferritin can be used by the developing embryo (Dunkov *et al*., 2002). This appears to differ for lepidopterans, since ferritin is not found in the ovaries of *G. mellonella* (Kim *et al*., 2001).

In summary, the roles of ferritin in insects are only beginning to be elucidated. The increase in holoferritin in the cell secretory pathway and in hemolymph following iron administration shows that insects load ferritin with iron and suggests that ferritin could be involved in the movement of iron among insect tissues. These findings, as well as the increase in ferritin expression following iron administration in insects from several orders, suggest that ferritin is the primary iron sequestering protein in these animals. Further, the consistent finding that iron elicits ferritin expression in insect gut tissues suggests that gut ferritin could serve roles in iron homeostasis, in iron excretion, and in the protection of some insects from iron overload. Finally, the increase in ferritin expression in eggs and ovaries of some insects indicates a role for ferritin in providing iron for the developing embryo.

4.10.3.2. Ferritin Subunit Messages and Protein Expression

Ferritin purified from *M. sexta* (Winzerling *et al*., 1995) and *C. ethlius* (Nichol and Locke, 1999) has a mass of ~660 kDA and contains multiple subunits of 24–36 kDa. N-terminus sequencing data indicate that these subunits are the products of two different genes and that the various size differences result from posttranslational processing. The lepidopteran LCH subunits contain *N*-glycosylation sites (Pham *et al*., 1996; Nichol and Locke, 1999; Kim *et al*., 2002), and those from *M. sexta* and *C. ethlius* bind concanavalin A (Nichol and Locke, 1989; Pham *et al*., 1996). Northern blot analyses indicate that LCH mRNA in lepidopterans is expressed in the fat body (Pham *et al*., 1996; Nichol and Locke, 1999; Kim *et al*., 2002), hemocytes (Pham *et al*., 1996), and midgut (Pham *et al*., 1996; Nichol and Locke, 1999; Kim *et al*., 2002), with greatest expression in the midgut (Pham *et al*., 1996; Kim *et al*., 2002). In *G. mellonella*, LCH message is expressed in the fat body and midgut and is increased with iron treatment (50 mM $FeCl_3$ with an artificial diet). However, an increase in the LCH protein is observed only in the midgut (Kim *et al*., 2001, 2002). The HCH mRNA also is expressed in these tissues (Kim *et al*., 2001) and increases after iron feeding. Neither mRNA is detectable in the ovaries or testes.

Purified midgut *N. lugens* (Hemiptera) ferritin has a molecular mass ~440 kDa and dissociates into three subunits of 20, 26, and 27 kDa (Du *et al*., 2000). N-terminus sequencing is successful for the 26 kDa, but not for the 20 or 27 kDa subunits. The

cDNA clone for the 26 kDa subunit encodes a protein of 236 amino acids with highest homology with the *M. sexta* LCH subunit (**Table 2**). The sequence has an *N*-glycosylation site and shows strong binding affinity to the lectin *Galanthus nivalis* agglutinin (GNA). The fact that this ferritin subunit is one of the most abundant GNA-binding proteins from the *N. lugens* midgut is of interest since GNA is toxic to *N. lugens* as shown by artificial diet bioassays. Perhaps one plant defense mechanism against insects involves interference with insect iron metabolism by the binding of GNA with ferritin. A second ferritin subunit cDNA clone was identified as part of a strategy to obtain nonabundant GNA-binding proteins from *N. lugens* midgut. This second clone encodes a 227 deduced amino acid sequence with a predicted molecular weight of 24 kDa and a glycosylation site.

In dipterans, ferritin is composed of at least four subunits. The HCH subunits are 24 kDa (Dunkov *et al.*, 1995), 25 kDa (Charlesworth *et al.*, 1997), and 26 kDa (Dunkov *et al.*, 1995), and the cDNA sequences have been obtained for both *Drosophila* (Charlesworth *et al.*, 1997) and *Aedes* (Dunkov *et al.*, 1995). Although the native subunits are greater in mass than predicted by their respective cDNA sequences, none has been shown to be glycosylated.

The expression of the HCH in *Drosophila* is low. It is found in second instar and wandering larvae gut tissues and a trace is seen in adults; it is not expressed in early pupae (Charlesworth *et al.*, 1997; Georgieva *et al.*, 2002b). In contrast, the HCH in *A. aegypti* is highly expressed. The 26 kDa HCH is expressed at all life stages and in all tissues and is responsive to iron administration (Dunkov *et al.*, 2002). Expression is increased in adult females fed an artificial meal with iron supplements or a blood meal. Blood feeding increases HCH expression in the gut wall, fat body, hemolymph, thorax, ovaries, and eggs. The 24 kDa HCH is expressed in mosquito larvae and pupae, increases at these stages following iron administration and increases in adult females following either an artificial meal supplemented with iron or blood feeding (Dunkov *et al.*, 2002). In adult females, expression of the 24 kDa subunit appears limited to the ovaries and eggs. In contrast to the 26 kDa subunit, expression of the 24 kDa subunit is minimal in sugar-fed adult females.

Although the HCH predominates in mosquitoes, the LCH is highly expressed in flies at all life stages (Georgieva *et al.*, 2002b). The cDNA sequence for the *Drosophila* LCH encodes a deduced amino acid sequence that predicts a mature polypeptide of ~23 kDa with an N-terminus that matches that of a 28 kDa subunit (Georgieva *et al.*, 2002b). Since the 28 kDa subunit binds concanavalin A, glycosylation could account for the mass of the native polypeptide. LCH mRNA is expressed in larvae, in light and dark pupae, and in adults. With the exception of dark pupae, the mRNA quantity is responsive to iron treatment. The protein is also present at all stages and, with the exception of eggs and third instar larvae, expression is enhanced modestly with iron treatment. In imagos, the LCH subunit is present in the hemolymph and fat body, gut, ovaries, and testes. Although LCH mRNA content is very low in the larval fat body, the protein is found in this tissue. Both the LCH message and protein are expressed in gut tissues and are strongly upregulated in response to iron (Georgieva *et al.*, 2002b).

A genomic clone of the *A. aegypti* LCH subunit was recently sequenced (Geiser *et al.*, 2003). Similar to the LCH of *Drosophila*, the mosquito cDNA sequence encodes a protein with an N-terminus that matches the *A. aegypti* 28 kDa subunit and the deduced amino acid sequence predicts a mature polypeptide of ~23 kDa. When *A. aegypti* embryonic cultured cells (Aag2 cells) are treated with iron, LCH mRNA is induced and this induction is sustained at 16 h. The LCH message is also expressed in larvae and elicited in blood-fed females in contrast to sugar-fed females. Dunkov *et al.* (2002) reported that the LCH protein is expressed in larvae and pupae only following iron or hemin treatment. Adult females express low levels of this subunit; however, expression increases in female hemolymph and thorax after a blood meal (Dunkov *et al.*, 2002).

Clearly, the expression of ferritin subunits in dipterans differs; perhaps iron exposure could help explain these differences. Flies are rarely exposed to high iron intake, whereas female mosquitoes are exposed to an iron load in a blood meal. In mosquitoes, the HCH subunit could be more highly expressed because rapid iron uptake into ferritin is a priority. In contrast, perhaps the LCH subunit predominates in flies because longer-term iron storage is desirable.

4.10.4. Control of Ferritin Expression

Several mechanisms are involved in the control of ferritin expression in insects. These include transcriptional control, alternative splicing, polyadenylation, and translational control.

4.10.4.1. Transcriptional Control

The limited knowledge available for the transcriptional control of vertebrate ferritin genes comes

from studies of the H genes (Torti and Torti, 2002). In mammalian cells, iron salts stimulate a threefold increase in H subunit mRNA and a fourfold increase in the synthesis of the protein (Coccia *et al.*, 1995). In addition to iron, other agents, such as hormones, cytokines, and certain chemicals, upregulate the transcription of the H gene (Chou *et al.*, 1986; Chazenbalk *et al.*, 1990b; Wei *et al.*, 1990; Miller *et al.*, 1991; Yokomori *et al.*, 1991; Tsuji *et al.*, 1995).

In *A. aegypti*, transcriptional control is relevant in the regulation of HCH gene expression (Pham *et al.*, 1999). HCH mRNA levels increase in Aag2 cells following exposure to iron as shown by Northern blot analyses. This could be attributed to an increase in the transcription of the HCH gene, a decrease in the degradation rate of HCH mRNA, or both. Since pretreatment of Aag2 cells with actinomycin D prevents upregulation of HCH message, the increase in this mRNA in response to iron is likely due to induced transcription rather than decreased mRNA degradation.

For vertebrate H genes, all known *cis*-regulatory elements are found upstream of the transcriptional start sites (Ponka *et al.*, 1998). The proximal promoter region of the vertebrate H gene contains a TATA box, a CCAAT box (C/EBP binding site), and one or more SP1 binding sites (Wei *et al.*, 1990; Miller *et al.*, 1991; Bevilacqua *et al.*, 1992; Beaumont *et al.*, 1994a, 1994b; Kwak *et al.*, 1995; Tsuji *et al.*, 1995; Ponka *et al.*, 1998). Current data indicate that additional regulatory elements located further upstream from the transcriptional start site are necessary for full activation (Ponka *et al.*, 1998). These elements include a 180 bp hemin response element located at 4.5 kb upstream of the transcriptional start site (Beaumont *et al.*, 1994b), an NF-E2 binding site that regulates the expression of genes involved in the heme biosynthetic pathway (Beaumont *et al.*, 1994a), and a Fer-2 site that is recognized by tumor necrosis factor-α (TNF-α) (Kwak *et al.*, 1995).

In *A. aegypti*, transient transfection expression assays of progressive deletions of the HCH gene promoter region reveal that the minimal promoter sequence necessary to sustain transcription *in vitro* lies between positions −40 and +50 (Pham *et al.*, 2003). These assays also show that the promoter of this gene is strong (30-fold stronger than the promoter of cytomegalovirus and 100-fold stronger than the SV40 promoter) and that, unlike vertebrate H genes, the *Aedes* HCH gene contains regulatory elements both upstream and downstream of the transcriptional start site. Results from DNase footprinting assays and MatInspector predictions suggest that, like the vertebrate H genes, the promoter of the mosquito ferritin gene contains a TATA-box and a C/EBP site (**Figure 1**). However, in the *Aedes* HCH gene, the C/EBP binding site is found downstream of the transcriptional start site. A putative element common in the promoters of cell cycle genes, E2F, also is present in the mosquito HCH gene. This element could be the analog of an E1A repressor site found in the vertebrate ferritin H genes (Tsuji *et al.*, 1999). Regulatory elements known for

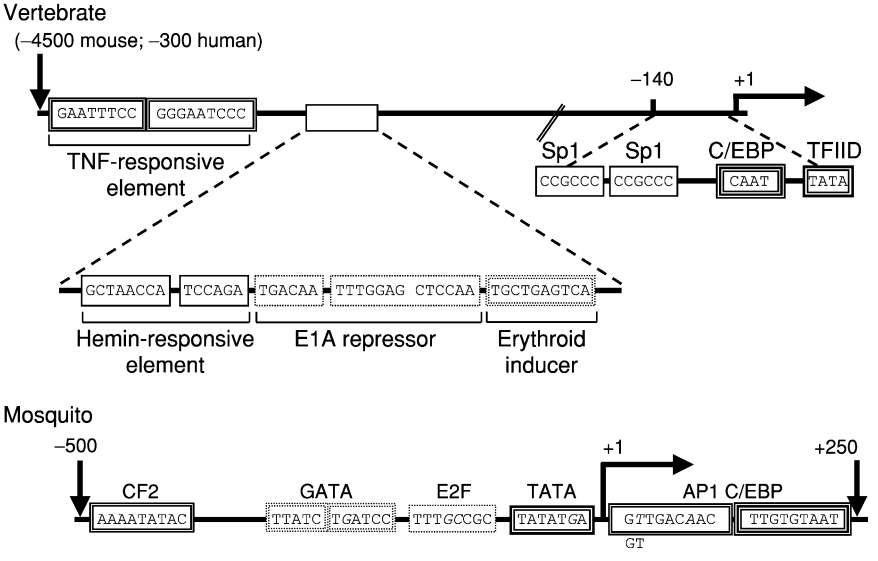


Figure 1 A diagram of the promotor regions of the genes for the vertebrate heavy chain subunit and the *Aedes* heavy chain homolog. The diagram for the vertebrate heavy chain gene promoter was adapted from information provided by Ponka *et al.* (1998) and Torti and Torti (2002). Boxes with the same border represent putative analogous *cis*-regulatory elements. +1, transcription start site; the nucleotide sequence inside the boxes are binding sites; numbers give positions using +1 as the transcriptional start site.

their involvement with erythropoiesis, GATA-binding sites and two cell-signaling sites, CF2 and AP-1, are also found in the mosquito HCH gene. These two latter putative sites might be the equivalents of the vertebrate Fer-2 site. Data from DNase footprinting assays and MatInspector predictions reveal that the regulatory elements described above are noncanonical.

Interestingly, the TATA box (TATAtga) of the *Aedes* HCH promotor does not match the known consensus sequence (TATAWAW) perfectly and the affinity of the TATA box binding protein (TBP) to TATA at this site is weak. Nonetheless, primer extension analysis indicates that transcription for the HCH gene initiates at one site in untreated cells, but switches to two sites in iron-treated cells (Pham, 2000; Pham *et al*., 2003). These data suggest that TBP recognizes the noncanonical TATA box under normal conditions and directs RNA pol II to initiate at a single start site. However, when cells are exposed to high levels of iron, TBP loses its accuracy in directing RNA pol II, resulting in RNA pol initiating at two start sites. The situation observed for iron treatment parallels transcriptional initiation seen for TATA-less promoters, where multiple transcriptional start sites are observed, perhaps as a result of a random response to the lack of a strong selector (Sitzler *et al*., 1991; Ince and Scotto, 1995; Lin *et al*., 2001). Since DNase footprinting analyses show that the recognition ability of TBP to TATA is not significantly altered under iron treatment, these data suggest that TBP can still recognize TATA in iron overload conditions, but loses the ability to direct RNA pol with accuracy. Thus, cellular iron concentration could alter transcription regulation by modifying interactions among *trans*-factors rather than between *trans*-factors and *cis*-elements.

For many years, the prevailing view in eukaryotic gene expression studies has been that response elements, specifically enhancers and silencers, are distance and orientation independent. However, recent work has challenged this tenet. Many newly found enhancers and silencers are both orientation and distance dependent (Surinya *et al*., 1998; Guevara-Garcia *et al*., 1999; Coughlin *et al*., 2000; Givogri *et al*., 2000; Lin and Chang, 2000; Wei and Brennan, 2000; Sun *et al*., 2001; Takahashi *et al*., 2001). Some elements even have the unusual property of switching activity with orientation. These elements act as an enhancer in one orientation and as a silencer in the other, as well as operating at dual promoters of two adjacent genes (Guevara-Garcia *et al*., 1999; Givogri *et al*., 2000). In the dipterans, the ferritin genes are located adjacent to, and in opposite directions from, each other (Dunkov and Georgieva, 1999; Holt *et al*., 2002). The enhancers and silencers of the HCH gene are orientation dependent (D.Q.-D. Pham, unpublished data). Preliminary data suggest that the response elements of these genes could play a role in the control of the expression of both LCH and HCH genes. Since the genes are located adjacent to each other, a switch in activity of some response elements could allow the elements to exert a dual and/or directional control over the ferritin genes.

4.10.4.2. Posttranscriptional Control: Alternative Splicing and Polyadenylation

Available evidence indicates that posttranscriptional control of ferritin synthesis also occurs in insects (Zhang *et al*., 2001b). Possible mechanisms for posttranscriptional control include alternative splicing (Lind *et al*., 1998; Georgieva *et al*., 1999), use of multiple polyadenylation sites (Lind *et al*., 1998; Georgieva *et al*., 1999, 2002b; Dunkov *et al*., 2002), and translational repression by an iron-responsive element (IRE) (Zhang *et al*., 2001a, 2002).

In mammals, the IRE is a *cis*-acting, stem–loop sequence that allows iron-mediated repression of ferritin translation (Leibold and Munro, 1988; Theil, 1994). A single IRE is found in the 5′-untranslated region (UTR) of mammalian ferritin mRNA (Caughman *et al*., 1988). An IRE also is present in the 5′-UTR of the ferritin HCH message of *D. melanogaster* (Charlesworth *et al*., 1997), *A. aegypti* (Dunkov *et al*., 1995), *A. gambiae* (Georgieva *et al*., 2002a), *N. lugens* (Du *et al*., 2000), and *G. mellonella* (Kim *et al*., 2001), as well as the HCH and LCH messages of *M. sexta* (Pham *et al*., 1996; Zhang *et al*., 2001a) and *C. ethlius* (Nichol and Locke, 1999).

In contrast to these ferritin messages, the LCH message of the dipterans, *Drosophila* (Georgieva *et al*., 2002b) and *Aedes* (Geiser *et al*., 2003), lacks a canonical IRE. Although not seen in mammals, ferritin subunits without IREs are found in other organisms (Dietzel *et al*., 1992). The LCH message in both dipterans is increased in response to iron. However, the protein does not appear to increase in proportion to message upregulation, suggesting that in addition to transcriptional control, other mechanisms control the synthesis of these subunits. This is supported by the finding that the *Drosophila* LCH is encoded by two mRNAs that differ in length as a result of using alternative polyadenylation sites (Georgieva *et al*., 2002b). Iron loading significantly increases the shorter message and induces both transcripts at all life stages except dark pupae. Vertebrate ferritin mRNA levels can be increased through the binding of proteins to conserved polypyrimidine

regions in the 3′-UTR that increase ferritin mRNA stability (Ai and Chau, 1999), but whether this is the case for these messages in this insect is unknown.

In contrast to the *Drosophila* LCH message, the *Drosophila* HCH message has a canonical IRE (Charlesworth *et al*., 1997). Both Lind *et al*. (1998) and Georgieva *et al*. (1999) reported that this IRE could be removed by alternative splicing. In fact, eight messages for the ferritin HCH exist in *Drosophila* that originate from three alternative splicing sites coupled with either of the two polyadenylation sites (Lind *et al*., 1998). In all cases, alternative splicing removes the IRE. Iron treatment increases splicing activity in a tissue-specific manner with the greatest increase in spliced messages occurring in the gut (Georgieva *et al*., 1999). These authors suggest that when a high level of HCH is desirable for the rapid storage of iron, alternative splicing of the IRE allows unhindered expression of this subunit.

In contrast to the *Drosophila* HCH message, Dunkov *et al*. (2002) reported no alternative splicing of the *Aedes* HCH message. They found only two messages for *Aedes* HCH that occur as a result of using different polyadenylation sites. Both messages increase in females following iron supplementation or blood feeding (Dunkov *et al*., 2002). The longer mRNA increases in concentration progressively after blood feeding and is sustained for 7 days. While the shorter mRNA increases progressively, it exceeds the levels of the longer message at 48 h and then declines. Both mRNAs are present in ovaries, midgut, and thorax, with greatest expression noted in midgut. Only the larger mRNA is observed in Malpighian tubules and fat body.

4.10.4.3. Translational Control

In mammals, repression of ferritin synthesis occurs when intracellular iron levels are low because an iron regulatory protein, IRP1 or IRP2, binds to the ferritin mRNA IRE and prevents translation (Gray and Hentze, 1994; Muckenthaler *et al*., 1998a). When intracellular iron levels increase and iron storage is desirable, ferritin synthesis increases as IRP1 binding activity declines (Eisenstein, 2000) and IRP2 is rapidly degraded (Guo *et al*., 1995). Although no measurable difference in IRP1 message or protein levels generally occurs in response to iron administration (Tang *et al*., 1992; Chen *et al*., 1997), binding activity declines because a cubane (4Fe–4S) iron sulfur cluster forms in the IRP1 core that prevents IRP1–IRE interaction (Beinert *et al*., 1996; Haile, 1999). Iron sulfur cluster assembly allows IRP1 to respond to iron levels, changes in oxidative stress, and redox potential (Cairo *et al*., 1995; Bouton, 1999; Eisenstein, 2000). When the cluster is present, IRP1 becomes cytoplasmic aconitase (Beinert *et al*., 1996), and recent work indicates that this enzyme is important in the generation of cellular NADPH (Narahari *et al*., 2000).

In 1990, a protein with binding activity to transcripts of the vertebrate ferritin IRE was detected in cytoplasmic extracts of *D. melanogaster* Schneider cells (Rothenberger *et al*., 1990). IRP1 was subsequently cloned and sequenced from *Drosophila* (Muckenthaler *et al*., 1998b), *Aedes* (Zhang *et al*., 2002), and *Manduca* (Zhang *et al*., 2001b). The predicted proteins all show high homology to the mammalian IRP1 (Zhang *et al*., 2001b). The residues necessary for iron sulfur cluster formation and aconitase activity are universally conserved, whereas a 72 amino acid insert that distinguishes the mammalian IRP2 from IRP1 is missing from the insect IRP sequences. The *Drosophila* IRP1 is expressed throughout the embryo and at all developmental stages (Muckenthaler *et al*., 1998b) and the mosquito IRP1 is expressed at all developmental stages (Zhang *et al*., 2002). No IRP2 has been identified in insects despite an exhaustive search.

Current data indicate that the insect IRP1 can regulate mRNA translation by binding to the insect IRE. Recombinant *M. sexta* or *A. aegypti* IRP1 (Zhang *et al*., 2001a, 2002) binds to transcripts of the insect or vertebrate ferritin IRE, and both repress ferritin translation *in vitro* by specific interaction with the IRE (Zhang *et al*., 2001a; D. Zhang and J.J. Winzerling, unpublished data). Recently, we found that an iron sulfur cluster forms in the recombinant *M. sexta* IRP1 exposed to iron and sulfide *in vitro* (Gailer, George, Pickering, and Winzerling, unpublished data). When the cluster is present, the IRP1 will not prevent the translation of *M. sexta* ferritin HCH *in vitro* (D. Zhang and J.J. Winzerling, unpublished data). *In vivo* data also support the postulation that IRP–IRE interaction is involved in iron-mediated control of ferritin synthesis. When *M. sexta* larvae are injected with iron, fat body IRP1–IRE binding activity declines without a change in IRP1 mRNA or protein expression and hemolymph ferritin increases (Winzerling *et al*., 1995; Zhang *et al*., 2001b). Similarly, blood-fed *A. aegypti* females show increased ferritin synthesis without a change in IRP1 mRNA expression (Dunkov *et al*., 1995; Zhang *et al*., 2002).

The mechanism whereby IRP–IRE interaction allows repression of ferritin translation in insects was studied recently (Nichol and Winzerling, 2002). In vertebrates, the ferritin IRE is within 60 nucleotides of the mRNA cap site; complete repression of ferritin translation occurs because

IRP–IRE interaction prevents recruitment of the full ribosomal apparatus (Gray and Hentze, 1994; Muckenthaler *et al*., 1998a). In constructs where the IRE is positioned more distant from the cap site, the ribosomal apparatus assembles and scans the message, but pauses at the region of IRP1–IRE interaction resulting in only partial repression of translation. Insect ferritin mRNA IREs are located distant from the cap site, yet complete repression of translation occurs. Current evidence suggests that the secondary structure of the insect ferritin mRNA could bring the IRE sufficiently close to the cap site to allow complete translational repression by the IRP1 (Nichol and Winzerling, 2002).

Interestingly, a 5′ UTR IRE also is found in the mRNA of the succinate dehydrogenase subunit b (SDhb, *D. melanogaster*) (Kohler *et al*., 1995; Gray *et al*., 1996; Melefors, 1996). *In vivo* synthesis of SDhb is increased following iron treatment of *Drosophila* cells indicating that the IRE is an active control site (Gray *et al*., 1996; Melefors, 1996). Under low iron conditions, the SDhb message is distributed to a nonpolysomal pool (Kohler *et al*., 1995). The SDhb IRE also confers iron-mediated translational control of the synthesis of a reporter protein in transfected cells (Kohler *et al*., 1995; Gray *et al*., 1996). These findings support the concept that the IRE is an active site of iron-mediated translational control in insects.

4.10.5. Iron, Oxidative Stress, and Ferritin

Ferritin synthesis in vertebrates is responsive to several factors in addition to intracellular iron concentration. These factors include some hormones (Chazenbalk *et al*., 1990a, 1990b), infection, oxidative stress (Beinert and Kiley, 1999; Bouton, 1999; Epsztejn *et al*., 1999; Haile, 1999; Cheng *et al*., 2000; Tsuji *et al*., 2000; Mueller *et al*., 2001), and hemin (Lin *et al*., 1990; Coccia *et al*., 1992; Beaumont *et al*., 1994a).

Free heme is a powerful generator of ROS that can damage biological molecules (Oliveira *et al*., 2002). Oliveira and colleagues found that, in *R. prolixis,* heme from the blood meal is transported in the hemolymph bound to a heme-binding protein (Oliveira *et al*., 2000; Braz *et al*., 2001, 2002). The relationship of heme to ferritin synthesis in blood-feeding insects is of considerable interest because in mammals ferritin serves as a protective agent against oxidative stress (Galvani *et al*., 1995; Cairo *et al*., 1996), and exposure to ROS alters ferritin expression (Cairo *et al*., 1995; Lobreaux *et al*., 1995; Haile, 1999; Tsuji *et al*., 2000).

Treatment of Aag2 cells with hemin increases LCH and HCH messages significantly in a time-dependent manner (Geiser *et al*., 2003). In *A. aegypti* females, hemin added to an artificial meal also increases expression of the LCH subunit in pupae and the HCH 24 kDa subunit in larvae, pupae, and adult females (Dunkov *et al*., 2002). Interestingly, although treatment with hemin elicits expression of the HCH 26 kDA subunit in pupae, expression of the 26 kDa subunit is not induced in larvae and in adult females it is decreased. Taken together, these data show that heme induces ferritin expression in mosquitoes in a stage-specific manner and suggest that the processing of the 24 kDa subunit is influenced by hemin. The mechanisms responsible for these changes are not known, but could involve induction by both iron and oxidative stress.

Oxidative stress imposed by exposure of Aag2 cells to H_2O_2 upregulates HCH and LCH mRNAs significantly; induction is time and dose dependent and follows a similar pattern to that observed when the cells are treated with hemin (Geiser *et al*., 2003). Treatment of mammalian cells with H_2O_2 also results in upregulation of ferritin message by acting at the transcriptional level (Tsuji *et al*., 2000). It seems probable that the increase in LCH and HCH messages observed for mosquito cells is mediated at the transcriptional level as well. In mammals, although oxidative stress enhances ferritin mRNA expression, it simultaneously downregulates ferritin synthesis at the translational level by increased IRP1–IRE interaction (Tsuji *et al*., 2000). The overall change in ferritin levels that results from these two opposing effects appears to be related to the type and strength of the stimulus. Both nitric oxide (NO) and H_2O_2 enhance IRP–IRE interaction by different mechanisms (Pantopoulos and Hentze, 1995) that result in disassembly of the iron sulfur cluster (Weiss *et al*., 1993, 1998; Pantopoulos and Hentze, 1995; Wardrop *et al*., 2000). Whether similar mechanisms exist in insects is unknown.

In mammals, infection results in increased IRP1–IRE binding activity without a change in IRP1 protein (Phillips *et al*., 1996; Weiss *et al*., 1997). This appears to be the case in insects as well. Mosquito cells exposed to lipopolysaccaharide show an immune response accompanied by an increase in IRP1–IRE binding activity (Zhang *et al*., 2002). Injection of lepidopteran larvae with water results in a significant increase in fat body IRP1–IRE binding activity accompanied by a significant fall in hemolymph ferritin (Zhang *et al*., 2001b). Interestingly, NO levels are increased in mosquitoes in response to a blood meal infected with parasites (Luckhart *et al*., 1998). The increase in IRP1–IRE

interaction in insects in response to infection or injury could be related to an effect of NO on the IRP1. Alternatively, when *A. gambiae* cells are exposed to H_2O_2, thioredoxin reductase also is induced (Dimopoulos *et al.*, 2002). Work with mammalian cells indicates that NO and thioredoxin reductase cooperate to modulate changes in IRP1–IRE interaction. If infection of insects with a blood meal induces both NO and thioredoxin reductase, these compounds could work together to enhance IRP1–IRE interaction and reduce ferritin synthesis at a time when iron is supplied in the blood meal.

Recent work on the effects of infection and oxidative stress in *A. gambiae* shows that these two factors primarily influence expression of different gene clusters (Dimopoulos *et al.*, 2002). Moreover, different types of infectious agents provoke expression of different gene clusters. Such information provides the opportunity to identify the role of the genes involved in these important response mechanisms. It will be interesting to see if genes regulated by iron overlap with genes of these clusters.

Acknowledgments

Research for this publication was supported by the US Public Health Service (GM 56812, GM55886, and GM 65861), the Agricultural Experiment Station and the College of Agriculture and Life Sciences and the Center for Insect Science of the Arizona Research Laboratories at the University of Arizona, and the University of Wisconsin-Parkside. The authors thank E. Kohlhepp, Dawn Geiser, and Jonathan Mayo for helpful comments on the manuscript; and acknowledge that parts of the text are reworked from Nichol *et al.* (2002).

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., *et al.*, **2000**. The genome sequence of *Drosophila melanogaster*. *Science 287*, 2185–2195.

Ai, L.S., Chau, L.Y., **1999**. Post-transcriptional regulation of H-ferritin mRNA. *J. Biol. Chem. 274*, 30209–30214.

Aisen, P., Enns, C., Wessling-Resnick, M., **2001**. Chemistry and biology of eukaryotic iron metabolism. *Int. J. Biochem. Cell Biol. 33*, 940–959.

Andrews, S.C., Arosio, P., Bottke, W., Briat, J.-F., Von Darl, M., *et al.*, **1992**. Structure, function, and evolution of ferritins. *J. Inorg. Biochem. 47*, 161–174.

Bartfeld, N.S., Law, J.H., **1990**. Isolation and molecular cloning of transferrin from the tobacco hornworm, *Manduca sexta*. *J. Biol. Chem. 265*, 1–8.

Beaumont, C., Jones, R., Seyhan, A., Grandchamp, B., **1994a**. A hemin-inducible enhancer lies 4.5 Kb upstream of the mouse ferritin H subunit gene. *Adv. Exp. Med. Biol. 356*, 211–218.

Beaumont, C., Seyhan, A., Yachou, A.K., Grandchamp, B., Jones, R., **1994b**. Mouse ferritin H subunit gene: functional analysis of the promoter and identification of an upstream regulatory element active in erythroid cells. *J. Biol. Chem. 269*, 20281–20288.

Beinert, H., Kennedy, M.C., Stout, C.D., **1996**. Aconitase as iron–sulfur protein, enzyme, and iron-regulatory protein. *Chem. Rev. 96*, 2335–2373.

Beinert, H., Kiley, P.J., **1999**. Fe–S proteins in sensing and regulatory functions. *Curr. Opin. Chem. Biol. 3*, 152–157.

Bevilacqua, M.A., Giordano, M., D'Agostino, P., Santoro, C., Cimino, F., *et al.*, **1992**. Promoter for the human ferritin heavy chain-encoding gen (*FERH*): structure and functional characterization. *Gene 111*, 255–260.

Bouton, C., **1999**. Nitrosative and oxidative modulation of iron regulatory proteins. *Cell. Mol. Life Sci. 55*, 1043–1053.

Braz, G.R., Abreu, L., Masuda, H., Oliveira, P.L., **2001**. Heme biosynthesis and oogenesis in the blood-sucking bug, *Rhodnius prolixus*. *Insect Biochem. Mol. Biol. 31*, 359–364.

Braz, G.R., Moreira, M.F., Masuda, H., Oliveira, P.L., **2002**. *Rhodnius* heme-binding protein (RHBP) is a heme source for embryonic development in the blood-sucking bug *Rhodnius prolixus* (Hemiptera, Reduviidae). *Insect Biochem. Mol. Biol. 32*, 361–367.

Briat, J.F., **1996**. Roles of ferritin in plants. *J. Plant Nutr. 19*, 1331–1342.

Cairo, G., Castrusini, E., Minotti, G., Bernellizazzera, A., **1996**. Superoxide and hydrogen peroxide-dependent inhibition of iron regulatory protein activity: a protective stratagem against oxidative injury. *FASEB J. 10*, 1326–1335.

Cairo, G., Tacchini, L., Pogliaghi, G., Anzon, E., Tomasi, A., *et al.*, **1995**. Induction of ferritin synthesis by oxidative stress: transcriptional and post-transcriptional regulation by expansion of the "free" iron pool. *J. Biol. Chem. 270*, 700–703.

Capurro, M.d.L., Iughetti, P., Ribolla, P.E., de Bianchi, A.G., **1996**. *Musca domestica* hemolymph ferritin. *Arch. Insect Biochem. Physiol. 32*, 197–207.

Caughman, S.W., Hentze, M.W., Rouault, T.A., Harford, J.B., Klausner, R.D., **1988**. The iron-responsive element is the single element responsible for iron-dependent translation regulation of ferritin biosynthesis. *J. Biol. Chem. 263*, 19048–19052.

Charlesworth, A., Georgieva, T., Gospodov, I., Law, J.H., Dunkov, B.C., *et al.*, **1997**. Isolation and properties of *Drosophila melanogaster* ferritin: molecular cloning of a cDNA that encodes one subunit, and localization of the gene on the third chromosome. *Eur. J. Biochem. 247*, 470–475.

Chasteen, N.D., Harrison, P.M., **1999**. Mineralization in ferritin: an efficient means of iron storage. *J. Struct. Biol. 126*, 182–194.

Chazenbalk, G.D., Wadsworth, H.L., Foti, D., Rapoport, B., **1990a**. Thyrotropin and adenosine 3′,5′-monophosphate stimulate the activity of the ferritin-H promoter. *Mol. Endocrinol. 4*, 1117–1124.

Chazenbalk, G.D., Wadsworth, H.L., Rapoport, B., **1990b**. Transcriptional regulation of ferritin H messenger RNA levels in FRTL rat thyroid cells by thyrotropin. *J. Biol. Chem. 265*, 666–670.

Chen, O.S., Schalinske, K.L., Eisenstein, R.S., **1997**. Dietary iron intake modulates the activity of iron regulatory proteins and the abundance of ferritin and mitochondrial aconitase in rat liver. *J. Nutr. 127*, 238–248.

Cheng, Q., Gonzalez, P., Zigler, J.S., Jr., **2000**. High level of ferritin light chain mRNA in lens. *Biochem. Biophys. Res. Commun. 270*, 349–355.

Chou, C.C., Gatti, R.A., Fuller, M.L., Concannon, P., Wong, A., *et al.*, **1986**. Structure and expression of ferritin genes in a human promyelocytic cell line that differentiates *in vitro*. *Mol. Cell. Biol. 6*, 566–573.

Coccia, E.M., Profita, V., Fiorucci, G., Romeo, G., Affabris, E., *et al.*, **1992**. Modulation of ferritin H-chain expression in Friend erythroleukemia cells: transcriptional and translational regulation by hemin. *Mol. Cell. Biol. 12*, 3015–3022.

Coccia, E.M., Stellacci, E., Orsatti, R., Testa, U., Battistini, A., **1995**. Regulation ferritin H-chain expression in differentiating Friend leukemia cells. *Blood 86*, 1570–1579.

Collin, O., Thomas, D., Flifla, M., Quintana, C., Gouranton, J., **1988**. Characterization of a ferritin isolated from the midgut epithelial cells of a homopteran insect, *Philaenus spumarious* L. *Biol. Cell 63*, 297–305.

Corsi, B., Perrone, F., Bourgeois, M., Beaumont, C., Panzeri, M.C., *et al.*, **1998**. Transient overexpression of human H- and L-ferritin chains in COS cells. *Biochem. J. 330*, 315–320.

Coughlin, B.C., Teixeira, S.M.R., Kirchhoff, L.V., Donelson, J.E., **2000**. Amastin mRNA abundance in *Trypanosoma cruzi* is controlled by a 3′-untranslated region position-dependent *cis*-element and an untranslated region-binding protein. *J. Biol. Chem. 275*, 12051–12060.

DeRusso, P.A., Philpott, C.C., Iwai, K., Mostowski, H.S., Klausner, R.D., *et al.*, **1995**. Expression of a constitutive mutant of iron regulatory protein 1 abolishes iron homeostasis in mammalian cells. *J. Biol. Chem. 270*, 15451–15454.

Dietzel, J., Hirzmann, M., Preis, D., Symmons, P., Kunz, W., **1992**. Ferritins of *Schistosoma mansoni*: sequence comparison and expression in female and male worms. *Mol. Biochem. Parasitol. 50*, 245–254.

Dimopoulos, G., Christophides, G.K., Meister, S., Schultz, J., White, K.P., *et al.*, **2002**. Genome expression analysis of *Anopheles gambiae*: responses to injury, bacterial challenge, and malaria infection. *Proc. Natl Acad. Sci. USA 99*, 8814–8819.

Du, J., Foissac, X., Carss, A., Gatehouse, A.M.R., Gatehouse, J.A., **2000**. Ferritin acts as the most abundant binding protein for snowdrop lectin in the midgut of rice brown planthoppers (*Nilaparvata lugens*). *Insect Biochem. Mol. Biol. 30*, 297–305.

Dunkov, B.C., Georgieva, T., **1999**. Organization of the ferritin genes in *Drosophila melanogaster*. *DNA Cell Biol. 18*, 937–944.

Dunkov, B.C., Georgieva, T., Yoshiga, T., Hall, M., Law, J.H., **2002**. *Aedes aegypti* ferritin heavy chain homologue: feeding of iron or blood influences message levels, lengths and subunit abundance. *J. Insect Sci. 2*, 7–17.

Dunkov, B.C., Zhang, D., Choumarov, K., Winzerling, J.J., Law, J.H., **1995**. Isolation and characterization of mosquito ferritin and cloning of a cDNA that encodes one subunit. *Arch. Insect Biochem. Physiol. 29*, 293–307.

Eisenstein, R.S., **2000**. Iron regulatory proteins and the molecular control of mammalian iron metabolism. *Annu. Rev. Nutr. 20*, 627–662.

Epsztejn, S., Glickstein, H., Picard, V., Slotki, I.N., Breuer, W., *et al.*, **1999**. H-ferritin subunit overexpression in erythroid cells reduces the oxidative stress response and induces multidrug resistance properties. *Blood 94*, 3593–3603.

Gallois, B., d'Estaintot, B.L., Michaux, M.-A., Dautant, A., Granier, T., *et al.*, **1997**. X-ray structure of recombinant horse L-chain apoferritin at 2.0 Å resolution: implication for stability and functions. *J. Biol. Inorg. Chem. 2*, 360–367.

Galvani, P., Colleoni, M., Origgi, M., Santagostino, A., **1995**. Mitochondrial toxicity of iron and the protective role of ferritin on dopaminergic PC12 cell line. *Tox. In Vitro 9*, 365–368.

Geiser, D.L., Chavez, C.A., Flores-Manguia, R.F., Winzerling, J.J., Pham, D.Q., **2003**. *Aedes aegypti* ferritin A cytotoxic protector against iron and oxidative challenge? *Eur. J. Biochem. 270*, 1–8.

Georgieva, T., Dunkov, B.C., Dimov, S., Ralchev, K., Law, J.H., **2002b**. *Drosophila melanogaster* ferritin: cDNA encoding a light chain homologue, temporal and tissue specific expression of both subunit types. *Insect Biochem. Mol. Biol. 32*, 295–302.

Georgieva, T., Dunkov, B.C., Harazanova, N., Ralchev, K., Law, J.H., **1999**. Iron availability dramatically alters the distribution of ferritin subunit messages in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA 96*, 2716–2721.

Georgieva, T., Dunkov, B., Law, J.H., **2002a**. *Anopheles gambia* secreted ferritin: characterization of cDNAs encoding two subunits, mRNA expression patterns, and gene organization. In: 4th Int. Symp. on Molecular Insect Science, Tuscon, AZ.

Givogri, M.I., Kampf, K., Schonmann, V., Campagnoni, A.T., **2000**. Identification of a novel silencer that regulates the myelin basic protein gene in neural cells. *Gene 252*, 183–193.

Gray, N.K., Hentze, M.W., **1994**. Iron regulatory protein prevents binding of the 43S translation pre-initiation complex to ferritin and eALAS mRNAs. *EMBO J. 13*, 3882–3891.

Gray, N.K., Pantopoulous, K., Dandekar, T., Ackrell, B.A., Hentze, M.W., **1996**. Translational regulation of mammalian and *Drosophila* citric acid cycle enzymes via iron-responsive elements. *Proc. Natl Acad. Sci. USA 93*, 4925–4930.

Guevara-Garcia, A., Lopez-Bucio, J., Herrera-Estrella, L., **1999**. The mannopine synthase promoter contains vectorial *cis*-regulatory elements that act as enhancers and silencers. *Mol. Gen. Genet. 262*, 608–617.

Guex, N., Diemand, A., Peitsch, M.C., **1999**. Protein modeling for all. *Trends Biochem. Sci. 24*, 364–367.

Guex, N., Peitsch, M.C., **1997**. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis 18*, 2714–2723.

Guo, B., Phillips, J.D., Yu, Y., Leibold, E.A., **1995**. Iron regulates the intracellular degradation of iron regulatory protein 2 by the proteasome. *J. Biol. Chem. 270*, 21645–21651.

Haile, D.J., **1999**. Regulation of genes of iron metabolism by the iron-response proteins. *Am. J. Med. Sci. 318*, 230–240.

Harrison, P.M., Arosio, P., **1996**. The ferritins: molecular properties, iron storage function and cellular regulation. *Biochem. Biophys. Acta 1275*, 161–203.

Hempstead, P.D., Yewdall, S.J., Fernie, A.R., Lawson, D.M., Artymiuk, P.J., *et al.*, **1997**. Comparison of the three-dimensional structures of recombinant human H and horse L ferritins at high resolution. *J. Mol. Biol. 268*, 424–448.

Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., *et al.*, **2002**. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science 298*, 129–149.

Hubbard, S.R., Mohammadi, M., Schlessinger, J., **1998**. Autoregulatory mechanisms in protein-tyrosine kinase. *J. Biol. Chem. 273*, 11987–11990.

Huebers, H.A., Huebers, E., Finch, C., Webb, B.A., Truman, J.W., *et al.*, **1988**. Iron binding proteins and their roles in the tobacco hornworm *Manduca sexta* (L). *J. Comp. Physiol. Biochem. 158*, 291–300.

Ince, T.A., Scotto, K.W., **1995**. Differential utilization of multiple transcription start points accompanies the overexpression of the P-glycoprotein-encoding gene in Chinese hamster lung cells. *Gene 156*, 287–290.

Kakhlon, O., Gruenbaum, Y., Cabatchik, Z.L., **2001**. Repression of ferritin expression increases the labile iron pool, oxidative stress and short-term growth of human erythroleukemia cells. *Blood 97*, 2863–2871.

Kakhlon, O., Gruenbaum, Y., Cabantchik, Z.I., **2002**. Ferritin expression modulates cell cycle dynamics and cell responsiveness to H-ras-induced growth via expansion of the labile iron pool. *Biochem. J. 363*, 431–436.

Kato, J., Niitsu, Y., **2002**. Recent advances in molecular iron metabolism: translational disorders of ferritin. *Int. J. Hematol. 76*, 1–5.

Kehrer, J.P., **2000**. The Haber–Weiss reaction and mechanisms of toxicity. *Toxicology 149*, 43–50.

Keim, C.N., Cruz-Landim, C., Carneiro, F.G., Farina, M., **2002**. Ferritin in iron containing granules from the fat body of the honeybees *Apis mellifera* and *Scaptotrigona postica*. *Micron 33*, 53–59.

Kim, B.S., Lee, C.S., Seol, J.Y., Yun, C.Y., Kim, H.R., **2002**. Cloning and expression of a 32 kDa ferritin from *Galleria mellonella*. *Arch. Biochem. Physiol. 51*, 80–90.

Kim, B.S., Yun, C.Y., Yeo, S.M., Lee, H.J., Kim, H.R., **2001**. Cloning and expression of a ferritin subunit for *Galleria mellonella*. *Arch. Biochem. Physiol. 47*, 8–17.

Kohler, S.A., Henderson, B.R., Kuhn, L.C., **1995**. Succinate dehydrogenase b mRNA of *Drosophila melanogaster* has a functional iron-responsive element in its 5′-untranslated region. *J. Biol. Chem. 270*, 30781–30786.

Kwak, E.L., Larochelle, D.A., Beaumont, C., Torti, S.V., Torti, F.M., **1995**. Role for NF-kappa B in the regulation of ferritin H by tumor necrosis factor-alpha. *J. Biol. Chem. 270*, 15285–15293.

Lawson, D.M., Treffry, A., Artymiuk, P.J., Harrison, P.M., Yewdall, S.J., *et al.*, **1989**. Identification of the ferroxidase centre in ferritin. *FEBS Lett. 254*, 207–210.

Lawson, D.M., Artymiuk, P.J., Yewdall, S.J., Smith, J.M.A., Livingston, J.C., *et al.*, **1991**. Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts. *Nature 349*, 541–544.

Leibold, E.A., Munro, H.N., **1988**. Cytoplasmic protein binds *in vitro* to a highly conserved sequence in the 5′ untranslated region of ferritin heavy- and light-subunit mRNAs. *Proc. Natl Acad. Sci. USA 85*, 2171–2175.

Levi, S., Corsi, B., Rovida, E., Cozzi, A., Santambrogio, P., *et al.*, **1994**. Construction of a ferroxidase center in human ferritin L-chain. *J. Biol. Chem. 269*, 30334–30339.

Lin, D.-L., Chang, C., **2000**. Identification of an essential *cis*-acting element (TR2-PACE) in the 5′ promoter of human TR2 orphan receptor gene. *Endocrine 12*, 89–95.

Lin, J.J., Daniels-McQueen, S., Patino, M.M., Gaffield, L., Walden, W.E., *et al.*, **1990**. Derepression of ferritin messenger RNA translation by hemin *in vitro*. *Science 247*, 74–77.

Lin, Y., Ince, T.A., Scotto, K.W., **2001**. Optimization of a versatile *in vitro* transcription assay for the expression of multiple start site TATA-less promoters. *Biochemistry 40*, 12959–12966.

Lind, M.I., Ekengren, S., Melefors, O., Soderhall, K., **1998**. *Drosophila* ferritin mRNA: alternative RNA splicing regulates the presence of the iron-responsive element. *FEBS Lett. 436*, 476–482.

Lobreaux, S., Thoiron, S., Briat, J.F., **1995**. Induction of ferritin synthesis in maize leaves by an iron-mediated oxidative stress. *Plant J. 8*, 443–449.

Locke, M., **1991**. Apoferritin in the vacuolar system of insect hemocytes. *Tissue Cell 23*, 367–375.

Locke, M., **2003**. Surface membranes, Golgi complexes, and vaculoar systems. *Annu. Rev. Entomol. 48*, 1–27.

Locke, M., Ketola-Pirie, C., Leung, H., Nichol, H., **1991**. Vacuolar apoferritin synthesis by the fat body of an insect (*Calpodes ethlius*). *J. Insect Physiol. 37*, 297–309.

Locke, M., Leung, H., **1984**. The induction and distribution of an insect ferritin: a new function for the endoplasmic reticulum. *Tissue Cell 16*, 739–766.

Locke, M., Nichol, H., **1992**. Iron economy in insects. *Annu. Rev. Entomol. 37*, 195–215.

Luckhart, S., Vodovotz, Y., Cui, L., Rosenberg, R., **1998**. The mosquito *Anopheles stephensi* limits malaria parasite development with inducible synthesis of nitric oxide. *Proc. Natl Acad. Sci. USA 95*, 5700–5705.

Melefors, O., **1996**. Translational regulation *in vivo* of the *Drosophila melanogaster* mRNA encoding succinate dehydrogenase iron protein via iron responsive elements. *Biochem. Biophys. Res. Commun. 221*, 437–441.

Miller, L.L., Miller, S.C., Torti, S.V., Tsuji, Y., Torti, F.M., **1991**. Iron-independent induction of ferritin H chain by tumor necrosis factor. *Proc. Natl Acad. Sci. USA 88*, 4946–4950.

Muckenthaler, M., Gray, N.K., Hentze, M.W., **1998a**. IRP-1 binding to ferritin mRNA prevents the recruitment of the small ribosomal subunit by the cap-binding complex eIF4F. *Mol. Cell 2*, 383–388.

Muckenthaler, M., Gunkel, N., Frishman, D., Cyrklaff, A., Tomancak, P., *et al.*, **1998b**. Iron-regulatory protein-1 (IRP-1) is highly conserved in two invertebrate species: characterization of IRP-1 homologues in *Drosophila melanogaster* and *Caenorhabditis elegans*. *Eur. J. Biochem. 254*, 230–237.

Mueller, S., Pantopoulos, K., Hubner, C.A., Stremmel, W., Hentze, M.W., **2001**. IRP1 activation by extracellular oxidative stress in the perfused rat liver. *J. Biol. Chem. 276*, 23192–23196.

Narahari, J., Ma, R., Wang, M., Walden, W.E., **2000**. The aconitase function of iron regulatory protein 1. Genetic studies in yeast implicate its role in iron-mediated redox regulation. *J. Biol. Chem. 275*, 16227–16234.

Nichol, H., Law, J.H., Winzerling, J.J., **2002**. Iron metabolism in insects. *Annu. Rev. Entomol. 47*, 535–559.

Nichol, H.K., Locke, M., **1989**. The characterization of ferritin in an insect. *Insect Biochem. Mol. Biol. 19*, 587–602.

Nichol, H.K., Locke, M., **1990**. The localization of ferritin in insects. *Tissue Cell 22*, 767–777.

Nichol, H., Locke, M., **1999**. Secreted ferritin subunits are of two kinds in insects molecular cloning of cDNAs encoding two major subunits of secreted ferritin from *Calpodes ethlius*. *Insect Biochem. Mol. Biol. 29*, 999–1013.

Nichol, H., Winzerling, J.J., **2002**. Structured RNA upstream of insect cap distal iron responsive elements enhances iron regulatory protein-mediated control of translation. *Insect Biochem. Mol. Biol. 32*, 1699–1710.

Oliveira, M.F., Silva, J.R., Dansa-Petretski, M., de Souza, W., Braga, C.M., *et al.*, **2000**. Haemozoin formation in the midgut of the blood-sucking insect *Rhodnius prolixus*. *FEBS Lett. 477*, 95–98.

Oliveira, M.F., Timm, B.L., Machaldo, E.A., Miranda, E.A., Attias, M., *et al.*, **2002**. On the pro-oxidant effect of haemozoin. *FEBS Lett. 512*, 139–144.

Pantopoulos, K., Hentze, M.W., **1995**. Nitric oxide signaling to iron-regulatory protein: direct control of ferritin mRNA translation and transferrin receptor mRNA stability in transfected fibroblasts. *Proc. Natl Acad. Sci. USA 92*, 1267–1271.

Pascoa, V., Oliveira, P.L., Dansa-Petretski, M., Silva, J.R., Alvarenga, P.H., *et al.*, **2002**. *Aedes aegypti* peritrophic matrix and its interaction with heme during blood digestion. *Insect Biochem. Mol. Biol. 32*, 517–533.

Peitsch, M.C., **1995**. Protein modeling by E-mail. *Biotechnology 13*, 658–666.

Peitsch, M.C., **1996**. ProMod and Swiss-Model: internet-based tools for automated comparative protein modelling. *Biochem. Soc. Trans. 24*, 274–279.

Peitzsch, R.M., McLaughlin, S., **1993**. Binding of acylated peptides and fatty acids to phospholipid-vesicles pertinence to myristoylated proteins. *Biochemistry 32*, 10436–10443.

Pham, D.Q.-D., **2000**. Molecular modeling of insect ferritins. *In Silico Biol., 5*31–544.

Pham, D.Q.-D., Shaffer, J.J., Chavez, C.A., Douglass, P.L., **2003**. Identification and mapping of the promoter for the gene encoding the ferritin heavy-chain homologue of the yellow fever mosquito *Aedes aegypt*i. *Insect Biochem. Mol. Biol. 33*, 51–62.

Pham, D.Q.-D., Winzerling, J.J., Dodson, M.S., Law, J.H., **1999**. Transcriptional control is relevant in the modulation of mosquito ferritin synthesis by iron. *Eur. J. Biochem. 266*, 236–240.

Pham, D.Q.-D., Zhang, D., Hufnagel, D.H., Winzerling, J.J., **1996**. *Manduca sexta* hemolymph ferritin: cDNA sequence and mRNA expression. *Gene 172*, 255–259.

Phillips, J.D., Kinikini, D.V., Yu, Y., Guo, B., Leibold, E.A., **1996**. Differential regulation of IRP1 and IRP2 by nitric oxide in rat hepatoma cells. *Blood 87*, 2983–2992.

Ponka, P., Beaumont, C., Richardson, D.R., **1998**. Function and regulation of transferrin and ferritin. *Sem. Hem. 35*, 35–54.

Rost, B., Casadio, R., Fariselli, P., Sander, C., **1995**. Transmembrane helices predicted at 95% accuracy. *Protein Sci. 4*, 521–533.

Rost, B., Sander, C., **1993**. Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol. 232*, 584–599.

Rost, B., Sander, C., **1994a**. Conservation and prediction of solvent accessibility in protein families. *Proteins 20*, 216–226.

Rost, B., Sander, C., **1994b**. Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins 19*, 55–72.

Rothenberger, S., Mullner, E.W., Kuhn, L.C., **1990**. The mRNA-binding protein which controls ferritin and

transferrin receptor expression is conserved during evolution. *Nucleic Acids Res. 18*, 1175–1179.

Santambrogio, P., Levi, S., Cozzi, A., Corsi, B., Arosio, P., **1996**. Evidence that the specificity of iron incorporation into homopolymers of human ferritin L- and H-chains is conferred by the nucleation and ferroxidase centres. *Biochem. J. 314*, 139–144.

Sitzler, S., Oldenburg, I., Petersen, G., Bautz, E.K.F., **1991**. Analysis of the promoter region of the housekeeping gene *DmRP140* by sequence comparison of *Drosophila melanogaster* and *Drosophila virilis*. *Gene 100*, 155–162.

Sun, G., Tomita, H., Shakkottai, V., Gargus, J.J., **2001**. Genomic orgnization and promoter analysis of human KCNN3 gene. *J. Hum. Genet. 46*, 463–470.

Surinya, K., Cox, T.C., May, B.K., **1998**. Identification and characterization of a conserved erythroid-specific enhancer located in intron 8 of the human 5–aminolevulinate synthase 2 gene. *J. Biol. Chem. 273*, 16798–16809.

Takahashi, K., Nishiyama, C., Nishiyama, M., Okumura, K., Ra, C., *et al.*, **2001**. A complex composed of USF1 and USF2 activates the human FcεRI α chain expression *via* a CAGCTG element in the first intron. *Eur. J. Immunol. 31*, 590–599.

Tang, C.K., Chin, J., Harford, J.B., Klausner, R.D., Rouault, T.A., **1992**. Iron regulates the activity of the iron-responsive element binding protein without changing its rate of synthesis or degradation. *J. Biol. Chem. 267*, 24466–24470.

Theil, E.C., **1987**. Ferritin: structure, gene regulation, and cellular function in animals, plants, and microorganisms. *Annu. Rev. Biochem. 56*, 289–315.

Theil, E.C., **1994**. Iron regulatory elements (IREs): a family of mRNA non-coding sequences. *Biochem. J. 304*, 1–11.

Torti, F.M., Torti, S.V., **2002**. Regulation of ferritin genes and protein. *Blood 99*, 3505–3516.

Trikha, J., Theil, E.C., Allewell, N.M., **1995**. High resolution crystal structures of amphibian red-cell L ferritin: potential roles for structural plasticity and solvation in function. *J. Mol. Biol. 248*, 949–967.

Trikha, J., Waldo, G.S., Lewandowski, F.A., Ha, Y., Theil, E.C., *et al.*, **1994**. Crystallization and structural analysis of bullfrog red cell L-subunit ferritins. *Proteins 18*, 107–118.

Trowler, D.A., Gordon, J.I., Adams, S.P., Glaser, L., **1988**. The biology and enzymology of eukaryotic protein acylation. *Annu. Rev. Biochem. 57*, 69–99.

Tsuji, Y., Akebi, N., Lam, T.K., Nakabeppu, Y., Torti, S.V., *et al.*, **1995**. FER-1, an enhancer of the ferritin H gene and a target of E1A-mediated transcriptional repression. *Mol. Cell. Biol. 15*, 5152–5164.

Tsuji, Y., Ayaki, H., Whitman, S.P., Morrow, C.S., Torti, S.V., *et al.*, **2000**. Coordinate transcriptional and translational regulation of ferritin in response to oxidative stress. *Mol. Cell. Biol. 20*, 5818–5827.

Tsuji, Y., Moran, E., Torti, S.V., Torti, F.M., **1999**. Transcriptional regulation of the mouse ferritin H gene: involvement of p300/CBP adaptor proteins in FER-1 enhancer activity. *J. Biol. Chem. 274*, 7501–7507.

Wardrop, S.L., Watts, R.N., Richardson, D.R., **2000**. Nitrogen monoxide activates iron regulatory protein 1 RNA-binding activity by two possible mechanisms: effect on the [4Fe–4S] cluster and iron mobilization from cells. *Biochemistry 39*, 2748–2758.

Wei, W., Brennan, M.D., **2000**. Polarity of transcriptional enhancement revealed by an insulator element. *Proc. Natl Acad. Sci. USA 97*, 14518–14523.

Wei, Y., Miller, S.C., Tsuji, Y., Torti, S.V., Torti, F.M., **1990**. Interleukin 1 induces ferritin heavy chain in human muscle cells. *Biochem. Biophys. Res. Commun. 169*, 289–296.

Weiss, G., Bogdan, C., Hentze, M.W., **1997**. Pathways for the regulation of macrophage iron metabolism by the anti-inflammatory cytokines IL-4 and IL-13. *J. Immunol. 158*, 420–425.

Weiss, G., Goossen, B., Doppler, W., Fuchs, D., Pantopoulos, K., *et al.*, **1993**. Translational regulation via iron-responsive elements by the nitric oxide/NO-synthase pathway. *EMBO J. 12*, 3651–3657.

Weiss, G., Widner, B., Zoller, H., Schobersberger, W., Fuchs, D., **1998**. Immune response and iron metabolism. *Br. J. Anaesth. 81*(Suppl. 1), 6–9.

Welch, K.D., Reilly, C.A., Aust, S.D., **2002**. The role of cysteine residues in the oxidation of ferritin. *Free Rad. Biol. Med. 33*, 399–408.

Winzerling, J.J., Law, J.H., **1997**. Comparative nutrition of iron and copper. *Annu. Rev. Nutr. 17*, 501–526.

Winzerling, J.J., Nez, P., Porath, J., Law, J.H., **1995**. Rapid and efficient isolation of transferrin and ferritin from *Manduca sexta*. *Insect Biochem. Mol. Biol. 25*, 217–224.

Yokomori, N., Iwasa, Y., Aida, K., Inoue, M., Tawata, M., *et al.*, **1991**. Transcriptional regulation of ferritin messenger ribonucleic acid levels by insulin in cultured rat glioma cells. *Endocrinology 128*, 1474–1480.

Zhang, D., Albert, D.W., Kohlhepp, P., Pham, D.Q.-D., Winzerling, J.J., **2001a**. Repression of *Manduca sexta* ferritin synthesis by IRP1/IRE interaction. *Insect Mol. Biol. 10*, 531–539.

Zhang, D., Dimopoulos, G., Wolf, A., Minana, B., Kafatos, F.C., *et al.*, **2002**. Cloning and molecular characterization of two mosquito iron regulatory proteins. *Insect Biochem. Mol. Biol. 32*, 579–589.

Zhang, D., Ferris, C., Gailer, J., Kohlhepp, P., Winzerling, J.J., **2001b**. *Manduca sexta* IRP1: molecular characterization and *in vivo* response to iron. *Insect Biochem. Mol. Biol. 32*, 85–96.

4.11 Tick-Talk, the Cellular and Molecular Biology of *Drosophila* Circadian Rhythms

P H Taghert and Y Lin, Washington University Medical School, Saint Louis, MO, USA

© 2005, Elsevier BV. All Rights Reserved.

4.11.1. Introduction

Circadian rhythms describe daily cycles of behavioral, physiological, or biochemical events that display a precise period, and that can persist without reference to environmental cues. They arise from endogenous biological clocks, the gears of which are driven by cycling genes and proteins, and whose free-running periods approximate 24 h. Environmental cues that register predictable time changes, such as the daily light/dark (LD) or temperature cycles (see **Chapter 3.11**), normally adjust the phase and period of the clock to maintain precise synchrony with the rotation of the earth. The registration of internal clocks with external events is termed "entrainment" (Pittendrigh, 1974).

Forward mutagenesis screens in *Drosophila* have identified a small handful of proteins whose principles of structural design subserve the central clock function, a daily rhythm of changes in their levels. Mutations in each of these genes result in circadian rhythm abnormalities, ranging from alterations in circadian period to complete arrhythmicity. The analysis of these proteins and their cognate genes has produced a wealth of information and insight. In sum, these findings have translated into detailed biochemical and genetic interactions. They have also begun to define general principles underlying the fundamental clock mechanisms. This subject has been reviewed extensively (e.g., recently by Young and Kay, 2001; Ashmore and Sehgal, 2003; Hall,

2003; Helfrich-Förster, 2003; Van Gelder *et al.*, 2003). That level of overview testifies to the sustained successes of circadian biology researchers, and to the general interest the field is able to stimulate.

In circadian biology, work on *Drosophila* has contributed the most to mechanistic studies of the central circadian pacemaker (the clockworks). However, there are two other important aspects of circadian biology that have begun to receive increasing attention from researchers in the past few years, and which also merit continued review: first, how circadian rhythms are synchronized with environmental cycles (the input end); and second, how they organize more distal rhythmic events (the output end). Therefore, this review is divided into sections that consider in turn clock mechanisms, their inputs, and their outputs. Separating circadian rhythms in this way facilitates presentation, but it also creates an impression of seams that likely do not correspond to biological reality. For example and as reviewed below, recent work has indicated that "input factors" can also act as central clockwork components, depending on the pacemakers studied. Likewise, *bona fide* "output factors" can also act as input factors, depending on which cells receive the timing signal. These issues reflect and represent complexities that presently are challenging to consider, and that eventually will be resolved and rationalized. At such a time, circadian biology will probably be presented and reviewed more seamlessly. In the meantime, the canonical three-part organization serves the present purposes well, and we have adopted it once more.

4.11.2. The Central Oscillator Described by Studies of *Drosophila*

4.11.2.1. An Overview of the Circadian Clockwork Mechanism in *Drosophila*

A brief description of the clockwork mechanism is as follows. *period* and *timeless* transcription are driven via a common *cis*-promoter element (the E-box) by the helix–loop–helix transcription factors *Clock* and *cycle*, working *in trans*. In the early night, Period (PER) and Timeless (TIM) accumulate in the cytoplasm; however, as individual proteins they are rapidly degraded. By a process of posttranslational modifications and protein–protein interaction, the TIM and PER are stabilized and their concentrations increase. Those increases promote the dimerization of TIM and PER, and the phosphorylation of TIM by Shaggy (SGG); the PER : TIM complex is imported (along with the Double-time (DBT)) kinase into the nucleus in the late night, where it binds to and inactivates the Clock : Cycle (CLK : CYC) dimer, causing transcription to halt. The PER : TIM dimer is destabilized by DBT phosphorylation, and PER and TIM are degraded via the proteosome. Their declining levels permit CLK : CYC dimers once again to bind to E-boxes and activate *per* and *tim* transcription. In the sections immediately following, these events and some of the evidence underlying the model are reviewed in greater detail.

4.11.2.2. The *period* and *timeless* Genes: Identification and Molecular Rhythms

4.11.2.2.1. *period* – the founding member of the clock genes The *period* gene was identified in pioneering genetic screens for mutations that affect circadian period (Konopka and Benzer, 1971). These and subsequent screens (e.g., Hamblen-Coyle *et al.*, 1989; Konopka *et al.*, 1994) recovered *per* alleles that either shorten or lengthen the period of rhythmicity, or eliminate it altogether (Baylies *et al.*, 1987; Yu *et al.*, 1987). Following its molecular cloning (Jackson *et al.*, 1986; Citri *et al.*, 1987), a conceptual translation related PER to the transcription factor proteins ARNT, AHR, and SIM by a novel sequence domain (Crews *et al.*, 1988; Hoffman *et al.*, 1991; Burbach *et al.*, 1992; Reyes *et al.*, 1992). That region was called the PAS domain, a name that acknowledges its founding members (*P*ER, *A*RNT, and *S*IM). Biochemical studies showed that the PER PAS domain mediates homodimerization (Huang *et al.*, 1993) and heterodimerization with the TIM protein (Gekakis *et al.*, 1995). PER lacks motifs typical of DNA-binding proteins (Jackson *et al.*, 1986; Citri *et al.*, 1987). *per* RNA and protein levels display robust diurnal and circadian rhythms (Siwicki *et al.*, 1988; Hardin *et al.*, 1990, 1992a; Zerr *et al.*, 1990; Edery *et al.*, 1994a). During a 24-h cycle under LD conditions, *per* RNAs are detectable between ZT11 and ZT18 (ZT = zeitgeber time where ZT0 = time of lights-on and ZT12 = time of lights-off). The *per* RNA rhythm displays a peak early in the evening, during ZT13–16. The trough to peak amplitude of this rhythm is about 10-fold. The levels of PER protein also display a circadian rhythm, but their accumulation and peak lag behind the RNA cycle by 6–8 h (Hardin *et al.*, 1990; Zerr *et al.*, 1990). The mechanisms and significance of separated peaks in *per* RNA versus PER protein has been a key point of analysis, which is described below.

4.11.2.2.2. *timeless* – the second circadian clock gene The second central clock gene isolated in flies was called *timeless* (*tim*). As with *per*, a null

mutation at *tim* results in arrhythmic individuals, indicating that *tim* is necessary to generate rhythmic behavior (Sehgal *et al*., 1994). There are at least 10 reported *tim* alleles including arrhythmic, short, long, and ultralong phenotypes (Sehgal *et al*., 1994; Rutila *et al*., 1996; Matsumoto *et al*., 1999; Rothenfluh *et al*., 2000a, 2000b). *tim* encodes a large protein of novel structure whose only recognizable domains were separate acidic and basic regions (Myers *et al*., 1995). *tim* RNA cycles roughly in phase with *per* RNA (Sehgal *et al*., 1995); TIM protein levels peak around ZT17–19 and PER protein levels peak a short while later, at about ZT19–21 (Zeng *et al*., 1996). The *tim* gene's function supports *per* RNA cycling (Sehgal *et al*., 1994) and PER protein cycling (Price *et al*., 1995). Together these observations provided important early clues that TIM and PER probably interact in their timekeeping functions.

Are the observed molecular rhythms of *per* and *tim* integral to the manifestation of behavioral rhythms? If so, it would follow that they should "track" alterations of behavioral rhythms that are observed in period-altering mutant stock. That prediction was borne out repeatedly. For example, PER protein in the per^S stock displayed altered peak times corresponding to the altered period of behavior in both LD and in constant darkness (DD; Zerr *et al*., 1990). Significantly, when alterations in PER or TIM levels were experimentally induced, the phase of behavioral rhythms was shifted (Edery *et al*., 1994b; Lee *et al*., 1996; Myers *et al*., 1996; Zeng *et al*., 1996; Suri *et al*., 1999). Those combined correlations and causal effects demonstrated that circadian behavioral rhythms are true read-outs of *per* and *tim* molecular rhythms.

4.11.2.3. The Autoregulatory Feedback Loop Model

The earliest genetic evidence suggested that the *per* gene product operates close to the heart of the circadian cycling machinery, but at what level it participates remained a mystery until a seminal discovery about a decade ago by Hardin *et al*. (1990). They showed that the *period* gene product regulates its own expression: this observation provided the essential clue to developing a model of the circadian clock mechanism. The PER protein feedbacks to affect *per* mRNA levels as indicated by the activity of a wild-type *per* transgene in a *per* null mutant. In the per^{01} stock, *per* RNA does not cycle and is stable at about 50% of its normal peak amplitude. Transgenic *per* activity restored cycling properties to that of per^{01} RNA – the mutant *per* RNA now displayed up to fourfold cycle amplitude (from a normal ~10-fold range). The findings led directly to a model in which PER protein affects a negative regulation on its own transcription, hence, a negative autoregulatory feedback loop. This concept was extended to incorporate the fundamental influence of TIM (Sehgal *et al*., 1995), and together these results suggested that fluctuating PER and TIM protein levels are a consequence of, and contribute to, the fluctuating levels of *per* and *tim* RNA (Hardin *et al*., 1990; Sehgal *et al*., 1995; Marrus *et al*., 1996). Cycles of *per* and *tim* gene expression therefore form the central elements of the *Drosophila* pacemaker – a feedback loop that meets the criteria set for a self-sustaining oscillator (Hardin *et al*., 1992a).

4.11.2.4. Mechanisms of Clock Gene Expression

4.11.2.4.1. Clock gene cycling largely reflects changes in transcription The initial observations of *per* RNA cycles were pursued to establish whether such fluctuations reflect transcriptional or posttranscriptional mechanisms. While both explanations have validity, evidence suggests that fluctuations in *per* gene transcription are the primary source for *per* mRNA oscillations. So and Rosbash (1997) showed that the rates of both *per* and *tim* transcription were circadianly regulated through the use of run-on assays. Further evidence of transcriptional control of the *per* gene was revealed by demonstrations that its upstream sequences confer circadian cycling onto heterologous (reporter) genes (Hardin *et al*., 1992a; Brandes *et al*., 1996). Hao *et al*. (1997, 1999) pursued the definition of *per* regulatory sequences by progressively reducing test fragments; they identified a 69-bp fragment capable of driving high-amplitude mRNA cycling under both LD and constant dark DD conditions. That transcriptional activity is *per* protein (PER)-dependent, but is not dependent on a conserved E-box found within the 69-bp fragment (Hao *et al*., 1997). Even more impressive was the demonstration that this small regulatory fragment, and even its isolated E-box, could produce correct spatial "*per*-like" expression (Hao *et al*., 1999; Darlington *et al*., 2000), and drive wild-type *per* sequences in a regulated manner sufficient to rescue behavioral arrhythmicity in per^{01} flies (Hao *et al*., 1999). Similar promoter analyses, including identification of relevant E-boxes, have also been reported for *tim* (McDonald *et al*., 2001).

4.11.2.4.2. Clock gene cycling also reflects post-transcriptional regulation Not only does *per* RNA rely upon transcriptional mechanisms, it also depends in part on a circadian regulation of *per* RNA stability (So and Rosbash, 1997; Stanewsky

et al., 1997; Chen *et al*., 1998; Suri *et al*., 1999). For example, the rate of *per* transcription rises faster than does the rate of *per* mRNA accumulation, suggesting phase-dependent differences in *per* mRNA stability (So and Rosbash, 1997). Modeling of the data indicates that the posttranscriptional effect contributes 20–40% of the total amplitude of the rhythm. Both TIM and PER proteins are likely to contribute to this stabilization of *per* RNAs (Suri *et al*., 1999). A posttranscriptional mechanism regulating *per* RNA was also invoked to explain phase differences in expression between PER-LUC fusion proteins versus PER-LUC fusion genes (the latter lacking PER amino acids; Stanewsky *et al*., 1997). The indicated *cis*-sequences map to the 5′ untranslated region of the *per* mRNA, within the first intron (Stanewsky *et al*., 2002). In all, there is good evidence to include posttranscriptional mechanisms in modeling the daily fluctuations in clock gene cycling, specifically, a clock-regulated temporal stabilization of *per* mRNA during its daily upswing in the morning.

4.11.2.5. *per* and *tim* RNA Cycles are not Required for the Circadian Autoregulatory Loop

The evidence discussed above showed that oscillations in PER are necessary for oscillation in *per* RNA (Hardin *et al*., 1990), but is the opposite also true, i.e., is the cycling of *per* RNA needed for PER protein rhythms and for rhythmic behavior? Several observations indicate the answer to this is no. Frisch *et al*. (1994) partially rescued behavioral rhythmicity using a *per* rescue construct that contained no 5′ upstream sequence. The resultant PER levels clearly cycled, while transgenic *per* RNA displayed only a low amplitude (~twofold) rhythm. Cheng and Hardin (1998) addressed the question by driving *per* with a constitutive promoter (*rhodopsin1*) in the eye. They rescued PER cycling in that tissue, with no evidence of cycling by the transgenic *per* RNA. Likewise, Vosshall and Young (1995) also rescued *per* arrhythmicity using a *glass* promoter that is known to be constitutively active, although transgenic *per* RNA levels were not directly measured in that study. Finally, Yang and Sehgal (2001) extended this idea to show that *tim* RNA cycles also are not required for behavioral rhythmicity. Even in the case of a per^{01}; tim^{01} double null mutant fly, behavioral rhythms could be partially rescued by constitutive expression of both *tim* and *per* genes. Each of these studies supports the conclusion that circadian oscillators arise through some posttranscriptional mechanism(s). Together they deny an absolute necessity for *per* or *tim* RNA cycles. A prediction from this general conclusion is that if PER or TIM protein cycles are swamped, then behavioral rhythms and pacemaker protein cycling should be attenuated or lost. That proved to be true following the use of very strong, constitutive promoters to drive PER and TIM to high levels in wild-type stocks (e.g., Zeng *et al*., 1994; Yang and Sehgal, 2001; Blanchardon *et al*., 2001).

per and *tim* RNA cycles, though not strictly required for a manifestation of rhythmicity, probably serve several useful functions (Cheng and Hardin, 1998). The strength and/or phasing of rhythmicity, in the absence of high amplitude clock gene RNA cycling, are often abnormal (e.g., Frisch *et al*., 1994; Vosshall and Young, 1995; Yang and Sehgal, 2001). Clock gene RNA cycling is also likely to be critical for the mechanisms that reset circadian phase according to light. Current models of that reset mechanism suggest different outcomes, depending on *tim* RNA levels (Hunter-Ensor *et al*., 1996; Lee *et al*., 1996; Myers *et al*., 1996; Zeng *et al*., 1996; see Section 4.11.3.2.1 for a description of these models).

4.11.2.6. Nuclear Translocation

4.11.2.6.1. PER and TIM nuclear translocation In considering the potential functions of PER and TIM in the clock mechanism, their subcellular locations were a good basis on which to formulate hypotheses. If the proteins remain cytoplasmic, the putative negative feedback on *per* RNA (Section 4.11.2.3) must have agency via other intermediary proteins. If, on the other hand, PER and/or TIM had a nuclear distribution, that would allow for consideration of their potential direct and/or indirect roles in such transcriptional regulation. All the available evidence suggests PER and TIM are nuclear proteins for at least a significant portion of the daily cycle. Using specific antibodies, PER and TIM were localized to the cytoplasm, but also to the nucleus at predictable times *in vivo* (Saez and Young, 1988; Siwicki *et al*., 1988; Zerr *et al*., 1990; Liu *et al*., 1992; Curtin *et al*., 1995; Shafer *et al*., 2002).

4.11.2.6.2. The importance of PER and TIM coexpression for nuclear translocation PER and TIM require each other's presence for nuclear translocation to occur normally. Evidence for these requirements came from genetic observations: PER is absent from the nucleus without TIM (Vosshall *et al*., 1994; Price *et al*., 1995; Saez and Young, 1996) and TIM's nuclear localization is likewise dependent on *per* function (Hunter-Ensor *et al*., 1996; Myers *et al*., 1996). This mutual requirement was strikingly illustrated in cultured (S2) *Drosophila* cells. When either TIM or PER proteins were

expressed independently, they accumulated in the cytoplasm; however, when coexpressed, both proteins translocated to the nuclei (Saez and Young, 1996).

The requirement for coexpression was most easily explained by a direct interaction between the PER and TIM proteins. Several observations *in vivo* and *in vitro* supported this model. Vosshall *et al*. (1994) studied PER–β-galactosidase (PER–β-gal) fusion proteins *in vivo* and noted that fusion proteins lacking PER amino acids 95 and 529 localized to nuclei in both wild-type and tim^{01} strains of flies. In contrast, fusion proteins containing this region of PER were nuclear in wild-type, but cytoplasmic in tim^{01} flies. This suggested that sequences mapping in the deleted section of PER normally inhibit its own nuclear localization, in the absence of *tim* activity (Vosshall *et al*., 1994). By the simplest model, TIM binds PER in this region, or permits some other factor to do so, and thus facilitates PER nuclear translocation.

4.11.2.6.3. The importance of PER and TIM heterodimerization for nuclear translocation Evidence for physical interactions between PER and TIM were first demonstrated *in vitro*. The *tim* gene was isolated in a two-hybrid screen for PER-interacting proteins and, in that same yeast interaction assay, TIM binds the PAS domain of PER (Gekakis *et al*., 1995). By deletion analysis, specific regions were identified in PER and TIM that are required for (1) nuclear entry and that (2) promote cytoplasmic retention (Saez and Young, 1996). These results indicated that each protein contains a nuclear localization signal (NLS) and a cytoplasmic localization domain (CLD). Remarkably, TIM also binds to the PER CLD region, as indicated by GST pull-down assays (Saez and Young, 1996). The same PER region that promotes cytoplasmic retention is also the site of interaction between the two proteins. In addition, the PER[L] protein isoform, which is defective in TIM binding, exhibits a delay in nuclear entry *in vivo* (Gekakis *et al*., 1995). Thus, suppression of cytoplasmic localization is accomplished by direct interaction of PER and TIM via sites that otherwise discourage nuclear entry. These observations helped establish the hypothesis that a critical checkpoint in the circadian cycle is the requirement for cytoplasmic assembly of a PER/TIM heterodimer (or complex) as a condition for nuclear transport of either protein. A consequence of this mechanism is that rates of PER and TIM synthesis and accumulation should influence the efficiency of PER/TIM heterodimerization and therefore the timing of nuclear localization.

In vivo, TIM and PER also interact in a predictable manner as a function of time of day, and PER appears to be the rate-limiting factor. By employing the technique of coimmunoprecipitation, Lee *et al*. (1996) and Zeng *et al*. (1996) showed that PER and TIM form a 1 : 1 heterodimer. Remarkably, at ZT20, ≥80% of the total amount of PER is bound to TIM in a 1 : 1 stoichiometric relationship. The evidence suggests that TIM is usually present in excess (1.5–1.8 times more TIM than PER) and is thus not rate-limiting; further, TIM is present in PER immunodepleted extracts (Zeng *et al*., 1996; Lee *et al*., 1998). Genetic evidence agrees with this conclusion: Smith and Konopka (1981) showed that decreasing *per* dosage lengthens circadian period, while increasing *per* dosage shortens it. PER is a circadian state variable in that its levels tell the organism the time of day. In contrast, TIM is not a state variable, as the clock is not sensitive to *tim* dosage (Rothenfluh *et al*., 2000a; Ashmore *et al*., 2003).

4.11.2.6.4. A recent revision to the model of the obligate PER : TIM heterodimer The simplest model predicts that PER : TIM dimers begin to enter the nucleus late in the day, and accumulate to a peak late at night. By several criteria heterodimer formation promotes that stable translocation and was therefore considered a requirement. But, in revision to the original model, PER is found in the nucleus of the principal brain pacemaker neurons 3–4 h prior to the nuclear detection of TIM (Shafer *et al*., 2002). The brain pacemakers in question are the small lateral cells (vLNs; see Section 4.11.4.2.1) and PER arrives in the nucleus at a time consistent with its participation in transcriptional repression (ZT18). This observation calls into the question the hypothesis of a requirement for the obligate PER : TIM heterodimer formation to enter the nucleus. That hypothesis was based primarily on biochemical analyses of head extracts, which presents a primary contribution by events occurring in photoreceptor pacemakers. An alternative mechanism (at least within LN pacemaker neurons) suggests that a PER : TIM heterodimer may enter the nucleus, but that TIM specifically may not be retained as efficiently as PER. Supporting observations were made by Ashmore *et al*. (2003). In S2 cells, TIM, but not PER, is subject to active nuclear export, as evidenced by the fact that TIM can enter the nucleus without PER, if nuclear export is blocked. Together these observations suggest that PER in the vLN pacemaker neurons may initially function to keep TIM in the cytoplasm as dimer formation progresses. It is then imported into the nucleus at a net rate faster than TIM and, later

in the cycle, it is required to permit TIM nuclear accumulation.

4.11.2.7. Posttranslational Modifications – Phosphorylation

If the PER : TIM dimer were able to accumulate in the nucleus directly following translation, negative regulation of *per* and *tim* gene expression would be exerted before the *per* and *tim* RNA levels had reached maximal levels, thereby dampening or eliminating oscillations (e.g., Sehgal *et al*., 1994; Leloup and Goldbeter, 1998). PER and TIM stability in the cytoplasm during the early evening therefore appears to be a major point of regulation to effect a lag in protein accumulation. The rate at which these proteins enter and/or accumulate in the nucleus is also likely to be critical. These points of regulation are processes likely to involve posttranslational modifications, and the one that has received the most experimental attention to date is phosphorylation.

4.11.2.7.1. PER and TIM phosphorylation PER displays a substantial decrease in gel mobility between ZT15 and ZT24 and that cycle continues in DD (Edery *et al*., 1994a). Phosphorylation underlies that mobility change, as the shift is entirely sensitive to phosphatases. During the period of phosphorylation, mobility changes appear continuously or repeatedly, which suggested the occurrence of multiple sequential phosphorylation events, perhaps in an interdependent fashion (Edery *et al*., 1994a). The most highly phosphorylated forms of PER occur at times of night when the proteins are predominantly nuclear. The functional significance of the phosphorylation of relevant clock proteins was further indicated by the demonstration of altered periodicities of phosphorylation rhythms that occur in per^S and per^L flies. The directions of the altered phosphorylation rhythms matched the directions of the altered behavioral rhythms. TIM also displays phosphorylation-dependent mobility shifts as a function of time of day, although the extent is not as dramatic as the case of PER, as indicated by the percentage mobility change (Myers *et al*., 1996; Zeng *et al*., 1996). To date, three kinases have been implicated in the modifications of PER and TIM, and their analysis has greatly informed the current models of circadian mechanisms.

4.11.2.7.2. Double-time A principal determinant of the subcellular movements and stability of PER over the course of the circadian cycle is the protein kinase encoded by the *double-time* (*dbt*) locus (Price *et al*., 1998). DBT is a kinase related to the vertebrate delta and epsilon isoforms of casein kinase I (Kloss *et al*., 1998). A mammalian ortholog (CK1 ε) is also known to have a circadian function, corresponding to the naturally occurring *tau* mutation of hamsters, which displays a short period (Lowrey *et al*., 2000). Genetic analyses of *dbt* functions in *Drosophila* suggest that DBT contributes to the generation of circadian rhythmicity by regulating multiple steps of clock biochemistry. To date, three main effects have been suggested. First, DBT helps to determine the ~8-h lag normally occurring between peaks of *per* RNA and protein. It binds to monomeric PER in the cytoplasm and slows its accumulation via phosphorylation, and later degradation (Kloss *et al*., 1998; Price *et al*., 1998). Second, a role for *dbt* in the turnover of nuclear PER has been proposed. PER is lost from wild-type nuclei but persists longer in *dbt* mutant nuclei (Price *et al*., 1998). Hypo-phosphorylated PER thus accumulates to high levels in *dbt* mutants and becomes a constitutive repressor in DD (Price *et al*., 1998; Rothenfluh *et al*., 2000b; Bao *et al*., 2001; Kloss *et al*., 2001; Suri *et al*., 2001). Third, DBT appears to affect the rate of nuclear accumulation of PER by delaying the rate at which PER translocates or is stabilized upon translocation (Bao *et al*., 2001; Suri *et al*., 2001).

Several alleles of *dbt* have been recovered, based mostly on chemical mutagenesis screens to detect circadian period alterations in heterozygous mutant flies (Price *et al*., 1998; Rothenfluh *et al*., 2000b; Suri *et al*., 2001). The semidominant alleles dbt^S and dbt^L display strong effects on circadian period (shortened and lengthened, respectively). Like per^S, the dbt^S mutation increases the amplitude of the phase-response curve (PRC) and shortens its period (Bao *et al*., 2001). These alleles influence the rate of PER phosphorylation in concert with their effects on behavior: phosphorylation is advanced in dbt^S and delayed in dbt^L. The rate of TIM phosphorylation appears much less affected. The basis for dbt^S and dbt^L semidominance is unknown: extra copies of wild-type *dbt* do not appreciably affect circadian period (Price *et al*., 1998). Kloss *et al*. suggest their semidominance may reflect a stable stoichiometric interaction between PER and DBT *in vivo*.

Cyclic associations of PER with DBT are critical in producing stereotyped changes in PER phosphorylation state and its stability. The *dbt* gene is highly pleiotropic and is allelic to the gene *discs large* (Zilian *et al*., 1999); homozygous mutants display pupal lethality. Antibody localization of DBT indicates that in the brain it is widely expressed, and found in those neurons that also express PER and TIM (Kloss *et al*., 2001). *dbt* RNA and DBT protein

levels do not cycle; however, DBT displays a robust circadian translocation from cytoplasm to nucleus in the LN pacemaker neurons (Kloss *et al*., 2001). DBT nuclear accumulation depends on expression of PER and not of TIM, which implies a physical association between DBT and PER. That inference is supported by genetic evidence that *per* short alleles specifically suppress the arrhythmic phenotype of the dbt^{AR} allele (Rothenfluh *et al*., 2000b). Furthermore, DBT interacts directly with PER *in vitro* and also within S2 cells. DBT: PER interactions *in vivo* can be detected by coimmunoprecipitations in the cytoplasm of tim^{01} heads (these are interactions with the residual PER that remains in that mutant). DBT is also detected in PER : TIM complexes, but does not appear to interact directly with TIM, as indicated by measurements in per^{01} heads (Kloss *et al*., 2001). Together these data suggest DBT is closely associated with monomeric PER, and later in the nucleus DBT is again closely associated with PER when it is complexed to TIM. One model has TIM inhibiting DBT phosphorylation of PER in the cytoplasm (Kloss *et al*., 2001). Then, TIM is slowly lost from the nuclear complex, which allows greater repression, but also more PER phosphorylation by DBT. Hyperphosphorylated PER is targeted for degradation.

Thus, DBT appears to promote good temporal order in the circadian system by delaying PER's normal trafficking and thus enforcing the lag of several hours normally seen between the times of peak *per* RNA versus peak PER protein. Promoting the instability of monomeric PER proteins, DBT thus allows PER accumulation only in conjunction with high titers of TIM. Lack of DBT function leads to a continuous, rather than a stably periodic, nuclear transport of PER : TIM complexes. The result of a continuous transport process would be a feedback loop at equilibrium rather than a dynamic molecular oscillator (Kloss *et al*., 1998). Accordingly, high amplitude cycles of *per* and *tim* RNA can only be achieved with normal DBT functions. Interestingly, in several *dbt* alleles, the lag between *per* RNA and protein peaks is nearly lost under constant conditions, despite the retention of behavioral rhythmicity (Bao *et al*., 2001; Suri *et al*., 2001). Both sets of authors conclude that the lag of total PER accumulation may not be especially critical, in comparison to the relative amount of phosphorylated PER in the nucleus.

A broader role for DBT in the definition of PER's spatial pattern of expression throughout the brain was suggested by observations reported by Price *et al*. (1998). PER expression in the hypomorphic allele dbt^{P} was widened to include regions of the brain not significantly stained in a wild-type background. They suggest that the effect is likely not due to changes in *per* transcription or RNA stability. Rather, they infer that normally there is a low level of *per* expression in a set of neurons, which is larger than that normally described. Diminution in the rate of PER degradation (here by loss of DBT activity) permits its visualization. Specifically, they relate this inferred pattern to similar ones seen with PER-BGAL fusion proteins (Kaneko *et al*., 1997 and see Section 4.11.4.2.1).

4.11.2.7.3. Shaggy The involvement of a second kinase called Shaggy (SGG)/Glycogen Synthase Kinase-3 was revealed in a gain-of-function screen by Martinek *et al*. (2001). They screened thousands of stocks for possible defects in circadian period when single genes were overexpressed in clock cells (using *tim*-GAL4). Overexpression of *sgg* produced a strong reduction in circadian period, and in complimentary experiments, reducing *sgg* levels lengthened period. SGG overexpression also shortened the cycle of TIM and PER nuclear entry to an extent predicted by the shortened behavioral rhythms. In addition, SGG overexpression also shortened the delay phase of the PRC, suggesting SGG affects a cytoplasmic event in the molecular oscillation mechanism. The primary biochemical event dependent on SGG appears to be TIM phosphorylation: hypomorphic *sgg* mutations alter TIM phosphorylation, and *sgg* overexpression elevates TIM phosphorylation *in vivo* (Martinek *et al*., 2001). The mammalian ortholog GSK3b phosphorylates TIM *in vitro*.

Martinek *et al*. suggest that, like DBT, SGG's principal influence is on the nuclear translocation of the TIM : PER heterodimer. By that scenario, these two kinases are prime actors, working in concert but at cross-purposes, to help determine the rate at which PER's negative feedback regulation occurs. While DBT retards that translocation to the nucleus, SGG accelerates it. Remarkably, in vertebrates, the ortholog of DBT and SGG coparticipate in the Wnt signaling pathway (Peifer *et al*., 1994; Siegfried *et al*., 1994; Sakanaka *et al*., 1999), signaling that appears otherwise unrelated to circadian periodicity.

4.11.2.7.4. Casein kinase II The analysis of PER phosphorylation in *dbt* mutants suggested the participation by other kinases (Price *et al*., 1998; Suri *et al*., 2001). Recently, two groups reported genetic evidence to implicate separate subunits of casein kinase II (CK2) in that pacemaker phosphorylation pathway (J.M. Lin *et al*., 2002; Akten *et al*., 2003).

The CK2 holoenzyme is a tetramer composed of two α (catalytic) and two β (regulatory) subunits. In *Drosophila*, CK2-α and CK2-β mutants both lengthen circadian period in loss-of-function states, while overexpression of the β form leads to a shortened period (Akten *et al*., 2003). CK2-β corresponds to a mutant called *Andante* that was originally identified in a genetic screen for alterations in circadian period (Konopka *et al*., 1991). Based on analogy to biochemical studies with the human enzyme, the β mutation is predicted to diminish the rate of association between CK2 subunits; experimentally, it leads to lowered levels of the β subunit but not the α subunit (Akten *et al*., 2003). A dominant mutation called *Timekeeper* (*Tik*), corresponding to CK2-α was also identified due to a period-altering phenotype (J.M. Lin *et al*., 2002), and it displays a genetic interaction with *per*. A revertant strain *TikR* behaved like a strong hypomorph. CK2 enzyme levels were reduced in *Tik* and *TikR* heterozygotes.

Bacterially expressed CK2-α directly phosphorylates PER *in vitro*, and to a lesser extent TIM also (Zeng *et al*., 1996; J.M. Lin *et al*., 2002). However, PER and TIM phosphorylation levels (mobilities) were not appreciably different in CK2-β mutant tissues (Akten *et al*., 2003). There are two molecular phenotypes associated with loss of CK2-β function. The first is increased PER/TIM abundance (Akten *et al*., 2003), consistent with a hypothesis that CK2 activity promotes PER and TIM degradation, akin to the actions of *dbt* and *sgg* (Zeng *et al*., 1996; Price *et al*., 1998; cf. Rothenfluh *et al*., 2000b; Martinek *et al*., 2001). The second phenotype is a delayed nuclear translocation of both PER and TIM proteins (J.M. Lin *et al*., 2002; Akten *et al*., 2003). It is likely that the mechanism by which CK2 affects circadian period involves this delayed PER/TIM nuclear entry. Both the molecular and behavioral oscillations are phase delayed by about 2 h. Interestingly, CK2 delays PER nuclear entry only in the small vLNs, but not in the large vLNs (Akten *et al*., 2003). Both groups suggest that CK2 might act both in the cytoplasm and nucleus of clock cells to determine the timing of nuclear entry and/or stability of clock proteins.

Both enzyme subunits are expressed preferentially in vLN pacemaker neurons and both display constitutive cytoplasmic expression throughout the day. Intriguingly, CK2 immunosignals are not restricted to the vLN cell bodies, but can also be found at high level in their axonal terminals (J.M. Lin *et al*., 2002). Those authors speculate that the enzyme has roles in addition to PER phosphorylation, perhaps in the processing or release of the neuropeptide PDF (described in Section 4.11.4.2.1).

4.11.2.7.5. PER degradation Phosphorylated PER is targeted for degradation by DBT and perhaps by other kinases (Price *et al*., 1998; Bao *et al*., 2001; Kloss *et al*., 2001; Suri *et al*., 2001). Two studies have shown that hyperphosphorylated PER is targeted to the proteasome by interactions with Slimb, an F-box/WD40-repeat protein functioning in the ubiquitin-proteasome pathway (Grima *et al*., 2002; Ko *et al*., 2002). In S2 cells, Slimb interacts preferentially with phosphorylated PER; overexpression of Slimb *in vivo* produces circadian behavioral phenotypes.

4.11.2.7.6. TIM degradation TIM is degraded in the proteasome via a tyrosine kinase-dependent step (Naidoo *et al*., 1999). This process is discussed below in the context of circadian photosensitivity (Section 4.11.3.2.1).

4.11.2.8. The Positive Clock Factors

4.11.2.8.1. Clock and cycle In flies, the *Clock* and *cycle* genes encode basic helix–loop–helix (bHLH) transcription factors and are critical for circadian rhythmicity and transcription of the *per* and *tim* genes (Allada *et al*., 1998; Bae *et al*., 1998; Darlington *et al*., 1998; Rutila *et al*., 1998). They are both members of the PER-ARNT-SIM (PAS) superfamily of transcription factors and their mammalian orthologs appear to play comparable functional roles (reviews: Stanewsky, 2003; Van Gelder *et al*., 2003). In both *Clk* and *cyc* mutants, nuclear run-on assays indicate that *per* transcription is severely reduced (Allada *et al*., 1998; Rutila *et al*., 1998). In those backgrounds, *per* and *tim* levels are as low as their normal trough levels, and much lower than the depressed amounts found in *per* or *tim* mutants, indicating that *Clk* and *cyc* are "epistatic to and upstream of *per* and *tim*" (Rutila *et al*., 1998). Of the two molecules, the sequence attributes of CLK suggest that it is likely to contain the functional transcriptional activator domains (Rutila *et al*., 1998).

4.11.2.8.2. The CLK : CYC dimer activates clock gene expression CLK and CYC activate *per* and *tim* transcriptions as a heterodimer, primarily through binding to *cis*-regulatory sequences, specifically termed as E-boxes, and to neighboring sequences. In S2 cells, *Clk* increased transcription of *tim-luciferase or per-luciferase* reporters in a manner that depended on the presence of the E-box in each promoter (Darlington *et al*., 1998). Lyons *et al*. (2000) used deletion analysis to scan the E-box containing 69 bp *per* regulatory fragment defined by Hao *et al*. (1997) and their results suggested

that the CLK–CYC target site actually extends beyond the E-box.

In S2 cells, transcriptional activity by CLK : CYC was repressed by coexpression of both *per* and *tim*, but not by either alone. In addition, *per* and *tim* expression without concomitant *Clk* expression did not affect *tim-luciferase* levels. CLK and CYC proteins bind to the E-box in a manner that depends on the presence of both proteins (Lee *et al*., 1999). PER, TIM, or both were able to block such binding, and did so efficiently when applied in a 1 : 1 molar ratio. Those data suggested that inhibition is due to binding of the inhibitors to one or both components of the CLK : CYC dimer. The available evidence suggests that repression by PER : TIM does not involve destabilization of the CLK : CYC heterodimer (Lee *et al*., 1999). *In vivo,* the rhythmic activity of the 69-bp *per* promoter fragment described by Hao *et al*. (1997) was dependent on both CLK and CYC (Allada *et al*., 1998; Rutila *et al*., 1998). These and other observations help to describe a model of a negative feedback loop in greater detail: CLK : CYC heterodimers transcriptionally activate the *per* and *tim* genes, whose protein products form a complex that subsequently represses that transcriptional activity (cf. Huang *et al*., 1993).

Clk RNA cycles in antiphase to those of *per* and *tim* RNAs; its levels peak late at night to early in the morning (ZT23 to ZT4), while *cyc* RNA levels do not exhibit rhythmicity (Bae *et al*., 1998). In the absence of either PER or TIM, *Clk* rhythms are abolished and its levels set at approximately trough values, suggesting that PER and TIM can function (directly or indirectly) as transcriptional activators (see Section 4.11.2.9.1). In agreement with the feedback model, the CLOCK protein undergoes circadian fluctuations in abundance, in phase with Cryptochrome (CRY) and in antiphase with PER and TIM. It is phosphorylated throughout a daily cycle, and interacts with PER, TIM, and/or the PER–TIM complex during the night, but not during most of the day (Lee *et al*., 1998). The time course for the accumulation of CLK and its RNA products is similar – this is in contrast to the phase lag seen for PER and TIM versus their RNAs. The peak to trough amplitudes of *Clk* cycles are higher in LD than DD (Lee *et al*., 1998), similar to differences seen for *per* and *tim* (Bae *et al*., 1998). These data all reflect an influence of light on the amplitude of clock gene expression rhythms (cf. Y. Lin *et al*., 2002; Section 4.11.4.1.2).

4.11.2.8.3. In the PER : TIM complex, which protein represses the CLK : CYC heterodimer? Most evidence to date suggests that PER is the agent of repression (Marrus *et al*., 1996; So and Rosbash, 1997; Lee *et al*., 1998; Rothenfluh *et al*., 2000c). A range of possible interactions are observed between TIM, PER, and CLK (Lee *et al*., 1998). Depending on the time in a daily cycle, PER, TIM, and preferentially the PER : TIM complex, can interact directly with CLK or a CLK-containing complex. In anti-CLK coimmunoprecipitations, PER and TIM are first detected at ZT12, followed by increases in amounts that reach peak values at ZT23.9. Also, compared to PER and TIM, CLK is present in limiting amounts during the night. Thus, the interaction of PER and TIM with CLK is restricted mainly to nighttime hours. At ZT20, the majority of the PER and TIM proteins that interact with CLK are in the form of a heterodimeric PER : TIM complex.

In per^{01}, TIM is still found in anti-CLK immunoprecipitants; in contrast, no PER is found associated with CLK in tim^{01}. Either there is too little PER in that mutant or PER requires TIM to bind CLK. Do PER or TIM provide repressing activity or do they only do so when working in concert? TIM is not able to repress CLK : CYC transcription without PER in S2 cells, although, *in vitro* TIM disrupts CLK : CYC binding to an E-box (Lee *et al*., 1999). Evidence that PER, and not TIM, is the transcriptional repressor of *Clk* came from studies of per^{S}. In that mutant, PER levels in the nucleus fall more quickly than do TIM levels in the nucleus (Marrus *et al*., 1996). This "per^{S} effect," which advances the subsequent *per* RNA peak, is consistent with the notion that monomeric PER might be the major transcriptional repressor. Likewise, PER cycling, but not TIM cycling, is strongly affected (i.e., displayed lowered amplitude) in the per^{01} background (So and Rosbash, 1997), suggesting that TIM levels are less tightly coupled to the transcriptional feedback loop and that the PER monomer or the PER–TIM heterodimer provides the relevant agent of negative feedback. In agreement with that general conclusion, Ashmore *et al*. (2003) showed that when TIM is localized to the nucleus of S2 cells in the absence of PER (by blocking nuclear export), it was not able to inhibit CLK : CYC transcriptional activation of an E-box reporter.

4.11.2.8.4. CLK levels and the phase of gene activation If CLK and CYC represent the positively acting factors in the circadian oscillation, the simplest model predicts that CLK levels should determine the rising phase of *per* and *tim* RNA cycles. In other words, CLK levels "should" commence at ZT5 and "should" peak between ZT10 and ZT12. However, and paradoxically, the levels of CLK decrease precisely when *per* and *tim* transcripts

accumulate; they are in roughly antiphase temporal alignment in LD (cf. So and Rosbash, 1997). Therefore, the CLK abundance cycle cannot be the main determinant of time at which *per* and *tim* expression increases. An explanation for this apparent paradox is suggested by Lee *et al.* (1998) invoking the importance of preexisting PER : TIM complexes. By this scenario, PER : TIM are extant at early phases of the CLK abundance cycle, and can therefore depress the transcriptional activity of the CLK : CYC heterodimer when CLK levels begin to rise.

4.11.2.9. A Second "Interlocked" Circadian Transcriptional Loop

4.11.2.9.1. Evidence for the second loop There are two interlocked negative feedback loops underlying circadian cycling (Glossop *et al.*, 1999). In the first loop (introduced in Section 4.11.2.3), genes that are activated by CLK are turned off by the PER : TIM complex; for example, the transcription of *per* and *tim* is turned off by PER : TIM repression of CLK : CYC activity (Darlington *et al.*, 1998; Lee *et al.*, 1999). In the second loop, genes that are repressed by CLK are turned on by the PER : TIM complex; for example, *Clk* transcription is activated by PER : TIM repression of CLK : CYC activity (Glossop *et al.*, 1999). The loops are interlocked in that both involve PER : TIM attenuation of CLK regulatory activity. *Clk* RNA levels peak just after dawn, roughly in antiphase with the peaks of *per* and *tim* RNAs (Bae *et al.*, 1998). The findings that underlie definition of the second loop involve *Clk* RNA measurements in single versus double mutant backgrounds. For example, *Clk* RNA levels are constitutively low in per^{01} and tim^{01} suggesting that the PER : TIM complex is required to activate *Clk* transcription (Bae *et al.*, 1998). However, *Clk* RNA levels are surprisingly high in Clk^{Jrk} and in cyc^{01}, indicating negative feedback by CLK on its own expression. Because *Clk* RNA levels are high even in per^{01}; Clk^{Jrk} or per^{01}; cyc^{01} double mutants, PER : TIM must act to activate *Clk* by derepressing CLK-mediated repression (Glossop *et al.*, 1999). In addition, the high levels of *Clk* mRNA in the double mutant stocks imply that a separate *Clk* activator is present (Glossop *et al.*, 1999). Blau, Young, Hardin, and colleagues have produced a series of reports that implicate two related proteins, VRILLE (VRI) and PDP1, as transcription factors that are targets of CLK, and that help to close this second circadian loop (Glossop *et al.*, 1999, 2003; Kim *et al.*, 2002; Cyran *et al.*, 2003).

VRI and PDP1 are hypothesized to have opposite actions: VRI is a repressor and PDP1 is an activator of *Clk* transcription. In doing so, they contribute to a sharpening of rhythmic transcriptional regulation for *Clk*, and possibly also for other clock-controlled genes. VRI levels peak 3–6 h before that of PDP1, and that phase difference is essential to the model of iterated actions by a repressor of *Clk*, then by an activator of *Clk*. In addition, the phenotypic consequences of increasing or decreasing *vri* levels genetically are complimentary to those manipulating *pdp1* levels. For example, raising *vri* levels coupled with a decrease in *pdp1* gene dosage leads to a synergistic lengthening of the circadian period (Cyran *et al.*, 2003). Significantly, VRI and PDP1 compete for access to a binding site on the *Clk* promoter, on which they act as predictable repressor and activator *in vitro*, respectively. These and other data strongly support a model whereby *Clk* activates *vri* and *pdp1*, and these two targets feedback with gene specific delays to generate a proper phasing for *Clk* transcriptional activity.

In further comment on the larger scope of the circadian mechanism, Allada *et al.* (2003) put forward the hypothesis that circadian rhythm amplitude and circadian rhythm period or phase are controlled separately. The posttranscriptional protein phosphorylation feedback loop (involving PER, TIM, SGG, DBT, and CK2) is primarily responsible for period determination. Its alleles display large period changes, whereas *Clk* and *cyc* do not. cyc^{0} does produce increased periods as a heterozygote (Rutila *et al.*, 1998) but that could be ascribed to a dominant negative gain of function. However, a *cyc* deletion/+phenotype is identical, indicating that the cyc^{0}/+phenotype is not due to a dominant mutant effect but rather to a dosage sensitive effect on period. A similar effect of a *Clk* deletion on circadian period is problematic due to its removal of neighboring clock-relevant genes, specifically *pdp1* (Allada *et al.*, 2003). So this proposition is challenging, but the issue remains unresolved.

4.11.2.9.2. The phase of the *Clk* RNA peaks is not a critical feature In spite of the exquisite control evidently in place to ensure the proper phasing of *Clk* RNA expression, there is good evidence that such RNA cycling is not strictly required for normal circadian periodicity in molecular cycling or in behavioral rhythmicity (Kim *et al.*, 2002). The evidence rests on misexpression of *Clk* using the *per*-GAL4 system to drive *Clk* at an inappropriate phase. Such a manipulation does not cause a disturbance in the period or phasing of molecular cycles or of behavioral rhythms. Analogous to the arguments that *per* and *tim* RNA cycles are not strictly required for behavioral or molecular

rhythmicity (cf., Section 4.11.2.5), these observations further the hypothesis that posttranscriptional events are central to the operation of a functional circadian clock. It is noteworthy that swamping *Clk* RNA cycles did produce alterations in certain aspects of light sensitivity (e.g., the phase-altering responses to light pulses delivered in the night). Such observations led to the surprising conclusion that *Clk* regulates the direct response of the locomotor system to light.

4.11.2.10. *In Situ* Expression of Clock Proteins

A great deal has been learned about the functions of clock proteins like PER by examining the cells and tissues in which they are expressed, and by analyzing their subcellular localization. Numerous studies over the years have detailed such information for PER and other clock proteins (Liu *et al*., 1988; Saez and Young, 1988; Siwicki *et al*., 1988; Zerr *et al*., 1990; Ewer *et al*., 1992; Liu *et al*., 1992; Kaneko *et al*., 1997; Rachidi *et al*., 1997; Kloss *et al*., 1998; Price *et al*., 1998; Blau and Young, 1999; Kaneko *et al*., 2000a; Cyran *et al*., 2003). PER is found in various cell types in the brain, typically associated with TIM (review: Helfrich-Förster, 2003), and also in diverse tissues (e.g., Hege *et al*., 1997; Plautz *et al*., 1997; Giebultowicz 1999; Giebultowicz *et al*., 2000; Myers *et al*., 2003). However, only in the cases of the antennal chemosensory cells (see **Chapter 3.15**) and prothoracic glands (see **Chapter 3.11**) have these peripheral oscillators been linked to functional outputs (Krishnan *et al*., 1999; Myers *et al*., 2003). PER's prominent nuclear localization was noted early on (Saez and Young, 1988; Liu *et al*., 1992) and are described in greater detail below, in considering both input pathways to the circadian pacemaking centers (Sections 4.11.3.3.2 and 4.11.3.4.1) and the output pathways (Sections 4.11.4.2.1 and 4.11.4.3). In a related vein, the issue of timing regarding when clock gene action "is required" was addressed by a conditional rescue protocol (Ewer *et al*., 1990). For adult behavioral rhythmicity, it was determined that *period* gene action is needed during the performance of the behavior, and not during antecedent developmental stages.

4.11.2.11. How to Build a Pacemaker: Lessons from the Misexpression of Clock Proteins

The proteins of the clock mechanism are, for the most part, limited in their spatial expression (e.g., Blau and Young, 1999; Kaneko and Hall, 2000; Kloss *et al*., 2001; Helfrich-Förster, 2003; Zhao *et al*., 2003). Recent experiments have asked whether the normal differentiation of a pacemaker cell is triggered by expression of a single "master" clock gene (Zhao *et al*., 2003). Zhao *et al*. misexpressed *Clk* and *per* (separately) and monitored for subsequent ectopic pacemaker activity. In most cases, misexpression of UAS-*Clk* was lethal (driven by promoters like *tim*-GAL4 or *per*-GAL4); these results may reflect the fact that such drivers likely produce patterns of expression greater than the "normal" patterns of their proteins (cf. Section 4.11.4.2.1). However, some drivers produced viable adults: when their brains were examined for other circadian gene expression, new patterns of pacemaker activity were observed (Zhao *et al*., 2003). Moreover, these ectopic pacemakers show diurnal variation of clock gene expression with phases similar to those displayed by normal pacemakers. These results were found using either a "circadian gene" driver (*cry*-GAL4; see Section 4.11.3.2.2) or a "noncircadian gene" driver (MJ162a; primarily expressed in mushroom bodies). Ectopic cycling was maintained in DD and locomotor rhythms in LD were disrupted, indicating that the ectopic clocks were autonomous, functionally activated, and somehow had access to locomotor centers.

Are these ectopic sites truly naïve for circadian gene expression or do they represent latent clocks that have, for example, low levels of *per* gene expression (cf. Price *et al*., 1998) and hence may be biased towards display of pacemaker function? The answer would lead to different conclusions regarding how pacemaker cells are organized. If the ectopic cells were truly naïve, then the selective expression of an additional regulatory molecule (e.g., CLK) could tip the balance and start a cascade that leads to an organized pacemaker function. If that scenario were true, then perhaps *Clk* does so in normal pacemaker cells as well during their differentiation. Because similar misexpression of *per* did not have this effect, *Clk* appears to have special properties in this regard (Zhao *et al*., 2003). Thus, if there are "master" clock genes, *Clk* may be unique in this respect, or it may simply be the first to be demonstrated. Further misexpression studies will help to resolve this significant point.

4.11.2.12. An Obligate Role for Membrane Excitability to Maintain Molecular Cycling in Pacemaker Cells

Physiological studies of cultured pacemaker neurons have measured rhythmic changes in membrane properties. In the snail *Bulla*, isolated retinal pacemakers display a robust rhythm in membrane conductance (Michel *et al*., 1993), which underlies a circadian rhythm of optic nerve activity. Electrical events at the membrane do more than reflect

an underlying oscillation: manipulations of the neurons' membrane potential affect both the period and phase of that rhythm (Block *et al*., 1995). That observation suggested that modulation of electrical activity in pacemakers represents both a circadian output and circadian input.

New data from genetic manipulations in *Drosophila* corroborate that hypothesis, and suggest that the effects of membrane excitability on pacemaker function *in vivo* may indeed be substantial (Nitabach *et al*., 2002). Nitabach *et al*. misexpressed a dominant negative form of a potassium channel (a "constitutively active" open rectifier channel). Such a molecule is predicted to enforce "electrical silence" by clamping the membrane at or near the K^+ equilibrium potential. Significantly, the neurons did not appear to be adversely affected by such treatment. As expected, when the critical vLN pacemaker neurons were silenced by driving UAS-*dORK* with *pdf*-GAL4, rhythmic behavior in DD was abolished (see Section 4.11.4.2.1 for a description of vLN cells). More surprising was the disappearance of molecular oscillations in those vLNs in DD and the retention of their oscillations in LD (Nitabach *et al*., 2002). These results indicate an essential role for electrical activity in the cycling of the clock, at least in the absence of environmental cues. This method permits designs to silence select neuronal populations to study their selective contributions to synaptic circuits. Further, the specific results force a reconsideration of the concept of a cell-autonomous circadian oscillator. Depending on the underlying mechanism, an absolute requirement for electrical activity may indicate that circadian pacemaker neurons depend on each other not just for synchronization, but also for sustained activity.

4.11.3. Molecular and Cellular Pathways for Circadian Photosensitivity

4.11.3.1. Introduction

The biological clock must be synchronized to local time, and it gains much of that information by sampling the intensity, spectral composition, and total daily duration of ambient light conditions. These features vary predictably as a function of time of day, with intensity high in midday, while low at dusk and dawn. Ambient light is also enriched for blue wavelengths during transition times, but for reds and greens at midday (Hall, 2000). It is likely therefore, that several mechanisms (including different photopigments) contribute to behavioral synchronization.

Studies of the earliest events in the fly's circadian photic sensitivity has now circumscribed a focus around the action of the blue-light photopigment CRY. That molecule acts primarily to change the status or activity of TIM. In addition, other photopigments collude with CRY. There is also now considerable genetic and anatomical evidence to implicate numerous parallel neural pathways as functional routes by which photic input, directly or indirectly, accesses critical pacemaking neurons. This section outlines recent progress made that addresses these critical aspects of clock function in *Drosophila*. This subject has been reviewed carefully over the past few years (Hall, 2000; Foster and Helfrich-Förster, 2001; Shafer, 2001; Ashmore and Sehgal, 2003) and the description here will be brief.

4.11.3.2. Light and Cryptochrome

4.11.3.2.1. A model for the light resetting mechanism Identification of the principal clock proteins permitted consideration of which molecular aspects display light sensitivity, and so may contribute to resetting and subsequent entrainment. There is now strong consensus that the TIM protein levels are lost rapidly upon exposure to light (Hunter-Ensor *et al*., 1996; Lee *et al*., 1996; Myers *et al*., 1996; Zeng *et al*., 1996). While TIM protein levels fall abruptly with light onset, PER levels decline at a much slower pace (Zerr *et al*., 1990; Edery *et al*., 1994a; Zeng *et al*., 1996). Neither *tim* nor *per* RNAs display such light sensitivity. TIM levels are light-sensitive within the first hour of the light signal, which is consistent with the pace at which light resets *Drosophila* behavioral rhythms. Decreases in TIM levels induced by pulses of light likely reflect a mechanism similar to that underlying the constitutive low TIM levels in flies kept in constant light (Price *et al*., 1995). Naidoo *et al*. (1999) reported that TIM degradation in response to light in the vLN pacemakers is mediated by the proteasome, and that TIM is phosphorylated and ubiquitinated prior to its degradation. Ubiquitination of TIM in response to light can also be observed in cultured S2 cells, despite the fact that TIM levels are nevertheless maintained.

The molecular significance of losing TIM in response to light is based on observations indicating that PER and TIM influence each other's appearance in the nucleus. PER and TIM display a reciprocal dependence in order to efficiently accumulate in the nucleus (Vosshall and Young, 1995 Hunter-Ensor *et al*., 1996;; Ashmore *et al*., 2003). The effect of light on TIM levels establishes a link between an environmental Zeitgeber to an alteration in a critical clock component, and that linkage is manifest as an

altered behavioral rhythm. A light pulse experienced early at night delays the onset of the subsequent cycle, whereas light experienced late in the night advances phase (Pittendrigh, 1974). According to the model, phase advances result from a premature decrement in TIM levels, which can only be reversed with sufficient time to allow additional *tim* translation. Likewise, light-triggered destruction of TIM late at night is predicted to hasten the rate at which PER levels decline and so hasten the onset of the next cycle.

A second model to explain the effects of light on the circadian machinery highlights negative regulation by light on the activity of a (presumed) PER : TIM heterodimer in repressing CLK : CYC transcriptional activity (Ceriani *et al*., 1999). This hypothesis is based on studies of the relevant clock molecules following their functional expression in S2 tissue culture cells. The authors found that the proposed photopigment CRY (see below) is associated with PER : TIM-containing complexes, and that it blocked PER : TIM derepression of CLK : CYC activity. This CRY effect was light-dependent and, importantly, it was efficient in the absence of TIM degradation, which is a feature not seen in S2 cells. Thus, in this cell line, photic regulation of the negative feedback loop (and by inference, the resetting of circadian phase) occurs despite the lack of TIM degradation. Lin *et al*. (2001) argue that such a degradation-independent phenomenon, because it may be specific to S2 cells, may not reflect the mechanistic details that occur within endogenous pacemaker cells (like the vLNs). Further, they contend that a substantial resetting only occurs by a process that changes a clock component (e.g., TIM levels) with a duration that outlasts the stimulus.

4.11.3.2.2. TIM's light sensitivity reflects activation by CRY Having identified a light-sensitive component of the negative feedback loop, efforts to dissect the initial events in circadian photoreception turned to the definition of the photopigments. Experimental evidence suggests that TIM is not directly light-sensitive, and that light-triggered degradation reflects an indirect action via the molecule CRY. Cryptochromes absorb light and transmit electromagnetic signals, using pterin and flavin adenine dinucleotide (FAD) as chromophore/cofactors (reviews: Sancar, 2000; Van Gelder, 2002). Cryptochromes are evolutionarily conserved and structurally related to the DNA repair enzyme photolyases – light-dependent enzymes that utilize flavin cofactors to repair DNA. In flies, CRY were studied by genomic methods: details of *cry* expression and its activities were reported by several groups (Emery *et al*., 1998; Egan *et al*., 1999; Ishikawa *et al*., 1999). Additionally, the *cry* locus was revealed by mutational analysis of loci that affected a *per-luciferase* daily rhythm (Stanewsky *et al*., 1998).

Quoting here from Sancar's (2000) review of the field, an explanation of the origin of the curious name is as follows: "The pigment system(s)… has been nicknamed 'cryptochrome' because of [its] importance in cryptogamic plants and its cryptic nature." Cryptochromes have 20–25% sequence identity with microbial photolyases and 40–60% sequence identity with the (6–4) photolyase. The sequence similarities between CRYs from various sources is remarkably high: human, *Drosophila*, and *Arabidopsis* CRYs share about 60% sequence identity (Todo, 1999). However, CRYs from different sources have nonhomologous C-terminal extensions ranging from very short ones in *Drosophila* to as many as 240 amino acids in plants. The two human CRYs are 73% identical to each other, but exhibit no sequence similarity within the C-terminal 75 amino acids. It is thought that this domain may bind to regulatory molecules (Cashmore *et al*., 1999; Todo, 1999).

Because declines in TIM levels are largely shaped by light, the absence of a *tim-luc* rhythm in cry^b suggested that CRY may normally contribute to TIM's photosensitivity. In addition, *per-luc* and *tim-luc* cycling was restored when cry^b mutants were exposed to rhythmic temperature cycles (Stanewsky *et al*., 1998). That result suggested that the basic clock mechanism is intact in such mutants, and that the behavioral defects derive from an inability to synchronize pacemaker activity to light cycles. Ceriani *et al*. (1999) reported that CRY and TIM interact in yeast in a light-dependent manner. Further, they showed that CRY is found in molecular complexes within S2 tissue culture cells that also contain TIM, and that CRY prevents TIM : PER transcriptional repression in S2 cells in a light-dependent manner. When expressed in S2 cells, CRY is degraded in response to light by the proteasome through a mechanism that requires electron transport (Lin *et al*., 2001). Various CRY mutant proteins (including CRY-B) were not so degraded. Does light promote TIM degradation by promoting or attenuating its interactions with CRY? TIM degradation is blocked by coexpression with mutant CRY (CRY-B), and TIM ubiquitination precedes CRY degradation. Thus, it appears more likely that in response to light, CRY transmits a signal that leads to TIM ubiquitination and degradation (Lin *et al*., 2001).

4.11.3.2.3. A second model of CRY activation and its interactions CRY may also interact directly with PER. Rosato *et al.* (2001) found genetic and molecular levels of interaction between CRY and PER. First they found that per^S; cry^b double mutant flies are able to detect light, but are deficient in the transmission of light information to the clock mechanism in a temperature-dependent manner. In contrast to the findings of Ceriani *et al.* (1999), they also showed that CRY and a truncated (TIM-interacting) PER interact in yeast cells, in a light-dependent manner. Furthermore, CRY can be coimmunoprecipitated along with full-length PER from S2 cells. Because CRY can interact with PER in the cytoplasm of S2 cells kept in the dark, it is possible that CRY can adopt "active forms" without the benefit of light. Therefore light activation, according to this model, involves derepression of an additional CRY-binding factor that prevents CRY from interactions with TIM or PER in the nucleus. There is evidence to suggest that CRY is not its own repressor, and that repression is exerted by other effectors acting at the CRY C-terminus (Rosato *et al.*, 2001).

4.11.3.2.4. The redox-based mechanisms of CRY activation Flavoenzymes like CRY catalyze redox processes because flavin itself is a reducing agent and can exist in different redox states (Sancar, 2000). Lin *et al.* (2001) tested the hypothesis that CRY has similar properties and that these may be important for circadian photoreception. CRY signaling in plants requires redox activity and is mediated, at least in part, by the flavin moiety bound to CRY. This is based on the finding that diphenylene iodonium (DPI), which inhibits the transport of electrons from reduced flavin (i.e., a reagent that blocks the transfer of electrons from reduced flavin prevents CRY degradation by light), was effective in blocking CRY-mediated photic signaling in *Arabidopsis* (Long and Jenkins, 1998). DPI also attenuated light-dependent CRY degradation in S2 cells, suggesting that an intramolecular conversion is required for this light response. Light-induced TIM ubiquitination precedes CRY degradation and is increased when electron transport is blocked. Thus, Lin *et al.* (2001) proposed that inhibition of electron transport may "lock" CRY in an active state by preventing signaling required either to degrade CRY or to convert it to an inactive form.

Further support for a role of the redox model of CRY action came from studies by Froy *et al.* (2002). Based on the crystal structures of photolyases from *E. coli* and *A. nidulans* (Park *et al.*, 1995; Tamada *et al.*, 1997), they targeted for mutagenesis conserved CRY residues predicted to participate in flavin binding. They then tested CRYs ability to block PER : TIM repression of CLK : CYC activated transcription in S2 cells. Alteration of three of the four flavin-binding residues abrogated light responsiveness in CRY. Additionally, two Trp residues, which are predicted to participate in intramolecular reduction, were also implicated in CRY function by similar criteria. These data support a multistep redox model for CRYs light sensitivity. In contrast, alterations of these same residues were without effect on the transcriptional repressive activity displayed by mouse CRY1, which to date has not been demonstrated to show photic sensitivity in its role in the SCN (Froy *et al.*, 2002).

4.11.3.3. What is CRY Doing?

4.11.3.3.1. CRY is a circadian photoreceptor In wild-type flies, *cry* RNA cycles in phase with *Clk* in LD and weakly in DD; CRY protein levels also cycle but only in LD (Emery *et al.*, 1998). From analysis in various mutant backgrounds, it was learned that *cry* transcription is under clock control and is also light-regulated at the translational or posttranslational levels. Additional copies of *cry*, in an otherwise wild-type background, enhances photosensitivity to phase shifting by short pulses of light at night (Emery *et al.*, 1998; Ishikawa *et al.*, 1999). In particular, the delay portion of the PRC is more affected than is the advance portion.

The cry^b mutant is a strong hypomorph, with a mutation that maps to a conserved flavin-binding site; CRY protein is absent or at very low levels in the cry^b mutant (Stanewsky *et al.*, 1998). In these CRY-depleted *Drosophila*, the entrainment by short light pulses is impaired (Stanewsky *et al.*, 1998). However, the clock still entrains to light–dark cycles, probably due to light input from the visual system. In addition, it can follow ~8-h shifts in LD phase (jet lag experiments) with only slight impairment. Significantly, under constant light conditions these animals are still rhythmic, unlike wild-type flies, which rapidly become arrhythmic (Emery *et al.*, 2000a). The site of *cry*s actions maps anatomically to the brain and not the eye: *cry* behavioral phenotypes are largely rescued by expressing *cry* under the control of a *tim*-GAL4 promoter, while a photoreceptor-specific GAL4 does not confer such a rescue (Emery *et al.*, 2000b). Significantly, restricting *cry* expression to just PDF neurons (with *pdf*-GAL4, see Section 4.11.4.2) partially rescues the LL phenotype (Emery *et al.*, 2000b) suggesting that CRY photosensitivity is principally required in just those few neurons for normal entrainment.

4.11.3.3.2. For CRY action(s), location may be everything A full description of CRY expression is still lacking, as no antibody-based localization has yet been reported. By *in situ* hybridization *cry* appears to be expressed in the vLNs and perhaps other pacemakers (Egan *et al*., 1999). *cry*-GAL4 promoter patterns indicate the *cry* gene expression is normally found within pacemaker neurons (Emery *et al*., 2000b), as well as in numerous peripheral tissues including the gut, antennae, and Malpighian tubules (Ivanchenko *et al*., 2001; Levine *et al*., 2002a), and in other neurons not traditionally considered to be pacemakers (Zhao *et al*., 2003). In spite of the fact that PER and TIM cycling is heavily damped in head extracts, animals are paradoxically rhythmic in DD. A likely explanation lies in the fact that PER and TIM in fact still cycle in vLNs and dLNs in both LD and DD (Stanewsky *et al*., 1998): this persistent rhythmicity in the pacemaker neurons that are critical for behavior probably explains the persistent behavioral rhythmicity.

The hypothesis that CRY is a circadian photoreceptor in *Drosophila* contrasts with the current conception of the two mammalian CRYs' roles as components of the core negative feedback loop (e.g., Froy *et al*., 2002). It is worth noting, therefore, that the photoreceptor function for CRY is a phenotype ascribed to its role in the central vLN pacemaker neurons. It remains possible that CRY plays different roles in different pacemaking tissues of the fly and several observations support that possibility. First, the differences in PER and TIM rhythmicity of *cry* mutant LNs versus photoreceptors proves that different pacemaker cells can display significant differences in the details of their molecular oscillations. Second, autonomous clocks present in peripheral tissues can be studied with isolated body parts (e.g., legs) and by measuring rhythms in *per-luc* or *tim-luc* (Plautz *et al*., 1997). The phenotype of the circadian oscillator present within those peripheral tissues often appears more severely affected by the *cry* mutation than the central oscillator. Those observations have led to the generalization that CRY may be a remote photosensor for the central clock, but may have a role closer to the actual clockworks in other oscillators.

The strongest evidence to suggest that CRY can be a clockwork component derives from observations in antennae, which display a circadian rhythm of sensitivity (Krishnan *et al*., 1999), and a tissue-autonomous *per-luc* rhythm (Plautz *et al*., 1997). In *cry* mutants, the amplitude of antennal *per-luc* cycles was severely diminished (Levine *et al*., 2002a; Krishnan *et al*., 2001). Surveying other, isolated peripheral tissues with the *per-luc* reporter, Levine *et al*. (2000a) reported that the cry^b mutation significantly decreased the number of rhythmic specimens in most cases, and also decreased the amplitude and altered the phases of rhythmicity. Likewise, Ivanchenko *et al*. (2001) examined a *tim-luc* reporter and found that the cycling of TIM was abolished in larval Malpighian tubules in DD, but it persisted in larval LNs of cry^b flies. From these and other observations, they concluded that CRY is involved in TIM-mediated entrainment of both central and peripheral clocks, but that it is also an indispensable component of peripheral oscillators. Dissecting these different roles for CRY, depending on cellular location, represents a substantial challenge for future research.

4.11.3.4. Cells Mediating Circadian Photoreception

4.11.3.4.1. Neuroanatomy In *Drosophila*, the photosensitive cells that feed light information to pacemaking centers are widespread. The presence or function of the compound eyes is not necessary to synchronize the fly's rhythms by light input (e.g., Wheeler *et al*., 1993; Yang *et al*., 1998), and the insect circadian system is sensitive to phototransduction from extraocular sources (Truman, 1976). However, the compound eyes are involved, as flies lacking them or lacking functional components (e.g., *norpA* – Pearn *et al*., 1996) display lowered sensitivity to entrainment by light (Helfrich-Förster, 1997; Stanewsky *et al*., 1998; Yang *et al*., 1998). In the retina, some or all of the eight retinal photoreceptors may signal via synapses in the optic lobes to the large vLNs, which form a tangential projection across the distal medulla (Helfrich-Förster, 1995; Chen *et al*., 1999; Taghert *et al*., 2000).

Candidate extraocular photoreceptors have now been identified and functionally implicated (Helfrich-Förster *et al*., 2002). In the periphery, the candidates include two extraretinal groups, the ocelli and the H-B Eyelet (Hofbauer and Buchner, 1989). The Eyelet resides at the posterior margin of the compound eye in a subretinal position and forms a specialized pigmented organ with cells that have numerous microvilli arranged into coherent rhabdomeres (Yasuyama and Meinertzhagen, 1999). The candidate photoreceptive cells in the brain include the LNs that likely express CRY (Emery *et al*., 2000b), and the PER-expressing dorsal neurons that likely express CRY and perhaps also a different photopigment (Malpel *et al*., 2002; Veleri *et al*., 2003).

Both Bolwig's neurons and Eyelet neurons express Rh5- and Rh6-opsin-like immunoreactivity

(Yasuyama and Meinertzhagen, 1999; Malpel *et al*., 2002), and they express a phospholipase (Malpel *et al*., 2002) encoded by *norpA* (Pearn *et al*., 1995). The Bolwig's neurons do not stain for antihistamine (Pollack and Hofbauer, 1991), in contrast to Eyelet cells (Yasuyama and Meinertzhagen, 1999), which also stain for anticholine acetyltransferase (Yasuyama and Meinertzhagen, 1999). The hypothesis that Eyelet neurons are cholinergic is further supported by pharmacological experiments on isolated LNs (Wegener *et al*., 2004). A circadian photoreceptor function for Eyelet was indicated in flies lacking compound eyes but retaining Eyelet (*sine oculis, so* – Helfrich-Förster *et al*., 2002); the action spectrum of light entraining the residual rhythmicity in *so* flies displayed a peak at 480 nm, different from the peak sensitivity of CRY. Although Eyelet photoreceptors alone were able to cause entrainment in the absence of compound eyes and of CRY, such entrainment was not robust under the circumstances tested.

4.11.3.4.2. Development of Eyelet The developmental origins of the Eyelet were subject to speculation until recently. It appeared likely that it was derived from Bolwig's organ, as suggested by the common positions of their axonal terminations in the accessory medulla (Gibbs and Truman, 1998; Yasuyama and Meinertzhagen, 1999). In larvae, Bolwig's neuron axons are in close proximity to the dendrites of PDF-expressing LNs (Kaneko *et al*., 1997; Malpel *et al*., 2002) and are likely to mediate circadian entrainment (Kaneko *et al*., 2000b). Helfrich-Förster *et al*. (2002) established a direct link to Eyelet neurons following expression of a *Kruppel-lacZ* reporter that is specific to the Bolwig's organ neurons. The metamorphic reorganization of Eyelet is not complete until adult eclosion – this suggests that Eyelet functions to help entrain locomotor rhythms, but not eclosion rhythms (see **Chapter 3.1**). Thus, the adult circadian system is dependent on the photoreceptive structures of the larva with the added contributions of photoreceptors derived from metamorphic tissues.

The hypothesis that multiple photoreceptors combine to help entrain the *Drosophila* circadian clock was indicated by analyses of various single and double mutant stocks (e.g., *cry*; *norpA*). The double mutant stock lacks function of both CRY and the compound eyes and ocelli. Such flies display less photic sensitivity than did either single mutant stock. Nevertheless, these double mutants maintained some sensitivity to entraining photic inputs (Stanewsky *et al*., 1998). Only when all known ocular and extraocular photoreceptors were genetically eliminated (by a *cry; glass* double mutation) were flies unable to respond at all to LD cycles (Helfrich-Förster *et al*., 2001). The *glass* gene is required for normal development of the compound eyes, ocelli, H-B Eyelet, and DN1 neurons. The recent research mentioned above has established the range of photoreceptors and photopigments involved in circadian photoreception in flies and begun to indicate their individual roles. Of course, new information always produces a generation of new questions (e.g., what are the qualitative and quantitative contributions of the various photoreceptors?) and much remains to be resolved to give a clear description of the parallel functions. This point was suggested by Helfrich-Förster *et al*. (2001): "Apparently, cryptochrome and rhodopsins of Eyelet and the compound eyes collude in a complex manner to entrain the adult fruit fly's circadian activity to the 24 h day."

4.11.3.5. Transmitter Sensitivity of Pacemaker Neurons

Two recent studies have indicated the identities of small molecule transmitters that are likely to provide critical inputs to the vLN pacemaker neurons. Ionotropic GluR-IB receptor subunits are preferentially expressed in vLNs of the adult head (Volkner *et al*., 2000), suggesting a role for glutamate signaling (see **Chapter 5.2**) in mediating entrainment mechanisms. In addition, isolated vLNs display pharmacological properties that are consistent with the expression of nicotinic Ach receptors that signal via calcium (Wegener *et al*., 2004). The latter results are congruent with the demonstration of cholinergic signaling (see **Chapter 6.2**) by Eyelet cells (Yasuyama and Meinertzhagen, 1999). In the cockroach, Petri *et al*. (2002) have shown that injections of GABA and of the neuropeptide Masallatostatin both resulted in stable phase-dependent resetting of the circadian locomotor activity. The PRCs for the transmitter effects matched those of the light PRC, suggesting that these substances may contribute to neural signaling underlying photic entrainment in this insect.

4.11.3.6. Other Entrainment Mechanisms: Sociality

Environmental inputs in addition to light can synchronize internal clocks: the most prominent is temperature (review: Rensing and Ruoff, 2002). A recent report indicated that *Drosophila* clocks are also set by social interactions (Levine *et al*., 2002b). The authors determined that the phase coherence of daily activity (a measure of synchronicity, where phase peak equaled the peak of daily

locomotion) was stronger among flies that had been jointly housed prior to testing, versus those housed individually. In effect, social experience had promoted the resetting of the activity clock. Likewise, the phase coherence of wild-type flies was diminished by the interactions (co-housing) with arrhythmic per^o flies (at a 4 : 1 and 2 : 1 ratio). Interestingly, the effect was lost when the populations were at equal strength. Levine *et al.* (2002b) favor the hypothesis that the inferred social interactions predominantly reflect chemosensory sensations. Genetic analyses and "conditioned air" tests were both consistent with a mechanism whereby short-lived volatile substances signal to con-specifics (see **Chapter 3.13**) and represent social experiences that help to reset the biological clock.

4.11.4. Output Studies

At some level, clock information morphs into output signals by which rhythms are synchronized and/or evoked in downstream centers. The substance of circadian output can take many forms (reviews: Jackson *et al.*, 2001; Taghert, 2001; Park, 2002), among which three levels can be described. At the most proximal level, circadian output regulates rhythmic processes in the pacemaker cell itself, and is represented by cycling gene expression within that cell. Estimates of the percentage of the genome that is subject to circadian control have ranged from one to many per cent (review: by Dunlap, 1999). Leaving the bounds of single cells, circadian output also describes cellular interactions that take place between pacemakers. Because pacemakers influence each other's rhythms, this circadian output transforms into substantial circadian input, affecting the phase and/or amplitude of subsequent cycles. Finally, circadian output is translated to signals that affect cells and tissues which are not themselves capable of generating pacemaker activities. The latter rely on timing information from the biological clock for their normal rhythmic physiology.

The literature in this field is reviewed here, with special emphasis on *Drosophila*, using the three-part categorization. To begin with, the results of several microarray studies of *Drosophila* circadian rhythmicity are summarized. Next, cellular descriptions of circadian pacemakers, their hierarchies, and the nature of signals used to coordinate their interactions are disscussed. Finally, several emerging areas of research activity (e.g., eclosion, drug sensitivity, and sleep), in which new information has become available to understand how circadian output is manifested as behavioral rhythms, are outlined.

4.11.4.1. Microarray Studies

4.11.4.1.1. Whole genome analysis of cycling gene expression Because many of the known circadian genes display regular fluctuations in their transcript levels, a method for discovering new components of the molecular clock is to search for rhythmic transcripts – transcripts which, in the absence of environmental cues, continue to display identical, daily waveform patterns. Such screens have been performed previously, including the examination of anonymous cDNAs by time-series Northern blots (Van Gelder *et al.*, 1995), enhancer trap screening of rhythmic signals with the luciferase reporter (Stempfl *et al.*, 2002), and a subtracted cDNA library strategy to identify time-dependent, differentially expressed transcripts (Rouyer *et al.*, 1997). With the advent of microarrays and the ability to perform, in effect, thousands of Northern blots in parallel, the circadian field quickly adopted microarray technology to continue this search.

Within the period of a year, five published studies described the extent of transcript rhythms in *Drosophila* (MacDonald *et al.*, 2001; Claridge-Chang *et al.*, 2001; Ceriani *et al.*, 2002; Y. Lin *et al.*, 2002; Ueda *et al.*, 2002). These studies have generated a treasure-trove of data, which is amenable to future meta-analysis by virtue of the commonalities in their experimental designs. All groups extracted transcripts specifically from fly heads, used the Affymetrix oligonucleotide microarray platform to monitor the activity of ~13 500 genes, and performed temporal sampling at a resolution of 4-h intervals. Most of the studies performed light-entrained (LD) microarray experiments and all studies also monitored transcript levels under constant conditions (DD). Some differences such as the use of fly strains with differing genetic backgrounds were unlikely to result in material differences with respect to the identification of novel clock candidates, as the underlying molecular components should be identical between strains.

4.11.4.1.2. A current lack of consensus on the number of cycling genes Even with so many common elements spanning these studies, the results of the analysis were in sharp disagreement. One study identified 378 rhythmic genes under DD conditions, yet of these 378, only 28 were identified by at least three studies, and only five were agreed upon by all five studies. These numbers illustrate the degree to which false positives may populate the various lists of rhythmic genes in *Drosophila*. Two salient issues distinguished these five reports and may have resulted in such disparities. One was the extent of

time in which flies were kept under constant conditions before the collection of DD data. Light-driven effects in flies may take several days to extinguish, as evidenced, for example, by the residual behavioral rhythms seen in the first 2 DD days of the otherwise arrhythmic $disco^{1}$ and pdf^{01} mutants (Hardin *et al*., 1990; Helfrich-Förster, 1998; Renn *et al*., 1999). Most studies collected data on the first day of constant darkness, whereas Ceriani *et al*. collected data from the first two consecutive DD days and Y. Lin *et al*. collected replicate cycles on the third day of DD. In contrast to the >115 genes identified by the other studies, Ceriani *et al*. found that 63 genes remained rhythmic 2 days into DD, and Y. Lin *et al*. found 22 genes that cycled on the third day of DD. The correlation between more cycles of free-run conditions and fewer identified rhythmic genes is notable. This time-dependent decrement in rhythmicity may reflect a damping effect on molecular rhythms when pacemakers become de-synchronized, as observed in peripheral clocks (Plautz *et al*., 1997). However, that explanation is not valid in this case as Y. Lin *et al*. found that all of the canonical clock genes – *per, tim, vri*, and *Clk* – continued to cycle robustly 3 days into DD.

The other significant difference between the studies was the treatment of statistics and methods of analysis. To identify rhythmic waveforms, the methods ranged from spectral analysis (Claridge-Chang *et al*., 2001) to idealized cosine curve fitting (McDonald *et al*., 2001; Ceriani *et al*., 2002; Ueda *et al*., 2002) to direct measurement of intercycle reproducibility of waveforms with autocorrelation analysis (Y. Lin *et al*., 2002). A concise treatment on the strengths and weaknesses of each strategy was recently detailed by Levine *et al*. (2002c).

It is reasonable to assume that all five studies have correctly identified at least some unambiguous, robustly cycling transcripts. Therefore, until a unified, meta-analysis is performed on the pooled data, a measure of confidence may be afforded by relying upon genes commonly identified across multiple studies. To balance between the degree of agreement amongst the various studies, and the inclusion of the canonical rhythmic genes *per, tim, takeout, vri*, and *Clk*, **Table 1** lists the 28 genes that were identified by at least three groups to display rhythms in DD. No functional patterns among the known or predicted gene functions seem yet to emerge from the list.

Table 1 The list of *Drosophila* genes that were commonly identified by at least three microarray studies as circadianly cycling

Gene	*Location*	*Known/predicted function*
Clk	66A12-66A12	RNA polymerase II transcription factor
Cyp18a1	17D1-17D1	Cytochrome P450, CYP18A1; cytochrome P45; EC:1.14.14.1
Cyp4d21	28A6-28B1	Cytochrome P450, CYP4D21; cytochrome P45
Cyp6a21	51D2-51D2	Cytochrome P450, CYP6A21; cytochrome P45
Pdh	72F1-72F1	Enzyme
Slob	28B1-28B3	Signal transduction
Ugt35b	86D5-86D5	UDP-glucuronosyltransferase; UDP-glucuronosyltransferase; EC:2.4.1.17
per	3B4-3B4	Transcription factor
tim	23F3-23F5	RNA polymerase II transcription factor
vri	25D4-25D4	RNA polymerase II transcription factor
BcDNA:GH02901	13A5-13A5	Long-chain-fatty-acid-CoA-ligase
CG1441	46C5-46C5	
CG4784	72F1-72F1	Cuticle protein; structural protein
CG4919	94C3-94C3	
CG5156	21F2-21F2	
CG5798	93C1-93C1	Ubiquitin thiolesterase; endopeptidase
CG5945	34A11-34A11	
CG9645	88B3-88B3	Endopeptidase
CG9649	88B3-88B3	Endopeptidase
CG10513	96C7-96C7	
CG10553	96C8-96C8	
CG11407	92B3-92B3	Luciferase-like; enzyme
CG11796	77C1-77C1	
CG11853	96C4-96C4	
CG11891	96C6-96C7	
CG14275	29B1-29B1	
CG15093	55F1-55F1	3-Hydroxyisobutyrate dehydrogenase-like; enzyme
CG17386	51A1-51A1	RNA binding

4.11.4.1.3. The clock also controls the basal level of expression for thousands of genes To control against the false-positive identification of rhythms, all groups performed time-series experiments with arrhythmic flies bearing mutations in clock components such as per^{01} and Clk^{Jrk}. A surprising additional observation from these data was that such mutations produced broad effects on noncycling, basal expression of genes across the *Drosophila* genome. MacDonald *et al*., Claridge-Change *et al*., Ueda *et al*., and Ceriani *et al*. observed significant differences in gene expression between Clk^{Jrk} mutants and control flies. These changes are likely to represent the absence of normal clock-controlled processes, but may also include gain-of-function consequences that reflect the dominant-negative character of the Clk^{Jrk} allele (Allada *et al*., 2003). Claridge-Change *et al*. and Y. Lin *et al*. found that normal expression levels of large swaths of the fly genome depended on *per* and *tim* functions. Furthermore, Y. Lin *et al*. observed that light influenced the expression levels of hundreds of genes when comparing expression levels of control flies under LD versus DD conditions. These light effects were largely dependent upon *per* function, as per^{01} mutants did not display comparable changes between LD and DD conditions.

Although the power of microarrays to generate candidates for future pursuit cannot be disputed, some caveats should be noted. Several limitations have been enumerated in the reports, such as the heavy dependence of probe sets upon the quality of genome annotation, and the relative insensitivity of microarrays to low-abundance and tissue-specific transcripts. Related to computational annotation is the issue of alternative splicing, a phenomenon thought to contribute heavily to the complexity of multicellular organisms (Harrison *et al*., 2002). Current generations of microarrays do not yet distinguish between splice variant forms of genes, thus limiting the accuracy of the circadian landscape as seen through a microarray filter. This point was exemplified recently, with the demonstration by Cyran *et al*. (2003) that only one of six splice variants of the new clock component *pdp1* displays transcript rhythms.

Is the limited manifestation of transcript rhythms as suggested by Y. Lin *et al*. and as indicated by the small group of genes coincident between the studies, a meaningful result? One possibility is that despite the existence of a limited number of rhythmic transcripts, a greater population of rhythmic proteins may exist. For example, rhythms in the Lark protein, a protein required for eclosion rhythms, are generated from a nonrhythmic transcript (McNeil *et al*., 1998). Another possibility is that beyond the immediate proximity of the core molecular clock, rhythms exist primarily at the physiological level. Perhaps this is demonstrated by the neurotransmitter PDF, which is neither transcriptionally nor translationally rhythmic, but is rhythmic in its release from axon terminals (J.H. Park *et al*., 2000; and see below). Likewise, several antecedent studies suggest that even high-amplitude rhythms of transcription in *per* and *tim* are not absolutely required for the production of rhythmic behavior (e.g., Frisch *et al*., 1994; Cheng and Hardin, 1998; Yang and Sehgal, 2001). In sum, these observations all point to the primacy of posttranscriptional events in defining circadian molecular oscillations. Thus, the "limited manifestation of transcript rhythms" may be a reasonable initial description of rhythmicity in the *Drosophila* transcriptome. Perhaps with future high-throughput techniques such as quantitative mass spectrometry, and the accumulation of more laborious, time-series immunostain-tracking of individual proteins, a broader sense of the *Drosophila* circadian clock will be achieved.

4.11.4.2. Pacemaker Neurons in the Fly Brain: Anatomy and Roles

4.11.4.2.1. Pacemaker cell types PER expression is found in numerous cells in numerous tissues (Liu *et al*., 1988, 1992; Siwicki *et al*., 1988; Zerr *et al*., 1990; Ewer *et al*., 1992; Kaneko *et al*., 1997; Rachidi *et al*., 1997) and the scale of the pattern depends on the method of visualization. PER immunostaining and *per* RNA *in situ* tend to feature small clusters of neurons in the lateral and dorsal brain of the adult, photoreceptors, as well as numerous glia cells in the optic lobes (e.g., Zerr *et al*., 1990, Ewer *et al*., 1992, Helfrich-Förster, 1995; Kaneko *et al*., 1997). In contrast, *per*- and *tim*-promoter fusions also display these same groups, plus many others not otherwise associated with circadian pacemaking (e.g., Kaneko and Hall, 2000). Price *et al*. (1998) observed that PER immunostaining was greatly increased in level and distribution (beyond the normal) in a strongly hypomorphic allele of *dbt*. That result is consistent with the hypothesis that PER is expressed normally, but highly unstable, in a large population of neurons in the larval brain.

The canonical PER neuronal groups comprise the lateral cells (LNs) which include the small and large vLNs (s-vLNs and l-vLNs), and dorsal cells (dLNs). There are four to five s-vLNs, of which four express the neuropeptide PDH/PDF (Helfrich-Förster, 1995; Kaneko *et al*., 1997; and see below). There are four to five l-vLNs and approximately seven dLNs.

In the dorsal brain, there are three additional groups: the DN1 cell group contains ~10 neurons; the DN2 group is composed of a pair of cells; and the DN3 group contains ~30 neurons. The axonal projections of all these neurons appear to target the neuropils of the dorsal protocerebrum, where they intermingle and may interact (Kaneko and Hall, 2000; Helfrich-Förster, 2003; Veleri *et al*., 2003). In addition, a pair of DN1s and some DN3s project to the accessory medulla and appear to provide reciprocal innervation to the s-vLNs. Other DN1s project caudally towards the suboesophageal region and perhaps to the ventral nerve cord.

4.11.4.2.2. Which PER neurons are necessary for rhythmic behavior? Given the several different pacemaker cell groups, which are necessary and which are sufficient to act as pacemakers to drive rhythmic behavior? Leak and Moore (2000) argue that the different rhythmic systems in mammals are controlled by separate sources of SCN efferents ("core" versus "shell"). Similar questions can now be applied to the insect brain, where smaller numbers of neurons may facilitate the cellular analysis. Several different genetic designs have tried to define the relative participation of the different pacemaker cells groups in *Drosophila*. Of these, the study of *disconnected* mutant flies was among the first and most substantive.

The *disconnected* (*disco*) gene in *Drosophila* encodes a widely expressed transcription factor that is required to establish normal sensory connections in the central nervous system (CNS) (Steller *et al*., 1987; Glossop and Sheperd, 1998). Approximately 5–10% of mutant animals display some retinal-brain connections. While Dushay *et al*. (1989) found that *disco* flies are arrhythmic for circadian behaviors, Hardin *et al*. (1992b) showed that PER levels fluctuate normally in homogenates of *disco* heads. Those findings suggested that *disco* acts on the output level because while the mutation eliminated behavioral rhythms, it left the "clock mechanism" intact. An alternative emphasis follows from considering abnormalities of PER expression in *disco* flies (Helfrich-Förster, 1998). PER is found in photoreceptors, DN pacemakers, and putative glia in *disco*, but is largely absent from LN pacemaker neurons. Blanchardon *et al*. (2001) subsequently reported that many LNs do in fact survive this mutant background, as indicated by expression of a P{GAL4} reporter line. Together these observations suggest that *disco* animals retain many functional clock centers (e.g., the retina, the DNs), but display poor behavioral rhythmicity due to the lack of functional clock centers (e.g., LNs) that normally organize daily locomotion. These animals were very useful to learn specific facts concerning output mechanisms, especially by restricting analysis to mutant animals displaying quasi-normal behavior. Helfrich-Förster (1998) monitored the behavior of hundreds of mutant animals: the few that displayed some rhythmic behavior were the only ones to also retain at least one s-vLN. Thus a strong correlation was made between the presence of a s-vLN and circadian locomotor behavior.

Mosaic analysis of *per* expression has also been influential in deciphering the hierarchies of pacemaker cells (Ewer *et al*., 1992). Frisch *et al*. (1994) showed that, in complimentary fashion to the *disco* phenotype, PER expression in just the LNs (both ventral and dorsal) was sufficient to display some rhythmic behavior. Evidence that the DNs also participate in rhythmic output comes from two studies. When PER was returned only to LNs and not to DNs of *per* mutant flies, the restored rhythms were abnormal in period and strength (Frisch *et al*., 1994): by inference, DNs may contribute the difference. Similarly, PER expression in accessory medulla neurons under the *glass* promoter in *per* mutant flies produced rhythmic behavior (Vosshall and Young, 1995). More recently, Veleri *et al*. (2003) described a *per-luc* transgene fusion that restored rhythmicity to *per* mutant flies under LD cycles, although not under constant conditions. In such flies, the only site of molecular oscillations was in the DN3 populations: strong evidence to suggest these DN neurons are sufficient to provide pacemaker activity for behavior, at least in LD.

Peng *et al*. (2003) and Allada *et al*. (2003) addressed the same question by asking whether limited expression of transgenic *cyc* or *Clk* could rescue either mutant phenotype. Peng *et al*. (2003) reported that in *cycle* mutant flies now expressing UAS-*cycle* under control of *pdf*-GAL4, PDF neurons alone now displayed proper rhythms of *tim* RNA, suggesting that UAS-*cyc* had indeed rescued molecular rhythmic in mosaic fashion. However, the behavior was not rescued, suggesting that other activity in other pacemaker neurons is required. It should be remembered that *cyc* mutant flies display aberrant PDF neuronal morphology (J.H. Park *et al*., 2000), and that aberrant PDF neuronal morphology is correlated with arrhythmic behavior (Helfrich-Förster, 1998). Allada *et al*. (2003) used UAS-*Clk* to rescue two different mutant alleles of *Clk*. They found that a *cry*-GAL4 line that is expressed in both vLNs and dLNs (Emery *et al*., 2000b; but see also Zhao *et al*., 2003) partially rescued the behavioral rhythmicity of *Clk* mutants. *pdf*-GAL4 : UAS-*Clk* did not provide such rescue. These results, while not yet

definitive, are consistent with the hypothesis that pacemaker actions of the ventral and dorsal LNs are both necessary and sufficient for *Drosophila* to display behavioral rhythms.

Within the vLN group, five lines of evidence (all indirect) suggest that the s-vLNs may be especially important and the l-vLNs largely unimportant as circadian pacemakers. First, the daily fluctuation in PDF immunostaining of vLN terminals occurs only in the case of the small cell subset (J.H. Park *et al.*, 2000; see further discussion below). Second, the effects of *Clock* and *cyc* mutations on *pdf* RNA levels are seen only in small and not l-vLNs (J.H. Park *et al.*, 2000). Third, in a study of the effects of *disco* mutations on rhythmicity, Helfrich-Förster (1998) confirmed that most or all *per*-expressing neurons (and hence all *pdf* vLNs) were typically undetected in *disco* animals (cf. Blanchardon *et al.*, 2001). She reported a small minority of animals ($n = 4$) that retained rhythmicity: PDF-expressing s-vLN were visible (as few as one neuron) only in this minority. A l-vLN was also found in a single rhythmic individual, but its processes projected in a s-vLN fashion, to the dorsal protocerebrum. Fourth, Yang and Sehgal (2001) and Shafer *et al.* (2002) reported fluctuations in PER and TIM immunostaining levels under constant darkness in small, but not in large, vLNs. Finally, cry^b mutants fail to maintain rhythms of PER or TIM in DD, in any pacemakers but the s-vLNs (Stanewskey *et al.*, 1998). These diverse and consistent observations form a compelling hypothesis to indicate a special role for the s-vLNs. Direct confirmation of that hypothesis would be a useful step in narrowly defining the pacemaker cell hierarchies within the brain. To do so will require further and more precise manipulation of these neuronal populations.

4.11.4.2.3. PER expression in other insects Several groups have used *period* gene expression as a method of surveying the number and position of potential circadian pacemakers in other insects. In addition, several of these comparative studies have addressed the potential conservation of circadian pacemaker cell types across insect phylogeny by examining possible coexpression of PER with PDH/PDF. Frisch *et al.* (1994) used anti-PER antibodies (against a conserved region of the *Drosophila* PER) to stain the brain of a beetle and reported finding many immunoreactive neuronal groups, some of which had nuclear labeling. In addition to the neurons, they also found many labeled glia in the optic lobes, reminiscent of glial *per* expression in the *Drosophila* brain (cf. Ewer *et al.*, 1992). Comparisons of anti-PER and anti-PDH stains suggested there was overlap in stained populations, but that many PER-positive did not express PDF. Reppert *et al.* (1994) cloned a *per* ortholog from the silkmoth *A. pernyii*, and reported that antibodies to the moth PER protein labeled the cytoplasm of eight neurosecretory neurons in the protocerebrum (Sauman and Reppert, 1996). Surprisingly, the nuclei of these neurons were never so labeled, the cells were distinct from the PDF-immunoreactive neurons, and no other cells displayed immunostaining. Interestingly, this anti-moth PER antibody when used in *Drosophila* produced a pattern of PER-like immunoreactivity that was highly similar to that previously found with antibodies to the fly PER (Levine *et al.*, 1995). In addition, the moth protein was functional within *Drosophila*: it could rescue the *per* behavioral phenotype when expressed as a transgene in *per* mutant *Drosophila* (Levine *et al.*, 1995).

The theme of differences between the patterns of PER in *Drosophila* versus other insects has been repeated in more recent studies, specifically in the moth *Manduca* (Wise *et al.*, 2002) and the honeybee (Bloch *et al.*, 2003). In both studies, the endogenous PER proteins were used to raise specific antibodies. In *Manduca*, widespread *per* expression was found in numerous neurons and glia. Many expressing cells displayed both nuclear and cytoplasmic staining, although evidence for rhythmic expression was only found in glia. Analogous to the observations in *A. pernyii*, four neurosecretory cells in the *pars lateralis* of each brain hemisphere exhibited both nuclear and cytoplasmic staining with anti-PER antibodies. These cells were identified as Ia(1) neurosecretory cells that express neuropeptide corazonin immunoreactivity. The accessory medulla contained ~100 neurons expressing *per* RNA but no immunoreactivity. No correspondence of *per* expression to PDH/PDF expression was evident in any part of the brain.

Likewise, in the honeybee brain, PER immunosignals were prominently found in small sets of protocerebral neurosecretory cells, but not within PDH-immunoreactive neurons of the medulla. The latter appeared to have dendrites within the accessory medulla and to project to dorsal protocerebrum, but to lack clock (e.g., *period*) protein expression. Bloch *et al.* concluded that: "... although clock proteins are conserved across insect groups, there is no universal pattern of co-expression that allows ready identification of pacemaker neurons within the insect brain." Finally, using an anti-*Drosophila* PER antiserum and anti-PDH, Zavodska *et al.* (2003) examined several insect

orders and found consistent expression of both immunosignals in most representatives. However, they reported no correspondence between PER and PDH immunoreactivities. Examples of the lack of nuclear PER staining may reflect a noncircadian function in those cells, or perhaps offer instances when PER's negative feedback functions may not involve its direct participation in the nucleus (cf. So and Rosbash, 1997). Nevertheless, it remains a challenge for the field to compare the different PER-expressing neuronal populations in different insects, and to rationalize these in terms of neuronal circuits underlying rhythmic behavior.

4.11.4.3. PDF: A Circadian Transmitter

4.11.4.3.1. Expression Interest in the pigment dispersing factor (PDF) by circadian biologists stems directly from the demonstration of its expression within specific circadian pacemaker neurons of *Drosophila* (Helfrich-Förster, 1995). Its expression within insect tissues, especially in areas associated with circadian pacemakers, was first described several years earlier (Homberg *et al.*, 1991; Nässel *et al.*, 1991, 1993; Helfrich-Förster and Homberg, 1993) using antibodies to crustacean pigment dispersing hormone (PDH, discussed below in Section 4.11.4.3.2). PDH immunosignals are normally found in a limited set of brain and ventral nerve cord cell types. Fundamental elements of the cellular pattern are found throughout insect orders, with salient differences also noted (e.g., Helfrich-Förster *et al.*, 1998; Zavodska *et al.*, 2003). Based on its hormonal activities in crustacea (see **Chapter 3.16** and see below), there is good evidence to hypothesize that PDF in *Drosophila* is a secreted neuropeptide and hence a true output signal for certain clock neurons (review: Taghert and Veenstra, 2003). However, recent studies in the cricket indicate that PDH/PDF may also have a nuclear location in some cell types (Chuman *et al.*, 2002). These careful *in vivo* observations were also supported by studies of PDF transfected into mammalian cells, where the functionality of a putative nuclear localization signal on the PDF precursor was tested. This work underscores the importance of testing all assumptions rigorously and the need to address a range of plausible hypotheses.

In *Drosophila*, PDF expression is limited to the CNS. Within the CNS, there are two PDF cell types in the larva and three cell types in the adult. In larvae, the brain contains four to five LNs that display molecular pacemaker properties (Kaneko *et al.*, 1997). LNs are the likely precursors of the s-vLNs of adults (Helfrich-Förster, 1997). In addition, the larval ventral nerve cord contains a prominent set of four to six large neuroendocrine neurons in the terminal abdominal segments that appear to release PDF into the hemolymph (cf. Persson *et al.*, 2001). The abdominal cells do not display clock properties. In the adult, the vLN group is enlarged by the differentiation of the l-vLNs, which project axons tangentially across a distal layer of the medulla. The abdominal neuroendocrine PDF cell group is maintained. Finally, the third group in adults is a transiently occurring population: a pair of cells in the suboesophageal ganglion of the adult that do not express clock properties and disappear by the second day of adult life. Despite the fact that larvae produce no known circadian output, larval LNs display all the molecular hallmarks of functional pacemakers (e.g., Price *et al.*, 1998). Using anti-PDH (crustacean PDH), an additional two to three cells in the larval protocerebrum and three to five cells in the adult protocerebrum are found. These appear to represent crossreactivity with another, so far unidentified, substance as these cells do not stain with *pdf in situ* methods (J.H. Park *et al.*, 2000) and they retain immunostaining in *pdf* mutant animals (Renn *et al.*, 1999).

4.11.4.3.2. The PDH/PDF family of peptides PDH peptides were first studied in crustacea where they cause diurnal movements of pigment granules in retinal cells and their dispersion in epithelial chromatophores. In 1971, a factor that caused the dispersion of distal retinal pigment was purified from eyestalk extracts of the prawn *Pandalus borealis* (Fernlund, 1971). Upon sequencing, the factor was revealed to be an 18 amino acid peptide with an amidated C-terminus and a free N-terminus (Fernlund, 1976). Originally called light adapting distal retinal pigment hormone (DRPH), it was renamed pigment-dispersing hormone (thus, Panbo-α-PDH), because it also translocates the pigments in the chromatophores centrifugally (Kleinholz, 1975). About a decade later, a second PDH was chemically identified from the eyestalks of the fiddler crab *Uca pugilator*, a so-called β-PDH that differs from α-PDH in six positions (Rao *et al.*, 1985). To date, PDHs from 15 crustacean species are known (see **Chapter 3.16**).

Extracts from heads of insects were able to elicit a dispersion of pigments in the epidermis of eyestalkless fiddler crabs (Rao *et al.*, 1987). This bioassay then served to isolate the active factor from the grasshopper *Romalea microptera*, a modified β-PDH. Pigment-dispersing factors (PDFs) have since been identified from several insect groups and orthologs identified in the genomes of *D. melanogaster*

(Park and Hall, 1998), and in the mosquito *A. gambiea* (Riehle *et al.*, 2002). The distribution of PDF/PDH appears limited to arthropods to date, but there is one report of immunoreactivity in a mollusk (Elekes and Nässel, 1999).

The PDH/PDF family of peptides displays a large amount of structural conservation; including the length (18 amino acid residues), the N-terminal Asn and C-terminal amidated Ala residues, as well as conserved amino acids at several internal positions. Interestingly, *pdh* genes in crustaceans have undergone gene duplication and up to three forms, for example, two α-PDHs and one β-PDH, are found in *Pandalus jordani* (Ohira *et al.*, 2002). *Drosophila* contains only a single *pdf* gene and there is no evidence to suggest that PDFs in *Drosophila* affect pigment dispersion. In various animals including *Drosophila*, the PDF peptide is predicted to be synthesized following posttranslational processing of a ~100 amino acid preproPDF precursor. Nothing is known about the actual biosynthesis of PDF. The general organization of PDF precursors (Ohira *et al.*, 2002) features a signal peptide that is followed immediately by a precursor-associated peptide (PAP) of unknown function, a di- or tri-basic cleavage site and the PDH/PDF octadecapeptide with an N-terminal Gly for amidation, and a mono- or di- or tri-basic cleavage site prior to the translation stop signal. The PAP does not display evolutionary conservation in either its length or primary sequence.

4.11.4.3.3. Cellular release of PDH/PDF There is a rhythm in antibody staining for PDF in the terminals of the s-vLNs of the adult *Drosophila* brain (J.H. Park *et al.*, 2000 – discussed further below). This observation is thought to reflect a daily rhythm of release. In addition, Kaneko *et al.* (2000a) and Blanchardon *et al.* (2001) reported that overexpression of an active tetanus toxin in *pdf* neurons of *Drosophila* (with which to cleave synaptobrevin and so reduce evoked release; Sweeney *et al.*, 1995) was not effective in disrupting circadian behavior. It was predicted that disruption of transmitter release by those neurons would have a strong behavioral effect due to results seen when either the *pdf* gene was mutant or the cells ablated (see below). The lack of a behavioral phenotype in these experiments may be explained by a lack of sensitivity by the PDF peptidergic release system to tetanus toxin. It may be that *Drosophila* peptidergic release relies on molecules different from synaptobrevin. However, there are two points that suggest caution in accepting that interpretation. For another peptidergic system of *Drosophila*, release of the eclosion hormone by two identified neurons is sensitive to this toxin (McNabb *et al.*, 1997) (see **Chapter 3.1**). That result suggests at least some *Drosophila* peptidergic neurons are likely to employ synaptobrevins in the exocytosis of peptidergic secretory granules. In addition, the *UAS-tetanus toxin* system produces only incomplete cleavage of synaptobrevin (Sweeney *et al.*, 1995), leaving open the possibility that the levels of toxin expression were not sufficient to effect a complete block of release in this instance.

4.11.4.3.4. Control of PDF by the clockworks In mammals, several SCN transmitters display diurnal and/or circadian variation in their expression reflecting control by the clockworks (e.g., Jin *et al.*, 1999). To what extent is PDF expression controlled by clockwork genes? Given a circadian role for PDF signaling, it is now important to ask how the circadian clock produces a diurnal PDF signal. Several cellular phenomena represent potential points of clock regulation: these include *pdf* transcription and translation, the electrical activity of *pdf*-expressing neurons, and the sensitivity of PDF-receptive neurons. We now know that PDF expression is regulated by components of the circadian clock, but the details of that regulation are still emerging and they reveal a number of unexpected features. Genetic observations show that *pdf* RNA is positively regulated by the transcription factors *Clock* and *cyc* (J.H. Park *et al.*, 2000) and is also regulated by *vrille* (Blau and Young, 1999). The effect of the dominant negative Clk^{Jrk} allele appeared more severe than that of *cyc* alleles, consistent with the hypothesis that *Clk* may have additional partners in mediating its control on *pdf* (J.H. Park *et al.*, 2000). As yet, no factors have been shown to regulate *pdf* expression by direct transcriptional activation assays. In the specific case of regulation by CLK : CYC, their actions appear to be indirect (J.H. Park *et al.*, 2000). The regulators *Clk* and *vri* produce distinct effects on PDF expression: continuous expression of *vri* produced a decrease in PDF levels in larval pacemaker neurons (vLNs) but no effect on *pdf* RNA levels (Blau and Young, 1999).

A remarkable feature of this clockwork regulation of PDF expression is its exquisite cell type specificity: it is seen only in clock neurons (not in the abdominal PDF neurons) and, more specifically, only in some clock neurons (the s-vLNs). In spite of its evident clock-controlled transcriptional regulation, *pdf* RNA does not fluctuate on a daily basis, nor does PDF immunoreactivity vary in large-scale fashion (Park and Hall, 1998). This conclusion was subsequently supported by many microarray experiments (op. cit. Section 4.11.4.1). *Clk* and *cyc*

mutants also displayed aberrations in PDF s-vLN axonal processes, consistent with a normal role in those cells' morphological differentiation (J.H. Park *et al.*, 2000). Such coordinate control of transmitter expression and axonal morphology by a single transcription factor may reflect a broader developmental theme, akin to similar demonstrations for other nonclock *Drosophila* neurons (Thor and Thomas, 1997; Allan *et al.*, 2003). In sum, the available data suggest substantial clockwork control of PDF expression. It also indicates that neither *pdf* transcription nor its translation contribute extensively to the principal gating mechanism within the s-vLN neurons.

There is substantial evidence for a daily rhythm of PDF release. In houseflies, PDH injections cause a swelling of L1 and L2 (lamina) axonal diameters that mimics a diurnal change of axon caliber these neurons normally display (Pyza and Meinertzhagen, 1996): an effect consistent with a daytime release of PDF. *Drosophila* exhibits similar diurnal and circadian changes in laminar axonal diameters and shape (Pyza and Meinertzhagen, 1999). Pyza and Meinertzhagen (1997) also reported that PDH varicosities in the optic lobes display a circadian rhythm in size and spacing: fewer in the subjective day, again consistent with a daily, daytime PDF release. PDF immunoreactivity within s-vLN terminals of *Drosophila* varies in a diurnal and circadian fashion (J.H. Park *et al.*, 2000). It is high in the subjective day (with a peak ~1 h after the start of the subjective day) and low in the subjective night. These findings are also consistent with a hypothesis of release of PDF during the subjective day. This predicted daily PDF release event displays clock influence, in that the period of the staining variation is sensitive to a period length-altering allele of *per* (J.H. Park *et al.*, 2000). By way of speculation, the daily changes in PDF neurite morphology of the larger flies may find a parallel in the phenotype of the fragile X-related protein gene (*dFMR-1*: Dockendorf *et al.*, 2002; Inoue *et al.*, 2002; Morales *et al.*, 2002). As discussed below, *dFMR-1* mutant animals display both circadian phenotypes and alterations in PDF cell branching.

The *per* and *tim* mutations did not affect *pdf* RNA levels, but did affect the level of PDF expression in the s-vLN terminals. Surprisingly, they have opposite effects in this regard. A per^0 allele caused consistently low PDF staining, while a tim^0 allele caused consistently high PDF staining (J.H. Park *et al.*, 2000). In sum, observations to date indicate multiple levels of control by the clockworks on PDF, specifically within the s-vLNs. *Clk* and *cyc* appear to affect *pdf* gene expression and PDF cell differentiation, while *vri, per*, and *tim* appear to affect a later step(s) in PDF expression, perhaps involving rhythmic transport, processing and/or release. A more complete definition of *pdf* regulation and an understanding of how its fluctuations contribute to gated PDF signaling represent important future goals.

4.11.4.3.5. PDF physiology and genetics *Drosophila* mutants for *pdf* were discovered resident among laboratory stocks of long standing (Renn *et al.*, 1999). These animals contain a nonsense mutation in the signal sequence of preproPDF and are protein nulls. The mutant animals appear and behave normally in most respects. PDF neurons are present and appear fully differentiated in the mutant background. However, the circadian clock-regulated behavior of *pdf* mutant animals is highly abnormal: while they entrain to light signals, a large majority of the population displays arrhythmicity under constant darkness (DD). The arrhythmicity of *pdf* mutants is not evident for 1 to 3 cycles of DD. This is unlike the phenotype of animals bearing mutations in clock genes like *period* or *timeless* for which arrhythmic behavior is evident as soon as the animal is placed in DD. Transgenic expression of wild-type *pdf* sequences restored expression of PDF peptide and rhythmic behavior to a great extent. Ablation of the PDF neurons (affected by genetic targeting) in an otherwise wild-type background produced a behavioral phenotype that was in all ways comparable to that produced by *pdf* mutant flies. A similar conclusion was reached by Blanchardon *et al.* (2001) who also ablated *pdf* neurons genetically using a specific GAL4 insertion that prominently features that cell group. Additionally, Helfrich-Förster *et al.* (2000) reported that overexpression of PDF by the UAS : GAL4 system in neurons projecting to the dorsal brain resulted in severe arrhythmicity. That result is consistent with the hypothesis that activation of PDF receptor(s) affects circadian locomotor activity, and that the timing or level of such activation is important to the signaling.

Together these observations indicate that the phenotypic deficits exhibited by *pdf* mutant animals could be attributed in large part to the absence of the *pdf* gene product, and they lead to several conclusions. First, they contribute directly to the hypothesis that the PDF-expressing vLNs are the primary pacemakers underlying control of daily locomotion (cf. Frisch *et al.*, 1994; Vosshall and Young, 1995). Second, they indicate that *pdf* is the sole functional output of vLN pacemakers and that PDF is the principal circadian transmitter in

Drosophila. Third, they predict that other neurons provide auxiliary pacemaking mechanisms and that they must release other (e.g., non-PDF) transmitters (cf. Taghert *et al*., 2001; Allada *et al*., 2003; Helfrich-Förster, 2003; Peng *et al*., 2003; Veleri *et al*., 2003).

4.11.4.3.6. PDF effects on pacemaker synchronization The phenotype of *pdf* mutant flies is consistent with either of two extreme cellular models. First, *pdf* signaling may be required to maintain synchronization and/or oscillation among/in pacemaker neurons. Additionally, *pdf* may be a factor that couples the pacemaker network to premotor centers that govern rhythmic behaviors. These models are not mutually exclusive. Three sets of studies have addressed the first possibility. Petri and Stengl (1997) injected PDH into cockroach brains and reported a significant effect on the phase of subsequent cycles of locomotor activity. PDH caused a phase-dependent ~3-h phase delay consistent with it being an input signal that helps sharpen the synchronization of circadian pacemaker neurons. The maximal effect was during the late subjective day, and the shape of the phase-response curve suggested that PDH presents a nonphotic input to the clock, perhaps especially important with respect to coupling the clock outputs from both sides of the brain. In similar fashion, the *Drosophila* mutant for *pdf* entrained to LD cycles, but with a phase-advanced activity peak (Renn *et al*., 1999). One interpretation of the latter observation is that *pdf* normally acts to delay the phase of the wild-type activity rhythm. That would be congruent with its pharmacological activity in the larger insects. In this regard, an important role for PDF may therefore be as clock input – to help synchronize the phases of disparate clock neurons. These studies suggest that other transmitters are likely to play similar input roles, promoting either clock phase advances or delays (cf. Volkner *et al*., 2000; Petri *et al*., 2002; Wegener *et al*., 2004).

More recently, Peng *et al*. (2003) analyzed the pattern of *tim* RNA *in situ* in the *Drosophila* brain as a measure of molecular pacemaker function in control versus *pdf* mutant animals. On DD5, control brains displayed a strong rhythm of *tim in situ* signals in each of the several sites. In contrast, the amplitude of cycling was much diminished in *pdf* mutant brains. The authors concluded that *pdf* was required to maintain high amplitude, molecular rhythms as well as synchronized activity among different pacemaker groups, akin to a recent hypothesis describing the role of neuropeptide VIP receptors in the mammalian SCN (Harmar *et al*., 2002). They speculate that the effect they observe may be related to the demonstration that membrane electrical activity is required by *Drosophila* pacemaker neurons in order to maintain molecular rhythms (Nitabach *et al*., 2002). A different conclusion was reached by Blanchardon *et al*. (2001) who monitored PER immunostaining on DD2-3 in brains of flies that had had the PDF-expressing vLNs (and several other) neurons genetically ablated. That group reported normal PER staining fluctuations in dLN and DN neuronal groups. Two technical differences between these studies are notable and may explain the differences – the amount of time flies were in DD before analysis and the nature of the probes (*tim in situ* probe versus anti-PER antibodies).

4.11.4.3.7. *pdf* and geotaxis Toma *et al*. (2002) reported on studies of highly inbred *Drosophila* stocks that display either a positive or negative geotaxic bias. Flies called "Hi" typically distribute at higher perches of a vertical maze; the opposite is true of flies called "Lo." Control animals typically occupy intermediate positions. The authors reasoned that such behavioral biases likely reflect polygenic differences and so they employed microarray technology to determine which genes showed average level differences between the two stocks. Surprisingly, *pdf* was among the genes found to be reliably different. Further, when tested in the same geotaxic maze device, *pdf* nulls displayed a negative geotaxic bias as great as that of "Hi" flies. Other clock mutants did not display a similar phenotype. Remarkably, the geotaxic test score was a function of *pdf* gene dosage; this presumably suggests that the behavior is sensitive to several quantitative levels of PDF peptide. Given the limited sites of PDF expression in the adult animal (essentially two sets of neurons, see above), it should be possible to narrow down the cellular sites of action using molecular genetic techniques.

4.11.4.3.8. PDF receptors There is currently no direct information on the identity of receptors that mediate PDF effects in any organism. Such data would help address many outstanding questions regarding the sites and mechanisms of PDF actions. Peng *et al*. (2003) reported recently on the distribution of biotinylated PDF in the *Drosophila* CNS. They applied the labeled probe to fixed brains in whole mount and described a pattern of binding that almost entirely overlapped the pattern of *tim-GAL4* gene expression. That result is consistent with an intriguing hypothesis proposed by those authors – that within the CNS, PDF receptors are

largely associated with neuronal pacemakers and support their synchronization. However, the report also indicated a need for more characterization of these putative PDF receptor sites: the molecular specificity was uncertain, as the staining effects could not be competed out by unlabeled PDF, even when applied at a 5000-fold excess. In addition, the staining patterns displayed an inconsistent character.

Structure-activity studies of the crustacean PDH have revealed possible facts about its interactions with a putative receptor and/or with degradative enzymes (Riehm and Rao, 1982). N-terminal deletions of PDH, removing as many as five amino acids, still resulted in the retention of some (weak) biological activity. Progressively re-extending the molecule back towards the N-terminus restores activity incrementally. Loss of the C-terminal amide resulted in ~300-fold loss of activity (Riehm *et al.*, 1985).

Because all small neuropeptides act via G protein-coupled receptors (GPCRs; see **Chapter 5.5**), it is reasonable to think that PDF receptors also act in this manner. Nery and Castrucci (1997) suggested that PDH effects in crustacean chromatophores reflect G's activation following binding to a GPCR. Hewes and Taghert (2001) found evidence for 44 (±1) peptide GPCRs in the *Drosophila* genome. Those assignments rested on the assumption that *Drosophila* peptide GPCRs are related to those of other animals by sequence similarities within the highly conserved seven transmembrane domains. Pharmacological study of many of these molecules following functional expression has proceeded using different methods, including signaling assays that rely on a promiscuous G protein (e.g., Cazzamali and Grimmelikhuijzen, 2002; Meeusen *et al.*, 2002), assays of membrane currents in *Xenopus* (e.g., Park *et al.*, 2002), assays of competitive radioligand binding (e.g., Johnson *et al.*, 2003a), and assays of ligand-dependent translocation of β-arrestin-GFP (Johnson *et al.*, 2003b). To date, roughly 26 of the 44 indicated peptide receptors have been "de-orphaned" by one or more methods. Hopefully, identification and analysis of the PDF receptor(s) will occur in the near future.

4.11.4.4. Second Messenger Systems Mediating PDF and/or Circadian Output

4.11.4.4.1. cAMP Only a few studies have been conducted on the signaling of PDH in crustaceans. It is suggested that pigment dispersion is achieved by ligand binding to a G_s protein-coupled receptor, resulting in the activation of adenylate cyclase and the increase of the intracellular concentration of cAMP, which in turn activates a cAMP-dependent protein kinase (Nery and Castrucci, 1997) (see also **Chapter 3.2**). There is genetic evidence supporting the role of the catalytic subunit of protein kinase A (PKA – called *DCO*) in circadian output controlling locomotor rhythms (Majercak *et al.*, 1997), as well as for a type II PKA regulatory subunit (S.K. Park *et al.*, 2000). However, there is no information yet available to indicate at what cellular/synaptic level PKA may be acting, nor whether this action reflects PDF action. Finally, Belvin *et al.* (1999) reported that *dCREB2* cycles in circadian fashion, and that it may help sustain rhythmicity of *period* gene expression by affecting both circadian period and amplitude. These observations suggest that cAMP signaling is likely important at several levels in the circadian system.

4.11.4.4.2. Nf1/MAP kinase Williams *et al.* (2001) showed an involvement by several putative signaling components that act downstream of PDF in mediating *Drosophila* circadian output. The *Drosophila neurofibromatosis 1* (*Nf1*) gene product neurofibromin regulates both Ras and cyclic AMP (cAMP) (The *et al.*, 1997). *Nf1* mutants exhibited virtually no rhythmic behavior under constant conditions. However, PER and TIM protein, and *tim* RNA levels in the head continued to oscillate normally in those flies, suggesting a defect downstream of the clockworks. Likewise, PER and TIM proteins also continued to oscillate in the larval LN pacemaker neurons. Furthermore, attempts at phenotypic rescue by restoring wild-type *Nf1* only to the *pdf*-expressing vLN neurons was not effective. Reduced Ras activity partially rescued the circadian rhythm defect of *Nf1* mutants, suggesting that NF1 normally regulates Ras in control of this behavior. A link to PDF action was made by demonstrating that phospho-MAP kinase levels were substantially elevated in *Nf1* mutants, that such kinase levels cycled diurnally, and that such cycling was reduced in *pdf* mutants. The authors concluded that PDF driven circadian output is mediated at least in part by NF1 activity through a Ras/MAP kinase-dependent signaling pathway (Williams *et al.*, 2001) (see also **Chapter 3.2**).

4.11.4.5. Contributions to Circadian Output by Other Transmitters

4.11.4.5.1. Other neuropeptides It follows from the chemical heterogeneity of pacemaker neurons (PDF is only expressed by a subset of them) that other transmitters must play important roles in the circuits underlying circadian output. Taghert *et al.* (2001) presented evidence to suggest that these

include other amidated peptides different from PDF. In particular, certain peptidylglycine alpha-hydroxylating monooxygenase (*PHM*) mosaic animals displayed a more severe locomotor defect than did *pdf* mutant animals. PHM is an enzyme responsible for the initial step in C-terminal α-amidation, a posttranslational modification displayed by more than 90% of all known *Drosophila* neuropeptides (Hewes and Taghert, 2001). Importantly, amidation is typically required for full biological activity of neuropeptides (see Riehm *et al*., 1985, for an example focusing on PDH), and it is specific to secretory peptides. The amidated peptide corazonin is a good candidate for further consideration as it is found in many insects to be co-localized with PER proteins (e.g., Wise *et al*., 2002; Bloch *et al*., 2003).

Ecdysial behaviors, which are gated by the circadian clock in some cases (e.g., adult eclosion; Truman, 1992) (see also **Chapter 3.1**), are regulated by several neuropeptides that act in a complex fashion (Ewer and Reynolds, 2002). Release of these factors (including eclosion hormone and CCAP) may be closely (perhaps directly) controlled by clock neurons (Park *et al*., 2003).

4.11.4.5.2. Other transmitters Two sets of studies highlight circadian changes in the conventional transmitter signaling in the fly. Wide field serotinergic neurons (and PDH-containing neurons) have both been proposed to modulate visual input in the optic ganglia (Nässel *et al*., 1991; Pyza and Meinertzhagen, 1997). To investigate possible rhythms in functional aspects of the visual system, Chen *et al*. (1999) described a diurnal and circadian change in the amplitude of ON and OFF transients in the electroretinogram (ERG) of the fly retina. Consistent with its circadian-regulated release, serotonin injections enhanced the amplitude of ERG transients, and antagonists decreased them. Andretic *et al*. (1999) found a clock-controlled change in dopamine receptor expression linked to locomotor output (also discussed below in Section 4.11.4.6.5).

4.11.4.6. Evidence for Additional Processes Related to Circadian Output

4.11.4.6.1. The clock controlling rhythmic eclosion Truman and Riddiford (1970) first described the control of adult eclosion by a circadian clock located within the silkmoth brain. They used brain extirpation and reimplantation techniques to demonstrate that an endocrine signal (since identified as the peptide "eclosion hormone," EH; see **Chapter 3.1**) is released by the brain according to a signal gated by a photoperiod-sensitive clock. Recent studies of the moth have identified specific EH neurons, permitting their comparison to putative sites of *period* gene expression (Sauman and Reppert, 1996). EH neurons do not appear to be intrinsic circadian pacemakers; instead they are likely driven by higher order pacemakers, which may themselves include other neurosecretory neurons. Myers *et al*. (2003) re-examined the cellular basis for rhythmic eclosion. They found that circadian oscillations in both the vLNs and the prothoracic gland (PG, a peripheral endocrine organ; see **Chapter 3.3**) were required for normal rhythmic eclosion. Further, the action of LNs on eclosion and on the rhythms of PER and TIM in the PG was shown to be mediated by its (presumed) secretion of the neuropeptide PDF. The manipulations used affected both clock PDF neurons (vLNs) and nonclock neurons (abdominal neuroendocrine neurons). It is therefore possible that the vLN PDF brain neurons regulate the PG directly (via hemolymph-borne PDF) or indirectly via the neurons themselves controlled by the vLNs.

4.11.4.6.2. A humoral signal controlling daily locomotion Handler and Konopka (1979) performed brain transplants similar to those performed by Truman and Riddiford (1970), but in the much smaller *Drosophila* to examine the possible influence of brain secretions on rhythmic activity. They found that in a small number of instances, brains from per^S animals could restore rhythmicity when transplanted into animals that were otherwise arrhythmic due to their carrying a per^{01} allele. The implicated molecule(s) remain unidentified. In summary, there is evidence of diffusible (hormonal) signals normally influencing rhythmic locomotor and eclosion behaviors in insects.

4.11.4.6.3. Fragile X-related protein Several groups have recently implicated the *dFMR-1* gene product as a component of circadian output. *dFMR-1* encodes an RNA-binding protein related to the mammalian fragile X protein. In *Drosophila*, it is a determinant of neuronal morphology (Zhang *et al*., 2001; Dockendorf *et al*., 2002; Morales *et al*., 2002; Lee *et al*., 2003; Schenck *et al*., 2003). Each group determined that *dFMR-1* mutant flies were largely arrhythmic in constant darkness (Dockendorf *et al*., 2002; Inoue *et al*., 2002; Morales *et al*., 2002). Inoue *et al*. (2002) found that the mutants had normal eclosion rhythms, while the other two groups reported either phase (Dockendorf *et al*., 2002), or phase and amplitude (Morales *et al*., 2002)

disturbances in the eclosion rhythm. The vLNs display specific axonal branching defects (Morales *et al*., 2002), although Dockendorf *et al*. (2002) reported that these were not observed uniformly. Rhythms of CREB activity were much reduced in mutants (Morales *et al*., 2002), but rhythmic *per* expression was maintained (Morales *et al*., 2002). Interestingly, the double mutant *futsch, dFMR-1* suppressed *dFMR-1* phenotypes in motor neurons (synaptic overgrowth and defects in neurotransmission; Zhang *et al*., 2001). However, that genotype did not suppress the *dFMR-1* circadian phenotype (Dockendorf *et al*., 2002). In sum, *dFMR-1* mutants maintained clock gene expression, but displayed alterations in behavioral rhythms and in the morphology of pacemaker neurons, and alterations in at least one molecular index of clock output. Determining where and how *dFMR-1* acts will provide valuable information concerning which circadian output mechanisms may be defined further.

4.11.4.6.4. *takeout* Further downstream, the *takeout* gene contributes to the animals' starvation responses: this gene responds to starvation with an upregulation of transcription that is dependent on clock gene expression (Sarov-Blat *et al*., 2000). Its transcription is dependent on expression of most of the major clock genes, but with a distinct delayed phase (So *et al*., 2000). Additional information concerning *takeout* functions comes from Dauwalder *et al*. (2002) who report that the gene is controlled by the sex determination pathway (see **Chapter 1.7**), and that a loss-of-function allele affects male courtship. Further, they argue that *takeout* is a member of a large family of genes encoding secreted factors that bind small lipophilic molecules. There exist apparent differences in the reported patterns of *takeout* gene expression in these various reports. Further studies are needed to help form a consensus regarding *takeout* functions, and so better understand its role in the circadian system.

4.11.4.6.5. *Lark* Several lines of evidence demonstrate that the Lark protein participates in construction of a circadian gate for rhythmic eclosion behavior (Newby and Jackson, 1996; McNeil *et al*., 1998). *Lark* heterozygotes alter the phase of the eclosion gate, but not its period (*Lark* homozygotes die during embryogenesis, likely due to loss of other functions). Furthermore, *Lark* rhythm defects are behavior specific in that they affect eclosion, but not daily locomotor activity. This protein has the molecular signature of an RNA binding protein and is widely expressed in the nucleus of most cells. Significantly, the protein accumulates in the cytoplasm of specific peptidergic (CCAP) neurons (Zhang *et al*., 2000). In other insects, these neurons have been implicated in triggering eclosion behavior motor patterns (Ewer and Truman, 1996; Park *et al*., 2003). Identifying the signal transduction pathways regulated by circadian transmitters and relating Lark actions to the elaboration of a circadian gate are challenges for the near future.

4.11.4.6.6. Circadian regulation of drug sensitivity in *Drosophila* Hirsh and colleagues have related circadian outputs to drug sensitivity using a decapitated fly assay that allows drug application directly to the CNS. Decapitated flies retain some CNS structures, i.e., the ventral nerve cord, which directly controls somatic musculature and gut motility. Quinpirole, a D2-like dopamine receptor agonist, induces reflexive locomotion, and is most effective during the subjective night (Andretic and Hirsh, 2000). Those studies indicated that dopamine receptor responsiveness is under circadian control and depends on the normal function of the *period* gene. Flies also show behavioral responsiveness to free base cocaine and sensitization to repeated cocaine doses. Four of five circadian genes tested (*per*, *Clk*, *cyc*, and *dbt*) altered cocaine sensitization responsiveness (Andretic *et al*., 1999). Similar phenomena were later described in the mouse (Abarca *et al*., 2002). Mutant flies also show a lack of tyrosine decarboxylase induction normally seen with cocaine administration. Interestingly, tim^{01} mutants displayed normal cocaine responses. Andretic *et al*. (1999) deduce that an unidentified PER binding partner is specifically involved in regulation of drug responsiveness. This result also suggests that drug responsiveness is likely regulated by *per* expression in a set of cells distinct from those involved in circadian function.

4.11.4.6.7. Circadian output and sleep The biological clock controls many rhythmic output processes, some overt and some obscure. Among these, sleep is arguably the most obvious. For many, it is also the most mysterious. Studies that focus on *Drosophila* rest as a model began in earnest several years ago (Hendricks *et al*., 2000; Shaw *et al*., 2000). Those empirical findings have provided a conceptual framework for which rest in flies can formally correspond to mammalian sleep (reviews: Hendricks, 2003; Shaw, 2003). A significant element in that concept is the demonstration that flies display a robust sleep homeostat – a mechanism to increase the duration of sleep (sleep rebound) within a daily cycle in response to some sleep deprivation experienced in past cycles.

Given the rhythmic nature of sleep, the possible effects of circadian clock mutations on *Drosophila* sleep has been an early and enduring point of interest. Two sets of observations describe the current understanding of the relationship between the circadian and sleep mechanisms in *Drosophila*. First, sleep rebound following a 3-, 6-, or 9-h deprivation protocol is displayed in the cycle immediately following, and it occurs principally during the subjective day (Hendricks *et al*., 2000; Shaw *et al*., 2000). Hence, sleep homeostasis is normally gated by the circadian clock. Beyond this permissive feature, does the clock provide regulatory control over sleep mechanisms? The second set of observations suggests the answer is yes. Clock gene mutants display predictable alterations in the dynamics and extent of sleep homeostasis (Shaw *et al*., 2002; Hendricks *et al*., 2003). There is clear consensus from observing mutant phenotypes that at least some of the core clock elements may directly regulate aspects of sleep. For example, *per* and *Clk* mutants display 100% recovery from sleep deprivation with one daily cycle, whereas control stocks only recover ~40% in that same time period. Additionally, *cyc* and *tim* mutant flies display increased or decreased amounts of sleep rebound after deprivation that reflect the amount of deprivation experienced. The phenotype of *cyc* mutant flies in response to sleep deprivation is especially intriguing, and can include exaggerated sleep rebound, increases in baseline sleep amount (set-point), and lethality. That syndrome suggests that not all components of the molecular pathway, as understood in a circadian context, participate equally in a sleep context. Nevertheless, these data together strongly imply that there is a close mechanistic association between elements of the circadian clock and the sleep homeostat. With this strong foundation, we can anticipate that further genetic analysis in *Drosophila* will make contributions to help define hypotheses by which to describe underlying mechanisms.

Acknowledgments

We are very grateful to Erik Johnson, Russ Van Gelder, and Paul Shaw for helpful discussions, and for reading sections of earlier drafts. We apologize to colleagues whose work was not covered for lack of space or oversight. We thank David Van Essen whose pun we refashioned to compose a title for this chapter. YL is supported by the Washington University Medical School MSTP Program. Work in PT's laboratory is supported by grants from the NINCDS (NS21749) and by the NIMH (MH067122).

References

Abarca, C., Albrecht, U., Spanagel, R., **2002**. Cocaine sensitization and reward are under the influence of circadian genes and rhythm. *Proc. Natl Acad. Sci. USA 99*, 9026–9030.

Akten, B., Jauch, E., Genova, G.K., Kim, E.Y., Edery, I., *et al*., **2003**. A role for CK2 in the *Drosophila* circadian oscillator. *Nature (Lond.) Neurosci. 6*, 251–257.

Allada, R., Kadener, S., Nandakumar, N., Rosbash, M., **2003**. A recessive mutant of *Drosophila Clock* reveals a role in circadian rhythm amplitude. *Eur. Mol. Biol. Org. J. 22*, 3367–3375.

Allada, R., White, N.E., So, W.V., Hall, J.C., Rosbash, M., **1998**. A mutant *Drosophila* homolog of mammalian clock disrupts circadian rhythms and transcription of *period* and *timeless*. *Cell 93*, 791–804.

Allan, D.W., St. Pierre, S.E., Miguel-Aliaga, I., Thor, S., **2003**. Specification of neuropeptide cell identity by the integration of retrograde BMP signaling and a combinatorial transcription factor code. *Cell 113*, 73–86.

Andretic, R., Chaney, S., Hirsh, J., **1999**. Requirement of circadian genes for cocaine sensitization in *Drosophila*. *Science (Wash.) 285*, 1066–1068.

Andretic, R., Hirsh, J., **2000**. Circadian modulation of dopamine receptor responsiveness in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA 97*, 1873–1878.

Ashmore, L.J., Sathyanarayanan, S., Silvestre, D.W., Emerson, M.M., Schotland, P., *et al*., **2003**. Novel insights into the regulation of the timeless protein. *J. Neurosci. 23*, 7810–7819.

Ashmore, L.J., Sehgal, A., **2003**. A fly's eye view of circadian entrainment. *J. Biol. Rhyth. 18*, 206–216.

Bae, K., Lee, C., Sidote, D., Chuang, K.Y., Edery, I., **1998**. Circadian regulation of a *Drosophila* homolog of the mammalian *Clock* gene: PER and TIM function as positive regulators. *Mol. Cell. Biol. 18*, 6142–6151.

Bao, S., Rihel, J., Bjes, E., Fan, J.Y., Price, J.L., **2001**. The *Drosophila double-timeS* mutation delays the nuclear accumulation of period protein and affects the feedback regulation of *period* mRNA. *J. Neurosci. 21*, 7117–7126.

Baylies, M.K., Bargiello, T.A., Jackson, F.R., Young, M.W., **1987**. Changes in abundance or structure of the *per* gene product can alter periodicity of the *Drosophila* clock. *Nature (Lond.) 326*, 390–392.

Belvin, M.P., Zhou, H., Yin, J.C., **1999**. The *Drosophila dCREB2* gene affects the circadian clock. *Neuron 22*, 777–787.

Blanchardon, E., Grima, B., Klarsfeld, A., Chelot, E., Hardin, P.E., *et al*., **2001**. Defining the role of *Drosophila* lateral neurons in the control of circadian rhythms in motor activity and eclosion by targeted genetic ablation and PERIOD protein overexpression. *Eur. J. Neurosci. 13*, 871–888.

Blau, J., Young, M.W., **1999**. Cycling *vrille* expression is required for a functional *Drosophila* clock. *Cell 99*, 661–671.

Block, G., Geusz, M., Khalsa, S., Michel, S., Whitmore, D., **1995**. Cellular analysis of a molluscan retinal biological clock. *Ciba Found. Symp. 183*, 51–60.

Bloch, G., Solomon, S.M., Robinson, G.E., Fahrbach, S.E., **2003**. Patterns of PERIOD and pigment-dispersing hormone immunoreactivity in the brain of the European honeybee (*Apis mellifera*), age- and time-related plasticity. *J. Comp. Neurol. 464*, 269–284.

Brandes, C., Plautz, J.D., Stanewsky, R., Jamison, C.F., Straume, M., *et al.*, **1996**. Novel features of *Drosophila period* transcription revealed by real-time luciferase reporting. *Neuron 16*, 687–692.

Burbach, K.M., Poland, A., Bradfield, C.A., **1992**. Cloning of the Ah-receptor cDNA reveals a distinctive ligand-activated transcription factor. *Proc. Natl Acad. Sci. USA 89*, 8185–8189.

Cashmore, A.R., Jarillo, J.A., Wu, Y.J., Liu, D., **1999**. Cryptochromes, blue light receptors for plants and animals. *Science (Wash.) 284*, 760–765.

Cazzamali, G., Grimmelikhuijzen, C.J., **2002**. Molecular cloning and functional expression of the first insect FMRFamide receptor. *Proc. Natl Acad. Sci. USA 99*, 12073–12078.

Ceriani, M.F., Darlington, T.K., Staknis, D., Mas, P., Petti, A.A., *et al.*, **1999**. Light-dependent sequestration of TIMELESS by CRYPTOCHROME. *Science (Wash.) 285*, 553–556.

Ceriani, M.F., Hogenesch, J.B., Yanovsky, M., Panda, S., Straume, M., *et al.*, **2002**. Genome-wide expression analysis in *Drosophila* reveals genes controlling circadian behavior. *J. Neurosci. 22*, 9305–9319.

Chen, B., Meinertzhagen, I.A., Shaw, S.R., **1999**. Circadian rhythms in light-evoked responses of the fly's compound eye, and the effects of neuromodulators 5-HT and the peptide PDF. *J. Comp. Physiol. (A) 185*, 393–404.

Chen, Y., Hunter-Ensor, M., Schotland, P., Sehgal, A., **1998**. Alterations of *per* RNA in noncoding regions affect periodicity of circadian behavioral rhythms. *J. Biol. Rhyth. 13*, 364–379.

Cheng, Y., Hardin, P.E., **1998**. *Drosophila* photoreceptors contain an autonomous circadian oscillator that can function without *period* mRNA cycling. *J. Neurosci. 18*, 741–750.

Chuman, Y., Matsushima, A., Sato, S., Tomioka, K., Tominaga, Y., *et al.*, **2002**. cDNA cloning and nuclear localization of the circadian neuropeptide designated as pigment-dispersing factor PDF in the cricket *Gryllus bimaculatus*. *J. Biochem. (Tokyo) 131*, 895–903.

Citri, Y., Colot, H.V., Jacquier, A.C., Yu, Q., Hall, J.C., *et al.*, **1987**. A family of unusually spliced biologically active transcripts encoded by a *Drosophila* clock gene. *Nature (Lond.) 326*, 42–47.

Claridge-Chang, A., Wijnen, H., Naef, F., Boothroyd, C., Rajewsky, N., *et al.*, **2001**. Circadian regulation of gene expression systems in the *Drosophila* head. *Neuron 32*, 657–671.

Crews, S.T., Thomas, J.B., Goodman, C.S., **1988**. The *Drosophila single-minded* gene encodes a nuclear protein with sequence similarity to the *per* gene product. *Cell 52*, 143–151.

Curtin, K.D., Huang, Z.J., Rosbash, M., **1995**. Temporally regulated entry of the *Drosophila period* protein contributes to the circadian clock. *Neuron 14*, 365–372.

Cyran, S.A., Buchsbaum, A.M., Reddy, K.L., Lin, M.C., Glossop, N.R., *et al.*, **2003**. vrille, Pdp1, and dClock form a second feedback loop in the *Drosophila* circadian clock. *Cell 112*, 329–341.

Darlington, T.K., Lyons, L.C., Hardin, P.E., Kay, S.A., **2000**. The period E-box is sufficient to drive circadian oscillation of transcription *in vivo*. *J. Biol. Rhyth. 15*, 462–471.

Darlington, T.K., Wager-Smith, K., Ceriani, M.F., Staknis, D., Gekakis, N., *et al.*, **1998**. Closing the circadian loop, CLOCK-induced transcription of its own inhibitors per and tim. *Science (Wash.) 280*, 1599–1603.

Dauwalder, B., Tsujimoto, S., Moss, J., Mattox, W., **2002**. The *Drosophila takeout* gene is regulated by the somatic sex-determination pathway and affects male courtship behavior. *Genes Devel. 16*, 2879–2892.

Dockendorf, T.C., Su, H.S., McBride, S.M., Yang, Z., Choi, C.H., *et al.*, **2002**. *Drosophila* lacking *dfmr1* activity show defects in circadian output and fail to maintain courtship interest. *Neuron 34*, 973–984.

Dunlap, J., **1999**. Molecular bases for circadian clocks. *Cell 96*, 271–290.

Dushay, M.S., Rosbash, M., Hall, J.C., **1989**. The *disconnected* visual system mutations in *Drosophila melanogaster* drastically disrupt circadian rhythms. *J. Biol. Rhyth. 4*, 1–27.

Edery, I., Rutila, J.E., Rosbash, M., **1994b**. Phase shifting of the circadian clock by induction of the *Drosophila period* protein. *Science (Wash.) 263*, 237–240.

Edery, I., Zwiebel, L.J., Dembinska, M.E., Rosbash, M., **1994a**. Temporal phosphorylation of the *Drosophila* period protein. *Proc. Natl Acad. Sci. USA 91*, 2260–2264.

Egan, E.S., Franklin, T.M., Hilderbrand-Chae, M.J., McNeil, G.P., Roberts, M.A., *et al.*, **1999**. An extraretinally expressed insect cryptochrome with similarity to the blue light photoreceptors of mammals and plants. *J. Neurosci. 19*, 3665–3673.

Elekes, K., Nässel, D.R., **1999**. Pigment-dispersing hormone-like immunoreactive neurons in the central nervous system of the gastropods, *Helix pomatia* and *Lymnaea stagnalis*. *Cell Tiss. Res. 295*, 339–348.

Emery, P., So, W.V., Kaneko, M., Hall, J.C., Rosbash, M., **1998**. CRY, a *Drosophila* clock and light-regulated cryptochrome, is a major contributor to circadian rhythm resetting and photosensitivity. *Cell 95*, 669–679.

Emery, P., Stanewsky, R., Hall, J.C., Rosbash, M., **2000a**. A unique circadian-rhythm photoreceptor. *Nature (Lond.) 404*, 456–457.

Emery, P., Stanewsky, R., Helfrich-Förster, C., Emery-Le, M., Hall, J.C., *et al.*, **2000b**. *Drosophila* CRY is a deep brain circadian photoreceptor. *Neuron 26*, 493–504.

Ewer, J., Frisch, B., Hamblen-Coyle, M.J., Rosbash, M., Hall, J.C., **1992**. Expression of the *period* clock gene

within different cell types in the brain of *Drosophila* adults and mosaic analysis of these cells' influence on circadian behavioral rhythms. *J. Neurosci. 12*, 3321–3349.

Ewer, J., Hamblen-Coyle, M., Rosbash, M., Hall, J.C., **1990**. Requirement for *period* gene expression in the adult and not during development for locomotor activity rhythms of imaginal *Drosophila melanogaster*. *J. Neurogenet. 7*, 31–73.

Ewer, J., Reynolds, S.E., **2002**. Neuropeptide control of molting in insects. In: Pfaff, D.W., Arnold, A.P., Fahrbach, S.E., Etgen, A.M., Rubin, R.T. (Eds.), Hormones, Brain and Behavior, vol. 35. Academic Press, San Diego, pp. 1–92.

Ewer, J., Truman, J.W., **1996**. Increases in cyclic 3′,5′-guanosine monophosphate (cGMP) occur at ecdysis in an evolutionarily conserved crustacean cardioactive peptide-immunoreactive insect neuronal network. *J. Comp. Neurol. 370*, 330–341.

Fernlund, P., **1971**. Chromactivating hormones of *Pandalus borealis*, isolation and purification of a light-adapting hormone. *Biochim. Biophys. Acta 237*, 519–529.

Fernlund, P., **1976**. Structure of a light-adapting hormone from the shrimp, *Pandalus borealis*. *Biochim. Biophys. Acta 439*, 17–25.

Foster, R.G., Helfrich-Förster, C., **2001**. The regulation of circadian clocks by light in fruitflies and mice. *Philos. Trans. R. Soc. Lond. B Biol. Sci. 356*, 1779–1789.

Frisch, B., Fleissner, G., Brandes, C., Hall, J.C., **1994**. Staining in the brain of *Pachymorpha sexguttata* mediated by an antibody against a *Drosophila* clock-gene product, labeling of cells with possible importance for the beetle's circadian rhythms. *Cell Tiss. Res. 286*, 411–429.

Frisch, B., Hardin, P.E., Hamblen-Coyle, M.J., Rosbash, M., Hall, J.C., **1994**. A promoterless *period* gene mediates behavioral rhythmicity and cyclical per expression in a restricted subset of the *Drosophila* nervous system. *Neuron 12*, 555–570.

Froy, O., Chang, D.C., Reppert, S.M., **2002**. Redox potential, differential roles in CRY and mCRY1 functions. *Curr. Biol. 12*, 147–152.

Gekakis, N., Saez, L., Delahaye-Brown, A.-M., Myers, M.P., Sehgal, A., *et al.*, **1995**. Isolation of *timeless* by PER protein interaction: defective interaction between *timeless* protein and long-period mutant PER^{L}. *Science (Wash.) 270*, 811–815.

Gibbs, S.M., Truman, J.W., **1998**. Nitric oxide and cyclic GMP regulate retinal patterning in the optic lobe of *Drosophila*. *Neuron 20*, 83–93.

Giebultowicz, J.M., **1999**. Insect circadian clocks, is it all in their heads? *J. Insect Physiol. 45*, 791–800.

Giebultowicz, J.M., Stanewsky, R., Hall, J.C., Hege, D.M., **2000**. Transplanted *Drosophila* excretory tubules maintain circadian clock cycling out of phase with the host. *Curr. Biol. 10*, 107–110.

Glossop, N.R., Houl, J.H., Zheng, H., Ng, F.S., Dudek, S.M., *et al.*, **2003**. VRILLE feeds back to control circadian transcription of *Clock* in the *Drosophila* circadian oscillator. *Neuron 37*, 249–261.

Glossop, N.R., Lyons, L.C., Hardin, P.E., **1999**. Interlocked feedback loops within the *Drosophila* circadian oscillator. *Science (Wash.) 286*, 766–768.

Glossop, N.R., Shepherd, D., **1998**. *Disconnected* mutants show disruption to the central projections of proprioceptive neurons in *Drosophila melanogaster*. *J. Neurobiol. 36*, 337–347.

Grima, B., Lamouroux, A., Chelot, E., Papin, C., Limbourg-Bouchon, B., *et al.*, **2002**. The F-box protein slimb controls the levels of clock proteins period and timeless. *Nature (Lond.) 420*, 178–182.

Hall, J.C., **2000**. Cryptochromes, sensory reception, transduction and clock functions subserving circadian systems. *Curr. Opin. Neurobiol. 10*, 456–466.

Hall, J.C., **2003**. Genetics and molecular Biology of rhythms in *Drosophila* and other insects. *Adv. Genet. 48*, 1–280.

Hamblen-Coyle, M., Konopka, R.J., Zwiebel, L.J., Colot, H.V., Dowse, H.B., *et al.*, **1989**. A new mutation at the *period* locus of *Drosophila melanogaster* with some novel effects on circadian rhythms. *J. Neurogenet. 5*, 229–256.

Handler, A.M., Konopka, R.J., **1979**. Transplantation of a circadian pacemaker in *Drosophila*. *Nature (Lond.) 279*, 236–238.

Hao, H., Allen, D.L., Hardin, P.E., **1997**. A circadian enhancer mediates PER-dependent mRNA cycling in *Drosophila melanogaster*. *Mol. Cell. Biol. 17*, 3687–3693.

Hao, H., Glossop, N.R., Lyons, L., Qiu, J., Morrish, B., *et al.*, **1999**. The 69 bp circadian regulatory sequence (CRS) mediates *per*-like developmental, spatial, and circadian expression and behavioral rescue in *Drosophila*. *J. Neurosci. 19*, 987–994.

Hardin, P.E., Hall, J.C., Rosbash, M., **1990**. Feedback of the *Drosophila period* gene product on circadian cycling of its messenger RNA levels. *Nature (Lond.) 343*, 536–540.

Hardin, P.E., Hall, J.C., Rosbash, M., **1992a**. Circadian oscillations in *period* gene mRNA levels are transcriptionally regulated. *Proc. Natl Acad. Sci. USA 89*, 11711–11715.

Hardin, P.E., Hall, J.C., Rosbash, M., **1992b**. Behavioral and molecular analyses suggest that circadian output is disrupted by *disconnected* mutants in *D. melanogaster*. *EMBO J. 11*, 1–6.

Harmar, A.J., Marston, H.M., Shen, S., Spratt, C., West, K.M., *et al.*, **2002**. The VPAC 2 receptor is essential for circadian function in the mouse suprachiasmatic nuclei. *Cell 10*, 497–508.

Harrison, P.M., Kumar, A., Lang, N., Snyder, M., Gerstein, M., **2002**. A question of size: the eukaryotic proteome and the problems in defining it. *Nucleic Acids Res. 30*, 1083–1090.

Hege, D.M., Stanewsky, R., Hall, J.C., Giebultowicz, J.M., **1997**. Rhythmic expression of a PER-reporter in

the Malpighian tubules of decapitated *Drosophila*, evidence for a brain-independent circadian clock. *J. Biol. Rhyth. 12*, 300–308.

Helfrich-Förster, C., **1995**. The *period* clock gene is expressed in central nervous system neurons which also produce a neuropeptide that reveals the projections of circadian pacemaker cells within the brain of *Drosophila melanogaster. Proc. Natl Acad. Sci. USA 92*, 612–616.

Helfrich-Förster, C., **1997**. Development of pigment-dispersing hormone-immunoreactive neurons in the nervous system of *Drosophila melanogaster. J. Comp. Neurol. 380*, 335–354.

Helfrich-Förster, C., **1998**. Robust circadian rhythmicity of *Drosophila melanogaster* requires the presence of lateral neurons, a brain-behavioral study of *disconnected* mutants. *J. Comp. Physiol. (A) 182*, 435–453.

Helfrich-Förster, C., **2003**. The neuroarchitecture of the circadian clock in the brain of *Drosophila melanogaster. Microsc. Res. Tech. 62*, 94–102.

Helfrich-Förster, C., Edwards, T., Yasuyama, K., Wisotzki, B., Schneuwly, S., *et al.*, **2002**. The extraretinal eyelet of *Drosophila*, development, ultrastructure, and putative circadian function. *J. Neurosci. 22*, 9255–9266.

Helfrich-Förster, C., Homberg, U., **1993**. Pigment-dispersing hormone-immunoreactive neurons in the nervous system of wild-type *Drosophila melanogaster* and of several mutants with altered circadian rhythmicity. *J. Comp. Neurol. 337*, 177–190.

Helfrich-Förster, C., Stengl, M., Homberg, U., **1998**. Organization of the circadian system in insects. *Chronobiol. Int. 15*, 567–594.

Helfrich-Förster, C., Tauber, M., Park, J.H., Muhlig-Versen, M., Schneuwly, S., *et al.*, **2000**. Ectopic expression of the neuropeptide pigment-dispersing factor alters behavioral rhythms in *Drosophila melanogaster. J. Neurosci. 20*, 3339–3353.

Helfrich-Förster, C., Winter, C., Hofbauer, A., Hall, J.C., Stanewsky, R., **2001**. The circadian clock of fruit flies is blind after elimination of all known photoreceptors. *Neuron 30*, 249–261.

Hendricks, J.C., **2003**. Genetic Models in Applied Physiology, Invited Review, Sleeping flies don't lie, the use of *Drosophila melanogaster* to study sleep and circadian rhythms. *J. Appl. Physiol. 94*, 1660–1672.

Hendricks, J.C., Finn, S.M., Panckeri, K.A., Chavkin, J., Williams, J.A., *et al.*, **2000**. Rest in *Drosophila* is a sleep-like state. *Neuron 25*, 129–138.

Hendricks, J.C., Lu, S., Kume, K., Yin, J.C., Yang, Z., *et al.*, **2003**. Gender dimorphism in the role of *cycle* (*BMAL1*) in rest, rest regulation, and longevity in *Drosophila melanogaster. J. Biol. Rhyth. 18*, 12–25.

Hewes, R.S., Taghert, P.H., **2001**. Neuropeptides and neuropeptide receptors in the *Drosophila melanogaster* genome. *Genome Res. 11*, 1126–1142.

Hofbauer, A., Buchner, E., **1989**. Does *Drosophila* have seven eyes? *Naturwissen 76*, 335–336.

Hoffman, E.C., Reyes, H., Chu, F.F., Sander, F., Conley, L.H., *et al.*, **1991**. Cloning of a factor required for activity of the Ah (dioxin) receptor. *Science (Wash.) 252*, 954–958.

Homberg, U., Davis, N.T., Hildebrand, J.G., **1991**. Peptide-immunocytochemistry of neurosecretory cells in the brain and retrocerebral complex of the sphinx moth *Manduca sexta. J. Comp. Neurol. 303*, 35–52.

Huang, Z.J., Edery, I., Rosbash, M., **1993**. PAS is a dimerization domain common to *Drosophila* period and several transcription factors. *Nature (Lond.) 364*, 259–262.

Hunter-Ensor, M., Ousley, A., Sehgal, A., **1996**. Regulation of the *Drosophila* protein timeless suggests a mechanism for resetting the circadian clock by light. *Cell 84*, 677–685.

Inoue, S., Shimoda, M., Nishinokubi, I., Siomi, M.C., Okamura, M., *et al.*, **2002**. A role for the *Drosophila fragile X-related* gene in circadian output. *Curr. Biol. 12*, 1331–1335.

Ishikawa, T., Matsumoto, A., Kato, T., Jr., Togashi, S., Ryo, H., *et al.*, **1999**. CRY is a *Drosophila* photoreceptor protein implicated in light entrainment of circadian rhythm. *Genes Cells 4*, 57–65.

Ivanchenko, M., Stanewsky, R., Giebultowicz, J.M., **2001**. Circadian photoreception in *Drosophila*, functions of cryptochrome in peripheral and central clocks. *J. Biol. Rhyth. 16*, 205–215.

Jackson, F.R., Bargiello, T.A., Yun, S.H., Young, M.W., **1986**. Product of *per* locus of *Drosophila* shares homology with proteoglycans. *Nature (Lond.) 320*, 185–188.

Jackson, F.R., Schroeder, A.J., Roberts, M.A., McNeil, G.P., Kume, K., *et al.*, **2001**. Cellular and molecular mechanisms of circadian control in insects. *J. Insect Physiol. 47*, 822–842.

Jin, X., Shearman, L.P., Weaver, D.R., Zylka, M.J., de Vries, G.J., *et al.*, **1999**. A molecular mechanism regulating rhythmic output from the suprachiasmatic circadian clock. *Cell 96*, 57–68.

Johnson, E.C., Bohn, L.M., Barak, L.S., Birse, R.T., Nässel, D.R., *et al.*, **2003b**. Identification of *Drosophila* neuropeptide receptors by G protein-coupled receptors-beta-arrestin2 interactions. *J. Biol. Chem. 278*, 52172–52178.

Johnson, E.C., Garczynski, S.F., Park, D., Crim, J.W., Nässel, D.R., *et al.*, **2003a**. Identification and characterization of a G protein-coupled receptor for the neuropeptide proctolin in *Drosophila melanogaster. Proc. Natl Acad. Sci. USA 100*, 6198–6203.

Kaneko, M., Hall, J.C., **2000**. Neuroanatomy of cells expressing clock genes in *Drosophila*, transgenic manipulation of the *period* and *timeless* genes to mark the perikarya of circadian pacemaker neurons and their projections. *J. Comp. Neurol. 422*, 66–94.

Kaneko, M., Hamblen, M.J., Hall, J.C., **2000b**. Involvement of the *period* gene in developmental time-memory: effect of the per^{Short} mutation on phase shifts induced by light pulses delivered to *Drosophila* larvae. *J. Biol. Rhyth. 15*, 13–30.

Kaneko, M., Helfrich-Förster, C., Hall, J.C., **1997**. Spatial and temporal expression of the *period* and *timeless* genes in the developing nervous system of *Drosophila*, newly identified pacemaker candidates and novel features of clock gene product cycling. *J. Neurosci. 17*, 6745–6760.

Kaneko, M., Park, J.H., Cheng, Y., Hardin, P.E., Hall, J.C., **2000a**. Disruption of synaptic transmission or clock-gene-product oscillations in circadian pacemaker cells of *Drosophila* cause abnormal behavioral rhythms. *J. Neurobiol. 43*, 207–233.

Kim, E.Y., Bae, K., Ng, F.S., Glossop, N.R., Hardin, P.E., *et al.*, **2002**. *Drosophila* CLOCK protein is under posttranscriptional control and influences light-induced activity. *Neuron 34*, 69–81.

Kleinholz, L.H., **1975**. Purified hormones from the crustacean eyestalk and their physiological specificity. *Nature (Lond.) 258*, 256–257.

Kloss, B., Price, J.L., Saez, L., Blau, J., Rothenfluh, A., *et al.*, **1998**. The *Drosophila* clock gene *double-time* encodes a protein closely related to human casein kinase I epsilon. *Cell 94*, 97–107.

Kloss, B., Rothenfluh, A., Young, M.W., Saez, L., **2001**. Phosphorylation of period is influenced by cycling physical associations of double-time, period, and timeless in the *Drosophila* clock. *Neuron 30*, 699–706.

Ko, H.W., Jiang, J., Edery, I., **2002**. Role for Slimb in the degradation of *Drosophila* Period protein phosphorylated by Doubletime. *Nature (Lond.) 420*, 673–678.

Konopka, R.J., Benzer, S., **1971**. Clock mutants of *Drosophila melanogaster. Proc. Natl Acad. Sci. USA 68*, 2112–2116.

Konopka, R.J., Hamblen-Coyle, M.J., Jamison, C.F., Hall, J.C., **1994**. An ultrashort clock mutation at the *period* locus of *Drosophila melanogaster* that reveals some new features of the fly's circadian system. *J. Biol. Rhyth. 9*, 189–216.

Konopka, R.J., Smith, R.F., Orr, D., **1991**. Characterization of *Andante*, a new *Drosophila* clock mutant, and its interactions with other clock mutants. *J. Neurogenet. 7*, 103–114.

Krishnan, B., Dryer, S.E., Hardin, P.E., **1999**. Circadian rhythms in olfactory responses of *Drosophila melanogaster. Nature (Lond.) 400*, 375–378.

Krishnan, B., Levine, J.D., Lynch, M.K., Dowse, H.B., Funes, P., *et al.*, **2001**. A new role for *cryptochrome* in a *Drosophila* circadian oscillator. *Nature (Lond.) 411*, 313–317.

Leak, R.K., Moore, R.Y., **2000**. Topographic organization of suprachiasmatic nucleus projection neurons. *J. Comp. Neurol. 433*, 312–334.

Lee, C., Bae, K., Edery, I., **1998**. The *Drosophila* CLOCK protein undergoes daily rhythms in abundance, phosphorylation, and interactions with the PER-TIM complex. *Neuron 21*, 857–867.

Lee, C., Bae, K., Edery, I., **1999**. PER and TIM inhibit the DNA binding activity of a *Drosophila* CLOCK-CYC/dBMAL1 heterodimer without disrupting formation of the heterodimer, a basis for circadian transcription. *Mol. Cell. Biol. 19*, 5316–5325.

Lee, A., Li, W., Xu, K., Bogert, B.A., Su, K., *et al.*, **2003**. Control of dendritic development by the *Drosophila fragile X-related* gene involves the small GTPase Rac1. *Development 130*, 5543–5552.

Lee, C., Parikh, V., Itsukaichi, T., Bae, K., Edery, I., **1996**. Resetting the *Drosophila* clock by photic regulation of PER and a PER-TIM complex. *Science (Wash.) 271*, 1740–1744.

Leloup, J.C., Goldbeter, A., **1998**. A model for circadian rhythms in *Drosophila* incorporating the formation of a complex between the PER and TIM proteins. *J. Biol. Rhyth. 13*, 70–87.

Levine, J.D., Funes, P., Dowse, H.B., Hall, J.C., **2002a**. Advanced analysis of a *cryptochrome* mutation's effects on the robustness and phase of molecular cycles in isolated peripheral tissues of *Drosophila*. *BMC Neurosci. 3*, 5.

Levine, J.D., Funes, P., Dowse, H.B., Hall, J.C., **2002b**. Resetting the circadian clock by social experience in *Drosophila melanogaster. Science (Wash.) 298*, 2010–2012.

Levine, J.D., Funes, P., Dowse, H.B., Hall, J.C., **2002c**. Signal analysis of behavioral and molecular cycles. *BMC Neurosci. 3*, 1.

Levine, J.D., Sauman, I., Imbalzano, M., Reppert, S.M., Jackson, F.R., **1995**. Period protein from the giant silkmoth *Antheraea pernyi* functions as a circadian clock element in *Drosophila melanogaster. Neuron 15*, 147–157.

Lin, F.J., Song, W., Meyer-Bernstein, E., Naidoo, N., Sehgal, A., **2001**. Photic signaling by cryptochrome in the *Drosophila* circadian system. *Mol. Cell. Biol. 21*, 7287–7294.

Lin, J.M., Kilman, V.L., Keegan, K., Paddock, B., Emery-Le, M., *et al.*, **2002**. A role for casein kinase 2alpha in the *Drosophila* circadian clock. *Nature (Lond.) 420*, 816–820.

Lin, Y., Han, M., Shimada, B., Wang, L., Gibler, T.M., *et al.*, **2002**. Influence of the *period*-dependent circadian clock on diurnal, circadian, and aperiodic gene expression in *Drosophila melanogaster. Proc. Natl Acad. Sci. USA 99*, 9562–9567.

Liu, X., Lorenz, L., Yu, Q.N., Hall, J.C., Rosbash, M., **1988**. Spatial and temporal expression of the *period* gene in *Drosophila melanogaster. Genes Devel. 2*, 228–238.

Liu, X., Zwiebel, L.J., Hinton, D., Benzer, S., Hall, J.C., *et al.*, **1992**. The *period* gene encodes a predominantly nuclear protein in adult *Drosophila*. *J. Neurosci. 12*, 2735–2744.

Long, J.C., Jenkins, G.I., **1998**. Involvement of plasma membrane redox activity and calcium homeostasis in the UV-B and UV-A/blue light induction of gene expression in *Arabidopsis*. *Plant Cell 10*, 2077–2086.

Lowrey, P.L., Shimomura, K., Antoch, M.P., Yamazaki, S., Zemenides, P.D., *et al.*, **2000**. Positional syntenic cloning and functional characterization of the mammalian circadian mutation *tau*. *Science (Wash.) 288*, 483–492.

Lyons, L.C., Darlington, T.K., Hao, H., Houl, J., Kay, S.A., *et al.*, **2000**. Specific sequences outside the E-box are required for proper *per* expression and behavioral rescue. *J. Biol. Rhyth. 15*, 472–482.

Majercak, J., Kalderon, D., Edery, I., **1997**. *Drosophila melanogaster* deficient in protein kinase A manifests behavior-specific arrhythmia but normal clock function. *Mol. Cell. Biol. 17*, 5915–5922.

Malpel, S., Klarsfeld, A., Rouyer, F., **2002**. Larval optic nerve and adult extra-retinal photoreceptors sequentially associate with clock neurons during *Drosophila* brain development. *Development 129*, 1443–1453.

Marrus, S.B., Zeng, H., Rosbash, M., **1996**. Effect of constant light and circadian entrainment of per^s flies, evidence for light-mediated delay of the negative feedback loop in *Drosophila*. *Eur. Mol. Biol. Org. J. 15*, 6877–6886.

Martinek, S., Inonog, S., Manoukian, A.S., Young, M.W., **2001**. A role for the segment polarity gene *shaggy/GSK-3* in the *Drosophila* circadian clock. *Cell 105*, 769–779.

Matsumoto, A., Tomioka, K., Chiba, Y., Tanimura, T., **1999**. tim^{rit} lengthens circadian period in a temperature-dependent manner through suppression of PERIOD protein cycling and nuclear localization. *Mol. Cell. Biol. 19*, 4343–4354.

McDonald, M.J., Rosbash, M., **2001**. Microarray analysis and organization of circadian gene expression in *Drosophila*. *Cell 107*, 567–578.

McDonald, M.J., Rosbash, M., Emery, P., **2001**. Wild-type circadian rhythmicity is dependent on closely spaced E boxes in the *Drosophila timeless* promoter. *Mol. Cell. Biol. 21*, 1207–1217.

McNabb, S.L., Baker, J.D., Agapite, J., Steller, H., Riddiford, L.M., *et al.*, **1997**. Disruption of a behavioral sequence by targeted death of peptidergic neurons in *Drosophila*. *Neuron 19*, 813–823.

McNeil, G.P., Zhang, X., Genova, G., Jackson, F.R., **1998**. A molecular rhythm mediating circadian clock output in *Drosophila*. *Neuron 20*, 297–303.

Meeusen, T., Mertens, I., Clynen, E., Baggerman, G., Nichols, R., *et al.*, **2002**. Identification in *Drosophila melanogaster* of the invertebrate G protein-coupled FMRFamide receptor. *Proc. Natl Acad. Sci. USA 99*, 15363–15368.

Michel, S., Geusz, M.E., Zaritsky, J.J., Block, G.D., **1993**. Circadian rhythm in membrane conductance expressed in isolated neurons. *Science (Wash.) 259*, 239–241.

Morales, J., Hiesinger, P.R., Schroeder, A.J., Kume, K., Verstreken, P., *et al.*, **2002**. *Drosophila* fragile X protein, DFXR, regulates neuronal morphology and function in the brain. *Neuron 34*, 961–972.

Myers, M.P., Wager-Smith, K., Wesley, C.S., Young, M.W., Sehgal, A., **1995**. Positional cloning and sequence analysis of the *Drosophila* clock gene, *timeless*. *Science (Wash.) 270*, 805–808.

Myers, M.P., Wager-Smith, K., Rothenfluh-Hilfiker, A., Young, M.W., **1996**. Light-induced degradation of TIMELESS and entrainment of the *Drosophila* circadian clock. *Science (Wash.) 271*, 1736–1740.

Myers, E.M., Yu, J., Sehgal, A., **2003**. Circadian control of eclosion. Interaction between a central and peripheral clock in *Drosophila melanogaster*. *Curr. Biol. 13*, 526–533.

Naidoo, N., Song, W., Hunter-Ensor, M., Sehgal, A., **1999**. A role for the proteasome in the light response of the timeless clock protein. *Science (Wash.) 285*, 1737–1741.

Nässel, D.R., Shiga, S., Mohrherr, C.J., Rao, K.R., **1993**. Pigment-dispersing hormone-like peptide in the nervous system of the flies *Phormia* and *Drosophila*, immunocytochemistry and partial characterization. *J. Comp. Neurol. 331*, 183–198.

Nässel, D.R., Shiga, S., Wikstrand, E.M., Rao, K.R., **1991**. Pigment-dispersing hormone-immunoreactive neurons and their relation to serotonergic neurons in the blowfly and cockroach visual system. *Cell Tiss. Res. 266*, 511–523.

Nery, L.E., Castrucci, A.M., **1997**. Pigment cell signaling for physiological color change. *Comp. Biochem. Physiol. A 118*, 1135–1144.

Newby, L.M., Jackson, F.R., **1996**. Regulation of a specific circadian clock output pathway by lark, a putative RNA-binding protein with repressor activity. *J. Neurobiol. 31*, 117–128.

Nitabach, M.N., Blau, J., Holmes, T.C., **2002**. Electrical silencing of *Drosophila* pacemaker neurons stops the free-running circadian clock. *Cell 109*, 485–495.

Ohira, T., Nagasawa, H., Aida, K., **2002**. Molecular cloning of cDNAs encoding two pigment-dispersing hormones and two corresponding genes from the Kuruma Prawn (*Penaeus japonicus*). *Mar. Biotechnol. (NY) 4*, 463–470.

Park, H.W., Kim, S.T., Sancar, A., Deisenhofer, J., **1995**. Crystal structure of DNA photolyase from *Escherichia coli*. *Science (Wash.) 268*, 1866–1872.

Park, J., **2002**. Downloading central clock information in *Drosophila*. *Mol. Neurobiol. 26*, 217–233.

Park, J.H., Hall, J.C., **1998**. Isolation and chronobiological analysis of a neuropeptide *pigment-dispersing factor* gene in *Drosophila melanogaster*. *J. Biol. Rhyth. 13*, 219–228.

Park, J.H., Helfrich-Förster, C., Lee, G., Liu, L., Rosbash, M., *et al.*, **2000**. Differential regulation of circadian pacemaker output by separate clock genes in *Drosophila*. *Proc. Natl Acad. Sci. USA 97*, 3608–3613.

Park, J.H., Schroeder, A.J., Helfrich-Förster, C., Jackson, F.R., Ewer, J., **2003**. Targeted ablation of CCAP neuropeptide-containing neurons of *Drosophila* causes specific defects in execution and circadian timing of ecdysis behavior. *Development 130*, 2645–2656.

Park, S.K., Sedore, S.A., Cronmiller, C., Hirsh, J., **2000**. Type II cAMP-dependent protein kinase-deficient *Drosophila* are viable but show developmental, circadian, and drug response phenotypes. *J. Biol. Chem. 275*, 20588–20596.

Park, Y., Kim, Y.J., Adams, M.E., **2002**. Identification of G protein-coupled receptors for *Drosophila* PRXamide peptides, CCAP, corazonin, and AKH supports a theory of ligand-receptor coevolution. *Proc. Natl Acad. Sci. USA 99*, 11423–11428.

Pearn, M.T., Randall, L.L., Shortridge, R.D., Burg, M.G., Pak, W.L., **1995**. Molecular, biochemical, and electrophysiological characterization of *Drosophila norpA* mutants. *J. Biol. Chem. 271*, 4937–4945.

Pearn, M.T., Randall, L.L., Shortridge, R.D., Burg, M.G., Pak, W.L., **1996**. Molecular, biochemical, and electrophysiological characterization of *Drosophila norpA* mutants. *J. Biol. Chem. 271*, 4937–4945.

Peifer, M., Pai, L.M., Casey, M., **1994**. Phosphorylation of the *Drosophila* adherens junction protein Armadillo: roles for wingless signal and zeste-white 3 kinase. *Dev. Biol. 16*, 543–556.

Peng, Y., Stoleru, D., Levine, J.D., Hall, J.C., Rosbash, M., **2003**. *Drosophila* free-running rhythms require intercellular communication. *PLoS Biological 1*, E13. Epub 2003 Sep 15.

Persson, M.G., Eklund, M.B., Dircksen, H., Muren, J.E., Nässel, D.R., **2001**. Pigment-dispersing factor in the locust abdominal ganglia may have roles as circulating neurohormone and central neuromodulator. *J. Neurobiol. 48*, 19–41.

Petri, B., Homberg, U., Loesel, R., Stengl, M., **2002**. Evidence for a role of GABA and Mas-allatotropin in photic entrainment of the circadian clock of the cockroach *Leucophaea maderae*. *J. Exp. Biol. 205*, 1459–1469.

Petri, B., Stengl, M., **1997**. Pigment-dispersing hormone shifts the phase of the circadian pacemaker of the cockroach *Leucophaea maderae*. *J. Neurosci. 17*, 4087–4093.

Pittendrigh, C.S., **1974**. Circadian oscillations in cells and circadian organization of multicellular systems. In: Schmitt, F.O., Worden, F.G. (Eds.), The Neurosciences Third Study Program. MIT Press, Cambridge, MA, pp. 437–458.

Plautz, J.D., Kaneko, M., Hall, J.C., Kay, S.A., **1997**. Independent photoreceptive circadian clocks throughout *Drosophila*. *Science (Wash.) 278*, 1632–1635.

Pollack, I., Hofbauer, A., **1991**. Histamine-like immunoreactivity in the visual system and brain of *Drosophila melanogaster. Cell Tiss. Res. 266*, 391–398.

Price, J.L., Blau, J., Rothenfluh, A., Abodeely, M., Kloss, B., *et al.*, **1998**. *double-time* is a novel *Drosophila* clock gene that regulates PERIOD protein accumulation. *Cell 94*, 83–95.

Price, J.L., Dembinska, M.E., Young, M.W., Rosbash, M., **1995**. Suppression of PERIOD protein abundance and circadian cycling by the *Drosophila* clock mutation *timeless*. *Eur. Mol. Biol. Org. J. 14*, 4044–4049.

Pyza, E., Meinertzhagen, I.A., **1996**. Neurotransmitters regulate rhythmic size changes amongst cells in the fly's optic lobe. *J. Comp. Physiol. A 178*, 33–45.

Pyza, E., Meinertzhagen, I.A., **1997**. Neurites of period-expressing PDH cells in the fly's optic lobe exhibit circadian oscillations in morphology. *Eur. J. Neurosci. 9*, 1784–1788.

Pyza, E., Meinertzhagen, I.A., **1999**. Daily rhythmic changes of cell size and shape in the first optic neuropil in *Drosophila melanogaster. J. Neurobiol. 40*, 77–88.

Rachidi, M., Lopes, C., Benichou, J.-C., Rouyer, F., **1997**. Analysis of *period* circadian expression in the *Drosophila* head by *in situ* hybridization. *J. Neurogenet. 11*, 255–263.

Rao, K.R., Mohrherr, C.J., Riehm, J.P., Zahnow, C.A., Norton, S., *et al.*, **1987**. Primary structure of an analog of crustacean pigment-dispersing hormone from the lubber grasshopper *Romalea microptera*. *J. Biol. Chem. 262*, 2672–2675.

Rao, K.R., Riehm, J.P., Zahnow, C.A., Kleinholz, L.H., Tarr, G.E., *et al.*, **1985**. Characterization of a Pigment-Dispersing Hormone is eyestalks of the Fiddler crab *Uca pugilator. Proc. Natl Acad. Sci. USA 82*, 5319–5322.

Renn, S.C., Park, J.H., Rosbash, M., Hall, J.C., Taghert, P.H., **1999**. A *pdf* neuropeptide gene mutation and ablation of PDF neurons each cause severe abnormalities of behavioral circadian rhythms in *Drosophila*. *Cell 99*, 791–802.

Rensing, L., Ruoff, P., **2002**. Temperature effect on entrainment, phase shifting, and amplitude of circadian clocks and its molecular bases. *Chronobiol. Int. 19*, 807–864.

Reppert, S.M., Tsai, T., Roca, A.L., Sauman, I., **1994**. Cloning of a structural and functional homolog of the circadian clock gene *period* from the giant silkmoth *Antheraea pernyi*. *Neuron 13*, 1167–1176.

Reyes, H., Reisz-Porszasz, S., Hankinson, O., **1992**. Identification of the Ah receptor nuclear translocator protein (Arnt) as a component of the DNA binding form of the Ah receptor. *Science (Wash.) 256*, 1193–1195.

Riehle, M.A., Garczynski, S.F., Crim, J.W., Hill, C.A., Brown, M.R., **2002**. Neuropeptides and peptide hormones in *Anopheles gambiae*. *Science (Wash.) 298*, 172–175.

Riehm, J.P., Rao, K.R., **1982**. Structure-activity relationships of a pigment-dispersing crustacean neurohormone. *Peptides 3*, 643–647.

Riehm, J.P., Rao, K.R., Semmes, O.J., Jorenby, W.H., Hintz, M.F., *et al.*, **1985**. C-terminal deletion analogs of a crustacean pigment-dispersing hormone. *Peptides 6*, 1051–1056.

Rosato, E., Codd, V., Mazzotta, G., Piccin, A., Zordan, M., *et al.*, **2001**. Light-dependent interaction between *Drosophila* CRY and the clock protein PER mediated by the carboxy terminus of CRY. *Curr. Biol. 11*, 909–917.

Rothenfluh, A., Abodeely, M., Young, M.W., **2000a**. Isolation and analysis of six *timeless* alleles that cause short- or long-period circadian rhythms in *Drosophila*. *Genetics 156*, 665–675.

Rothenfluh, A., Abodeely, M., Young, M.W., **2000b**. Short-period mutations of *per* affect a *double-time-*

dependent step in the *Drosophila* circadian clock. *Curr. Biol. 10*, 1399–1402.

Rothenfluh, A., Young, M.W., Saez, L., **2000c**. A TIMELESS-independent function for PERIOD proteins in the *Drosophila* clock. *Neuron 26*, 505–514.

Rouyer, F., Rachidi, M., Pikielny, C., Rosbash, M., **1997**. A new gene encoding a putative transcription factor regulated by the *Drosophila* circadian clock. *Eur. Mol. Biol. Org. J. 16*, 3944–3954.

Rutila, J.E., Suri, V., Le, M., So, W.V., Rosbash, M., *et al.*, **1996**. The *timSL* mutant of the *Drosophila* rhythm gene *timeless* manifests allele-specific interactions with *period* gene mutants. *Neuron 17*, 921–929.

Rutila, J.E., Suri, V., Le, M., So, W.V., Rosbash, M., *et al.*, **1998**. CYCLE is a second bHLH-PAS clock protein essential for circadian rhythmicity and transcription of *Drosophila period* and *timeless*. *Cell 93*, 805–814.

Saez, L., Young, M.W., **1988**. *In situ* localization of the per clock protein during development of *Drosophila melanogaster. Mol. Cell. Biol. 8*, 5378–5385.

Saez, L., Young, M.W., **1996**. Regulation of nuclear entry of the *Drosophila* clock proteins period and timeless. *Neuron 17*, 911–920.

Sakanaka, C., Leong, P., Xu, L., Harrison, S.D., Williams, L.T., **1999**. Casein kinase iepsilon in the *wnt* pathway: regulation of beta-catenin function. *Proc. Natl Acad. Sci. USA 96*, 12548–12552.

Sancar, A., **2000**. CRYPTOCHROME, The second photoactive pigment in the eye and its role in circadian photoreception. *Annu. Rev. Biochem. 69*, 31–67.

Sarov-Blat, L., So, W.V., Liu, L., Rosbash, M., **2000**. The *Drosophila takeout* gene is a novel molecular link between circadian rhythms and feeding behavior. *Cell 101*, 647–656.

Sauman, I., Reppert, S.M., **1996**. Circadian clock neurons in the silkmoth *Antheraea pernyi*, novel mechanisms of Period protein regulation. *Neuron 17*, 889–900.

Schenck, A., Bardoni, B., Langmann, C., Harden, N., Mandel, J.L., *et al.*, **2003**. CYFIP/Sra-1 controls neuronal connectivity in *Drosophila* and links the Rac1 GTPase pathway to the fragile X protein. *Neuron 38*, 887–898.

Sehgal, A., Price, J.L., Man, B., Young, M.W., **1994**. Loss of circadian behavioral rhythms and *per* RNA oscillations in the *Drosophila* mutant *timeless*. *Science (Wash.) 263*, 1603–1606.

Sehgal, A., Rothenfluh-Hilfiker, A., Hunter-Ensor, M., Chen, Y., Myers, M.P., *et al.*, **1995**. Rhythmic expression of *timeless*, a basis for promoting circadian cycles in *period* gene autoregulation. *Science (Wash.) 270*, 808–810.

Shafer, O.T., **2001**. Blind clocks reveal elusive light input pathway in *Drosophila*. *Trends Neurosci. 24*, 627–628.

Shafer, O.T., Rosbash, M., Truman, J.W., **2002**. Sequential nuclear accumulation of the clock proteins period and timeless in the pacemaker neurons of *Drosophila melanogaster. J. Neurosci. 22*, 5946–5954.

Shaw, P., **2003**. Awakening to the behavioral analysis of sleep in *Drosophila*. *J. Biol. Rhyth. 18*, 4–11.

Shaw, P.J., Cirelli, C., Greenspan, R.J., Tononi, G., **2000**. Correlates of sleep and waking in *Drosophila melanogaster. Science (Wash.) 287*, 1834–1837.

Shaw, P.J., Tononi, G., Greenspan, R.J., Robinson, D.F., **2002**. Stress response genes protect against lethal effects of sleep deprivation in *Drosophila*. *Nature (Lond.) 417*, 287–291.

Siegfried, E., Wilder, E.L., Perrimon, N., **1994**. Components of wingless signalling in *Drosophila*. *Nature (Lond.) 367*, 76–80.

Siwicki, K.K., Eastman, C., Petersen, G., Rosbash, M., Hall, J.C., **1988**. Antibodies to the period gene product of *Drosophila* reveal diverse tissue distribution and rhythmic changes in the visual system. *Neuron 1*, 141–150.

Smith, R.F., Konopka, R.J., **1981**. Circadian clock phenotypes of chromosome aberrations with a breakpoint at the *per* locus. *Mol. Gen. Genet. 183*, 243–251.

So, W.V., Rosbash, M., **1997**. Post-transcriptional regulation contributes to *Drosophila clock* gene mRNA cycling. *EMBO J. 16*, 7146–7155.

So, W.V., Sarov-Blat, L., Kotarski, C.K., McDonald, M.J., Allada, R., *et al.*, **2000**. *takeout*, a novel *Drosophila* gene under circadian clock transcriptional regulation. *Mol. Cell. Biol. 20*, 6935–6944.

Stanewsky, R., **2003**. Genetic analysis of the circadian system in *Drosophila melanogaster* and mammals. *J. Neurobiol. 54*, 111–147.

Stanewsky, R., Jamison, C.F., Plautz, J.D., Kay, S.A., Hall, J.C., **1997**. Multiple circadian-regulated elements contribute to cycling *period* gene expression in *Drosophila*. *EMBO J. 16*, 5006–5018.

Stanewsky, R., Kaneko, M., Emery, P., Beretta, B., Wager-Smith, K., *et al.*, **1998**. The *cryb* mutation identifies cryptochrome as a circadian photoreceptor in *Drosophila*. *Cell 95*, 681–692.

Steller, H., Fischbach, K.F., Rubin, G., **1987**. *disconnected*, a locus required for neuronal pathway formation in the visual system of *Drosophila*. *Cell 50*, 1139–1153.

Stempfl, T., Vogel, M., Szabo, G., Wulbeck, C., Liu, J., *et al.*, **2002**. Identification of circadian-clock-regulated enhancers and genes of *Drosophila melanogaster* by transposon mobilization and luciferase reporting of cyclical gene expression. *Genetics 160*, 571–593.

Suri, V., Hall, J.C., Rosbash, M., **2001**. Two novel *doubletime* mutants alter circadian properties and eliminate the delay between RNA and protein in *Drosophila*. *J. Neurosci. 20*, 7547–7555.

Suri, V., Lanjuin, A., Rosbash, M., **1999**. TIMELESS-dependent positive and negative autoregulation in the *Drosophila* circadian clock. *EMBO J. 18*, 675–686.

Sweeney, S.T., Broadie, K., Keane, J., Niemann, H., O'Kane, C.J., **1995**. Targeted expression of tetanus toxin lighr chain in *Drosophila* specifically eliminates synaptic transmission and causes behavioral defects. *Neuron 14*, 341–351.

Taghert, P.H., **2001**. How does the circadian clock send timing information to the brain? *Sem. Cell. Devel. Biol. 12*, 329–341.

Taghert, P.H., Hewes, R.S., Park, J.H., O'Brien, M.A., Han, M., *et al.*, **2001**. Multiple amidated neuropeptides are required for normal circadian locomotor rhythms in *Drosophila*. *J. Neurosci. 21*, 6673–6686.

Taghert, P.H., Roberts, M.E., Renn, S.C., Jacobs, P.S., **2000**. Metamorphosis of tangential visual system neurons in *Drosophila*. *Devel. Biol. 222*, 471–485.

Taghert, P.H., Veenstra, J.A., **2003**. Neuropeptide signaling in *Drosophila*. *Adv. Genet. 49*, 1–65.

Tamada, T., Kitadokoro, K., Higuchi, Y., Inaka, K., Yasui, A., *et al.*, **1997**. Crystal structure of DNA photolyase from *Anacystis nidulans*. *Nat. Struct. Biol. 4*, 887–891.

The, I., Hannigan, G.E., Cowley, G.S., Reginald, S., Zhong, Z., *et al.*, **1997**. Rescue of a *Drosophila* NF1 mutant phenotype by protein kinase A. *Science (Wash.) 276*, 791–794.

Thor, S., Thomas, J.B., **1997**. The *Drosophila islet* gene governs axon pathfinding and neurotransmitter identity. *Neuron 18*, 397–409.

Todo, T., **1999**. Functional diversity of the DNA photolyase/blue light receptor family. *Mutat. Res. 434*, 89–97.

Toma, D.P., White, K.P., Hirsch, J., Greenspan, R.J., **2002**. Identification of genes involved in *Drosophila melanogaster* geotaxis, a complex behavioral trait. *Nature (Lond.) Genetics 31*, 349–353.

Truman, J.W., **1976**. Extraretinal photoreception in insects. *Photophysiology 23*, 215–225.

Truman, J.W., **1992**. The eclosion hormone system of insects. *Prog. Brain Res. 92*, 361–374.

Truman, J.W., Riddiford, L.M., **1970**. Neuroendocrine control of ecdysis in silkmoths. *Science (Wash.) 167*, 1624–1626.

Ueda, H.R., Matsumoto, A., Kawamura, M., Iino, M., Tanimura, T., *et al.*, **2002**. Genome-wide transcriptional orchestration of circadian rhythms in *Drosophila*. *J. Biol. Chem. 277*, 14048–14052.

Van Gelder, R.N., **2002**. Tales from the crypt(ochromes). *J. Biol. Rhyth. 17*, 110–120.

Van Gelder, R.N., Herzog, E.D., Schwartz, W.J., Taghert, P.H., **2003**. Circadian rhythms, in the loop at last. *Science (Wash.) 300*, 1534–1535.

Van Gelder, R.N., von Zastrow, M.E., Yool, A., Dement, W.C., Barchas, J.D., *et al.*, **1995**. Extent and character of circadian gene expression in *Drosophila melanogaster*, identification of twenty oscillating mRNAs in the fly head. *Curr. Biol. 5*, 1424–1436.

Veleri, S., Brandes, C., Helfrich-Förster, C., Hall, J.C., Stanewsky, R., **2003**. A self-sustaining, light-entrainable circadian oscillator in the *Drosophila* brain. *Curr. Biol. 13*, 1758–1767.

Volkner, M., Lenz-Bohme, B., Betz, H., Schmitt, B., **2000**. Novel CNS glutamate receptor subunit genes of *Drosophila melanogaster*. *J. Neurochem. 75*, 1791–1799.

Vosshall, L.B., Price, J.L., Sehgal, A., Saez, L., Young, M.W., **1994**. Block in nuclear localization of period protein by a second clock mutation, *timeless*. *Science (Wash.) 263*, 1606–1609.

Vosshall, L.B., Young, M.W., **1995**. Circadian rhythms in *Drosophila* can be driven by *period* expression in a restricted group of central brain cells. *Neuron 15*, 345–360.

Wegener, C., Hamasaka, Y., Nässel, D.R., **2004**. Acetylcholine increases intracellular Ca^{2+} via nicotinic receptors in cultured PDF-containing clock neurons of *Drosophila*. *J. Neurophysiol. 91*, 912–923.

Wheeler, D.A., Hamblen-Coyle, M.J., Dushay, M.S., Hall, J.C., **1993**. Behavior in light–dark cycles of *Drosophila* mutants that are arrhythmic, blind, or both. *J. Biol. Rhythms 8*, 67–94.

Williams, J.A., Su, H.S., Bernards, A., Field, J., Sehgal, A., **2001**. A circadian output in *Drosophila* mediated by neurofibromatosis-1 and Ras/MAPK. *Science (Wash.) 293*, 2251–2256.

Wise, S., Davis, N.T., Tyndale, E., Noveral, J., Folwell, M.G., *et al.*, **2002**. Neuroanatomical studies of *period* gene expression in the hawkmoth, *Manduca sexta*. *J. Comp. Neurol. 447*, 366–380.

Yang, Z., Emerson, M., Su, H.S., Sehgal, A., **1998**. Response of the timeless protein to light correlates with behavioral entrainment and suggests a nonvisual pathway for circadian photoreception. *Neuron 21*, 215–223.

Yang, Z., Sehgal, A., **2001**. Role of molecular oscillations in generating behavioral rhythms in *Drosophila*. *Neuron 29*, 453–467.

Yasuyama, K., Meinertzhagen, I.A., **1999**. Extraretinal photoreceptors at the compound eye's posterior margin in *Drosophila melanogaster*. *J. Comp. Neurol. 412*, 193–202.

Young, M.W., Kay, S.A., **2001**. Time zones: a comparative genetics of circadian clocks. *Nat. Rev. Genet. 2*, 702–715.

Yu, Q., Jacquier, A.C., Citri, Y., Hamblen, M., Hall, J.C., *et al.*, **1987**. Molecular mapping of point mutations in the *period* gene that stop or speed up biological clocks in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA 84*, 784–788.

Zavodska, R., Sauman, I., Sehnal, F., **2003**. Distribution of PER protein, pigment-dispersing hormone, prothoracicotropic hormone, and eclosion hormone in the cephalic nervous system of insects. *J. Biol. Rhyth. 18*, 106–122.

Zeng, H., Hardin, P.E., Rosbash, M., **1994**. Constitutive overexpression of the *Drosophila* period protein inhibits period mRNA cycling. *Eur. Mol. Biol. Org. J. 13*, 3590–3598.

Zeng, H., Qian, Z., Myers, M.P., Rosbash, M., **1996**. A light-entrainment mechanism for the *Drosophila* circadian clock. *Nature (Lond.) 380*, 129–135.

Zerr, D.M., Hall, J.C., Rosbash, M., Siwicki, K.K., **1990**. Circadian rhythms of period protein immunoreactivity in the CNS and the visual system of *Drosophila*. *J. Neurosci. 10*, 2749–2762.

Zhang, X., McNeil, G.P., Hilderbrand-Chae, M.J., Franklin, T.M., Schroeder, A.J., *et al.*, **2000**. Circadian regulation of the lark RNA-binding protein within identifiable neurosecretory cells. *J. Neurobiol. 45*, 14–29.

Zhang, Y.Q., Bailey, A.M., Matthies, H.J., Renden, R.B., Smith, M.A., *et al.*, **2001**. *Drosophila fragile X-related* gene regulates the MAP1B homolog Futsch to control synaptic structure and function. *Cell 107*, 591–603.

Zhao, J., Kilman, V.L., Keegan, K.P., Peng, Y., Emery, P., *et al.*, **2003**. *Drosophila* clock can generate ectopic circadian clocks. *Cell 113*, 755–766.

Zilian, O., Frei, E., Burke, R., Brentrup, D., Gutjahr, T., *et al.*, **1999**. *double-time* is identical to *discs overgrown*, which is required for cell survival, proliferation and growth arrest in *Drosophila* imaginal discs. *Development 126*, 5409–5420.

4.12 Insect Transposable Elements

Z Tu, Virginia Tech, Blacksburg, VA, USA

© 2005, Elsevier BV. All Rights Reserved.

4.12.1. Introduction

More than half a century ago, Barbara McClintock's observation of unstable mutations in maize led to the discovery of two mobile genetic elements, *Activator* (*Ac*) and *Dissociator* (*Ds*) (McClintock, 1948, 1950). The discovery of these mobile segments of DNA, later named as transposable elements (TEs), set forth the revolutionary concept of a fluid and dynamic genome. Five decades later as biology is entering the genomic era, the tremendous diversity of TEs and their potential impact are just being appreciated.

Being mobile, TEs have the ability to replicate and spread in the genome as primarily "selfish" genetic units (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). They tend to occupy significant portions of the eukaryotic genome. For example, at least 46% of the human genome (Lander *et al.*, 2001) and 16% of the euchromatic portion of the newly reported malaria mosquito genome (Holt *et al.*, 2002) are TE-derived sequences. The relative abundance and diversity of TEs have contributed to the differences in the structure and size of eukaryotic genomes (Kidwell, 2002). Recent evidence suggests

that the "selfish" property may have enabled TEs to provide the genome with potent agents to generate genetic and genomic plasticity (Kidwell and Lisch, 2000). For example, TEs may have reshaped the human genome by ectopic rearrangements, by creating new genes, and by modifying and shuffling existing genes (Lander *et al.*, 2001). In some cases, TEs had been co-opted to perform critical functions in the biology of their host. One well-documented example is the generation of the extensive array of immunoglobulins and T cell receptors by V(D)J recombination, which is believed to have evolved from an ancient transposition system (Gellert, 2002). Another example is the maintenance of telomeric structures in *Drosophila melanogaster* by site-specific insertions of two TEs (Pardue and DeBaryshe, 2002). An emerging view is that genomes are dynamic systems, within which diverse TEs evolve. The "selfish" TEs could evolve a wide spectrum of relationships with their hosts, ranging from "junk parasites" to "molecular symbionts" (Brookfield, 1995; Kidwell and Lisch, 2000).

The intricate dynamic between TEs and their host genomes is further complicated by the fact that some TEs are capable of crossing species barriers to spread in a new genome. Such a process is referred to as horizontal (or lateral) transfer, which is distinct from the vertical transmission of genetic material from ancestral species/organisms to their offspring. Horizontal transfer may be an important part of the life cycle of some TEs and it may contribute to their continued success during evolution (Silva *et al.*, 2004).

From an applied perspective, TEs have been used as tools to genetically manipulate cells/organisms, taking advantage of their ability to integrate cognate DNA in the genome. A well-known example is the transformation system derived from the *D. melanogaster P* transposable element, which has been instrumental to our understanding of this model genetic organism by providing transformation and mutagenesis tools (see **Chapter 4.13**). In addition, some TEs have been used as genetic markers for mapping and population studies, taking advantage of their dimorphic insertion states (presence and absence of an insertion) and their interspersed distribution in the genome. For example, the human *Alu* elements have been shown to be useful population genetic markers (Batzer *et al.*, 1994; Batzer and Deininger, 2002; Salem *et al.*, 2003). The presence and absence of insertions of short interspersed TEs at different genomic loci have also been used as molecular systematics markers to trace the explosive speciation of the cichlid fishes and other vertebrates (Shedlock and Okada, 2000; Terai *et al.*, 2003).

The focus of this chapter is on recent advances in the study of insect TEs. A brief introduction on TE classification and transposition mechanisms will be provided, followed by sections that describe the current approaches to study insect TEs and sections that highlight the diversity and evolutionary dynamics of TEs in insect genomes. Applications of TEs in genetic and molecular analysis of insects will be discussed in the end. Readers may consult recent reviews (Silva *et al.*, 2004) (see **Chapter 4.13**) and the second edition of a book on mobile DNA (Craig *et al.*, 2002) for details on related topics.

4.12.2. Classification and Transposition Mechanisms of Eukaryotic TEs

TEs can be categorized as Class I RNA-mediated or Class II DNA-mediated elements according to their transposition mechanisms (Finnegan, 1992). The transposition of RNA-mediated TEs involves a reverse transcription step, which generates cDNA from RNA molecules (Eickbush and Malik, 2002). The cDNA molecules are integrated in the genome, allowing replicative amplification. The transposition of DNA-mediated elements is directly from DNA to DNA and does not involve an RNA intermediate (Craig, 2002). In most cases, both classes of TEs will create target site duplication (TSD) upon their insertion in the genome (**Figure 1**). Both classes can be further categorized into different groups and all groups of TEs discussed here have been found in various species of insects. There have been several recent reviews on different groups of TEs in both classes (Deininger and Roy-Engel, 2002; Eickbush and Malik, 2002; Feschotte *et al.*, 2002; Robertson, 2002).

4.12.2.1. Class I RNA-Mediated TEs

RNA-mediated TEs include long terminal repeat (LTR) retrotransposons, non-LTR retrotransposons, and short interspersed repetitive/nuclear elements (SINEs). Non-LTR retrotransposons are also referred to as retroposons or long interspersed repetitive/nuclear elements (LINEs). The structural features of the three groups of RNA-mediated TEs are illustrated in **Figure 2** using representatives from different insects. All RNA-mediated TEs produce RNA transcripts that are reverse transcribed into cDNA to be integrated in the genome (Eickbush and Malik, 2002). Detailed mechanisms used by LTR and non-LTR retrotransposons are elegantly described in two recent reviews (Eickbush, 2002; Voytas and Boeke, 2002).

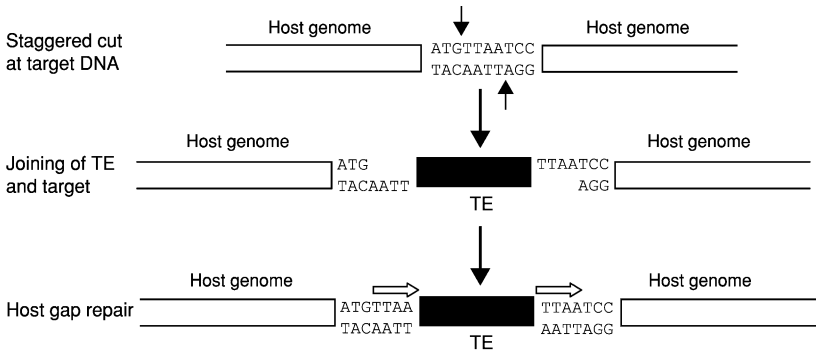


Figure 1 Mechanism of generating target site duplication (TSD). The TSDs, which are marked by open arrows, are not part of the TE sequence. They are target sequences duplicated upon a TE insertion. Most TEs create TSDs although the *Helitron* transposons and some non-LTR retrotransposons do not. See recent reviews (Kapitonov and Jurka, 2001; Craig, 2002; Eickbush and Malik, 2002) for illustrations of the transposition mechanisms of different types of TEs.

4.12.2.1.1. LTR retrotransposons LTR retrotransposons transpose through a mechanism much like that used by retroviruses. The LTRs in the LTR retrotransposons are generally 200 to 500 bp long and are involved in all aspects of their life cycle that includes providing promoter sequences and transcription termination signals (Eickbush and Malik, 2002). As shown in **Figure 2**, LTR retrotransposons encode a *pol*(polymerase)-like protein that contains reverse transcriptase (RT), ribonuclease H (RNase H), protease (PR), and integrase (IN) domains that are important for their retrotransposition. The RT domain performs the key function of reverse transcription and its sequence has been used for phylogenetic classification of LTR retrotransposons into four clades including Ty1/*copia*, Ty3/*gypsy*, BEL, and DIRS (Eickbush and Malik, 2002). The IN domain is responsible for inserting the cDNA copy into the host genome. In addition to the *pol*-like protein, LTR retrotransposons encode an additional protein related to the retroviral *gag* (group-associated antigene or group-specific antigen) protein that binds nucleic acid or forms the nucleocapsid shell. Some LTR retrotransposons also have an *env*(envelope)-like fragment that encodes a transmembrane receptor-binding protein that allows the transmission of retroviruses. Some of the LTR retrotransposons that encode an *env* protein are in fact retroviruses (Eickbush and Malik, 2002).

4.12.2.1.2. Non-LTR retrotransposons Non-LTR retrotransposons, or LINEs, or retroposons are generally 3 to 8 kb long and have been found in virtually all eukaryotes studied. Like the LTR retrotransposons, most non-LTR retrotransposons also have a *pol*-like protein that includes a RT domain that is essential for their retrotransposition. The RT domain has been used for phylogenetic classification of non-LTR retrotransposons into 17 clades, most of which probably date back to the Precambrian era, approximately 600 million years ago (Malik *et al.*, 1999; Eickbush and Malik, 2002; Biedler and Tu, 2003). Some elements also have an RNase H and/or AP endonuclease (APE) domain encoded in the *pol*-like open reading frame (ORF). In addition to the *pol*-like protein, many non-LTR retrotransposons encode a protein that is related to the retroviral *gag* protein. Studies of a *gag*-like protein from L1 retrotransposon in mice show that it acts as a nucleic acid chaperone (Martin and Bushman, 2001). Other typical structural characteristics found in various non-LTR retrotransposons are internal Pol II promoters and 3′ ends containing AATAAA polyadenylation signals, poly(A) tails, or simple tandem repeats. Target primed reverse transcription has been proposed as the mechanism of retrotransposition for R2 of *Bombyx mori* and may be generally true for all non-LTR elements (Luan *et al.*, 1993; Eickbush, 2002). Because they transpose by target primed reverse transcription, some non-LTR retrotransposons could rely rather heavily on host DNA repair mechanisms, and this relationship with the host may give non-LTR retrotransposons some flexibility with regard to the domains required in an autonomous element (Eickbush and Malik, 2002). Some non-LTR retrotransposons such as R2 are site-specific because their endonucleases make precise cleavage at specific targets (Xiong and Eickbush, 1988; Eickbush, 2002).

4.12.2.1.3. SINEs SINEs are generally between 100 and 500 bp long and they do not have any coding potential. SINEs may have been borrowing the

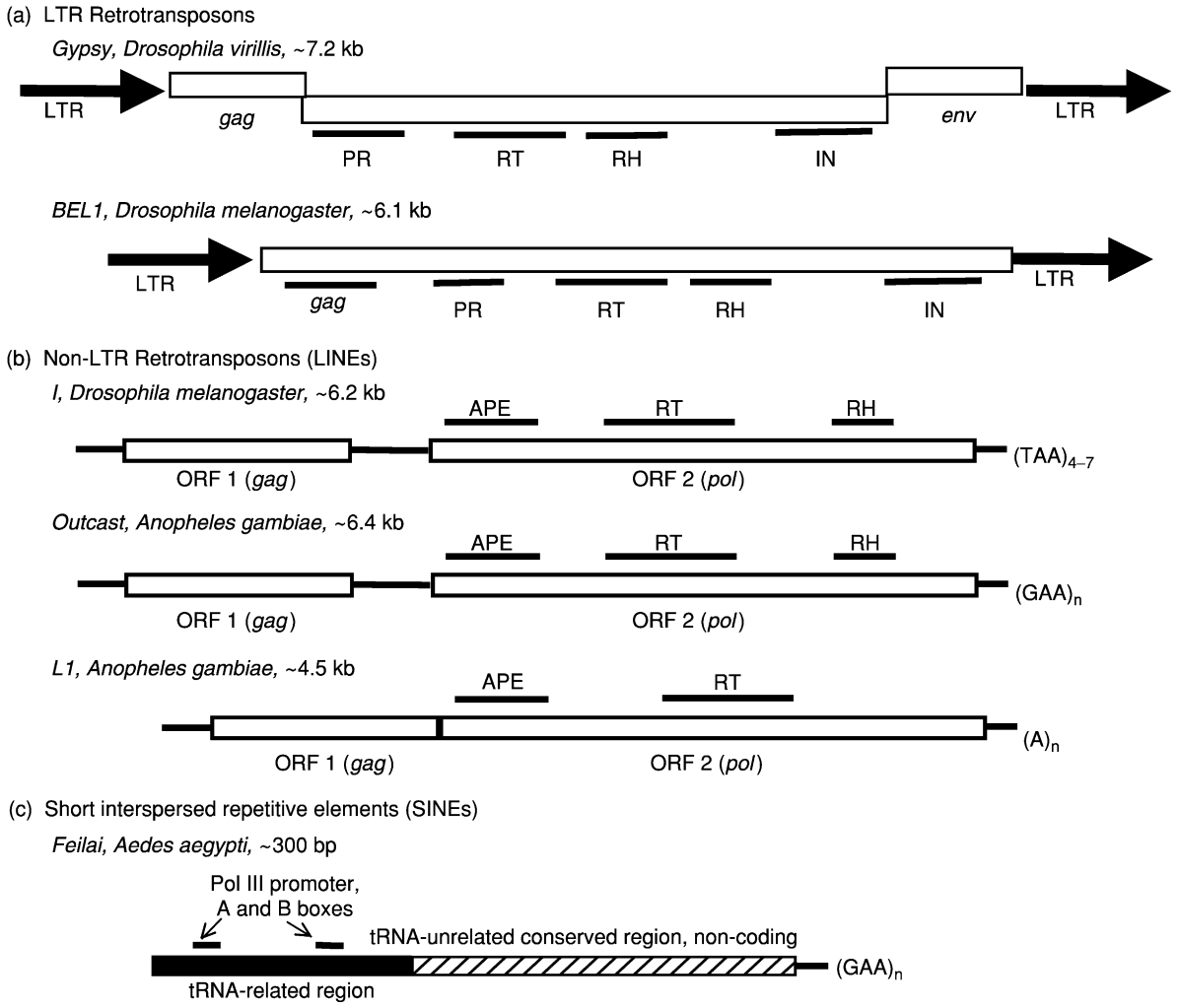


Figure 2 Structural characteristics of representative Class I RNA-mediated TEs in insects. Representatives are shown from three major groups: (a) long terminal repeat (LTR) retrotransposons, (b) non-LTR retrotransposons, and (c) short interspersed repetitive elements (SINEs). The name of each representative element, its host species, and its approximate length are shown as the heading. Open reading frames (ORFs) are shown as open boxes. The elements are not drawn to scale. References for information on these RNA-mediated elements are the following. *BEL1*, Davis and Judd (1995); *Feilai*, Tu (1999); *Gypsy*, Mizrokhi and Mazo (1991); *I*, Fawcett *et al.* (1986); *L1*, Biedler and Tu (2003); *Outcast*, Biedler and Tu (2003). *env*, envelope protein; *gag*, group-associated antigene or group-specific antigen; *pol*, polymerase-like protein; APE, apurinic-apyrimidinic endonuclease; IN, integrase; LINEs, long interspersed repetitive elements; LTR, long terminal repeat; PR, protease; RH, RNase H; RT, reverse transcriptase.

retrotransposition machinery from autonomous non-LTR retrotransposons, which may be facilitated by similar sequences or structures at the 3′ ends of a SINE and its "partner" non-LTR retrotransposon (Ohshima *et al.*, 1996; Okada and Hamada, 1997). Unlike non-LTR retrotransposons that use internal Pol II promoters, SINE transcription is directed from their own Pol III promoters that are similar to those found in small RNA genes. SINEs can be further divided into three groups based on similarities of their 5′ sequences to different types of small RNA genes. Elements such as the primate *Alu* family share sequence similarities with 7SL RNA (Jurka, 1995) while most other SINEs belong to a different group that shares sequence similarities to tRNA molecules (Adams *et al.*, 1986; Okada, 1991; Tu, 1999). Recently, a new group of SINEs named *SINE3* have been discovered in the zebrafish genome, which share similarities to 5S rRNA (Kapitonov and Jurka, 2003a). Some non-LTR retrotransposons tend to generate truncated copies due to incomplete reverse transcription during cDNA synthesis. Although these short copies of RNA-mediated TEs are also called SINEs (Malik and Eickbush, 1998), they should not be confused with the true SINEs that use Pol III promoters.

4.12.2.2. Class II DNA-Mediated TEs

DNA-mediated TEs include cut-and-paste DNA transposons (**Figure 3**), miniature inverted-repeat TEs (MITEs) (**Figure 4**), and a recently discovered group called *Helitrons* (Kapitonov and Jurka, 2001, 2003b). Their transposition is directly from DNA to DNA, and does not involve an RNA intermediate (Craig, 2002).

4.12.2.2.1. Cut-and-paste DNA transposons DNA transposons such as *P*, *hobo*, and *mariner* are usually characterized by 10–200 bp terminal inverted repeats (TIRs) flanking one or more ORFs that encode a transposase. They usually transpose by a cut-and-paste mechanism and their copy number can be increased through a repair mechanism (Finnegan, 1992; Craig, 2002). As shown in **Figure 3**, cut-and-paste DNA transposons can be subdivided into several families or superfamiles according to their transposase sequences and these families/superfamilies are also characterized by TSDs of specific sequence or length. The families/superfamilies that have been found in insects include *IS630-Tc1-mariner, hAT, piggyBac, PIF/Harbinger, P,* and *Transib* (Shao and Tu, 2001; Robertson, 2002; Kapitonov and Jurka, 2003b).

4.12.2.2.2. MITEs MITEs are widely distributed in plants, vertebrates, and invertebrates (Oosumi *et al*., 1995; Wessler *et al*., 1995; Smit and Riggs, 1996; Tu, 1997, 2001a). Most MITEs share common structural characteristics such as TIRs, small size, lack of coding potential, AT richness, and the potential to form stable secondary structures (Wessler *et al*., 1995). MITEs may have been "borrowing" the transposition machinery of autonomous DNA transposons by taking advantage of shared TIRs (MacRae and Clegg, 1992; Feschotte and Mouches, 2000b; Zhang *et al*., 2001). However, an alternative hypothesis suggests that they may transpose by a hairpin DNA intermediate produced from the folding back of single-stranded DNA during replication, which may better explain how MITEs could achieve immensely high copy numbers in some genomes (Izsvak *et al*., 1999). The evolutionary origin of MITEs is still not clear (see Section 4.12.4.3.3). One obvious source of MITEs would be internal deleted autonomous DNA transposons (Feschotte and Mouches, 2000b). In this case, MITEs are basically nonautonomous deletion derivatives of DNA transposons. Recent studies show that the similarities between many MITEs and their putative autonomous partners are restricted to the TIRs (Feschotte *et al*., 2003) (**Figure 4**). Although subsequent loss of autonomous partners in the genome remains a possible explanation for the lack of internal sequence similarity between MITEs and autonomous DNA transposons, two other explanations are perhaps more plausible. First, MITEs could originate *de novo* from chance mutation or recombination events resulting in the association of TIRs flanking unrelated segments of DNA (MacRae and Clegg, 1992; Tu, 2000; Feschotte *et al*., 2003). Alternatively, MITEs could originate from abortive gap-repair following the transposition of DNA transposons, which has been shown to occasionally introduce transposon-unrelated sequences (Rubin and Levy, 1997).

4.12.2.2.3. *Helitrons*: the rolling-circle DNA transposons *Helitrons* and related transposons have recently been discovered in insects and plants, which appear to use a rolling-circle mechanism of transposition (Le *et al*., 2000; Kapitonov and Jurka, 2001; 2003b). Instead of cut-and-paste transposase, *Helitrons* encode proteins similar to helicase, ssDNA-binding protein, and replication initiation protein. These proteins facilitate the rolling-circle replication of *Helitrons*, a mechanism previously described for the bacterial *IS91* transposons (Garcillan-Barcia *et al*., 2002).

4.12.2.3. Related Topics

4.12.2.3.1. *Foldback* elements *Drosophila Foldback* elements are characterized by very long inverted repeats (Truett *et al*., 1981). It is not known how *Foldback* elements transpose, although the presence of long inverted repeats indicates a possible DNA-mediated mechanism. Some researchers group *Foldback* elements as a distinct class, namely Class III (Kaminker *et al*., 2002).

4.12.2.3.2. What is a family? Before we move ahead with discussions on the discovery and diversity of insect TEs in the next two sections, it may be helpful to clarify the use of the term "family" in the context of TEs. The term "family" is often used to refer to a group of related TEs in diverse organisms that usually share conserved amino acid sequences in their transposase or RT and one such example is the *mariner* family. A TE also consists of many copies generated by transposition events in a genome, and therefore these related copies are sometimes also referred to as a family. Some families consist of multiple distinct groups that are subdivided into subfamilies. Obviously, relatedness is a relative concept in evolution and thus a working definition is needed in each case until a universal family definition is developed.

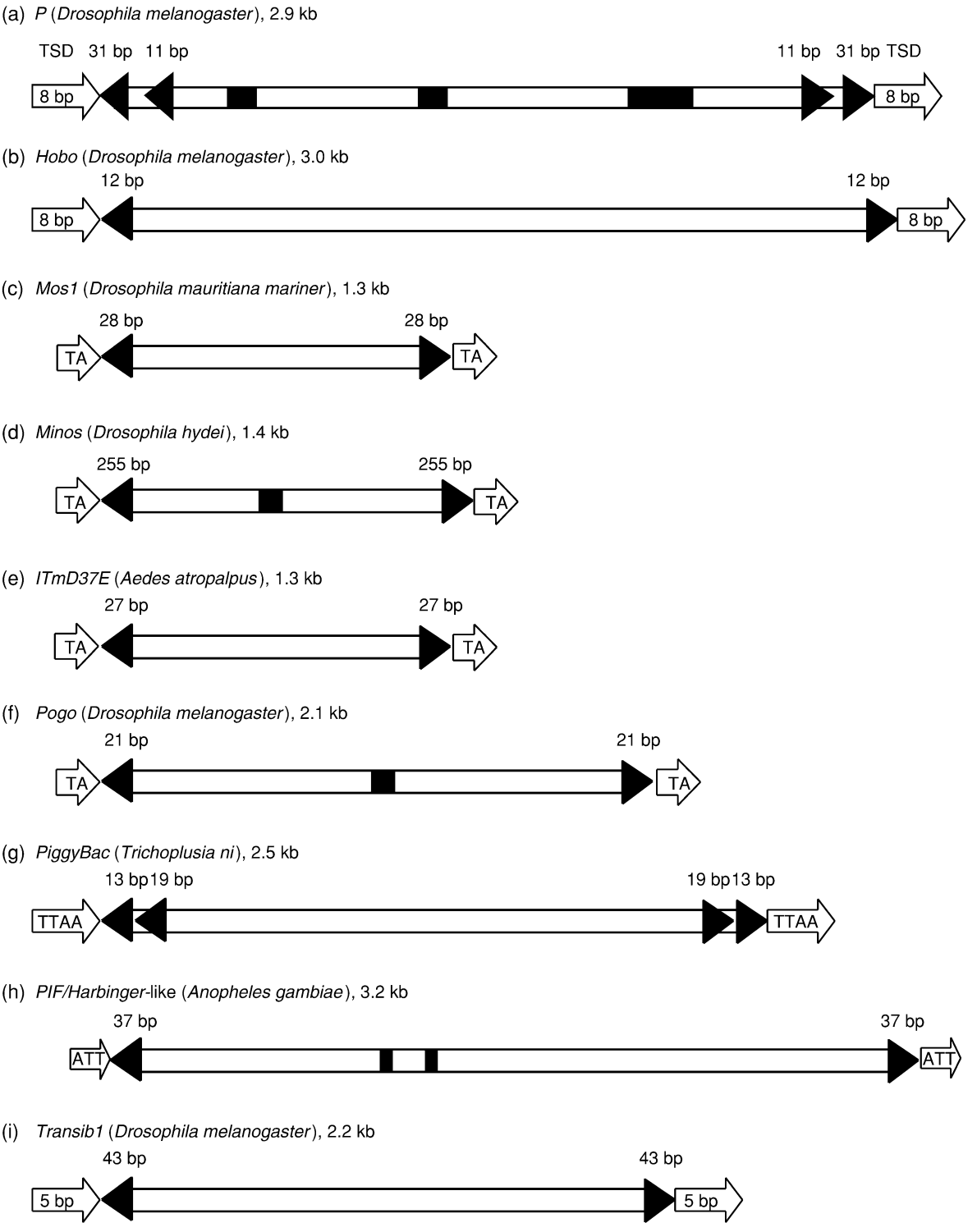


Figure 3 Structural characteristics of representative cut-and-paste DNA transposons in insects. The name of each representative transposon, its host species, and its approximate length are shown as the heading of each panel. Open arrows indicate target site duplications (TSDs). Filled triangles indicate terminal and subterminal inverted repeats. The lengths of these inverted repeats are marked. Exons are shown as open boxes and introns are shown as filled black boxes. The 5′ and 3′ untranslated regions are not shown. The elements are not drawn to scale. References for information on these transposons are the following: *P*, Engels (1989) and Rio (2002); *hobo*, Streck *et al.* (1986); *mariner Mos1*, Jacobson *et al.* (1986); *Minos*, Franz and Savakis (1991); *ITmD37E*, Shao and Tu (2001); *Pogo*, Tudor *et al.* (1992) and Feschotte and Mouches (2000b); *piggyBac*, Cary *et al.* (1989); *PIF/Harbinger*-like, Biedler *et al.* unpublished data; *Transib1*, Kapitonov and Jurka (2003b).

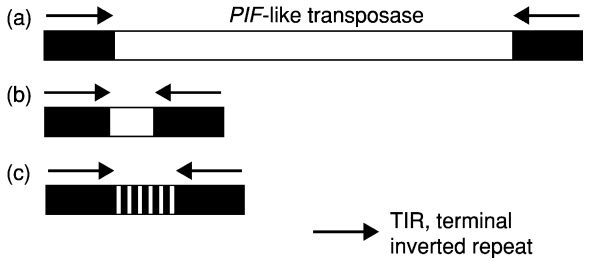


Figure 4 Relationship between *PIF/Harbinger*-like DNA transposons and related MITEs in *Anopheles gambiae*. (a) Structural features of *PIF/Harbinger*-like DNA transposons. There are multiple *PIF/Harbinger*-like DNA transposon families in *A. gambiae*, all of which have the characteristic AT-rich 3 bp TSDs (Biedler *et al.*, unpublished data). Although the TIRs are highly conserved between different copies of a family, they are variable between different *PIF/Harbinger*-like transposon families. (b) Structural features of deletion derivatives of *PIF/Harbinger*-like DNA transposons. Only two elements are apparent deletion derivatives of two different *PIF/Harbinger*-like transposons. (c) Structural features of *PIF/Harbinger*-related MITEs. These putative MITEs share similar TSDs and TIRs with *PIF/Harbinger*-like DNA transposons, although there are no internal sequence similarities between them. Twenty families of *PIF/Harbinger*-realted MITEs belong to this category. Note that the term ''family'' is used here to refer to a group of similar copies.

4.12.3. Methods to Uncover and Characterize Insect TEs

4.12.3.1. Early Discoveries and General Criteria

Before the availability of the large amount of genomic sequence data, TEs were often discovered by serendipitous observations during genetic experiments. As described above (see Section 4.12.1), McClintock's observation of the unstable mutations in maize led to the discovery of two mobile genetic elements *Ac* and *Ds*, although the molecular characterization of these elements came many years later (review: Fedoroff, 1989). Similarly, the observation of an unstable *white-peach* eye-color mutation in *Drosophila mauritiana* led to the discovery of the *mariner* transposon (Hartl, 1989). The *piggyBac* transposon was discovered as an insertion in a baculovirus after passage through a cell line of the cabbage looper *Trichoplusia ni* (review: Fraser, 2000). In a slightly different vein, the *D. melanogaster P* and *I* elements were discovered because of their association with a genetic phenomenon called hybrid dysgenesis, which refers to a group of abnormal traits including high mutation rates and sterility in crosses of certain strains (Kidwell, 1977; Finnegan, 1989). The genetic mutations described above, albeit rare, tend to identify active transposition events that resulted from active TEs in the genome.

The repetitive nature of TEs can also be used for their discovery and isolation, although not all repetitive elements are TEs. When DNA sequences are available, TEs can be identified on the basis of either similarity to known TEs, or common structural characteristics. In some cases, evidence for past TE insertion events could be identified on the basis of sequence analysis, which further supports the mobile history of a particular element. Although the criteria and methods described in this section are not unique to insects, it may be necessary to visit this topic here because of the lack of a systematic review on these issues and because of the growing interests in TE analysis in the current genomic environment.

4.12.3.2. Experimental Methods to Isolate and Characterize TEs

Several experimental methods have been used to discover TEs on the basis of their repetitive nature. Although relatively straightforward, these methods may not clearly distinguish between TEs and other repetitive sequences in the genome. In other words, the repeats discovered using these methods are not always TEs. One way to discover repeats in the genome is to isolate visible bands in an agarose gel running a sample of restriction enzyme-digested genomic DNA. This is based on the assumption that only highly reiterated sequences containing two or more conserved recognition sites for the restriction enzyme will produce a visible band among the smear of digested genomic DNA, and the bands can be cut out from the gel and purified for cloning and sequencing. Another approach to search for repeats is to screen a genomic library using labeled genomic DNA as a probe. This approach can be used effectively to identify abundant or highly repetitive sequences in the genome, which is based on the principle that only the repetitive fraction of the genome will produce a sufficient amount of labeled fragments that will generate hybridization signals during the screening (Gale, 1987; Cockburn and Mitchell, 1989). A third method is to use *Cot* analysis to help identify repetitive sequences in the genome, which is based on DNA reassociation kinetics (Adams *et al.*, 1986; Peterson *et al.*, 2002). For example, *Cot* analysis of genomic DNA can be performed to isolate the moderately repetitive portion of the genome that tends to contain TEs, and one can construct a subgenomic library using this fraction of genomic DNA to search for potential TEs. Several methods can be used to identify and isolate TEs on the basis of information derived from related TEs. For example, homologous TE probes may be used in Southern blotting and genomic library

screening experiments to identify related TEs. Polymerase chain reaction (PCR) analysis using primers that are conserved between related TEs can also be used to isolate different members of a TE family.

4.12.3.3. Computational Approaches to Discover and Analyze TEs

Insect science is rapidly entering a new era as demonstrated by the publication of the *D. melanogaster* genome sequence (Adams *et al.*, 2000), the report of the genome assembly of the African malaria mosquito *Anopheles gambiae* (Holt *et al.*, 2002), and the recent progress in genome projects of a number of other insects including the yellow fever mosquito *Aedes aegypti*, the honeybee *Apis mellifera*, the silkworm *Bombyx mori*, and the tobacco budworm *Heliothis virescens* (Kaufman *et al.*, 2002; Transgenesis and Genomics of Invertebrate Organisms, 2003). This genome revolution is producing an ever-expanding sea of data that can be explored using a bioinformatics approach to identify interspersed TEs. As described below, a few new tools have been developed which represent a shift from merely masking TEs (e.g., RepeatMasker; review: Jurka, 2000) to the discovery, annotation, and genomic analysis of TEs. The use of bioinformatics tools provides great advantages by allowing analysis of TEs in the entire genome and by allowing quick surveys of a large number of TE families to identify the most promising candidates for discovering active TEs (see Section 4.12.5) and for population analysis (see Section 4.12.9). It should be noted that these approaches are not limited to fully sequenced genomes. Because of the repetitive nature of TEs, sequences from a small fraction of a genome tend to contain a large number of TEs that may be discovered using the bioinformatics approaches described here. Of course, greater numbers of sequences and longer assembly would be beneficial in analyzing low copy number or long TE sequences. It should also be mentioned that no high-end computing facilities are required for the majority of these bioinformatics programs.

4.12.3.3.1. Homology-dependent approaches Searching for TEs in a genome on the basis of similarities to known elements discovered in different species is relatively straightforward. However, given the diversity and abundance of TEs, systematic computational approaches are necessary for efficient and comprehensive analysis. One such program was reported that uses profile hidden Markov models to find all sequences matching the full-length RT with the conserved FYXDD motif common to all reverse transcriptases (Berezikov *et al.*, 2000). A BLAST-based systematic approach to simultaneously identify and classify TEs is developed recently (**Figure 5**). This approach incorporates multiquery BLAST (Altschul *et al.*, 1997) and a few computer program modules (freely available at jaketu.biochem.vt.edu) that organize BLAST output, retrieve sequence fragment, and mask database for identified TEs. The method was successfully used to discover and characterize non-LTR retrotransposons in the *A. gambiae* genome assembly (Biedler and Tu, 2003). The strategy is explained using the example of a representative non-LTR retrotransposon as shown in **Figure 5a**. The potentially comprehensive nature of this approach was demonstrated as two new clades were identified (Biedler and Tu, 2003) (see Section 4.12.4). The inclusive nature of this approach was further indicated when observed that non-LTRs across all existing clades were identified using the reiterative approach with a single *D. melanogaster* representative in the *Jockey* clade as the starting query. The reiterative feature of the strategy described in this study provides an opportunity for further automation by linking the modules described in **Figure 5a**. An early version of a fully automated program was tested, which was named TEpipe, to identify all non-LTR retrotransposon families in the *A. gambiae* genome. Preliminary results showed that 102 of the 104 non-LTR retrotransposon families were identified in one run of the program which takes less than 2 h on a Linux workstation. The robustness of the family classification assigned by the program was confirmed using a few independent tests. Alignment of the nucleotide sequences plus flanking genomic sequence of each family was performed with ClustalW (Thompson *et al.*, 1994) to determine transposon boundaries, full-length elements, and TSDs (**Figure 5b**). The alignment process has recently been automated. The TEpipe approach should work for any TE with coding capacity in any genome.

4.12.3.3.2. Homology-independent approaches A suite of computer programs has been developed recently to search large databases rapidly for sequences with characteristics of MITEs (Tu, 2001a). The key program, FINDMITE1, searches the database for inverted repeats flanked by user-defined direct repeats within a specified distance range (**Figure 6**). The program uses the idea of the Knuth–Morris–Pratt string-matching algorithm (Tu, 2001a) to speed up the pattern match shifts. FINDMITE1 was used to uncover eight novel families of MITEs in *A. gambiae* (Tu, 2001a). Improvements have been made so that the new version of

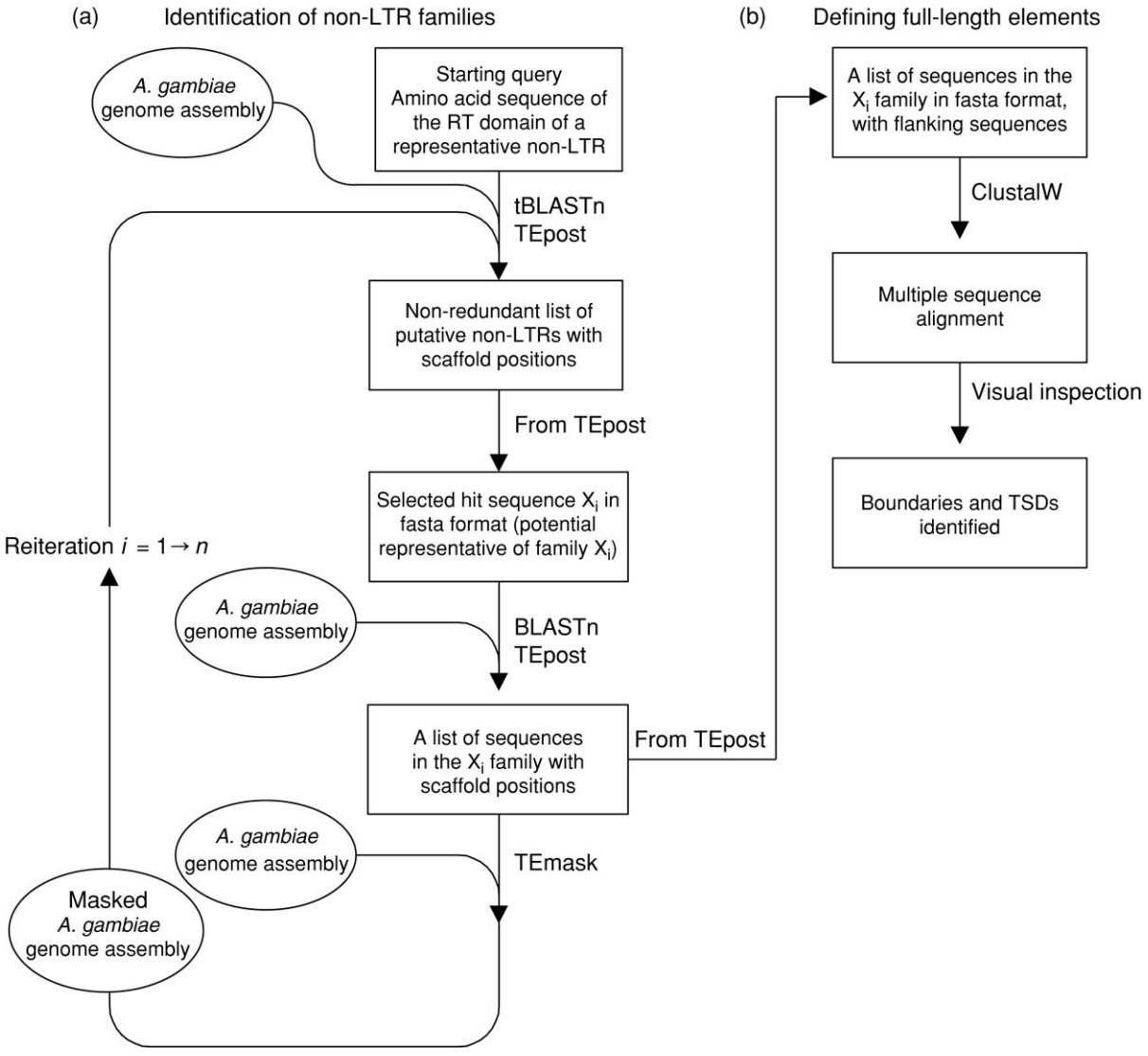


Figure 5 Design of TEpipe, a pipeline program for simultaneous identification and classification of TEs on the basis of sequence similarities. The example shown here is the strategy used to analyze all non-LTR retrotransposons in the *Anopheles gambiae* genome (Biedler and Tu, 2003). (a) Strategy to identify non-LTR retrotransposon families. (b) Strategy to define full-length elements. Ovals indicate databases used for searches. Rectangles indicate input/output files. Program modules are in bold beside arrows. Modules written in our laboratory are available at the website jaketu.biochem.vt.edu. (Biedler and Tu, 2003). TSDs, target site duplications. (Reproduced with permission from Biedler, J., Tu, Z., **2003**. Non-LTR retrotransposons in the African malaria mosquito, *Anopheles gambiae*: unprecedented diversity and evidence of recent activity. *Mol. Biol. Evol. 20*, 1811–1825; © Oxford University Press.)

FindMITE handles whole-genome sequence databases better, incorporates downstream analysis, and produces fewer false positive results which could be overwhelming in some cases. A program module AlignMITE has been developed to identify candidate MITEs that share the same TIRs and automatically align them by calling ClustalW (Thompson *et al.*, 1994) to determine MITE boundaries and full-length elements. Another program module, MITEInsertion, has been developed to uncover evidence of MITE insertion. Because a whole-genome database is not required, this systematic approach could have broad applications for the analysis of the model genomes as well as the vast majority of the less sequenced genomes. Although FINDMITE1 was originally designed to discover and analyze MITEs, it is possible to use it for the identification and characterization of DNA transposons that also contain TIRs. It is possible to use TEpipe to survey for the transposase-encoding sequences and use FINDMITE to analyze the TIRs and TSDs.

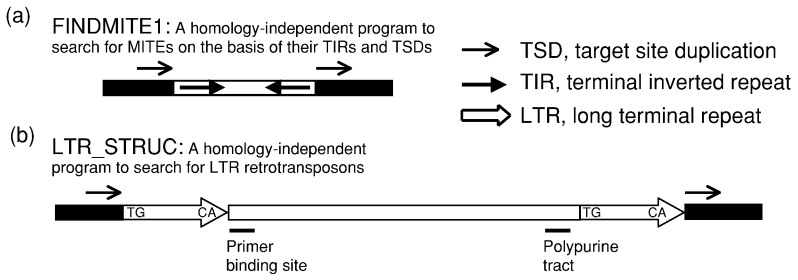


Figure 6 Design of two programs to search for different groups of TEs on the basis of shared structural features. (a) FINDMITE1 is a C program designed to rapidly search a database for sequences that have the characteristics of MITEs (Tu, 2001a). The program searches sequences in the database for inverted repeats flanked by user-defined direct repeats within a specified distance range. FINDMITE1 is available at our website (jaketu.biochem.vt.edu). The newly developed MITEpipe program, which further reduces false positive signals and incorporates FINDMITE1 with downstream analysis, will be available soon at the same website. (b) LTR_STRUC is a program that identifies LTR retrotransposons on the basis of the presence of long terminal repeats (most LTRs contain TG. . .CA termini), target site duplications (TSDs), and additional information such as primer binding site and polypurine tract (McCarthy and McDonald, 2003). The program is available as a console application from its authors.

LTR_STRUC is a program that identifies LTR retrotransposons on the basis of the presence of LTRs (most LTRs contain TG. . .CA termini), TSDs, and additional information such as primer binding site and polypurine tract (McCarthy and McDonald, 2003) (**Figure 6b**). Although it is not designed to uncover solo LTRs or truncated LTR retrotransposons, the program offers a rapid and efficient approach to systematically identify and characterize LTR retrotransposons in a given genome and it can be used as a discovery tool for new families of LTR retrotransposons. The program is available as a console application from its authors.

Recently, an automated program named RECON has been reported (Bao and Eddy, 2002), which identifies TE sequences on the basis of their repetitive nature in the genome. RECON uses a multiple sequence alignment algorithm which represents a significant improvement to previous methods based on the same strategy. Because it does not rely on sequence homology or structural information, RECON is potentially comprehensive and may be able to identify all repetitive sequences. By the same token, RECON is computationally intensive as the size of the genome database increases. A two-step approach in which a large number of repeats are identified and masked first using a small fraction of the genome sequence may resolve this potential problem.

4.12.4. Diversity and Characteristics of Insect TEs

4.12.4.1. Overview

Virtually all classes and types of eukaryotic TEs have been found in insects. Insect TEs such as *copia*, *gypsy*, *I*, *R1*, *P*, *mariner*, *hobo*, *piggyBac*, and *transib* are the founding members of several diverse families/superfamilies that were later shown to have broad distributions in eukaryotes. In addition, recent studies revealed a few novel and intriguing TEs in insects that are described in detail below. **Tables 1** and **2** provide a relatively extensive, but by no means exhaustive, compilation of the two classes of TEs in insects. **Figures 2** and **3** depict the characteristics of representatives of different RNA-mediated TEs and the cut-and-paste DNA transposons. In addition to the two "fully sequenced" dipterans, a few other dipteran species and a lepidopteran species (*Bombyx mori*) represent the sources from which the majority of TEs have been discovered and analyzed. As more insect genome projects move ahead and studies on insect TEs expand, more TEs from diverse species will be discovered which will expand our horizon and offer new insights from a comparative genomics perspective. In this section, a detailed account of the characteristics of all different families of TEs will not be discussed. Instead the focus will be on recent advances and interesting features of some novel insect TEs. In anticipation of the explosion of TE discovery in the near future, my laboratory will periodically upload updated **Tables 1** and **2** on a website (jaketu. biochem.vt.edu).

4.12.4.2. RNA-Mediated TEs

4.12.4.2.1. LTR retrotransposons The distribution of LTR retrotransposons in insects is shown in **Table 1**. The structural characteristics of representatives of three major groups found in insects, Ty1/*copia*, Ty3/*gypsy*, and BEL, are shown in **Figure 2**. In addition to the LTRs that contain the

Table 1 RNA-mediated transposable elements in insects

Superfamily (clade)	*Element and reference*[b]	*Order*	*Organism*[c,d]
I. LTR retrotransposons[a]			
Ty1/copia	*Mosqcopia* (Tu, unpublished data), *Zebedee* (Warren *et al.*, 1997)	Diptera	*Aedes aegypti*
	copia-like (Holt *et al.*, 2002), *mtanga* (Rohr *et al.*, 2002)	Diptera	*Anopheles gambiae*
	1731, copia, Dm88, frogger	Diptera	*Drosophila melanogaster*
	Yokozuna (Ohbayashi *et al.*, 1998)	Lepidoptera	*Bombyx mori*
Ty3/gypsy	*Woot* (Beeman *et al.*, 1996)	Coleoptera	*Tribolium castaneum*
	Beagle, gypsy-like, Cruiser, Osvaldo, Springer (Holt *et al.*, 2002)	Diptera	*Anopheles gambiae*
	Ozymandias (Hill *et al.*, 2001)	Diptera	*Anopheles gambiae*
	yoyo (Zhou and Haymer, 1998)	Diptera	*Ceratitis capitata*
	17.6, 297, 412, accord, blastopia, blodd, Burdock, Circe, gtwin, gypsy, HMS Beagle, Idefix, invader, McClintock, mdg1, mag3, micropia, opus, qbert, Quasimodo, rover, springer, Stalker, Tabor, Tirant, Transpac, Zam	Diptera	*Drosophila melanogaster*
	Judo, Karate (Abe *et al.*, 2002)	Lepidoptera	*Bombyx mandarina*
	Kabuki (Abe *et al.*, 2000), *Mag* (Garel *et al.*, 1994)	Lepidoptera	*Bombyx mori*
	Lydia (Pfeifer *et al.*, 2000)	Lepidoptera	*Lymantria dispar*
	TED (Friesen and Nissen, 1990)	Lepidoptera	*Trichoplusia ni*
BEL	*MosqNinja* (Tu, unpublished data)	Diptera	*Aedes aegypti*
	Moose, Pao-like (Holt *et al.*, 2002)	Diptera	*Anopheles gambiae*
	aurora, Bel, diver, GATE, roo, rooA	Diptera	*Drosophila melanogaster*
	Yamato (Abe *et al.*, 2002)	Lepidoptera	*Bombyx mandarina*
	Pao (Xiong *et al.*, 1993), *Kamikaze, Yamato* (Abe *et al.*, 2001)	Lepidoptera	*Bombyx mori*
II. Non-LTR retrotransposons[a]			
R4	*Ag-R4_* 1 (Biedler and Tu, 2003)	Diptera	*Anopheles gambiae*
	Dong (Xiong and Eickbush, 1993)	Lepidoptera	*Bombyx mori*
R2	*R2* (Eickbush and Malik, 2002)	Diptera	*Drosophila melanogaster*
	R2Bm (Xiong and Eickbush, 1988)	Lepidoptera	*Bombyx mori*
L1	*Ag-L1_* 1–5 (Biedler and Tu, 2003)	Diptera	*Anopheles gambiae*
L2	*Ag-L2_* 1–3 (Biedler and Tu, 2003)	Diptera	*Anopheles gambiae*
RTE	*JAM1* (GenBank Z86117)	Diptera	*Aedes aegypti*
	*Ag-Jammin_* 1–2 (Biedler and Tu, 2003)	Diptera	*Anopheles gambiae*
R1	*RT1, RT2, Ag-R1 _*1–11 (Biedler and Tu, 2003)	Diptera	*Anopheles gambiae*
	R1, RT1 (Waldo)	Diptera	*Drosophila melanogaster*
	R1Bm (Xiong and Eickbush, 1988), *SART1, TRAS1* (Anzai *et al.*, 2001)	Lepidoptera	*Bombyx mori*
LOA	*Lian* (Tu *et al.*, 1998)	Diptera	*Aedes aegypti*
	Bilbo, Baggins1 (Kapitonov and Jurka, 2003b)	Diptera	*Drosophila melanogaster*
	LOA (Felger and Hunt, 1992)	Diptera	*Drosophila silvestris*
I	*MosqI* (Tu and Hill, 1999)	Diptera	*Aedes aegypti*
	*Ag-I_* 1–7 (Biedler and Tu, 2003)	Diptera	*Anopheles gambiae*
	I, I2, You	Diptera	*Drosophila melanogaster*
Jockey	*JuanA* (Mouches *et al.*, 1992)	Diptera	*Aedes aegypti*
	*Ag-Jockey_* 1–13, *Ag-Jen _*1–12, (Biedler and Tu, 2003)	Diptera	*Anopheles gambiae*
	NCR1Cth1 (Blinov *et al.*, 1997)	Diptera	*Chironomus thummi*
	JuanC (Agarwal *et al.*, 1993)	Diptera	*Culex pipiens*
	BS, Doc, F, G, Helena, Het-A, Jockey, Juan, TART, X	Diptera	*Drosophila melanogaster*
	AMY (*BMC1, L1Bm*; Abe *et al.*, 1998)	Lepidoptera	*Bombyx mori*
	LDT1 (Garner and Slavicek, 1999)	Lepidoptera	*Lymantria dispar*
CR1	*T1, Q, Ag-CR1 _*1–29 (Biedler and Tu, 2003)	Diptera	*Anopheles gambiae*
	DMCR1A (Kapitonov and Jurka, 2003b)	Diptera	*Drosophila melanogaster*
Loner	*Ag-Loner_* 1–3 (Biedler and Tu, 2003)	Diptera	*Anopheles gambiae*
Outcast	*Ag-Outcast_* 1–11 (Biedler and Tu, 2003)	Diptera	*Anopheles gambiae*
Not classified	*Sponge* (Biedler and Tu, 2003)	Diptera	*Anopheles gambiae*
	CM-gag (Bensaadi-Merchermek *et al.*, 1997)	Diptera	*Culex pipiens*
	Kurosawa, Kendo (Abe *et al.*, 2002)	Lepidoptera	*Bombyx mandarina*
III. SINEs			
tRNA-related SINEs	*Feilai* (Tu, 1999)	Diptera	*Aedes aegypti*
	SINE200 (Holt *et al.*, 2002)	Diptera	*Anopheles gambiae*
	Cp1 (Liao *et al.*, 1998)	Diptera	*Chironomus pallidivittatus*

Continued

Table 1 Continued

Superfamily (clade)	*Element and reference*[b]	*Order*	*Organism*[c,d]
	Twin (Feschotte *et al.*, 2001)	Diptera	*Culex pipiens*
	Bm1 (Adams *et al.*, 1986)	Lepidoptera	*Bombyx mori*
	Lm1 (Bradfield *et al.*, 1985)	Orthoptera	*Locusta migratoria*
Unclassified SINEs?	*Maque* (Tu, 2001b)	Diptera	*Anopheles gambiae*
	DINE-1 (Locke *et al.*, 1999)	Diptera	*Drosophila melanogaster*
	IE (Sun *et al.*, 1991; Lampe and Willis, 1994)	Lepidoptera	*Hyalophora cecropia*
IV. Penelope-like retrotransposons			
Penelope	*Penelope-like* (Biedler and Tu, unpublished data)	Diptera	*Aedes aegypti*
	Penelope (Evgen'ev *et al.*, 1997)	Diptera	*Drosophila virilis*

[a]Classifications of LTR and non-LTR retrotransposons are according to Eickbush and Malik (2002).
[b]The citation is not always the original report of a TE. In some cases a review or an article describing the phylogenetic placement of the TE is cited instead. References for all *D. melanogaster* TEs are from Kaminker *et al.* (2002) unless otherwise noted. There are several PCR surveys of a few RNA-mediated TEs in different insects that are not listed in the table (Booth *et al.*, 1994; Rongnoparut *et al.*, 1998; Cook *et al.*, 2000).
[c]TEs in *Drosophila* species other than *D. melanogaster* are not listed unless they are the founding member of a group of TEs.
[d]In some cases, TEs from only one species are listed although they were found in several closely related species.

promoter and transcription termination sequences, the LTR retrotransposons have a flexible structure that allows gain, loss, and perhaps rearrangement of functional domains (Eickbush and Malik, 2002). While the RT and a few other protein domains perform the key function of making double-stranded cDNA, the acquisition of the IN activity allowed the integration of the cDNA in a way much like the mechanisms employed by DNA transposons. In fact, the IN domain and some of the prokaryotic and eukaryotic DNA transposases are believed to share a common origin (Capy *et al.*, 1997; and see below). The acquisition of the *env*-like protein by some LTR retrotransposons such as *gypsy* confers the ability to leave the cell and become infectious retroviruses (Eickbush and Malik, 2002).

4.12.4.2.2. Non-LTR retrotransposons Twelve of the 17 clades of non-LTR retrotransposons have been found in insects (Eickbush and Malik, 2002; Biedler and Tu, 2003) (**Table 1** and **Figure 7**). In fact, the founding members of many of these clades were discovered in insects. The characterization and classification of 104 families of non-LTR retrotransposons in *A. gambiae* have been recently reported (Biedler and Tu, 2003). The 104 *A. gambiae* families represent divergent lineages in eight previously established clades (*R4*, *L1*, *RTE*, *R1*, *L2*, *CR1*, *Jockey*, and *I*) and two new clades, *Loner* and *Outcast*. Representation appears to be biased toward the most derived clades in non-LTR retrotransposon evolution, especially the *CR1* and *Jockey* clades, with 31 and 25 families respectively. All of the 31 *A. gambiae CR1* elements are grouped together with the *Drosophila CR1* being a sister branch. On the other hand, there appear to be three groups (I, II, and III) of the *A. gambiae Jockey* elements, which may have different sister elements either from other mosquitoes or other dipterans (Biedler and Tu, 2003).

4.12.4.2.3. SINEs SINEs have not been extensively investigated in insects. Only a small number of them have been found as shown in **Table 1**. Insect SINEs characterized so far all belong to the tRNA-related group. The structural features of *Feilai*, a SINE discovered in *Aedes aegypti*, include a tRNA-related promoter region, a tRNA-unrelated conserved region, and a triplet tandem repeat at its 3′ end (**Figure 2**). The *Twin* SINE family, which was discovered in *Culex pipiens* (Feschotte *et al.*, 2001), consists of two tRNA-related regions separated by a 39 bp spacer. *SINE200* from *A. gambiae* contains only one of the two conserved boxes found in tRNA-related Pol III promoters (Tu, unpublished data). *DINE-1*, a SINE from *D. melanogaster*, lacks the structural features of typical SINEs (Locke *et al.*, 1999). *Feilai*, *SINE200*, and two other SINEs *BM1* and *Lm1* (**Table 1**) are all highly repetitive in their respective genomes. *Twin* is only moderately repetitive with a copy number of 500. *Cp1*, a *Chironomus pallidivittatus* SINE, inserts specifically to centromeric tandem repeats (Liao *et al.*, 1998).

4.12.4.2.4. Two intriguing families: *Maque* and *Penelope* A family of very short interspersed repetitive elements named *Maque* has recently been found in *A. gambiae*. There are approximately 220

Table 2 DNA-mediated transposable elements in insects

Superfamily/family	*Defining features*	*Element and reference*[a]	*Order*	*Organism*[b,c]
I. Cut-and-paste transposons				
P	8 bp TSD, conserved transposase	*AgaP* (8 subfamilies: Sarkar *et al.*, 2003)	Diptera	*Anopheles gambiae*
		P (Canonical, M, O, T), *ProtoP (Hoppel), ProtoP_B* (Kapitonov and Jurka, 2003b)	Diptera	*Drosophila melanogaster*
		Lu-P1, Lu-P2 (Perkins and Howells, 1992)	Diptera	*Lucilia cuprina*
		P (Lee *et al.*, 1999)	Diptera	*Musca domestica*
hAT (hobo-Ac-Tam3)	8 bp TSD, conserved transposase	*hAT*-type (Holt *et al.*, 2002)	Diptera	*Anopheles gambiae*
		hopper (Handler, 2003)	Diptera	*Bactrocera dorsalis*
		Homer (Pinkerton *et al.*, 1999)	Diptera	*Bactrocera tryoni*
		hobo (Calvi *et al.*, 1991)	Diptera	*Drosophila melanogaster*
		Hermit (Coates *et al.*, 1996)	Diptera	*Lucilia cuprina*
		Hermes (Warren *et al.*, 1994)	Diptera	*Musca domestica*
IS630-Tc1-mariner	TA TSD, DDE(D) catalytic triad			
Tc1	DD34E	*Quetzal* (Ke *et al.*, 1996)	Diptera	*Anopheles albimanus*
		Crusoe (Hill *et al.*, 2001), *DD34E, Tiang, Topi, Tsessebe* (Holt *et al.*, 2002)	Diptera	*Anopheles gambiae*
		Minos (Franz and Savakis, 1991)	Diptera	*Drosophila hydei*
		Bari1-2, HB, S, S2, Tc1 (Kaminker *et al.*, 2002); *Tc3*-like (Shao *et al.*, 2001)	Diptera	*Drosophila melanogaster*
		Unnamed element (Mikitani *et al.*, 2000)	Lepidoptera	*Bombyx mori*
mariner[d]	DD34D	*mariner* (Holt *et al.*, 2002)	Diptera	*Anopheles gambiae*
		D.mauritiana.mar1 (Robertson, 2002)	Diptera	*Drosophila mauritiana*
		mariner2 (Kaminker *et al.*, 2002)	Diptera	*Drosophila melanogaster*
		A.mellifera.mar1 (Robertson, 2002)	Hymenoptera	*Apis mellifera*
		H.cecropia.mar1 (Robertson, 2002)	Lepidoptera	*Hyalophora cecropia*
ITmD37D (maT)	DD37D	*ITmD37D* (Holt *et al.*, 2002)	Diptera	*Anopheles gambiae*
		MdmaT1 (Claudianos *et al.*, 2002)	Diptera	*Musca domestica*
		Bmmar1 (Robertson and Asplund, 1996), *Bmmar6* (Robertson and Walden, 2003)	Lepidoptera	*Bombyx mori*
ITmD41D	DD41D	*Crmar2* (Gomulski *et al.*, 2001)	Diptera	*Ceratitis rosa*
		Tcp3.2 (Arends and Jehle, 2002; Robertson and Walden, 2003)	Lepidoptera	*Cydia pomonella*
ITmD37E	DD37E	*A.aegypti.ITmD37E* (Shao and Tu, 2001)	Diptera	*Aedes aegypti*
		A.gambiae.ITmD37E (Shao and Tu, 2001)	Diptera	*Anopheles gambiae*
Pogo[e]	DDxD, long C-terminus	*Pogo*-like (Holt *et al.*, 2002)	Diptera	*Anopheles gambiae*
		Pogo (Tudor *et al.*, 1992)	Diptera	*Drosophila melanogaster*
piggyBac	TTAA TSD, conserved transposase	*piggyBac* (Holt *et al.*, 2002; Sarkar *et al.*, 2003)	Diptera	*Anopheles gambiae*
		piggyBac (Handler and McCombs, 2000)	Diptera	*Bactrocera dorsalis*
		looper1 (Kapitonov and Jurka, 2002)	Diptera	*Drosophila melanogaster*
		piggyBac (Sarkar *et al.*, 2003)	Lepidoptera	*Bombyx mori*
		piggyBac (Cary *et al.*, 1989)	Lepidoptera	*Trichoplusia ni*

Continued

Table 2 Continued

Superfamily/family	*Defining features*	*Element and reference*[a]	*Order*	*Organism*[b,c]
PIF-harbinger	3 bp TSD, conserved transposase	*PIF/harbinger*-like (Biedler *et al.*, unpublished data)	Diptera	*Anopheles gambiae*
Transib	5 bp TSD, conserved transposase	*Transib1_AG* (Kapitonov and Jurka, 2003b)	Diptera	*Anopheles gambiae*
		Transib1-4 (Kapitonov and Jurka, 2003b), *Hopper* (Bernstein *et al.*, 1995)	Diptera	*Drosophila melanogaster*
Not classified		*Ikirara* (Romans *et al.*, 1998)	Diptera	*Anopheles gambiae*
		TECth1 (Wobus *et al.*, 1990)	Diptera	*Chironomus thummi*
II. Rolling circle transposons				
Helitron	No TSD, similarity to helicase	*Helitron1_AG, Helitron 2_AG* (Kapitonov and Jurka, 2003b)	Diptera	*Anopheles gambiae*
		Helitron (Kapitonov and Jurka, 2003b)	Diptera	*Drosophila melanogaster*
III. MITEs				
*mTA*TA TSD	*DEC* (Braquart *et al.*, 1999)	Cleoptera	*Tenebrio molitor*	
		Pony (Tu, 2000), *Wujin* (Tu, 1997); *mTA_1–8* (Mao and Tu, unpublished data)	Diptera	*Aedes aegypti*
		TA-I-Ag, TA-II-Ag, TA-III-Ag, TA-IV-Ag, TA-V-Ag (Tu, 2001a)	Diptera	*Anopheles gambiae*
		Mikado, Milord, Mimo, Mirza, Nemo (Feschotte *et al.*, 2002)	Diptera	*Culex pipiens*
		mPogo (Feschotte *et al.*, 2002)	Diptera	*Drosophila melanogaster*
m3bp (*Tourist*-like MITEs)	3 bp TSD	*m3bp_1–10* (Mao and Tu, unpublished data)	Diptera	*Aedes aegypti*
		Joey, TAA-I-Ag, TAA-II-Ag (Tu, 2001a), *m3bp_1–19* (Biedler *et al.*, unpublished data)	Diptera	*Anopheles gambiae*
		A MITE in a NOS gene (Luckhart and Rosenberg, 1999)	Diptera	*Anopheles stephensi*
m4bp	4 bp (often TTAA) TSD	*Wukong, Wuneng* (Tu, 1997), *m4bp_1–6* (Mao and Tu, unpublished data)	Diptera	*Aedes aegypti*
m7bp	7 bp TSD	*m7bp_1–3* (Mao and Tu, unpublished data)	Diptera	*Aedes aegypti*
m8bp	8 bp TSD	*m8bp_1–10* (Mao and Tu, unpublished data)	Diptera	*Aedes aegypti*
		8bp-I-Ag (Tu, 2001a), *Pegasus* (Besansky *et al.*, 1996)	Diptera	*Anopheles gambiae*
		Mar, Vege (Holyoake and Kidwell, 2003)	Diptera	*Drosophila willistoni*
m9bp	9 bp TSD	*m9bp_1–2* (Mao and Tu, unpublished data)	Diptera	*Aedes aegypti*
Not classified		*Microuli* (TTAA TSD, SIR: Tu and Orphanidis, 2001)	Diptera	*Aedes aegypti*
		MEC (5 bp TSD: Blinov *et al.*, 1991)	Diptera	*Chironomus thummi*
		Mint (CA TSD, SIR: Feschotte *et al.*, 2002)	Diptera	*Culex pipiens*
		SGM-IS (SIR: Miller *et al.*, 2000)	Diptera	*Drosophila obscura*

[a]The citation is not always the original report of a TE. In some cases a review or an article describing the phylogenetic placement of the TE is cited instead. *Drosophila Foldback* elements are not listed. There are several PCR surveys of a few DNA-mediated TEs in different insects that are not listed in the table (Daniels *et al.*, 1990; Robertson, 1993; Bigot *et al.*, 1994; Clark and Kidwell, 1997; Imwong *et al.*, 2000; Green and Frommer, 2001).

[b]TEs in *Drosophila* species other than *D. melanogaster* are not listed unless they are the founding member of a group of TEs.

[c]In some cases, TEs from only one species are listed although they were found in several closely related species.

[d]*mariner* sequences, full-length or fragment, have been found in a wide range of insects. The DD34D mariners consist of 6 subfamilies briggase, cecropia, elegans, irritans, mauritiana, and mellifera. See Shao and Tu (2001) for a reclassification of the *IS630-Tc1-mariner* superfamily. See Robertson (1993, 2002) for reviews on *mariner* evolution.

[e]The status of *Pogo* is not certain. It is either considered a member of the *IS630-Tc1-mariner* superfamily or a separate superfamily (Shao and Tu, 2001; Robertson, 2002).

copies of *Maque*. Only approximately 60 bp long, *Maque* has the appearance of a distinct transposition unit. The majority of *Maque* elements were flanked by 9–14 bp TSDs. *Maque* has several characteristics of non-LTR retrotransposons such as TSDs of variable length, imprecise 5′ terminus, and CAA simple repeats at the 3′ end. The evolutionary origin of *Maque* and the differences between *Maque* and other known retro-elements including SINEs is not yet known. The 5′ end of *Maque* represents a strong stop position that caused frequent premature termination of reverse transcription is suggested (Tu, 2001b). Although no autonomous non-LTR retrotransposons have been found that share similar 3′ sequences with *Maque*, there is a family of non-LTR retrotransposons, *Ag-I-2* (Biedler and Tu, 2003) that has the same CAA tandem repeats at their 3′ termini. It is possible that short sequences such as *Maque* that contain just the RT recognition signal could potentially contribute to the genesis of some primordial SINEs (Tu, 2001b).

Penelope, another intriguing family, was discovered as a TE involved in the hybrid dysgenesis of crosses between field-collected and laboratory strains of *D. virilis* (Evgen'ev *et al*., 1997). It has a RT that is grouped with the RT from telomerase (Arkhipova *et al*., 2003). More strikingly, members of the *Penelope* family in bdelloid rotifers are able to retain their introns, which is inconsistent with a transposition mechanism involving an RNA intermediate. It was proposed that the Uri endonuclease domain found in all *Penelope*-like elements may allow them, at least in part, to use a DNA-mediated mechanism similar to that used by group I introns (Arkhipova *et al*., 2003). On the basis of these unique features, *Penelope* was classified as a unique group that is distinct from LTR and non-LTR retrotransposons.

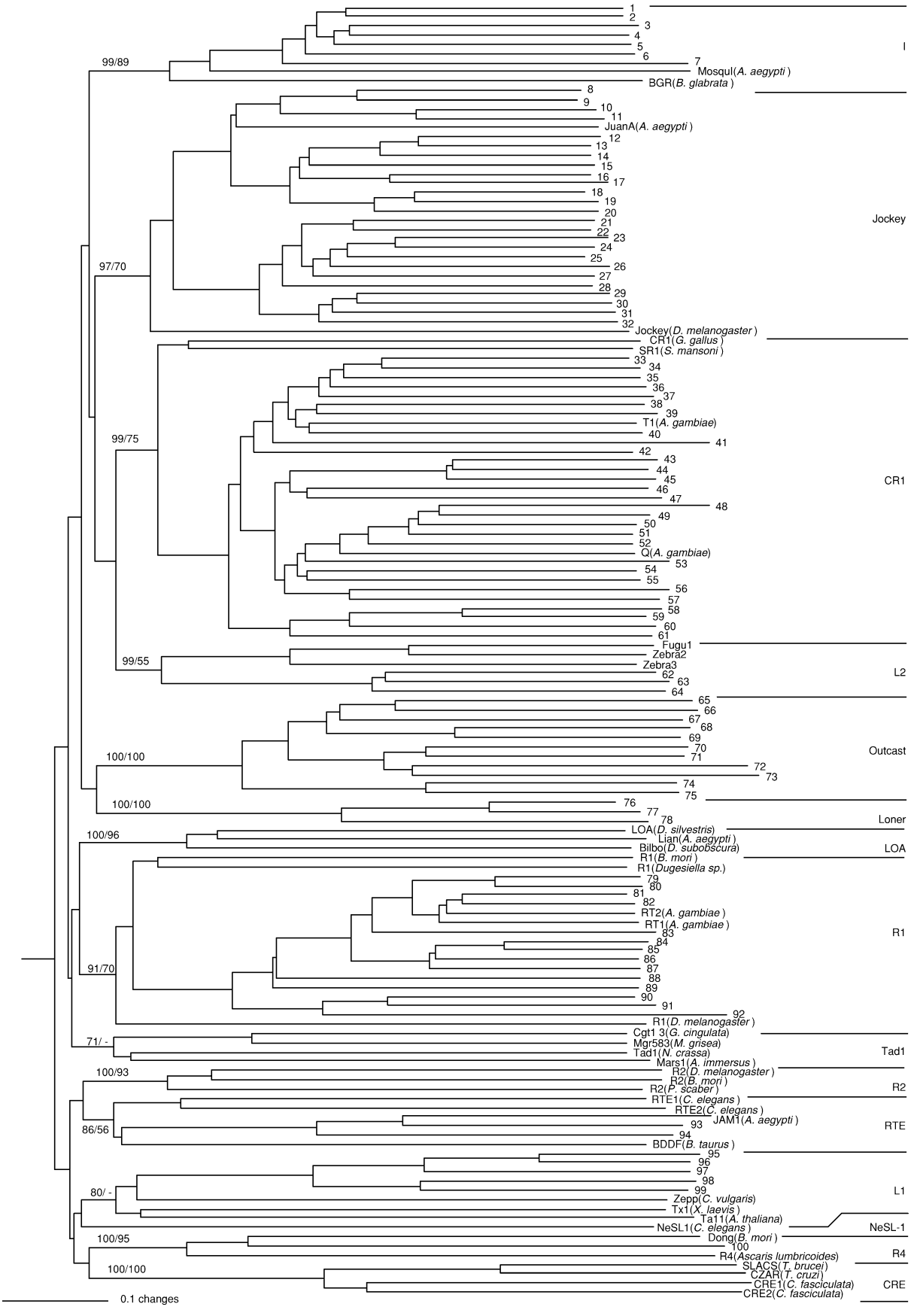
4.12.4.3. DNA-Mediated TEs

4.12.4.3.1. Cut-and-paste DNA transposons The majority of DNA-mediated TEs are believed to transpose by a cut-and-paste mechanism. This group of elements is characterized by 10–200 bp TIRs flanking one or more ORFs that encode a transposase, the enzyme that performs the excision and integration (cut and paste) of the cognate TE DNA. Their copy number may be increased through a repair mechanism, or by transposing ahead of a replication fork (Finnegan, 1992; Craig, 2002).

Different families/superfamilies of insect TEs in this group are listed in **Table 2**. Again, several insect TEs are the founding members of their respective families/superfamilies that have broad distributions. The families/superfamilies that have been found in insects include *IS630-Tc1-mariner, hAT, piggyBac, PIF/Harbinger, P*, and *Transib* (Shao and Tu, 2001; Robertson, 2002; Kapitonov and Jurka, 2003b). Conserved transposase sequences and TSDs of specific sequence or length are the hallmarks of each family/superfamily. *In vitro* experiments have shown that, for a few cut-and-paste DNA transposons characterized so far, the transposase alone is sufficient to direct the transposition reactions by interacting with the TIRs and the insertion target (Plasterk *et al*., 1999). However, the transposition of *P* elements requires a host protein (inverted repeat binding protein, IRBP) that binds to its TIRs (Badge and Brookfield, 1997; Rio, 2002). This requirement may explain the lack of success in the early attempts to use *P* elements to transform insects other than *Drosophila*.

The structural characteristics of representative elements from each family are shown in **Figure 3** (see **Chapter 4.13**). A few TE families including *PIF/Harbinger* (Biedler *et al*., unpublished data) and *Transib* (Kapitonov and Jurka, 2003b), which were discovered in insects only recently, will be highlighted in this section. Recent expansion and reclassification of the *IS630-Tc1-mariner* superfamily will also be discussed (Shao and Tu, 2001). In a separate section on MITEs (see Section 4.12.4.3.2), two new groups of mosquito MITEs will be described that have 7 and 9 bp TSDs respectively (Mao and Tu, unpublished data). The discovery of MITEs with 9 bp TSDs provides a preliminary indication of the possible existence of *Mutator*-like transposons in insects, which have so far been only found in plants (Walbot and Rudenko, 2002). On the other hand, the 7 bp TSD MITEs could potentially lead to the discovery of an entirely new family of eukaryotic DNA transposons because 7 bp TSD elements have until now only been found in bacteria (Krebs *et al*., 1990; Mahillon and Chandler, 1998) (see Section 4.12.4.3.2.2).

4.12.4.3.1.1. Discovery of the* ITmD37E *transposon and the reclassification of the* IS630-Tc1-mariner (ITm) *superfamily It was shown previously that some prokaryotic *IS* elements, eukaryotic *Tc1* and *mariner* transposons, and eukaryotic retrotransposons and retroviruses form a megafamily which shares similar signature sequences or motifs in the catalytic domain of their respective transposase and IN (Capy *et al*., 1996, 1997). The common motif for this transposase–integrase megafamily is a conserved D(Asp)DE(Glu) or DDD catalytic triad. The distance between the first two Ds is variable while the distance between the last two residues in the catalytic triad is mostly invariable


1
2
3
4
5
6
7
Mosqul(A. aegypti)
BGR(B. glabrata)
8
9
10
11
JuanA(A. aegypti)
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
Jockey(D. melanogaster)
CR1(G. gallus)
SR1(S. mansoni)
33
34
35
36
37
38
39
T1(A. gambiae)
40
41
42
43
44
45
46
47
48
49
50
51
52
Q(A. gambiae)
53
54
55
56
57
58
59
60
61
Fugu1
Zebra2
Zebra3
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
LOA(D. silvestris)
Lian(A. aegypti)
Bilbo(D. subobscura)
R1(B. mori)
R1(Dugesiella sp.)
79
80
81
82
RT2(A. gambiae)
RT1(A. gambiae)
83
84
85
86
87
88
89
90
91
92
R1(D. melanogaster)
Cgt1 3(G. cingulata)
Mgr583(M. grisea)
Tad1(N. crassa)
Mars1(A. immersus)
R2(D. melanogaster)
R2(B. mori)
R2(P. scaber)
RTE1(C. elegans)
RTE2(C. elegans)
JAM1(A. aegypti)
93
94
BDDF(B. taurus)
95
96
97
98
99
Zepp(C. vulgaris)
Tx1(X. laevis)
Ta11(A. thaliana)
NeSL1(C. elegans)
Dong(B. mori)
100
R4(Ascaris lumbricoides)
SLACS(T. brucei)
CZAR(T. cruzi)
CRE1(C. fasciculata)
CRE2(C. fasciculata)
99/89
97/70
99/75
99/55
100/100
100/100
100/96
91/70
71/ -
100/93
86/56
80/ -
100/95
100/100
I
Jockey
CR1
L2
Outcast
Loner
LOA
R1
Tad1
R2
RTE
L1
NeSL-1
R4
CRE
0.1 changes

for a given transposon family in eukaryotes, indicating functional importance. Within this megafamily, the eukaryotic DNA transposon families *Tc1* and *mariner* and the bacterial *IS630* element and its relatives in prokaryotes and ciliates comprise a superfamily, the *IS630-Tc1-mariner* superfamily, which is based on overall transposase similarities and a common TA dinucleotide insertion target (Henikoff, 1992; Doak *et al*., 1994; Robertson and Lampe, 1995; Capy *et al*., 1996; Shao and Tu, 2001). *Tc1*-like elements identified in fungi, invertebrates, and vertebrates all contain a DD34E motif while most *mariner* elements identified in flatworms, insects, and vertebrates contain a DD34D motif. A few TEs that contain DD37D and DD39D motifs were previously regarded as basal subfamilies, the max subfamily and mori subfamily respectively, of the *mariner* family (Robertson, 2002). A novel transposon, *ITmD37E*, was recently reported in a wide range of mosquito species (Shao and Tu, 2001). The *ITmD37E* transposases contain a conserved DD37E catalytic motif, which is unique among the reported transposons of the ITm superfamily. Sequence comparisons and phylogenetic analyses suggest that *ITmD37E* is a novel family (**Figure 8**). In addition, our phylogenetic analyses show that the mori subfamily (DD37D) and max subfamily (DD39D) of *mariner* may also be classified as two distinct families, namely the *ITmD37D* and *ITmD39D* families. In fact, *ITmD37D* (previously mori subfamily of *mariner*) is more closely related to *Tc1* (DD34E) than other *mariner* elements (**Figure 8**). The recognition of the three new families *ITmD37E, ITmD37D*, and *ITmD39D* is consistent with the fact that they share family-specific catalytic motifs and similar TIRs. Claudianos and colleagues also noticed the need for reclassification of the DD37D transposons and named them the *maT* family (Claudianos *et al*., 2002). Finally, a group of transposons that contain a DD41D catalytic motif have been found in the medfly *Ceratitis rosa*, establishing yet another family (Gomulski *et al*., 2001; Robertson and Walden, 2003), namely the *ITmD41D* family. In summary, according to recent analyses, the ITm superfamily can be organized in seven families including *ITmD37E*, *ITmD37D*, *ITmD39D*, *ITmD41D*, *Tc1*, *mariner*, *pogo*, and an unresolved clade which includes bacterial *IS630*-like elements and some fungal and ciliate transposons (**Figure 8**). *pogo* is an interesting case as it has a unique N-terminal DNA-binding domain and a long C-terminal domain rich in acidic residues, although it contains a DDxD catalytic domain related to the ITm transposons (Smit and Riggs, 1996).

4.12.4.3.1.2. Two new families: **PIF/Harbinger *and* Transib** *PIF/Harbinger* and *Transib* are two newly discovered families of DNA transposons in insects. *PIF* and related TEs *Harbinger* and *Tc8* have previously been found in plants and nematodes (Kapitonov and Jurka, 1999; Le *et al*., 2001; Zhang *et al*., 2001) and they may have mobilized *Tourist*, one of the most abundant MITEs in plants (Zhang *et al*., 2001; Jiang *et al*., 2003). Analysis of the *A. gambiae* genome uncovered approximately 30 families of *PIF/Harbinger*-like transposons, although the majority of them are "fossil"-like sequences (Biedler *et al*., unpublished data). Like all other *PIF/Harbinger*-like transposons, the *A. gambiae PIF/Harbinger*-like elements are characterized by 3 bp TSDs and transposase sequences similar to that of the bacterial *IS5* elements (Zhang *et al*., 2001). A second new family, the *Transib* transposons, are so far found only in *D. melanogster* and *A. gambiae* (Kapitonov and Jurka, 2003b). They use a unique transposase and generate 5 bp TSDs. *Transib* transposons have resided in these two genomes for a long time as extensive *in silico* reconstruction had to be employed to trace back the ancestral full-length copies (Kapitonov and Jurka, 2003b).

Figure 7 Phylogenetic analysis of diverse non-LTR retrotransposons found in *Anopheles gambiae*. Phylogenetic analysis classifies 104 families of *A. gambiae* elements into two new clades and eight previously defined clades. The two new clades, *Loner* and *Outcast*, are in bold. Shown here is the neighbor-joining tree constructed using alignment of approximately 260 amino acids of the reverse transcriptase (RT) domain from non-LTR retrotransposons of *A. gambiae* and representative elements in different clades. The tree was rooted using RTs of three prokaryotic Group II introns (not shown). Maximum parsimony was also used which produced a similar phylogenetic tree (not shown). Confidence of the groupings was estimated using 500 bootstrap replications for both methods. The first and second numbers at a particular node represent the bootstrap values derived from 500 neighbor-joining and maximum parsimony analysis, respectively. Only the values for the major groupings (clades) that are above 50% are shown. The scale at bottom left indicates amino acid divergence. The names of elements from previously established clades are given but names of new *A. gambiae* non-LTR families are omitted to save space (see Biedler and Tu, 2003). Previously reported *A. gambiae* non-LTRs in the tree are *Q*, *T1*, *RT1*, and *RT2* (Besansky, 1990; Besansky *et al*., 1992, 1994). (Reproduced with permission from Biedler, J., Tu, Z., **2003**. Non-LTR retrotransposons in the African malaria mosquito, *Anopheles gambiae*: unprecedented diversity and evidence of recent activity. *Mol. Biol. Evol. 20*, 1811–1825; © Oxford University Press.)

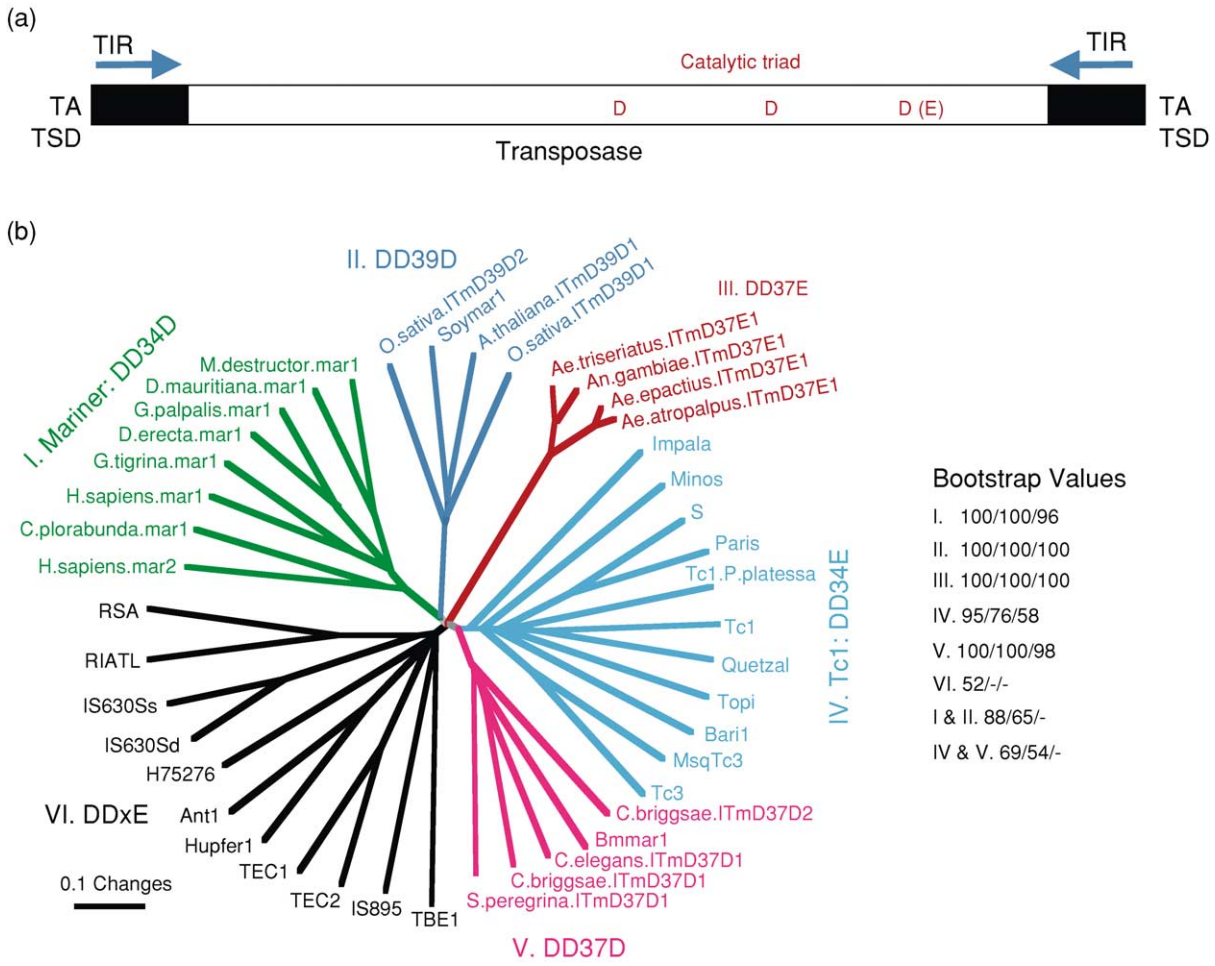


Figure 8 Structural features and classification of the *IS630-Tc1-mariner* superfamily. (a) Structural features. The catalytic triad in the transposase is highlighted. The characteristic TA target site duplications (TSDs) flanking an *IS630-Tc1-mariner* are shown. The terminal inverted repeats (TIRs) specify the boundries of the element. Possible introns are not shown. (b) Phylogenetic relationship between members of the *IS630-Tc1-mariner* superfamily on the basis of the catalytic domain. The alignment used here was previously described (Shao and Tu, 2001). The tree shown here is an unrooted phylogram constructed using a minimum evolution algorithm. Two additional methods, neighbor-joining and maximum parsimony, were also used. Confidence of the groupings was estimated using 500 bootstrap replications. The bootstrap value represents percent of times that branches were grouped together at a particular node. The first, second, and third numbers represent the bootstrap value derived from minimum evolution, neighbor-joining, and maximum parsimony analysis, respectively. Only the values for major groupings are shown. Various colors indicate different clades. All phylogenetic analyses were conducted using PAUP 4.0 b8 (Swofford, 2001). Note that a recently described group of transposons that contain a DD41D catalytic triad was not included here. They are a distinct group related to the DD37D transoposons (Gomulski *et al.*, 2001). (Modified from Shao, H., Tu, Z., **2001**. Expanding the diversity of the *IS630-Tc1-mariner* superfamily: discovery of a unique DD37E transposon and reclassification of the DD37D and DD39D transposons. *Genetics 159*, 1103–1115.)

4.12.4.3.2. MITEs As shown in Table 2, MITEs that share similar TSDs and TIRs with DNA transposons in the *IS630-Tc1-mariner, hAT, piggyBac, PIF/Harbinger* families have been found in three species of mosquitoes. A MITE that generates a specific TA TSD has also been reported in a coleopteran (Braquart *et al.*, 1999). Two *hAT*-like MITEs have been recently found in *D. willistoni* (Holyoake and Kidwell, 2003) and a deletion-derivative of the *pogo* transposon has been found in *D. melanogaster* (Feschotte *et al.*, 2002).

The relationship between *PIF/Harbinger*-like DNA transposons and related *Tourist*-like MITEs in *A. gambiae* is interesting (Biedler *et al.*, unpublished data). There are multiple *PIF/Harbinger*-like DNA transposon families in *A. gambiae*, all of which have the characteristic AT-rich 3 bp TSDs. Although the TIRs are highly conserved between different copies of a family, they are often variable between different *PIF/Harbinger*-like transposon families. Only two MITE families were found to be apparent deletion derivatives of two *PIF/Harbinger*-like

transposons (Figure 4). On the other hand, 20 families of the *Tourist*-like MITEs were found to share similar TSDs and TIRs with their respective *PIF/Harbinger*-like transposons although no similarities were found between the internal sequences of these MITEs and any of the *PIF/Harbinger*-like transposons. The implication of these results on the evolution of MITEs is discussed elsewhere (see Sections 4.12.2 and 4.12.6).

4.12.4.3.2.1. **m9bp,** ***MITEs with 9 bp TSDs*** Two novel groups of MITEs have been discovered in *Aedes aegypti*. One group, *m9bp*, includes two families of putative MITEs that have 9 bp TSDs (Mao and Tu, unpublished data). These two families are moderately repetitive and evidence of insertion has been found for both families. Several plant MITEs have 9 bp TSDs (Charrier *et al*., 1999; Feschotte *et al*., 2002) and they may be mobilized by *Mutator*-like transposons only found in plants (Walbot and Rudenko, 2002). The discovery of 9 bp TSD MITEs in *Aedes aegypti* is the first in animals, which may indicate a broader distribution of the *Mutator*-like transposons.

4.12.4.3.2.2. **m7bp,** ***MITEs with 7 bp TSDs*** Three families of putative MITEs that have 7 bp TSDs (*m7bp_1, m7bp_2, m7bp_3*) have been found in *Aedes aegypti* (Mao and Tu, unpublished data). These families are moderately repetitive and evidence of insertion has been found for all three families. These 7 bp TSD MITEs could potentially lead to the discovery of an entirely new family of eukaryotic DNA transposons, as 7 bp TSD elements have until now only been found in bacteria (Krebs *et al*., 1990; Mahillon and Chandler, 1998). No full-length protein-encoding DNA transposons that produce 7 bp TSDs have been identified in *A. aegypti* yet. However, this is within expectation considering the fragmented nature of the BAC-end sequences in the current *Aedes aegypti* database.

4.12.4.3.2.3. **Microuli,** ***a miniature subterminal inverted-repeat TE*** *Microuli* is a family of small (~200 bp) and highly AT rich (68.8–72.6%) TEs found in *Aedes aegypti* that do not have any coding capacity (Tu and Orphanidis, 2001). There is a 61 to 62 bp internal subterminal inverted repeat as well as a 7 bp subterminal inverted repeat 11 bp from the two termini. In addition, there are three imperfect subterminal direct repeats near the 5′ end. All of the above characteristics clearly resemble the structural features of MITEs. The only feature that separates *Microuli* from MITEs is that *Microuli* elements lack TIRs. Therefore, we use the phrase "miniature subterminal inverted-repeat transposable elements," or MSITEs, to refer to the structural characteristics of the *Microuli* elements. Short insertion sequences that contain subterminal inverted repeats but lack TIRs have been identified in the genomes of rice and a *Culex* mosquito (Song *et al*., 1998; Feschotte and Mouches, 2000a). Fourteen of the 19 nucleotides at the 5′ (and only 5′) terminus of *Microuli* are identical to the TIR of *Wuneng*, a previously characterized MITE in *Aedes aegypti* (Tu, 1997). Both *Microuli* and *Wuneng* insert specifically into the TTAA target. It has been suggested that MITEs and the autonomous DNA transposons share the same transposition machinery based on common TIRs (Feschotte *et al*., 2002). Then how did *Microuli* transpose without the TIRs? The three subterminal direct repeats could potentially be the binding sites for transposases because subterminal inverted repeats and subterminal direct repeats have been shown to bind transposases in several autonomous DNA transposons (Morgan and Middleton, 1990; Beall and Rio, 1997; Becker and Kunze, 1997). It remains unclear how the termini of *Microuli* are determined at the strand cleavage step without the TIR. The TTAA target duplication plus a 3 bp TIR are essential for the excision of the autonomous transposon *piggyBac* (Bauser *et al*., 1999). Therefore, it is possible that *Microuli* may also be able to use the TTAA target sequence as part of the signal for recombination. It is tempting to hypothesize that some MITEs could evolve from MSITEs through mutation and/or recombination events at the termini which would result in TIRs.

4.12.4.3.3. ***Helitrons*** *Helitrons* have been found in *D. melanogaster* and *Anopheles gambiae*, which use a rolling-circle mechanism of transposition (Kapitonov and Jurka, 2001; Kapitonov and Jurka, 2003b) (**Table 2**). Insect *Helitrons* has several characteristics including short specific terminal sequences (5′ TC and 3′ CTAG), a 3′ hairpin, and the lack of TSDs. Instead of a cut-and-paste transposase, *Helitron1* in *A. gambiae* encodes an intronless protein including domains similar to helicase and replication initiation protein. There are approximately 100 copies of *Helitron* elements in *A. gambiae* that form 10 distinct families (Kapitonov and Jurka, 2003b).

In summary, the diversity demonstrated by insect TEs encompasses all classes and types of eukaryotic elements. Analyses of insect TEs have more than once led to discoveries that broadly impacted the field of TE research. Because of extensive genetic studies, *D. melanogaster* has long served as the launching pad for the discovery of novel families

of TEs in eukaryotes. The advent of genomics has provided the opportunity for the discovery of novel TEs in a wide range of organisms including non-drosophilid insects. As demonstrated above, greater TE diversity may be revealed and new insights may be gained from these genomic analyses.

4.12.5. Search for Active TEs in Insects

Active TEs may be used as tools for genetic manipulation of insects in basic and applied research (see Section 4.12.9). In addition, the behavior of TEs in host genomes and their spread in natural populations may be studied by monitoring active TE families. It is therefore highly desirable to isolate active copies of TEs. As described above (see Section 4.12.3), TEs discovered from observations of genetic mutations tend to result from active transposition events. Although several active TEs were discovered in this manner, this discovery process relies heavily on fortuitous events. Several methods that, when used in concert, could provide a systematic approach to uncover active TEs in insect genomes are described below.

4.12.5.1. Identification of Potentially Active TEs on the Basis of Bioinformatics Analysis

As discussed above, the ongoing genome revolution has produced an immense amount of sequence data from which diverse TEs can be identified in various insect genomes. The computational programs described above (see Section 4.12.3) can greatly facilitate the discovery and characterization of a large number of TE families. Unfortunately, the vast majority of TEs have accumulated inactivating mutations during evolution, rendering the discovery of active TEs the difficult task of "finding needles in a haystack." Bioinformatics analysis can provide leads to potentially active candidates that can be studied further. For example, using a semi-automated reiterative search strategy, many potentially active families of non-LTR retrotransposons in the *A. gambiae* genome were identified (Biedler and Tu, 2003). Here candidate families were identified based on sequence characteristics, which include the presence of full-length elements, intact ORFs, multiple copies with high nucleotide identity, and the presence of TSDs. High nucleotide identity indicates recent amplification from a source element, without enough time for divergence caused by nucleotide substitution and other mutations. It should be emphasized that sequence analysis can only provide leads for further analysis. For example, high sequence identity between copies of a TE family may not always indicate recent transposition activity because it can also result from gene conversion events.

4.12.5.2. Detection of TE Transcription

Transcription is a required step during transposition of the RNA-mediated TEs. Although DNA-mediated TEs do not use RNA as an intermediate, transcription is required for production of transposase proteins. Therefore, the detection of transcription may offer further support for an active family in both classes of TEs. Transcription can be inferred if a match was found in an expressed sequence tag (EST) database to a TE sequence from the same organism. For example, 21 families of non-LTR retrotransposons had significant hits when BLAST searches were carried out against over 94 000 *A. gambiae* ESTs downloaded from NCBI (Biedler and Tu, 2003). Transcription of TEs can also be detected experimentally by real time polymerase chain reaction (RT-PCR), Northern blot, and even microarray. The source of mRNA may affect the outcome of these experiments because the activity of some TEs may be temporally and spatially controlled. It has been shown that TE activity can be elevated during the culturing of mammalian and plant cells (Wessler, 1996; Grandbastien, 1998; Liu and Wendel, 2000; Kazazian and Goodier, 2002). Different cell lines are available for a number of insect species. One caveat of the above approach is that, transcripts shown by either experimental detection or EST analysis could arise from spurious transcription. These transcripts could originate by transcription from a nearby host promoter.

4.12.5.3. Detection of Transposition Events by TE Display

TE display (Van den Broeck *et al.*, 1998; Casa *et al.*, 2000; Biedler *et al.*, 2003) is a sensentive and reproducible experimental method to detect TE insertions. We have used it recently to study insertion site polymorphisms of endogenous insect TEs (Biedler *et al.*, 2003) (see Sections 4.12.7 and 4.12.9). It is a modified form of amplified fragment length polymorphism (AFLP), the difference being that one primer is designed according to a TE sequence (**Figure 9**). First, genomic DNA is digested using a four-base restriction enzyme such as BfaI and then ligated to an adapter sequence. This is followed by two rounds of PCR. During the second PCR, a radioactive-labeled nested primer specific for the TE sequence is used following a touch-down protocol. After the second PCR, products are run on a sequencing gel and visualized by autoradiography. TE display is a powerful tool for genome-wide analysis of TE insertions and for detection of new insertions due to transposition (De Keukeleire *et al.*,

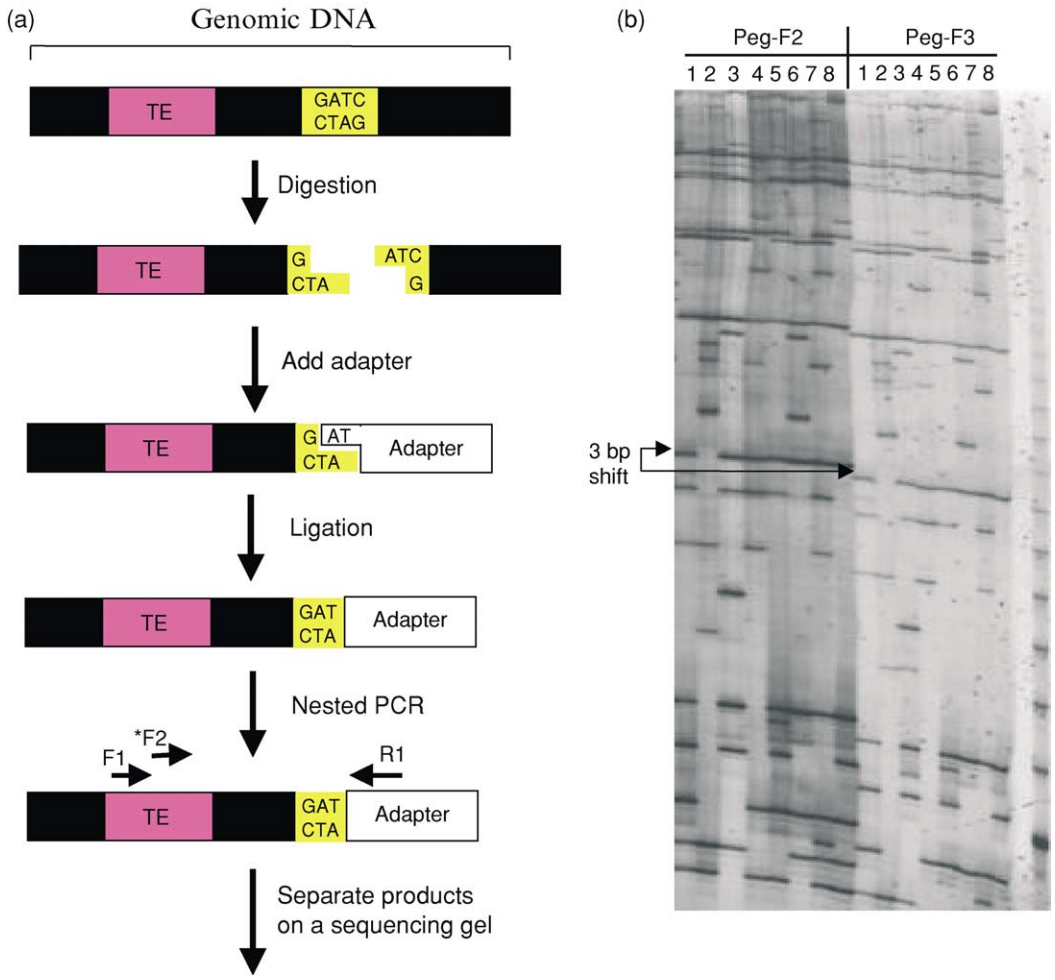


Figure 9 TE display, a method to scan multiple insertion sites of a TE in the genome. (a) Principle of TE display, which is a modified form of amplified fragment length polymorphism (AFLP). The difference is that TE-specific primers (F1 and F2) are used in addition to the adaptor primer (R1). F2 is labeled as shown by the asterisk. (b) Partial image of a TE display using primers for the *Pegasus* element with eight female individuals from an *Anopheles gambiae* colony (GAMCAM) originally collected from Cameroon (Biedler *et al.*, 2003). The eight samples on the left are amplified with a *Pegasus*-specific primer Peg-F2. The eight samples on the right are the same as those on the left except they were amplified with primer Peg-F3, which is designed to amplify a product smaller by three bases. The three base shift is clearly observable. A 10 bp ladder is shown on the right. Bands from a TE display gel were reamplified and sequenced, showing that they contained *Pegasus* sequences as well as flanking genomic and adapter sequences in the expected order (not shown). Comigrating bands among different individuals had the same flanking genomic sequence, indicating that they were from the same genomic locus. TE displays have been also developed using two highly reiterated SINEs, *SINE200* in *A. gambiae* and *Feilai* in *Aedes aegypti* (not shown). (Modified from Biedler, J., Qi, Y., Holligan, D., Della Torre, A., Wessler, S., *et al.*, **2003**. Transposable element (TE) display and rapid detection of TE insertion polymorphism in the *Anopheles gambiae* species complex. *Insect Mol. Biol. 12*, 211–216.)

2001). It offers a higher degree of sensitivity and resolution than genomic Southern blot analysis.

TE display has been used to detect somatic cell transposition (De Keukeleire *et al.*, 2001), simply by looking for the presence of new bands that represent newly transposed copies of a TE. A caveat of this approach is that a change in a restriction site may also result in new TE display bands. The same method may also be used to identify germline transposition by comparing TE display patterns of parent insects with the patterns of a large number of offspring. Alternatively, one could take advantage of the possibility that some TEs are activated in cell culture. Using TE display, one may be able to identify active families by comparing the relative abundance of a TE in cultured cells with that in individuals from different strains of the same species (Jiang *et al.*, 2003). This approach is based on the assumption that some TE families may be more active in cultured cells than in live organisms.

After TE display, one side of the new insertional copy and its associated flanking sequence can be recovered because the band can be reamplified and sequenced. To recover the entire sequence of the

newly inserted copy, one has to rely on mapping the insertion site to the genome if the genome sequence is available. Otherwise, further PCR or screening of a genomic library is necessary to recover the entire copy and both sides of the flanking sequence. The identified insertional copy itself may or may not be active because it could have been mobilized by a *trans*-acting protein that is encoded by an active TE in the genome. However, the experiment described above can demonstrate *in vivo* or *ex vivo* transposition events and the recovered insertional copy may lead to the identification of the autonomous active TE on the basis of shared *cis*-acting sequences such as TIRs.

4.12.5.4. Detection of Transposition Events by Inverse PCR

Actively transposing DNA-mediated TEs can be identified as extrachromsomal DNA in the form of linear or circular intermediates or byproducts (Arca *et al.*, 1997; Gorbunova and Levy, 1997) (see **Chapter 4.13**). Using a set of outward-orienting primers within the TE, the circular extrachromosomal copies may be amplified, which could serve as evidence of active excision or transposition. However, head-to-head copies of the same TE in the genome could also produce PCR products when outward-orienting primers are used, which must be ruled out by sequencing and further analysis. Extrachromosomal circles of *Hermes* (see **Chapter 4.13**) and *Pegasus* (Coy and Tu, unpublished data) have been identified. Imprecise excision had occurred when *Pegasus* is excised from its genomic location within the *A. gambiae* genome.

4.12.5.5. Transposition Assay, Reconstruction, and Genetic Screen

Transposition assays can be used to directly assess the functionalities of both the *cis*-(TIRs) and the *trans*- (transposase) components of a DNA transposon, allowing the demonstration of autonomous transposition events (see **Chapter 4.13**). In addition, transposition assays have also been established for the detection of retrotransposition of non-LTR retrotransposons (Jensen *et al.*, 1994; Ostertag *et al.*, 2000). Recently, a molecular reconstruction approach has been developed to restore inactivated copies of a vertebrate transposon, *Sleeping Beauty* (Ivics *et al.*, 1997), but such an approach requires extensive phylogenetic analysis and elaborate reconstructions (Ivics *et al.*, 1997). As an alternative, a genetic screen based on a bacterial system has been developed to identify hyperactive copies of an insect *mariner* transposon among randomly mutated copies (Lampe *et al.*, 1999). This approach can potentially be used to screen for active copies of transposons that do not require specific host factors. In summary, the progress in insect genome projects and the development and application of the methods described in this section will greatly facilitate searches for active TEs. The task of "finding needles in a haystack" could potentially be replaced by targeted and more efficient investigations.

4.12.6. Evolution of Insect TEs

The evolutionary dynamics of TEs are complex, which is in part due to their replication and their interactions with the host genome. The intricate dynamics between TEs and their host genomes are further complicated by the ability of some TEs to cross species barriers and spread in a new genome by horizontal transfer. Horizontal transfer may be an important part of the life cycle of some TEs and contribute to their continued success during evolution (Silva *et al.*, 2004). While the broad distribution of both RNA-mediated TEs and DNA-mediated TEs in all eukaryotic groups is evidence of the long-term evolutionary success of TEs, the two TE classes may have adopted different strategies, for which several insect TEs in both classes provide good examples.

4.12.6.1. Genomic Considerations of TE Evolution

It has been hypothesized that TE insertions may present three types of potentially deleterious effects including: (1) insertional mutagenesis which may disrupt gene function and/or regulation; (2) transcriptional/translational cost of the production of TE transcripts and proteins; and (3) ectopic recombination between homologous copies of TEs in different chromosomal locations that may result in duplication, deletion, and a new linkage relationship between genes (Nuzhdin, 1999; Bartolome *et al.*, 2002; Kidwell and Lisch, 2002; Rizzon *et al.*, 2002; Petrov *et al.*, 2003). The costs of having TEs may also include the cost associated with DNA replication when TEs occupy a large fraction of the genome. Obviously, these hypotheses are not mutually exclusive. This section discusses the intragenomic dynamics of TE–host interaction. The population dynamics affecting the spread of TEs in insects, which is also important for TE evolution, will be discussed below (see Section 4.12.7).

4.12.6.1.1. Self-regulation of insect TEs Both TE-driven mechanisms (self-regulation) and host-driven mechanisms (host control) have been shown to affect TE activities in insects. Self-regulation has been shown for *mariner* and *P* elements in *Drosophila*

(Hartl *et al*., 1997; Kidwell and Lisch, 2001). In the case of the *Drosophila P* element, self-regulation is achieved through the activities of at least two types of element-encoded repressors. In the case of *mariner*, several mechanisms may be involved including overproduction inhibition (an increase in the amount of transposase results in a decrease in net transposase activity), missense mutation effects (defective transposase encoded by missense copies interfering with functional transposase), and titration effects by inactive copies. In this regard, it is interesting to note that several hyperactive mutants of an active *mariner*, originally discovered in the horn fly, have been isolated (Lampe *et al*., 1999). This suggests that the horn fly *mariner* has not evolved for maximal activity.

4.12.6.1.2. Host control of TEs in insects Host control of TE activity can either be targeted at a particular family of TEs or a large group of TEs in general. Family-specific host control has been demonstrated in several cases (Labrador and Corces, 2002). The best example so far is the control of *gypsy* activity by a genetic locus *flamenco* in *D. melanogaster* (Bucheton, 1995). Homozygous female mutants of *flamenco*, which is an X-linked recessive gene that represses the transposition of *gypsy*, produce progeny that show high rates of *gypsy* transposition. The relief of the suppression of *gypsy* activity in *flamenco* mutants, which are also called permissive mutants, functions through maternal factors. Another example of host control is the *P* element which only transposes in the germline because correct splicing of the *P* transcript only occurs in germ cells. The tissue-specificity of the *P* element allows it to transmit efficiently to the next generation while minimizing potential damage to the host. Two host genes have been implicated in the inhibition of *P* activity in somatic cells (Siebel *et al*., 1992, 1995).

RNA interference (RNAi), a mechanism that confers posttranscriptional degradation of mRNA on the basis of homology to small fragments of double-stranded RNA, has been implicated as a host defense mechanism against a broad spectrum of TEs in the nematode *Caenorhabditis elegans* (Ketting *et al*., 1999; Tabara *et al*., 1999). RNAi has been shown to function in *Drosophila* and a few other insects as well (Misquitta *et al*., 1999; Hammond *et al*., 2000). The potential role of RNAi as a general control method against TE activity in *Drosophila* has recently been proposed on the basis of cosuppression of the *I* element by an increasing number of *I*-related transgenes (Jensen *et al*., 1999; Labrador and Corces, 2002). It has been proposed that RNAi has evolved as a host defense mechanism against the invasion by TEs and viruses (review: Fedoroff, 2002). In addition to posttranscriptional degradation, this defense system may also be involved in the formation of transcriptionally inactive heterochromatin where TEs and other repeats concentrate (Couzin, 2002).

4.12.6.1.3. Nonrandom distribution of insect TEs Patterns of nonrandom TE distribution have been shown in both *D. melanogaster* and *A. gambiae* (Bartolome *et al*., 2002; Holt *et al*., 2002; Kapitonov and Jurka, 2003b). TEs tend to accumulate in heterochromatin. Such a distribution bias could result either from preferential TE insertion, or selection against insertions in euchromatic regions, or both. Bartolome and colleagues suggested that the abundance of TEs is more strongly associated with local recombination rates (Bartolome *et al*., 2002), which are low in heterochromatic regions, rather than with gene density. They argue that this association is consistent with the hypothesis that selection against harmful effects of ectopic recombination is a major force opposing TE spread. However, selection against insertional mutagenesis is also at work as shown by the absence of insertions in coding regions. The insertional bias of *P* elements has been demonstrated recently during genome-scale *P* mutagenesis analysis (Spradling *et al*., 1995, 1999). Therefore insertion bias may contribute to the biased pattern of TE distribution in insects. A related topic here is the suggestion that concentrations of TE insertions in the *Drosophila* Y chromosome may have contributed to the evolutionary process leading to its inactivation (Labrador and Corces, 2002). It should be noted that not all TEs have a bias towards heterochromatic or recombination-deprived regions. On the basis of analysis of limited gene sequences, it was shown that *Aedes aegypti* MITEs tend to be associated with the noncoding regions within or near genes (Tu, 1997), which is similar to what has been observed for plant MITEs (Zhang *et al*., 2000).

4.12.6.1.4. Autonomous and nonautonomous TEs In addition to the genomic interactions described above (see Section 4.12.6.1.3), most TEs have to contend with the fact that defective copies are often generated during or after transposition. This process could contribute to self-regulation as discussed above (see Section 4.12.6.1.1). It can also lead to total inactivation and ultimate extinction of a TE as the inactive TE population eventually overwhelms the active copies (Eickbush and Malik, 2002). Therefore, the replicative ability that is responsible for the success of the TE may also lead

to its inactivation in a genome. Interestingly, some nonautonomous TEs have been very successful with regard to amplification, although the mechanisms that contribute to their success are not entirely clear. For example, SINEs found in insect genomes including *Aedes aegypti, Anopheles gambiae*, and *Bombyx mori* (**Table 1**) all contain thousands of copies. Similarly, most of the MITEs found in insects are also highly reiterated although low-copy-number families are also found (Tu, 1997, 2000, 2001a). The small size of MITEs and SINEs may confer less deleterious effects on the host, either because they are less efficient substrates for homologous recombination (Petrov *et al.*, 2003) or because their impact on neighboring genes may be less severe. Therefore reduced selection pressure as well as other properties inherent to MITEs and SINEs could contribute to their apparent success. It is a fascinating question as to how SINEs and MITEs affect the evolution of the autonomous TEs that mobilize them.

4.12.6.2. Vertical Transmission and Horizontal Transfer of Insect TEs

As described above (see Section 4.12.6.1), the replicative ability that is responsible for the success of a TE may in some cases lead to its inactivation by generating defective copies or by activating host control mechanisms. This is especially true for DNA transposons. Therefore, the ability to escape the above vertical inactivation by invading a new genome would greatly enhance the evolutionary success of DNA transposons. However, not all TEs have adopted this life cycle of invasion, amplification, senescence, and new invasion (Hartl *et al.*, 1997; Eickbush and Malik, 2002; Robertson, 2002) (Figure 10).

4.12.6.2.1. Detection of horizontal transfer The occurrence of horizontal transfer can be supported in various degrees by three types of evidence (Silva *et al.*, 2004). First, detection of elements with a high level of sequence similarity in divergent taxa will offer strong support for horizontal transfer although variable rates of sequence change should be considered. Second, detection of phylogenetic incongruence between TEs and their hosts will also provide relatively strong support for horizontal transfer. However, this alone will not be convincing, especially in light of the high levels of intragenomic diversity of TE families observed in insects. In other words, the existence of multiple TE lineages could confound the phylogenetic analysis as paralogous lineages may be treated as orthologous ones. Finally, horizontal transfer may be inferred when "patchy" distribution of a TE among closely related taxa is observed. This type of support is weak as loss of

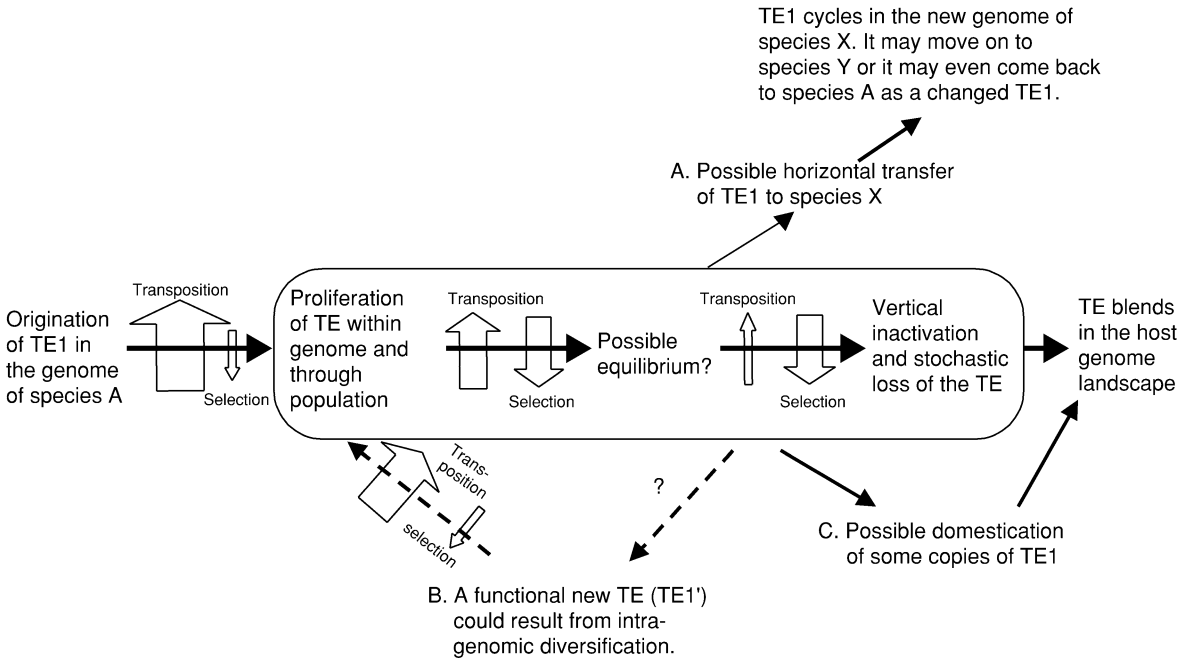


Figure 10 A model of the evolutionary dynamics of TEs in eukaryotic genomes. This hypothetical model incorporates recent work by several groups (Hartl *et al.*, 1997; Lampe *et al.*, 2001; Silva *et al.*, 2004). Some aspects of this model are better suited for DNA transposons that generally have a high propensity for horizontal transfer. Three possible alternatives to inactivation and stochostic loss are highlighted (A, B, and C).

TEs from sister taxa may result from a phenomenon similar to assortment of an ancestral polymorphism (Silva *et al*., 2004).

4.12.6.2.2. Horizontal transfer and vertical transmission in insects: differences between different groups of TEs The first case of eukaryotic horizontal transfer was reported in *Drosophila*, where the *P* element was shown to have invaded the *D. melanogaster* genome during the last century from a species in the *D. willistoni* group (review: Kidwell, 1992). Evidence for this lateral event includes all three types of support described above and is therefore widely accepted (Silva *et al*., 2004). Further analyses of a large number of *P* element sequences from a number of *Drosophila* species showed that many horizontal transfer events must have occurred to account for the current distribution pattern of *P* in *Drosophila* (Silva and Kidwell, 2000; Silva *et al*., 2004). Another spectacular case of horizontal transfer, which was also initially discovered in insects, involves the *mariner* transposons. The *mariner* transposon family has been implicated in hundreds or more horizontal transfer events among a wide range of animal species including a large number of insects across different orders (Robertson, 1993, 2002). It has been shown that the newly discovered *ITmD37E* transposons have also been involved in relatively frequent horizontal transfer events among different mosquito species (Shao and Tu, 2001, unpublished data). Finally, horizontal transfers of *hobo* and *piggyBac* has been shown in flies and moths (Bonnivard *et al*., 2000; Handler and McCombs, 2000). It has been proposed that horizontal transfer may be a key link in the life cycle of those DNA TEs that are not intricately associated with their host biology, allowing them to escape the vertical inactivation and stochastic loss (Hartl *et al*., 1997). Thus, these types of TEs may be regarded as "resident aliens" in a host (Plasterk *et al*., 1999). It has been shown that selection of some *mariner* transposases acts only during horizontal transfer, which is consistent with the invasion–inactivation–escape model (Lampe *et al*., 2003).

Horizontal transfer has also been shown for LTR retrotransposons in insects. A clear example involves the *copia* element in *Drosophila* (Jordan *et al*., 1999). The *copia* elements in *D. melanogaster* and *D. willistoni*, two divergent species that separated more than 40 million years ago, showed less than 1% nucleotide difference. It appears that *copia* jumped from *D. melanogaster* to *D. willistoni*. Horizontal transfer of the *Drosophila gypsy* element has also been reported (Terzian *et al*., 2000; Vazquez-Manrique *et al*., 2000).

Possible horizontal transfer of the *Drosophila* non-LTR retrotransposon *Jockey* was suggested on the basis of phylogenetic incongruence (Mizrokhi and Mazo, 1990). However, Malik and colleagues recently showed that an analysis that takes into account the rates of evolution was consistent with vertical transmission of the *Jockey* elements in *Drosophila* (Malik *et al*., 1999). Eickbush and Malik further suggest that no convincing evidence exists for the horizontal transfer of any non-LTR retrotransposon during the past 600 million years (Eickbush and Malik, 2002). In addition, strict vertical transmission of two non-LTR retrotransposons *R1* and *R2* have been shown in *Drosophila* (Eickbush and Eickbush, 1995). In these cases, the TE phylogenies well reflect those of the host species. Furthermore, it was shown that the relationship between *R2* elements from several insect orders including Diptera, Lepidoptera, Coleoptera, and Hymenoptera is consistent with vertical transmission (Eickbush, 2002).

4.12.6.2.3. Possible reasons for differing propensities for horizontal transfer Two reasons have been proposed to explain the apparent differences in the prevalence of horizontal transfer events between DNA transposons and non-LTR retrotransposons (Eickbush and Malik, 2002). The first is that DNA transposons need horizontal transfer for their long-term survival but non-LTR retrotransposons appear not to be dependent on such rare evolutionary events. Defective copies of DNA transposons retain the ability to be transposed as long as they have the *cis*-acting signals such as the TIRs. This indiscrimination leads to the inevitable fate of inactivation of the entire transposon family. Therefore, horizontal transfer offers a much-needed escape from the above vertical inactivation, which greatly enhances the evolutionary success of DNA transposons. On the other hand, it has been shown that the RT of non-LTR retrotransposons tends to associate with the mRNA molecules from which they were translated (Wei *et al*., 2001). This *cis*-preference would bias transposition events in favor of the active elements, thus providing a mechanism to sustain the non-LTR retrotransposons. However, the *cis*-preference is not enough to prevent the highly successful retrotransposition of SINEs in insects (Adams *et al*., 1986; Tu, 1999) and other organisms (Lander *et al*., 2001), which presumably borrows the retrotransposition machinery from non-LTR retrotransposons. It will be interesting to see how SINEs affect the evolution of their

non-LTR retrotransposon "partners." The second explanation is that DNA transposons may be more predisposed to horizontal transfer than non-LTR retrotransposons. The transposition process of DNA transposons involves an extrachromosomal DNA intermediate, which may facilitate horizontal transfer (Eickbush and Malik, 2002) (see Section 4.12.6.2.4). DNA transposons use their transposase for integration, which may be less dependent on host repair machinery than non-LTR retrotransposons and thus not as restricted to its original host.

LTR retrotransposons form an extrachromosomal DNA intermediate and use transposase-like IN for integration. Therefore, LTR retrotransposons have access to the same horizontal transfer mechanisms as the DNA transposons although their life cycle may not require horizontal transfer because defective copies are not thought to be a major factor (Eickbush and Malik, 2002). It should be noted that the above are general statements and that the propensity for horizontal transfer may vary among individual families within the three groups discussed here.

4.12.6.2.4. Mechanisms of horizontal transfer Mechanisms of horizontal transfer are poorly understood although direct transfer of the extrachromosomal DNA intermediate and indirect transfer through a viral vector have been proposed as possible mechanisms (Eickbush and Malik, 2002; Silva *et al*., 2004). Geographical and temporal overlap between the donor and recipient host species may be essential. An intriguing case of horizontal transfer of a *mariner* element between a parasitoid wasp and its lepidopteran host offers a good example of such overlap (Yoshiyama *et al*., 2001).

4.12.6.3. Other Possible Evolutionary Strategies

In addition to horizontal transfer and vertical extinction, recent studies suggest that there might be a third way, or an alternative strategy, which may be adopted by some TEs (Lampe *et al*., 2001). On the basis of the loss of interaction between *mariner* transposons of slightly changed TIRs, it was proposed that intraspecific or intragenomic diversification of *mariner* transposons may allow the newly diverged *mariner* to start a new lineage. Although this requires the coevolutionary events to occur in both the transposase and the TIRs, this scenario would provide the transposon the opportunity to escape vertical inactivation because it is now virtually a brand new element in a virgin genome because of the loss of interaction between itself and its relatives in the genome. Genome sequencing has provided increasing opportunity to survey the diversity of different families of TEs. Our recent analysis showed a large number of lineages of non-LTR retrotransposons of the *CR1* and *Jockey* clades in *A. gambiae* (Biedler and Tu, 2003) (**Figure 7**). Given the presence of multiple recently active lineages within the *CR1* and *Jockey* clades, it is tempting to speculate that the observed diversity may be driven by competition among different non-LTR families or by attempts to escape suppressive mechanisms imposed by the host.

On the other hand, some TEs are recruited for host functions and thus become "domesticated" (Lander *et al*., 2001; Kidwell and Lisch, 2002). This type of molecular domestication is the ultimate case of trading "freedom" for "security." It allows TEs to sustain and positively impact the host, examples of which will be discussed below (see Section 4.12.8). Strictly speaking, these domesticated TEs are no longer TEs. However, it is theoretically possible that these "domesticated" TEs could revert back to their "old ways" on rare occasions.

4.12.6.4. Understanding the Intragenomic Diversity of Insect TEs

High levels of TE diversity have been reported in the two insect genomes in which large-scale systematic analyses have been carried out (Holt *et al*., 2002; Kaminker *et al*., 2002; Kapitonov and Jurka, 2003b; Biedler and Tu, 2003) (**Figures 4** and **7**). The evolutionary process that generated this diversity may also be quite diverse. It is possible that the evolution of some TEs may be a complex mix of both vertical transmission and horizontal transfer events. Parsing out the results of intragenomic diversification from those of horizontal transfer events may require additional data from related species. Understanding the process responsible for the intragenomic diversity of insect TEs and the potential interactions between different TE families in insect genomes will be both challenging and rewarding. A summary of the current hypothesis on TE evolution is illustrated in **Figure 10**. Some aspects of this model are better suited for DNA transposons that have a high propensity for horizontal transfer.

4.12.7. TEs in Insect Populations

4.12.7.1. Fundamental Questions and Practical Relevance

In general, the increase of TE copy number through transposition is balanced by selective forces against the potential genetic load of TEs on host fitness (Nuzhdin, 1999). The control of transposition rate of TEs and other mechanisms to minimize their deleterious effects has been discussed (see Section 4.12.6). The population dynamics affecting the

spread of TEs in insect genomes is described. Earlier work on TEs in *Drosophila* populations suggest that the copy numbers in euchromatic regions are low and most euchromatic copies exist at very low frequency (<5%) in the population (Petrov *et al.*, 2003). This is interpreted as evidence of selection against individual TE copies in nature. It has been hypothesized that this selection is against three types of potentially overlapping deleterious effects by TEs including insertional mutagenesis, transcriptional or translational cost, and ectopic recombinations between similar copies of TEs in different chromosomal regions (Nuzhdin, 1999; Bartolome *et al.*, 2002; Kidwell and Lisch, 2002; Petrov *et al.*, 2003). One of the major questions in this field has been parsing out the main factors containing the spread of TEs in natural populations. From an applied perspective, TEs have been proposed as tools to genetically drive the spread of beneficial genes through insect populations to control infectious diseases (e.g., Ashburner *et al.*, 1998; Alphey *et al.*, 2002). For such a sophisticated approach to work, it is important to understand the population dynamics of TEs in their insect hosts.

4.12.7.2. Experimental Approaches

In situ hybridization of the *Drosophila* polytene chromosomes has been the main workhorse in studies of the population dynamics of different TE families (Charlesworth and Langley, 1989). Although extremely useful, this method obviously only works in species with accessible polytene chromosomes and it is not efficient in detecting short regions of sequence similarities to the probe (Petrov *et al.*, 2003). An alternative method, genomic Southern blotting, was also used to study TE insertions and excisions in *Drosophila* (Maside *et al.*, 2001). Although Southern blotting is a good method to estimate TE copy numbers, it is not reliable when the numbers are high. In addition, it may not be able to detect low-frequency sites in cases where multiple small insects have to be pooled to obtain enough high-quality genomic DNA. However, a single *Drosophila* can provide enough DNA for a Southern blot experiment.

TE display, a genome-scale detection method for TE insertions (see Section 4.12.5.3), has been used recently in *Drosophila* and two sibling species of mosquitoes *A. gambiae* and *A. arabiensis* (Biedler *et al.*, 2003; Yang and Nuzhdin, 2003). Because it is a PCR-based method, TE display allows investigation of multiple TE families using genomic DNA isolated from an individual insect. Although extremely powerful, this method cannot reliably distinguish homozygous from heterozygous insertions. However, TE display will allow recovery of a specific insertion site by sequencing the corresponding band. The sequence flanking the TE copy can be used to locate the specific site by searching the genome database if available or by inverse PCR. Therefore, sequences flanking a TE at a specific locus can be used as primers to amplify genomic DNA isolated from an individual sample (**Figure 11**). When the PCR products are run on an agarose gel, individuals with insertions at both alleles will show a single high molecular mass band while individuals with no insertions at either allele will give a single low molecular mass band (**Figure 11**). Individuals that have heterozygous alleles will yield both bands. Thus this locus-specific approach is codominant. When the genome sequence or a bulk of genomic sequences are available for an insect species, it is not absolutely necessary to couple locus-specific PCR with TE display because a number of TE insertion sites will already have been available for analysis (Petrov *et al.*, 2003). However, TE display can facilitate the investigation by providing a rapid scan of a large number of loci and by providing initial assessment of the level of polymorphism.

4.12.7.3. Recent Advances

Population studies, mainly of *Drosophila* TEs, suggest that selection against ectopic recombination may be the major factor in containing TE copy number in nature (Bartolome *et al.*, 2002; Kidwell and Lisch, 2002; Rizzon *et al.*, 2002; Petrov *et al.*, 2003). However, this issue apparently is quite complex (Biemont *et al.*, 1997). Although there appears to be evidence rejecting the effect of TE expression as a selection mechanism (Yang and Nuzhdin, 2003), the relative importance of insertional mutagenesis as a force to contain TE spread in *Drosophila* populations is still debated (Biemont *et al.*, 1997; Bartolome *et al.*, 2002; Petrov *et al.*, 2003).

TE insertions were shown to be generally at very low frequencies in natural populations of *D. melanogaster* (Charlesworth and Langley, 1989), which is consistent with the hypothesis of natural selection acting against the TE copies. However, recent surveys in *D. melanogaster* revealed an unexpected number of sites that are either fixed or of high frequency (Petrov *et al.*, 2003; Yang and Nuzhdin, 2003). In addition, an analysis of a retrotransposon *Osvaldo* in *D. buzzatii* (Labrador *et al.*, 1998) has shown a number of euchromatic sites with high occupancy in all populations from the Iberian Peninsula. The authors suggest that genetic drift is

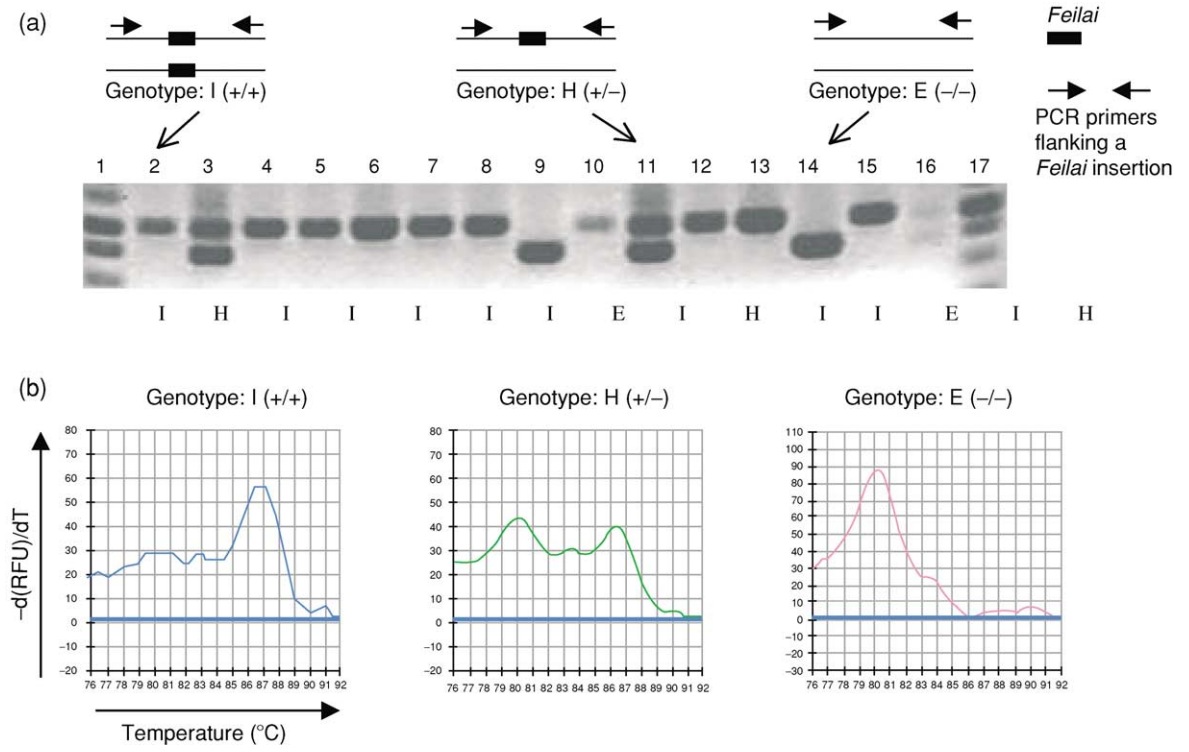


Figure 11 Locus-specific PCR for the detection of SINE insertion polymorphism. Shown here are images from analysis of a *Feilai* insertion in the yellow fever mosquito, *Aedes aegypti*. (a) Different genotypes revealed by PCR and agarose gel electrophoresis. Lanes 1 and 17 are size markers. Lanes 2–16 are PCR products from 15 individual mosquitoes collected in Thailand. Three genotypes and their corresponding banding patterns are shown. I, homozygous insertion (+/+); E, homozygous empty (−/−); H, heterozygotes (+/−). (b) The same three genotypes at the same *Feilai* insertion site detected using melt-curve analysis on an iCycler from Bio-Rad (Hercules, CA). The peaks represent PCR products. Higher melting temperature reflect larger and/or GC-rich PCR products. −d(RFU)/dT represents the rate of change of the relative fluorescence units (RFU) with time (T).

the main force responsible for the high frequency distribution of *Osvaldo* as a result of the founder effect. Using TE display, it has been found that fixed and high-occupancy sites of MITEs and SINEs are common in *A. gambiae* and *Aedes aegypti* strains and natural populations (Qi and Tu, unpublished data). The small size of MITEs and SINEs may confer less deleterious effects on the host either because they are less efficient substrates for ectopic recombination (Petrov *et al.*, 2003) or because their impact on neighboring genes may be less severe. Further analyses of different types of TEs in the natural populations of a wide range of insects will allow us to examine the validity and applicability of some of the fundamental conclusions drawn from the analysis of *Drosophila* TEs. The use of TE display and the advent of insect genomics provide an exciting opportunity for such investigations.

It may be reasonable to assume that TE copy number and insertion frequency in a population are highly dynamic parameters that can be influenced by TE-specific factors such as their intrinsic ability for transposition and self-regulation, by species- or genome-specific factors such as deletion and recombination rate and genomic control of transposition, and of course by effective population size and other ecological factors. As most population surveys only reflect a cross-section during evolution, the relative stage of a TE in its life cycle (**Figure 10**) is also an important consideration (Vieira *et al.*, 1999).

4.12.8. Impact of TEs in Insects

4.12.8.1. TEs and Genome Size and Organization

TEs are integral and significant components of eukaryotic genomes (Berg and Howe, 1989; Sherratt, 1995; Kidwell and Lisch, 2000; Craig *et al.*, 2002). For example, at least 46% of the human genome is TE-derived sequences (Lander *et al.*, 2001). In the two "completed" insect genomes, TEs occupy approximately 22% in *D. melanogaster* (Kapitonov and Jurka, 2003b) and at least 16% of the euchromatic region in *Anopheles gambiae* (Holt *et al.*, 2002). It has been proposed that the differing TE abundance may account for the "C value

paradox," which reflects the discrepancy between genome size or DNA content of an organism (as indicated by the C value) and its biological complexity (Kidwell, 2002). In other words, organisms with similar genetic/biological complexity may have huge variations in genome size due to differences in TE content. In higher eukaryotes, the interspersed repetitive sequences such as TEs are often organized in two distinct patterns relative to single copy DNA. One pattern of organization is the "short period interspersion" where single copy DNA is often interrupted every 1–3 kb by repetitive elements (Davidson *et al*., 1975). The other is the "long period interspersion" in which single copy DNA in the euchromatic region is less interrupted by repetitive sequences (Davidson *et al*., 1975). These two different patterns are probably shaped by the diversity, varied distribution, and abundance of repetitive sequences in the genome.

As noted above, TEs occupy approximately 22% of the *D. melanogaster* genome, which is threefold higher than previously reported (Kapitonov and Jurka, 2003b). Most TEs in *D. melanogaster* are relatively young (<20 million years old) and they consitute a major component of the paracentromeric regions. Detailed comparisons of TE age, distribution, and abundance are under way between the *D. melanogaster* and the *A. gambiae* genomes. Further comparisons that include other genomes such as *Aedes aegypti*, *Apis mellifera*, and *Bombyx mori* will be extremely informative. Some preliminary information gleaned from a comparative analysis of different species of mosquitoes are briefly described. Genome size and organization vary significantly between different mosquitoes. The genome of *Anopheles gambiae* is 270 Mb in size, and it is organized in a pattern of "long period interspersion" (Rai and Black, 1999). In contrast, the *Aedes aegypti* genome is 800 Mb in size, and it is organized in a pattern of mainly "short period interspersion" (Warren and Crampton, 1991). The relative abundance of MITEs and SINEs in these two mosquito genomes is interesting. The copy number of *Anopheles gambiae* MITEs ranges from 40 to 1340, much lower compared with 400 to 10 000 copies of MITEs in *Aedes aegypti* (Tu, 1997, 2000, 2001a). Similarly, the *Aedes aegypti* SINE family *Feilai* has 60 000 copies while *SINE200* in *Anopheles gambiae* has fewer than 6000 copies (Tu, 1999, unpublished data). The difference in the relative abundance of MITEs and SINEs may have contributed to the different organizations of the mosquito genomes and reflect different types of interactions between the hosts and these widespread TEs. Because of their small size, the copy number differences in MITEs may not explain the bulk of differences in the size of these two mosquito genomes. The relative abundance of other TEs, different deletion rates, simple repeat expansion, and large-segment duplication are among other possible contributing factors (Petrov, 2001). In *Anopheles gambiae*, TEs including MITEs are significantly more concentrated in heterochromatic regions (Holt *et al*., 2002). Alternatively, indications that a few families of MITEs are associated with the noncoding regions of genes in *Aedes aegypti* are presented (Tu, 1997). It is possible that unlike many other TEs, MITEs may be tolerated in gene-rich euchromatic regions in *A. aegypti*. MITEs have been found near plant genes where they could potentially affect gene regulation and/or define chromatin domains through structures such as matrix attachment regions (Wessler *et al*., 1995; Tikhonov *et al*., 2000). Varied amounts of repetitive elements have also been shown to be a major factor for the nearly threefold intraspecific differences of genome size in populations of the Asian tiger mosquito, *A. albopictus* (Rai and Black, 1999). Such an intraspecific variation provides a rare opportunity to study the contribution of TEs to genome evolution at the initial stage of the evolutionary process.

As noted before, TEs can induce chromosomal rearrangements through ectopic recombination and other mechanisms (Gray, 2000). Such rearrangements could result in gross reorganizations of the chromosomes, which may have a significant evolutionary impact as suggested by McClintock (McClintock, 1984). A DNA TE named *Odysseus* was found adjacent to the distal breakpoint of a naturally occurring paracentric chromosomal inversion that is characteristic of *Anopheles arabiensis*, one of the cryptic species in the *A. gambiae* complex (Mathiopoulos *et al*., 1998). Similar evidence of TE involvement in chromosomal rearrangement has been reported in *Drosophila* (Lim and Simmons, 1994; Caceres *et al*., 2001).

4.12.8.2. Evolutionary Impact

The tremendous potential for TEs to generate genetic diversity has long been recognized (McClintock, 1956). They can cause spontaneous mutation, recombination, chromosomal rearrangement, and hybrid dysgenesis (Bregliano *et al*., 1980; Engels, 1989; Shiroishi *et al*., 1993; Mathiopoulos *et al*., 1998). However, the long-term evolutionary impact of TEs is less clear. TEs have been generally regarded as "selfish" DNA since the early 1980s (Doolittle and Sapienza, 1980; Orgel and Crick, 1980) as opposed to the original "controlling elements"

hypothesis stating that TEs provide the physical basis controlling gene action and mutation (McClintock, 1956). This change of attitude was mainly due to the realization that TEs are somewhat "independent" genetic units and that their replicative ability allows them to spread even when they are not beneficial to the host organism. The question of whether the "selfish" TEs are just "junk DNA" to the host, or whether they can play important and even adaptive roles in organismal evolution is at the heart of the debate. Thanks in part to recent studies of insect TEs, it is increasingly clear that there may be a middle ground between the "junk DNA" and the "controlling elements" hypotheses. The host genome can be viewed as an ecological community with complex host–TE and TE–TE interactions (Brookfield, 1995; Kidwell and Lisch, 2000). Given the opportunistic nature of the evolutionary process, a "selfish" element could develop a wide spectrum of relationships with their host. They could be a "junk parasite," a "molecular symbiont," or something in between.

TE insertions are often under negative selection and their activities are under tight control by themselves and their host genomes (Bucheton, 1995; Hartl *et al*., 1997; Lin and Avery, 1999). Therefore the mere presence of TEs may represent a powerful genetic force with which the genome has been evolving. The possibility that RNAi has evolved as a host defense mechanism against the invasion by TEs and viruses and the potential involvement of RNAi in the formation of transcriptionally inactive heterochromatin are good examples of the host reacting to TE activities (Couzin, 2002). DNA transposons such as the *P* element in *Drosophila* can induce mutations by insertion as well imprecise excision in the coding or regulatory regions of genes (Kidwell and Lisch, 2002). On a larger scale, TE insertions can serve as substrates for homologous recombination that can result in chromosomal deletion, duplication, and inversion. Polymorphic chromosomal inversions are common in *Drosophila* and *A. gambiae*. TEs are often implicated in the generation of these inversions (Kidwell and Lisch, 2002) (see Section 4.12.8.1). It is proposed that these natural chromosomal inversions may result in reproductive isolation and perhaps genetic changes that allow *A. gambiae* to exploit a range of ecological niches (Mathiopoulos *et al*., 1999).

Recent evidence also indicates that some TEs may be co-opted to contribute to organismal biology. One of the early examples supporting this concept was the discovery that the telomere in *Drosophila* was maintained by site-specific insertions of two retrotransposons *HET-A* and *TART* (Biessmann *et al*., 1992; Levis *et al*., 1993). In addition, some TEs such as MITEs are often found near genes where they may change the expression profile of nearby host genes (Tu, 1997). The recruitment of TE insertions to produce diversity in gene expression profiles may have a better chance to confer a selective advantage when it occurs in duplicated genes. A fascinating genome-wide analysis showed that retrotransposition was responsible for the creation of a significant number of new functional genes in *Drosophila* (Betran *et al*., 2002). Finally, Some TE copies can be "domesticated" and take on a host function (see Section 4.12.6.3). For example, one of the *P* element repeats appears to have evolved the function of transcription factors (Miller *et al*., 1995). Other examples of "domesticated" TEs found in noninsect species include the *RAG1* and *RAG2* genes involved in V(D)J recombination in vertebrate lymphocytes and the more than 40 new genes derived from TEs in the human genome (Kidwell and Lisch, 2000; Lander *et al*., 2001; Gellert, 2002). On the basis of their broad distribution in bacteria, archea, and eukaryotes (Craig *et al*., 2002), it is safe to assume that TEs have long been, and will continue to be, evolving together with the immensely diverse life forms on this planet.

4.12.9. Applications of Insect TEs

4.12.9.1. Endogenous TEs and Genetic Manipulation of Insects

P element-based transgenic and mutagenesis tools in *D. melanogaster* have played a major role in the tremendous success of this tiny fly as a model organism for genetic analysis. A limited number of DNA transposons such as *hobo*, *mariner, minos,* and *piggyBac*, which have a broader host range than *P*, are being developed as tools in insects (see **Chapter 4.13**). Further analysis of TEs in insect genomes may expand the pool of active DNA transposons, which may be used to generate a set of tools with diverse features that can be used collectively for a variety of genetic analysis in different insects. In addition to simply transforming an insect, active TEs mentioned above are used to construct specific vectors to be used in gene trapping, enhancer trapping, and genome-wide insertional mutagenesis studies (Spradling *et al*., 1999; Klinakis *et al*., 2000; Horn *et al*., 2003). These analyses are powerful new ways to investigate gene function and regulation on a genome scale.

In addition to providing new active TEs to be used as tools for genetic manipulations of insects, a better understanding of endogenous TEs will allow

better-informed usage of current transposon-based genetic tools. Interactions between exogenous and endogenous transposons that share similar TIRs have been shown to be a potential problem (Sundararajan *et al.*, 1999; Jasinskiene *et al.*, 2000). Such interactions could be significant in light of the discovery of a diverse range of DNA TEs in a few insects in which genetic manipulation is actively pursued (Tu, 2001a; Holt *et al.*, 2002; Mao and Tu, unpublished data). For example, the diverse families of endogenous MITEs and DNA transposons could act as potential substrates for transposition if the introduced transposon uses similar TIRs. Thus, analyses of endogenous insect TEs will lead to better-informed design of transposon-based transformation tools that reduce instability resulting from interactions with endogenous TEs (Ashburner *et al.*, 1998; Atkinson *et al.*, 2001). It is also hoped that the non-Mendelian inheritance of TEs could help beneficial transgenes sweep through insect populations as the *P* element did in *Drosophila* (Ribeiro and Kidwell, 1994; Engels, 1997). Such a strategy is being investigated in the context of driving refractory genes into mosquito populations to control mosquito-borne infectious diseases (Ashburner *et al.*, 1998; Alphey *et al.*, 2002). A better understanding of endogenous TEs may be important to help achieve sustained success of such sophisticated genetic approaches.

4.12.9.2. SINE Insertion Polymorphism as Genetic Markers

The genetic differences and the pattern of gene flow between insect populations are of fundamental importance to a number of entomological questions ranging from evolution to practical applications. Single nucleotide polymorphisms (SNPs) are being developed as powerful markers for population genetic analysis, especially for insects with a large amount of sequence data available (Berger *et al.*, 2001). However, for most other insects, markers such as microsatellite, restriction fragment length polymorphism (RFLP) (Severson *et al.*, 1993; Yan *et al.*, 1999), cDNA-based single strand conformational polymorphism (SSCP) (Fulton *et al.*, 2001), mitochondrial (mtDNA), and ribosomal genes (rDNA) are still the tools of choice. For many insect species, finding a useful set of genetic markers is not trivial (Fagerberg *et al.*, 2001). It is clear that additional genome-wide markers are needed. Polymorphic insertion sites of interspersed TEs are potentially rich sources of genetic markers for population and genetic mapping studies. SINEs, which are especially useful for reasons described below are discussed. In population studies, sequences flanking a SINE at a specific locus are used as primers to amplify genomic DNA isolated from an individual sample. When the PCR products are run on an agarose gel, the genotype of an individual will be revealed on the basis of the number and size of bands (**Figure 11**) (see Section 4.12.7.2). Thus this locus-specific PCR assay may be used as codominant markers that reveal the dimorphism (insertion versus noninsertion) at a specific site. SINEs including the human *Alu* elements have been shown to be powerful genetic markers using the above method (Batzer *et al.*, 1994; Stoneking *et al.*, 1997; Carroll *et al.*, 2001; de Pancorbo *et al.*, 2001; Nasidze *et al.*, 2001; Roy-Engel *et al.*, 2001; Batzer and Deininger, 2002; Salem *et al.*, 2003). The ability of SINE insertion polymorphic markers to differentiate recently separated human populations is a good indication of their power (de Pancorbo *et al.*, 2001; Nasidze *et al.*, 2001; Watkins *et al.*, 2001). The locus-specific PCR assay of SINE insertions (**Figure 11**) has a few potential advantages over the popular microsatellite markers. The same SINE insertions are identical by descent. The probability that different SINEs of the same size independently insert into the same chromosomal location may be negligible (Stoneking *et al.*, 1997; York *et al.*, 1999; de Pancorbo *et al.*, 2001; Watkins *et al.*, 2001). However, the same signals in microsatellite, RFLP, SSCP, and melting curve single nucleotide polymorphism (McSNP) can only be considered identical by state, with the possibility of different origins. Moreover, the ancestral state of the SINE insertion polymorphism is likely to be the absence of a SINE because of general lack of excision, although exceptions do exist (Medstrand *et al.*, 2002). The ability to distinguish the ancestral versus derived states provides additional resolving power to address population genetic questions (York *et al.*, 1999). Because of the significant size difference of the PCR products and the dimorphic nature of the alleles, SINE insertions can be easily scored using agarose gels or melt-curve analysis, which can be developed for high-throughput analysis. Finally, a minimal amount of template is required for the PCR assay, allowing the use of a single insect for hundreds if not thousands of assays. One potential limitation of this approach could be the removal of a SINE insertion by rare recombination or gene conversion events, which may be confused with the noninsertion state. Such events may be revealed during the PCR analysis. Sequence analysis at the insertion site will also allow the investigation of this possibility. A better understanding of the evolutionary and population dynamics of SINEs will further provide theoretical foundations for this

potentially powerful method (Batzer and Deininger, 2002).

The recent development of a TE-anchored PCR approach, or TE display (see Section 4.12.5.3), has made it possible to directly screen for TE insertion polymorphism in a few species including insects (Roy *et al.*, 1999; Casa *et al.*, 2000, 2002; De Keukeleire *et al.*, 2001; Biedler *et al.*, 2003; Yang and Nuzhdin, 2003) (**Figure 9**). TE display efficiently scans a large number of loci in the genome, which makes it a very good tool for genotyping. However, as a population genetic tool, it has a major limitation common for dominant markers such as AFLP and randomly amplified polymorphic DNA (RAPD), namely the inability to distinguish between heterozygous and homozygous insertions, rendering the detection of population genetic structure difficult. However, TE display can be used as a direct screen to identify potential polymorphic insertion sites. Therefore TE display in conjunction with the development of locus-specific PCR markers that are codominant (**Figure 11**) will help TE insertion polymorphism markers reach their full potential as population genomic tools for insects. The use of TE insertion polymorphic markers in insect population studies is at an early stage. TE display and TE-based locus-specific PCR are used as genome-wide markers to investigate the genetic differences between different populations of *A. gambiae sensu stricto*, a species in the *A. gambiae* complex, and to study its ongoing speciation. These population genomic studies also have significant practical implications because genetic heterogeneity between mosquito populations can affect both the efficiency of disease transmission and the relative value of mosquito control measures (Coluzzi *et al.*, 2002; della Torre *et al.*, 2002). It was shown that different regions of the *A. gambiae* genome behave very differently with regard to introgression and gene flow (della Torre *et al.*, 2002). One important advantage of TE-based tools is the ability to scan a large number of TE insertions across different regions of the genome. Our initial TE display using *A. gambiae SINE200* showed significant differences at multiple sites between two forms of *A. gambiae sensu stricto*, namely the M and S forms (Coluzzi *et al.*, 2002). Polymorphic insertion sites for locus-specific PCR are isolated. Similar assays are also developed for *Aedes aegypti* using a SINE named *Feilai* (Tu, 1999). The development of TE-based population genomic tools for *A. aegypti* will be especially significant considering the lack of useful microsatellite markers in this species. The likely broad distribution of SINEs in insects makes them potentially rich sources of genetic markers for many different species.

4.12.9.3. SINE Insertions as Phylogenetic Markers

TE insertions, more specifically SINE insertions, have recently been used as molecular systematic tools to trace the evolutionary relationship between whales and Artiodactyla (Shimamura *et al.*, 1997; Nikaido *et al.*, 1999) and between salmonid fishes (Murata *et al.*, 1993). Perhaps the most impressive use of SINE insertions is the resolution of the evolutionary relationship of one of the major tribes of the African cichlid fishes which have evolved through an explosive adaptive radiation (Takahashi *et al.*, 1998; Terai *et al.*, 2003). To obtain TE insertion information for molecular systematics, locus-specific PCR described above is used. Here an RNA-mediated TE such as a SINE is again better suited because its transposition does not involve excison. Therefore, the ancestral state is known to be the absence of a SINE (Shedlock and Okada, 2000). The basic concept of this approach is illustrated in **Figure 12**. The insertion state can be easily determined using an agarose gel or melt-curve analysis as shown in **Figure 11**. When a fixed insertion is found at a particular site in species 1 and 2 but not in species 3, it can be inferred that species 1 and 2 are sister taxa. One of the requirements of the above approach is that the TE insertion site should be fixed within a species. TE display, a fingerprinting method described earlier (**Figure 9**), may be used to search for such sites. Potential problems such as nonspecific deletions, gene conversions, and sorting of ancestral polymorphisms can either be detected during the PCR and gel electrophoresis analysis or mitigated by surveying multiple loci (Shedlock and Okada, 2000). Sequences of the SINEs themselves in the loci used for the systematic analysis may provide further

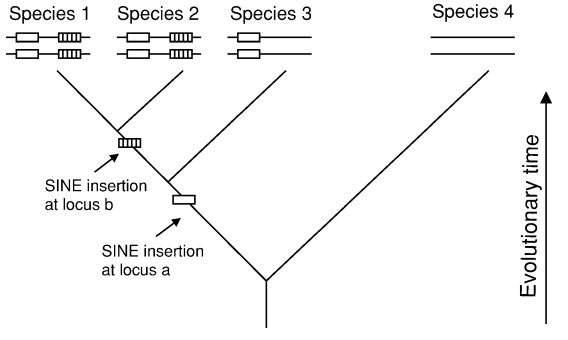


Figure 12 A schematic illustration of the principle of using SINE insertions as molecular systematic markers. Monophyletic relationship between species may be inferred on the basis of shared SINE insertions. See Section 4.12.9 for detailed explanations.

phylogenetic information. The usefulness of a particular SINE family in molecular systematics studies is dependent on its distribution and lifespan in the taxonomic group of interest. The tremendous diversity of insect species offers interesting challenges to evolutionary biologists. For example, a number of medically and economically important insect organisms exist as cryptic species complexes (Munstermann and Conn, 1997; Krzywinski and Besansky, 2003). New phylogenetic tools that can be integrated with methods using conventional characters such as morphology and DNA sequences will undoubtedly be of significance. Although SINEs have not been extensively studied in insects, highly repetitive SINEs have already been characterized in different orders including many medically and economically important species. Therefore SINEs may provide useful markers for molecular systematic analysis of insects, one of the most diverse groups of life forms on this planet.

In summary, TEs have been successfully used as vectors for genetic manipulation of insects and other organisms. A better understanding of insect TEs will allow better-informed usage of the currently available TE-based tools. The application of TEs as population and phylogenetic markers is at an early stage. Although these markers are promising tools, their scope of application, their resolving power and reliability depend on a better understanding of the population and evolutionary dynamics of the TEs. Therefore the fundamental studies dicussed in previous sections also have significant implications for the applications of TE-based molecular tools.

4.12.10. Summary

The last few years have witnessed an explosive growth both in the development of TE-based genetic and molecular tools and in our fundamental understanding of the diversity and impact of TEs in insects. Studies of TEs in insects, especially in *D. melanogaster*, have more than once led to discoveries that broadly impacted the field of TE research. The availability of new bioinformatics and experimental tools and the expanding genome revolution provide an exciting opportunity for the discovery of novel TEs in a wide range of insects and for the in-depth molecular and genomic analysis of these mobile genetic elements. Only through comparative genomic approaches, can we achieve a full appreciation of the complex and intricate dynamics governing the evolution of diverse TEs in insect genomes. From an applied perspective, systematic analysis of TEs in many insect species has significant economic and health implications because of the importance of these insects in disease transmission and agriculture.

Acknowledgments

I thank F. Collins, G. Gasperi, L. Gomulski, A. Handler, M. Kidwell, D. Lampe, H. Robertson, A. Sarkar, and J. Silva for sharing their manuscripts before publication. I thank H. Hagedorn, A. Handler, M. Kidwell, D. Lisch, S. Nuzhdin, and J. Silva for helpful comments. I thank J. Biedler, Monique Coy, Senay Sengul, and Rui Yang for help with the manuscript and for interesting discussions. Work in the author's laboratory is supported by NIH grants AI42121 and AI53203, Jeffress Foundation, and the Virginia Agricultural Experimental Station.

References

Abe, H., Ohbayashi, F., Shimada, T., Sugasaki, T., Kawai, S., *et al.*, **1998**. A complete full-length non-LTR retrotransposon, BMC1, on the W chromosome of the silkworm *Bombyx mori*. *Genes Genet. Syst. 73*, 353–358.

Abe, H., Ohbayashi, F., Shimada, T., Sugasaki, T., Kawai, S., *et al.*, **2000**. Molecular structure of a novel gypsy-Ty3-like retrotransposon (Kabuki) and nested retrotransposable elements on the W chromosome of the silkworm *Bombyx mori*. *Mol. Gen. Genet. 263*, 916–924.

Abe, H., Ohbayashi, F., Sugasaki, T., Kanehara, M., Terada, T., *et al.*, **2001**. Two novel Pao-like retrotransposons (Kamikaze and Yamato) from the silkworm species *Bombyx mori* and *B. mandarina*: common structural features of Pao-like elements. *Mol. Genet. Genomics 265*, 375–385.

Abe, H., Sugasaki, T., Kanehara, M., Ohbayashi, F., Shimada, T., *et al.*, **2002**. Nested retrotransposons on the W chromosome of the wild silkworm *Bombyx mandarina*. *Insect Mol. Biol. 11*, 307–314.

Adams, D.S., Eickbush, T.H., Herrera, R.J., Lizardi, P.M., **1986**. A highly reiterated family of transcribed oligo(A)-terminated, interspersed DNA elements in the genome of *Bombyx mori*. *J. Mol. Biol. 187*, 465–478.

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., *et al.*, **2000**. The genome sequence of *Drosophila melanogaster*. *Science 287*, 2185–2195.

Agarwal, M., Bensaadi, N., Salvado, J.C., Campbell, K., Mouches, C., **1993**. Characterization and genetic organization of full-length copies of a LINE retroposon family dispersed in the genome of *Culex pipiens* mosquitoes. *Insect Biochem. Mol. Biol. 23*, 621–629.

Alphey, L., Beard, C.B., Billingsley, P., Coetzee, M., Crisanti, A., *et al.*, **2002**. Malaria control with genetically manipulated insect vectors. *Science 298*, 119–121.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., *et al.*, **1997**. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res. 25*, 3389–3402.

Anzai, T., Takahashi, H., Fujiwara, H., **2001**. Sequence-specific recognition and cleavage of telomeric repeat (TTAGG)(n) by endonuclease of non-long terminal repeat retrotransposon TRAS1. *Mol. Cell Biol. 21*, 100–108.

Arca, B., Zabalou, S., Loukeris, T.G., Savakis, C., **1997**. Mobilization of a *Minos* transposon in *Drosophila melanogaster* chromosomes and chromatid repair by heteroduplex formation. *Genetics 145*, 267–279.

Arends, H.M., Jehle, J.A., **2002**. Homologous recombination between the inverted terminal repeats of defective transposon TCp3.2 causes an inversion in the genome of *Cydia pomonella* granulovirus. *J. Gen. Virol. 83*, 1573–1578.

Arkhipova, I.R., Pyatkov, K.I., Meselson, M., Evgen'ev, M.B., **2003**. Retroelements containing introns in diverse invertebrate taxa. *Nat. Genet. 33*, 123–124.

Ashburner, M., Hoy, M.A., Peloquin, J.J., **1998**. Prospects for the genetic transformation of arthropods. *Insect Mol. Biol. 7*, 201–213.

Atkinson, P.W., Pinkerton, A.C., O'Brochta, D.A., **2001**. Genetic transformation systems in insects. *Annu. Rev. Entomol. 46*, 317–346.

Badge, R.M., Brookfield, J.F., **1997**. The role of host factors in the population dynamics of selfish transposable elements. *J. Theoret. Biol. 187*, 261–271.

Bao, Z., Eddy, S.R., **2002**. Automated *de novo* identification of repeat sequence families in sequenced genomes. *Genome Res. 12*, 1269–1276.

Bartolome, C., Maside, X., Charlesworth, B., **2002**. On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster. Mol. Biol. Evol. 19*, 926–937.

Batzer, M.A., Deininger, P.L., **2002**. *Alu* repeats and human genomic diversity. *Nat. Rev. Genet. 3*, 370–379.

Batzer, M.A., Stoneking, M., Alegria-Hartman, M., Bazan, H., Kass, D.H., *et al.*, **1994**. African origin of human-specific polymorphic *Alu* insertions. *Proc. Natl Acad. Sci. USA 91*, 12288–12292.

Bauser, C.A., Elick, T.A., Fraser, M.J., **1999**. Proteins from nuclear extracts of two lepidopteran cell lines the ends of TTAA-specific transposons *piggyBac* and *tagalong*. *Insect Mol. Biol. 8*, 223–230.

Beall, E.L., Rio, D.C., **1997**. *Drosophila P*-element transposase is a novel site-specific. *Genes Devel. 11*, 2137–2151.

Becker, H.A., Kunze, R., **1997**. Maize *Activator* transposase has a bipartite DNA binding domain recognizes subterminal sequences and the terminal inverted. *Mol. Gen. Genet. 254*, 219–230.

Beeman, R.W., Thomson, M.S., Clark, J.M., DeCamillis, M.A., Brown, S.J., *et al.*, **1996**. *Woot*, an active gypsy-class retrotransposon in the flour beetle, *Tribolium castaneum*, is associated with a recent mutation. *Genetics 143*, 417–426.

Bensaadi-Merchermek, N., Cagnon, C., Desmons, I., Salvado, J.C., Karama, S., *et al.*, **1997**. *CM-gag*, a transposable-like element reiterated in the genome of *Culex pipiens* mosquitoes, contains only a *gag* gene. *Genetica 100*, 141–148.

Berezikov, E., Bucheton, A., Busseau, I., **2000**. A search for reverse transcriptase-coding sequences reveals new non-LTR retrotransposons in the genome of *Drosophila melanogaster. Genome Biol. 1*, RESEARCH0012.

Berg, D.E., Howe, M.J. (Eds.), **1989**. Mobile DNA. American Society for Microbiology Press, Washington, DC.

Berger, J., Suzuki, T., Senti, K.A., Stubbs, J., Schaffner, G., *et al.*, **2001**. Genetic mapping with SNP markers in *Drosophila*. *Nat. Genet. 29*, 475–481.

Bernstein, M., Lersch, R.A., Subrahmanyan, L., Cline, T.W., **1995**. Transposon insertions causing sex-lethal activity in *Drosophila melanogaster* affect Sx1 sex-specific transcript splicing. *Genetics 139*, 631–648.

Besansky, N.J., **1990**. A retrotransposable element from the mosquito *Anopheles gambiae*. *Mol. Cell Biol. 10*, 863–871.

Besansky, N.J., Bedell, J.A., Mukabayire, O., **1994**. *Q*: a new retrotransposon from the mosquito *Anopheles gambiae*. *Insect Mol. Biol. 3*, 49–56.

Besansky, N.J., Mukabayire, O., Bedell, J.A., Lusz, H., **1996**. *Pegasus*, a small terminal inverted repeat transposable element found in the *white* gene of *Anopheles gambiae*. *Genetica 98*, 119–129.

Besansky, N.J., Paskewitz, S.M., Hamm, D.M., Collins, F.H., **1992**. Distinct families of site-specific retrotransposons occupy identical positions in the rRNA genes of *Anopheles gambiae*. *Mol. Cell Biol. 12*, 5102–5110.

Betran, E., Thornton, K., Long, M., **2002**. Retroposed new genes out of the X in *Drosophila*. *Genome Res. 12*, 1854–1859.

Biedler, J., Qi, Y., Holligan, D., Della Torre, A., Wessler, S., *et al.*, **2003**. Transposable element (TE) display and rapid detection of TE insertion polymorphism in the *Anopheles gambiae* species complex. *Insect Mol. Biol. 12*, 211–216.

Biedler, J., Tu, Z., **2003**. Non-LTR retrotransposons in the African malaria mosquito, *Anopheles gambiae*: unprecedented diversity and evidence of recent activity. *Mol. Biol. Evol. 20*, 1811–1825.

Biemont, C., Tsitrone, A., Vieira, C., Hoogland, C., **1997**. Transposable element distribution in *Drosophila*. *Genetics 147*, 1997–1999.

Biessmann, H., Valgeirsdottir, K., Lofsky, A., Chin, C., Ginther, B., *et al.*, **1992**. HeT-A, a transposable element specifically involved in "healing" broken chromosome ends in *Drosophila melanogaster. Mol. Cell. Biol. 12*, 3910–3918.

Bigot, Y., Hamelin, M.H., Capy, P., Periquet, G., **1994**. *Mariner*-like elements in hymenopteran species: insertion site and distribution. *Proc. Natl Acad. Sci. USA 91*, 3408–3412.

Blinov, A.G., Sobanov, Y.V., Gaidamakova, E.K., Bogachev, S.S., Kolesnikov, N.N., *et al.*, **1991**. MEC: a transposable element from *Chironomus thummi* (Diptera). *Mol. Gen. Genet. 229*, 152–154.

Blinov, A.G., Sobanov, Y.V., Scherbik, S.V., Aimanova, K.G., **1997**. The *Chironomus* (*Camptochironomus*) *tentans* genome contains two non-LTR retrotransposons. *Genome 40*, 143–150.

Bonnivard, E., Bazin, C., Denis, B., Higuet, D., **2000**. A scenario for the *hobo* transposable element invasion, deduced from the structure of natural populations of *Drosophila melanogaster* using tandem TPE repeats. *Genet. Res. 75*, 13–23.

Booth, D.R., Ready, P.D., Smith, D.F., **1994**. Isolation of non-LTR retrotransposon reverse transcriptase-like sequences from phlebotomine sandflies. *Insect Mol. Biol. 3*, 89–96.

Bradfield, J.Y., Locke, J., Wyatt, G.R., **1985**. An ubiquitous interspersed DNA sequence family in an insect. *DNA 4*, 357–363.

Braquart, C., Royer, V., Bouhin, H., **1999**. *DEC*: a new miniature inverted-repeat transposable element from the genome of the beetle *Tenebrio molitor. Insect Mol. Biol. 8*, 571–574.

Bregliano, J.C., Picard, G., Bucheton, A., Pelisson, A., Lavige, J.M., *et al.*, **1980**. Hybrid dysgenesis in *Drosophila melanogaster. Science 207*, 606–611.

Brookfield, J.F., **1995**. Transposable elements as selfish DNA. In: Sherratt, D.J. (Ed.), Mobile Genetic Elements. Oxford University Press, Oxford, pp. 130–153.

Bucheton, A., **1995**. The relationship between the *flamenco* gene and *gypsy* in tame a retrovirus. *Trends Genet. 11*, 349–353.

Caceres, M., Puig, M., Ruiz, A., **2001**. Molecular characterization of two natural hotspots in the *D. buzzatii* genome induced by transposon insertions. *Genome Res. 11*, 1353–1364.

Calvi, B.R., Hong, T.J., Findley, S.D., Gelbart, W.M., **1991**. Evidence for a common evolutionary origin of inverted repeat transposons in *Drosophila* and plants: *hobo*, *Activator*, and *Tam3*. *Cell 66*, 465–471.

Capy, P., Langin, T., Higuet, D., Maurer, P., Bazin, C., **1997**. Do the integrases of LTR-retrotransposons and class II element transposases have a common ancestor? *Genetica 100*, 63–72.

Capy, P., Vitalis, R., Langin, T., Higuet, D., Bazin, C., **1996**. Relationships between transposable elements based upon the integrase-transposase domains: is there a common ancestor? *J. Mol. Evol. 42*, 359–368.

Carroll, M.L., Roy-Engel, A.M., Nguyen, S.V., Salem, A.H., Vogel, E., *et al.*, **2001**. Large-scale analysis of the *Alu Ya5* and *Yb8* subfamilies and their contribution to human genomic diversity. *J. Mol. Biol. 311*, 17–40.

Cary, L.C., Goebel, M., Corsaro, B.G., Wang, H.G., Rosen, E., *et al.*, **1989**. Transposon mutagenesis of baculoviruses: analysis of *Trichoplusia ni* transposon *IFP2* insertions within the *FP*-locus of nuclear polyhedrosis viruses. *Virology 172*, 156–169.

Casa, A.M., Brouwer, C., Nagel, A., Wang, L., Zhang, Q., *et al.*, **2000**. Inaugural article: the MITE family *heartbreaker* (*Hbr*): molecular markers in maize. *Proc. Natl Acad. Sci. USA 97*, 10083–10089.

Casa, A.M., Mitchell, S.E., Smith, O.S., Register, J.C., III, Wessler, S.R., *et al.*, **2002**. Evaluation of *Hbr* (MITE) markers for assessment of genetic relationships among maize (*Zea mays* L.) inbred lines. *Theoret. Appl. Genet. 104*, 104–110.

Charlesworth, B., Langley, C.H., **1989**. The population genetics of *Drosophila* transposable elements. *Annu. Rev. Genet. 23*, 251–287.

Charrier, B., Foucher, F., Kondorosi, E., d'Aubenton-Carafa, Y., Thermes, C., *et al.*, **1999**. *Bigfoot*, a new family of MITE elements characterized from the genus. *Plant J. 18*, 431–441.

Clark, J.B., Kidwell, M.G., **1997**. A phylogenetic perspective on *P* transposable element evolution in *Drosophila*. *Proc. Natl Acad. Sci. USA 94*, 11428–11433.

Claudianos, C., Brownlie, J., Russell, R., Oakeshott, J., Whyard, S., **2002**. *maT*: a clade of transposons intermediate between *mariner* and *Tc1*. *Mol. Biol. Evol. 19*, 2101–2109.

Coates, C.J., Johnson, K.N., Perkins, H.D., Howells, A.J., O'Brochta, D.A., *et al.*, **1996**. The hermit transposable element of the Australian sheep blowfly, *Lucilia cuprina*, belongs to the *hAT* family of transposable elements. *Genetica 97*, 23–31.

Cockburn, A.F., Mitchell, S.E., **1989**. Repetitive DNA interspersion patterns in Diptera. *Arch. Insect Biochem. Physiol. 10*, 105–113.

Coluzzi, M., Sabatini, A., della Torre, A., Di Deco, M.A., Petrarca, V., **2002**. A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science 298*, 1415–1418.

Cook, J.M., Martin, J., Lewin, A., Sinden, R.E., Tristem, M., **2000**. Systematic screening of *Anopheles mosquito* genomes yields evidence for a major clade of *Pao*-like retrotransposons. *Insect Mol. Biol. 9*, 109–117.

Couzin, J., **2002**. Breakthrough of the year: small RNAs make big splash. *Science 298*, 2296–2297.

Craig, N., **2002**. Mobile DNA: an introduction. In: Craig, N., Craigie, R., Gellert, M., Lambowitz, A. (Eds.), Mobile DNA II. American Society for Microbiology Press, Washington, DC, pp. 3–11.

Craig, N., Craigie, R., Gellert, M., Lambowitz, A. (Eds.), **2002**. Mobile DNA II. American Society for Microbiology Press, Washington, DC.

Daniels, S.B., Chovnick, A., Boussy, I.A., **1990**. Distribution of *hobo* transposable elements in the genus *Drosophila*. *Mol. Biol. Evol. 7*, 589–606.

Davidson, E.H., Galau, G.A., Angerer, R.C., Britten, R.J., **1975**. Comparative aspects of DNA organization in Metazoa. *Chromosoma 51*, 253–259.

Davis, P.S., Judd, B.H., **1995**. Molecular characterization of the 3C region between white and roughest loci of *Drosophila melanogaster. Drosophila Inf. Serv. 76*, 130–134.

De Keukeleire, P., Maes, T., Sauer, M., Zethof, J., Van Montagu, M., *et al.*, **2001**. Analysis by transposon display of the behavior of the *dTph1* element family during ontogeny and inbreeding of *Petunia hybrida*. *Mol. Genet. Genomics 265*, 72–81.

de Pancorbo, M.M., Lopez-Martinez, M., Martinez-Bouzas, C., Castro, A., Fernandez-Fernandez, I., *et al.*, **2001**. The Basques according to polymorphic *Alu* insertions. *Hum. Genet. 109*, 224–233.

Deininger, P., Roy-Engel, A., **2002**. Mobile elements in plant and animal genomes. In: Craig, N., Craigie, R., Gellert, M., Lambowitz, A. (Eds.), Mobile DNA II. American Society for Microbiology Press, Washington, DC, pp. 1074–1092.

della Torre, A., Costantini, C., Besansky, N.J., Caccone, A., Petrarca, V., *et al.*, **2002**. Speciation within *Anopheles gambiae*: the glass is half full. *Science 298*, 115–117.

Doak, T.G., Doerder, F.P., Jahn, C.L., Herrick, G., **1994**. A proposed superfamily of transposase genes: transposon-like ciliated protozoa and a common "D35E" motif. *Proc. Natl Acad. Sci. USA 91*, 942–946.

Doolittle, W.F., Sapienza, C., **1980**. Selfish genes, the phenotype paradigm and genome evolution. *Nature 284*, 601–603.

Eickbush, D.G., Eickbush, T.H., **1995**. Vertical transmission of the retrotransposable elements *R1* and *R2* during the evolution of the *Drosophila melanogaster* species subgroup. *Genetics 139*, 671–684.

Eickbush, T., **2002**. *R2* and related site-specific non-long terminal repeat retrotransposons. In: Craig, N., Craigie, R., Gellert, M., Lambowitz, A. (Eds.), Mobile DNA II. American Society for Microbiology Press, Washington, DC, pp. 813–835.

Eickbush, T., Malik, H., **2002**. Origins and evolution of retrotransposons. In: Craig, N., Craigie, R., Gellert, M., Lambowitz, A. (Eds.), Mobile DNA II. American Society for Microbiology Press, Washington, DC, pp. 1111–1144.

Engels, W.R., **1989**. *P* elements in *Drosophila melanogaster.* In: Berg, D.E., Howe, M.J. (Eds.), Mobile DNA. American Society for Microbiology Press, Washington, DC, pp. 437–484.

Engels, W.R., **1997**. Invasions of *P* elements. *Genetics 145*, 11–15.

Evgen'ev, M.B., Zelentsova, H., Shostak, N., Kozitsina, M., Barskyi, V., *et al.*, **1997**. *Penelope*, a new family of transposable elements and its possible role in hybrid dysgenesis in *Drosophila virilis*. *Proc. Natl Acad. Sci. USA 94*, 196–201.

Fagerberg, A.J., Fulton, R.E., Black, W.C., **2001**. Microsatellite loci are not abundant in all arthropod genomes: analyses in the hard tick, *Ixodes scapularis* and the yellow fever mosquito, *Aedes aegypti*. *Insect Mol. Biol. 10*, 225–236.

Fawcett, D.H., Lister, C.K., Kellett, E., Finnegan, D.J., **1986**. Transposable elements controlling I-R hybrid dysgenesis in *D. melanogaster* are similar to mammalian LINEs. *Cell 47*, 1007–1015.

Fedoroff, N., **1989**. Maize transposable elements. In: Berg, D.E., Howe, M.J. (Eds.), Mobile DNA. American Society for Microbiology Press, Washington, DC, pp. 375–411.

Fedoroff, N., **2002**. Control of mobile DNA. In: Craig, N., Craigie, R., Gellert, M., Lambowitz, A. (Eds.), Mobile DNA II. American Society for Microbiology Press, Washington, DC, pp. 997–1007.

Felger, I., Hunt, J.A., **1992**. A non-LTR retrotransposon from the Hawaiian *Drosophila*: the *LOA* element. *Genetica 85*, 119–130.

Feschotte, C., Fourrier, N., Desmons, I., Mouches, C., **2001**. Birth of a retroposon: the *Twin* SINE family from the vector mosquito *Culex pipiens* may have originated from a dimeric tRNA precursor. *Mol. Biol. Evol. 18*, 74–84.

Feschotte, C., Mouches, C., **2000a**. Recent amplification of miniature inverted-repeat transposable elements in the vector mosquito *Culex pipiens*: characterization of the *Mimo* family. *Gene 250*, 109–116.

Feschotte, C., Mouches, C., **2000b**. Evidence that a family of miniature inverted-repeat transposable elements (MITEs) from the *Arabidopsis thaliana* genome has arisen from a *pogo*-like DNA transposon. *Mol. Biol. Evol. 17*, 730–737.

Feschotte, C., Swamy, L., Wessler, S.R., **2003**. Genome-wide analysis of *mariner*-like transposable elements in rice reveals complex relationships with stowaway miniature inverted repeat transposable elements (MITEs). *Genetics 163*, 747–758.

Feschotte, C., Zhang, X., Wessler, S., **2002**. Miniature inverted-repeat transposable elements and their relationship to established DNA transposons. In: Craig, N., Craigie, R., Gellert, M., Lambowitz, A. (Eds.), Mobile DNA II. American Society for Microbiology Press, Washington, DC, pp. 1147–1158.

Finnegan, D.J., **1989**. The *I* factor and *I-R* hybrid dysgenesis in *Drosophila melanogaster.* In: Berg, D.E., Howe, M.J. (Eds.), Mobile DNA. American Society for Microbiology Press, Washington, DC, pp. 503–517.

Finnegan, D.J., **1992**. Transposable elements. *Curr. Opin. Genet. Devel. 2*, 861–867.

Franz, G., Savakis, C., **1991**. *Minos*, a new transposable element from *Drosophila hydei*, is a member of the *Tc1*-like family of transposons. *Nucl. Acids Res. 19*, 6646.

Fraser, M.J., **2000**. The TTAA-specific family of transposable elements: identification, functional characterization, and utility for transformation of insects. In: Handler, A.M., James, A.A. (Eds.), Insect Transgenesis: Methods and Applications. CRC Press, Boca Raton, FL, p. 249.

Friesen, P.D., Nissen, M.S., **1990**. Gene organization and transcription of TED, a lepidopteran retrotransposon integrated within the baculovirus genome. *Mol. Cell Biol. 10*, 3067–3077.

Fulton, R.E., Salasek, M.L., DuTeau, N.M., Black, W.C.T., **2001**. SSCP analysis of cDNA markers provides a dense linkage map of the *Aedes aegypti* genome. *Genetics 158*, 715–726.

Gale, K.R., **1987**. Characterization of the mosquito, *Aedes aegypti*, at the molecular level of genetics. Ph.D. thesis, University of Liverpool.

Garcillan-Barcia, M.P., Bernales, I., Mendiola, M.V., De La Cruz, F., **2002**. IS91 rolling-circle transposition. In: Craig, N., Craigie, R., Gellert, M., Lambowitz, A. (Eds.), Mobile DNA II. American Society for Microbiology Press, Washington, DC, pp. 891–904.

Garel, A., Nony, P., Prudhomme, J.C., **1994**. Structural features of *mag*, a *gypsy*-like retrotransposon of *Bombyx mori*, with unusual short terminal repeats. *Genetica 93*, 125–137.

Garner, K.J., Slavicek, J.M., **1999**. Identification of a non-LTR retrotransposon from the gypsy moth. *Insect Mol. Biol. 8*, 231–242.

Gellert, M., **2002**. V(D)J recombination. In: Craig, N., Craigie, R., Gellert, M., Lambowitz, A. (Eds.), Mobile DNA II. American Society for Microbiology Press, Washington, DC, pp. 705–729.

Gomulski, L.M., Torti, C., Bonizzoni, M., Moralli, D., Raimondi, E., *et al.*, **2001**. A new basal subfamily of *mariner* elements in *Ceratitis rosa* and other tephritid flies. *J. Mol. Evol. 53*, 597–606.

Gorbunova, V., Levy, A.A., **1997**. Circularized *Ac/Ds* transposons: formation, structure and fate. *Genetics 145*, 1161–1169.

Grandbastien, M.A., **1998**. Activation of plant retrotransposons under stress condiitons. *Trends Plant Sci. 3*, 181–187.

Gray, Y.H., **2000**. It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. *Trends Genet. 16*, 461–468.

Green, C.L., Frommer, M., **2001**. The genome of the Queensland fruit fly *Bactrocera tryoni* contains multiple representatives of the *mariner* family of transposable elements. *Insect Mol. Biol. 10*, 371–386.

Hammond, S.M., Bernstein, E., Beach, D., Hannon, G.J., **2000**. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature 404*, 293–296.

Handler, A.M., **2003**. Isolation and analysis of a new *hopper* hAT transposon from the dorsalis white eye strain. *Genetica 118*, 17–24.

Handler, A.M., McCombs, S.D., **2000**. The *piggyBac* transposon mediates germ-line transformation in the Oriental fruit fly and closely related elements exist in its genome. *Insect Mol. Biol. 9*, 605–612.

Hartl, D., **1989**. Transposable element *mariner* in *Drosophila* species. In: Berg, D.E., Howe, M.J. (Eds.), Mobile DNA. American Society for Microbiology Press, Washington, DC, pp. 531–536.

Hartl, D.L., Lozovskaya, E.R., Nurminsky, D.I., Lohe, A.R., **1997**. What restricts the activity of *mariner*-like transposable elements. *Trends Genet. 13*, 197–201.

Henikoff, S., **1992**. Detection of *Caenorhabditis* transposon homologs in diverse. *New Biol. 4*, 382–388.

Hill, S.R., Leung, S.S., Quercia, N.L., Vasiliauskas, D., Yu, J., *et al.*, **2001**. Ikirara insertions reveal five new *Anopheles gambiae* transposable elements in islands of repetitious sequence. *J. Mol. Evol. 52*, 215–231.

Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., *et al.*, **2002**. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science 298*, 129–149.

Holyoake, A.J., Kidwell, M.G., **2003**. *Vege* and *Mar*: two novel *hAT* MITE families from *Drosophila willistoni*. *Mol. Biol. Evol. 20*, 163–167.

Horn, C., Offen, N., Nystedt, S., Hacker, U., Wimmer, E.A., **2003**. *piggyBac*-based insertional mutagenesis and enhancer detection as a tool for functional insect genomics. *Genetics 163*, 647–661.

Imwong, M., Sharpe, R.G., Kittayapong, P., Baimai, V., **2000**. Distribution of the transposable element *mariner* in anopheline mosquitoes. *Heredity 85*, 271–276.

Ivics, Z., Hackett, P.B., Plasterk, R.H., Izsvak, Z., **1997**. Molecular reconstruction of *Sleeping Beauty*, a *Tc1*-like transposon from fish, and its transposition in human cells. *Cell 91*, 501–510.

Izsvak, Z., Ivics, Z., Shimoda, N., Mohn, D., Okamoto, H., *et al.*, **1999**. Short inverted-repeat transposable elements in teleost fish and implications for a mechanism of their amplification. *J. Mol. Evol. 48*, 13–21.

Jacobson, J.W., Medhora, M.M., Hartl, D.L., **1986**. Molecular structure of a somatically unstable transposable element in *Drosophila*. *Proc. Natl Acad. Sci. USA 83*, 8684–8688.

Jasinskiene, N., Coates, C.J., James, A.A., **2000**. Structure of *hermes* integrations in the germline of the yellow fever mosquito, *Aedes aegypti*. *Insect Mol. Biol. 9*, 11–18.

Jensen, S., Cavarec, L., Dhellin, O., Heidmann, T., **1994**. Retrotransposition of a marked *Drosophila* LINE-like I element in cells in culture. *Nucl. Acids Res. 22*, 1484–1488.

Jensen, S., Gassama, M.P., Heidmann, T., **1999**. Cosuppression of *I* transposon activity in *Drosophila* by *I*-containing sense and antisense transgenes. *Genetics 153*, 1767–1774.

Jiang, N., Bao, Z., Zhang, X., Hirochika, H., Eddy, S.R., *et al.*, **2003**. An active DNA transposon family in rice. *Nature 421*, 163–167.

Jordan, I.K., Matyunina, L.V., McDonald, J.F., **1999**. Evidence for the recent horizontal transfer of long terminal repeat retrotransposon. *Proc. Natl Acad. Sci. USA 96*, 12621–12625.

Jurka, J., **1995**. Origin and evolution of *Alu* repetitive elements. In: Maraia, R.J. (Ed.), The Impact of Short Interspersed Elements (SINEs) on the Host Genome. R. G. Landes, Austin, TX, pp. 25–41.

Jurka, J., **2000**. Repbase update: a database and an electronic journal of elements. *Trends Genet. 16*, 418–420.

Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J., Svirskas, R., *et al.*, **2002**. The transposable elements of the *Drosophila melanogaster* genomics perspective. *Genome Biol. 3*, RESEARCH0084–0084.

Kapitonov, V.V., Jurka, J., **1999**. Molecular paleontology of transposable elements from *Arabidopsis*. *Genetica 107*, 27–37.

Kapitonov, V.V., Jurka, J., **2001**. Rolling-circle transposons in eukaryotes. *Proc. Natl Acad. Sci. USA 98*, 8714–8719.

Kapitonov, V.V., Jurka, J., **2002**. *Looper1_DM*, a family of *Looper/piggyBac* DNA transposons in *D. melanogaster*. *Repbase Reps. 2*, 6.

Kapitonov, V.V., Jurka, J., **2003a**. A novel class of SINE elements derived from 5S rRNA. *Mol. Biol. Evol. 20*, 694–702.

Kapitonov, V.V., Jurka, J., **2003b**. Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. *Proc. Natl Acad. Sci. USA 100*, 6569–6574.

Kaufman, T.C., Severson, D.W., Robinson, G.E., **2002**. The *Anopheles* genome and comparative insect genomics. *Science 298*, 97–98.

Kazazian, H.H., Jr., Goodier, J.L., **2002**. LINE drive, retrotransposition and genome instability. *Cell 110*, 277–280.

Ke, Z., Grossman, G.L., Cornel, A.J., Collins, F.H., **1996**. *Quetzal*: a transposon of the *Tc1* family in the mosquito *Anopheles albimanus*. *Genetica 98*, 141–147.

Ketting, R.F., Haverkamp, T.H., van Luenen, H.G., Plasterk, R.H., **1999**. *Mut-7* of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell 99*, 133–141.

Kidwell, M.G., **1977**. Reciprocal differences in female recombination associated with dysgenesis in *Drosophila melanogaster*. *Genet. Res. 30*, 77–88.

Kidwell, M.G., **1992**. Horizontal transfer of *P* elements and other short inverted repeat transposons. *Genetica 86*, 275–286.

Kidwell, M.G., **2002**. Transposable elements and the evolution of genome size in eukaryotes. *Genetica 115*, 49–63.

Kidwell, M.G., Lisch, D.R., **2000**. Transposable elements and host genome evolution. *Trends Ecol. Evol. 15*, 95–99.

Kidwell, M.G., Lisch, D.R., **2001**. Perspective: transposable elements, parasitic DNA, and genome evolution. *Int. J. Org. Evol. 55*, 1–24.

Kidwell, M.G., Lisch, D.R., **2002**. Transposable elements as sources of genomic variation. In: Craig, N., Craigie, R., Gellert, M., Lambowitz, A. (Eds.), Mobile DNA II. American Society for Microbiology Press, Washington, DC, pp. 59–90.

Klinakis, A.G., Loukeris, T.G., Pavlopoulos, A., Savakis, C., **2000**. Mobility assays confirm the broad host-range activity of the transposable element and validate new transformation tools. *Insect Mol. Biol. 9*, 269–275.

Krebs, M.P., RajBhandary, U.L., Khorana, H.G., **1990**. Nucleotide sequence of ISH11, a new *Halobacterium halobium* element isolated from the plasmid pGRB1. *Nucl. Acids Res. 18*, 6699.

Krzywinski, J., Besansky, N.J., **2003**. Molecular systematics of *Anopheles*: from subgenera to subpopulations. *Annu. Rev. Entomol. 48*, 111–139.

Labrador, M., Corces, V., **2002**. Interactions between transposable elements and the host genome. In: Craig, N., Craigie, R., Gellert, M., Lambowitz, A. (Eds.), Mobile DNA II. American Society for Microbiology Press, Washington, DC, pp. 1008–1023.

Labrador, M., Seleme, M.C., Fontdevila, A., **1998**. The evolutionary history of *Drosophila buzzatii*. XXXIV. The distribution of the retrotransposon *Osvaldo* in original and colonizing populations. *Mol. Biol. Evol. 15*, 1532–1547.

Lampe, D.J., Akerley, B.J., Rubin, E.J., Mekalanos, J.J., Robertson, H.M., **1999**. Hyperactive transposase mutants of the Himar1 *mariner* transposon. *Proc. Natl Acad. Sci. USA 96*, 11428–11433.

Lampe, D.J., Walden, K.K., Robertson, H.M., **2001**. Loss of transposase-DNA interaction may underlie the divergence of mariner family transposable elements and the ability of more than one mariner to occupy the same genome. *Mol. Biol. Evol. 18*, 954–961.

Lampe, D.J., Willis, J.H., **1994**. Characterization of a cDNA and gene encoding a cuticular protein from rigid cuticles of the giant silkmoth, *Hyalophora cecropia*. *Insect Biochem. Mol. Biol. 24*, 419–435.

Lampe, D.J., Witherspoon, D.J., Soto-Adames, F.N., Robertson, H.M., **2003**. Recent horizontal transfer of mellifera subfamily mariner transposons into insect lineages representing four different orders shows that selection acts only during horizontal transfer. *Mol. Biol. Evol. 20*, 554–562.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., *et al.*, **2001**. Initial sequencing and analysis of the human genome. *Nature 409*, 860–921.

Le, Q.H., Turcotte, K., Bureau, T., **2001**. *Tc8*, a *Tourist*-like transposon in *Caenorhabditis elegans*. *Genetics 158*, 1081–1088.

Le, Q.H., Wright, S., Yu, Z., Bureau, T., **2000**. Transposon diversity in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA 97*, 7376–7381.

Lee, S.H., Clark, J.B., Kidwell, M.G., **1999**. A *P* element-homologous sequence in the house fly, *Musca domestica*. *Insect Mol. Biol. 8*, 491–500.

Levis, R.W., Ganesan, R., Houtchens, K., Tolar, L.A., Sheen, F.M., **1993**. Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell 75*, 1083–1093.

Liao, C., Rovira, C., He, H., Edstrom, J.E., **1998**. Site-specific insertion of a SINE-like element, *Cp1*, into centromeric tandem repeats from *Chironomus pallidivittatus*. *J. Mol. Biol. 280*, 811–821.

Lim, J.K., Simmons, M.J., **1994**. Gross chromosome rearrangements mediated by transposable elements in *Drosophila melanogaster*. *BioEssays 16*, 269–275.

Lin, R., Avery, L., **1999**. RNA interference: policing rogue genes. *Nature 402*, 128–129.

Liu, B., Wendel, J.F., **2000**. Retrotransposon activation followed by rapid repression in introgressed rice plants. *Genome 43*, 874–880.

Locke, J., Howard, L.T., Aippersbach, N., Podemski, L., Hodgetts, R.B., **1999**. The characterization of DINE-1, a short, interspersed repetitive element present on chromosome and in the centric heterochromatin of *Drosophila melanogaster*. *Chromosoma 108*, 356–366.

Luan, D.D., Korman, M.H., Jakubczak, J.L., Eickbush, T.H., **1993**. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell 72*, 595–605.

Luckhart, S., Rosenberg, R., **1999**. Gene structure and polymorphism of an invertebrate nitric oxide synthase gene. *Gene 232*, 25–34.

MacRae, A.F., Clegg, M.T., **1992**. Evolution of *Ac* and *Dsl* elements in select grasses (Poaceae). *Genetica 86*, 55–66.

Mahillon, J., Chandler, M., **1998**. Insertion sequences. *Microbiol. Mol. Biol. Rev. 62*, 725–774.

Malik, H.S., Burke, W.D., Eickbush, T.H., **1999**. The age and evolution of non-LTR retrotransposable elements. *Mol. Biol. Evol. 16*, 793–805.

Malik, H.S., Eickbush, T.H., **1998**. The RTE class of non-LTR retrotransposons is widely distributed in animals and is the origin of many SINEs. *Mol. Biol. Evol. 15*, 1123–1134.

Martin, S.L., Bushman, F.D., **2001**. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol. Cell Biol. 21*, 467–475.

Maside, X., Bartolome, C., Assimacopoulos, S., Charlesworth, B., **2001**. Rates of movement and distribution of transposable elements in *Drosophila melanogaster*: *in situ* hybridization vs. Southern blotting data. *Genet. Res. 78*, 121–136.

Mathiopoulos, K.D., della Torre, A., Predazzi, V., Petrarca, V., *et al.*, **1998**. Cloning of inversion breakpoints in the *Anopheles gambiae* complex traces a transposable element at the inversion junction. *Proc. Natl Acad. Sci. USA 95*, 12444–12449.

Mathiopoulos, K.D., della Torre, A., Santolamazza, F., Predazzi, V., Petrarca, V., *et al.*, **1999**. Are chromosomal inversions induced by transposable elements? A paradigm from the malaria mosquito *Anopheles gambiae*. *Parasitologia 41*, 119–123.

McCarthy, E.M., McDonald, J.F., **2003**. LTR_STRUC: a novel search and identification program for LTR retrotransposons. *Bioinformatics 19*, 362–367.

McClintock, B., **1948**. Mutable loci in maize. *Carnegie Inst. Wash. Yb 47*, 155–169.

McClintock, B., **1950**. The origin and behavior of mutable loci in maize. *Proc. Natl Acad. Sci. USA 36*, 344–355.

McClintock, B., **1956**. Controlling elements and the gene. *Cold Spring Harbor Symp. Quant. Biol. 8*, 58–74.

McClintock, B., **1984**. The significance of responses of the genome to challenge. *Science 226*, 792–801.

Medstrand, P., van de Lagemaat, L.N., Mager, D.L., **2002**. Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res. 12*, 1483–1495.

Mikitani, K., Sugasaki, T., Shimada, T., Kobayashi, M., Gustafsson, J.A., **2000**. The chitinase gene of the silkworm, *Bombyx mori*, contains a novel *Tc*-like transposable element. *J. Biol. Chem. 275*, 37725–37732.

Miller, W.J., Nagel, A., Bachmann, J., Bachmann, L., **2000**. Evolutionary dynamics of the SGM transposon family in the *Drosophila obscura* species group. *Mol. Biol. Evol. 17*, 1597–1609.

Miller, W.J., Paricio, N., Hagemann, S., Martinez-Sebastian, M.J., Pinsker, W., *et al.*, **1995**. Structure and expression of clustered *P* element homologues in *Drosophila subobscura* and *Drosophila guanche*. *Gene 156*, 167–174.

Misquitta, L., Paterson, B.M., **1999**. Targeted disruption of gene function in *Drosophila* by RNA interference (RNA-i): a role for nautilus in embryonic somatic muscle formation. *Proc. Natl Acad. Sci. USA 96*, 1451–1456.

Mizrokhi, L.J., Mazo, A.M., **1990**. Evidence for horizontal transmission of the mobile element *jockey* between distant *Drosophila* species. *Proc. Natl Acad. Sci. USA 87*, 9216–9220.

Mizrokhi, L.J., Mazo, A.M., **1991**. Cloning and analysis of the mobile element *gypsy* from *D. virilis*. *Nucl. Acids Res. 19*, 913–916.

Morgan, G.T., Middleton, K.M., **1990**. Short interspersed repeats from *Xenopus* that contain multiple motifs are related to known transposable elements. *Nucl. Acids Res. 18*, 5781–5786.

Mouches, C., Bensaadi, N., Salvado, J.C., **1992**. Characterization of a LINE retroposon dispersed in the genome of three non-sibling *Aedes* mosquito species. *Gene 120*, 183–190.

Munstermann, L.E., Conn, J.E., **1997**. Systematics of mosquito disease vectors (Diptera, Culicidae): impact of molecular biology and cladistic analysis. *Annu. Rev. Entomol. 42*, 351–369.

Murata, S., Takasaki, N., Saitoh, M., Okada, N., **1993**. Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. *Proc. Natl Acad. Sci. USA 90*, 6995–6999.

Nasidze, I., Risch, G.M., Robichaux, M., Sherry, S.T., Batzer, M.A., *et al.*, **2001**. *Alu* insertion polymorphisms and the genetic structure of human from the Caucasus. *Eur. J. Hum. Genet. 9*, 267–272.

Nikaido, M., Rooney, A.P., Okada, N., **1999**. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interpersed elements: hippopotamuses are the closest extant relatives of whales. *Proc. Natl Acad. Sci. USA 96*, 10261–10266.

Nuzhdin, S.V., **1999**. Sure facts, speculations, and open questions about the evolution of transposable element copy number. *Genetica 107*, 129–137.

Ohbayashi, F., Shimada, T., Sugasaki, T., Kawai, S., Mita, K., *et al.*, **1998**. Molecular structure of the *copia*-like retrotransposable element *Yokozuna* on the W chromosome of the silkworm, *Bombyx mori*. *Genes Genet. Syst. 73*, 345–352.

Ohshima, K., Hamada, M., Terai, Y., Okada, N., **1996**. The 3′ ends of tRNA-derived short interspersed repetitive elements are derived from the 3′ ends of long interspersed repetitive elements. *Mol. Cell Biol. 16*, 3756–3764.

Okada, N., **1991**. SINEs. *Curr. Opin. Genet. Devel. 1*, 498–504.

Okada, N., Hamada, M., **1997**. The 3′ ends of tRNA-derived SINEs originated from the 3′ ends of LINEs: a new example from the bovine genome. *J. Mol. Evol. 44* (Suppl. 1), S52–56.

Oosumi, T., Garlick, B., Belknap, W.R., **1995**. Identification and characterization of putative transposable DNA elements in solanaceous plants and *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA 92*, 8886–8890.

Orgel, L., Crick, F., **1980**. Selfish DNA: the ultimate parasite. *Nature 284*, 604–607.

Ostertag, E.M., Prak, E.T., DeBerardinis, R.J., Moran, J.V., Kazazian, H.H., Jr., **2000**. Determination of L1 retrotransposition kinetics in cultured cells. *Nucl. Acids Res. 28*, 1418–1423.

Perkins, H.D., Howells, A.J., **1992**. Genomic sequences with homology to the *P* element of *Drosophila melanogaster* occur in the blowfly *Lucilia cuprina*. *Proc. Natl Acad. Sci. USA 89*, 10753–10757.

Peterson, D.G., Schulze, S.R., Sciara, E.B., Lee, S.A., Bowers, J.E., *et al.*, **2002**. Integration of Cot analysis, DNA cloning, and high-throughput facilitates genome characterization and gene discovery. *Genome Res. 12*, 795–807.

Petrov, D.A., **2001**. Evolution of genome size: new approaches to an old problem. *Trends Genet. 17*, 23–28.

Petrov, D.A., Aminetzach, Y.T., Davis, J.C., Bensasson, D., Hirsh, A.E., **2003**. Size matters: non-LTR retrotransposable elements and ectopic recombination in *Drosophila*. *Mol. Biol. Evol. 20*, 880–892.

Pfeifer, T.A., Ring, M., Grigliatti, T.A., **2000**. Identification and analysis of *Lydia*, a LTR retrotransposon from *Lymantria dispar. Insect Mol. Biol. 9*, 349–356.

Pinkerton, A.C., Whyard, S., Mende, H.A., Coates, C.J., O'Brochta, D.A., *et al.*, **1999**. The Queensland fruit fly, *Bactrocera tryoni*, contains multiple members of the *hAT* family of transposable elements. *Insect Mol. Biol. 8*, 423–434.

Plasterk, R.H., Izsvak, Z., Ivics, Z., **1999**. Resident aliens: the *Tc1/mariner* superfamily of transposable elements. *Trends Genet. 15*, 326–332.

Rai, K.S., Black, W.C.T., **1999**. Mosquito genomes: structure, organization, and evolution. *Adv. Genet. 41*, 1–33.

Ribeiro, J.M., Kidwell, M.G., **1994**. Transposable elements as population drive mechanisms: specification of critical parameter values. *J. Med. Entomol. 31*, 10–16.

Rio, D., **2002**. *P* transposable elements in *Drosophila melanogaster.* In: Craig, N., Craigie, R., Gellert, M., Lambowitz, A. (Eds.), Mobile DNA II. American Society for Microbiology Press, Washington, DC, pp. 484–518.

Rizzon, C., Marais, G., Gouy, M., Biemont, C., **2002**. Recombination rate and the distribution of transposable elements *Drosophila melanogaster* genome. *Genome Res. 12*, 400–407.

Robertson, H.M., **1993**. The *mariner* transposable element is widespread in insects. *Nature 362*, 241–245.

Robertson, H.M., **2002**. Evolution of DNA transposons in eukaryotes. In: Craig, N., Craigie, R., Gellert, M., Lambowitz, A. (Eds.), Mobile DNA II. American Society for Microbiology Press, Washington, DC, pp. 1093–1110.

Robertson, H.M., Asplund, M.L., **1996**. Bmmar1: a basal lineage of the *mariner* family of transposable elements in the silkworm moth, *Bombyx mori*. *Insect Biochem. Mol. Biol. 26*, 945–954.

Robertson, H.M., Lampe, D.J., **1995**. Distribution of transposable elements in arthropods. *Annu. Rev. Entomol. 40*, 333–357.

Robertson, H.M., Walden, K.K., **2003**. Bmmar6, a second mori subfamily *mariner* transposon from *Bombyx mori*. *Insect Mol. Biol. 12*, 167–171.

Rohr, C.J., Ranson, H., Wang, X., Besansky, N.J., **2002**. Structure and evolution of *mtanga*, a retrotransposon actively expressed on the Y chromosome of the African malaria vector *Anopheles gambiae*. *Mol. Biol. Evol. 19*, 149–162.

Romans, P., Bhattacharyya, R.K., Colavita, A., **1998**. *Ikirara*, a novel transposon family from the malaria vector mosquito, *Anopheles gambiae*. *Insect Mol. Biol. 7*, 1–10.

Rongnoparut, P., Sirichotpakorn, N., Rattanarithikul, R., Linthicum, K.J., **1998**. Sequence heterogeneity in *copia*-like retrotransposons in *Anopheles* (Diptera: Culicidae) in Thailand. *J. Med. Entomol. 35*, 771–777.

Roy, A.M., Carroll, M.L., Kass, D.H., Nguyen, S.V., Salem, A.H., *et al.*, **1999**. Recently integrated human *Alu* repeats: finding needles in the haystack. *Genetica 107*, 149–161.

Roy-Engel, A.M., Carroll, M.L., Vogel, E., Garber, R.K., Nguyen, S.V., *et al.*, **2001**. *Alu* insertion polymorphisms for the study of human genomic diversity. *Genetics 159*, 279–290.

Rubin, E., Levy, A.A., **1997**. Abortive gap repair: underlying mechanism for *Ds* element formation. *Mol. Cell Biol. 17*, 6294–6302.

Salem, A.H., Kilroy, G.E., Watkins, W.S., Jorde, L.B., Batzer, M.A., **2003**. Recently integrated *Alu* elements and human genomic diversity. *Mol. Biol. Evol. 20*, 1349–1361.

Sarkar, A., Sengupta, R., Krzywinski, J., Wang, X., Roth, C., *et al.*, **2003**. *P* elements are found in the genomes of nematoceran insects of the genus *Anopheles*. *Insect Biochem. Mol. Biol. 33*, 381–387.

Sarkar, A., Sin, C., Hong, Y.S., Hogan, J.R., Fraser, M.J., *et al.*, **2003**. Molecular evolutionary analysis of the widespread *piggyBac* transposon family and related "domesticated" sequences. *Mol. Genet. Genomics 270*, 173–180.

Severson, D.W., Mori, A., Zhang, Y., Christensen, B.M., **1993**. Linkage map for *Aedes aegypti* using restriction fragment length polymorphisms. *J. Hered. 84*, 241–247.

Shao, H., Qi, Y., Tu, Z., **2001**. *MsqTc3*, a *Tc3*-like transposon in the yellow fever mosquito *Aedes*. *Insect Mol. Biol. 10*, 421–425.

Shao, H., Tu, Z., **2001**. Expanding the diversity of the *IS630-Tc1-mariner* superfamily: discovery of a unique DD37E transposon and reclassification of the DD37D and DD39D transposons. *Genetics 159*, 1103–1115.

Shedlock, A.M., Okada, N., **2000**. SINE insertions: powerful tools for molecular systematics. *BioEssays 22*, 148–160.

Sherratt, D.J., **1995**. Mobile Genetic Elements. Oxford University Press, Oxford.

Shimamura, M., Yasue, H., Ohshima, K., Abe, H., Kato, H., *et al.*, **1997**. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. *Nature 388*, 666–670.

Shiroishi, T., Sagai, T., Moriwaki, K., **1993**. Hotspots of meiotic recombination in the mouse major histocompatibility complex. *Genetica 88*, 187–196.

Siebel, C.W., Admon, A., Rio, D.C., **1995**. Soma-specific expression and cloning of PSI, a negative regulator of *P* element pre-mRNA splicing. *Genes Devel. 9*, 269–283.

Siebel, C.W., Fresco, L.D., Rio, D.C., **1992**. The mechanism of somatic inhibition of *Drosophila P*-element pre-mRNA splicing: multiprotein complexes at an exon pseudo-5′ splice site control U1 snRNP binding. *Genes Devel. 6*, 1386–1401.

Silva, J.C., Kidwell, M.G., **2000**. Horizontal transfer and selection in the evolution of *P* elements. *Mol. Biol. Evol. 17*, 1542–1557.

Silva, J.C., Loreto, E.L., Clark, J.B., **2004**. Factors that affect the horizontal transfer of transposable elements. *Curr. Issues Mol. Biol. 6*, 57–71.

Smit, A.F., Riggs, A.D., **1996**. Tiggers and DNA transposon fossils in the human genome. *Proc. Natl Acad. Sci. USA 93*, 1443–1448.

Song, W.Y., Pi, L.Y., Bureau, T.E., Ronald, P.C., **1998**. Identification and characterization of 14 transposon-like noncoding regions of members of the Xa21 family of disease genes in rice. *Mol. Gen. Genet. 258*, 449–456.

Spradling, A.C., Stern, D., Beaton, A., Rhem, E.J., Laverty, T., *et al.*, **1999**. The Berkeley Drosophila Genome Project gene disruption project: single *P*-element insertions mutating 25% of vital *Drosophila* genes. *Genetics 153*, 135–177.

Spradling, A.C., Stern, D.M., Kiss, I., Roote, J., Laverty, T., *et al.*, **1995**. Gene disruptions using *P* transposable elements: an integral component of the *Drosophila* genome project. *Proc. Natl Acad. Sci. USA 92*, 10824–10830.

Stoneking, M., Fontius, J.J., Clifford, S.L., Soodyall, H., Arcot, S.S., *et al.*, **1997**. Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res. 7*, 1061–1071.

Streck, R.D., MacGaffey, J.E., Beckendorf, S.K., **1986**. The structure of *hobo* transposable elements and their insertion. *EMBO J. 5*, 3615–3623.

Sun, S.C., Asling, B., Faye, I., **1991**. Organization and expression of the immunoresponsive lysozyme gene in the giant silk moth, *Hyalophora cecropia*. *J. Biol. Chem. 266*, 6644–6649.

Sundararajan, P., Atkinson, P.W., O'Brochta, D.A., **1999**. Transposable element interactions in insects: cross-mobilization of *hobo* and *Hermes*. *Insect Mol. Biol. 8*, 359–368.

Swofford, D.L., **2001**. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, MA.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., *et al.*, **1999**. The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell 99*, 123–132.

Takahashi, K., Terai, Y., Nishida, M., Okada, N., **1998**. A novel family of short interspersed repetitive elements (SINEs) from cichlids: the patterns of insertion of SINEs at orthologous loci support the proposed monophyly of four major groups of cichlid fishes in Lake Tanganyika. *Mol. Biol. Evol. 15*, 391–407.

Terai, Y., Takahashi, K., Nishida, M., Sato, T., Okada, N., **2003**. Using SINEs to probe ancient explosive speciation: "hidden" radiation of African cichlids? *Mol. Biol. Evol. 20*, 924–930.

Terzian, C., Ferraz, C., Demaille, J., Bucheton, A., **2000**. Evolution of the *Gypsy* endogenous retrovirus in the *Drosophila melanogaster* subgroup. *Mol. Biol. Evol. 17*, 908–914.

Thompson, J.D., Higgins, D.G., Gibson, T.J., **1994**. ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res. 22*, 4673–4680.

Tikhonov, A.P., Bennetzen, J.L., Avramova, Z.V., **2000**. Structural domains and matrix attachment regions along colinear chromosomal segments of maize and sorghum. *Plant Cell 12*, 249–264.

Transgenesis and Genomics of Invertebrate Organisms, **2003**. Reports of the 4th International Workshop, Pacific Grove, CA, 11–15 May 2003. http://www.nises.affrc.go.jp/sgp/SGP.htm; www.genoptera.com/press-release/PR20030403.html.

Truett, M.A., Jones, R.S., Potter, S.S., **1981**. Unusual structure of the FB family of transposable elements in *Drosophila*. *Cell 24*, 753–763.

Tu, Z., **1997**. Three novel families of miniature inverted-repeat transposable elements are associated with genes of the yellow fever mosquito, *Aedes aegypti*. *Proc. Natl Acad. Sci. USA 94*, 7475–7480.

Tu, Z., **1999**. Genomic and evolutionary analysis of *Feilai*, a diverse family of highly reiterated SINEs in the yellow fever mosquito, *Aedes aegypti*. *Mol. Biol. Evol. 16*, 760–772.

Tu, Z., **2000**. Molecular and evolutionary analysis of two divergent subfamilies of a novel miniature inverted repeat transposable element in the yellow fever mosquito, *Aedes aegypti*. *Mol. Biol. Evol. 17*, 1313–1325.

Tu, Z., **2001a**. Eight novel families of miniature inverted repeat transposable elements in the African malaria mosquito, *Anopheles gambiae*. *Proc. Natl Acad. Sci. USA 98*, 1699–1704.

Tu, Z., **2001b**. *Maque*, a family of extremely short interspersed repetitive elements: characterization, possible mechanism of transposition, and evolutionary implications. *Gene 263*, 247–253.

Tu, Z., Hill, J.J., **1999**. *MosquI*, a novel family of mosquito retrotransposons distantly related to the *Drosophila I* factors, may consist of elements of more than one origin. *Mol. Biol. Evol. 16*, 1675–1686.

Tu, Z., Isoe, J., Guzova, J.A., **1998**. Structural, genomic, and phylogenetic analysis of *Lian*, a novel family of non-LTR retrotransposons in the yellow fever mosquito, *Aedes aegypti*. *Mol. Biol. Evol. 15*, 837–853.

Tu, Z., Orphanidis, S.P., **2001**. *Microuli*, a family of miniature subterminal inverted-repeat transposable elements (MSITEs): transposition without terminal inverted repeats. *Mol. Biol. Evol. 18*, 893–895.

Tudor, M., Lobocka, M., Goodell, M., Pettitt, J., O'Hare, K., **1992**. The *pogo* transposable element family of *Drosophila melanogaster*. *Mol. Gen. Genet. 232*, 126–134.

Van den Broeck, D., Maes, T., Sauer, M., Zethof, J., De Keukeleire, P., *et al.*, **1998**. Transposon display identifies individual transposable elements in high copy number lines. *Plant J. 13*, 121–129.

Vazquez-Manrique, R.P., Hernandez, M., Martinez-Sebastian, M.J., de Frutos, R., **2000**. Evolution of *gypsy* endogenous retrovirus in the *Drosophila obscura* species group. *Mol. Biol. Evol. 17*, 1185–1193.

Vieira, C., Lepetit, D., Dumont, S., Biemont, C., **1999**. Wake up of transposable elements following *Drosophila simulans* worldwide colonization. *Mol. Biol. Evol. 16*, 1251–1255.

Voytas, D., Boeke, J., **2002**. Ty1 and Ty5 of *Saccharomyces cerevisiae*. In: Craig, N., Craigie, R., Gellert, M., Lambowitz, A. (Eds.), Mobile DNA II. American Society for Microbiology Press, Washington, DC, pp. 631–662.

Walbot, V., Rudenko, G., **2002**. *MuDR/Mu* transposable elements of maize. In: Craig, N., Craigie, R., Gellert, M., Lambowitz, A. (Eds.), Mobile DNA II. American Society for Microbiology Press, Washington, DC, pp. 533–564.

Warren, A.M., Crampton, J.M., **1991**. The *Aedes aegypti* genome: complexity and organization. *Genet. Res. 58*, 225–232.

Warren, A.M., Hughes, M.A., Crampton, J.M., **1997**. Zebedee: a novel *copia*-Ty1 family of transposable elements in the genome of the medically important mosquito *Aedes aegypti*. *Mol. Gen. Genet. 254*, 505–513.

Warren, W.D., Atkinson, P.W., O'Brochta, D.A., **1994**. The *Hermes* transposable element from the house fly, *Musca domestica*, is a short inverted repeat-type element of the *hobo, Ac*, and *Tam3* (*hAT*) element family. *Genet. Res. 64*, 87–97.

Watkins, W.S., Ricker, C.E., Bamshad, M.J., Carroll, M.L., Nguyen, S.V., *et al.*, **2001**. Patterns of ancestral human diversity: an analysis of *Alu*-insertion and restriction-site polymorphisms. *Am. J. Hum. Genet. 68*, 738–752.

Wei, W., Gilbert, N., Ooi, S.L., Lawler, J.F., Ostertag, E.M., *et al.*, **2001**. Human L1 retrotransposition: *cis* preference versus *trans* complementation. *Mol. Cell Biol. 21*, 1429–1439.

Wessler, S.R., **1996**. Turned on by stress: plant retrotransposons. *Curr. Biol. 6*, 959–961.

Wessler, S.R., Bureau, T.E., White, S.E., **1995**. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Curr. Opin. Genet. Devel. 5*, 814–821.

Wobus, U., Baumlein, H., Bogachev, S.S., Borisevich, I.V., Panitz, R., *et al.*, **1990**. A new transposable element in *Chironomus thummi*. *Mol. Gen. Genet. 222*, 311–316.

Xiong, Y., Burke, W.D., Eickbush, T.H., **1993**. Pao, a highly divergent retrotransposable element from *Bombyx mori* containing long terminal repeats with tandem copies of the putative R regions. *Nucleic Acids Res. 21*, 2117–2123.

Xiong, Y., Eickbush, T.H., **1988**. The site-specific ribosomal DNA insertion element *R1Bm* belongs to a class of non-long-terminal-repeat retrotransposons. *Mol. Cell Biol. 8*, 114–123.

Xiong, Y., Eickbush, T.H., **1993**. *Dong*, a non-long terminal repeat (non-LTR) retrotransposable element from *Bombyx mori*. *Nucl. Acids Res. 21*, 1318.

Yan, G., Romero-Severson, J., Walton, M., Chadee, D.D., Severson, D.W., **1999**. Population genetics of the yellow fever mosquito in Trinidad: comparisons of amplified fragment length polymorphism (AFLP) and restriction fragment length polymorphism (RFLP) markers. *Mol. Ecol. 8*, 951–963.

Yang, H.P., Nuzhdin, S.V., **2003**. Fitness costs of Doc expression are insufficient to stabilize its copy number in *Drosophila melanogaster*. *Mol. Biol. Evol. 20*, 800–804.

York, D.S., Blum, V.M., Low, J.A., Rowold, D.J., Puzyrev, V., *et al.*, **1999**. Phylogenetic signals from point mutations and polymorphic *Alu* insertions. *Genetica 107*, 163–170.

Yoshiyama, M., Tu, Z., Kainoh, Y., Honda, H., Shono, T., *et al.*, **2001**. Possible horizontal transfer of a transposable element from host to parasitoid. *Mol. Biol. Evol. 18*, 1952–1958.

Zhang, Q., Arbuckle, J., Wessler, S.R., **2000**. Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family *Heartbreaker* into genic regions of maize. *Proc. Natl Acad. Sci. USA 97*, 1160–1165.

Zhang, X., Feschotte, C., Zhang, Q., Jiang, N., Eggleston, W.B., *et al.*, **2001**. *P* instability factor: an active maize transposon system associated with the amplification of *Tourist*-like MITEs and a new superfamily of transposases. *Proc. Natl Acad. Sci. USA 98*, 12572–12577.

Zhou, Q., Haymer, D.S., **1997**. Molecular structure of *yoyo*, a *gypsy*-like retrotransposon from the mediterranean fruit fly, *Ceratitis capitata*. *Genetica 101*, 167–178.

Relevant Websites

http://jaketu.biochem.vt.edu – programs for TE searches from Virginia Tech.

http://www.genetics.wustl.edu – includes program RECON for identifying TE sequences.

4.13 Transposable Elements for Insect Transformation

A M Handler, US Department of Agriculture, Agricultural Research Service, Gainesville, FL, USA
D A O'Brochta, University of Maryland Biotechnology Institute, College Park, MD, USA

Published by Elsevier BV.

4.13.1. Introduction

The ability to create genetically transformed organisms has played a central role in the history of modern genetics, in particular, to our understanding of gene expression and development. Indeed, the pioneering transformation experiments of *Pneumococcus* by Griffith (1928) and subsequent systematic analyses by Avery *et al.* (1944), that showed transformation from a "rough" to a "smooth" bacterial cell wall phenotype, were instrumental in defining DNA as the inherited genetic material. The importance of these initial transformation experiments to prokaryotic genetic analysis was widely appreciated, and continued studies by many other laboratories laid the foundation for modern molecular biology.

The importance of "transformation" technology to eukaryotic genetic studies was apparent, and several of the initial attempts to create transgenic animals were performed in insects, though the means of achieving and assessing insect transformation were not straightforward. The primary reasons why these early attempts to create transgenic insects were largely unsuccessful were the inability to isolate and reproduce individual genetic elements that could be used as transformation vectors and markers, and the lack of efficient means of introducing DNA into germ cells. Most of the initial studies of insect transformation relied on soaking embryos or larvae, with visible mutant phenotypes, in solutions of total wild-type genomic DNA in hopes of reverting the mutant phenotype. The first experiments performed on *Bombyx* and *Ephestia* met with some success where mutant wing color pattern phenotypes were reverted in some organisms, though inheritance was inconsistent and transformation events could not be confirmed unequivocally

(Caspari and Nawa, 1965; Nawa and Yamada, 1968; Nawa *et al*., 1971). Similar results were obtained in *Drosophila* studies (Fox and Yoon, 1966, 1970; Fox *et al*., 1970), and for all of these initial experiments, it is most likely that the observed phenotypic changes resulted from extrachromosomal maintainance of introduced DNA in the somatic tissue by an unknown mechanism. A different approach involving the microinjection of wild-type genomic DNA into embryos homozygous for a recessive eye color mutation (*vermilion*), resulted in transformants with a reversion to the normal red eye color phenotype. While the reversion event was genetically mapped away from the mutant locus, a thorough molecular analysis to verify a transformation event, before the lines were lost, was not achieved and so the nature of the phenomenon observed in this experiment remains unexplained (Germeraad, 1976).

In the mid-1970s a turning point in insect science occurred with the extension of molecular genetic analysis to *Drosophila melanogaster*. These early studies and subsequent studies not only provided many of the tools and reagents necessary for developing and critically assessing genetic transformation in insects, but they also emphasized the need for a technology that would facilitate a more complete understanding of the genes being isolated using recombinant DNA methods. One technology that was clearly needed was a means to integrate DNA molecules into the chromosomes of germ cells where it remained stable, resulting in heritable germline transformation. The simple introduction of raw linearized DNA into preblastoderm embryos in the hope of fortuitous recombination into host chromosomes was clearly not reliable. Interest was growing, however, in the use of mobile genetic elements as vectors for DNA integration, including retrotransposons and transposons that were being isolated in *Drosophila* for the first time. Foremost among these was the *P* transposable element, isolated from certain mutant alleles of the *white* gene. The subsequent testing and success of transformation mediated by the *P* element in the *Drosophila* germline proved to be a dramatic turning point in the genetic analysis of an insect species. The eventual impact of this technology on understanding genetic mechanisms in all eukaryotic systems cannot be understated. The success with *P* in *Drosophila* gave hope that this system could be straightforwardly extended to genetic manipulation of other insect species, and especially those highly important to agriculture and human health. While there was reason for optimism, we now realize that this was a naive expectation given what we now understand about the natural history of *P* elements relative to other Class II transposable elements, in particular its extremely limited distribution and its dependence on species-specific host factors. The inability of *P* to function in non-drosophilids, however, was a motivating force to more completely understand transposon regulation and the identification and testing of new vector systems. These included other transposable elements, as well as viral and bacterial vectors.

The development of routine methods for insect gene-transfer was probably delayed by a decade due to attention being focused exclusively on the *P* element. Yet, this delay has resulted in a more varied toolbox of vectors and markers that now allow nearly routine transformation for many important species, and the potential for transformation of most insects (see Handler, 2001). Indeed, some of the tools developed for testing the *P* element, in particular embryonic mobility assays, are now routinely used for initial tests for function of other vectors in an insect species before more laborious and time-consuming transformation experiments are attempted.

The creation of this varied toolbox was first related to the potential need for different vector and marker systems for different insect species. We now realize that the future of genetic analysis will depend on multiple vector and marker systems for each of these species, since genomics and functional genomics studies will require multiple systems for DNA integration and reporters for gene expression. Indeed, germline transformation is essential for the insertional mutagenesis and functional genomics studies that are critical underpinnings for both assessing genomic architecture and relating sequences to gene expression. Notably, the continuing functional analysis of the *Drosophila* genome now relies on the vectors and markers, described in this chapter, that were first developed for nondrosophilid insect species.

4.13.2. *P* Element Transformation

4.13.2.1. *P* Element

The use of transposable element-based vectors for *Drosophila* transformation followed the discovery of short inverted terminal repeat-type elements similar to the *Activator* (*Ac*) element discovered in maize by McClinotck (see Federoff, 1989). The first such element to be discovered in insects was the *P* element, the factor responsible for hybrid dysgenesis that occurred in crosses of males from a *P* strain (containing *P* factor) with females from an

M strain (devoid of *P* factor) females (Kidwell *et al.*, 1977). The identification of *P* sequences resulted from the molecular analysis of *P*-induced *white* mutations that occurred in dysgenic hybrids (Rubin *et al.*, 1982). While the initial *P* elements isolated as insertion sequences were incomplete, nonautonomous elements, complete functional elements were later isolated and characterized by O'Hare and Rubin (1983).

P is 2907 bp in length with 31 bp inverted terminal repeats (ITRs) and 11 bp subterminal inverted repeats that occur approximately 125 bp from each terminus (**Figure 1**). Other repeat sequences exist within *P*, but their functional significance, if any, remains unknown. A defining signature for *P*, as with other transposable elements, is the nature of its insertion site which consists of an 8 bp direct repeat duplication. The extensive use of *P* for transformation and transposon mutagenesis has shown the element to have a distinctly nonrandom pattern of integration. It is now clear that *P* elements are blind to a significant fraction of the genome and new gene vectors are being employed in *Drosophila* to complement these limitations. *P* elements and all transposable elements currently used as insect gene vectors belong to a general group of transposable elements known as Class II short inverted terminal repeat transposons (see Finnegan, 1989). These elements transpose via a DNA intermediate and generally utilize a cut-and-paste mechanism that creates a duplication of the insertion site. These are distinguished from Class I elements, or retrotransposons, that have long direct terminal repeats (LTRs) and transpose via reverse transcription an RNA intermediate.

The original use of *P* for germline transformation was accomplished by inserting a marker gene within the element so that it did not disrupt activity of the terminal sequences or the transposase gene. The $rosy^{+}$ gene was inserted at the 3′ end of the transposase-coding region, but upstream of the 3′ subterminal inverted repeat sequence. Plasmids containing this vector were injected into preblastoderm (syncytial) embryos homozygous for ry^{-} so the *P* vector could transpose into germ cell nuclei. Germline transformation events were identified in the following generation (G_1) by virtue of reversion of the mutant ry^{-} eye color phenotype to wild-type. These experiments not only proved the feasibility of transposon-mediated transformation, but also permitted structure–function relationships within the *P* element to be defined (Karess and Rubin, 1984). The *P* transcriptional unit was found to be composed of four exons separated by three introns. Further analysis determined that the *P* transposase function is cell-type specific owing to differential transcript splicing of the third intron that is limited to the germline. The lack of splicing in the soma results in production of nonfunctional truncated polypeptides in this tissue (Rio *et al.*, 1986).

While the original *P* vector allowed efficient transformation, the presence of a functional transposase gene within the vector made the system self-mobilizable (autonomous) and inherently unstable, allowing potential excision or transposition of the original insertion event. Subsequent vector development resulted in a binary system in which the vector transposase was deleted or made defective by insertion of a marker gene. The ability of the transposase

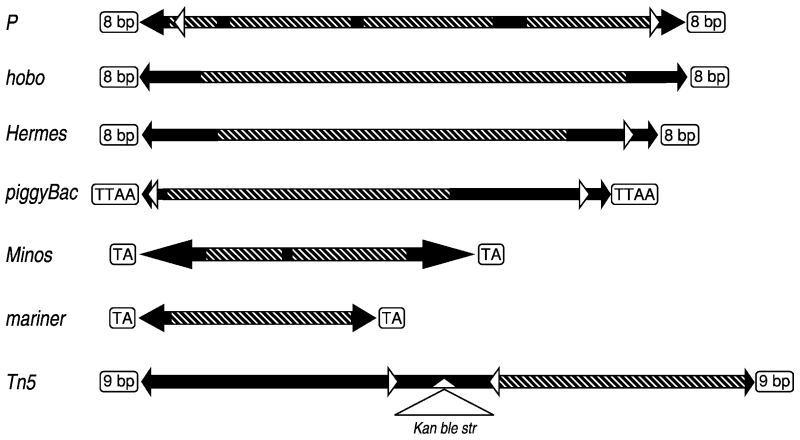


Figure 1 Diagram of transposable elements currently in use for the germline transformation of insect species. The left arms represent the 5′ termini and right arms represent the 3′ termini. Transposon sizes and specific internal elements are shown in relative positions but are not at precise scale. Major structural elements include duplicated insertion sites (open boxes); inverted terminal repeat sequences (black arrowheads); internal subterminal repeat sequences (white arrowheads); transposase coding region (boxed diagonals); and intron sequences (black boxes). The *Tn5* element is a composite transposable element consisting of two functional elements flanking three antibiotic resistance genes. Refer to text for specific details on nucleotide lengths and relative positions.

to act *in trans* allowed transposase to be provided by a separate plasmid (helper), that could facilitate vector integrations when cointroduced with the vector-containing plasmid into the same nucleus (Rubin and Spradling, 1982). Integrations would remain stable if the helper did not integrate, but the original helpers, such as pπ25.1, were autonomous *P* elements themselves that could integrate along with the vector. While helper integration was diminished by injecting much higher concentrations of vector plasmid, this possibility was only eliminated with the creation of defective helpers having one or both of their terminal sequences deleted (known as "wings-clipped" helpers). The first of these was pπ25.7wc, which was immobilized by deletion of 3′ terminal sequences (Karess and Rubin, 1984). This prototype vector system served as a model for binary systems of nonautonomous vector: helper elements used for all the transposon-based transformation systems currently in use (**Table 1**).

A notable characteristic of *P* elements was not only their discontinuous distribution within the species (*P* and *M* strains), but their discontinuous interspecific distribution. Based on its distribution patterns it has become apparent that *P* was recently introduced into *D. melanogaster* from *D. willistoni* by an unknown mechanism (Daniels and Strausbaugh, 1986). Regardless of the mechanism since the 1950s, *P* elements have thoroughly invaded wild populations of *D. melanogaster* (Anxolabéhère *et al.*, 1988), and without the existence of *M* strain laboratory stocks that were removed from nature before this time, the development of *P* vectors might never have been realized. This is due to the repression of *P* mobility in *P*-containing strains that was first observed in hybrid dysgenesis studies, which also showed that movement was not repressed in *M* strains devoid of *P*. The basis for *P* strain repression appears to be due to a number of factors including repressor protein synthesis, transposase titration by resident defective elements, and regulation of transposase gene transcription (Handler *et al.*, 1993b; Simmons *et al.*, 2002). As will be discussed further on, other vector systems in use have thus far been shown to be widely functional in several orders of insects, and the presence of the same or related transposon in a host insect does not necessarily repress vector transposition. In this and several other aspects, the *P* vector system appears to be the exception rather than the rule for transposon-mediated gene transfer in insects.

4.13.2.2. *P* Vectors and Markers

Regardless of regulatory differences between *P* and other transposon vector systems currently in use, methods developed for *P* transformation of *Drosophila* serve as a paradigm for all other insect vector systems. Those familiar with *Drosophila* transformation will be in the best position to attempt these methods in other insects. Current techniques developed for other insect species are variations on a theme, although as we describe, considerable modifications have been made. Several comprehensive reviews are available for more specific details on the structure, function, and use of *P* for transformation in *Drosophila*, which are highly relevant to the understanding and use of other vector systems (see Karess, 1985; Spradling, 1986; Engels, 1989; Handler and O'Brochta, 1991). Particularly useful are the books and method manuals by Ashburner (1989a, 1989b) that review the various vectors, markers, and methodologies used for *Drosophila* transformation, as well as early techniques used to manipulate *Drosophila* embryos. This information is especially applicable to other insect systems.

The first consideration for transformation is the design of vector and helper plasmids, and the marker system used for transformant selection. The first *P* vectors and helpers were actually autonomous vectors, which was probably a useful starting point since the actual sequence requirements for vector mobility and transposase function were unknown. As noted, the first nonautonomous helper had a 3′ terminal deletion that prevented its transposition, providing greater control over vector stability. However, this source of transposase was inefficient, until it was placed under *hsp70* regulation which allowed transposase induction by heat shock (Steller and Pirrotta, 1986). All other vector system helper constructs have similarly taken advantage of heat shock promoters, mostly from the *D. melanogaster hsp70* gene, but other *hsp* promoters have been tested including those from the host species being transformed. Other constitutive promoters such as those from the genes for *actin* and *α1-tubulin* have proven successful for helper transposase regulation, and will be discussed further on.

While sufficient transposase production is critical for transposition, the structure of the vector is equally important, and for some, very subtle changes from the autonomous vector can dramatically decrease or eliminate mobility. These variations include critical sequences (typically in the termini and subtermini), and placement and amount of exogenous DNA inserted within the termini. For some vectors the amount of plasmid DNA external to the vector can affect transposition rates. Subsequent to the initial test of several *P* vectors, the terminal sequence requirements for *P* mobility were determined to

Table 1 Transposon-mediated germline transformation

Transposon	*Host species*	*Marker*	*Reference*
Hermes	*Aedes aegypti*	*Dm-cinnabar*$^+$	Jasinskiene *et al.* (1998)
		actin5C-EGFP	Pinkerton *et al.* (2000)
	Culex quinquefasciatus	actin5C-EGFP	Allen *et al.* (2001)
	Ceratitis capitata	*Cc-white*$^+$	Michel *et al.* (2001)
	Drosophila melanogaster	*Dm-white*$^+$	O'Brochta *et al.* (1996)
		actin5C-EGFP	Pinkerton *et al.* (2000)
		3xP3-EGFP	Horn *et al.* (2000)
	Stomoxys calcitrans	actin5C-EGFP	O'Brochta *et al.* (2000)
	Tribolium castaneum	3xP3-EGFP	Berghammer *et al.* (1999)
hobo	*Drosophila melanogaster*	*Dm*-mini-*white*$^+$	Blackman *et al.* (1989)
	Drosophila virilis	*Dm*-mini-*white*$^+$	Lozovskaya *et al.* (1996); Gomez and Handler (1997)
mariner (*Mos1*)	*Aedes aegypti*	*Dm-cinnabar*$^+$	Coates *et al.* (1998)
	Drosophila melanogaster	*Dm-white*$^+$	Garza *et al.* (1991); Lidholm *et al.* (1993)
		3xP3-EGFP	Horn *et al.* (2000)
	Drosophila virilis	*Dm-white*$^+$	Lohe and Hartl (1996a)
Minos	*Anopheles stephensi*	actin5C-EGFP	Catteruccia *et al.* (2000b)
	Ceratitis capitata	*Cc-white*$^+$	Loukeris *et al.* (1995b)
	Drosophila melanogaster	*Dm-white*$^+$	Loukeris *et al.* (1995a)
P	*Drosophila melanogaster*	*Dm-rosy*$^+$	Rubin and Spradling (1982)
		Dm-white$^+$	Hazelrigg *et al.* (1984); Pirrotta *et al.* (1985)
		Dm-hsp70-mini-*white*$^+$	Klemenz *et al.* (1987)
		pUChsneo	Steller and Pirrotta (1985)
	Drosophila simulans	*Dm-rosy*$^+$	Scavarda and Hartl (1984)
piggyBac	*Aedes aegypti*	*Dm-cinnabar*$^+$	Lobo *et al.* (2002)
		3xP3-EGFP	Kokoza *et al.* (2001)
	Anastrepha suspensa	PUb-nls-EGFP	Handler and Harrell (2000)
	Anopheles albimanus	PUb-nls-EGFP	Perera *et al.* (2002)
	Anopheles gambiae	hr5-ie1:EGFP	Grossman *et al.* (2001)
	Anopheles stephensi	actin5C-DsRed	Nolan *et al.* (2002)
	Athalia rosae	*Bm*A3-EGFP, *hsp70*-GFP	Sumitani *et al.* (2003)
	Bactrocera dorsalis	*Cc-white*$^+$	Handler and McCombs (2000)
		PUb-nls-EGFP	Handler and McCombs (unpublished data)
	Bombyx mori	*Bm*A3-EGFP	Tamura *et al.* (2000)
		3xP3-EGFP	Thomas *et al.* (2002); Uhlirova *et al.* (2002)
	Ceratitis capitata	*Cc-white*$^+$	Handler *et al.* (1998)
		PUb-nls-EGFP	Handler and Krasteva (unpublished data)
		PUb-DsRed1	Handler and Krasteva (unpublished data)
	Cochliomyia hominivorax	PUb-nls-EGFP	Allen *et al.* (2004)
	Drosophila melanogaster	*Dm-white*$^+$, PUb-nls-EGFP	Handler and Harrell (1999)
		PUb-DsRed1	Handler and Harrell (2001)
		3xP3-EGFP	Horn *et al.* (2000)
		3xP3-EYFP	Horn and Wimmer (2000)
		3xP3-ECFP	Horn and Wimmer (2000)
		3xP3-DsRed	Horn *et al.* (2002)
	Lucilia cuprina	PUb-nls-EGFP	Heinrich *et al.* (2002)
	Musca domestica	3xP3-EGFP	Hediger *et al.* (2000)
	Pectinophora gossypiella	*Bm*A3-EGFP	Peloquin *et al.* (2000)
	Tribolium castaneum	3xP3-EGFP	Berghammer *et al.* (1999); Lorenzen *et al.* (2003)
Tn5	*Aedes aegypti*	3xP3-DsRed	O'Brochta (unpublished data)

EGFP, enhanced green fluorescent protein; GFP, green fluorescent protein; EYFP, enhanced yellow fluorescent protein; ECFP, enhanced cyan fluorescent protein.

include 138 bp of the 5′ end and 216 bp of the 3′ end. While the inverted repeat sequences within these terminal regions are identical, the adjacent sequences were found not to be interchangeable in terms of vector mobility (Mullins *et al*., 1989). Of interest was the discovery that the strongest binding affinity for the *P* transposase was at sequences approximately 50 bp internal to the terminal repeats (Rio and Rubin, 1988; Kaufman *et al*., 1989). While the minimal sequences required for mobility may be used in vectors, typically the rate of mobility decreases with the decreased length of terminal sequence. Specific sequences may be required for binding of transposase or other nuclear factors, and conformational changes needed for recombination may be dependent upon sequence length and position.

P vector mobility was also found to be influenced by the amount of exogenous DNA inserted between the termini, with transformation frequency diminishing with increasing size. Initial tests with 8 kb vectors marked with *rosy* yielded transformation frequencies of approximately 50% per fertile G_0, while use of 15 kb vectors resulted in 20% frequencies (see Spradling, 1986). Larger vectors could transpose but frequencies approached 1% or less.

Of equal importance to creating an efficient vector system is having marker genes and appropriate host strains that will allow efficient and unambiguous identification or selection of transgenic individuals. Indeed, the genetic resources available for *Drosophila* also provided cloned wild-type DNA and appropriate mutant hosts for use in visible mutant-rescue marker systems that made testing *P* transformation possible. As noted, the first of these used the ry^+ eye color gene, but this required a relatively large genomic fragment of nearly 8 kb. The *white* (*w*) eye color gene was then tested, but this required a genomic sequence that was longer than *ry*, and resultant transformation frequencies were relatively low (Hazelrigg *et al*., 1984; Pirrotta *et al*., 1985). New *w* markers, known as mini-*white*, that had the large first intron deleted decreased the marker insert to 4 kb resulting in much more efficient transformation, and placing the mini-*white* marker under *hsp70* regulation increased efficiencies further (Klemenz *et al*., 1987). Use of *w* markers, especially in CaSpeR vectors (Pirrotta, 1988), has been a mainstay of *Drosophila* transformation, yet expression of the *w* gene in particular is subject to position effect variegation/suppression (PEV) that typically diminishes eye pigmentation. PEV, indeed, was originally discovered as a result of translocating w^+ proximal to heterochromatin (Green, 1996), and it routinely manifests itself in w^- flies transformed with w^+. This effect has been observed with the use of eye pigmentation markers in several other insect species as well.

Other markers based upon chemical selections or enzymatic activity were also developed for *Drosophila*, though none has found routine use. These included alcohol dehydrogenase (*Adh*) (Goldberg *et al*., 1983) and dopa decarboxylase (*Ddc*) (Scholnick *et al*., 1983) that complemented existing mutations, and neomycin phosphotransferase (NPT or *neo*) (Steller and Pirrotta, 1985), β-galactosidase (Lis *et al*., 1983), organophosphorus dehydrogenase (*opd*) (Benedict *et al*., 1995), and dieldrin resistance (*Rdl*) (ffrench-Constant *et al*., 1991) which are dominant selections not requiring preexisting mutations (see ffrench-Constant and Benedict, 2000).

4.13.2.3. *P* Transformation of Non-Drosophilids

Given the straightforward procedures for transforming *Drosophila* with *P* elements, there were high expectations that the system would function in other insects. The ability to test this was facilitated by the development of the neomycin (G418)-resistance marker system (Steller and Pirrotta, 1985), and neomycin resistance-containing *P* vectors were widely tested in tephritid flies and mosquitoes (see Walker, 1990; Handler and O'Brochta, 1991). Unfortunately, the neomycin resistance system was generally unreliable, and recovery of resistant individuals that were not transgenic was common. In three mosquito species, however, neomycin resistant transgenic insects were recovered but they arose from rare transposition-independent recombination events (Miller *et al*., 1987; McGrane *et al*., 1988; Morris *et al*., 1989). Other dominant chemical resistance markers, including *opd* and *Rdl*, which had had some success in *Drosophila*, were also tested, but no transformation events could be verified in other insects. A major limitation of these experiments was that, given the numerous variables involved, it was impossible to determine which components in the system were failing. This limitation led to efforts to determine systematically whether the transposon vector system was, indeed, functional in host embryos, which resulted in the development of rapid transposon mobility assays as described below. The first of these assays tested *P* excision in drosophilid and non-drosophilid embryos, revealing that *P* function decreased in drosophilids as a function of relatedness to *D. melanogaster*, with no function evident in non-drosophilids (O'Brochta and Handler, 1988; Handler *et al*., 1993a). These results were the first indication that for transposon-mediated germline transformation to succeed in non-drosophilids, new vector

systems would have to be created from existing and newly discovered transposon systems.

4.13.3. Excision and Transposition Assays for Vector Mobility

Assessing the ability of an insect gene vector to function in a particular species can be challenging. The procedures required to create a transgenic insect using transposable element-based gene vectors require a great deal of technical skill and the ability to perform basic genetic manipulations. Depending on the insect, its generation time, and its amenability to being reared in the laboratory, the process of genetic transformation can be quite lengthy. At the early stages of developing non-drosophilid transformation technology there was little experience in manipulating and injecting the embryos of the various non-drosophilid species of insects. In addition, the genetic markers available to select for, or recognize, transgenic insects were limited and none could be confidently expected to function optimally in the species being tested at that time. Consequently, early efforts to test the functionality of potential gene vectors, by attempting to create transgenic insects, required simultaneous success in dealing with a number of daunting challenges. The failure of these efforts to yield a transgenic insect could not, unfortunately, be ascribed to the failure of any one particular step in the process (see Handler and O'Brochta, 1991). These efforts, therefore, did not represent an isolated test of the gene vector since failure to obtain a transgenic insect might have been due to a failure in DNA delivery, expression of the genetic marker, or the failure of the transposable element vector system. Technology development under these conditions was very difficult. What was needed was an experimental system that permitted the activity of the transposable element system to be assessed in the species of interest independent of any prospective genetic marker system and DNA delivery system. Such a system was developed for investigating the mobility properties of the *D. melanogaster P* element, and was very adaptable to other transposable element and insect systems (**Figure 2**).

The system developed for *P* elements involved transfecting *Drosophila* cells with a mixture of two plasmids – one containing a *P* element inserted into the coding region of the LacZα peptide of a common cloning vector, and a second containing the *P* element transposase gene under the regulatory control

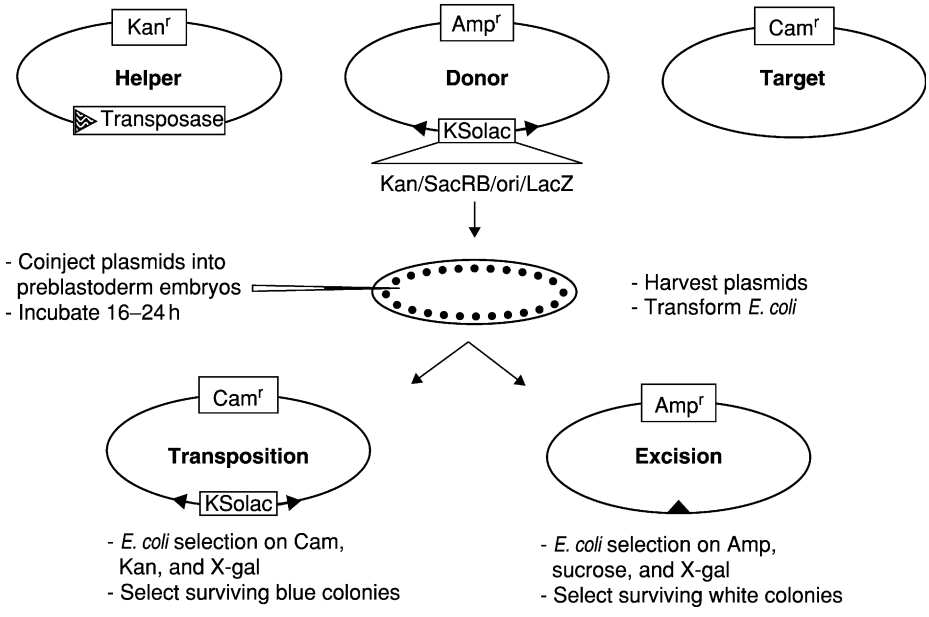


Figure 2 Plasmid-based transposable element mobility assays. A mixture of three plasmids is coinjected into preblastoderm embryos to insure incorporation into nuclei. After approximately 24 h the plasmids are extracted from the embryos and introduced into *E. coli*. Transient expression of the transposase gene on the helper plasmid in the developing embryos results in the production of functional transposase. If the transposase catalyzes excision and transposition of the element, excision will result in the loss of element-specific markers on the donor plasmid. In the example shown sucrose sensitivity, β-galactosidase activity, and kanamycin resistance are lost, and others could be used. Transposition results in the target plasmid acquiring all of the element-specific markers. In this example the target plasmid is from a Gram-positive bacteria and incapable of replicating in *E. coli* unless it acquires the origin of replication present on the element. Assays can be completed in 3 days and rates of movement of 0.001% or greater are routinely detectable.

of a strong promoter (Rio *et al*., 1986). Transient expression of the transposase gene resulted in the production of transposase, catalyzing the excision of *P* elements from the "excision indicator plasmids." Subsequent recovery of the injected plasmids from the cells, followed by their introduction into an appropriate strain of *Escherichia coli,* permitted plasmids that had lost the *P* element through excision to be recognized by virtue of their restored LacZα peptide coding capacity. This transient *P* element excision assay was readily adaptable to use in *Drosophila* embryos through the process of direct microinjection of preblastoderm embryos, and it played a critical role in assessing the functionality of the *P* element system in a variety of drosophilid and non-drosophilid insect systems (O'Brochta and Handler, 1988).

As originally configured the excision assay only permitted the identification and recovery of excision events that resulted in the restoration of the open reading frame of the LacZα peptide reporter gene. Various modifications in this basic assay were adopted that permitted precise and imprecise excisions to be identified and recovered (O'Brochta *et al*., 1991). For example, marker genes such as the *E. coli* LacZα peptide coding region, *E. coli supF*, sucrase (*SacRB*) from *Bacillus subtilis*, and streptomycin sensitivity, were incorporated into the transposable element (O'Brochta *et al*., 1991; Coates *et al*., 1997; Sundararajan *et al*., 1999). Plasmids recovered that lacked marker gene expression were usually excision events. Further refinements of the excision assay involved the use of transposable element-specific restriction endonuclease sites as a means for selecting for excision events. Digesting plasmids recovered from embryos with restriction endonucleases with sites only in the transposable element was a very powerful method of physically removing plasmids that had not undergone excision from the pool of plasmids recovered from embryos and used to transform *E. coli*. Each restriction site was essentially a single dominant genetic marker and therefore transposable elements with multiple restriction sites provided a very powerful system for selecting against plasmids that have not undergone excision (O'Brochta, unpublished data).

Continued development of element mobility assays lead to assays in which interplasmid transposition could be measured. These assays involved the coinjection of a transposase–encoding helper plasmid, an element "donor" plasmid and a "target" plasmid. Typically the target plasmid contains a gene whose inactivation results in a selectable phenotype. For example, the *SacRB* gene has been used since its inactivation eliminates sucrose sensitivity. If the donor element also contains unique genetic markers then transposition events would lead to a recombinant plasmid with a new combination of a variety of markers (Saville *et al*., 1999). Perhaps the most powerful transposition assay developed for assessing transposable elements in insect embryos involved the use of a genetic marker cassette containing a plasmid origin of replication, an antibiotic resistance marker, and the LacZα peptide coding region, in combination with a target consisting of a Gram-positive plasmid (pGDV1) (Sarkar *et al*., 1997a). pGDV1 contains an origin of replication that cannot function in *E. coli,* although it does have a chloramphenicol resistance gene that is functional in this species. Transposition of the marked transposable element into pGDV1 converts it into a functional replicon in *E. coli*. Because of the absolute *cis*-dependence of origins of replication, and the complete inability of pGDV1 to replicate in *E. coli,* transposition events can be readily detected even at low frequencies.

Transient mobility assays are now a standard for defining vector competence in insect embryos, and in particular, when assessing a vector in a species for the first time. For this application, transposition assays provide the most information relevant to the potential for successful germline transformation, and can be used as a system to test helper construct function. As noted below, however, there may be differing constraints on plasmid and chromosomal transpositions for particular transposons. The use of these assays for analyzing transposon function is discussed in more detail in the relevant sections below.

Embryonic assays also provide an essential test system for assessing potential transgene instability by mobilizing or cross-mobilizing systems within a host genome, which is a critical information for the risk analysis of transgenic insects being considered for release. The importance of excision assays for this purpose became evident by the *hobo* excision assays in *Musca domestica* that revealed the existence of the *Hermes* element (Atkinson *et al*., 1993), and the subsequent assays that defined the interaction between the two transposons (Sundararajan *et al*., 1999). Since cross-mobilizing systems do not always promote precise exicisions, assays that reveal imprecise as well as precise excisions are most sensitive for this purpose. Since successful transposition may depend on precise excision, transposition assays may only reveal the existence of mobilizing systems that have a high level of functional relatedness.

4.13.4. Transformation Marker Systems

The availability and development of selectable marker systems has played a large part in recent

advancements in insect transformation, that have been equal to the importance of vector development. The rapid implementation and expansion of *P* transformation in *Drosophila* was possible, in large part, due to the availability of several eye color mutant-rescue systems. These systems depend on the transgenic expression of the dominant-acting wild-type gene for an eye color mutation present in the host strain (see Sarkar and Collins, 2000). Successful transformation of non-drosophilid species was similarly dependent upon the development of analogous systems, with the first transformations in the medfly, *Ceratitis capitata*, and a mosquito, *Aedes aegypti*, relying on *white* and *cinnabar* mutant-rescue systems. While chemical resistance markers were used initially for non-drosophilid transformation, and can be highly useful for specific applications, their inefficiency and inconsistency when used alone provided ambiguous results for several species (see ffrench-Constant and Benedict, 2000). Eye color marker systems are generally efficient and reliable, and cloned wild-type genes from *Drosophila* often complement orthologous mutant alleles in other insects. However, only a handful of species have stable mutant strains that can serve as suitable hosts for mutant-rescue strategies. The most significant advancement in marker gene development for the wide use of insect transformation has been the development of fluorescent protein markers (Higgs and Sinkins, 2000; Horn *et al*., 2002). As dominant-acting neomorphs, that do not depend on preexisting mutations, they are directly useful in almost all host strains. When compared to the *white* eye color marker in *Drosophila*, the enhanced green fluorescent protein (EGFP) gene seemed to be less affected by position effect suppression, and thus has the additional advantage of more reliable detection (Handler and Harrell, 1999). Certainly for the forseeable future, fluorescent protein markers will continue to be the markers of choice for most insect transformation strategies.

4.13.4.1. Eye Color Markers

The first insect transformations used mutant-rescue systems to identify transformant individuals, but in these experiments total genomic DNA was used, rather unreliably, to complement mutations in the respective host strains. The most reliable of these, however, was reversion of the *vermilion* eye color mutation in *D. melanogaster* (Germeraad, 1976). The success of the initial *P* element transformations in *Drosophila* also depended on reversion of eye color mutant strains, but the use of cloned *rosy* and *white* genomic DNA within the vector plasmid allowed for much greater efficiency and reliability. The first non-drosophilid transformations in medfly (Loukeris *et al*., 1995b; Handler *et al*., 1998; Michel *et al*., 2001) similarly relied on use of the wild-type medfly *white* gene cDNA that was placed under *Drosophila hsp70* regulation (Zwiebel *et al*., 1995). This gene complemented a mutant allele in a *white eye* medfly host strain that was isolated more than 20 years earlier. The medfly *white* gene also complemented the orthologous gene mutation in the oriental fruit fly, yielding in one line a nearly complete reversion (Handler and McCombs, 2000). The first transformations in *A. aegypti* used a kynurenine hydroxylase-*white* mutant host strain, but for these tests the complementing marker was the *D. melanogaster cinnabar* gene (Cornel *et al*., 1997). The *D. melanogaster vermilion* gene, that encodes tryptophan oxygensase (*to*) has also been used to complement the orthologous *green* eye color mutation in *M. domestica* (White *et al*., 1996), and the *Anopheles gambiae* tryptophan oxygensase gene complements *vermilion* in *Drosophila* (Besansky *et al*., 1997). The *vermilion* and *cinnabar* orthologs have also been cloned from *Tribolium*, and while the *white* mutation in this species is complemented by tryptophan oxygensase, no preexisting eye color mutation is complemented by kynurenine hydroxylase (Lorenzen *et al*., 2002). The use of eye color mutant-rescue systems has certainly been critical to initial advances in insect transformation, and these markers should have continued utility for those species that have been successfully tested. The use of these markers, however, for the development of insect transformation in other species will be limited by the availability of suitable mutant host strains.

4.13.4.2. Chemical Selections

Previous to the development of mutant-rescue marker systems, transformant selections in non-drosophilid insects focused on genes that could confer resistance to particular chemicals or drugs. Importantly, these types of selections could be used for screening transformants en masse by providing the selectable chemical or drug in culture media. Ideally, only transformant individuals would survive the selection, allowing the rapid screening of large numbers of G_1 insects. For vectors that are inefficient and insects that are difficult to rear, the efficient screening of populations can be essential to identifying transformant individuals. The first drug resistance selection tested used the bacterial neomycin phosphotransferase gene (NPT II or *neomycin*^{r}) that conferred resistance by inactivation of the neomycin analog G418 (or Geneticin) (Steller

and Pirrotta, 1985). This seemed straightforward since the selection and *hsneo* marker system (putting NPT II under heat shock regulation) was already developed and tested in *Drosophila* for mass transformant screens, and the bacterial resistance gene was thought to be functional in most eukaryotes. The initial *P* transformations in *Drosophila* using the pUChsneo vector were generally reliable; however, the marker was not easily transferable to other species. G418 resistance was highly variable, most likely due to species differences in diet, physiology, and symbiotic bacteria, and indeed, variations in resistance in transformed *Drosophila* have been attributed to strains of yeast used in culture media (Ashburner, 1989a). Other chemical resistance markers, including *opd* conferring resistance to paraoxan (Phillips *et al*., 1990; Benedict *et al*., 1995), and the gene for dieldrin resistance (*Rdl*) (ffrench-Constant *et al*., 1991), that were initially tested in *Drosophila* were also problematic when tested in other species. These failures were due in large part to ineffective vector systems, but a common attribute in these studies was the selection of individuals having nonvector related or natural resistance to the respective chemical. While naturally resistant insects could be selected out by molecular tests in primary transformant screeens, the recurrence of resistant insects in subsequent generations would make use of the transgenic strains highly impractical.

While the problems cited made chemical resistance selections frustrating for several species, and they have not been used for any recent transformation experiments, some successes were reported and the need for mass screening still exists. The initial tests for *P* transformation in several mosquito species used the pUChsneo vector with G418 resistant transformants being selected, though transformation frequencies were low and all of them resulted from fortuitous recombination events and not *P*-mediated transposition (Miller *et al*., 1987; McGrane *et al*., 1988; Morris *et al*., 1989). Nonetheless, chemical selections can be very powerful, and if reliable, they would dramatically improve the efficiency of transformation screens for most insects. It is quite possible that many species will not be amenable to current transformation techniques without markers that allow selection en masse. A potential means of increasing the reliability of chemical resistance screens would be to link a resistance marker to a visible marker within the vector. Initial G_1 transformants could be screened en masse by chemical resistance, with surviving individuals verified as transformants and maintained in culture by use of the visible marker. This type of marking is tested by linking the *hsneo* construct with a red fluorescent protein marker in the *piggyBac* vector. Thus far initial results in *Drosophila* are highly encouraging (Handler and Harrell, unpublished data).

Of the enzyme systems tested for chemical selection in *Drosophila* that might be extended to other insects, the *Adh* system might have the most promise (Goldberg *et al*., 1983). An *Adh* marker gene can complement the *adh* mutation in *Drosophila*, eliminating lethal sensitivity to ethanol treatment in mutant hosts. An *adh* gene has been cloned from the medfly, and a strategy has been developed to use it for genetic sexing by male-specific overexpression (Christophides *et al*., 2001). Conceivably a similar strategy could be extended to transformant selections, though its use would be limited to medfly and possibly other tephritid species.

4.13.4.3. Fluorescent Protein Markers

The dramatic advancement of insect transformation in recent years has been due, primarily, to the development of fluorescent protein markers which are dominant-acting neomorphs that do not depend on preexisting mutations. The first of these to be tested was the green fluorescent protein (GFP) gene that was isolated from the jellyfish *Aequorea victoria* (Prasher *et al*., 1992) and which initially exhibited heterologous function in the nematode *Caenorhabditis elegans* (Chalfie *et al*., 1994). GFP expression was then tested in transformant *Drosophila* where it was used as a reporter for gene expression (Plautz *et al*., 1996; Hazelrigg *et al*., 1998), and several other species for both *in vivo* and *in vitro* studies. GFP was first tested in non-drosophilid insects as marker for Sindbis viral infection in *Aedes aegypti* (Higgs *et al*., 1996), and the dramatic somatic expression of GFP in adults was highly encouraging for the further use of GFP for germline transformants.

This possibility was first tested in *Drosophila* using a construct that linked EGFP to a *polyubiquitin* promoter and nuclear localizing sequence (Lee *et al*., 1988; Davis *et al*., 1995; Handler and Harrell, 1999). A control for transformation, that also allowed a direct comparison of EGFP expression to that from the visible mini-*white* marker, was the initial use of a *piggyBac* vector, pB[Dm*w*, PUbnlsEGFP], that linked the two markers in a mutant *white* strain. The results from this experiment indicated that not only was the PUbnslEGFP marker efficient and easily detectable under epifluorescense optics, but that many of the G_1 transformants that expressed GFP did not express a detectable level

of *white*$^{+}$. Presumably the chromosomal position effects that suppressed *white* had a negligible effect on GFP expression. This result was highly encouraging for the use of GFP as a marker in non-drosophilids and several subsequent transformation experiments used EGFP regulated by a variety of promoters in *piggyBac, Hermes*, and *Minos* vectors. Notably, this allowed germline transformation to be tested in several species that otherwise have no visible marker systems, such as the Caribbean fruit fly, *Anastrepha suspensa*, which was transformed with pB[PUbnlsEGFP] (Handler and Harrell, 2000). This vector was subsequently tested in the Australian sheep blowfly, *Lucilia cuprina* (Heinrich *et al*., 2002) and the mosquito *Anopheles albimanus* (Perera *et al*., 2002). Similarly, a *Hermes* vector marked with EGFP regulated by the *Drosophila actin5C* promoter was first tested in *Drosophila* (Pinkerton *et al*., 2000), and was used then to efficiently select transformants in *Aedes aegypti* (Pinkerton *et al*., 2000), *Stomoxys calcitrans* (O'Brochta *et al*., 2000), and *Culex quinquefasciatus* (Allen *et al*., 2001). A *Minos* vector marked with *actin5C*-EGFP was used to select *Anopheles stephensi* transformants (Catteruccia *et al*., 2000b), and a *piggyBac* vector marked with EGFP under *Bombyx actin 3A* promoter regulation was used to transform the lepidopteran species *Bombyx mori* (Tamura *et al*., 2000) and *Pectinophora gossypiella* (Peloquin *et al*., 2000).

Both the *polyubiquitin* and *actin* promoters have activity in all tissues throughout development, making insects marked in this fashion particularly useful for some applications such as the marking of insects used in biocontrol release programs (see Handler, 2002b). However, the detection of these markers can be limited due to quenching or obstruction by melanized cuticle or scales, and fluorescent protein expression regulated by a strong tissue-specific promoter can be valuable for difficult insects and particular applications. Foremost among these markers have been a series of fluorescent protein constructs regulated by the artificial 3xP3 promoter derived from the *Drosophila eyeless* gene (Sheng *et al*., 1997; Horn *et al*., 2000). The markers express strongly from the larval nervous system and adult eyes and ocelli. A 3xP3–EGFP marker within *piggyBac* was first used to transform *D. melanogaster* and *Tribolium castaneum* (Berghammer *et al*., 1999) and has since been used in *M. domestica* (Hediger *et al*., 2000), *Aedes aegypti* (Kokoza *et al*., 2001), and the sawfly *Athalia rosae* (Sumitani *et al*., 2003). The particular strengths and weaknesses for a marker construct such as 3xP3–EGFP is evident from experiments where it allowed transformant selection in *Bombyx* embryos prior to larval hatching (Thomas *et al*., 2002), while it was undetectable in *Aedes aegypti* adults having normal eye pigmentation (Kokoza *et al*., 2001). It must, therefore, be recognized that the utility of fluorescent protein markers must be considered in the context of the host insect's structure and physiology during development.

GFP expression is less sensitive to position effect suppression than eye color markers, yet there is much evidence for quantitative and qualitative variabilities in fluorescent protein expression from transgenes. It is likely that tissue-specific variations are primarily due to position effects, while expression of new tissue phenotypes are due to proximal enhancer effects. Polyubiquitin-regulated EGFP expression is most intense in the thoracic flight mucles in *Drosophila* and tephritid fruit fly adults. In Caribbean fruit fly transgenic adult lines, EGFP was only observed in the thorax, but spectrofluorometric assays revealed as much as fivefold differences in fluorescence between lines having the same number of vector integrations (Handler and Harrell, 1999), and differences in expression that are stable within lines are often observed by inspection. In contrast to typical thoracic expression in tephritid flies, adult PUb-EGFP expression in *Lucilia* was limited to female ovaries (Heinrich *et al*., 2002), and a PUb-DsRed transgenic medfly line exhibits most intense expression in the tarsi while another line expresses in abdominal tracheal apertures at the dorsal–ventral midline (Handler and Krasteva, unpublished data). In *Tribolium*, 3xP3–EGFP expresses typically from the eyes and brain, though several lines are atypical with one having muscle-specific expression throughout development (Lorenzen *et al*., 2003). In *Anopheles stephensi*, the 3xP3–EGFP marker showed atypical expression in the pylorus, epidermal cells and in a subset of cells in the rectum (O'Brochta, Kim, and Koo, unpublished data).

The use of GFP will certainly be continued for transformant identification in many, if not most, other species where transformation is tested. Yet continuing studies in species already transformed will require multiple marking systems that are distinguishable from one another, especially when coexpressed. This will allow the detection of multiple independent transgenes when used in concert for conditional gene expression systems and gene discovery methods such as enhancer traps (Bellen *et al*., 1989; Wilson *et al*., 1989; Brand *et al*., 1994). After testing 3xP3–EGFP, the 3xP3 promoter was linked to the GFP redshifted variants that emit blue (BFP), cyan (CFP), and yellow (YFP) fluorescence; these were

tested in *Drosophila*, and have proven useful individually as reporters and for identifying transformants (Horn and Wimmer, 2000). BFP and GFP have distinct enough emission spectra to be used together, though BFP photobleaches quickly and is not useful for many applications. While use of EGFP with ECFP is also problematic, ECFP and EYFP can be distinguished when using appropriate filter sets. For details on appropriate filter sets for particular applications see Horn *et al.* (2002) and the website for Chroma Technology Corp. (Chroma, 2004) which manufactures filters for most of the stereozoom fluorescence microscopes used for insect studies.

The most spectrally distinct fluorescent protein distinct from GFP and its variants is a red fluorescent protein, known as DsRed, isolated from the Indo-Pacific sea coral *Discosoma striata* (Matz *et al.*, 1999). It was first tested in insects by linking it to the *polyubiquitin* promoter in a *piggyBac* vector (pB[PUb-DsRed1]) and tested in *Drosophila*, where it exhibited highly intense expression (Handler and Harrell, 2001). Importantly, DsRed expression was completely distinguishable from EGFP when the two transgenic lines were interbred, and when coexpressed as an hsp70-Gal4/UAS-DsRed reporter in lines having vectors marked with EGFP. DsRed has since been incorporated into several mosquito and fruit fly species (Nolan *et al.*, 2002; Handler, unpublished data).

Both EGFP and DsRed are highly stable and generally resistant to photobleaching, and could be detected in tephritid flies several weeks after death, though DsRed1 was relatively the more stable of the two. This is highly advantageous for the use of these genes as markers for released insects that might only be retrieved several weeks after death in traps. A drawback for fluorescent proteins, and DsRed in particular, is that they require oligomerization and slow maturation that can take up to 48 h, resulting in low intensity in early development. However, variants of DsRed with shorter maturation times (Campbell *et al.*, 2002), and new fluorescent proteins with enhanced properties for specific applications are becoming available on a consistent basis (see Matz *et al.*, 2002).

4.13.4.3.1. Detection methods for fluorescent proteins Once heterologous expression of GFP in nematodes was discovered it was realized that use of the marker for whole-body analysis of gene expression would require an optical system allowing a large depth of field and a stage with working space for culture plates. Up to this time, most epifluoresence systems were linked to compound or inverted microscopes that had limited field depth and capability to manipulate organisms under observation. This led to the development of an epifluorescence module using a mercury lamp that could be attached to a Leica stereozoom microscope system. Most major microscope manufacturers now market integrated epifluorescent stereozoom microscopes with capabilities for several filter systems.

A lower cost alternative for GFP screening is use of a lamp module using ultra bright blue light emitting diodes (LEDs) with barrier filters that attaches to the objective lens of most stereozoom microscopes (BLS Ltd., Budapest, Hungary). It costs considerably less than a mercury lamp system, but at present, it only has capabilities for detecting GFP and YFP.

The use of fluorescent protein markers, and especially multiple markers will be greatly aided by the use of fluorescence activated embryo sorters. A device first developed to sort *Drosophila* embryos expressing GFP (Furlong *et al.*, 2001) has been modified and commercially marketed for *Drosophila* and other organisms by Union Biometrica (Somerville, MA, USA). The latest sorting machines are highly sensitive having the ability not only to distinguish different fluorescent proteins, but also to discriminate between levels of fluoresence from the same protein. Thus, these systems may have enormous importance to the straightforward screening for transgenics, and more sophisticated assays such as those for enhancer traps. Practical applications could include the screening of released transgenic insects caught in traps (in systems adapted for adults), or for genetic-sexing of embryos having a Y-linked or male-specific fluorescent marker.

4.13.5. Transposon Vectors

4.13.5.1. *Hermes*

4.13.5.1.1. Discovery, description, and characteristics *Hermes* is a member of the *hAT* family of transposable elements and is related to the *hobo* element of *D. melanogaster*, the *Ac* element from maize, *Zea mays*, and the *Tam3* element from *Antirrhinum majus* (Warren *et al.*, 1994). The initial interest in this family of elements by those interested in creating new insect gene vectors stemmed from two observations. First, during the middle and late 1980s the mobility characteristics of the *Ac/Ds* element system were being extensively studied because the element was recognized as having great potential to serve as a gene-analysis and gene-finding tool in maize and other plants. In addition, the mobility properties of *Ac/Ds* were being extensively tested in species of plants other than maize and in almost

every case evidence for *Ac/Ds* mobility was obtained (Fedoroff, 1989). *Ac/Ds* appeared to be a transposable element with a very broad host range, unlike, for example, the *P* element from *D. melanogaster*, which only functions in closely related species (O'Brochta and Handler, 1988). Because transposable elements with broad host ranges were of interest to those attempting to develop insect transformation technology, *Ac*-like elements warranted attention. The second significant observation at this time was that the *hobo* element from *D. melanogaster* had notable DNA sequence similarity to *Ac/Ds*, suggesting that it was a distant relative of this broadly active element (Calvi *et al.*, 1991). Investigation into the host range of *hobo* using plasmid-based mobility assays (as described above) ensued (O'Brochta *et al.*, 1994). It was during the investigation of *hobo* that *Hermes* was discovered (Atkinson *et al.*, 1993). Atkinson *et al.* (1993) performed plasmid-based *hobo* excision assays in embryos of *M. domestica* as part of an initial attempt to assess the host range of *hobo*. Assays were performed in the presence of *hobo*-encoded transposase and *hobo* excision events were recovered suggesting that *hobo*, like *Ac/Ds*, would have a broad host range. However, when the assays were performed without providing *hobo*-encoded transposase *hobo*, excision events were still recovered in *M. domestica*. The movement of *hobo* in the absence of *hobo*-transposase was completely dependent upon the inverted terminal repeats of *hobo* and the resulting excision events had all of the characteristics of a transposase-mediated process. It was proposed that *M. domestica* embryos contained a *hobo* transposase activity and that this activity arose from the transposase gene of an endogenous *hobo*-like transposable element (Atkinson *et al.*, 1993). These investigators were eventually able to confirm their hypothesis and the element they discovered was called *Hermes* (Warren *et al.*, 1994).

Hermes is 2749 bp in length and is organized like other Class II transposable elements in that it contains ITRs and a transposase-coding region (**Figure 1**). It contains 17 bp ITRs with 10 of the distal 12 nucleotides being identical to the 12 bp ITRs of *hobo*. *Hermes* encodes for a transposase with a predicted size of 72 kDa and based on the amino acid sequence is 55% identical and 71% similar to *hobo* transposase (Warren *et al.*, 1994). The cross-mobilization of *hobo* by *Hermes* transposase that was proposed by Atkinson *et al.* (1993) was tested directly by Sundararajan *et al.* (1999). These investigators used plasmid-based excision assays in *D. melanogaster* embryos to show that *hobo* transposase could mobilize *Hermes* elements and that *Hermes* transposase could mobilize *hobo* elements (Sundararajan *et al.*, 1999). The phenomenon of cross-mobilization has important implications for the future use of transposable element-based gene vectors in non-drosophilid insects and will be discussed below. As is typical of transposable elements, *Hermes* is present as a middle repetitive sequence within the genomes of multiple strains of *M. domestica* and in all populations examined there appeared to be full-length copies of the element. The natural history of this element within *M. domestica* has not been investigated and its mobility properties within this species and the existence of any regulatory system remain unknown.

4.13.5.1.2. Patterns of integration The integration behavior of *Hermes* has been examined in a variety of contexts. Sarkar *et al.* (1997a, 1997b) tested the ability of *Hermes* to transpose, using a plasmid-based assay, in five species of Diptera. They recovered transpositions of *Hermes* in the target plasmid at a frequency of approximately 10^{-3} in all species tested. In addition they examined the distribution of 127 independent transposition events into the 2.8 kb plasmid used as a target in their assay and observed a distinctly nonrandom pattern of integrations. Most notable was the existence of three sites that were targets for *Hermes* integration 10 or more times each. In an experiment in which any site used two or more times was considered a hot spot for integration, the three sites used 10 or more times constitute sites with unusual characteristics. The precise nature of those characteristics however could not be defined. The sites shared four of eight nucleotides of the target site in common (GTNNNNAC); however, other sites with this nucleotide composition were not equally attractive as integration sites indicating that other factors must be influencing target choice. Saville *et al.* (1999) demonstrated that sequences flanking *hobo* integration hot spots were critical for determining the targeting characteristics of a site. These investigators were able to move an 8 bp *hobo* target site from plasmid to plasmid without losing its target characteristics as long as they included 20 bp of flanking sequence on each side of the target. It was suggested that proximity to a preferred integration site increased the likelihood of a site being used as a target (Sarkar *et al.*, 1997a). They found that sites 80 and 160 bp flanking the integration hot spot were also preferred integration sites. The authors suggested that nucleosomal organization of the target contributes significantly to the target site selection process and contributed to the local juxtaposition of hot spots and flanking DNA.

4.13.5.1.3. Structure–function relationships Many Class II transposable elements contain a distinct amino acid motif within their catalytic domains consisting of two aspartate residues and a glutamate. This DD35E motif can be found in many but not all Class II transposable elements. The presence of this motif in *Hermes* transposase has been unclear. Bigot *et al*. (1996) proposed the existence of a DDE motif among members of the *hAT* family. However, they proposed that the second aspartate was replaced by a serine in *Ac*, *hobo*, and *Hermes*. Capy *et al*. (1996) concluded that *hAT* elements, like *P* elements from *Drosophila*, do not contain the DDE motif based on sequence alignments, and Lerat *et al*. (1999) supported this conclusion based on the lack of similarity in predicted secondary structure of the transposase of members of the *mariner/Tc* superfamily and *hobo* transposase. Michel *et al*. (2002) examined experimentally the importance of D402, S535, and E572 to the proper functioning of *Hermes* transposase. They found that mutations D402N and E572Q abolished transposase activity while the mutations S535A and S535D had no effect on transposase activity. The work of Michel *et al*. (2002) provided the first experimental data to support the hypothesis that the positive charge of residues D402 and E572 are required for transposition. The authors concluded, based on these data, that D402, S535, and E572 do not constitute the catalytic center of *Hermes* transposase because one of the residues was not essential for activity. Therefore, *Hermes* (and *hAT* elements in general) do not appear to be of the DD35E type of transposable elements, making them distinct from *mariner/Tc* elements.

Because *Hermes* transposase acts within the nucleus it is expected to contain a nuclear localization signal to direct the mature transposase from the ribosome to the nucleus. Deletion and site-directed mutagenesis analysis were performed that demonstrated that the *Hermes* nuclear localization signal is located at the amino acid end of the protein and divided among three domains (Michel and Atkinson, 2003).

The ITRs of transposable elements play an essential role in their mobility. Altering the sequence of ITRs can, depending on the element, lead to loss of function, hyperactivity of the element, or switching the mode of transposition from a cut-and-paste mechanism to a replicative mechanism. *Hermes* contains imperfect ITRs with a 2 bp mismatch within the ITR (Warren *et al*., 1994). In addition, a naturally occurring polymorphism in the terminal nucleotide of the right 3′ ITR exists. Elements with a cytidine in the terminal position of the right ITR have no activity within *D. melanogaster* but are capable of undergoing an aberrant form of transposition in mosquitoes. Small pentanucleotide motifs in the subterminal regions of both *Hermes* and *hobo* have been found to be important for the mobilization of *Hermes* and *hobo*. The sequences GTGGC and GTGAC are interspersed throughout the subterminal region of the element, and similar repeats are present in the subterminal regions of *Ac* and are known to be transposase binding sites. In *Hermes*, altering a single repeat can eliminate transpositional activity (Atkinson *et al*., 2001).

Hermes transposase is capable of dimerizing and one region of the protein critical for dimerization is located in the C-terminus of the protein including amino acids 551–569. This region is not only essential for dimerization but is also required for transposition activity. A second region that affects dimerization is located in the N-terminus of the protein within the first 252 amino acids of the transposase. However, this region apparently plays a nonspecific role in dimerization (Michel *et al*., 2003).

4.13.5.1.4. Host range of *Hermes* *Hermes* has a wide insect host range and has been found to function (as measured by either plasmid-based mobility assays or by germline transformation) in at least 13 species of insects including 11 flies, one beetle, and one moth (Atkinson *et al*., 2001). *Hermes* functions rather efficiently in *D. melanogaster* and transforms this species at rates of 20–40% (O'Brochta *et al*., 1996). In all other species tested the efficiency of transformation was considerably lower and tended to be less than 10%. For example, *T. castenaeum* was transformed at a rate of 1%, *A. aegypti* at 5%, *C. quinquefasciatus* at 11%, *C. capitata* at 3%, and *S. calcitrans* at 4% (Atkinson *et al*., 2001). In all insects except mosquitoes, *Hermes* appeared to use a standard cut-and-paste type mechanism as is typical of most Class II transposable elements. Such integrations are characterized by the movement of only those sequences delimited by the ITRs, and the integrated elements are flanked by direct duplications of 8 bp. Integration of *Hermes* into the germline of *A. aegypti* and *C. quinquefasciatus* appears to occur by a noncanonical mechanism resulting in the integration of DNA sequences originally flanking the element on the donor plasmid. The amount of flanking DNA that accompanies the integration of *Hermes* in these mosquito species varies. In some cases two tandem copies of the *Hermes* element were transferred to the chromosome and each copy was separated by plasmid DNA sequences (Jasinskiene *et al*., 2000). Although

these transposition reactions are unusual they are dependent upon *Hermes* transposase since the introduction of *Hermes*-containing plasmid DNA in the absence of *Hermes* transposase failed to yield transformation events. The germline integration behavior of *Hermes* in mosquitoes is not unique; however, other elements being used as gene vectors such as *mariner* and *piggyBac* have occasionally shown similar behavior in *A. aegypti* (O'Brochta, unpublished data). Transposition assays performed with plasmids in developing mosquito embryos and in mosquito cell lines showed that *Hermes* could transpose via a canonical cut-and-paste type mechanism under these conditions (Sarkar *et al.*, 1997b). The basis for the difference in types of integration events between plasmid-based transposition assays and chromosomal integrations is unknown but may reflect differences in somatic and germ cells. In *Aedes*, canonical cut-and-paste transposition has been readily detected in the somatic tissues of insects containing an autonomous element. Germline transposition in these same insects has not been detected. It has been suggested that mosquitoes might contain endogenous *hAT* elements that affect the ability of *Hermes* elements to be integrated precisely. An alternative suggestion is that *Hermes* may have a second mode of transposition as do the transposable elements *Tn7*, *IS903*, and *Mu*, which utilize a replicative mechanism of integration. Such a mechanism would result in integration products that resemble those observed in the germline of *A. aegypti* and *C. quinquefasciatus*. Replicative transposition of *Hermes* has not been demonstrated experimentally and direct tests of the "alternate mechanism" hypothesis have not been reported.

4.13.5.1.5. Postintegration behavior Once integrated into the genome of *D. melanogaster*, *Hermes* maintains its ability to be remobilized and has shown mobility characteristics that are similar to other transposable elements. Following the introduction of an autonomous *Hermes* element in which the transposase gene was under *hsp70* promoter regulation, and also contained an EGFP marker gene under constitutive regulatory control of the a*ctin5C* promoter, Guimond *et al.* (2003) found that the element continued to transpose in the germline at a rate of 0.03 jumps per element per generation. The element used in this study was also active in the somatic tissue and they used this as a means of collecting approximately 250 independent transposition events. Analysis of somatic integration events revealed a number of interesting patterns. First, they found that transpositions were clustered around the original integration event. On average 39% of the *Hermes* transpositions recovered were intrachromosomal and 17% were within the same numbered polytene chromosome division. Ten percent of the new insertions were at sites within 2 kb of the donor element, indicating that *Hermes*, like other transposable elements, shows the characteristic of local hopping. Local hopping refers to the tendency of some elements to preferentially integrate into closely linked sites. Local hopping has been described for a number of elements and is likely to be a general characteristic of Class II transposable elements although the mechanistic basis for this behavior is unknown. Certain regions of the *D. melanogaster* genome, as defined by numbered divisions of the polytene chromosomes, are preferred as integration sites, with these regions being repeatedly targeted by *Hermes*. The observed clustering of independent transposition events in regions of the chromosome seems to reflect undefined aspects of the transposition process that might be influenced by the chromatin landscape. With one exception, the clustering observed by Guimond *et al.* (2003) was not correlated with any common feature of the chromosomes or the genes within a region. This type of nonrandom pattern of integration with regional differences has also been reported for other elements. Interestingly, there does not seem to be any strong correlation between the preferred insertion-site regions of the elements *P*, *hobo*, and *Hermes*, at least with respect to chromosome 3 of *D. melanogaster* (see figure 7 from Guimond *et al.*, 2003). Guimond *et al.* (2003) also observed a notable clustering of integrations in polytene chromosome division 5. Eight of the 11 integration events recovered from division 5 (3.2% of all the transposition events examined) were within the 2.7 kb segment of DNA upstream of the cytoplasmic actin gene, *actin5C*. This same 2.7 kb segment of the 5′ regulatory region of a*ctin5C* was also present within the autonomous *Hermes* element, as a promoter for the EGFP marker, which the investigators tracked as it jumped within the genome.

The strong clustering of transpositions in a target sequence that is homologous to a sequence contained within the vector has been referred to as "homing." This type of target site selection bias was first described for *P* elements and has been reported on a number of occasions. It was initially reported as a strong bias in the integration site distribution of a number of primary germline integration events in which a *P* element containing the *engrailed* gene preferentially integrated into the *engrailed* region of the host genome (Hama *et al.*, 1990; Kassis *et al.*, 1992). A similar biasing of integration site selection was also observed with *P* elements containing *Bithorax*

and *Antennapedia* regulatory sequences (Engstrom *et al*., 1992; Bender and Hudson, 2000). Taillebourg and Dura (1999) reported a remarkable example of homing of a remobilized *P* element in *D. melanogaster*. This element contained either an 11 kb or 1.6 kb fragment of the 5′ region of the *linotte* gene, and it was found that 20% of the remobilized elements integrated into the 5′ region of the *linotte* gene. Insertions in this case were highly localized and most occurred within a 36 bp fragment of the *linotte* regulatory region. *Hermes* homing indicates that the phenomenon is not element specific, but may be a general characteristic of Class II elements. Guimond *et al*. (2003) suggested that homing was a special case of local hopping, and the physical proximity between donor elements and target sites seems to underlie the phenomenon of local hopping. The presence of transgene regulatory sequences (e.g., *actin5C* 5′ region) may promote tethering of the donor elements to similar regulatory regions via proteins with common DNA binding sites. Deliberate tethering or transposable elements to selected sequences may be a means to regulate target site selection and to minimize the detrimental mutagenic effects of transposable element integration (Bushman, 1994; Kaminski *et al*., 2002).

The postintegration behavior of the same autonomous *Hermes* element described above in *A. aegypti* had quite different characteristics. In this case germline transposition of the autonomous *Hermes* element was never detected, and it should be noted that the primary integration events in the germline involved the integration of DNA sequences flanking the element (Jasinskiene *et al*., 1998, 2000). Despite the fact that the element was intact and that functional transposase was expressed, the element was immobile in the germline. This was not the case, however, in the soma of *A. aegypti* where *Hermes* excision and cut-and-paste transpositions were readily detected. Transposition events in the soma had all of the hallmarks of Class II cut-and-paste integration. Only those sequences precisely delimited by the ITRs moved and integration resulted in the creation of 8 bp direct duplications at the target site. Excision of *Hermes* was imprecise and led, in some cases, to the creation of small deletions. The basis for the difference in behavior of the *Hermes* element in the germline versus the somatic tissue of *A. aegypti* is unknown. Clearly the postintegration behavior of *Hermes* in this species will influence how this element will be employed, and in situations were germline stability is essential, *Hermes* will be particularly useful. It will not be useful in its present form for constructing gene-finding tools such as enhancer and promoter traps that rely heavily on transposable element vector remobilization to be effective.

4.13.5.1.6. Extrachromosomal forms of *Hermes* Excision of *Hermes* in *M. domestica*, and autonomous *Hermes* integrations in *D. melanogaster* and *A. aegypti*, lead to the formation of circularized *Hermes* elements in which the terminal inverted repeats are jointed end-to-end in various ways following the excision reaction (Atkinson and O'Brochta, unpublished data). The most common configuration results in the ends being joined end-to-end with a short spacer sequence between them. The spacer sequence was most often 1, 3, or 4 bp but could also be as much as 200 bp. The extrachromosomal *Hermes* elements found in *M. domestica* are particularly interesting because they have been found in all populations tested and in great abundance in somatic tissue. These data provide evidence for the somatic activity of *Hermes* in the insect from which it was originally isolated. Circularized forms of excised transposable elements of a number of types have been reported in the past (Sundraresan and Freeling, 1987). For example, circularized forms of *Ac/Ds* have been described as well as *Minos* (Arca *et al*., 1997; Gorbunova and Levy, 1997), yet the significance of extrachromosomal forms of transposable elements has remained unclear. In some cases the circularized elements do not contain intact ITRs and consequently the elements are not expected to be integration competent. Based on rather limited data it has generally been concluded that such forms represent byproducts of aborted or interrupted transposition reactions. A study of the extrachromosomal forms of *Hermes* suggests that these elements may have some biological significance. Circularized *Hermes* elements with intact ITRs are integration-competent, potentially allowing them to contribute to forward transposition (Atkinson and O'Brochta, unpublished data). The ability of circularized forms of excised *Hermes* elements to reintegrate may have an impact on the transmission potential of this element. The existence of a large pool of functional extrachromosomal transposable elements may have implications for the ability of the element to be transferred horizontally and may provide an additional means of vertical transmission (e.g., maternally inherited), both of which will potentially enhance the elements ability to increase in frequency within a population. The biology of extrachromosomal *Hermes* elements needs to be investigated further.

4.13.5.1.7. *hAT* elements have been found in other insects The Queensland fruit fly, *Bactrocera tryoni*, contains members of at least two distinct *hAT*-like transposable elements (Pinkerton *et al.*, 1999). *Homer* is a 3789 kb element whose sequence is 53% identical to *Hermes* and 54% identical to *hobo*. The transposase coding region is approximately 53% identical and 71% similar to the transposases of *Hermes* and *hobo*. Similarly, the ITRs of *Homer*, which are 12 bp in length, are identical to those of the *hobo* and *Hermes* elements at 10 of 12 positions. There are also *Homer*-like elements within *B. tryoni*. There are fewer than ten copies per genome, and while these elements have not been fully characterized, a conceptual translation of the transposase of this *Homer*-like element reveals 48% identity and 66% similarity to the transposase of *Homer*. These *Homer*-like elements are as similar to *hobo* as they are to *Homer*. Although *Homer* appears to be weakly functional in *D. melanogaster* based on plasmid-based excision assays, all *Homer*-like elements contain inactivating frameshift mutaions.

The blowfly *L. cuprina* contains a nonfunctional *hAT* element called *hermit*. *Hermit* was initially found by low stringency hybridization screening of an *L. cuprina* genomic library using a DNA probe homologous to *hobo* (Coates *et al.*, 1996). *Hermit* is 2716 bp and contains perfect 15 bp ITRs, the distal 12 of which are identical to the *hobo* ITRs at 10 of 12 positions. Although inactive because of frameshift mutations within the transposase coding region, its amino acid sequence is 42% identical and 64% similar to *hobo* transposase. *Hermit* is unusual in that it is present as a unique sequence within *L. cuprina*, in contrast to multiple copies that exist for most transposons. Although present only once within this species it does appear to have arisen within the genome as a result of transposition since the existing copy of the element is flanked by an 8 bp direct duplication of a sequence that is similar to the consensus target site duplication derived from other *hAT* elements. *Hermit* appears to have become inactivated soon after integrating into the *L. cuprina* genome.

Several *hAT* elements have been discovered in tephritid fruit flies using a polymerase chain reaction (PCR) approach similar to that used to discover *Hermes* (Handler and Gomez, 1996). Of these elements, a complete *hAT* transposon (*hopper*) was isolated from a genomic library of the wild Kahuku strain of the Oriental fruit fly, *B. dorsalis*, using the *B. dorsalis hobo*-related element (Bd-HRE) PCR product as a hybridization probe (Handler and Gomez, 1997). A complete 3120 kb element was isolated having 19 bp ITRs. However, the putative transposase-coding region was frameshifted and it did not have a duplicated 8 bp insertion site, suggesting that it had accumulated mutations and was nonfunctional. The Kahuku sequence was used to isolate additional *hopper* elements using an inverse and direct PCR approach, and a new 3131 bp *hopper* was isolated from the *B. dorsalis white eye* strain (Handler, 2003). This element has an uninterrupted coding region and an 8 bp duplicated insertion site. Notably, *hopper* is highly diverged from all other known insect *hAT* elements and its transposase is distantly yet equally related to the coding regions of *hobo* and *Ac*. Of the terminal 12 nucleotides only five are identical to those of *hobo*, while six are identical to the ITRs of *Homer* (*B. tryoni*). *hopper* also exists in the melonfly, *B. cucurbitae*, and another *hAT* element originally discovered in the melonfly, that is closely related to *hobo* and *Hermes*, also exists in *B. dorsalis* (Handler and Gomez, 1996).

hAT elements have been also reported in the human malaria vector *A. gambiae*. Approximately 25 copies of sequences that resemble *hAT* transposases were discovered although none appeared to be part of an intact transposable element. More recently, however, search criteria were used based on unique aspects of *hAT* transposable elements such as length and spacing of ITRs and the characteristics of *hAT* element target sites. This search revealed a *hAT* element in *A. gambiae* that contained perfect 12 bp ITRs flanked by 8 bp direct duplications and a 603 amino acid transposase open reading frame that appeared to contain no internal stop codons. This element (*Herves*) is most closely related to *hopper* and the ability of this element to excise and transpose in *A. gambiae* or other species has not been determined (Atkinson and Arenburger, personal communication).

4.13.5.2. *piggyBac*

4.13.5.2.1. Discovery of *piggyBac* and other TTAA-specific elements Similar to several other insect transposable element systems, the *piggyBac* element was discovered fortuitously in association with a mutant phenotype. However, unlike all the other transposons used for insect transformation, the mutant phenotype was the result of a functional element that had transposed into an infectious organism. Fraser and colleagues (see Fraser, 2000) discovered several Few Polyhedra (FP) mutations in the baculoviruses, *Autographa californica* nucleopolyhedrovirus (AcNPV), and *Galleria mellonella* nucleopolyhedrovirus (GmNPV), after passage through the *Trichoplusia ni* cell line TN-368 (Fraser *et al.*, 1983, 1985). Among these elements that inserted specifically into tetranucleotide TTAA sites

was *piggyBac* (then named IFP2), which transposed into AcNPV. Although it might be assumed that IFP2 was an autonomous functional element based on its mobility, another TTAA insertion-site element, *tagalong* (then called TFP3), discovered in AcNPV and GmNPV, was later found not to have a transposase coding region and thus had to be mobilized by another TFP3 or related element. Autonomous functional elements have not yet been found for *tagalong*, though the original IFP2 *piggyBac* element was indeed functional (Wang *et al*., 1989; Wang and Fraser, 1993). All the *piggyBac* elements discovered in TN-368 were found to be identical, having a length of 2472 kb with 13 bp perfect ITRs and 19 bp subterminal repeats located 31 bp from the 5′ ITR and 3 bp from the 3′ ITR (Cary *et al*., 1989) (**Figure 1**). The transposase coding region exists as a single reading frame of 2.1 kb that encodes a protein with a predicted molecular mass of 64 kDa. The functionality of *piggyBac* and the precise nature of its transposition was further verified by a series of viral and plasmid transposition and excision assays. A *piggyBac* indicator plasmid was marked with *polh/lacZ*, and assays in the fall armyworm, *Spodoptera frugiperda* cell line SF21AE showed that the original *piggyBac* element, within the p3E1.2 plasmid, could mobilize the marked element. These assays proved that the 3E1 *piggyBac* element encoded a functional transposase, and defined the element's TTAA insertion-site specificity and the precise nature of its transposition. Importantly, these assays also showed directly that *piggyBac* could be mobilized in other lepidopteran species (Fraser *et al*., 1995), indicating that it might function similarly as a vector for germline transformation. This was a critical realization given the failure of *P* to be mobilized in non-drosophilids, which was consistent with its failure as a vector in these species.

4.13.5.2.2. *piggyBac* transformation The failure of *P* vectors to transform non-drosophilid species made the testing of other available transposon systems a high priority. The other systems found to be functional in non-drosophilids, however, were first tested successfully for gene transfer vector function in *Drosophila*. For *piggyBac*, germline transformation was first attempted in the Mediterranean fruit fly, *Ceratitis capitata*. This was possible due to the availability of a marker system that had been tested previously by medfly transformation with the *Minos* transposon vector. The medfly *white* gene cDNA was linked to the *Drosophila hsp70* promoter, and was used as a mutant-rescue system in a white eye host strain (Loukeris *et al*., 1995b; Zwiebel *et al*., 1995). In the absence of data for the minimal sequence requirements for *piggyBac* mobility, the first *piggyBac* vector was constructed by insertion of the 3.6 kb *hsp-white* cDNA marker into the unique *Hpa*I site within *piggyBac* in the p3E1.2 plasmid. None of the *piggyBac* sequence was deleted though the insertion interrupted the coding region eliminating transposase function. Construction of the first helper was a simple deletion of the 5′ ITR sequence resulting from a *Sac*I digestion and religation of p3E1.2. There is some uncertainty as to whether the upstream *Sac*I site cuts within the *piggyBac* promoter (Cary *et al*., 1989), yet transposase expression was indeed sufficient to support germline transpositions from the vector plasmid. The first experiment with this helper in the medfly resulted in one transgenic line at a transformation frequency of 5% per fertile G_0. However, sibling sublines exhibited two and three independent integrations (Handler *et al*., 1998). This experiment with a *piggyBac*-regulated helper was repeated with five additional G_1 lines isolated, but at approximately the same frequency. These attempts at *piggyBac* transformation yielded relatively low transformation frequencies, but it was notable that a lepidopteran transposon vector system had autonomous function in a dipteran species.

Subsequent to the medfly transformation, *piggyBac* transformation was tested in *Drosophila* using the mini-*white* marker from that species (Handler and Harrell, 1999). Using the self-regulated helper, transformants were isolated at a similar frequency of 1–3%, but tests with a *hsp70*-regulated transposase increased the frequency to above 25%, consistent with *P* and *hobo* transformations using heat shock promoted transposase.

Given that *piggyBac* was first isolated from a lepidopteran species, there was some optimism that it would be functional as a vector in other moth species. Function was first tested by transposition assays in the pink bollworm, *Pectinophora gossypiella* (Thibault *et al*., 1999), which then led to successful germline transformation of this species using the phspBac helper and a vector marked with EGFP regulated by the *Bombyx actinA3* promoter (Peloquin *et al*., 2000). Concurrent experiments were also performed in the silkmoth *B. mori* using a similar *actinA3*-regulated EGFP marker, but for this species transformation was achieved with an *actinA3*-regulated transposase helper (Tamura *et al*., 2000). While these are the only moth species reported to be transformed with *piggyBac*, several other dipteran species have been transformed, as well as species in the orders Coleoptera and Hymenoptera. The dipteran species transformed include

several of medical and agricultural importance, such as the mosquitoes *Aedes aegypti* (Kokoza *et al.*, 2001; Lobo *et al.*, 2002), *A. gambiae* (Grossman *et al.*, 2001), *A. albimanus* (Perera *et al.*, 2002), and *A. stephensi* (Nolan *et al.*, 2002), and the tephritid fruit flies *Anastrepha suspensa* (Handler and Harrell, 2000) and *Bactrocera dorsalis* (Handler and McCombs, 2000). Other transformed dipterans include *M. domestica* (Hediger *et al.*, 2000) and *L. cuprina* (Heinrich *et al.*, 2002). Of particular importance at this time, has been the use of *piggyBac* to transform a coleopteran, the red flour beetle, *Tribolium castaneum* (Berghammer *et al.*, 1999; Lorenzen *et al.*, 2003), and a hymenopteran, the sawfly *Athalia rosae* (Sumitani *et al.*, 2003). Notably, all of these species were primarily transformed using a helper regulated by the *Drosophila hsp70* promoter, and with vectors marked with EGFP, though other fluorescent proteins have since been used for some as well.

Although most of these transformations occurred at frequencies between 3% and 5% per fertile G_0, dramatic differences between species have been observed as well, and in some of the same species performed by different laboratories. A single transformant line was reported for *Anopheles gambiae*, at a frequency of approximately 1% (Grossman *et al.*, 2001), while transformation in *A. albimanus* occurred at frequencies ranging from 20% to 40% (Perera *et al.*, 2002). The first transformations of *Tribolium* occurred at an unusually high frequency of 60% (Berghammer *et al.*, 1999).

Many of the transformations were preceded by testing *piggyBac* function by embryonic transposition assays that were first developed for *piggyBac* mobility in the pink bollworm (Thibault *et al.*, 1999). As discussed previously, these assays can rapidly assess the relative mobility of *piggyBac* in a specific host species in a few days. Positive results from these assays provided some assurance that more tedious and time-consuming transformation experiments had some likelihood of success. For some studies the assays also were used to test promoter function in helper plasmids, or provided insights into insertion site specificity, or determined the likelihood of a particular vector construct retaining function in the absence of specific sequences (Lobo *et al.*, 2001). For example, *piggyBac* helper promoters were tested by transposition assays and germline transformation in *D. melanogaster* and *L. cuprina* (Li *et al.*, 2001a; Heinrich *et al.*, 2002). It was found that in *Drosophila*, an *hsp70*-regulated helper yielded the highest transposition frequency, while a constitutive *α1-tubulin*-regulated helper was more effective for germline transformation. By comparison, in *Lucilia* the *hsp70* helper was most effective for both plasmid and germline transpositions, while the *Drosophila α1-tub* helper failed to support transformation. Transposition assays have also shown target site preferences among the TTAA sites within the pGDV1 target plasmid, and assays in *Drosophila* indicated a bias for sites having A or T nucleotides at positions −3, −1, +1, and +3 relative to TTAA (Li *et al.*, 2001a). However, a sequence analysis of 45 genomic integrations sites in *Tribolium*, after *piggyBac* vector remobilization, failed to show this bias (Lorenzen *et al.*, 2003), which may be an indication of species specificity for insertion site preference.

Mobility assays also provide a rapid means of testing sequence requirements for vector mobility, which allow modifications for more efficient vector function. Since vector mobility is known to be affected negatively with increasing size, this information should allow minimal vectors to be created that retain optimal function. However, minimal sequence requirements for plasmid transpositions may differ from those for chromosomal transposition. For example, excision and transposition assays performed in *Trichoplusia ni* embryos showed that the *piggyBac* inverted terminal repeat and subterminal repeat sequences were sufficient for transposition (35 bp from the 5′ terminus and 63 bp from the 3′ terminus), but that an outside spacer region between the ITRs of greater than 40 bp is necessary for optimal transposition from a plasmid (Li *et al.*, 2001b). Use of similar vectors in *Drosophila*, however, did not result in germline transformants (Handler, unpublished data). The minimal sequence requirements for *piggyBac* transformation verified thus far for *Drosophila* are 300 bp from the 5′ terminus and 250 bp from the 3′ terminus (Li, Fraser, and Handler, unpublished data).

4.13.5.2.3. Phylogenetic distribution of *piggyBac* and implications for transgene stability Unlike most other transposons used for transformation, *piggyBac* is not an apparent member of a larger family, or superfamily of related elements such as the *mariner/Tc* or *hAT* families. Until recently, the only *piggyBac* elements known were the functional elements orginally discovered in *T. ni* (Fraser *et al.*, 1983). Thus, an unexpected finding from the Southern analysis of *B. dorsalis* transformants was that 8 to 10 *piggyBac*-related elements exist in the host strain genome (Handler and McCombs, 2000). PCR analysis of internal coding sequence indicated that these were nearly identical elements, though none has been found to be identical to *piggyBac*, nor are their coding regions consistent with transposase

functionality. The isolation of complete *B. dorsalis piggyBac* elements as genomic clones and by inverse PCR indicate that complete elements exist with conserved terminal and subterminal sequences that are integrated into duplicated TTAA insertion sites. *Bactrocera dorsalis* is part of a larger species complex and PCR analysis indicated that, indeed, *piggyBac* exists throughout the complex having nucleotide sequence identities of 92% among all the elements sequenced, with individual identities of 96–99% to one another and the *T. ni* 3E1 *piggyBac* (Handler, unpublished data).

4.13.5.2.4. *piggyBac* in other insects The evolutionary distance between *T. ni* and *Bactrocera* strongly suggests that the transposon moved between these species by recent horizontal transmission, and the separation of their geographical habitats raises the possibility that this movement may have been mediated by intermediary species. A Southern blot survey for *piggyBac* in more than 50 species showed the most clear evidence for multiple *piggyBac* elements in *S. frugiperda*, but hybridization patterns suggested that most of the elements are defective and nonfunctional (Handler, unpublished data; see Handler, 2002a). Evidence for *piggyBac* in other insects and other organisms, including mosquitoes and humans, comes from recent sequence data from genome projects. Although discrete sequence similarities suggests that *piggyBac* has an ancient history, there is little evidence at present to indicate that *piggyBacs* have coevolved as functional elements, and related complete elements have yet to be discovered.

The first functional *piggyBac* elements were discovered in a *T. ni* cell line, but little analysis has been done to characterize *piggyBac* in the organismal genome. Recent hybridization analysis of *piggyBac* from larval *T. ni* genomes indicates that *piggyBac* exists, and that its general structure is consistent with full-length functional elements. However, a PCR survey of these genomic sequences has only identified nonfunctional elements, having a level of identity no greater than many of the elements found in *Bactrocera* (>96%) (Zimowska and Handler, unpublished data). If functional *piggyBacs* do not exist *in vivo*, this could be explained by genomic instability that may arise from a highly active transposon. Functional elements may create a genetic load resulting in organismal lethality, which is more easily withstood in cell lines. If this is the case, it is therefore intriguing to consider how the functional element arose in the cell line.

It is also intriguing to consider how horizontal transmission of *piggyBac* may have occurred, considering that the element was originally discovered by virtue of its transposition into an infectious baculovirus. This could potentially explain a distribution among lepidopterans, but not the apparent recent movement between moths and flies. Understanding the interspecies movement of *piggyBac*, as well as all other vectors used for practical application, will be critical to undertanding and eliminating risk associated with the release of transgenic insects.

4.13.5.3. *mariner*

4.13.5.3.1. Discovery, description, and characteristics The *mariner* element was first discovered as an insertion element responsible for the *white*-peach (w^{pch}) mutant allele of *D. mauritiana* (Haymer and Marsh, 1986; Jacobson *et al.*, 1986). This particular allele was interesting when discovered because it was highly unstable with reversions to wild type occurring at a frequency of approximately 10^{-3} per gene per generation. *white*-peach individuals also had a high frequency of mosaic eyes, at an approximate frequency of 10^{-3}, suggesting somatic instability. Molecular analysis of the w^{pch} allele indicated that it was the result of a 1286 bp transposable element insertion into the 5′ untranslated leader region of the *white* gene (Jacobson *et al.*, 1986) (**Figure 1**). The *mariner* element is a Class II type transposable element with 28 bp imperfect inverted repeats with four mismatches. The element recovered from w^{pch} contained a single open reading frame capable of encoding a 346 amino acid polypeptide (Jacobson *et al.*, 1986). While the original w^{pch} was highly unstable, another strain of *D. mauritiana* was discovered in which mosaicism of the eyes occurred in every fly (Bryan *et al.*, 1987). This mosacism factor was found to be heritable and was referred to as *Mos1* (Mosaic eyes). *Mos1* was a dominant autosomal factor on chromosome 3 and was subsequently found to be identical to *mariner* except for six amino acid differences in the putative transposase coding region (Medhora *et al.*, 1988). *Mos1* encodes for a functional transposase while the 346 amino acid polypeptide of the w^{pch} *mariner* element was not a functional transposase.

One of the most notable characteristics of *mariner* and *mariner*-like elements (*MLE*s) is their widespread distribution. *MLE*s are found not only in insects and invertebrates but also in vertebrates and plants (Robertson, 2000; Robertson and Zumpano, 1997). Not long after the *D. mauritiana mariner* elements were described, a related element was discovered in the *cecropin* gene of the moth *Hyalophora cecropia* (Lidholm *et al.*, 1991). Based on the sequence comparison between the *mariner* elements from *D. mauritiana* and *H. cecropia*, Robertson (1993) designed degenerate PCR primers

and surveyed 404 species of insects for the presence of related sequences. He found that 64 of the genomes examined contained *MLE*s, and within this group are five subgroups referred to as the *mauritiana, cecropia, mellifera, irritans*, and *capitata* subgroups (Robertson and MacLeod, 1993). Since that original analysis insect *MLE*s have continued to be discovered and currently there are two additional subgroups recognized, known as *mori* and *briggsae* (Lampe *et al*., 2000). Additional subgroups are likely to be recognized in the future as additional representatives of this family of elements are found. Elements from different subgroups are typically about 50% identical at the nucleotide sequence level while the transposases encoded by elements from different subgroups are usually between 25% and 45% identical at the amino acid level. A notable feature of the phylogenetic relationships of the *MLE*s is their incongruence with the phylogenetic relationships of the insects from which they were isolated. The implication is that many of these elements were introduced into their host genome via a horizontal gene transfer event (Robertson and Lampe, 1995a). The abundant examples of horizontal transfer of *mariner* elements have led to the conclusion that such transfers occur relatively frequently. Hartl *et al*. (1997) estimated that the rate of horizontal transmission of *MLE*s is about the same as the rate of speciation, at least within the *D. melanogaster* species subgroup. The widespread occurrence of horizontal transmission of *MLE*s has been proposed to be critical for the long-term survival of these elements. Horizontal transmission provides a means for invading naive genomes where element proliferation can occur before inactivating influences of mutation and host regulation can occur (Hartl *et al*., 1997).

Although hundreds of *MLE*s have been reported, only two (*Mos1* from *D. mauritiana* and *Himar1* from *Haematobia irritans*) have been demonstrated to be functional or active. *Haemotobia irritans* contains approximately 17 000 copies of *Himar1*, although all of the copies examined were highly defective. Functional elements could be reconstructed based on the consensus sequence of *Himar1* and then constructed by modifying the closely related *Cpmar1* element from the green lacewing, *Chrysoperla plorabunda,* to match the *Himar* consensus sequence (Robertson and Lampe, 1995b; Lampe *et al*., 1998). Purification of the transposase from a bacterial expression system and its use in an *in vitro* mobility assay demonstrated the functionality of the *Himar1* protein and the ITRs of the element (Lampe *et al*., 1996).

4.13.5.3.2. Structure–function relationships The transposases of *MLE*s belong to a large group of integrases and transposases that share a significant feature of their catalytic domains. Specifically, *MLE*s contain the highly conserved DD35E motif within the active site of the protein (Robertson, 2000). This part of the active site interacts with a divalent cation that is essential for catalysis. Transposase binds to the ITRs of the element, and gel retardation assays were used to assess the binding activity of eight mutant transposases with deletions at the N- or C-termini (Auge-Gouillou *et al*., 2001a). It was possible to show that amino acids 1–141 were sufficient for binding to the ITRs. The ITR binding domain of *Mos1* transposase differs somewhat from that of *Tc1* elements in that it is composed of two different structural motifs, a helix–turn–helix motif and an α-helical region (Auge-Gouillou *et al*., 2001a).

The ITRs of *Mos1* are not identical and differ in sequence at four positions, which have effects on the activity of the element *in vitro*. Auge-Gouillou *et al*. (2001b) reported a 10-fold higher affinity of *Mos1* transposase for the 3′ ITR compared to the 5′ ITR. In addition, modified 5′ ITRs that were made to resemble 3′ ITRs at one of the four variable positions resulted in an increase in transposase binding. These investigators also showed that a *Mos1* element with two 3′ ITRs had 104 times the transposition activity of the native ITRs (Auge-Gouillou *et al*., 2001b). This hyperactive double-ended configuration has not been tested *in vivo*.

Hyperactive transposase mutants of the *Himar1* transposase have been reported (Lampe *et al*., 1999) and one of the mutants contains two amino acid changes (at positions 131 and 137) in the ITR binding domain of the protein. Although not tested directly, it is possible that these hyperactive mutants result in increased binding of the transposase and consequently higher rates of movement. Paradoxically, neither *Himar1* nor any of the hyperactive muatants shows any transpositional activity in insects (Lampe *et al*., 2000).

4.13.5.3.3. Host range of *mariner* The widespread distribution of *MLE*s in nature and the frequent examples of their horizontal transfer between species suggest that these elements have a broad host range. Empirical studies in which *Mos1* has been employed as a gene vector in a wide variety of organisms supports this conclusion. *Mos1* has been used successfully to create transgenic *D. melanogaster* (Lidholm *et al*., 1993), *D. virilis* (Lohe and Hartl, 1996a), and *Aedes aegypti* (Coates *et al*.,

1998). In each of these species the frequency of transformation was approximately 5%. This element has also been used to create transgenic *B. mori* cells in culture (Wang *et al*., 2000). In addition to transgenic insects, *Mos1* has been used to create transgenic *Leishmania* (Gueiros-Filho and Beverley, 1997), *Plasmodium* (Mamoun *et al*., 2000), zebrafish (Fadool *et al*., 1998), and chickens (Sherman *et al*., 1998). Similarly the *Himar1* element has been shown to function in *E. coli* (Rubin *et al*., 1999), Archaebacteria (Zhang *et al*., 2000), and human cells (Zhang *et al*., 1998). However, this element has not been shown to be active in *D. melanogaster* or any other insect species, for reasons that are not clear (Lampe *et al*., 2000).

4.13.5.3.4. Postintegration behavior The postintegration behavior of *Mos1* has been investigated in *D. melanogaster* and *A. aegypti*. *mariner* gene vectors used to create transgenic *D. melanogaster* have been found to be uncommonly stable even in the presence of functional transposase. Lidholm *et al*. (1993) created two lines of transgenic *D. melanogaster* with a *mariner* vector derived from *Mos1* and containing the mini-*white* gene as a genetic marker. When these lines were crossed to *Mos1* transposase-expressing lines, eye mosaicism was found in only 1% of the progeny, while these same *Mos1* expressing lines resulted in 100% mosaicism of the w^{pch} element. Similarly, germline transposition occurred at rates of less than 1% (Lidholm *et al*., 1993), and Lohe *et al*. (1995) reported similar evidence for postintegration stability of *mariner* vectors. Lozovsky *et al*. (2002) suggested, after investigating the postintegration mobility of a number of *mariner* vectors containing different genetic markers in different locations within the element, that *mariner* mobility is highly dependent upon critical spacing of subterminal sequences and ITRs. They found that vectors with simple insertions of exogenous DNA of varying lengths and in varying positions showed levels of somatic and germline excision that were at least 100-fold lower than that observed with uninterrupted *mariner* elements. Only vectors consisting of two, almost complete, elements flanking the marker gene showed detectable levels of both somatic and germline mobility. Approximately 10% of the insects with these composite vectors had mosaic eyes when transposase was present. Germline excision rates of approximately 0.04% were observed in these same insects. Again, these values are considerably less than those reported for uninterrupted elements. In addition to the potential importance of subterminal sequence spacing (Lozovsky *et al*., 2002), Lohe and Hartl (2002) suggested that efficient mobilization of *mariner in vivo* also depends on the presence of critical sequences located quite distant from the ITRs. Based on the mobility characteristics of about 20 *mariner* elements with a wide range of internal deletions, they concluded that there are three regions within the element that play an important role in *cis*. Region I is approximately 350 bp in length and is located 200 bp from the left 5′ ITR. Region II is approximately 50 bp in length and located approximately 500 bp from the right 3′ ITR. Region III is about 125 nucleotides in length and located approximately 200 bp from the right ITR (Lohe and Hartl, 2002). While the presence of subterminal sequences that play a critical role in the movement of many Class II transposable elements is not unusual, what is uncommon in the case of *mariner* is the location of these *cis*-critical sequences. Their dispersed distribution within the element is unique, and consequently, manipulating the element for the purposes of creating gene vectors and associated tools without disrupting these important relationships may be difficult.

The postintegration mobility of *Mos1* can also be regulated by nonstructural aspects of the system including "overproduction inhibition" and "dominant-negative complementation." Increasing the copy number of *Mos1* in the genome resulted in a 25% decrease in the rate of germline excision. Copy number increases in *Mos1* presumably lead to increased transposase levels and, by an unknown mechanism, to the inhibition of excision (Lohe and Hartl, 1996b). High concentrations of transposase may lead to nonspecific associations of the protein resulting in inactive oligomers of transposase. In addition, the presence of mutated forms of *Mos1* transposase can repress the activity of functional transposase. Because the transposases of other transposable elements act as dimers or multimers it is thought that mutated *Mos1* transposases may become incorporated into multimers with functional transposases, thereby inactivating the entire complex (Lohe and Hartl, 1996b).

The possibility that transposase overproduction may negatively affect its own activity is a highly important concept in terms of vector system development. Most systems have the helper transposase under strong promoter regulation to optimize transpositional activity, though this may, indeed, be counterproductive. For *mariner* vectors, and potentially other systems, optimal transformation may require testing various helper promoters and a range of plasmid concentrations.

The postintegration mobility properties of *mariner* were also examined in *A. aegypti* (Wilson *et al*.,

2003). As part of an effort to create an enhancer trapping and gene discovery technology for *A. aegypti*, they created nonautonomous *mariner*-containing lines and lines expressing *Mos1* transposase. By creating heterozygotes between these two lines, they attempted to detect and recover germline transposition events, but only a single germline transposition event was recovered after screening 14 000 progeny. Somatic transpositions were detected, and while precise estimates of rates of somatic transposition were not possible because of the detection method, the authors observed fewer than one event per individual which they estimated to be an indication of a very low rate of movement. The vectors used by Wilson *et al.* (2003) resembled the simple vectors reported by Lozovsky *et al.* (2002) which had apparently disrupted spacing of the ITRs, and partial deletions of *cis*-critical sequences described by Lohe and Hartl (2002).

While the postintegration stability of *mariner* has been described in two species and appears to be a general mobility characteristic of this element, and not a reflection of a species-specific host effect, paradoxical observations remain to be explained. First, the use of *mariner* as a primary germline transformation vector in non-drosophilid insects and in non-insect systems is an effective means for creating transgenic organisms. Indeed, the host range of *mariner* as a gene transformation vector is unrivaled by any of the other gene vectors currently employed for insect transformation. *mariner* has been used as a gene vector in microbes, protozoans, insects, and vertebrates. The rates of germline transformation using *mariner*-based vectors in insects is approximately 10% or less, and is comparable to the efficiency of *Hermes*, *Minos*, and *piggyBac* gene vectors. This raises the question of whether *mariner* vectors present on plasmids behave the same as *mariner* vectors integrated into insect chromosomes. Given the rates of germline integration from plasmids it appears that the *mariner* vectors being used are not suffering from "critical spacing/critical sequence" defects. In addition, the *in vitro* behavior of *mariner* also differs from the behavior of chromosomally integrated elements. Tosi and Beverly (2000) demonstrated that only 64 nucleotides from the left end, and 33 nucleotides from the right end, of *mariner* were essential for transposition of a 1.1 kb vector *in vitro*. The rate of transposition of a minimal *mariner* vector *in vitro* was only two-fold less than that of a vector containing essentially a complete *mariner* element. These results suggest that *mariner* mobility has relatively simple sequence requirements and that the role of subterminal sequences is minimal *in vitro*. These apparently conflicting data suggest that host factors may play an important role in the transposition process *in vivo*, and may influence the relative importance of *cis* sequences in the *mariner* transposition process. The broad distribution of *MLE*s and host range of *mariner*/*Mos1* suggest, however, that host factors play little role in the movement of these elements.

The postintegration behavior of *mariner*/*Mos1* seems to indicate that this element will not be a good candidate for developing gene-finding tools such as promoter/enhancer trapping and transposon tagging systems in *A. aegypti* or perhaps other insects. On the other hand, if a high level of post-integration stability is desired, then *mariner* is an appropriate element to consider in insects. The potential of this element to be lost through excision or transposition is low, even in the presence of functional *mariner* transposase. As currently configured and used, *mariner* vectors may be considered as suicide vectors in insects since they essentially become dysfunctional upon integration.

4.13.5.3.5. *MLE*s have been found in other insects While hundreds of *MLE*s have been described, few have been shown to be functional. The original *mariner* element from the *white*-peach allele was transpositionally competent although it did not produce a functional transposase. *Mos1* is a functional autonomous element and has been the basis for constructing all *mariner* gene vectors that function in insects. *Himar1* is a functional element from the *irritans* subgroup that was reconstructed based on multiple sequence comparisons of elements within this group. It has not been shown to be functional in insects despite significant efforts to do so. Lampe *et al.* (2000) report that at least eight other elements from the other subgroups are likely to be active or made active by minor modifications.

4.13.5.4. *Minos*

The first germline transformation of a non-drosophilid insect mediated by a transposon-based vector system was achieved with the *Minos* element. *Minos* was originally isolated as a fortuitous discovery in *D. hydei* during the sequencing of the non-coding region of a ribosomal gene (Franz and Savakis, 1991). *Minos* was found to be a 1.4 kb element having, unlike the other Class II transposons used as vectors, relatively long ITRs of 255 bp, with its transcriptional unit consisting of two exons (**Figure 1**). Additional *Minos* elements were isolated from *D. hydei* having small variations of one or two nucleotides, though the new elements had a transition change that restored the normal reading frame allowing translation of a functional transposase.

The sequence homology, general structure, and TA insertion-site specificty placed *Minos* within the *Tc* transposon family (Franz *et al*., 1994). *Minos* was first used to transform *D. melanogaster* with *Minos*-mediated events demonstrated by sequencing insertion sites and remobilization of integrations (Loukeris *et al*., 1995a). The first non-drosophilid transformation with *Minos* was achieved in a medfly *white eye* host strain using a cDNA clone for the medfly *white* gene as a marker (Zwiebel *et al*., 1995), at an approximate frequency of 1–3% per fertile G_0 (Loukeris *et al*., 1995b). *Minos* transposition was subsequently demonstrated in dipteran and lepidopteran cell lines (Klinakis *et al*., 2000; Catteruccia *et al*., 2000a), with germline transformation reported for *Anopheles stephensi* (Catteruccia *et al*., 2000b) and *D. virilis* (Megna and Cline, personal communication). Recently transformation frequencies have been substantialy increased in *Drosophila* and medfly by the use of *in vitro* synthesized transposase mRNA as helper (Kapetanaki *et al*., 2002).

Although *Minos* has not been widely used for insect transformation, embryonic and cell line mobility assays in several insect species in the Diptera, Lepidoptera, and Orthoptera have indicated a broad range of function. Notably, *Minos* transposition in the cricket *Gryllus bimaculatus* was driven by transposase regulated by a *Gryllus* actin gene promoter, and not by the *Drosophila hsp70* promoter that has been widely used in dipterans (Zhang *et al*., 2002). The broad function of the *Minos* vector is further supported by its ability to transpose in a mouse germline (Drabek *et al*., 2003).

Minos structure places it within the *mariner*/*Tc* transposon superfamily, though knowledge of the distribution of *Minos* is thus far limited to the genus *Drosophila* (Arca and Savakis, 2000). In *Drosophila*, *Minos* is clearly widely distributed in the *Drosophila* and *Sophophora* subgenera, though discontinuously in the *Sophophora*. As noted for the *hAT*, *mariner*, and *piggyBac* elements, *Minos* may have also undergone horizontal transfer between *Drosophila* species.

4.13.5.5. *Tn5*

Tn5 is one of a number of very well-characterized transposable elements from prokaryotes. Recently, hyperactive forms of this element have been created in the laboratory that have proven to be the basis for the development of a number of commercially useful genomics tools (Goryshin and Reznikoff, 1998; Epicentre, 2004). *Tn5*-based genomics tools can be used in a wide variety of bacterial species and given the system's independence from host-encoded factors, might be applicable to eukaryotic systems as well (Goryshin *et al*., 2000). Efforts to use *Tn5* as an insect gene vector have been successful.

Tn5 is a prokaryotic transposon 5.8 kb in length, and it is often referred to as a composite transposon because it consists of five independently functional units (review: Reznikoff, 2000) (**Figure 1**). It contains three antibiotic resistance genes that are flanked by 1.5 kb inverted repeat sequences. Each inverted repeat is actually a copy of an *IS50* insertion sequence that are themselves functional transposons. Each *IS50* element contains 19 bp terminal sequences known as outside end (OE) and inside end (IE), and while OE and IE are very similar, they are not identical. *IS50* also encodes for two proteins: transposase (Tnp) is 476 amino acids long and catalyzes transposition while the second protein is an inhibitor of transposition (Inh). The *IS50* elements present at each end of *Tn5* are not identical and only *IS50R* is fully functional. *IS50L* contains an ochre codon that prematurely terminates the Tnp and Inh proteins resulting in a loss of function of both proteins.

The transposition reaction and all of the components involved in the reaction have been studied in great detail (Reznikoff *et al*., 1999). Transposition proceeds by a cut-and-paste process involving binding of Tnp to the end sequences followed by dimerization of the bound Tnp to form a synaptic complex. Cleavage at the ends of the element results in an excised transposon with bound transposase that interacts with a target DNA molecule. Strand transfer results in the integration of the element into the target, and *in vitro*, this reaction requires only a donor element, a target DNA molecule, transposase, and Mg^{2+} (Goryshin and Reznikoff, 1998). Modifications of both the transposase and the terminal 19 bp sequences have lead to the creation of *Tn5* elements consisting of little more than two copies of end sequences that can be mobilized a 1000-fold more efficiently than an unmodified *Tn5* element. This hyperactive *Tn5* system has been developed into a powerful tool for genetic analysis of a variety of organisms. *Tn5* has been attractive as a broad host range genomics tool because its pattern of integration is random and its biochemical requirements very simple. *Tn5* has been shown to function in a variety of bacterial and nonbacterial systems.

Current insect transformation protocols consist of microinjecting a mixture of two plasmids into preblastoderm embryos (see Section 4.13.6.3). One plasmid contains a nonautonomous transposable element with the transgenes and genetic markers of interest while the second plasmid contains a copy of the transposase gene. Transient expression of the

transposase gene is required postinjection and is followed by element excision and integration. Previous experiments examining the frequency of element excision of elements, such as *Hermes*, *mariner*, *Minos*, and *piggyBac*, from plasmids injected into insect embryos along with helper plasmids indicated that only one plasmid per thousand injected underwent an excision event. Therefore, 99.9% of the donor plasmids introduced into insect embryos will contribute nothing to the transformation efforts. The introduction of preexcised elements configured as active intermediates, such as synaptic complexes, was considered a means to permit higher integration rates and overall efficiency of transformation.

Transgenic *Aedes aegypti* were created using a *Tn5* vector containing DsRed under the regulatory control of the 3xP3 promoter (Rowan *et al*., 2004). Preexcised vectors in the form of synaptic complexes were injected into preblastoderm embryos. Nine hundred adults were obtained from the injected embryos and families consisting of approximately 10 G_0 individuals were established. Two families of G_0 individuals produced transgenic progeny for an estimated transformation frequency of 0.22% (2/900). Analysis of the transgenic progeny showed that multiple integrations of *Tn5* occurred in each line. The patterns of integrations were complex with evidence of the *Tn5* vector integrating into *Tn5* vector sequences. The integration of the vector into copies of itself followed by the integration of the resulting concatamers was very unusual, and in no case was a simple cut-and-paste integration of the *Tn5* vector found with characteristic 9 bp direct duplications flanking the element. The complex pattern of *Tn5* integration was thought to be a direct consequence of injecting preassembled intermediates, that were inactive in the absence of Mg^{2+}. Therefore, as soon as the synaptic complexes were injected they became activated and the first target sequences the elements were likely to encounter were other *Tn5* synaptic complexes. At the time of injection, *A. aegypti* embryos only contain approximately four to eight nuclei making genomic target DNA relatively rare. Furthermore, the synaptic complexes injected were expected to have a very short half-life. Therefore, although active intermediates were being introduced, a number of factors contributed to the inefficiency observed with this system including a short half-life of the active intermediate and low numbers of genomic target sequences. Injecting binary plasmid systems (as is done with *Hermes*, *mariner*, *Minos*, and *piggyBac*), while relatively inefficient in producing active transposition intermediates, achieves persistence over an extended period of time. Consequently, more target genomes are exposed to active vectors over a longer period of time, resulting in higher transformation rates. The limitations of injecting synaptic complexes is unlikely to be specific to the *Tn5* system and similar approaches with other insect gene vectors are likely to encounter similar problems. It should be noted, however, that the results of Rowan *et al*. (2004) demonstrate that *Tn5* is functional in insects and, while injecting active intermediates is not recommended, using *Tn5* in a more conventional binary plasmid system consisting of a donor and helper plasmids is likely to be a viable option for creating transgenic insects.

4.13.6. Transformation Methodology

The technical methodology for insect transformation has largely remained the same or only slightly modified from the techniques originally used to transform *Drosophila*. The references cited for *P* transformation are relevant to this, as well as several recent articles that focus on methods for non-drosophilid transformation (Handler and O'Brochta, 1991; Morris, 1997; Ashburner *et al*., 1998; Handler, 2000; Handler and James, 2000). The most variable aspect of this method is the preparation of embryos for DNA microinjection, though arguably, the lack of new techniques for DNA introduction has been the primary limitation in the more widespread use of the technology. While all successful insect transformations have utilized microinjection, variations on this method have been necessary for different types of embryos, and most of the procedures must be tested empirically and modified for particular insect species. This may be extended to different strains and for a variety of local ambient conditions including temperature and humidity. The apparatus for microinjection is usually the same for all species, though a wide variety of variations and modifications are possible and sometimes required. The basic equipment includes an inverted microscope or a stereozoom microscope with a mechanical stage having a magnification up to 60 to 80×; a micromanipulator that is adjustable in three axes with an appropriate needle holder; and a means to transmit the DNA into the egg. For dechorionated eggs, transmitted light allows precise positioning of the needle within the egg posterior, while direct illumination is needed for nondechorionated eggs that typically include mosquitoes and moths.

The standard for gene transfer methodology in general, and embryo microinjection in particular, was originally developed for *Drosophila*. The standard method involves collecting preblastoderm embryos

within 30 min of oviposition, and dechorionating them either manually or chemically. The timing of egg collection and DNA injection is related to the need to inject into preblastoderm embryos during a phase of nuclear divisions previous to cellularization. This allows the injected DNA to be taken up into the nuclei, and specifically into the primordial germ cell nuclei that are the gamete progenitors. For *Drosophila*, cellularization of the pole cells begins at approximately 90 min after fertilization at 25 °C, with blastoderm formation occurring about 30 min afterwards. The timing of these events and location of the pole cells varies among insects, and thus some knowledge of early embryogenesis in the desired host insect is highly advantageous. In the absence of this information for a particular species, the most prudent time of injection would be the earliest time after oviposition that does not compromise viability.

4.13.6.1. Embryo Preparation

Manual dechorionation of *Drosophila* eggs is achieved by gently rolling the eggs on double-stick tape with a forcep until the chorions peel off. While gentle on the eggs and requiring little desiccation time, manual dechorionation is tedious and has not been applicable to any other insect. Chemical dechorionation is typically achieved by soaking eggs in a 50% bleach solution (2.5% hypochlorite) for 2–4 min and washing at least three times in 0.02% Triton X-100. Tephritid fruit fly eggs usually have thinner chorions that can be dechorionated in 30% bleach (1.25% hypochlorite) in 2–3 min, but this must be determined empirically since they are easily overbleached resulting in death, either directly or after injection. Some species, such as *M. domestica*, can be only partially dechorionated, but bleached eggs can be released from the chorion by agitation. We have found the simplest and most precise method for bleach dechorionation with rapid washes is by using a 42 mm Buchner funnel with a filter flask attached to a water vacuum. Eggs can be washed into the funnel on filter paper and swirled within the funnel with the solution gently sucked out by regulating the water flow or the seal between the funnel and flask. The last wash is done on black filter paper that allows the eggs to be easily detected, which facilitates their mounting for injection (see below).

Many insects eggs cannot be dechorionated without a high level of lethality, and must be injected without dechorionation. These include most moth and mosquito species. *Drosophila* and tephritid flies can, similarly, be injected without dechorionation, and while embryo viability after injection is often lower than for dechorionated eggs, the frequency of transformation in surviving embryos is often higher. It is more difficult to determine a precise site for injection in nondechorionated eggs, though this can be aided by adding food coloring to the DNA injection mix.

After dechorionation, fruit fly embryos are typically placed on a thin strip (~1 mm) of double stick tape placed on a microscope slide or 22×30 mm cover slip, though use of a cover slip is more adaptable for subsequent operations. A thin strip of tape is suggested due to anecdotal reports of toxic solvents from the tape affecting survival, though some particular tapes are considered to be nontoxic (3M Double Coated Tape 415; 3M, St. Paul, MN, USA) and some are useful for particular applications such as aqueous conditions needed for mosquito eggs. Adhesives resistant to moisture include Toupee tape (TopStick^{TM}, Vapon Inc.) and Tegaderm (3M). When eggs are injected under oil, the tape strip is placed within a thick rectangle created with a wax pencil that can retain the oil. It is important that the wax fence not be breached by oil when overlaying the eggs, since the loss of oil will result in embryo death.

Where possible, eggs are placed on the tape in an orientation having their posterior ends facing outwards towards the needle, but at a slight angle. All fruit fly eggs must be desiccated to some extent before injection. The interior of the egg is normally under positive pressure, and yolk and injected DNA will invariably flow out after injection without desiccation. This will result in lethality, sterility (from loss of pole plasm), or the lack of transformation if the plasmid DNA is lost. The time and type of desiccation, however, must be evaluated empirically, and sometimes varied during the course of an injection period. A major factor for dechorionated eggs is the length of time they are kept on moist filter paper before being placed on the tape. Typically we desiccate embryos on one strip of tape (15–20 embryos) for 8 to 10 min. Depending on the ease of injection the time can be varied by 1–2 min. In ambient conditions that are humid, it may be necessary to desiccate in a closed chamber with a drying agent (e.g., drierite), with or without a gentle vacuum. An important consideration is that a very short variation in the time for desiccation can be the difference between perfect desiccation and overdesiccation resulting in death, and that the optimal desiccation time will vary for different eggs on the tape. Thus, it is unlikely that all the eggs will respond well to the set conditions, which must be modified so that the majority of eggs can be injected with DNA at a high level of survival and fertility. After the determined time for desiccation, the eggs

must be placed immediately under Halocarbon 700 oil, or oil of similar density, to stop the desiccation process. Desiccation of most nondechorionated eggs is more challenging and one approach is to soak eggs in 1 M NaCl for several minutes. In contrast, for nondechorionated mosquito eggs, desiccation can occur within 1–2 min after removal from water, which is evidenced by slight dimpling of the egg surface, and this must be observed to avoid overdesiccation. Due to the rapidity of desiccation, mosquito eggs are typically arranged on moist filter paper and blotted together onto a taped cover slip from above, and after desiccation, the eggs are submerged in Halocarbon oil. Nondechorionated eggs from many species do not require oil, and it may be lethal for some insects such as moths, yet oil submersion was helpful for the survival of *Drosophila* and tephritid flies.

4.13.6.2. Needles

The type of needle and its preparation is possibly the most important component of successful embryo injections. Most dechorionated fruit fly eggs can be injected easily with borosilicate needles, which are drawn out to a fine tip and broken off to a 1 to 2 µm opening. Opening the tip is typically achieved by scraping the needle against the edge of the slide carrying the eggs to be injected. Opening the needle by beveling, however, creates consistently sharp tips that are much more important for nondechorionated eggs, and stronger alumina-silicate and quartz needles also provide an improvement to easily pierce chorions or tough vitelline membranes. Beveled needles are also critical when a large tip opening is required for large plasmids that are susceptible to shearing. Preparation of borosilicate needles, pulled from 25 µl capillary stock that has been silanized, can be achieved with several types of vertical or horizontal needle pullers, and we find the Sutter Model P-30 (Sutter Instruments, Novato, CA, USA) vertical micropipette puller to be highly effective. Alumina-silicate needles, and certainly quartz needles, require more sophisticated pullers that allow for fine programmable adjustment of high filament temperatures and pulling force, and the Sutter Models P-97 and P-2000 fulfill this need. Several needle bevelers are available, with the Sutter BV-10 used by many laboratories.

4.13.6.3. DNA Preparation and Injection

A mixture of highly purified vector and helper plasmid DNA is essential to embryo survival. This is achieved most optimally by purifying plasmid twice through cesium chloride gradients or a solid-phase anion exchange chromatography column. These have the advantage of high yields of DNA, but the disadvantage of specialized equipment and long preparation times. Successful transformation has been achieved with plasmids prepared with silica-gel membrane kits from Qiagen Corp. (Valencia, CA, USA), but their successful use has been inconsistent, with failures possibly related to the type of host bacteria and its growth conditions. The Qiagen Endotoxin-free plasmid preparation systems allow additional purity, and this system is routinely used for successful plasmid injection.

Purified plasmid concentration must be titered accurately and verified by gel electrophoresis previous to injection mix preparation. Appropriate amounts of vector and helper plasmid are ethanol precipitated, washed several times in 70% ethanol, and resuspended in injection buffer. Injection buffer has typically been the same as that originally used for *Drosophila* (5 mM KCl, 0.1 mM sodium phosphate, pH 6.8), though this may not be optimal for other insects and embryo survival should be assessed by control injections. Total DNA concentration for injection should not exceed 1 mg ml^{-1}, using two- to fourfold higher concentration of vector to helper (e.g., 600 ng μl^{-1} vector to 200 ng μl^{-1} helper). Higher DNA concentrations are inadvisable since they are subject to shearing during injection and may clog the needle, and the nucleic acids and/or contaminants can be toxic to the embryo. High transposase levels may also have a negative effect on transposition, as with the overproduction–inhibition phenomenon observed with *mariner* (Lohe and Hartl, 1996b).

Previous to injection the DNA mixture should be filtered through a 0.45 µm membrane, or centrifuged before loading into the injection needle. Typically, DNA is back-filled into the injection needle using a drawn-out silanized 100 µl microcapillary, and a microliter of DNA should be sufficient for injecting hundreds of eggs.

4.13.6.3.1. DNA injection The microinjection of DNA into embryos requires a system that forces a minute amount of DNA through the needle in a highly controllable fashion. Remarkably, many *Drosophila* laboratories simply use a mounted syringe and tubing filled with oil connected to a needle holder, with manual pressure applied. This system is successful due to accumulated expertise and the efficiency of transformation in the species, but would probably be less useful for injecting more sensitive embryos that transform less easily. Regulated air-pressure systems are available that are economical and allow highly controlled and rapid DNA injection. We use the PicoPump from WPI that

is most versatile in allowing positive and negative (with vacuum) pressure, and a hold capability that prevents back flow into the needle resulting in clogging (especially by yolk). A less expensive system can be constructed from Clippard components (Clippard Instrument Laboratory, Inc., Cincinnati, OH, USA) that uses a simple air-pressure regulator and electronic valve and switch (see Handler, 2000). Needle holders from WPI can be used with both systems (MPH-3 and MPH-1, respectively).

All embryo injections are performed on a microscope with a mechanical stage, with the injection needle mounted on a micromanipulator. Microscopes first used for *Drosophila* transformation were inverted or compound microscopes, but the availability of a useful mechanical stage and stage adaptor for the Olympus SZ stereozoom microscopes makes this the most versatile choice (the Olympus stage can be mounted on most stereomicroscopes). The micromanipulator can be free-standing next to the stage or mounted on the microscope base. It allows the precise positioning of the needle at the desired point of entry into the egg, while the actual injection occurs by using the mechanical stage to push the egg onto the needle. Piezo Translators that were developed for rapid and automatic intracellular injection may be more efficient for some embryos, and will obviate the need for a mechanical stage (Peloquin *et al.*, 1997). The WPI MPM20 translator used with the PV820 PicoPump allows a fully automated system for egg penetration, DNA injection, and needle withdrawal.

4.13.6.4. Postinjection Treatment

After injection the cover slip can be placed in a covered petri dish (but not sealed) with moist filter paper. The use of square dishes with black filter paper seems to be most suitable for up to six cover slips and simple observation of the embryos and hatched larvae. For injected embryos submerged in oil, oxygen concentration may be a limiting factor for development, if not viability. This can be ameliorated by reducing the crowding of eggs on the cover slip, or by incubation in a portable hatbox tissue culture chamber that is humidified and under slight positive pressure with oxygen. For eggs without oil, oxygen saturation without pressure is advisable.

Most helper constructs have the transposase gene under heat shock regulation. The *Drosophila hsp70* promoter is a constitutive promoter that is active in the absence of heat shock (but also responds to anoxia which may occur in embryos under oil), and transformation is possible with most vectors with or without heat shock treatment. If heat shock is desirable, it should be noted that the optimal temperature varies for different species. For example, *hsp70* responds optimally at 37 °C in *Drosophila*, but at 39 °C in medfly (Papadimitriou *et al.*, 1998). Injected embryos should be incubated for at least 4–6 h after injection before heat shock, or after overnight incubation. Optimal temperatures for insect development vary, but the lowest temperatures possible can be beneficial to survival, and the injection process can slow development by 50% or more. Thus, larval hatching may be delayed considerably and hatching should be monitored for several days after the expected time before discarding embryos.

Hatched larvae can be placed on normal culture media, though they may be weak and require careful handling and soft diet. Rearing of putative transgenic lines is typically achieved by backcrossing to the parental line in small group matings, or individual mating if a determination of transformation frequency is required. Inbreeding of G_0s can minimize rearing efforts, but this may be complicated by high rates of infertility which is typically close to 50% after fruit fly injections.

4.13.7. Research Needs for Improved Transgenesis

4.13.7.1. DNA Delivery

Dramatic progress has been made in transformation technology for non-drosophilid insects, and it appears that the vectors and markers in use should be widely applicable. Nevertheless, transformation of many other insect species will be highly challenging, primarily due to limitations in the delivery of DNA into preblastoderm embryos. As noted, to date all successful non-drosophilid transformations have resulted from embryonic microinjection of DNA, but for many species current injection techniques are likely to result in high levels of lethality or sterility. Experimentation with alternative methods has been reported, though arguably, none has been tested exhaustively for germline transformation, or vector systems were used that are now known to be ineffective. The most promising method is biolistics where eggs are bombarded with micropellets encapsulated by DNA, which was first developed as a ballistics method to transform plant cells (Klein *et al.*, 1987). Ballistics is based upon a "shotgun" technique for bombardment, and it is the only non-injection method successfully used to transform an insect. This was a *P* transformation of *Drosophila*, though only a single transformant line was created

and the technique never gained wide applicability (Baldarelli and Lengyel, 1990). This was most likely due to the high efficiency of *P* transformation of *Drosophila* by microinjection, eliminating the need for an alternative technique. Mosquito eggs are considerably more difficult to inject, and a significant effort was made to modify a biolistics approach to DNA delivery in *A. gambiae*, using a burst of pressurized helium for bombardments (Miahle and Miller, 1994). This technique was effective in introducing plasmid DNA into mosquito eggs, yielding high levels of transient expression of a reporter gene. Biolistics was subsequently used for transient expression in specific tissues, allowing the testing of fibroin gene promoters in the *B. mori* silk gland (Horard *et al*., 1994; Kravariti *et al*., 2001). Recent advances have included the use of a rigid macrocarrier in the Bio-Rad PDS/1000-Helium biolistics apparatus, which minimizes the blast effect in soft tissue (Thomas *et al*., 2001). This allows greater micropellet penetration into insect tissues with improved survival. Despite these advances in delivering DNA into eggs and tissue, biolistics has yet to yield a germline transformant.

The only other method reported for DNA delivery is electroporation, which, like biolistics, has resulted in high levels of transient expression of plasmid-encoded genes in *Drosophila* (Kamdar *et al*., 1992), as well as in *Helicoverpa zea* and *M. domestica* (Leopold *et al*., 1996). Though transformation has not been reported, as with biolistics, it is not apparent that this was seriously tested or if functional vectors systems were used (certainly for non-drosophilids). Electroporation techniques have also advanced in recent years, with DNA transferred into many different tissue types from a variety of organisms using new electroporation chamber designs and electric field paramenters.

These recent advances with both biolistics and electroporation are highly encouraging that new efforts will have greater chances for success, and they deserve a high priority for testing. Both methods also have the advantage, if successful, of delivering DNA simultaneously to multiple embryos, ranging from hundreds to thousands depending on the species. This would be highly beneficial to all transformation experiments, but especially so for species that transform at low frequencies. These methods could also be used in cellularized embryos after blastoderm formation in insects having embryos that cannot be handled easily or collected in the preblastoderm stage.

Other approaches to DNA delivery can include the incorporation of vector/helper DNA into bacterial or viral carriers, that may be delivered by maternal injection or feeding. Variations on microinjection that might be required for ovoviviparous insects include maternal injection into ovaries or abdominal hemocoel (Presnail and Hoy, 1994), and the use of liposomes might allow injection into cellularized embryos (Felgner *et al*., 1987). All of these techniques should be reevaluated with the use of vectors and markers now known to be highly efficient in non-drosophilid systems.

4.13.7.2. Gene Targeting

The ability to target genes to specific or desired integration sites in the genome would be highly advantageous to the basic and applied uses of transgenic strains. The expression of transgenes in most vector integrations is affected negatively by chromosomal position effects, so target sites known to be devoid of, or insulated from, suppression elements could be utilized for optimal or consistent transgene expression. Target sites positioned in innocuous genomic regions could also eliminate random integrations into genes necessary for viability and fertility, eliminating costs to fitness in host strains. It will also be highly important to gene expression studies to achieve reliable methods for gene replacement or targeted transposition, which is especially important for systems where preexisting null mutations, or gene "knockouts," do not exist.

Gene targeting can be achieved, generally, in two ways. First is homologous recombination where an endogenous genomic sequence is replaced by recombination with homologous sequences within or surrounding the transgene. This results in targeted transposition which can be used for gene replacement, or for targeting to an innocuous genomic region. This approach has been effective in transforming lower eukaryotes, and plant and vertebrate systems (see Bollag *et al*., 1989) and has been reported to occur in *Drosophila* (Cherbas and Cherbas, 1997) and mosquito (Eggleston and Zhao, 2000) cell line studies. Homologous recombination can also occur in insects *in vivo*, but this is not routine and thus far, must be facilitated. In *B. mori,* female moths were infected with a modified AcNPV baculovirus that had its polyhedrin gene replaced with fibroin light chain-GFP gene fusion (Yamao *et al*., 1999). Progeny of the infected moths exhibited stable integration of the gene fusion into the genomic fibroin gene, with resulting GFP expression. This particular method relies on host susceptibility to baculovirus infection, though conceivably other pantropic or species-specific viruses could be used for a wider range of insects.

Homologous recombination was also achieved in *Drosophila* where linearized extrachromosomal

DNA was found to be recombinogenic with homologous sequences in the genome (Rong and Golic, 2000). This was achieved by using the *FRT*–FLP recombination system to create DNA circles, which were linearized at a rare endonuclease recognition site within the *FRT* sequences, resulting in a *yellow* marker gene integrating into its homologous chromosomal site. This system has potential application in any insect species that can be stably transformed, but it requires the integration of three components which can be tedious to achieve and may present fitness costs to the host. Though its use could have major importance to genetic studies in non-drosophilid insects, thus far it is not routinely used in *Drosophila*.

4.13.7.2.1. Site-specific recombination Site-specific recombination systems such as the *FRT*–FLP system from the 2 μm circle of yeast (Senecoff *et al.*, 1985) mentioned above, and the bacteriophage Cre/*lox* system (Hoess *et al.*, 1985) can be used for various types of gene targeting and chromosomal manipulation. Both systems function in *Drosophila* in which recombination occurs between specific sequences in the presence of a recombinase enzyme (Golic and Lindquist, 1989; Siegal and Hartl, 1996). For *FRT*, the recombination site consists of two 13 bp inverted repeats separated by an 8 bp spacer that specifically recombines with identical *FRT* sites in the presence of FLP recombinase. Depending upon the orientation of the *FRT* sites, the intervening sequence between them can be inverted or deleted by recombination (Golic and Golic, 1996; Golic *et al.*, 1997). When placed within a vector, such *FRT* rearrangements can allow several types of vector manipulation after genomic integration. Genes or sequences within the vector necessary for the initial transformation or selection, but deleterious to use of the transformed strain, can be deleted or inactivated (Dale and Ow, 1991; see Handler, 2002b). This may include the marker system used for selection (e.g., chemical resistance system) or even a transposase gene used in a single plasmid autonomous vector system. Expression of genes of interest can be similarly manipulated by placing *FRT* sites outside the gene and within an internal noncoding region. Of particular importance to transgene stability would be the rearrangement of vector sequences required for mobility, that typically would include the terminal and subterminal ITRs. If *FRT* site placement between the terminal sequences does not hinder the primary transposon vector integration, then subsequent subterminal ITR deletion could eliminate any secondary vector mobilization. Such mobilization or cross-mobilziation would have serious consequences for strain stability and function, as well as ecological risks due to unintended transmission of the vector into other organisms. Eliminating this possibility would provide a major advancement to the applied use of transgenic insects.

The ultimate use of recombination systems for improved transformant stability and transgene expression would be their development into a second generation of vectors that use an integrated recombination site as a stable chromosomal target. Plasmids having the same recombination site and a marker gene would be used as vectors that integrate by recombination in the presence of recombinase. The expectation is that these systems would be highly stable in eukaryotes, and specific target site loci could be selected that are minimally affected by position effect variegation/suppression. The internal 8 bp spacer sequence within the *FRT* can be varied, but only identical *FRT* sequences will recombine with one another. Thus, multiple independent *FRT* target sites can be incorporated into the same genome. Importantly, manipulations by *FRT* recombination will depend upon a controllable source of FLP recombinase, which can be provided as a separate transgene integration, or exogenously by DNA, RNA, or protein injection.

4.13.8. Summary

After concerted efforts for more than 30 years to achieve gene transfer in non-drosophilid insects, only in the last decade have these efforts been fruitful. Since 1995 the germline of nearly 20 species in four orders of insects have been transformed, and this number may be only limited by the insects of current experimental and applied interest. Unlike plant and vertebrate animal systems that allow relatively efficient genomic integration of introduced DNA, insect systems have generally relied on vector-mediated integrations, and the only vectors found reliable for germline transformation are those based on transposable elements. Curiously, the two main vector systems developed for routine use in *D. melanogaster*, and originally discovered in that species, *P* and *hobo*, have not been applicable as vectors to any other species. Yet, four other transposons found in non-*melanogaster* or non-drosophilid species are widely functional in insects, and for some, other organisms. Their discovery has been of enormous importance to the wider use of transformation technology, since little progress would have been made if most vector systems were specific to a particular host. Equal in importance to the advancements in vector development, have been concurrent progress in genetic marker discovery and development. This began with the finding

that cloned eye color genes from *Drosophila* could complement existing mutations in other insects, and has continued with the more recent use of several fluorescent protein genes that are widely applicable as markers for transformation and reporters for gene expression.

The advancement of these techniques comes at a fortuitous time when genomics is providing a wealth of genetic infomation and resources that might be used to create transgenic strains of pest and beneficial insects to control their population size and behavior. As part of these efforts, genetic transformation is also critical to functional genomics studies that will provide information essential to understanding the biological function of genetic material, and relating specific genomic elements to those functions. Techniques such as enhancer traps and transposon tagging, which rely on remobilizable insertional mutagenesis, are only possible with transposon-based vector systems, and other techniques such as RNA interference (RNAi) are greatly facilitated by these systems. Together, routine methods for transposon-mediated germline transformation and genomics analysis should provide the tools for dramatic progress in our understanding and control of insect species.

Acknowledgments

We gratefully acknowledge support from the USDA-NRI Competetitive Grants Program and National Institutes of Health. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture.

References

Allen, M.L., O'Brochta, D.A., Atkinson, P.W., Levesque, C.S., **2001**. Stable, germ-line transformation of *Culex quinquefasciatus* (Diptera: Culicidae). *J. Med. Entomol. 38*, 701–710.

Allen, M.L., Handler, A.M., Berkebile, D.R., Skoda, S.R., **2004**. *piggyBac* transformation of the New World screwworm, *Cochliomyia hominivorax*, produces multiple distinct mutant strains. *Med. Vet. Entomol. 18*, 1–9.

Anxolabéhère, D., Kidwell, M., Periquet, G., **1988**. Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile *P* elements. *Mol. Biol. Evol. 5*, 252–269.

Arca, B., Savakis, C., **2000**. Distribution of the transposable element *Minos* in the genus *Drosophila*. *Genetica 108*, 263–267.

Arca, B., Zabolou, S., Loukeris, T.G., Savakis, C., **1997**. Mobilization of a *Minos* transposon in *Drosophila melanogaster* chromosomes and chromatid repair by heteroduplex formation. *Genetics 145*, 267–279.

Ashburner, M., **1989a**. *Drosophila*: A Laboratory Handbook. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Ashburner, M., **1989b**. *Drosophila*: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Ashburner, M., Hoy, M.A., Peloquin, J., **1998**. Transformation of arthropods: research needs and long term prospects. *Insect Mol. Biol. 7*, 201–213.

Atkinson, P.W., Pinkerton, A.C., O'Brochta, D.A., **2001**. Genetic transformation systems in insects. *Annu. Rev. Entomol. 46*, 317–346.

Atkinson, P.W., Warren, W.D., O'Brochta, D.A., **1993**. The *hobo* transposable element of *Drosophila* can be cross-mobilized in houseflies and excises like the *Ac* element of maize. *Proc. Natl Acad. Sci. USA 90*, 9693–9697.

Auge-Gouillou, C., Hamelin, M.-H., Demattei, M.-V., Periquet, G., Bigot, Y., **2001a**. The ITR binding domain of the mariner Mos-1 transposase. *Mol. Gen. Genomics 265*, 58–65.

Auge-Gouillou, C., Hamelin, M.-H., Demattei, M.-V., Periquet, G., Bigot, Y., **2001b**. The wild-type conformation of the Mos-1 inverted terminal repeats is suboptimal for transposition in bacteria. *Mol. Gen. Genomics 265*, 51–57.

Avery, T., Macleod, C.M., McCarty, M., **1944**. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. I. Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J. Exp. Med. 79*, 137–158.

Baldarelli, R.M., Lengyel, J.A., **1990**. Transient expression of DNA after ballistic introduction into *Drosophila* embryos. *Nucl. Acids Res. 18*, 5903–5904.

Bellen, H.J., O'Kane, C.J., Wilson, C., Grossniklaus, U., Pearson, R.K., *et al.*, **1989**. *P*-element-mediated enhancer detection: a versatile method to study development in *Drosophila*. *Development 3*, 1288–1300.

Bender, W., Hudson, A., **2000**. *P* element homing in the *Drosophila* bithorax complex. *Development 127*, 3981–3992.

Benedict, M.Q., Salazar, C.E., Collins, F.H., **1995**. A new dominant selectable marker for genetic transformation: *Hsp70-opd*. *Insect Biochem. Mol. Biol. 25*, 1061–1065.

Berghammer, A.J., Klingler, M., Wimmer, E.A., **1999**. A universal marker for transgenic insects. *Nature 402*, 370–371.

Besansky, N.J., Mukabayire, O., Benedict, M.Q., Rafferty, C.S., Hamm, D.M., *et al.*, **1997**. The *Anopheles gambiae* tryptophan oxygenase gene expressed from a baculovirus promoter complements *Drosophila melanogaster vermilion*. *Insect Biochem. Mol. Biol. 27*, 803–805.

Bigot, Y., Auge-Gouillou, C., Periquet, G., **1996**. Computer analyses reveal a *hobo*-like element in the nematode *Caenorhabditis elegans*, which presents a conserved transposase domain common with the *Tc1-mariner* transposon family. *Gene 174*, 265–271.

Blackman, R.K., Macy, M., Koehler, D., Grimaila, R., Gelbart, W.M., **1989**. Identification of a fully functional *hobo* transposable element and its use for germ-line transformation of *Drosophila*. *EMBO J. 8*, 211–217.

Bollag, R.J., Waldman, A.S., Liskay, R.M., **1989**. Homologous recombination in mammalian cells. *Annu. Rev. Genet. 23*, 199–225.

Brand, A.H., Manoukian, A.S., Perrimon, N., **1994**. Ectopic expression in *Drosophila*. *Methods Cell Biol. 44*, 635–654.

Bryan, G.J., Jacobson, J.W., Hartl, D.L., **1987**. Heritable somatic excision of a *Drosophila* transposon. *Science 235*, 1636–1638.

Bushman, F.D., **1994**. Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. *Proc. Natl Acad. Sci. USA 91*, 9233–9237.

Calvi, B.R., Hong, T.J., Findley, S.D., Gelbart, W.M., **1991**. Evidence for a common evolutionary origin of inverted repeat transposons in *Drosophila* and plants: *hobo*, *Activator*, and *Tam*3. *Cell 66*, 465–471.

Campbell, R.E., Tour, O., Palmer, A.E., Steinbach, P.A., Baird, G.S., *et al.*, **2002**. A monomeric red fluorescent protein. *Proc. Natl Acad. Sci. USA 99*, 7877–7882.

Capy, P., Vitalis, R., Langin, T., Higuet, D., Bazin, C., **1996**. Relationships between transposable elements based upon the integrase–transposase domains: is there a common ancestor? *J. Mol. Evol. 42*, 359–368.

Cary, L.C., Goebel, M., Corsaro, H.H., Wang, H.H., Rosen, E., *et al.*, **1989**. Transposon mutagenesis of baculoviruses: analysis of *Trichoplusia ni* transposon IFP2 insertions within the FP-Locus of nuclear polyhedrosis viruses. *Virology 161*, 8–17.

Caspari, E., Nawa, S., **1965**. A method to demonstrate transformation in *Ephestia*. *Z. Naturforsch. 206*, 281–284.

Catteruccia, F., Nolan, T., Blass, C., Muller, H.-M., Crisanti, A., *et al.*, **2000a**. Toward *Anopheles* transformation: *Minos* element activity in anopehline cells and embryos. *Proc. Natl Acad. Sci. USA 97*, 2157–2162.

Catteruccia, F., Nolan, T., Loukeris, T.G., Blass, C., Savakis, C., *et al.*, **2000b**. Stable germline transformation of the malaria mosquito *Anopheles stephensi*. *Nature 405*, 959–962.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W., Prasher, D.C., **1994**. Green fluorescent protein as a marker for gene expression. *Science 263*, 802–805.

Cherbas, L., Cherbas, P., **1997**. "Parahomologous" gene targeting in *Drosophila* cells: an efficient, homology-dependent pathway of illegitimate recombination near a target site. *Genetics 145*, 349–358.

Christophides, G.K., Savakis, C., Mintzas, A.C., Komitopoulou, K., **2001**. Expression and function of the *Drosophila melanogaster* ADH in male *Ceratitis capitata* adults: a potential strategy for medfly genetic sexing based on gene-transfer technology. *Insect Mol. Biol. 10*, 249–254.

Chroma, **2004**. Chroma Technology Corp., Brattleboro, VT, USA. http://www.chroma.com.

Coates, C.J., Jasinskiene, N., Miyashiro, L., James, A.A., **1998**. *Mariner* transposition and transformation of the yellow fever mosquito, *Aedes aegypti*. *Proc. Natl Acad. Sci. USA 95*, 3742–3751.

Coates, C.J., Johnson, K.N., Perkins, H.D., Howells, A.J., O'Brochta, D.A., *et al.*, **1996**. The *hermit* transposable element of the Australian sheep blowfly, *Lucilia cuprina*, belongs to the *hAT* family of transposable elements. *Genetica 97*, 23–31.

Coates, C.J., Turney, C.L., Frommer, M., O'Brochta, D.A., Atkinson, P.W., **1997**. Interplasmid transposition of the *mariner* transposable element in non-drosophilid insects. *Mol. Gen. Genet. 253*, 728–733.

Cornel, A.J., Benedict, M.Q., Rafferty, C.S., Howells, A.J., Collins, F.H., **1997**. Transient expression of the *Drosophila melanogaster cinnabar* gene rescues eye color in the white eye (WE) strain of *Aedes aegypti*. *Insect. Biochem. Mol. Biol. 27*, 993–997.

Dale, E.C., Ow, D.W., **1991**. Gene transfer with subsequent removal of the selection gene from the host genome. *Proc. Natl Acad. Sci. USA 88*, 10558–10562.

Daniels, S.B., Strausbaugh, L.D., **1986**. The distribution of *P*-element sequences in *Drosophila*: the *willistoni* and *saltans* species groups. *J. Mol. Evol. 23*, 138–148.

Davis, I., Girdham, C.H., O'Farrell, P.H., **1995**. A nuclear GFP that marks nuclei in living *Drosophila* embryos; maternal supply overcomes a delay in the appearance of zygotic fluorescence. *Devel. Biol. 170*, 726–729.

Drabek, D., Zagoraiou, L., deWit, T., Langeveld, A., Roumpaki, C., *et al.*, **2003**. Transposition of the *Drosophila hydei Minos* transposon in the mouse germ line. *Genomics 81*, 108–111.

Eggleston, P., Zhao, Y., **2000**. Targeted transformation of the insect genome. In: Handler, A.M., James, A.A. (Eds.), Insect Transgenesis: Methods and Applications. CRC Press, Boca Raton, FL, pp. 29–52.

Engels, W.R., **1989**. *P* elements in *Drosophila melanogaster*. In: Berg, D.E., Howe, M.M. (Eds.), Mobile DNA. American Society of Microbiology, Washington, DC, pp. 437–484.

Engstrom, Y., Schneuwly, S., Gehring, W., **1992**. Spatial and temporal expression of an *Antennapedia LacZ* gene construct integrated into the endogenous *Antennapedia* gene of *Drosophila melanogaster*. *Roux's Arch. Devel. Biol. 201*, 65–80.

Epicentre, **2004**. Epicentre, Madison, WI, USA. http://www.epicentre.com.

Fadool, J.M., Hartl, D.L., Dowling, J.E., **1998**. Transposition of the *mariner* element from *Drosophila mauritiana* in zebrafish. *Proc. Natl Acad. Sci. USA 95*, 5182–5186.

Federoff, N., **1989**. Maize transposable elements. In: Berg, D.E., Howe, M.M. (Eds.), Mobile DNA. American Society for Microbiology, Washington, DC, pp. 375–411.

Felgner, P.L., Gadek, T.R., Holm, M., Roman, R., Chan, H.W., *et al.*, **1987**. Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure. Proc. *Natl Acad. Sci. USA 84*, 7413–7417.

ffrench-Constant, R.H., Benedict, M.Q., **2000**. Resistance genes as candidates for insect transgenesis. In: Handler, A.M., James, A.A. (Eds.), Insect Transgenesis: Methods and Applications. CRC Press, Boca Raton, FL, pp. 109–121.

ffrench-Constant, R.H., Mortlock, D.P., Shaffer, C.D., MacIntyre, R.J., Roush, R.T., **1991**. Molecular cloning and transformation of cyclodiene resistance in *Drosophila*: an invertebrate $GABA_A$ receptor locus. *Proc. Natl Acad. Sci. USA 88*, 7209–7213.

Finnegan, D.J., **1989**. Eucaryotic transposable elements and genome evolution. *Trends Genet. 5*, 103–107.

Fox, A.S., Duggleby, W.F., Gelbart, W.M., Yoon, S.B., **1970**. DNA-induced transformation in *Drosophila*: evidence for transmission without integration. *Proc. Natl Acad. Sci. USA 67*, 1834–1838.

Fox, A.S., Yoon, S.B., **1966**. Specific genetic effects of DNA in *Drosophila melanogaster*. *Genetics 53*, 897–911.

Fox, A.S., Yoon, S.B., **1970**. DNA-induced transformation in *Drosophila*: locus specificity and the establishment of transformed stocks. *Proc. Natl Acad. Sci. USA 67*, 1608–1615.

Franz, G., Loukeris, T.G., Dialektaki, G., Thompson, C.R.L., Savakis, C., **1994**. Mobile *Minos* elements from *Drosophila hydei* encode a two-exon transposase with similarity to the paired DNA-binding domain. *Proc. Natl Acad. Sci. USA 91*, 4746–4750.

Franz, G., Savakis, C., **1991**. *Minos*, a new transposable element from *Drosophila hydei*, is a member of the Tc-1-like family of transposons. *Nucl. Acids Res. 19*, 6646.

Fraser, M.J., **2000**. The TTAA-specific family of transposable elements: identification, functional characterization, and utility for transformation of insects. In: Handler, A.M., James, A.A. (Eds.), Insect Transgenesis: Methods and Applications. CRC Press, Boca Raton, FL, pp. 249–268.

Fraser, M.J., Brusca, J.S., Smith, G.E., Summers, M.D., **1985**. Transposon-mediated mutagenesis of a baculovirus. *Virology 145*, 356–361.

Fraser, M.J., Cary, L., Boonvisudhi, K., Wang, H.G., **1995**. Assay for movement of lepidopteran transposon IFP2 in insect cells using a baculovirus genome as a target DNA. *Virology 211*, 397–407.

Fraser, M.J., Smith, G.E., Summers, M.D., **1983**. Acquisition of host-cell DNA sequences by baculoviruses: relationship between host DNA insertions and FP mutants of *Autographa californica* and *Galleria mellonella* nuclear polyhedrosis viruses. *J. Virol. 47*, 287–300.

Furlong, E.E., Profitt, D., Scott, M.P., **2001**. Automated sorting of live transgenic embryos. *Nature Biotechnol. 19*, 153–156.

Garza, D., Medhora, M., Koga, A., Hartl, D.L., **1991**. Introduction of the transposable element *mariner* into the germline of *Drosophila melanogaster*. *Genetics 128*, 303–310.

Germeraad, S., **1976**. Genetic transformation in *Drosophila* by microinjection of DNA. *Nature 262*, 229–231.

Goldberg, D.A., Posakony, J.W., Maniatis, T., **1983**. Correct developmental expression of a cloned alcohol dehydrogenase gene transduced into the *Drosophila* germ line. *Cell 34*, 59–73.

Golic, K.G., Golic, M.M., **1996**. Engineering the *Drosophila* genome: chromosome rearrangements by design. *Genetics 144*, 1693–1711.

Golic, K.G., Lindquist, S.L., **1989**. The FLP recombinase of yeast catalyzes site-specific recombination in the *Drosophila* genome. *Cell 59*, 499–509.

Golic, M.M., Rong, Y.S., Petersen, R.B., Lindquist, S.L., Golic, K.G., **1997**. FLP-mediated DNA mobilization to specific target sites in *Drosophila* chromosomes. *Nucl. Acids Res. 25*, 3665–3671.

Gomez, S.P., Handler, A.M., **1997**. A *Drosophila melanogaster hobo-white*$^+$ vector mediates low frequency gene transfer in *D. virilis* with full interspecific *white*$^+$ complementation. *Insect Mol. Biol. 6*, 1–8.

Gorbunova, V., Levy, A.A., **1997**. Circularized *Ac/Ds* transposons: formation, structure and fate. *Genetics 145*, 1161–1169.

Goryshin, I.Y., Jendrisak, J., Hoffman, L.M., Meis, R., Reznikoff, W.S., **2000**. Insertional transposon mutagenesis by electroporation of released *Tn5* transposition complexes. *Nature Biotechnol. 18*, 97–100.

Goryshin, I.Y., Reznikoff, W.S., **1998**. *Tn5 in vitro* transposition. *J. Biol. Chem. 273*, 7367–7374.

Green, M.M., **1996**. The "Genesis of the White-Eyed Mutant" in *Drosophila melanogaster*: a reappraisal. *Genetics 142*, 329–331.

Griffith, F., **1928**. Significance of pneumococcal types. *J. Hyg., Camb. 27*, 113–159.

Grossman, G.L., Rafferty, C.S., Clayton, J.R., Stevens, T.K., Mukabayire, O., *et al.*, **2001**. Germline transformation of the malaria vector, *Anopheles gambiae*, with the *piggyBac* transposable element. *Insect Mol. Biol. 10*, 597–604.

Gueiros-Filho, F.J., Beverley, S.M., **1997**. Trans-kingdom transposition of the *Drosophila* element *mariner* within the protozoan *Leishmania*. *Science 276*, 1716–1719.

Guimond, N., Bideshi, D.K., Pinkerton, A.C., Atkinson, P.W., O'Brochta, D.A., **2003**. Patterns of *Hermes* transposition in *Drosophila melanogaster*. *Mol. Gen. Genet. 268*, 779–790.

Hama, C., Ali, Z., Kornberg, T.B., **1990**. Region-specific recombination and expression are directed by portions of the *Drosophila engrailed* promoter. *Genes Devel. 4*, 1079–1093.

Handler, A.M., **2000**. An introduction to the history and methodology of insect gene transfer. In: Handler, A.M., James, A.A. (Eds.), Insect Transgenesis: Methods and Applications. CRC Press, Boca Raton, FL, pp. 3–26.

Handler, A.M., **2001**. A current perspective on insect gene transfer. *Insect Biochem. Mol. Biol. 31*, 111–128.

Handler, A.M., **2002a**. Use of the *piggyBac* transposon for germ-line transformation of insects. *Insect Biochem. Mol. Biol. 32*, 1211–1220.

Handler, A.M., **2002b**. Prospects for using genetic transformation for improved SIT and new biocontrol methods. *Genetica 116*, 137–149.

Handler, A.M., **2003**. Isolation and analysis of a new *hopper hAT* transposon from the *Bactrocera dorsalis* white eye strain. *Genetica 118*, 17–24.

Handler, A.M., Gomez, S.P., **1996**. The *hobo* transposable element excises and has related elements in tephritid species. *Genetics 143*, 1339–1347.

Handler, A.M., Gomez, S.P., **1997**. A new *hobo*, *Activator*, *Tam*3 transposable element, *hopper*, from *Bactrocera dorsalis* is distantly related to *hobo* and *Ac*. *Gene 185*, 133–135.

Handler, A.M., Gomez, S.P., O'Brochta, D.A., **1993a**. A functional analysis of the *P*-element gene-transfer vector in insects. *Arch. Insect Biochem. Physiol. 22*, 373–384.

Handler, A.M., Gomez, S.P., O'Brochta, D.A., **1993b**. Negative regulation of *P* element excision by the somatic product and terminal sequences of *P* in *Drosophila melanogaster*. *Mol. Gen. Genet. 237*, 145–151.

Handler, A.M., Harrell, R.A., **1999**. Germline transformation of *Drosophila melanogaster* with the *piggyBac* transposon vector. *Insect Mol. Biol. 8*, 449–458.

Handler, A.M., Harrell, R.A., **2000**. Transformation of the Caribbean fruit fly with a *piggyBac* transposon vector marked with polyubiquitin-regulated GFP. *Insect Biochem. Mol. Biol. 31*, 199–205.

Handler, A.M., Harrell, R.A., **2001**. Polyubiquitin-regulated DsRed marker for transgenic insects. *Biotechniques 31*, 820–828.

Handler, A.M., James, A.A. (Eds.), **2000**. Insect Transgenesis: Methods and Applications. CRC Press, Boca Raton, FL.

Handler, A.M., McCombs, S.D., **2000**. The *piggyBac* transposon mediates germ-line transformation in the Oriental fruit fly and closely related elements exist in its genome. *Insect Mol. Biol. 9*, 605–612.

Handler, A.M., McCombs, S.D., Fraser, M.J., Saul, S.H., **1998**. The lepidopteran transposon vector, *piggyBac*, mediates germline transformation in the Mediterranean fruitfly. *Proc. Natl Acad. Sci. USA 95*, 7520–7525.

Handler, A.M., O'Brochta, D.A., **1991**. Prospects for gene transformation in insects. *Annu. Rev. Entomol. 36*, 159–183.

Hartl, D.L., Lohe, A.R., Lozovskaya, E.R., **1997**. Modern thoughts on an ancyent marinere: function, evolution, regulation. *Annu. Rev. Genet. 31*, 337–358.

Haymer, D.S., Marsh, J.L., **1986**. Germ line and somatic instability of a *white* mutation in *Drosophila mauritiana* due to a transposable element. *Devel. Genet. 6*, 281–291.

Hazelrigg, T., Levis, R., Rubin, G.M., **1984**. Transformation of *white* locus DNA in *Drosophila*: dosage compensation, zeste interaction, and position effects. *Cell 64*, 1083–1092.

Hazelrigg, T., Liu, N., Hong, Y., Wang, S., **1998**. GFP expression in *Drosophila* tissues: time requirements for formation of a fluorescent product. *Devel. Biol. 199*, 245–249.

Hediger, M., Niessen, M., Wimmer, E.A., Dübendorfer, A., Bopp, D., **2000**. Genetic transformation of the housefly *Musca domestica* with the lepidopteran derived transposon *piggyBac*. *Insect Mol. Biol. 10*, 113–119.

Heinrich, J.C., Li, X., Henry, R.A., Haack, N., Stringfellow, L., *et al.*, **2002**. Germ-line transformation of the Australian sheep blowfly *Lucilia cuprina*. *Insect Mol. Biol. 11*, 1–10.

Higgs, S., Sinkins, D.L., **2000**. Green fluorescent protein (GFP) as a marker for transgenic insects. In: Handler, A.M., James, A.A. (Eds.), Insect Transgenesis: Methods and Applications. CRC Press, Boca Raton, FL, pp. 93–108.

Higgs, S., Traul, D., Davis, B.S., Kamrud, K.I., Wilcox, C.L., *et al.*, **1996**. Green fluorescent protein expressed in living mosquitoes: without the requirement of transformation. *Biotechniques 21*, 660–664.

Hoess, R., Wierzbicki, A., Abremski, K., **1985**. Formation of small circular DNA molecules via an *in vitro* site-specific recombination system. *Gene 40*, 325–329.

Horard, B., Mange, A., Pelissier, B., Couble, P., **1994**. *Bombyx* gene promoter analysis in transplanted silk gland transformed by particle delivery system. *Insect Mol. Biol. 3*, 261–265.

Horn, C., Jaunich, B., Wimmer, E.A., **2000**. Highly sensitive, fluorescent transformation marker for *Drosophila* transgenesis. *Devel. Genes Evol. 210*, 623–629.

Horn, C., Schmid, B., Pogoda, F.S., Wimmer, E.A., **2002**. Fluorescent transformation markers for insect transgenesis. *Insect Biochem. Mol. Biol. 32*, 1221–1235.

Horn, C., Wimmer, E.A., **2000**. A versatile vector set for animal transgenesis. *Devel. Genes Evol. 210*, 630–637.

Jacobson, J.W., Medhora, M.M., Hartl, D.L., **1986**. Molecular structure of somatically unstable transposable element in *Drosophila*. *Proc. Natl Acad. Sci. USA 83*, 8684–8688.

Jasinskiene, N., Coates, C.J., Benedict, M.Q., Cornel, A.J., Rafferty, C.S., *et al.*, **1998**. Stable transposon mediated transformation of the yellow fever mosquito, *Aedes aegypti*, using the *Hermes* element from the housefly. *Proc. Natl Acad. Sci. USA 95*, 3743–3747.

Jasinskiene, N., Coates, C.J., James, A.A., **2000**. Structure of *Hermes* integrations in the germline of the yellow fever mosquito, *Aedes aegypti*. *Insect Mol. Biol. 9*, 11–18.

Kamdar, P., Von Allmen, G., Finnerty, V., **1992**. Transient expression of DNA in *Drosophila* via electroporation. *Nuc. Acids Res. 11*, 3526.

Kaminski, J.M., Huber, M.R., Summers, J.B., Ward, M.B., **2002**. Design of a nonviral vector for site-selective, efficient integration into the human genome. *FASEB J. 16*, 1242–1247.

Kapetanaki, M.G., Loukeris, T.G., Livadaras, I., Savakis, C., **2002**. High frequencies of *Minos* transposon mobilization are obtained in insects by using *in vitro*

synthesized mRNA as a source of transposase. *Nucl. Acids Res. 30*, 3333–3340.

Karess, R.E., **1985**. *P* element mediated germ line transformation of *Drosophila*. In: Glover, D.M. (Ed.), DNA Cloning, vol. 2. A Practical Approach. IRL Press, Oxford, pp. 121–141.

Karess, R.E., Rubin, G.R., **1984**. Analysis of *P* transposable element functions in *Drosophila*. *Cell 38*, 135–146.

Kassis, J.A., Noll, E., VanSickle, E.P., Odenwald, W.F., Perrimon, N., **1992**. Altering the insertional specificity of a *Drosophila* transposable element. *Proc. Natl Acad. Sci. USA 89*, 1919–1923.

Kaufman, P.K., Doll, R.F., Rio, D.C., **1989**. *Drosophila P* element transposase recognizes internal *P* element DNA sequences. *Cell 59*, 359–371.

Kidwell, M.G., Kidwell, J.F., Sved, J.A., **1977**. Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility, and male recombination. *Genetics 86*, 813–833.

Klein, T.M., Wolf, E.D., Wu, R., Sanford, J.C., **1987**. High-velocity microprojectiles for delivering nucleic acids into living cells. *Nature 327*, 70–73.

Klemenz, R., Weber, U., Gehring, W.J., **1987**. The white gene as a marker in a new *P*-element vector for gene transfer in *Drosophila*. *Nucl. Acids Res. 15*, 3947–3959.

Klinakis, A.G., Loukeris, T.G., Pavlopoulos, A., Savakis, C., **2000**. Mobility assays confirm the broad host-range activity of the *Minos* transposable element and validate new transformation tools. *Insect Mol. Biol. 9*, 269–276.

Kokoza, V., Ahmed, A., Wimmer, E.A., Raikhel, A.S., **2001**. Efficient transformation of the yellow fever mosquito *Aedes aegypti* using the *piggyBac* transposable element vector pBac[3xP3–EGFP afm]. *Insect Biochem. Mol. Biol. 31*, 1137–1143.

Kravariti, L., Thomas, J., Sourmeli, S., Rodakis, G.C., Mauchamp, B., *et al.*, **2001**. The biolistic method as a tool for testing the differential activity of putative silkmoth chorion gene promoters. *Insect Biochem. Mol. Biol. 31*, 473–479.

Lampe, D.J., Akerley, B.J., Rubin, E.J., Mekalanos, J.J., Robertson, H.M., **1999**. Hyperactive transposase mutants of the *Himar1 mariner* transposon. *Proc. Natl Acad. Sci. USA 96*, 11428–11433.

Lampe, D.J., Churchill, M.E., Robertson, H.M., **1996**. A purified *mariner* transposase is sufficient to mediate transposition *in vitro*. *EMBO J. 15*, 5470–5479.

Lampe, D.J., Grant, T.E., Robertson, H.M., **1998**. Factors affecting transposition of the *Himar1 mariner* transposon *in vitro*. *Genetics 149*, 179–187.

Lampe, D.J., Walden, K.K.O., Sherwood, J.M., Robertson, H.M., **2000**. Genetic engineering of insects with *mariner* transposons. In: Handler, A.M., James, A.A. (Eds.), Insect Transgenesis: Methods and Applications. CRC Press, Boca Raton, FL, pp. 237–248.

Lee, H., Simon, J.A., Lis, J.T., **1988**. Structure and expression of ubiquitin genes of *Drosophila melanogaster*. *Mol. Cell. Biol. 8*, 4727–4735.

Leopold, R.A., Hughes, K.J., DeVault, J.D., **1996**. Using electroporation and a slot cuvette to deliver plasmid DNA to insect embryos. *Genet. Anal. 12*, 197–200.

Lerat, E., Brunet, F., Bazin, C., Capy, P., **1999**. Is the evolution of transposable elements modular? *Genetica 107*, 15–25.

Li, X., Heinrich, J.C., Scott, M.J., **2001a**. *piggyBac*-mediated transposition in *Drosophila melanogaster*: an evaluation of the use of constitutive promoters to control transposase gene expression. *Insect Mol. Biol. 10*, 447–456.

Li, X., Lobo, N., Bauser, C.A., Fraser, M.J., Jr., **2001b**. The minimum internal and external sequence requirements for transposition of the eukaryotic transformation vector *piggyBac*. *Mol. Genet. Genomics 266*, 190–198.

Lidholm, D.-A., Gudmundsson, G.H., Boman, H.G., **1991**. A highly repetitive, *mariner*-like element in the genome of *Hyalophora cecropia*. *J. Biol. Chem. 266*, 11518–11521.

Lidholm, D.-A., Lohe, A.R., Hartl, D.L., **1993**. The transposable element *mariner* mediates germline transformation in *Drosophila melanogaster. Genetics 134*, 859–868.

Lis, J.T., Simon, J.A., Sutton, C.A., **1983**. New heat shock puffs and β-galactosidase activity resulting from transformation of *Drosophila* with an *hsp70–LacZ* hybrid gene. *Cell 35*, 403–410.

Lobo, N.F., Hua-Van, A., Li, X., Nolen, B.M., Fraser, M.J., Jr., **2002**. Germ line transformation of the yellow fever mosquito, *Aedes aegypti*, mediated by transpositional insertion of a *piggyBac* vector. *Insect Mol. Biol. 11*, 133–913.

Lobo, N., Li, X., Hua-Van, A., Fraser, M.J., Jr., **2001**. Mobility of the *piggyBac* transposon in embryos of the vectors of Dengue fever (*Aedes albopictus*) and La Crosse encephalitis (*Ae. triseriatus*). *Mol. Genet. Genomics 265*, 66–71.

Lohe, A.R., Hartl, D.L., **1996a**. Germline transformation of *Drosophila virilis* with the transposable element *mariner. Genetics 143*, 365–374.

Lohe, A.R., Hartl, D.L., **1996b**. Autoregulation of *mariner* transposase activity by overproduction and dominant-negative complementation. *Mol. Biol. Evol. 13*, 549–555.

Lohe, A.R., Hartl, D.L., **2002**. Efficient mobilization of *mariner in vivo* requires multiple internal sequences. *Genetics 160*, 519–526.

Lohe, A.R., Lidholm, D.-A., Hartl, D.L., **1995**. Genotypic effects, maternal effects and grandmaternal effects of immobilized derivatives of the transposable element *mariner. Genetics 140*, 183–192.

Lorenzen, M.D., Berghammer, A.J., Brown, S.J., Denell, R.E., Klingler, M., *et al.*, **2003**. *piggyBac*-mediated germline transformation in the beetle *Tribolium castaneum*. *Insect Mol. Biol. 12*, 433–440.

Lorenzen, M.D., Brown, S.J., Denell, R.E., Beeman, R.W., **2002**. Cloning and characterization of the *Tribolium castaneum* eye-color genes encoding tryptophan

oxygenase and kynurenine 3-monooxygenase. *Genetics 160*, 225–234.

Loukeris, T.G., Arca, B., Livadras, I., Dialektaki, G., Savakis, C., **1995a**. Introduction of the transposable element *Minos* into the germ line of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA 92*, 9485–9489.

Loukeris, T.G., Livadaras, I., Arca, B., Zabalou, S., Savakis, C., **1995b**. Gene transfer into the Medfly, *Ceratitis capitata*, with a *Drosophila hydei* transposable element. *Science 270*, 2002–2005.

Lozovskaya, E.R., Nurminsky, D.I., Hartl, D.L., Sullivan, D.T., **1996**. Germline transformation of *Drosophila virilis* mediated by the transposable element *hobo*. *Genetics 142*, 173–177.

Lozovsky, E.R., Nurminsky, D., Wimmer, E.A., Hartl, D.L., **2002**. Unexpected stability of *mariner* transgenes in *Drosophila*. *Genetics 160*, 527–535.

Mamoun, C.B., Guzman, I.Y., Beverly, S.M., Goldberg, D.E., **2000**. Transposition of the *Drosophila* element *mariner* within the human malaria parasite *Plasmodium falciparum*. *Mol. Biochem. Parasitol. 110*, 405–407.

Matz, M.V., Fradkov, A.F., Labas, Y.A., Savitsky, A.P., Zaraisky, A.G., *et al.*, **1999**. Fluorescent proteins from nonbioluminescent *Anthozoa* species. *Nature Biotechnol. 17*, 969–973.

Matz, M.V., Lukyanov, K.A., Lukyanov, S.A., **2002**. Family of the green fluorescent protein: journey to the end of the rainbow. *BioEssays 24*, 953–959.

McGrane, V., Carlson, J.O., Miller, B.R., Beaty, B.J., **1988**. Microinjection of DNA into *Aedes triseriatus* ova and detection of integration. *Am. J. Trop. Med. Hyg. 39*, 502–510.

Medhora, M.M., MacPeek, A.H., Hartl, D.L., **1988**. Excision of the *Drosophila* transposable element *mariner*: identification and characterization of the *Mos* factor. *EMBO J. 7*, 2185–2189.

Miahle, E., Miller, L.H., **1994**. Biolistic techniques for transfection of mosquito embryos (*Anopheles gambiae*). *Biotechniques 16*, 924–931.

Michel, K., Atkinson, P.W., **2003**. Nuclear localization of the *Hermes* transposase depends on basic amino acid residues at the N-terminus of the protein. *J. Cell. Biochem. 89*, 778–790.

Michel, K.A., O'Brochta,, D.A., Atkinson, P.W., **2002**. Does the proposed DSE motif form the active center in the *Hermes* transposase? *Gene 298*, 141–146.

Michel, K., O'Brochta, D.A., Atkinson, P.W., **2003**. The C-terminus of the *Hermes* transposase contains a protein multimerization domain. *Insect Biochem. Mol. Biol. 33*, 959–970.

Michel, K., Stamenova, A., Pinkerton, A.C., Franz, G., Robinson, A.S., *et al.*, **2001**. *Hermes*-mediated germline transformation of the Mediterranean fruit fly *Ceratitis capitata*. *Insect Mol. Biol. 10*, 155–162.

Miller, L.H., Sakai, R.K., Romans, P., Gwadz, R.W., Kantoff, P., *et al.*, **1987**. Stable integration and expression of a bacterial gene in the mosquito *Anopheles gambiae*. *Science 237*, 779–781.

Morris, A.C., **1997**. Microinjection of mosquito embryos. In: Crampton, J.M., Beard, C.B., Louis, C. (Eds.), Molecular Biology of Insect Disease Vectors: A Methods Manual. Chapman and Hall, London, pp. 423–429.

Morris, A.C., Eggleston, P., Crampton, J.M., **1989**. Genetic transformation of the mosquito *Aedes aegypti* by micro-injection of DNA. *Med. Vet. Entomol. 3*, 1–7.

Mullins, M.C., Rio, D.C., Rubin, G.M., **1989**. *Cis*-acting DNA sequence requirements for *P*-element transposition. *Cell 3*, 729–738.

Nawa, S., Sakaguchi, B., Yamada, M.A., Tsujita, M., **1971**. Hereditary change in *Bombyx* after treatment with DNA. *Genetics 67*, 221–234.

Nawa, S., Yamada, S., **1968**. Hereditary change in *Ephestia* after treatment with DNA. *Genetics 58*, 573–584.

Nolan, T., Bower, T.M., Brown, A.E., Crisanti, A., Catteruccia, F., **2002**. *piggyBac*-mediated germline transformation of the malaria mosquito *Anopheles stephensi* using the red fluorescent protein dsRED as a selectable marker. *J. Biol. Chem. 277*, 8759–8762.

O'Brochta, D.A., Atkinson, P.W., Lehane, M.J., **2000**. Transformation of *Stomoxys calcitrans* with a *Hermes* gene vector. *Insect Mol. Biol. 9*, 531–538.

O'Brochta, D.A., Gomez, S.P., Handler, A.M., **1991**. *P* element excision in *Drosophila melanogaster* and related drosophilids. *Mol. Gen. Genet. 225*, 387–394.

O'Brochta, D.A., Handler, A.M., **1988**. Mobility of *P* elements in drosophilids and nondrosophilids. *Proc. Natl Acad. Sci. USA 85*, 6052–6056.

O'Brochta, D.A., Warren, W.D., Saville, K.J., Atkinson, P.W., **1994**. Interplasmid transposition of *Drosophila hobo* elements in non-drosophilid insects. *Mol. Gen. Genet. 244*, 9–14.

O'Brochta, D.A., Warren, W.D., Saville, K.J., Atkinson, P.W., **1996**. *Hermes*, a functional non-drosophilid insect gene vector. *Genetics 142*, 907–914.

O'Hare, K., Rubin, G.M., **1983**. Structures of *P* transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. *Cell 34*, 25–35.

Papadimitriou, E., Kritikou, D., Mavroidis, M., Zacharopoulou, A., Mintzas, A.C., **1998**. The heat shock 70 gene family in the Mediterranean fruit fly *Ceratitis capitata*. *Insect Mol. Biol. 7*, 279–290.

Peloquin, J.J., Thibault, S.T., Schouest, L.P., Jr., Miller, T.A., **1997**. Electromechanical microinjection of pink bollworm *Pectinophora gossypiella* embryos increases survival. *Biotechniques 22*, 496–499.

Peloquin, J.J., Thibault, S.T., Staten, R., Miller, T.A., **2000**. Germ-line transformation of pink bollworm (Lepidoptera: Gelechiidae) mediated by the *piggyBac* transposable element. *Insect Mol. Biol. 9*, 323–333.

Perera, O.P., Harrell, R.A., Handler, A.M., **2002**. Germline transformation of the South American malaria vector, *Anopheles albimanus*, with a *piggyBac/EGFP* transposon vector is routine and highly efficient. *Insect Mol. Biol. 11*, 291–297.

Phillips, J.P., Xin, J.H., Kirby, K., Milne, C.P., Krell, P., *et al.*, **1990**. Transfer and expression of an

organophosphate insecticide degrading gene from *Pseudomonas* in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA 87*, 8155–8159.

Pinkerton, A.C., Michel, K., O'Brochta, D.A., Atkinson, P.W., **2000**. Green fluorescent protein as a genetic marker in transgenic *Aedes aegypti*. *Insect Mol. Biol. 9*, 1–10.

Pinkerton, A.C., Whyard, S., Mende, H.M., Coates, C.J., O'Brochta, D.A., *et al.*, **1999**. The Queensland fruit fly, *Bactrocera tryoni*, contains multiple members of the *hAT* family of transposable elements. *Insect Mol. Biol. 8*, 423–434.

Pirrotta, V., **1988**. Vectors for *P*-mediated transformation in *Drosophila*. In: Rodriguez, R.L., Denhardt, D.T. (Eds.), Vectors: A Survey of Molecular Cloning Vectors and their Uses. Butterworth, Oxford, pp. 437–456.

Pirrotta, V., Steller, H., Bozzetti, M.P., **1985**. Multiple upstream regulatory elements control the expression of the *Drosophila white* gene. *EMBO J. 4*, 3501–3508.

Plautz, J.D., Day, R.N., Dailey, G.M., Welsh, S.B., Hall, J.C., *et al.*, **1996**. Green fluorescent protein and its derivatives as versatile markers for gene expression in living *Drosophila melanogaster*, plant and mammalian cells. *Gene 173*, 83–87.

Prasher, D.C., Eckenrode, V.K., Ward, W.W., Prendergast, F.G., Cormier, M.J., **1992**. Primary structure of the *Aequorea victoria* green fluorescent protein. *Gene 111*, 229–233.

Presnail, J.K., Hoy, M.A., **1994**. Transmission of injected DNA sequences to multiple eggs of *Metaseiulus occidentalis* and *Amblyseius finlandicus* (Acari: Phytoseiidae) following maternal microinjection. *Exp. Appl. Acarol. 18*, 319–330.

Reznikoff, W.S., **2000**. *Tn5* transposition. In: Craig, N.L., Craige, R., Gellert, M., Lambowitz, A.M. (Eds.), Mobile DNA II. American Society for Microbiology, Washington, DC, pp. 403–422.

Reznikoff, W.S., Bhasin, A., Davies, D.R., Goryshin, I.Y., Mahnke, L.A., *et al.*, **1999**. *Tn5*: a molecular window on transposition. *Biochem. Biophys. Res. Commun. 266*, 729–734.

Rio, D.C., Laski, F.A., Rubin, G.M., **1986**. Identification and immunochemical analysis of biologically active *Drosophila P* element transposase. *Cell 44*, 21–32.

Rio, D.C., Rubin, G.M., **1988**. Identification and purification of a *Drosophila* protein that binds to the terminal 31-base-pair inverted repeats of the transposable element *P*. *Proc. Natl Acad. Sci. USA 85*, 8929–8933.

Robertson, H.M., **1993**. The *mariner* transposable element is widespread in insects. *Nature 362*, 241–245.

Robertson, H.M., **2000**. Evolution of DNA transposons in eukaryotes. In: Craig, N.L., Craige, R., Gellert, M., Lambowitz, A.M. (Eds.), Mobile DNA II. American Society for Microbiology, Washington, DC, pp. 1093–1110.

Robertson, H.M., Lampe, D.J., **1995a**. Distribution of transposable elements in arthropods. *Annu. Rev. Entomol. 40*, 333–357.

Robertson, H.M., Lampe, D.J., **1995b**. Recent horizontal transfer of a *mariner* element between Diptera and Neuroptera. *Mol. Biol. Evol. 12*, 850–862.

Robertson, H.M., MacLeod, E.G., **1993**. Five major subfamilies of *mariner* transposable elements in insects, including the Mediterranean fruit fly, and related arthropods. *Insect Mol. Biol. 2*, 125–139.

Robertson, H.M., Zumpano, K.L., **1997**. Molecular evolution of an ancient *mariner* transposon, *Hsmar1*, in the human genome. *Gene 205*, 203–217.

Rong, Y.S., Golic, K.G., **2000**. Gene targeting by homologous recombination in *Drosophila*. *Science 288*, 2013–2018.

Rowan, K., Orsetti, J., Atkinson, P.W., O'Brochta, D.A., **2004**. *Tn5* as an insect gene vector. *Insect Biochem. Molec. Biol. 34*, 695–705.

Rubin, E.J., Akerley, B.J., Novik, V.N., Lampe, D.J., Husson, R.N., *et al.*, **1999**. *In vivo* transposition of *mariner*-based elements in enteric bacteria and mycobacteria. *Proc. Natl Acad. Sci. USA 96*, 1645–1650.

Rubin, G.M., Kidwell, M.G., Bingham, P.M., **1982**. The molecular basis of *P-M* hybrid dysgenesis: the nature of induced mutations. *Cell 29*, 987–994.

Rubin, G.M., Spradling, A.C., **1982**. Genetic transformation of *Drosophila* with transposable element vectors. *Science 218*, 348–353.

Sarkar, A., Coates, C.J., Whyard, S., Willhoeft, U., Atkinson, P.W., *et al.*, **1997a**. The *Hermes* element from *Musca domestica* can transpose in four families of cyclorrhaphan flies. *Genetica 99*, 15–29.

Sarkar, A., Collins, F.C., **2000**. Eye color genes for selection of transgenicinsects. In: Handler, A.M., James, A.A. (Eds.), Insect Transgenesis: Methods and Applications. CRC Press, Boca Raton, FL, pp. 79–91.

Sarkar, A., Yardley, K., Atkinson, P.W., James, A.A., O'Brochta, D.A., **1997b**. Transposition of the *Hermes* element in embryos of the vector mosquito, *Aedes aegypti*. *Insect Biochem. Mol. Biol. 27*, 359–363.

Saville, K.J., Warren, W.D., Atkinson, P.W., O'Brochta, D.A., **1999**. Integration specificity of the *hobo* element of *Drosophila melanogaster* is dependent on sequences flanking the target site. *Genetica 105*, 133–147.

Scavarda, N.J., Hartl, D.L., **1984**. Interspecific DNA transformation in *Drosophila*. *Proc. Natl Acad. Sci. USA 81*, 7515–7519.

Scholnick, S.B., Morgan, B.A., Hirsh, J., **1983**. The cloned Dopa decarboxylase gene is developmentally regulated when reintegrated into the *Drosophila* genome. *Cell 34*, 37–45.

Senecoff, J.F., Bruckner, R.C., Cox, M.M., **1985**. The FLP recombinase of the yeast 2-micron plasmid: characterization of its recombination site. *Proc. Natl Acad. Sci. USA 82*, 7270–7274.

Sheng, G., Thouvenot, E., Schmucker, D., Wilson, D.S., Desplan, C., **1997**. Direct regulation of rhodopsin 1 by *Pax-6/eyeless* in *Drosophila*: evidence for a conserved function in photoreceptors. *Genes Devel. 11*, 1122–1131.

Sherman, A., Dawson, A., Mather, C., Gilhooley, H., Li, Y., *et al.*, **1998**. Transposition of the *Drosophila* element *mariner* into the chicken germ line. *Nature Biotechnol. 16*, 1050–1053.

Siegal, M.L., Hartl, D.L., **1996**. Transgene coplacement and high efficiency site-specific recombination with the *Cre/loxP* system in *Drosophila*. *Genetics 144*, 715–726.

Simmons, M.J., Haley, K.J., Grimes, C.D., Raymond, J.D., Niemi, J.B., **2002**. A *hobo* transgene that encodes the *P*-element transposase in *Drosophila melanogaster*: autoregulation and cytotype control of transposase activity. *Genetics 161*, 195–204.

Spradling, A.C., **1986**. *P* element-mediated transformation. In: Roberts, D.B. (Ed.), *Drosophila*: A Practical Approach. Oxford, IRL Press, pp. 174–197.

Steller, H., Pirrotta, V., **1985**. A transposable *P* vector that confers selectable G4l8 resistance to *Drosophila* larvae. *EMBO J. 4*, 167–171.

Steller, H., Pirrotta, V., **1986**. *P* transposons controlled by the heat shock promoter. *Mol. Cell Biol. 6*, 1640–1649.

Sumitani, M., Yamamoto, D.S., Oishi, K., Lee, J.M., Hatakeyama, M., **2003**. Germline transformation of the sawfly, *Athalia rosae* (Hymenoptera: Symphyta), mediated by a *piggyBac*-derived vector. *Insect Biochem. Mol. Biol. 33*, 449–458.

Sundararajan, P., Atkinson, P.W., O'Brochta, D.A., **1999**. Transposable element interactions in insects: crossmobilization of *hobo* and *Hermes*. *Insect Mol. Biol. 8*, 359–368.

Sundraresan, V., Freeling, M., **1987**. An extrachromosomal form of the *Mu* transposon of maize. *Proc. Natl Acad. Sci. USA 84*, 4924–4928.

Taillebourg, E., Dura, J.M., **1999**. A novel mechanism for *P* element homing in *Drosophila*. *Proc. Natl Acad. Sci. USA 96*, 6856–6861.

Tamura, T., Thibert, T., Royer, C., Kanda, T., Eappen, A., *et al.*, **2000**. A *piggyBac* element-derived vector efficiently promotes germ-line transformation in the silkworm *Bombyx mori* L. *Nature Biotechnol. 18*, 81–84.

Thibault, S.T., Luu, H.T., Vann, N., Miller, T.A., **1999**. Precise excision and transposition of *piggyBac* in pink bollworm embryos. *Insect Mol. Biol. 8*, 119–123.

Thomas, J.L., Bardou, J., L'hoste, S., Mauchamp, B., Chavancy, G., **2001**. A helium burst biolistic device adapted to penetrate fragile insect tissues. *J. Insect Sci. 1*, 9.

Thomas, J.L., Da Rocha, M., Besse, A., Mauchamp, B., Chavancy, G., **2002**. 3xP3–EGFP marker facilitates screening for transgenic silkworm *Bombyx mori* L. from the embryonic stage onwards. *Insect Biochem. Mol. Biol. 32*, 247–253.

Tosi, L.R.O., Beverly, S.M., **2000**. *Cis* and *trans* factors affecting *Mos1 mariner* evolution and transposition *in vitro*, and it potential for functional genomics. *Nucl. Acids Res. 28*, 784–790.

Uhlirova, M., Asahina, M., Riddiford, L.M., Jindra, M., **2002**. Heat-inducible transgenic expression in the silkmoth *Bombyx mori*. *Devel. Genes Evol. 212*, 145–151.

Walker, V.K., **1990**. Gene transfer in insects. *Adv. Cell Culture 7*, 87–124.

Wang, H.H., Fraser, M.J., **1993**. TTAA serves as the target site for TFP3 lepidopteran transposon insertions in both nuclear polyhedrosis virus and *Trichoplusia ni* genomes. *Insect Mol. Biol 1*, 1–7.

Wang, H.H., Fraser, M.J., Cary, L.C., **1989**. Transposon mutagenesis of baculoviruses: analysis of TFP3 lepidopteran insertions at the FP locus of nuclear polyhedrosis viruses. *Gene 81*, 97–108.

Wang, W., Swevers, L., Iatrou, K., **2000**. *Mariner* (*Mos1*) transposase and genomic integration of foreign gene sequences in *Bombyx mori* cells. *Insect Mol. Biol. 9*, 145–155.

Warren, W.D., Atkinson, P.W., O'Brochta, D.A., **1994**. The *Hermes* transposable element from the house fly, *Musca domestica*, is a short inverted repeat-type element of the *hobo*, *Ac*, and *Tam*3 (*hAT*) element family. *Genet. Res. 64*, 87–97.

White, L.D., Coates, C.J., Atkinson, P.W., O'Brochta, D.A., **1996**. An eye color gene for the detection of transgenic non-drosophilid insects. *Insect Biochem. Mol. Biol. 26*, 641–644.

Wilson, C., Pearson, R.K., Bellen, H.J., O'Kane, C.J., Grossniklaus, U., *et al.*, **1989**. *P*-element-mediated enhancer detection: an efficient method for isolating and characterizing developmentally regulated genes in *Drosophila*. *Genes Devel. 3*, 1301–1313.

Wilson, R., Orsetti, J., Klocko, A.D., Aluvihare, C., Peckham, E., *et al.*, **2003**. Post-integration behavior of a *Mos1* gene vector in *Aedes aegypti*. *Insect Biochem. Mol. Biol. 33*, 853–863.

Yamao, M., Katayama, N., Nakazawa, H., Yamakawa, M., Hayashi, Y., *et al.*, **1999**. Gene targeting in the silkworm by use of a baculovirus. *Genes Devel. 13*, 511–516.

Zhang, J.K., Pritchett, M.S., Lampe, D.J., Robertson, H.M., Metcalf, W.W., **2000**. *In vivo* transposon mutagenesis of the methanogenic archeon *Methanosarcina acetivorans C2A* using a modified version of the insect *mariner*-family transposable element *Himar1*. *Proc. Natl Acad. Sci. USA 97*, 9665–9670.

Zhang, H., Shinmyo, Y., Hirose, A., Mito, T., Inoue, Y., *et al.*, **2002**. Extrachromosomal transposition of the transposable element *Minos* in embryos of the cricket *Gryllus bimaculatus*. *Devel. Growth Differ. 44*, 409–417.

Zhang, L., Sankar, U., Lampe, D.J., Robertson, H.M., Graham, F.L., **1998**. The *Himar1 mariner* transposase cloned in a recombinant adenovirus vector is functional in mammalian cells. *Nucl. Acids Res. 226*, 3687–3693.

Zwiebel, L.J., Saccone, G., Zacharapoulou, A., Besansky, N.J., Favia, G., *et al.*, **1995**. The *white* gene of *Ceratitis capitata*: a phenotypic marker for germline transformation. *Science 270*, 2005–2008.

4.14 Insect Cell Culture and Recombinant Protein Expression Systems

P J Farrell, The University of Calgary, Canada
L Swevers and K Iatrou, Institute of Biology, National Centre for Scientific Research ''Demokritos'', Athens, Greece

© 2005, Elsevier BV. All Rights Reserved.

4.14.1. Introduction

In the era of genomics, proteomics, and biotechnology, protein expression systems play key roles. For example, characterization of a gene now requires overexpression of the encoded protein by heterologous expression systems to realize its functional and structural characterization. Furthermore, production of proteins for therapeutic purposes is dependent on the use of host organisms that are genetically engineered for protein production at high levels.

Several host organisms have been used in recent years for recombinant protein expression purposes. These include bacteria, yeast, mammals, and insects, as well as cell lines derived from various mammalian and insect species. Each of these expression systems has unique advantages and deficiencies and, as a result, is more suitable than the others for specific applications while not as good for others.

Bacterial expression systems, for example, can direct very high levels of protein expression but the recombinant proteins generally lack the posttranslational modifications that take place in eukaryotes. Thus, for eukaryotic proteins that require appropriate modifications for functionality, bacterial expression systems are considered inappropriate. Bacteria also do not have eukaryotic-type secretion systems and normally secreted heterologous proteins usually accumulate in the bacterial cytoplasm in the form of insoluble aggregates, known as "inclusion bodies," which make the purification of functional proteins extremely difficult.

Yeast-based expression systems can normally yield high levels of recombinant proteins. However, the expression levels for particular classes of proteins such as secreted or membrane-anchored proteins are, in general, much more limited than those for intracellular proteins. Moreover, the patterns of posttranslational modifications including glycosylation are more limited in yeast than in higher eukaryotes and this limitation often affects biological activity.

Transformed mammalian cells are considered to represent excellent means for expression of membrane-anchored and secreted proteins. However,

mammalian cells need to be maintained in expensive media supplemented with CO_2 under carefully controlled conditions. Most importantly, mammalian cell lines are also potential sources of harmful pathogens, which can limit their utility for production of proteins of pharmaceutical value. Finally, mammalian systems are labor intensive because it requires several months for the generation of clones overexpressing stably recombinant proteins.

Insect cell-based expression systems, which can be used for *in vitro* production of virtually all types of heterologous proteins, accumulate most of the advantages of the other expression systems while lacking most of their disadvantages. Insect cell-derived expression systems are currently divided into two types, (1) the baculovirus expression system, which employs insect cell lines as hosts for baculovirus expression vectors and recombinant protein production in batch mode, and (2) transformed insect cell lines that overexpress continuously proteins from nonlytic plasmid-based expression vectors.

Following the development of the Baculovirus Expression System (BES) in 1983, insect cells have been employed extensively for the production of a wide variety of recombinant proteins using various versions of baculovirus-based expression vectors, and excellent results have been obtained for various research and industrial applications. As a result, the baculovirus/insect cell expression system is currently considered as the industry workhorse for recombinant protein production. As explained below, however, an important limitation of this system relates to the production of secreted proteins and proteins that require extensive posttranslational modifications. Furthermore, the baculovirus expression system is a transient production system because host cells are lysed and killed during each infection cycle.

More recently, a series of baculovirus-free expression vector systems have also emerged, which direct high-level recombinant protein production in insect cell lines, either simply transfected or stably transformed with relevant expression plasmid DNAs, and circumvent the disadvantages of the baculovirus expression system. These insect cell-based expression systems can be divided into two subgroups, one employing dipteran insect cell lines and another that utilizes lepidopteran insect cell lines. For the latter, the incorporation of genetic elements obtained from baculoviruses in the context of plasmid vectors has allowed the generation of extremely powerful expression vectors that direct the generation of large quantities of recombinant proteins. In general, continuous insect cell-based expression systems show outstanding features for the production of recombinant proteins, which, among others, include:

1. A fast process from cDNA cloning to production of recombinant protein – stable cell lines expressing recombinant protein are developed within 1–2 months.
2. High yields of secreted recombinant proteins that may surpass those achieved by the baculovirus system and mammalian cell systems.
3. Utilization of cell lines derived from a multitude of insect species.
4. Easy purification of secreted proteins from serum-free media.
5. Full capacity of posttranslational modifications – the cells maintain their integrity throughout the production process.
6. Continuous protein production – the transgene is stably integrated into the genome of the cells and protein production occurs continuously because no lysis occurs.

Irrespective of whether it is associated with the use of baculovirus or plasmid expression vectors, the use of insect cells can overcome the shortcomings of other established protein expression workhorses. Among others, the advantages of insect cells over other expression systems include the capacity to produce higher levels of soluble recombinant proteins that are folded correctly and extensively modified posttranslationally, the fact that both baculoviruses and insect cell lines are nonpathogenic to humans, and that insect cells can be grown in medium free of protein or other, potentially pathogenic, animal products.

This article is intended as a practical review of the developments in both baculovirus and plasmid-based protein expression systems employing insect cells. New applications of insect cells are also discussed including surface display and high-throughput screening for the identification of bioactive compounds. For related information on this topic, particularly in relation to applications related to baculoviruses and expression systems derived from them, the reader is also referred to **Chapters 6.8** and **6.9**.

4.14.2. Baculovirus Expression Systems

4.14.2.1. Brief History of the Baculovirus Expression System

Insect cells have certainly gained most of their popularity as hosts for production of foreign proteins using recombinant baculoviruses as expression vectors. For this application, the Open Reading Frames (ORFs) of heterologous genes substitute for ORFs of nonessential genes of a baculovirus

or heterologous ORFs are placed under the control of either viral promoters or nonviral promoters inserted into "neutral" sites of the baculovirus genome. Following infection of cultured insect cells (or larvae) with the recombinant baculovirus, the heterologous gene is expressed. By taking advantage of the several powerful promoters of nonessential genes, the levels and quality of the recombinant protein obtained from insect cells infected with the relevant recombinant baculoviruses are frequently superior to those obtained from other expression systems.

The original description of a recombinant baculovirus for foreign protein expression appeared in a series of articles from Max Summer's laboratory at Texas A&M University in 1983. Their model baculovirus system was the *Autographa californica* nucleopolyhedrosis virus (AcNPV), a member of the nuclear polyhedrosis virus (NPV) subfamily of baculoviruses that have the unique characteristic of producing large proteinaceous occlusion bodies (OBs) during the very late stages of infection. OBs constitute over 25% of the total protein mass within an infected cell (Smith *et al.*, 1983a) and serve to protect the mature virions embedded within them from the damaging effects of the natural environment. OBs are mostly composed of a single protein, polyhedrin, whose gene was initially cloned from AcNPV (Smith *et al.*, 1983b). Deletions were introduced in the cloned polyhedrin gene in plasmids, and several mutant AcNPVs with an OB negative (OB^-) phenotype were generated by homologous recombination following cotransfection of insect cells with wild-type AcNPV DNA and the plasmid deletions (Smith *et al.*, 1983c). Although lacking the ability to synthesize the complete polyhedrin protein, OB^- AcNPVs were still able to infect cultured cells and produce progeny virus. Furthermore, the polyhedrin promoter was shown to remain as active in OB^- AcNPVs and to produce various forms of truncated polyhedrin protein at levels similar to those of the full length polyhedrin protein expressed by the wild-type AcNPV.

Having established that the highly expressed polyhedrin protein was not essential for the survival and propagation of AcNPV, the researchers at Texas A&M constructed various transfer vectors for inserting foreign genes into different positions of the polyhedrin gene region of the AcNPV genome by homologous recombination. Using human beta interferon (IFN-β) as a secreted reporter protein, OB^- recombinant viruses were obtained that expressed either chimeric polyhedrin-interferon proteins or authentic IFN-β under the control of the polyhedrin promoter. The recombinant IFN-β was secreted (with concomitant removal of its signal peptide), glycosylated and biologically active, and was expressed at levels up to 10 μg per ml or 100- to 500-fold higher than what was possible at that time using transformed mouse or bacterial expression systems, respectively (Smith *et al.*, 1983b).

The successful expression of IFN-β marked the start of the baculovirus/insect cell protein expression system. Other reports of high-level recombinant protein expression followed soon afterwards, which included the production of a fusion between polyhedrin and *Escherichia coli* β-galactosidase (Pennock *et al.*, 1984), human c-Myc (Myamoto *et al.*, 1985), and human interleukin-2 (Smith *et al.*, 1985).

4.14.2.2. Baculovirus Life Cycle

The baculovirus life cycle has been described in detail elsewhere (O'Reilly *et al.*, 1992; Vialard *et al.*, 1995; see **Chapter 6.8**).

Members of the *Baculoviridae* family are characterized by a rod-shaped enveloped virion and a circular, double-stranded and covalently closed DNA genome of 60–160 kbp (Mathews, 1982). They infect mainly insects, are not infectious to vertebrates, and are classified under two subfamilies: NPVs and granulosis viruses (GVs), which are members of the subfamily of *Eubaculovirinae* which comprises OB-forming viruses, while under the family of *Nudilibaculovirinae* are classified nonoccluded baculoviruses. Most baculoviruses have a very narrow host range, with the majority of the *Eubaculovirinae* infecting lepidopteran insects. NPVs are also the main constituents of expression vectors, with those species receiving the most attention being AcNPV, first isolated from larvae of the alfalfa looper *A. californica*, and the nuclear polyhedrosis virus of the silkworm *Bombyx mori* (BmNPV).

After an insect dies from an NPV infection, viral OBs, which serve to protect the virions embedded within them from the damaging effects in the natural environment such as UV radiation, are released from the carcass. When these OBs are ingested by an uninfected feeding insect, they are dissolved in the midgut (polyhedrin is solubilized due to the alkaline pH of the gut lumen), and the occluded virions are freed. Following fusion of the virions with midgut epithelial cells, the nucleocapsids migrate to the cell nuclei, where the genome is uncoated and transcription begins. A temporally regulated gene expression cascade ensues, that directs replication and formation of new virions (**Figure 1**).

Two forms of viruses are produced at different phases of the infection cycle: initially nucleocapsids are produced in the nucleus and enveloped upon

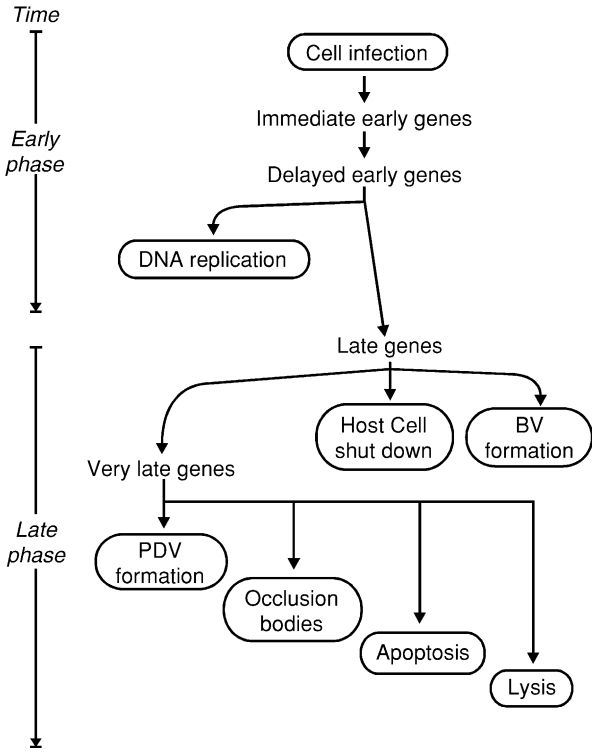


Figure 1 A simplified view of the viral gene expression cascade following baculovirus infection of an insect cell. The cascade is divided into two temporal halves, the first half terminating upon completion of DNA replication of the baculovirus genome, and the second one leading to the production of both budded virions (BVs) and polyhedra-derived virions (PDVs). The infection process ends with the lysis of the host cell.

budding from the cell's plasma membrane. The envelopes of these budded virions (BVs) are also distinguished by the presence of gp64 (also known as gp67), a baculovirus glycoprotein essential for BV entry into the cells by endocytosis (Whitford *et al.*, 1989; Monsma *et al.*, 1996). BVs are responsible for the systemic spread of the infection (secondary infection of neighboring cells or other tissues via the tracheal system of the infected insect). Later in the infection cycle, a second type of virion is generated. These are enclosed, either individually or in association with others, in an envelope of nuclear origin, prior to their incorporation into OBs. This second form, termed polyhedra-derived virions (PDVs), are distinguished from BVs by the association of several viral proteins, including gp41 (Whitford and Faulkner, 1992), p25 (Russell and Rohrmann, 1993), and p74, which is essential for PDV infectivity (Kuzio *et al.*, 1989), with the PDV envelope.

The baculovirus infection of insect cells derived from larval tissues and cultured *in vitro* is similar to that occurring in nature: susceptible cultured cells must be derived from susceptible insect species. OBs, appearing as obvious crystal structures under a simple light microscope, form in the nuclei and are released from lysed cells. Due to the nonalkaline nature of the insect cell culture medium (~pH 6.2), OBs remain insoluble and the PDVs are trapped within them. Thus, the infection is only spread to other cultured cells by BVs. PDVs extracted from OBs can also be used for the primary infection of cultured cells, albeit with significantly lower efficiency than BVs (Volkman and Summers, 1977).

The transcription of the 150 or so baculovirus genes following host cell infection is temporally regulated in four phases (**Figure 1**): immediate early and early genes are transcribed prior to viral DNA replication; late and very late genes are transcribed after replication. Immediate early gene promoters are activated by host factors immediately following uncoating, while delayed early, late, and very late gene promoters require the availability of viral transcription factors, synthesized during the previous phases, for transcriptional activation in the gene expression cascade. Approximately 20 activators of transcription have been identified in AcNPV including IE-1 (Guarino and Summers, 1986), IE-2 (or IE-N; Carson *et al.*, 1988), several late expression factors (LEFs; Lu and Miller, 1995; Rapp *et al.*, 1998) and one very late factor, VLF-1, that is necessary for transcription from the polyhedrin promoter (Yang and Miller, 1999) (for additional details, see **Chapter 6.8**).

4.14.2.3. Early Vectors and Techniques for Generating Recombinant Baculoviruses

Due to the large size of the baculovirus genome, it is difficult to routinely manipulate the baculovirus DNA for ligation of foreign genes directly into its genome. Until recently, the most common approach was to first clone the foreign gene into a plasmid "transfer vector," which is easy to manipulate using regular molecular biology techniques. The purpose of the transfer vector is to target an insertion site in the baculovirus genome for homologous recombination (via double-crossover). Thus, a generic transfer vector contains (1) a promoter, (2) a multiple cloning site for insertion of a heterologous ORF, (3) a poly(A) addition site, and (4) a few kilobase pairs of DNA sequence matching the destination locus in the baculovirus genome flanking elements 1–3. Usually, the destination locus is the polyhedrin region, because the polyhedrin gene promoter is very powerful and the polyhedrin gene itself is nonessential for the virus life cycle *in vitro*. Furthermore, the OB^- phenotype of recombinant viruses facilitates their selection. Most transfer vectors are designed

to simply swap, by homologous recombination, the polyhedrin ORF in the baculovirus genome with a foreign gene, while conserving the polyhedrin promoter and transcription termination [(poly(A) addition site)] regions. Various permutations of this basic approach have also been used (see below).

After the insertion of a foreign gene into the transfer vector is accomplished, the recombinant transfer vector is cotransfected with wild-type baculovirus DNA into insect cells. Within the nucleus, a relatively low frequency genetic exchange, by homologous recombination, occurs between the sequences of the transfer plasmid and the wild-type baculovirus DNA, and a small fraction of recombinant virus is generated. As an alternative to a transfer vector, PCR products generated with primers containing only 50 nucleotides of flanking DNA have also been used successfully, although the frequency of homologous recombination was significantly lower (Gritsun *et al.*, 1997). The recombinant viruses can be discriminated visually and isolated from wild-type ones in virus plaque purification assays either based on their OB^- phenotype (for the cases of poyhedrin ORF disruption) or by virtue of other visual or molecular phenotypes expressed by them (in the cases where the polyhedrin gene is not disrupted). Another widely used technique for isolating a recombinant virus employs end-point dilution cloning and identity confirmation by PCR, dot-blot hybridization, immunoblotting or activity assays. Once isolated, the recombinant virus is amplified by one or more infection cycles in insect cells. At this point the virus stock is ready for recombinant protein expression. The whole process of preparation of the recombinant transfer vector, cotransfection, purification, amplification, and verification of protein expression usually takes 4 to 6 weeks to complete.

4.14.2.3.1. Polyhedrin promoter As has been already mentioned above, the polyhedrin promoter is by far the most frequently exploited promoter for recombinant protein expression using baculovirus vectors. Initially, transfer vector pAc373 (Smith *et al.*, 1985) was used to target the AcNPV polyhedrin gene locus for foreign gene expression. This vector contained a deletion of approximately 182 bp between nucleotides −9 and +175 relative to the polyhedrin ATG translation start site. However, the early studies on the use of this specific and other related transfer vectors revealed that recombinant protein expression was less efficient when the transfer vectors encompassed disruptions in (1) the immediate vicinity of the normal polyhedrin translation start ATG codon, both in the 5′-UTR (Matsuura *et al.*, 1987; Possee and Howard, 1987) and the polyhedrin N-terminal region (Luckow and Summers, 1988, 1989); (2) the 15–20 nt upstream of the polyhedrin mRNA transcription initiation site (Possee and Howard, 1987; Rankin *et al.*, 1988); and, especially (3) the polyhedrin TAAG transcription initiation site itself (Possee and Howard, 1987; Rankin *et al.*, 1988). In addition, it was quickly discovered that the polyhedrin sequences downstream of the poly(A) addition site cannot be deleted extensively due to the presence of an ORF necessary for AcNPV replication (Possee *et al.*, 1991). To maximize foreign gene expression from the polyhedrin promoter, improved transfer vectors, such as pAcYM1 (Page, 1989) or pVL941 (Luckow and Summers, 1989), were designed subsequently, which targeted the polyhedrin locus while preserving, through point mutations (ATG to ATC or ATT), the overall context of the region surrounding the normal polyhedrin translation start site.

4.14.2.3.2. *p10* promoter The *p10* gene promoter, another powerful very late phase promoter, was also harnessed as an alternative to the polyhedrin promoter (Vlak *et al.*, 1988). P10 is a 10 kDa protein of ambiguous function, probably involved in lysis (van Oers *et al.*, 1993) or occlusion body formation (Williams *et al.*, 1989), that is not essential for the production of either budded or extracellular virus (Vlak *et al.*, 1988), at least *in vitro*. Due to the lack of a selectable phenotype from a recombinant virus targeting the *p10* locus, the β-galactosidase reporter protein was used to identify recombinant viruses. In a report related to comparative aspects of protein production using baculovirus-based vectors, the expression of β-galactosidase and a secreted reporter protein, juvenile hormone esterase (JHE), from recombinant AcNPVs using the *p10* gene promoter were found to be higher and 8 h earlier than that obtained with the polyhedrin promoter (Bonning *et al.*, 1994), while work from another group suggested comparable levels of expression (DiFalco *et al.*, 1997).

4.14.2.3.3. Basic protein promoter The basic protein is a late gene product associated with the viral DNA within the nucleocapsid. The harnessing of this promoter allows the expression of foreign genes at earlier times than those using the very late phase promoters of the polyhedrin and *p10* genes. As the presence of the basic protein gene was suspected to be essential for viral replication, its deletion was not expected to yield viable recombinant virus. For this reason, its promoter was duplicated and inserted into the polyhedrin gene locus (Hill-Perkins and

Possee, 1990). The promoter duplication was not reported to have an effect on virus stability. Maximal rates of expression of β-galactosidase from recombinant AcNPV using this promoter occurred 3 to 6 h earlier than from the polyhedrin gene promoter but the expression levels were reported to be lower than those obtained using the polyhedrin promoter. Later studies have found, however, that the yields of both β-galactosidase and JHE from recombinant AcNPVs employing the basic protein promoter were higher than those obtained using the polyhedrin promoter (Bonning *et al*., 1994).

4.14.2.3.4. Constitutive cellular and early viral promoters For the expression of recombinant proteins requiring extensive posttranslational modifications and secretion, the use of late and very late promoters is suboptimal due to the compromising of the normal cellular processes during the late stages of baculovirus infection (Jarvis and Summers, 1989). Moreover, for biopesticide applications involving the expression of insecticidal proteins by recombinant baculoviruses *in vivo* (see **Chapter 6.9**), it is desirable to be in a position to express insect-toxic proteins at an early time in the baculovirus infection cycle.

One alternative to the use of late promoters was the employment of a constitutively active cellular promoter. A recombinant BmNPV was generated that expressed chloramphenicol acetyl-transferase (CAT) reporter protein under control of the *B. mori* cytoplasmic actin promoter (Johnson *et al*., 1992). The CAT protein could be detected in the infected silkworm cells grown in culture as quickly as 5 h postinfection, compared to 20 h using the polyhedrin promoter, and the overall CAT yield was only threefold lower using the actin promoter relative to that obtained with the polyhedrin promoter. Furthermore, the actin promoter was also shown to function in cultured cells derived from other lepidopteran species such as *Spodoptera frugiperda* and *Choristoneura fumiferana* and could be potentially incorporated into other baculovirus species such as AcNPV.

Similarly, Morris and Miller (1992) generated several recombinant AcNPVs expressing CAT under control of promoter elements obtained from the following genes: *Drosophila melanogaster* hsp70, AcNPV IE-N, AcNPV IE-1, AcNPV IE-0, and AcNPV ETL. While CAT activity could be detected rapidly following infection in most cases, the levels of CAT produced using these promoters were one to two orders of magnitude lower compared to a conventional recombinant AcNPV that employed the polyhedrin promoter to drive foreign CAT expression. For secreted proteins, however, similar or higher overall levels of recombinant protein were obtained from recombinant AcNPVs employing the IE-1 promoter as the driver of foreign gene expression, as compared to a conventional baculovirus expression vector (Jarvis *et al*., 1996).

4.14.2.3.5. Multiple promoters There are two strategies to accommodate applications, where cells are required to synthesize more than one protein, e.g., for production of antibodies or virus-like particles. One method is to infect cultures with more than one recombinant virus, each expressing a unique protein. This requires a careful determination of the stoichiometric balance of each virus required for achieving consistent ratios of each recombinant protein. The second approach employs multiple transfer vectors for generating a single AcNPV expressing simultaneously several different proteins. Duplication of the polyhedrin gene within one recombinant AcNPV was shown to be feasible (Emery and Bishop, 1987), and this approach was used to produce Bluetongue virus core-like particles (French and Roy, 1990) and monoclonal antibodies (zu Putlitz *et al*., 1990). Transfer vectors have been also developed for the generation of a single recombinant AcNPV capable of producing up to five different recombinant proteins, by duplication of both the polyhedrin and p10 gene promoters, without causing genetic instability of the viral genome, at least over a course of six passages (Belyaev *et al*., 1995).

4.14.2.4. Improvements in Baculovirus Expression Vectors

The generation of recombinant baculoviruses using the original protocols is often technically difficult, time consuming, and labor intensive. First, the frequency of a successful homologous recombination event between the transfer vector and wild-type baculovirus DNA is typically in the range of 0.2–1% (O'Reilly *et al*., 1992), therefore one or more rounds of purification by plaque assay or dilution cloning are necessary for the isolation of a pure recombinant virus. Second, at least one further round of amplification is required in order to generate a suitable viral titer before the virus can be used in a small scale test for expression. In total, this usually takes at least 3–4 weeks and up to 6 weeks (O'Reilly *et al*., 1992; Luckow *et al*., 1993). However, over the last decade, several companies have marketed commercially available kits that have capitalized on ingenious molecular biology designs that circumvent some of the bottleneck problems

associated with the traditional methods of preparing recombinant baculoviruses and accelerate their identification and purification.

4.14.2.4.1. Linearized baculovirus DNA Linearization of AcNPV at the destination locus can increase the recovery of recombinant baculoviruses when the former is cotransfected with an appropriate transfer vector (Kitts *et al*., 1990). AcRP23.LacZ AcNPV DNA (Becton Dickinson/Pharmingen; Possee and Howard, 1987) contains a *lacZ* gene in place of the polyhedrin gene, in addition to a unique *Bsu*361 site that has been introduced downstream of the polyhedrin promoter. Cotransfection of insect cells with *Bsu*361-linearized AcRP23.LacZ and a transfer vector targeting the polyhedrin locus results in approximately 30% of the progeny viruses being recombinant via homologous recombination (Kitts *et al*., 1990). This reduces significantly the number of progeny viruses that have to be screened in order to achieve purification of recombinant ones. Furthermore, the isolation of recombinants in plaque assays or limiting dilution assays is facilitated by the fact that the recombinants do not produce LacZ and appear colorless, as opposed to the parental blue ones, in the presence of X-gal. Linearized AcUW1.LacZ DNA (Becton Dickinson/Pharmingen; Weyer *et al*., 1990) is a modified AcNPV DNA similar to AcRP23.LacZ but designed to be used with transfer vectors targeting the *p10* gene promoter.

BaculoGold™ DNA (Becton Dickinson/Pharmingen) and BacPAK™ (Clontech) are modified AcNPV DNAs with three *Bsu*36I sites inserted in the vicinity of the polyhedrin locus (Kitts and Possee, 1993). Digestion of the DNAs with *Bsu*36I yields linear DNAs encompassing a lethal deletion in the essential ORF 1629 (Possee and Howard, 1987) that maps downstream of the polyhedrin locus, and this precludes the generation of a viable virus upon self-ligation of the restricted parental viral genome. When provided with a transfer vector spanning the deleted region, homologous recombination in the transfected insect cells rescues the lethal deletion. Over 99% of viruses produced by the transfected cells are recombinants and only one round of purification is necessary. These DNAs also contain a *lacZ* gene at the polyhedrin locus, which is substituted for by the gene of interest in a successful recombination event. Thus, the isolation of recombinant baculoviruses by plaque or limiting dilution assays is facilitated by the lack of expression of LacZ and the presence of colorless infected cells upon addition of X-gal. BaculoGold™ Bright (Becton Dickinson/Pharmingen) recombinant baculoviruses express the green fluorescence protein (GFP), that can be visualized 24 h postinfection (p.i.), under the control of the *p10* gene promoter. This can facilitate the purification of recombinant virus by fluorescence activated cell sorting (FACS) sorting of single GPF positive cells. The titration process is also simplified by FACS analysis or by a serial dilution and visualization using fluorescent microscopy.

BacVector® (Novagen) and Bac-N-Blue™ (Invitrogen) are modified linearized AcNPV DNAs similar to BaculoGold™ and BacPAK™. They also contain a lethal deletion in ORF1629 that is rescued by a transfer vector. The relevant transfer vectors target the polyhedrin locus and contain reporter genes encoding either β-glucuronidase (pBac, Novagen) or β-galactosidase (pBlueBac, Invitrogen) under the control of the AcNPV ETL promoter. This allows for visual identification of a successful homologous recombination event by the scoring of blue (recombinant) plaques in a background of colorless ones following the addition of X-gluc or X-gal, respectively. Finally, BacVector-3000 is a modified linearized AcNPV DNA (Novagen) containing deletions of several nonessential genes that, apparently, compete with target protein production, in addition to deletions of the genes encoding two baculovirus degrading enzymes, cathepsin and chitinase, which are claimed to destabilize recombinant proteins.

4.14.2.4.2. Shuttle vectors Certainly one of the most ingenious and fastest methods to generate recombinant AcNPV is the Bac-to-Bac® expression system (Luckow *et al*., 1993), licensed to Invitrogen (formerly Life Technologies) by Monsanto. In this system, the recombination event between the AcNPV and transfer vector occurs in *E. coli* DH10B by site-specific transposase-mediated insertion, as opposed to homologous recombination in insect cells. The DH10Bac™ strain harbors a modified AcNPV genome, known as bacmid, that is essentially a large (136 kbp) shuttle plasmid containing, in addition to the baculovirus genome sequences, a kanamycin resistance marker, a low copy number mini-F replicon for retention and replication in *E. coli*, and the target site *att*Tn7 for the Tn7 bacterial transposon in the polyhedrin locus. A helper plasmid providing Tn7 transposase is also maintained in the DH10Bac™ strain by tetracyclcine resistance. When DH10Bac™ cells are transformed with a third plasmid containing the gene of interest flanked by Tn7 elements, transposition of the gene of interest occurs to the Tn7 target site in the polyhedrin locus of the bacmid. The successful transposition disrupts the expression of a

LacZ reporter gene in the bacmid, so that generation of a recombinant bacmid results in white *E. coli* colonies in blue background colonies arising from unaltered parent cells. The recombinant bacmid DNA is isolated from minipreps of white colonies and then used to transfect insect cells and generate viral stocks for small-scale recombinant protein expression. The whole process is reported to take 7–10 days (Luckow *et al.*, 1993).

Patel *et al.* (1992) also described a rapid method for generating recombinant baculoviruses by homologous recombination between a modified AcNPV genome propagated in yeast and a baculovirus transfer vector containing yeast sequences. With this yeast-baculovirus shuttle system, it is reported that recombinant viral stocks can be obtained within 10–12 days. This system is not available from a commercial supplier.

4.14.2.4.3. Direct cloning Several publications have also appeared describing direct cloning of the gene of interest into polyhedrin locus of the AcNPV genome (Ernst *et al.*, 1994; Lu and Miller, 1996). Modified AcNPV's were generated that could be linearized with restriction enzymes in the polyhedrin locus. Ligation with the foreign gene fragment *in vitro* and transfection of the ligation mix into insect cells resulted in the percentage of recombinant viruses approaching 100%. Invitrogen Corporation has recently developed Baculodirect™, a direct cloning strategy for generating recombinant AcNPV using the well-characterized lambda-phage site-specific recombination system (Gateway™) as opposed to a homologous recombination. In addition, the recombination event transferring the desired gene to the polyhedrin locus also includes the HSV thymidine kinase (TK) selection marker that is driven by the IE-1 promoter. Once the *in vitro* recombinase reaction is complete, the reaction contents are transfected into insect cells and selective pressure of ganciclovir is applied so that only those viruses harboring the TK resistance marker can replicate. Within 7 days, purified viral stocks are available for expression studies.

4.14.2.5. Cell Lines Used as Hosts for Baculovirus Expression Vectors

With the popularity of the BES and the fact that the vast majority of baculovirus expression vectors are based on AcNPV, it is not surprising that recent developments in insect cell culture have focused on cell lines derived from species that can support both the replication of AcNPV and the high level expression of foreign proteins under control of the AcNPV polyhedrin, p10 and basic protein gene promoters. AcNPV has a relatively broad host range among lepidopteran species and cell lines established from them, including the fall armyworm *S. frugiperda*, the cabbage looper *Trichoplusia ni*, the tobacco budworm *Heliothis virescens* and the gypsy moth *Lymantria dispar*. However, those cell lines that are the most widely used to date as hosts for protein expression include IPLB-Sf21AE (Vaughn *et al.*, 1977) established from *S. frugiperda* pupal ovaries, its subclone Sf9 (Summers and Smith, 1987) and BTI-Tn-5B1-4 or High Five™ cells established from *T. ni* embryos (Granados *et al.*, 1994). These cell lines are well characterized and have a history of good growth and recombinant protein expression in serum-free culture medium and in large-scale suspension cultures. In contrast to AcNPV, BmNPV has a narrow host range; Bm5 (Grace, 1967) and BMN-4 cells (Maeda, 1989) established from *B. mori* ovarian tissue cells are used as host cells for BmNPV infections. Although both Bm5 and BMN-4 lines can be propagated in serum-free media, Bm5 cells have superior growth characteristics in suspension culture (Keith *et al.*, 1999) and have been reported to produce recombinant protein in bioreactors (Zhang *et al.*, 1993).

4.14.2.6. Classes of Recombinant Proteins Expressed by Baculovirus Infected Insect Cells

Insect cells are capable of most types of posttranslational modifications that occur in mammalian cells. They include signal peptide cleavage, N- and O-linked glycosylation (although they may not be capable of a significant level of complex glycosylation), phosphorylation, acylation, palmitylation, myristoylation, prenylation, amidation, and carboxymethylation (see Luckow, 1991 for an extensive compilation to that date). Heterologous proteins are also correctly targeted to their subcellular destination including plasma membranes, cytosol, lysosomes, mitochondria, nuclei, or nucleoli. Hence, different classes of recombinant proteins from a variety of organisms have been expressed successfully using the baculovirus expression system. These include cytoplasmic proteins, structural proteins, lysozomal enzymes, membrane receptors, ion exchangers, transport proteins, and nuclear factors (**Table 1**).

Although it is unwise to make gross generalizations, some classes of proteins are certainly expressed at much higher levels than others, and each individual protein within a class will often have its own unique troublesome characteristics that a researcher will have to deal with: some proteins may form insoluble aggregates, others are more vulnerable to proteolysis (especially following

lysis when degradative enzymes are released), secretion may be inefficient, expression levels may be low and those proteins requiring a physiologically intact cell membrane may have to be harvested prematurely.

4.14.2.6.1. Intracellular proteins Clearly the most successful application of the baculovirus expression system is for expression of intracellular recombinant proteins. Expression levels over 100–200 mg l^{-1} are not uncommon (**Table 1**) and proteins may be released by the cells into the culture medium following virus-induced lysis to simplify purification. Insoluble aggregates of recombinant intracellular protein within the cell have occasionally been reported (Alnemri and Litwack, 1993; Rankl *et al*., 1994; Baldock *et al*., 2000). In rare cases, the fusion of an insect-derived signal peptide to the N-terminus of a normally intracellular targeted protein has been successful for secreting the intracellular protein (Mroczkowski *et al*., 1994; Laukkanen *et al*., 1996).

4.14.2.6.2. Secreted proteins Secreted proteins are usually expressed at much lower levels compared to intracellular proteins, usually in the order of 10 mg l^{-1} (**Table 1**). This is thought to be due to a breakdown in the secretory pathway in the late stages of infection, when foreign gene expression from late and very late promoters is maximal (Jarvis and Summers, 1989). Three approaches have been tried to improve the efficiency of secretion. In the first approach, the substitution of the native signal peptide of a heterologous secreted protein with an insect-specific signal peptide, such as those of the honeybee mellitin (Tessier *et al*., 1991; Jarvis *et al*., 1993), *Drosophila* cecropin B (Jarvis *et al*., 1993), AcNPV 64K (Jarvis *et al*., 1993), AcNPV gp67 (Golden *et al*., 1998; Murphy *et al*., 1993), AcNPV chitinase (Chen *et al*., 2000), and AcNPV egt (Murphy *et al*., 1993), was expected to make secretion more efficient. In some cases expression was improved dramatically using insect-specific signal peptides (Murphy *et al*., 1993; Chen *et al*., 2000), while in others the modifications proved ineffective (Jarvis *et al*., 1993; Golden *et al*., 1998). In the second approach, the use of an earlier promoter (but weaker than late and very late promoters) such as IE-1 to drive the expression of a heterologous secreted protein, before the secretory pathway was compromized, resulted in only a slight improvement in the expression level (Jarvis *et al*., 1996). The third approach was to coexpress a molecular chaperone protein such as immunoglobulin heavy chain binding protein (BiP), although no improvement in secretion of immunoglobulin was observed (Hsu *et al*., 1994). If glycosylation is not important for a particular secreted protein, one strategy to improve expression is to remove the signal sequence altogether. This will prevent its entry into the secretory pathway and disguise the protein as an intracellular protein. Unpublished results from Dr Bruce Hammock's laboratory (personal communication) have shown an improvement in the expression of active *H. virescens* JHE from 4 to 75 mg l^{-1} using this strategy (**Figure 2**).

4.14.2.6.3. Membrane proteins The reputation of the baculovirus expression system for lower expression levels of proteins destined for the secretory pathway compared to other proteins (Jarvis and Guarino, 1995) also applies to membrane proteins. This said, expression levels of membrane proteins are often higher than those achieved with other expression systems; the expression of membrane proteins in bacteria, for example, is largely unsuccessful. Certainly many laboratories have used the BES to express and purify a variety of membrane proteins. Some examples include ion exchangers (Li *et al*., 1992; Dale *et al*., 1996; Egger *et al*., 1999), G protein-coupled receptors (GPCRs) (Strosberg and Guillaume, 2000; reviewed by Bouvier *et al*., 1998; Carpentier *et al*., 2001), transport proteins (Smith *et al*., 1992; Tate, 1998; Gao *et al*., 1999; Miyasaka *et al*., 2001), adhesion molecules (Stoltenberg *et al*., 1993; Nagarajan and Selvaraj, 1999; Phan *et al*., 2001), and other membrane proteins (Xiao *et al*., 2000; Soares *et al*., 2001; Sun *et al*., 2002). Purified levels obtained are generally in the range of 0.1–10 mg l^{-1} (Chinni *et al*., 1998; Carpentier *et al*., 2001; Miras *et al*., 2001; Soares *et al*., 2001). There are numerous reports of either inactive, immature, misfolded membrane proteins or incompletely glycosylated forms of baculovirus-expressed membrane proteins (Pajot-Augy *et al*., 1995; Vasudevan *et al*., 1995; Loisel *et al*., 1997; Massotte *et al*., 1997; Tate, 1999; Sun *et al*., 2002). Improvements in the yields of active membrane proteins have been obtained by the coexpression of molecular chaperones including calreticulin, BiP, cannexin, and nina A (Tate, 1999; Miyasaka *et al*., 2001; Lenhard and Reilander, 1997).

For those studying membrane protein signaling events following ligand binding, there can be some difficulties using the BES. Obviously, there is the fact that an inherent property of the baculovirus is that it will ultimately kill the host cell. Thus, only a short window of opportunity exists for the study of receptor signaling events in live cells, between the moment of expression from a late- or very late-phase promoter and cell death. Some investigators have

Table 1 Expression levels of some secreted and intracellular proteins expressed using the baculovirus expression system. The expression levels reported are in mg per liter of *unprocessed* cell culture

Protein	*Baculovirus*	*Promoter*	*Level (mg l^{-1})*	*Culture conditions*	*Reference*
Secreted					
Hu tPA	AcNPV/Sf9	Polyhedrin	2.5	n.r.	Steiner *et al.* (1988)
Hu SEAP	AcNPV/High Five	Polyhedrin	13	n.r.	Davis *et al.* (1992)
Rat PCE	AcNPV/Sf9	Polyhedrin	25	Shake flask	DiPersio *et al.* (1992)
Hu IL-8	AcNPV/Sf9	Polyhedrin	0.5–2	T-flask	Kang *et al.* (1992)
Hu LTA_4H	AcNPV/Sf9	Polyhedrin	100	10 l bioreactor	Gierse *et al.* (1993)
Hu C9	AcNPV/Sf21	Polyhedrin	0.7	Shake flask	Tomlinson *et al.* (1993)
HIV gp120	AcNPV/Sf9	Polyhedrin	10–15	5 l reactor	Murphy *et al.* (1993)
Hu procolipase	AcNPV/High Five	Polyhedrin	15	Spinner flask	Lowe (1994)
Hu PSA	AcNPV/Sf9	Polyhedrin	2–4	30 l bioreactor	Kurkela *et al.* (1995)
In JHE	AcNPV/Sf21	Polyhedrin	12	6-well plate	Bonning and Hammock (1995)
In JHE	AcNPV/Sf21	P10	13	6-well plate	Bonning and Hammock (1995)
In JHE	AcNPV/Sf21	Basic protein	16	6-well plate	Bonning and Hammock (1995)
In JHE	AcNPV/High Five	Polyhedrin	32	Shake flask	Bonning and Hammock (1995)
In JHE	AcNPV/Sf21	Polyhedrin	22	Spinner flask	Bonning and Hammock (1995)
Hu IL-5	AcNPV/Sf9	Polyhedrin	5–15	3 l reactor	Brown *et al.* (1995)
Ms OSF-2	AcNPV/Sf9	Polyhedrin	40	15 cm dish	Sugiura *et al.* (1995)
Hu HCII	AcNPV/High Five	Polyhedrin	3	T-flask	Ciaccia *et al.* (1995)
GalNAc-transferase	AcNPV/Sf9	Polyhedrin	2–3	Shake flask	Homa *et al.* (1995)
Hu IL-5	AcNPV/Sf9	Polyhedrin	5–15	n.r.	Brown *et al.* (1995)
Hu LIF	AcNPV/Sf21	Polyhedrin	12	Roller bottle	Geisse *et al.* (1996)
Hu ApoA-I	AcNPV/Sf21	Polyhedrin	40–50	1 l reactor	Sorci-Thomas *et al.* (1996)
Hu prorennin	AcNPV/Sf9	Polyhedrin	40–50	23 l reactor	Mathews *et al.* (1996)
Ms FAK	AcNPV/Sf9	Polyhedrin	19	1 l suspension	Withers *et al.* (1996)
In transferrin	AcNPV/Sf9	Polyhedrin	10	Roller bottles	Winzerling *et al.* (1996)
Hu prorennin	AcNPV/Sf9	Polyhedrin	0.5	Spinner flasks	Mathews *et al.* (1996)
Ms GHBP	AcNPV/Sf21	Polyhedrin	17	Shake flasks	Thordarson *et al.* (1996)
Chk avidin	AcNPV/Sf9	Polyhedrin	12.2	Shake flask	Airenne *et al.* (1997)
Hu MMP9	AcNPV/High Five	Polyhedrin	>300	8 l bioreactors	George *et al.* (1997)
Hu α_1m	AcNPV/High Five	Polyhedrin	50–150	n.r.	Wester *et al.* (1997)
Mn CatK	AcNPV/Sf21	Polyhedrin	7–10	20 ml cultures	McQueney *et al.* (1998)
Hu STC	AcNPV/Sf9	Polyhedrin	125	60 l bioreactor	Zhang *et al.* (1998)
Hu GL	AcNPV/High Five	Polyhedrin	5–12	4 l bioreactor	Canaan *et al.* (1998)
Hu Apo A-I mutants	AcNPV/Sf21	Polyhedrin	1–8	Spinner flask	Sviridov *et al.* (1999)
Hu OPG-Fc	AcNPV/High Five	Polyhedrin	15	Shake flask	Willard *et al.* (2000)
Hu prolactin	AcNPV/High Five	Polyhedrin	53–60	T-flask	Das *et al.* (2000)
Hu prolactin	AcNPV/Sf9	Polyhedrin	172	1 l bioreactor	Pereira *et al.* (2001)
In proPAP	AcNPV/Sf21	Polyhedrin	0.37	T-flask	Wang *et al.* (2001)
Hu MMP-9	AcNPV/Sf9	Polyhedrin	0.8	Spinner flask	Sadatmansoori *et al.* (2001)
Hu amylase	AcNPV/Sf9	Polyhedrin	6–10	Roller bottle	Ragunath *et al.* (2002)
In proPAP-2	AcNPV/Sf21	Polyhedrin	5.9	suspension	Ji *et al.* (2003)
Hu IL-3	AcNPV/High Five	Polyhedrin	1.5–3.0	2.5 l bioreactor	Ding *et al.* (2003)
Intracellular					
CAT	AcNPV/Sf21	Polyhedrin	>100	T-flask	Luckow and Summers (1988)
VP6	AcNPV/Sf9	Polyhedrin	350	4 l bioreactor	Caron *et al.* (1990)
β-gal	AcNPV/High Five	Polyhedrin	>500	Shake flask	Wickham *et al.* (1992)
VP4	AcNPV/Sf9	Polyhedrin	50	Spinner flasks	Juarbe-Osorio (1993)
CAT	BmNPV/Bm5	Polyhedrin	250	1.5 l bioreactor	Zhang *et al.* (1993)
Hu ACL	AcNPV/Sf9	Polyhedrin	200	Spinner flasks	Lord *et al.* (1997)
NF6B1 and RelA	AcNPV/Sf9	Polyhedrin	>10	Spinner flasks	Coleman *et al.* (1997)
Hu aminoacylase I	AcNPV/Sf21	Polyhedrin	100–200	n.r.	Pittelkow *et al.* (1998)
Hu M/NEI	AcNPV/Sf9	Polyhedrin	20	15 l bioreactor	Cooley *et al.* (1998)
Hu Syk	AcNPV/Sf21	Polyhedrin	30–40	n.r.	Baldock *et al.* (2000)
EBNA1	AcNPV/Sf9	Polyhedrin	30	Spinner flask	Meij *et al.* (2000)
Hu PARP	AcNPV/Sf9	Polyhedrin	8–80	1 l bioreactor	Pereira *et al.* (2001)
KSHV Pol	AcNPV/Sf9	Polyhedrin	5–15	Spinner flasks	Dorjsuren *et al.* (2003)

Most reports are obtained from a survey of the Protein Expression and Purification journal (n.r., not reported; Hu, human; Ms, mouse; In, insect; Chk, chicken; Mn, monkey; tPA, tissue plasminogen activator; SEAP, placental secreted alkaline phosphatase; PCE, pancreatic cholesterol esterase; IL-8, interleukin-8; LTA_4H, leukotriene A_4 hydrolase; C9, complement protein 9; HIV gp120,

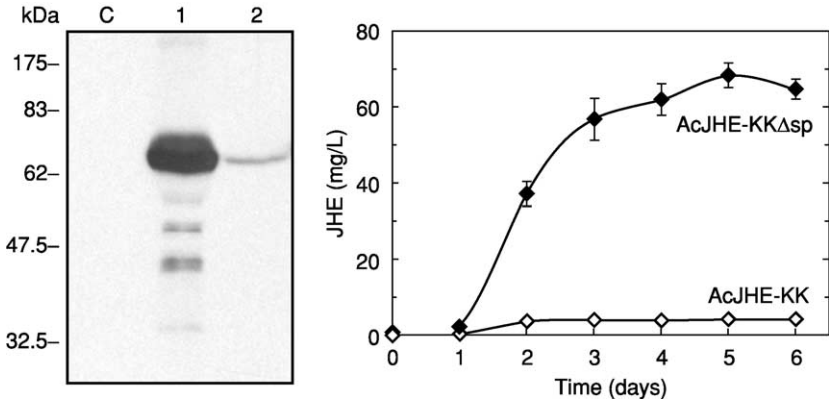


Figure 2 One technique to improve the expression level of secreted proteins from baculovirus infected insect cells involves the removal of the signal peptide encoded by the gene and the creation of an artificial ATG at the start of the mature polypeptide coding region in the expression vector. This modification prevents the nascent polypeptide from entering the secretory pathway. In this example, the yield of biologically active juvenile hormone esterase (JHE) present in the culture supernatant of recombinant AcNPV-infected Sf21 cells is enhanced approximately 20-fold when the JHE signal peptide is removed (AcJHE-KKΔsp) compared to the full length gene AcJHE-KK (Bonning *et al.*, 1997). Left panel: Western blot analysis of 5 μl culture media to confirm the relative JHE expression levels in static cultures. Lane C, control culture medium; lane 1, day 6 cultutre medium form AcJHE-KKΔsp infected Sf21 cells; lane 2, day 6 culture medium from AcJHE-KK infected Sf21 cells. Right panel: Batch production of JHE by AcJHE-KK or AcJHE-KKΔsp-infected Sf21 cells in serum-containing medium in 6-well plates over a 6-day period. Note that JHE produced by AcJHE-KKΔsp was not directed through the secretory pathway.

used membrane preparations from infected cells rather than the cells themselves to carry out their studies (Clawges *et al.*, 1997). In addition, coupling of heterologous receptors with endogenous insect cell host proteins, such as G proteins, has generally produced low responses upon ligand binding (see below). This problem has been circumvented by coexpressing a GPCR with specific G protein subunits in order to assess agonist binding (Butkerait *et al.*, 1995; Barr *et al.*, 1997; Bouvier *et al.*, 1998).

4.14.2.6.4. Compartmentalized proteins The BES has also been employed to synthesize and correctly process proteins targeted to subcellular compartments. A variety of mitochondrial proteins have been expressed using the BES, both membrane associated (Yet *et al.*, 1995; Bader *et al.*, 1998; Huang *et al.*, 2000) and located within the mitochondrial matrix (Wang and Kaguni, 1999; Holcomb *et al.*, 2000), although the latter reported incomplete processing and inefficient targeting. Lysosomal enzymes have also been successfully expressed (Tschantz *et al.*, 1999; Bromme and McGrath, 1996; Steed *et al.*, 1998), as has a nucleolar protein (Ren *et al.*, 1996) and a nuclear membrane protein (Bailer *et al.*, 1995).

4.14.2.7. Other Applications of the Baculovirus Expression System

4.14.2.7.1. Virus-like particles The baculovirus expression system has proven to be very effective for the production of heterologous virus-like and core-like particles (VLPs and CLPs) destined for use in immunization or structural studies. Some examples include the poliovirus VLP (Urakawa *et al.*, 1989), bluetongue virus CLP (French and Roy, 1990), Norwalk virus (Jiang *et al.*, 1992), the papillomavirus (Cann *et al.*, 1995), and herpes simplex virus (Newcomb *et al.*, 1999). VLPs and CLPs can be architecturally complex and comprising several structural proteins in different molar

human immunodeficiency virus glycoprotein 120; PSA, prostrate specific antigen; JHE, *Heliothis virescens* juvenile hormone esterase; IL-5, interleukin-5; OSF-2, murine osteoblast specific factor 2; HCII, heparin cofactor II; GalNAc-transferase, bovine UDP-GalNAc polypeptide *N*-acetyl galactosaminyl transferase; LIF, leukemia inhibitory factor; Apo A1, apolipoprotein A-I; FAK, focal adhesion kinase; GHBP, growth hormone binding protein; MMP9 matrix metalloproteinase-9; α_1m–α_1, microglobulin; CatK, cathepsin K; STC, stanniocalcin; GL, gastric lipase; OPG-Fc, osteoprotegerin-immunoglobulin Fc fusion; proPAP, *Manduca sexta* prophenoloxidase-activating proteinase precursor; proPAP-2, *M. sexta* prophenoloxidase-activating proteinase-2 precursor; IL-3, interleukin-3; CAT, bacterial chloramphenicol acetyl-transferase; VP6, rotavirus VP6; β-gal, beta galactosidase; VP4, rotavirus outer capsid protein; ACL, ATP citrate lyase; NFκB1/RelA, subunits of the NF-κB transcriptional activator of immunoglobulin κ light chain; M/NEI, monocyte/neutrophil elastase inhibitor; Syk, Syk protein; EBNAI, Epstein-Barr virus nuclear antigen 1; PARP, poly(ADP-ribose) polymerase; KSHV Pol, Kaposi's sarcoma-associated herpesvirus DNA polymerase).

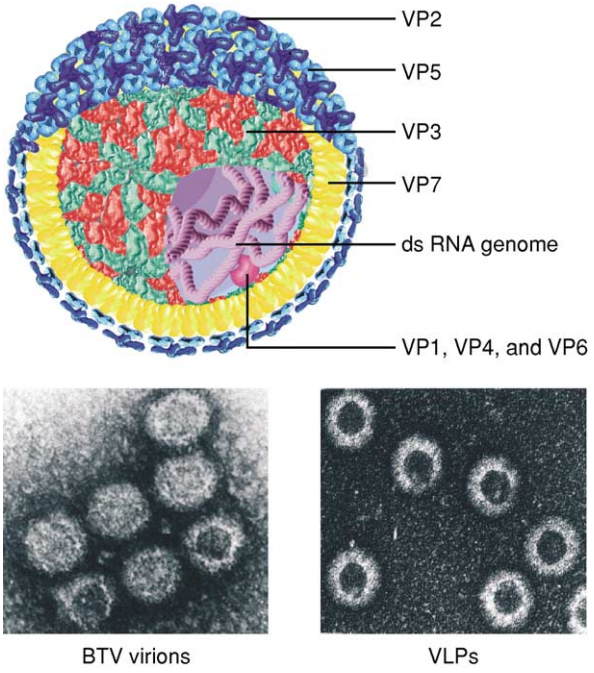


Figure 3 A comparison of electron micrographs of native blue-tongue virus (BTV) virions with recombinant BTV virus-like particles (VLPs) produced and self-assembled in baculovirus-infected insect cells. For optimal synthesis of the BTV VLPs, a quadruple gene expression vector was used to co-express BTV VP2, VP3, VP5, and VP7 proteins (Roy *et al.*, 1997). At the top of the figure, a model of the BTV virion is shown. Images kindly supplied by Dr Polly Roy (London School of Hygiene and Tropical Medicine, UK).

proportions. It was observed that the coexpression of the various VLP structural proteins in baculovirus-infected insect cells results in the self-assembly of the VLP (Urakawa *et al.*, 1989). Initially, the various structural proteins were produced by coinfection with individual recombinant baculoviruses. However, variation in the distribution of individual recombinant baculoviruses to each cell would result in differences in the quality of the VLP. Coexpression of proteins of the VLP or CLP using a single baculovirus with multiple promoters circumvents this problem (**Figure 3**; French and Roy, 1990; Belyaev *et al.*, 1995). The ratio of each protein expressed can be controlled through the use of baculovirus promoters of different strength (Roy *et al.*, 1997).

4.14.2.7.2. Surface display and antibody production The baculovirus virion has been developed as an eukaryotic alternative to the bacteriophage for the surface display of foreign proteins from an expression library and the selection of specific binding proteins (review: Grabherr *et al.*, 2001). By duplicating the essential major envelope glycoprotein gene encoding gp64 in the baculovirus genome, it was shown that foreign proteins (X) could be presented on the mature N-terminus of the duplicated X-gp64 (fusion) protein, in the envelope of BVs (Boublik *et al.*, 1995). A demonstration of the feasibility of using baculovirus display for the screening of a library has been described (Ernst *et al.*, 1998). Immunization of mice with purified virions displaying X-gp64 fusion proteins has also been useful for obtaining polyclonal and monoclonal antibodies against the antigen X (Lindley *et al.*, 2000). This procedure eliminates one bottleneck in the conventional antibody production process of having to generate purified protein X for immunization. Due to an immunostimulatory effect often associated with viral immunogens (Minev *et al.*, 1999), this method may be more effective than the conventional process for raising antibodies against certain antigens (Lindley *et al.*, 2000).

4.14.2.7.3. Expression of recombinant proteins in insect larvae Insect larvae have occasionally been employed, instead of cultured insect cells, as hosts for recombinant baculoviruses for the expression of recombinant proteins (Maeda *et al.*, 1985). Production in larvae can be inexpensive, high yielding, and easily scalable compared to cell culture. Yields around 1 mg per silkworm larva infected with recombinant BmNPV have been obtained for human interleukin-3 (Miyajima *et al.*, 1987), hepatitis B surface antigen (Higashihashi *et al.*, 1991), human CD66 antigens (Yamanaka *et al.*, 1996), grass carp growth hormone (Ho *et al.*, 1998), hepatitis E virus capsid protein (Sehgal *et al.*, 2003), and human fibroblast growth factor (Wu *et al.*, 2001). A bovine cardiac sodium–calcium exchanger has also been successfully expressed in *T. ni* larvae (Hale *et al.*, 1999). However, the recovery of the expressed protein may be more problematic from larvae than from tissue culture because of increased proteolysis in the insect hemolymph and contamination of preparations by abundant hemolymph proteins.

4.14.2.7.4. Baculovirus transducing and transforming vectors Because of their ability to enter efficiently host and nonhost insect cells alike, baculovirus vectors can be designed with the goal of overexpressing regulatory factors and studying their role during development. Because baculoviruses are normally lethal to the host tissues, expression studies are by necessity limited to a brief period of experimentation extended between the initial infection and the activation of the host cell lysis process (Iatrou and Meidinger, 1989). For lepidopteran insect systems, this problem can be

circumvented by the construction and use of mutant viruses that are unable to proceed through the late phases of the infection cycle. By contrast, infection of tissues of nonhost insects that are refractory to productive infection can result in more extended periods of expression and experimental observation (Oppenheimer *et al*., 1999).

Two approaches have been utilized for the generation of transgenic lepidopteran insects using baculoviruses. One exploits the fact that AcNPV can infect nonhost lepidopteran insect larvae without killing the insect (Mori *et al*., 1995; Oppenheimer *et al*., 1999). Due to this property, the AcNPV can be used as a vehicle to efficiently deliver a transgene to germline cells, which is then targeted to chromosome sites for homologous recombination and generation of transgenic progeny (Yamao *et al*., 1999). A second approach uses a recombinant baculovirus with a deleted *lef-8* gene (Iatrou *et al*., 2000). The *lef-8* gene encodes a subunit of the viral RNA polymerase (Passarelli *et al*., 1994) that is necessary for mRNA transcription from the promoters of late and very late-phase baculovirus genes. Loss-of-function mutations in this gene prevent the progression from the replication phase to the virulent phase of the baculovirus infection cycle (Shikata *et al*., 1998). Accordingly, deletions of the *lef-8* gene convert the baculovirus to a harmless, self-replicating extra-chromosomal entity (baculovirus artificial chromosome – BVAC), while host cells infected by it behave as normal ones (P.J. Farrell and K. Iatrou, unpublished results). Passage of the *lef-8* deficient baculovirus through a rescuing cell line that constitutively expresses the Lef-8 protein enables the production of BVAC inoculum that maintains full infectivity for and replicating capacity in the host cells (**Figure 4**).

The BVACs could potentially be used for prolonged protein expression by taking advantage of the high initial infection efficiency of cells by the baculovirus, the replication of the BVACs in them, and the subsequent mitotic transmission of the transgene-containing BVAC to the daughter cells. Furthermore, the prevention of both virus-induced damage to the secretory pathway and lysis would make BVACs more suitable for the expression of secreted and membrane proteins under the control of early viral or cellular promoters than conventional baculovirus expression vectors. Efforts are also being made to generate transgenic silkworms through the infection of germline cells with BVACs derived from BmNPV.

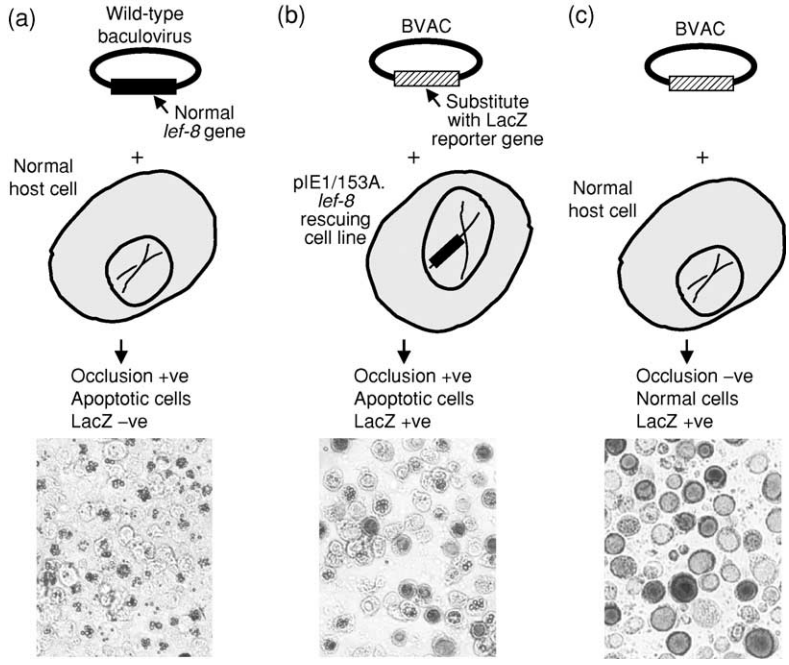


Figure 4 The potential for a nonlytic baculovirus is demonstrated. (a) Infection of normal insect cells by a wild-type baculovirus. The infected cells produce OBs (small, refractive objects) and eventually lyse. (b) A baculovirus artificial chromosome (BVAC) is created when the *lef-8* gene is eliminated and replaced with a reporter gene expression β-galactosidase (dark cells). When the *lef-8* gene product is supplied by a transformed cell line, the BVAC can complete the infection cycle that includes the formation of budded virions and occlusion bodies and cell lysis, in addition to the production of β-galactosidase. (c) The absence of the *lef-8* gene product, in normal permissive cells infected by a BVAC, prevents completion of the baculovirus infection cycle. OBs are absent, the cells appear and replicate normally, and also express β-galactosidase.

4.14.2.7.4. Transduction of mammalian cells Baculovirus virions were demonstrated to have the ability to enter certain cell lines derived from vertebrate species without evidence of viral gene expression (Volkman and Goldsmith, 1983; Carbonell and Miller, 1987). By incorporating mammalian expression cassettes into recombinant AcNPVs, several reports have appeared demonstrating that baculoviruses can serve as an efficient mode of gene transfer into a variety of primary and transformed mammalian cell lines, including those difficult to transfect using traditional methods, and express reporter proteins under the control of mammalian promoters (Hoffmann *et al*., 1995; Boyce and Bucher, 1996; Shoji *et al*., 1997; Yap *et al*., 1997; Condreay *et al*., 1999). Chromosomal integration and stable gene expression have also been demonstrated to be feasible by inclusion of a selectable marker (Condreay *et al*., 1999; Merrihew *et al*., 2001) or by integration signals from adeno-associated virus (AAV) into the baculovirus genome (Palombo *et al*., 1998). The advantages of using the baculovirus in this application include the high transfection efficiency (often greater than 90%; Condreay *et al*., 1999), low toxicity of the baculovirus to mammalian cells, and capacity of the baculovirus to carry large inserts. Recently, hybrids of the baculovirus and AAV were used to coinfect mammalian cells and produce high titer recombinant AAV for gene therapy applications (Sollerbrant *et al*., 2001).

4.14.3. Insect Cells as Hosts for Plasmid-Based Expression of Recombinant Proteins

4.14.3.1. Introduction

Several baculovirus-free, plasmid-based approaches to recombinant protein expression in insect cells have been developed. In this technique, a plasmid expression cassette harboring a gene of interest is introduced into the host insect cell line using a variety of transfection techniques. The recombinant protein will be transiently expressed for a few days after transfection whereupon the cells or their media containing the recombinant protein can be harvested. Stable cell lines expressing the recombinant protein of interest, either continuously or by induction, can be generated by applying an antibiotic resistance-selection scheme over a period of several weeks. The antibiotic resistance gene can either be present on the same expression cassette as the gene of interest or supplied by cotransfection with a separate plasmid. Antibiotic selection schemes conferring resistance to G418 (Jarvis *et al*., 1990), hygromycin B (Johansen *et al*., 1989), methotrexate (Shotkoski and Fallon, 1993), puromycin (McLachlin and Miller, 1997), and zeocin (Pfeifer *et al*., 1997) have all been used extensively with insect cell lines.

The production of recombinant proteins in transfected or stably transformed insect cells provides considerable advantages over both the baculovirus expression system and transformed mammalian cells for the case of applications involving secreted and membrane-anchored proteins: insect cell lines are safe to humans, introns are spliced correctly and efficiently from expressed genomic DNAs and lysis does not occur, therefore allowing continuous, as opposed to batch-type, expression of the proteins and limiting proteolysis and facilitating purification of the secreted proteins. Furthermore, insect cells can perform most essential posttranslational modifications as efficiently as mammalian ones, and membrane proteins can be expressed in a stable physiological environment. Finally, most insect cell lines can grow in serum-free media to high densities, and many insect cell lines, particularly of lepidopteran origin, are already well characterized in large-scale suspension culture because of their role as hosts in the baculovirus expression system.

Two major categories of plasmid-based expression systems have been developed utilizing promoters and cell lines derived from two different insect orders, Diptera and Lepidoptera. These are presented in more detail below.

4.14.3.2. Expression Systems Based on Dipteran Cell Lines

Dipteran expression vectors mostly utilize the strong constitutive actin 5C promoter of *D. melanogaster* (Angelichio *et al*., 1991) or the inducible metallothionein promoter of *Drosophila* (Johansen *et al*., 1989; Kovach *et al*., 1992; Millar *et al*., 1995; Hegedus *et al*., 1998; Zhang *et al*., 2001) to drive foreign protein expression, in conjunction to a hygromycin B antibiotic selection scheme to generate a polyclonal population of Schneider 2 cell lines (S2 cells; Schneider, 1972), which takes approximately 3 weeks to accomplish (Johansen *et al*., 1989). The latter system has been developed by SmithKline-Beecham Pharmaceuticals and made commercially available through Invitrogen Corporation as the Drosophila Expression System (DES®). *Drosophila* S2 cells can grow in serum-free medium in batch suspension culture to cell densities up to 15×10^6 cells ml^{-1}, although S2 cells are substantially smaller than most lepidopteran cell lines including Sf21, High Five™, and Bm5

Table 2 Expression levels of some secreted proteins from stably transformed insect cells

Protein	*Cells*	*Level (mg l^{-1})*	*Culture conditions*	*Reference*
D. melanogaster metallothionein promoter				
Hu IL-5	S2	22	n.r.	Johansen *et al.* (1995)
Modified HIV gp120	S2	2	n.r.	Culp *et al.* (1991)
Modified HIV gp120	S2	5–35	n.r.	Ivey-Hoyle *et al.* (1991)
Hu IgG_1	S2	1	n.r.	Kirkpatrick *et al.* (1995)
Hu SPC1	S2	3	n.r.	Denault *et al.* (2000)
Hu EPO	S2	2	Spinner flasks	Shin and Cha (2003)
AcNPV ie-1 promoter				
Hu tPA	Sf9	1.0	Static	Jarvis *et al.* (1990)
OpMNPV ie-2 promoter				
Ms IgG_1	Sf9	0.5–1	Spinner flasks	Li *et al.* (2001)
Modified Hu p97	Sf9	10	Spinner flasks	Hegedus *et al.* (1999)
Modified Hu Factor X	Sf9	18	Shake flasks	Pfeifer *et al.* (2001)
Hu plasminogen	S2	10–15	Spinner flasks	Nilsen and Castellino (1999)
B. mori cytoplasmic actin promoter, pIE1/153A vector				
Hu tPA	Bm5	135–160	Static-spinner flask	Farrell *et al.* (1999)
In JHE	Bm5	130–190	Static-spinner flask	Farrell *et al.* (1998)
Hu GM-CSF	High Five	27–46	Spinner flask-static	Keith *et al.* (1999)

n.r., not reported; Hu, human; Ms, mouse; In, insect; IL-5, Interleukin-5; HIV gp120, human immunodeficiency virus glycoprotein 120; IgG_1, immunoglobulin G_1; SPC1, subtilisin-like proprotein convertase 1; EPO, erythropoietin; tPA, tissue plasminogen activator; p97, melanotransferrin; JHE, *Heliothis virescens* juvenile hormone esterase; GM-CSF, granulocyte macrophage colony stimulating factor.

cells. Enzymes, membrane receptors, ion channels, viral antigens, and monoclonal antibodies have been successfully produced using these systems (see Table 2).

4.14.3.3. Expression Systems Based on Lepidopteran Cell Lines

Lepidopteran expression vectors utilize a variety of promoters derived from baculoviruses and lepidopteran cells, including the AcNPV *ie-1* promoter, the OpNPV *ie-2* promoter, and the *B. mori actin* gene promoter; these promoters function mainly in lepidopteran cell lines such as Sf21, Sf9, High Five™, or Bm5 cells.

4.14.3.3.1. Constitutive expression

4.14.3.3.1.1. Baculovirus immediate early promoters A lepidopteran insect cell-based expression system was initially developed by Dr Don Jarvis at Texas A&M University (Jarvis *et al.*, 1990). To generate stably transformed cell lines, Sf9 cells were cotransfected with a neomycin resistance plasmid, an expression vector employing the AcNPV immediate early gene promoter (*ie-1*) and a region containing mRNA polyadenylation signals (Guarino and Summers, 1987), followed by selection and isolation of G418 resistant clones over a period of 4 weeks following transfection. From the transformed insect cell clones, the expression levels of β-galactosidase (an intracellular protein) were approximately 100-fold lower than those obtained from a baculovirus expression vector employing the polyhedrin promoter, while the level of t-PA (a secreted protein) obtained was only two-fold lower. Despite the low values, these experiments demonstrated the potential of using stable lepidopteran cell lines for producing those recombinant proteins directed through the secretory pathway as an alternative to a baculovirus. The *ie-1* gene promoter was also shown to function in dipteran cell lines (Vanden Broeck *et al.*, 1995). Improvements in the expression vector were made by incorporating the AcNPV homologous repeat 5 (HR5) transcriptional enhancer element (Guarino *et al.*, 1986) upstream of the *ie-1* promoter (Jarvis *et al.*, 1996). HR5 was shown to act in *cis* to significantly stimulate the expression of a reporter protein from early baculovirus promoters such as *ie-1* and *p35* in transient expression assays (Rodems and Friesen, 1993; Pullen and Friesen, 1995). A series of vectors utilizing the HR5-enhanced AcNPV *ie-1* promoter are available from Novagen for stable or transient protein expression. A similar set of expression vectors has been developed based on the *Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus (OpMNPV) *ie-2* promoter (Theilmann and Stewart, 1992; Pfeifer *et al.*, 1997; Hegedus *et al.*, 1998), which can function in both lepidopteran and dipteran insect cell lines. These vectors are available from Invitrogen as the Insect Select™ vector set.

4.14.3.3.1.2. **Bombyx mori** ***cytoplasmic actin promoter*** A powerful lepidopteran expression cassette, pIE1/153A, was recently developed in the laboratory of Dr Kostas Iatrou at the University of Calgary. The cassette utilizes the silk moth (*B. mori*) cytoplasmic actin gene promoter (Mounier and Prudhomme, 1986; Johnson *et al.*, 1992), one of the stronger constitutive cellular promoters that is also active in a variety of transfected lepidopteran cell lines. Transcription from this cellular promoter was surprisingly found to be stimulated by the immediate early gene product (*ie-1*) of BmNPV, IE-1, a transcription factor capable of stimulating the *in vitro* rate of transcription from the actin promoter in *trans* by up to 100-fold (Lu *et al.*, 1996). Stimulation of the actin promoter by up to two orders of magnitude was also obtained by linking in *cis* the homologous repeat 3 (HR3) region of BmNPV in various orientations relative to the actin promoter (Lu *et al.*, 1997). Linkage of the *ie-1* gene and the HR3 enhancer element with the actin gene promoter in the pIE1/153A expression cassette resulted in a stimulation of foreign gene expression directed by the actin promoter by approximately 5000-fold in transient expression assays for two reporter proteins (Lu *et al.*, 1997). Stable cell lines have been generated by cotransfecting the cells with the expression cassette and a second plasmid conferring resistance to hygromycin B, puromycin, or G418. Reported expression levels of secreted proteins from stable cell lines transformed with this system have exceeded, by far, those obtained by the baculovirus expression system (Farrell *et al.*, 1998; Farrell *et al.*, 1999). Furthermore the expression plasmid functions in a wide variety of lepidopteran cell lines including those derived from *B. mori*, *T. ni*, *Plodia interpuntella*, *L. dispar*, *Mamestra brassicae*, *C. fumiferana*, and *S. frugiperda* (Keith *et al.*, 1999). The expression cassette is also suited to scaled-up transient expression protocols that can yield tens of milligrams of recombinant protein per liter in 5 days posttransfection (Farrel and Iatrou, 2004) (**Figure 5**), thus avoiding the time consuming and labour intensive process of stable antibiotic selection and the cloning of high expressors. Derivatives of the pIE1/153A expression cassette are available from Dr Iatrou.

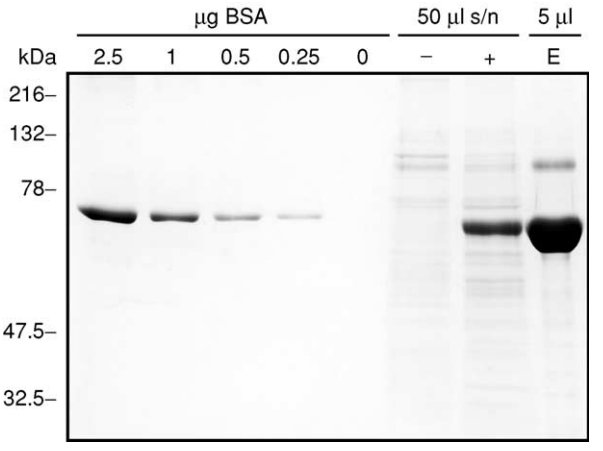


Figure 5 Transient expression and affinity purification of histidine-tagged human secreted alkaline phosphatase (SEAP; kindly provided by Eric Carpentier and Amine Kamen, Biotechnology Research Institute, Montreal, Canada) from a lipofectin-mediated transfection of High Five™ cells in suspension culture using the pIE1/153A expression cassette. By comparison to a series of bovine serum albumin mass standards on the Coomassie blue-stained gel, approximately 50 µg ml^{-1} of SEAP was present in the culture supernatant 5 days after transfection and production in protein-free medium ESF-921 (lane marked "+" compared to control marked "−"). Two milligram of SEAP was recovered from one elution fraction (E) following Ni-ion affinity purification from 50 ml of culture supernatant.

4.14.3.4. Insect Cell-Based Expression Using Inducible Promoter Systems

In contrast to constitutive promoters, the use of inducible promoters has the advantage that it allows the production of proteins that are toxic or growth-inhibitory to the cells. In such cases, cells are first grown in batch mode (rather than continuously) to high densities. When sufficiently high densities are reached, the inducer is added to the cell culture to induce high-level protein expression. Of importance to the choice of an inducible expression system is the induction ratio of gene expression. In the absence of inducer, promoter activity should be minimal as to prevent accumulation of toxic proteins during the growth phase of the cells. Addition of the inducer should, subsequently, result in the accumulation of high amounts of protein.

4.14.3.4.1. Heat shock promoter The *Drosophila hsp70* promoter (Thummel and Pirrotta, 1992) has been widely used to achieve inducible gene expression in transgenic flies. The promoter is heat-inducible and functions during all stages of development and in all tissues (Steller and Pirrotta, 1984; Spradling, 1986). It has a low basal expression coupled with high inducibility (>100-fold at 42 °C) (Huynh and Zieler, 1999). The promoter is not only functional in *Drosophila* cells but also in mosquito and lepidopteran cell lines (Zhao and Eggleston, 1999; Helgen and Fallon, 1990; Lan and Riddiford, 1997).

While the *hsp70* promoter has been used to mediate expression of membrane proteins (Shotkoski *et al.*, 1996) and to drive expression of "helper" antibiotic resistance genes for the establishment of

permanently transformed cell lines (Lycett and Crampton, 1993; Vanden Broeck *et al*., 1995) and transgenic mosquitoes (Miller *et al*., 1987), its application for the inducible expression of proteins at high levels for functional or structural characterization is considered limited. Long-term exposure to high temperature, which is considered necessary to reach the highest expression levels, is generally detrimental to the cell cultures. Also the translational induction to heat shock is reported to be much lower than the transcriptional induction (Cherbas *et al*., 1994), presumably because the heat shock interferes with posttranscriptional processing and translation.

Heat shock (*hsp70*) promoter-based reporter and expression constructs were also incorporated in baculovirus vectors to monitor infection of tissues and cells and to achieve protein production. In the context of the baculovirus genome in infected cells, however, gene expression from the *hsp70* promoter was constitutive rather than heat-inducible (Lee *et al*., 2000; Moto *et al*., 2003).

4.14.3.4.2. Metallothionein promoter Transcription of mammalian and *Drosophila* metallothionein genes is activated by heavy metal load via activation of the metal-responsive factor 1 (MTF-1) (Zhang *et al*., 2001). Inducible expression of genes via the metallothionein promoter is achieved in *Drosophila* S2 cells and other dipteran cell lines (Johansen *et al*., 1989; Kovach *et al*., 1992; Millar *et al*., 1995; Hegedus *et al*., 1998). Protein expression via the metallothionein promoter is induced by the addition of cadmium or copper ions (10–1000 μM) and induction levels of several 100-fold are achieved (Bunch *et al*., 1988; Otto *et al*., 1987; Zhang *et al*., 2001). The metallothionein promoter-based inducible system is considered suitable for the inducible expression of many different classes of proteins in dipteran cell lines.

A disadvantage of the metal ion-inducible expression system is that at the high metal ion concentrations required for achieving the highest expression levels, toxic effects are also observed on the cells, such as cytoplasmic vacuolization and granule formation as well as an associated reduction in their growth rate (Bunch *et al*., 1988). Thus, optimal protein production requires determination of the right balance between degree of induction of gene expression and general cellular toxicity.

While the metal ion-mediated induction works efficiently in *Drosophila* and mammalian cell lines (Zhang *et al*., 2001), in lepidopteran cell lines such as Sf9, Ld652Y, and Bm5, the metallothionein promoter was either not or only marginally functional (Hegedus *et al*., 1998; V. Douris, L. Swevers and K. Iatrou, unpublished observations).

4.14.3.4.3. Ecdysone-inducible promoters Many insect tissue culture cell lines are sensitive to the insect molting hormone 20-hydroxyecdysone (20E) and respond to it by morphological changes and a decline in growth rates (Sorbrier *et al*., 1989; Sohi *et al*., 1995; Smagghe *et al*., 2003). At the molecular level, 20E exerts its actions by binding to the ecdysone nuclear receptor (EcR) that exists as a heterodimer with the nuclear receptor Ultraspiracle (USP) (Thomas *et al*., 1993; Yao *et al*., 1993; Swevers *et al*., 1996) and the hormone-bound complex directly activates transcription at ecdysone-response elements (EREs) in promoters of target genes (Antoniewski *et al*., 1996; Vogtli *et al*., 1998; Wang *et al*., 1998; see **Chapter 3.5**). Early-response genes, which are directly regulated by the EcR/USP complex, are induced in insect cell lines after challenge with 20E (Sohi *et al*., 1995; Chen *et al*., 2002; Swevers *et al*., 2003b), indicating that all transduction elements for gene activation by 20E are present.

Basal reporter cassettes that contain multiple repeats of an ERE derived from the *Drosophila hsp27* promoter are 1000- to 2000-fold induced by μM quantities of 20E in *Drosophila* S2 cells as well as *Bombyx*-derived (lepidopteran) Bm5 cells (Koelle *et al*., 1991; Swevers *et al*., 2003b). The induced expression levels are comparable to those obtained by strong constitutive promoters (Swevers *et al*., 2003b) and silk moth (lepidopteran) transformed cell lines incorporating an ecdysteroid-inducible expression system have been described for production of recombinant proteins (Tomita *et al*., 2001) as well as high-throughput screening for potential 20E agonists and antagonists (Swevers *et al*., 2003b).

An advantage of the ecdysteroid-inducible expression system is that, in the absence of ligand, the EcR/USP heterodimer functions as a repressor of gene transcription (Tsai *et al*., 1999) and the basal expression levels in the absence of the inducer are very low. Because 20E is a natural hormone, toxicity effects on the cell cultures are virtually nonexistent. Nevertheless, it remains to be investigated in detail whether the physiological effects of 20E on the function of the cells could potentially interfere with their capacity for protein production. It is expected that this will not be the case: using the baculovirus/lepidopteran cell line expression system, it was reported that the addition of ecdysteroids actually increased recombinant protein production (Sarvari *et al*., 1990).

4.14.3.5. Classes of Recombinant Proteins Expressed by Transformed Insect Cells

4.14.3.5.1. Secreted proteins A variety of heterologous secreted proteins have been expressed from stable insect cell lines at similar or higher expression levels than can be obtained using baculovirus expression vectors (**Table 2**). Certainly the cellular or early-phase baculovirus promoters employed are not as transcriptionally powerful as late- and very late phase promoters such as those of the polyhedrin, p10 or basic protein genes. However, the fact that the cell's secretory pathway is not being damaged in the absence of viral infection, contributes to the improvement in the expression level and quality of the overexpressed protein compared to expression using recombinant baculovirus. The substitution of the native mammalian signal peptide coding sequence in a heterologous mammalian gene with a sequence encoding an insect-specific signal peptide was found to have no effect on the resulting expression level in transfected insect cells (Farrell *et al.*, 2000).

4.14.3.5.2. Membrane proteins As with secreted proteins, the absence of viral infection may provide a superior cellular environment for the production of membrane proteins which also traverse the secretory pathway. Several functional membrane proteins have been successfully expressed in stable insect cell lines including ion exchangers (Szerenscei *et al.*, 2000), transmitter gated ion channels (Joyce *et al.*, 1993; Millar *et al.*, 1994; Smith *et al.*, 1995; Atkinson *et al.*, 1996; Buckingham *et al.*, 1996), and GPCRs (Kleymann *et al.*, 1993; Swevers *et al.*, 2003a; see **Chapter 5.5**). As found with the baculovirus expression system (Bouvier *et al.*, 1998), coupling of heterologous GPCRs with endogenous insect G proteins may be inefficient following ligand binding, even in transformed insect cells (**Figure 6**). However, a strategy that has proven useful for the functional expression of heterologous GPCRs is the coexpression of mammalian G proteins in stably transformed insect cells (L. Swevers, K. Iatrou, E. Morou, N. Balatsos, and Z. Georgoussi, unpublished data; see also further below).

4.14.3.5.3. Intracellular proteins The use of transformed insect cells for the expression of intracellular proteins is of limited potential compared to the baculovirus expression system. There are a few reports on the concentrations of intracellular proteins obtained, however the promoters employed in stable gene expression from insect cells are considerably weaker than those utilized in baculovirus

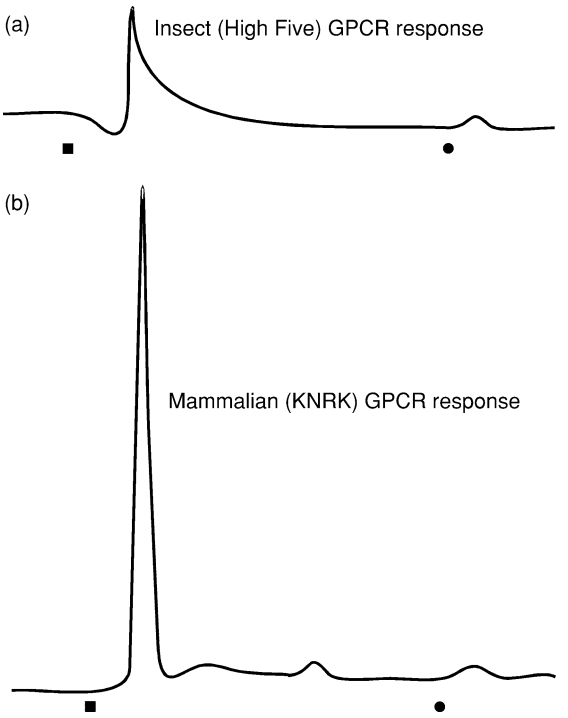


Figure 6 Fluorescent spectrophotometer outputs of calcium signaling assays to compare the G protein-coupled receptor (GPCR) response of a mammalian GPCR, the rat protease activated receptor 2 (rPAR2; Hollenberg *et al.*, 1996), expressed in stably transformed insect (panel a, High Five™) and mammalian (panel b, KNRK) cell lines. The amplitude reflects elevations in intracellular calcium with time, following exposure to the rPAR-2 peptide agonist, SLIGRL, at the time points denoted by the filled square and circle. Despite the fact that the receptor density was higher in the insect cell line (data not shown), the results imply that this mammalian GPCR does not couple well with the endogenous insect G proteins. Courtesy of Dr Morley Hollenberg, Mamoud Saifeddine and Bajhat Al-Ani (University of Calgary, Canada).

expression vectors (Jarvis *et al.*, 1990; Percival *et al.*, 1997; Pfeifer *et al.*, 1997). Furthermore, the purification of heterologous intracellular proteins, expressed at relatively low levels inside insect cells, from endogenous cellular protein mixtures is expected to be difficult. Secretion of the intracellular protein would facilitate the purification process. However, the mere fusion of an insect-specific signal peptide to the N-terminus of an intracellular protein failed to provide the means necessary to secrete intracellular proteins (Farrell *et al.*, 2000). It was realized that other biological signals in addition to a signal peptide are required for efficient secretion from insect and mammalian cells into an extracellular environment. These could be supplied by fusing the complete coding sequence of a secreted protein to that of an intracellular protein (**Figure 7**; Farrell *et al.*, 2000). The chimeric proteins produced are soluble and can be easily purified from

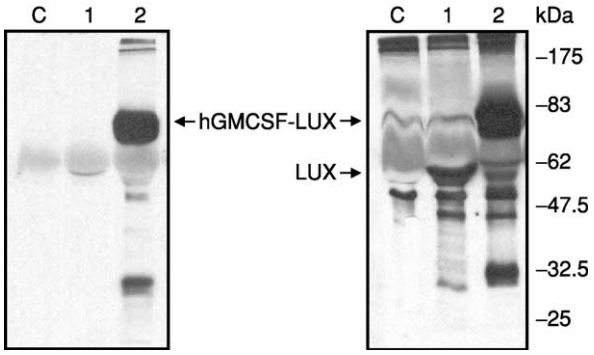


Figure 7 Secretion of luciferase from lepidopteran insect cells transfected with the pIE1/153A expression plasmid that encompasses a luciferase secretion module. Normally, luciferase (LUX) is expressed intracellularly, however, by fusing it with a secreted protein, in this case the human granulocyte macrophage colony stimulating factor (hGMCSF), the normally intracellular protein can be dragged into the supernatant as a chimera. High Five cell culture supernatants are shown in the SDS-PAGE/Western blots probed with (left) an anti-hGMCSF antibody and (right) an anti-LUX antibody. The lane marked ''C'' contains protein from a control transfection using the pIE1/153A expression vector without inserted transgene. Lanes 1 contain supernatants from cells transfected with vector pIE1/153.lux directing intracellular expression of luciferase, and Lanes 2 contain supernatants from cells transfected with vector pIE1/153A.hgmcsf-lux directing the secretion of the overexpressd hGMCSF-LUX fusion protein. Although some luciferase does leak from the cells naturally, a large proportion is secreted when expressed as a hGMCSF-LUX fusion protein.

the cell culture supernatant under nondenaturing conditions.

4.14.4. Insect Cell-Based High-Throughput Screening Systems

In cell-based detection systems, cell lines are engineered such that they respond to a biological stimulus by the generation of an easily detectable, fluorescent or luminescent, signal. Accordingly, cell-based detection systems integrate two relevant genetic elements, (1) an expression element that directs expression of the target molecule (e.g., a receptor) thus allowing the cells to respond specifically to bioactive compounds (e.g., ligands) that interact specifically with the target and initiate a downstream cell signaling response, and (2) a reporter element that allows the cells to generate an easily measurable response (fluorescent or luminescent signal) following activation of the expressed receptor by the bioactive compound.

Insect cell-based high throughput screening systems have been developed for two major types of receptors, nuclear hormone receptors (see **Chapters 3.4–3.6**) and membrane-anchored GPCRs (see **Chapter 5.5**). Because nuclear hormone receptors activate directly transcription at specific target sites in the DNA (Gronemeyer and Laudet, 1995), detection of nuclear receptor activation occurs by induction of expression of appropriate fluorescent (GFP or its variants) or luminescent (luciferase) reporter genes. On the other hand, GPCRs, which are subdivided into several classes according to the type of signal transduction pathway that they generate at the plasma membrane (cAMP up- or downregulation, release of Ca^{2+}, activation/inhibition of ion channels) (Hamm, 2001; Marinissen and Gudkind, 2001) require different reporter assays that are designed according to the specificity of each receptor class. Since the release of Ca^{2+} can be easily measured by fluorescent (e.g., employing the Ca^{2+}-sensitive fluorescent dyes fura-2 or fluo-3) or bioluminescent (coelenterazine/aequorin-based) methods (Grynkiewicz *et al.*, 1985; Knight *et al.*, 1991), activation of GPCRs that naturally function by Ca^{2+} release can be monitored by the detection of Ca^{2+} accumulation. For other GPCRs, the possibility exists to coexpress them with the mammalian $G\alpha_{16}$ protein, which can couple GPCRs of other transduction classes to the Ca^{2+} pathway (Kostenis, 2001).

4.14.4.1. Nuclear Receptors: The Ecdysone Receptor

Of the approximately 20 nuclear hormone receptors identified in the insect (*Drosophila* or *Anopheles*) genome (Adams *et al.*, 2000; Zdobnov *et al.*, 2002; see **Chapters 3.5** and **3.6**), only one has a clearly identified ligand. This is EcR, the receptor for the insect moulting hormone 20E (Escriva *et al.*, 1997). Because many insect cell lines express functional EcR receptors (Sohi *et al.*, 1995; Chen *et al.*, 2002; Swevers *et al.*, 2003a), they have the potential to be developed as screening systems for the 20E hormone (see **Chapter 3.4**).

A microplate-based bioassay, for the detection of 20E mimetic (agonistic and antagonistic) activities, was described that is based on changes in growth or morphology of the *Drosophila* B_{II} tumorous blood cell line (Clement *et al.*, 1993). This cell line is used to screen natural products from plants and fungi as well as collections of synthetic ecdysteroid derivatives for moulting and antimoulting hormone activity (Dinan *et al.*, 1997; Harmatha and Dinan, 1997; Dinan *et al.*, 2001; Harmatha *et al.*, 2002).

Another, more recently developed high throughput screening system for 20E agonists and antagonists, is based on silk moth-derived Bm5 cells that have been engineered to incorporate an ecdysteroid-responsive GFP reporter cassette in their genome

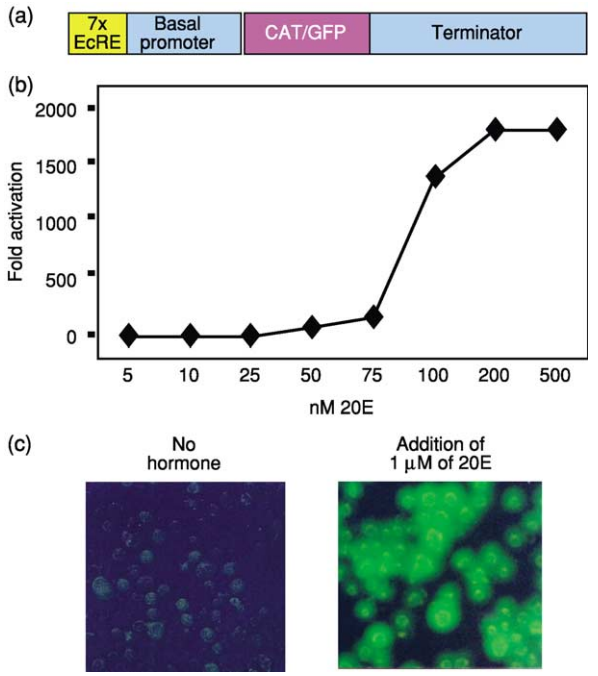


Figure 8 Establishment of a high-throughput screening system for ecdysone agonists and antagonists based on transformed silk moth-derived Bm5 cells. (a) Schematic presentation of the ecdysone-responsive reporter constructs. EcRE, ecdysone-responsive element; CAT, chloramphenicol acetyl transferase; GFP, green fluorescent protein. (b) Dose-response curve of the ecdysone-responsive CAT reporter construct in transient expression experiments. Shown is the fold inducibility in function at different hormone concentrations. (c) Induction of green fluorescence by 20E in a clonal transformed cell line that contains integrated copies of the ecdysone-responsive GFP reporter construct in their chromosomes.

(Swevers *et al.*, 2003b; **Figure 8**). The half-maximal response of activation of the reporter cassette is elicited by concentrations of 20E between 50 and 100 nM, which are comparable to the concentrations required for induction of morphological or growth inhibitory responses in other cell lines (Cherbas *et al.*, 1980; Dinan *et al.*, 1997). An important advantage of the ecdysteroid-responsive fluorescent cell lines is that the amount of fluorescence emitted by the cells can be easily detected and quantified from cells seeded in multiwell plates by a fluorescence plate reader, thus rendering the system amenable to high-throughput analysis. The ecdysteroid-inducible fluorescent Bm5 cell lines were used successfully to screen a collection of plant extracts as well as a chemical library of potential nonsteroidal agonists (dibenzoyl hydrazine derivatives) for detection of 20E agonist and antagonist activities (Swevers *et al.*, 2003b). The dibenzoyl hydrazine compounds selected by the high-throughput assay were also effective in larval toxicity tests, thus validating the usefulness of the fluorescent screening system for rapid selection of 20E mimics as potential new insecticides.

4.14.4.2. G Protein-Coupled Receptors

The majority of signal transduction by hormones or neurotransmitters is mediated by GPCRs (Vanden Broeck, 2001). In the two insects whose complete genome sequence is known, *D. melanogaster* and *Anopheles gambiae*, GPCRs comprise the largest number of genes, accounting for 1–2% of the total gene number (Hill *et al.*, 2002). Similarly, in the human genome more than 600 GPCR genes have been identified. Human GPCRs are considered important targets for drug development, and it is estimated that up to 40–50% of the available drugs are modulators of GPCR function (Kostenis, 2001).

Using the AcNPV-based baculovirus expression system, several types of mammalian GPCRs have been successfully expressed in *S. frugiperda* Sf9 and *T. ni* Hi5 cell lines (Vasudevan *et al.*, 1992; Ng *et al.*, 1995, 1997; Zhang *et al.*, 1995; Wehmeyer and Schulz, 1997; Ohtaki *et al.*, 1998). While the AcNPV-based expression system is not amenable to high throughput screening development because the cells are lysed at the end of the infection cycle, these studies nevertheless demonstrated that the receptors that were overexpressed were functional and could couple efficiently to downstream signaling elements such as ion channels and the adenylate cyclase system. Thus, using plasmid-based expression systems that do not result in cell lysis (Farrell *et al.*, 1998), transformed cell lines can be obtained that express constitutively functional GPCRs for screening of ligand mimetics.

Drosophila Schneider 2 cells have been used by several research groups to achieve permanent and functional expression of insect as well as mammalian GPCRs (Joyce *et al.*, 1993; Kleymann *et al.*, 1993; Millar *et al.*, 1995; Tota *et al.*, 1995; Vanden Broeck *et al.*, 1995; Buckingham *et al.*, 1996; Torfs *et al.*, 2000; Perret *et al.*, 2003). In the case of a tachykinin receptor from the stable fly *Stomoxys calcitrans*, activation resulted in the detection of Ca^{2+} release by fluorescent and luminescent methods, thus making the transformed cell line an effective system for the screening of tachykinin-like ligands (Torfs *et al.*, 2002).

In mammalian cell-based screening systems, the "promiscuous" Gα proteins, $G\alpha_{15}$ and $G\alpha_{16}$, have been used to couple the activation of any GPCR to the Ca^{2+} signaling pathway (Stables *et al.*, 1997; Kostenis, 2001). The presence of the $G\alpha_{16}$ protein results in the activation of phospholipase Cβ

(PLCβ) resulting in the production of the secondary messengers diacyl glycerol (DAG) and inositol (1,4,5)-triphosphate (IP_3). While DAG activates other downstream pathways through activation of protein kinase C, the action of IP_3 results in the release of Ca^{2+} from intracellular stores (Stables *et al.*, 1997), which can be subsequently detected easily by fluorescent or bioluminescent methods (Grynkiewicz *et al.*, 1985; Knight *et al.*, 1991). Cell lines overexpressing $G\alpha_{16}$ protein can, therefore, be used to screen for ligands of GPCRs that normally couple to other second messenger or to ion channels.

This strategy has been employed recently in the case of transformed silk moth Bm5 cells that overexpress the murine δ opioid receptor: coexpression of the human $G\alpha_{16}$ protein in these cells has resulted in the stimulation of the coupling of the mouse receptor to the PLCβ/IP_3 pathway, thus making the transformed cell line a very effective screening system for ligands of the δ opioid receptor (L. Swevers, K. Iatrou, E. Morou, N. Balatsos and Z. Georgoussi, unpublished data).

4.14.5. Glycosylation of Recombinant Proteins in Insect Cells

Considerable attention has been paid to the glycosylation patterns of heterologous recombinant proteins produced in insect cells, particularly N-linked glycosylation. Several excellent and comprehensive reviews in this area have been published by Jarvis *et al.* (1998), Altmann *et al.* (1999), Marchal *et al.* (2001), and Jarvis (2003). In these reviews, it is agreed that oligomannose (five to nine mannose residues; Altmann *et al.*, 1999) and fucosylated trimannose type carbohydrate (Jarvis *et al.*, 1998) moieties dominate the N-linked glycosylation pattern of heterologous secreted and membrane proteins expressed in insect cells. The lack of complex glycosylation, such as the presence of terminal sialic acid found on many mammalian glycoproteins, excludes the use of insect cells to express mammalian glycoproteins for some applications. One solution to this limitation is to engineer baculoviruses (Jarvis *et al.*, 1996; Wagner *et al.*, 1996) or insect cell lines (Hollister *et al.*, 1998) to express mammalian glycotransferases necessary for complex glycosylation of an over-expressed protein. The coexpression of two mammalian glycotransferases, β-1,4-galactosyltransferase and α-2,6-sialyltransferase, enabled an insect cell for the first time to produce an artificially sialylated baculovirus glycoprotein gp64 (Seo *et al.*, 2001). This research will almost certainly extend the utility of insect cells to the production of "mammalized" complex glycoproteins.

4.14.6. Conclusions and Future Perspectives

Baculoviruses and insect cell lines have acquired a significant biotechnological niche for the production of proteins of general scientific and pharmaceutical value. While baculovirus vectors provide a superior system for high level expression of intracellular proteins, their use for the production of proteins that require an intact cellular environment is considered to be suboptimal. In contrast, the recently developed expression systems that employ baculovirus-derived genetic elements and direct continuous high level expression in transformed insect cells, not only circumvent the major drawbacks of the baculovirus expression systems for secreted and membrane-anchored proteins but can also be employed as high-throughput screening tools for the identification of bioactive substances with defined biological specificities.

For the baculovirus system, great technical improvements have occurred, mostly with regard to the efficiency and ease of generating recombinant viruses, with the best case being the generation of recombinant viruses through direct cloning into bacterially isolated bacmid chromosomal DNA. Because of the high degree of sophistication already achieved, relatively little improvement in this system is expected in the coming years. By contrast, the use of plasmid-based expression systems for insect cells is relatively recent and important improvements and new applications are expected to arise in the future.

Plasmid-based expression systems are not, in general, considered optimal for the production of intracellular proteins, because their isolation requires cell lysis and purification from a multitude of contaminating cellular proteins. Improvements in the production process are expected to include the redirection of normally intracellular recombinant proteins toward the extracellular medium, from which they can be purified with relative ease, particularly if the medium is serum-free. To test this strategy, secretion modules derived from secreted proteins, such as JHE and granulocyte-macrophage colony-stimulating factor (GM-CSF), have been fused in-frame with the N-termini of intracellular proteins, such as the CAT enzyme and the BmCF1 nuclear receptor (Farrell *et al.*, 2000). Efficient secretion to the extracellular medium has been reported and, because the engineered proteins also contained a histidine tag and an enteropeptidase cleavage site, single-step purification

could be achieved by metal-ion affinity binding coupled with proteolytic cleavage (Farrell *et al.*, 2000). Future modifications in this method may include the use of alternative affinity tags and proteolytic sites as well as improved versions of the secretion modules.

More recently, *Drosophila* S2 cell lines have been used as expression hosts for the purification of multiprotein complexes that are dedicated to the execution of particular cellular functions (Forler *et al.*, 2003). Key factors involved in particular cellular processes (e.g., splicing of mRNA precursor molecules, RNA quality control, nuclear export) were expressed as double-tagged proteins, while the expression of endogenous untagged proteins was suppressed by RNA interference. Tagged proteins, together with assembled interacting factors, were purified by the "tandem-affinity purification" strategy and component proteins of purified complexes were identified by Western blot and mass spectrometry (Forler *et al.*, 2003). The technique will be valuable for the characterization of multiprotein functional complexes operating in metazoan species, in general, and in insect ones, in particular.

Transformation of insect cell lines with expression constructs still represents a cumbersome process requiring selection of resistant cell clones after addition of antibiotics. It is expected, however, that alternatives to the method of selecting cells containing genomic integrants of plasmid vectors by antibiotic selection, including the use of vectors capable of directing autonomous integration into the host genome, will be developed in the immediate future. In fact, recent reports suggest that the use of a plasmid encompassing the genomic sequences of the *Junonia coenia* densovirus (JcDNV) directs autonomous genomic integration in insect cells grown in culture, as well as *in vivo*, without affecting the normal physiology of the host cells in any detectable way (Bossin *et al.*, 2003). Furthermore, it has also been reported that this system offers the additional advantage of integrated copy number manipulation through directed deletion of viral genome sequences from the expression plasmid (Bossin *et al.*, 2003). Although the production capabilities of this system have yet to be tested with specific protein models, it is anticipated that the incorporation in it of promoter systems that are in current use as drivers of recombinant protein expression in insect cells, will expand further the current capabilities and advantages of the nonlytic systems utilizing insect cells as hosts for continuous high level production of recombinant proteins over lytic ones. The major advantage of the use of the densovirus-based transformation system is the ease of selection of transformed cells by FACS analysis through the detection of red fluorescence in the cell nuclei (achieved by the nuclear expression of DsRed/viral capsid protein fusions directed by the use of the viral P9 promoter; Bossin *et al.*, 2003). Finally, the detection of efficient densovirus amplification and densovirus-derived transgene expression in *Drosophila* larval and adult tissues (Royer *et al.*, 2001) indicates that the densovirus-based transformation system is not limited to lepidopteran cell lines but can be applied to dipteran cell lines as well.

In mammalian systems, many nuclear receptors and GPCRs have been characterized at the molecular level using transformed cell lines, and this has led to the discovery of lead compounds (and drugs) that interfere specifically at very low doses with the function of the receptors. Equivalent data on insect receptors, however, are currently lacking to a great extent. Although the identification of the natural ligands for orphan insect (*Drosophila*) GPCRs is now occurring at a good pace, the identification of the relevant ligands was made possible through the use of heterologous expression systems such as transformed mammalian cell lines or *Xenopus* oocytes (Birgül *et al.*, 1999; Meeusen *et al.*, 2002; Staubli *et al.*, 2002; Johnson *et al.*, 2003). In this regard, we expect that insect cell lines would provide better host expression systems for insect GPCRs and could be employed for a more effective screening for compounds that interfere with their functions. The same applies for the identification of ligands for the approximately 20 orphan nuclear receptors that exist in the insect genome (Adams *et al.*, 2000; Zdobnov *et al.*, 2002). The transcriptional properties of several of these receptors have been characterized (e.g., the HNF-4 receptor (Kapitskaya *et al.*, 1998), the FTZ-F1 receptor (Suzuki *et al.*, 2001), and the HR3 and E75 receptors (Swevers *et al.*, 2002)) and it should be relatively straightforward to design cell-based reporter systems for screening of ligands, particularly small molecules capable of blocking their functions. In this regard, it is worth noting that the *Drosophila* DHR38/USP receptor complex was recently found to be activated by various ecdysteroids (Baker *et al.*, 2003), illustrating the as yet unrealized potential of orphan nuclear receptors as targets for interference by small molecule ligands. Thus, like their mammalian counterparts, insect orphan receptors can be considered as targets for interference by small molecules, which can be developed into environmentally friendly insecticides that function as "endocrine disruptors," and insect cell-based expression systems are anticipated to play an instrumental role in the discovery effort.

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., *et al.*, **2000**. The genome sequence of *Drosophila melanogaster. Science 287*, 2185–2195.

Airenne, K.J., Oker-Blom, C., Marjomaki, V.S., Bayer, E.A., Wilchek, M., *et al.*, **1997**. Production of biologically active recombinant avidin in baculovirus-infected insect cells. *Protein Expr. Purif. 9*, 100–108.

Alnemri, E.S., Litwack, G., **1993**. The steroid binding domain influences intracellular solubility of the baculovirus overexpressed glucocorticoid and mineralocorticoid receptors. *Biochemistry 32*, 5387–5393.

Altmann, F., Staudacher, E., Wilson, I.B.H., Marz, L., **1999**. Insect cells as hosts for the expression of recombinant glycoproteins. *Glycoconjugate J. 16*, 109–123.

Angelichio, M.L., Beck, J.A., Johansen, H., Ivey-Hoyle, M., **1991**. Comparison of several promoters and polyadenylation signals for use in heterologous gene expression in cultured Drosophila cells. *Nucleic Acids Res. 19*, 5037–5043.

Antoniewski, C., Mugat, B., Delbac, F., Lepesant, J.-A., **1996**. Direct repeats bind the EcR/USP receptor and mediate ecdysteroid responses in *Drosophila melanogaster. Mol. Cell. Biol. 16*, 2977–2986.

Atkinson, A.E., Henderson, J., Hawes, C.R., King, L.A., **1996**. Efficient membrane targeting of the chick nicotinic acetylcholine receptor a-subunit in insect cell lines. *Cytotechnology 19*, 37–42.

Bader, B., Knecht, W., Fries, M., Loffler, M., **1998**. Expression, purification and characterization of histidine-tagged rat and human flavoenzyme dihydroorotate dehydrogenase. *Protein Expr. Purif. 13*, 414–422.

Bailer, S.M., Berlin, W.K., Starr, C.M., Hanover, J.A., **1995**. Characterization of nuclear pore protein P62 produced using a baculovirus. *Protein Expr. Purif. 6*, 546–554.

Baker, K.D., Shewchuk, L.M., Kozlova, T., Makishima, M., Hassell, A., *et al.*, **2003**. The *Drosophila* orphan nuclear receptor DHR38 mediates an atypical ecdysteroid signaling pathway. *Cell 113*, 731–742.

Baldock, D., Graham, B., Akhlaq, M., Graff, P., Jones, C.E., *et al.*, **2000**. Purification and characterization of human syk produced using a baculovirus expression system. *Protein Expr. Purif. 18*, 86–94.

Barr, A.J., Brass, L.F., Manning, D.R., **1997**. Reconstitution of receptors and GTP-binding regulatory proteins (G proteins) in Sf9 cells. *J. Biol. Chem. 272*, 2223–2229.

Belyaev, A.X., Hails, R.S., Roy, P., **1995**. High-level expression of five foreign genes by a single recombinant baculovirus. *Gene 156*, 229–233.

Birgül, N., Weise, C., Kreienkamp, H.-J., Richter, D., **1999**. Reverse physiology in *Drosophila*: identification of a novel allatostatin-like neuropeptide and its cognate receptor structurally related to the mammalian somatostatin/galanin/opioid receptor family. *EMBO J. 18*, 5892–5900.

Bonning, B.C., Hammock, B.D., **1995**. Use of juvenile hormone esterase as a novel reporter enzyme in the baculovirus expression system. *J. Vir. Meth. 51*, 103–114.

Bonning, B.C., Roelvink, P.W., Vlak, J.M., Possee, R.D., Hammock, B.D., **1994**. Superior expression of juvenile hormone esterase and β-galactosidase from the basic protein promoter of *Autographa califonica* nuclear polyhedrosis virus compared to the p10 protein and polyhedrin promoters. *J. Gen. Virol. 75*, 1551–1556.

Bonning, B.C., Ward, V.K., van Meer, M.M.M., Booth, T.F., Hammock, B.D., **1997**. Disruption of lysosomal targeting is associated with insecticidal potency of juvenile hormone esterase. *Proc. Natl Acad. Sci. USA 94*, 6007–6012.

Bossin, H., Fournier, P., Royer, C., Barry, P., Cérutti, P., *et al.*, **2003**. *Junonia coenia* densovirus-based vectors for stable transgene expression in Sf9 cells: influence of the densovirus sequences on genomic integration. *J. Virol. 77*, 11060–11071.

Boublik, Y., Di Bonito, P., Jones, I.M., **1995**. Eukaryotic virus display, engineering the major surface glycoprotein of the *Autographa californica* nuclear polyhedrosis virus (AcNPV) for the presentation of foreign proteins on the virus surface. *Bio/Technol. 13*, 1079–1084.

Bouvier, M., Menard, L., Dennis, M., Marullo, S., **1998**. Expression and recovery of functional G-protein-coupled receptors using baculovirus expression systems. *Curr. Opin. Biotechnol. 9*, 522–527.

Boyce, F.M., Bucher, N.L., **1996**. Baculovirus-mediated gene transfer into mammalian cells. *Proc. Natl Acad. Sci. USA 93*, 2348–2352.

Bromme, D., McGrath, M.E., **1996**. High level expression and crystallization of recombinant human cathespin S. *Protein Sci. 5*, 789–791.

Brown, P.M., Scheid, M.P., O'Neill, G.P., Tagari, P.C., Nocholson, D.W., **1995**. A single-step purification of biologically active recombinant human interleukin-5 from a baculovirus expression system. *Protein Expr. Purif. 6*, 63–71.

Buckingham, S.D., Matsuda, K., Hosie, A.M., Baylis, H.A., Squire, M.D., *et al.*, **1996**. Wild-type and insecticide-resistant homo-oligomeric GABA receptors of *Drosophila melanogaster* stably expressed in a *Drosophila* cell line. *Neuropharmacology 35*, 1393–1401.

Bunch, T.A., Grinblat, Y., Goldstein, L.S.B., **1988**. Characterization and use of the *Drosophila* metallothionein promoter in cultured *Drosophila melanogaster* cells. *Nucl. Acids Res. 16*, 1043–1061.

Butkerait, P., Zheng, Y., Hallak, H., Graham, T.E., Miller, H.A., *et al.*, **1995**. Expression of the human 5-hydroxytryptamine1A receptor in Sf9 cells. Reconstitution of a coupled phenotype by co-expression of mammalian G protein subunits. *J. Biol. Chem. 270*, 18691–18699.

Cann, P.L., Touze, A., Enogat, N., Leboulleux, D., Mougin, C., *et al.*, **1995**. Detection of antibodies against human papillomavirus (HPV) type 16 virions by enzyme-linked immunosorbent assay using recombinant HPV 16 L1

capsids produced by recombinant baculovirus. *J. Clin. Microbiol. 33*, 1380–1382.

Canaan, S., Dupuis, L., Riviere, M., Faessel, K., Romette, J.-L., *et al.*, **1998**. Purification and interfacial behavior of recombinant human gastric lipase produced from insect cells in a bioreactor. *Protein Expr. Purif. 14*, 23–30.

Carbonell, L.F., Miller, L.K., **1987**. Baculovirus interaction with nontarget organisms: a virusborne reporter gene is not expressed in two mammalian cell lines. *Appl. Environm. Microbiol. 53*, 1412–1417.

Caron, A.W., Archambault, J., Massie, B., **1990**. High-level protein production in bioreactors using the baculovirus-insect cell expression system. *Biotechnol. Bioeng. 36*, 1133–1140.

Carpentier, E., Lebesgue, D., Kamen, A.A., Hogue, M., Bouvier, M., *et al.*, **2001**. Increased production of active human$_{\beta 2}$-andrenergic/G$_{\forall S}$ fusion receptor in Sf-9 cells using nutrient limiting conditions. *Protein Expr. Purif. 23*, 66–74.

Carson, D.D., Summers, M.D., Guarino, L.A., **1988**. Functional mapping of an AcNPV immediate early gene which augments the expression of the IE-1 *trans*-activated 39k gene. *Virology 162*, 444–451.

Chen, Y., Jin, M., Goodrich, L., Smith, G., Coppola, G., *et al.*, **2000**. Purification and characterisation of human α-galactosidase A expression in insect cells using a baculovirus vector. *Protein Expr. Purif. 20*, 228–236.

Chen, J.H., Turner, P.C., Rees, H.H., **2002**. Molecular cloning and induction of nuclear receptors from insect cell lines. *Insect Biochem. Mol. Biol. 32*, 657–667.

Cherbas, L., Moss, R., Cherbas, P., **1994**. Transformation techniques for *Drosophila* cell lines. *Methods Cell Biol. 44*, 161–179.

Cherbas, L., Yonge, C.D., Cherbas, P., Williams, C.M., **1980**. The morphological response of Kc-H cells to ecdysteroids: hormonal specificity. *Wilhelm Roux's Archives 189*, 1–15.

Chinni, C., Bottomley, S.P., Duffy, E.J., Hemmings, B.A., Stone, S.R., **1998**. Expression and purification of the human thrombin receptor. *Protein Expr. Purif. 13*, 9–15.

Ciaccia, A.V., Cunningham, E.L., Church, F.C., **1995**. Characterization of recombinant heparin cofactor II expressed in insect cells. *Protein Expr. Purif. 6*, 806–812.

Clawges, H.M., Depree, K.M., Parker, E.M., Graber, S.G., **1997**. Human 5-HT1 receptor subtypes exhibit distinct G protein coupling behaviors in membranes from Sf9 cells. *Biochemistry 36*, 12930–12938.

Clement, C.Y., Bradbrook, D.A., Lafont, R., Dinan, L., **1993**. Assessment of a microplate-based bioassay for the detection of ecdysteroid-like or antiecdysteroid activities. *Insect Biochem. Mol. Biol. 23*, 187–193.

Coleman, T.A., Parmelee, D., Rao Thotakura, N., Nguyen, N., Burgin, M., *et al.*, **1997**. Production and purification of novel secreted human proteins. *Gene 190*, 163–171.

Cooley, J., Mathieu, B., Remold-O'Donnell, E., Mandle, R.J., **1998**. Production of recombinant human monocyte/neutrophil elastase inhibitor (rM/NEI). *Protein Expr. Purif. 14*, 38–44.

Condreay, J.P., Witherspoon, S.M., Clay, W.C., Kost, T.A., **1999**. Transient and stable gene expression in mammalian cells transduced with a recombinant baculovirus vector. *Proc. Natl Acad. Sci. USA 96*, 127–132.

Culp, J.S., Johansen, H., Hellmig, B., Beck, J., Matthews, T.J., *et al.*, **1991**. Regulated expression allows high level production and secretion of HIV-1 gp120 envelope glycoprotein in *Drosophila* Schneider cells. *Biotechnology 9*, 173–177.

Dale, W.E., Textor, J.A., Mercer, R.W., Simchowitz, L., **1996**. Expression and characterization of the human erythrocyte anion exchanger in a baculovirus/Sf-9 system. *Protein Expr. Purif. 7*, 1–11.

Das, T., Johns, P.W., Goffin, V., Kelly, P., Kelder, B., *et al.*, **2000**. High-level expression of biologically active human prolactin from recombinant baculovirus in insect cells. *Protein Expr. Purif. 20*, 265–273.

Davis, T.R., Munkenbeck Trotter, K., Granados, R.R., Wood, H.A., **1992**. Baculovirus expression of alkaline phosphatase as a reporter gene for evaluation of production glycosyltaion and secretion. *Biotechnology 10*, 1148–1150.

Denault, J.-B., Lazure, C., Day, R., Leduc, R., **2000**. Comparative characterization of two forms of recombinant Human SPC1 secreted from Schneider 2 cells. *Protein Expr. Purif. 19*, 113–124.

DiFalco, M.R., Bakopanos, E., Patricelli, M., Chan, G., Congote, L.F., **1997**. The influence of various insect cell lines p10 and polyhedrin promoters in the production of secreted insulin-like growth factor-interleukin-3 chimeras in the baculovirus expression system. *J. Biotechnol. 56*, 49–56.

Dinan, L., Bourne, P.C., Meng, Y., Sarker, S.D., Tolentino, R.B., *et al.*, **2001**. Assessment of natural products in the Drosophila melanogaster B(II) cell bioassay for ecdysteroid agonist and antagonist activities. *Cell. Mol. Life Sci. 58*, 321–342.

Dinan, L., Whiting, P., Girault, J.-P., Lafont, R., Dhadialla, T.S., *et al.*, **1997**. Cucurbitacins are insect steroid hormone antagonists acting at the ecdysteroid receptor. *Biochem. J. 327*, 643–650.

Ding, H., Griesel, C., Nimtz, M., Conradt, H.S., Weich, H.A., *et al.*, **2003**. Molecular cloning, expression, purification, and characterization of soluble full-length, human interleukin-3 with a baculovirus-insect cell expression system. *Protein Expr. Purif. 31*, 34–41.

DiPersio, L.P., Kissel, J.A., Hui, D.Y., **1992**. Purification of pancreatic choilesterol esterase expressed in recombinant baculovirus-infected Sf9 cells. *Protein Expr. Purif. 3*, 114–120.

Dorjsuren, D., Badralmaa, Y., Mikovits, J., Li, A., Fisher, R., *et al.*, **2003**. Expression and purification of recombinant Kaposi's sarcoma-associated herpesvirus

DNA polymerase using a Baculovirus vector system. *Protein Expr. Purif. 29*, 42–50.

Egger, M., Ruknudin, A., Niggli, E., Lederer, W.J., Schulze, D.H., **1999**. Ni^{2+} transport by the human Na^{+}/Ca^{2+} exchanger expressed in Sf9 cell. *Am. J. Physiol. 276*, 1184–1192.

Emery, V.C., Bishop, D.H.L., **1987**. The development of multiple expression vectors for high level synthesis of eukaryotic proteins expression of LCMV-N and AcNPV polyhedrin protein by a recombinant baculovirus. *Protein Engineer. 1*, 359–366.

Ernst, W., Grabherr, R., Wegner, D., Borth, N., Grassauer, A., *et al.*, **1998**. Baculovirus surface display construction and screening of a eukaryotic epitope library. *Nucleic Acids Res. 26*, 1718–1723.

Ernst, W.J., Grabherr, R.M., Katinger, H.W.D., **1994**. Direct cloning into the *Autographa californica* nuclear polyhedrosis virus for generation of recombinant baculoviruses. *Nucleic Acids Res. 22*, 2855–2856.

Escriva, H., Safi, R., Hänni, C., Langlois, M.-C., Saumitou-Laprade, P., *et al.*, **1997**. Ligand binding was acquired during evolution of nuclear receptors. *Proc. Natl Acad. Sci. USA 94*, 6803–6808.

Farrell, P.J., Behie, L.A., Iatrou, K., **2000**. Secretion of intracellular proteins from animal cells using a novel secretion module. *Proteins–Struct. Funct. Genetics 41*, 144–150.

Farrell, P.J., Behie, L.A., Iatrou, K., **1999**. Transformed insect cells new sources of human tissue plasminogen activator. *Biotechnol. Bioeng. 64*, 426–433.

Farrell, P.J., Iatrou, K., **2004**. Transfected insect cells in suspension culture rapidly yield moderate quantities of recombinant proteins in protein-free culture medium. *Protein Expr. Purif. 36*, 177–185.

Farrell, P.J., Lu, M., Prevost, J., Brown, C., Behie, L.A., *et al.*, **1998**. High level of expression recombinant glycoproteins from transformed Lepidopteran insect cells using a novel expression vector. *Biotechnol. Bioeng. 60*, 656–663.

Forler, D., Köcher, T., Rode, M., Gentzel, M., Izaurralde, E., *et al.*, **2003**. An efficient protein complex purification method for functional proteomics in higher eukaryotes. *Nature Biotechnol. 21*, 89–92.

French, T.J., Roy, P., **1990**. Synthesis of bluetongue virus (BTV) corelike particles by a recombinant baculovirus expressing the two major structural core proteins of BTV. *J. Virol. 64*, 1530–1536.

Gao, X., McLean, H., Caveney, S., Donly, C., **1999**. Molecular cloning and functional characterization of a GABA transporter from the CNS of the cabbage looper *Trichoplusia ni*. *Insect Biochem. Mol. Biol. 29*, 609–623.

Geisse, S., Gram, H., Kleuser, B., Kocher, H.P., **1996**. Eukaryotic expression systems a comparison. *Protein Expr. Purif. 8*, 271–282.

George, H.J., Marchand, P., Murphy, K., Wiswall, B.H., Dowling, R., *et al.*, **1997**. Recombinant human 92-kDa type IV collagenase/gelatinase from baculovirus-infected insect cells expression purification and characterization. *Protein Expr. Purif. 10*, 154–161.

Gierse, J.K., Luckow, V.A., Askonas, L.J., Duffin, K.L., Aykent, S., *et al.*, **1993**. High-level expression and purification of human leukotriene A_4 hydrolase from insect cells infected with a baculovirus vector. *Protein Expr. Purif. 4*, 358–366.

Golden, A., Austen, D.A., van Schravendijk, M.R., Sullivan, B.J., Kawasaki, A.S., *et al.*, **1998**. Effect of promoters and signal sequences on the production of secreted HIV-1 gp120 protein in the baculovirus system. *Protein Expr. Purif. 14*, 8–12.

Grabherr, R., Ernst, W., Oker-Blom, C., Jones, I., **2001**. Development in the use of baculoviruses for the surface display of complex eukaryotic proteins. *Trends Biotechnol. 19*, 231–236.

Grace, T.D.C., **1967**. Establishment of a line of cells from the silkworm *Bombyx mori*. *Nature 216*, 613.

Granados, R.R., Li, G., Derkensen, A.C.G., McKenna, K.A., **1994**. A new insect cell line from *Trichoplusia ni* (BTI-Tn-5B1-4) susceptible to *Trichloplusia ni* single enveloped nuclear polyhedrosis virus. *J. Inverteb. Pathol. 64*, 260–266.

Gritsun, T.S., Mikhailov, M.V., Roy, P., Gould, E.A., **1997**. A new rapid and simple procedure for direct cloning of PCR products into baculoviruses. *Nucleic Acids Res. 25*, 1864–1865.

Gronemeyer, H., Laudet, V., **1995**. Nuclear receptors. *Protein Profile 2*, 1173–1308.

Grynkiewicz, G., Poenie, M., Tsien, R.Y., **1985**. A new generation of Ca^{2+} indicators with greatly improved fluorescence properties. *J. Biol. Chem. 260*, 3440–3450.

Guarino, L.A., Gonzalez, M.A., Summers, M.D., **1986**. Complete sequence and enhancer function of the homologous DNA regions of *Autographa californica* nuclear polyhedrosis virus. *J. Virol. 60*, 224–229.

Guarino, L.A., Summers, M.D., **1986**. Functional mapping of a trans-activating gene required for expression of a baculovirus delayed-early gene. *J. Virol. 57*, 563–571.

Guarino, L.A., Summers, M.D., **1987**. Nucleotide sequence and temporal expression of a baculovirus regulatory gene. *J. Virol. 61*, 2091–2099.

Hale, C.C., Zimmerschied, J.A., Bliler, S., Price, E.M., **1999**. Large-scale expression of recombinant cardiac sodium-calcium exchange in insect larvae. *Protein Expr. Purif. 15*, 121–126.

Hamm, H., **2001**. How activated receptors couple to G proteins. *Proc. Natl Acad. Sci. USA 98*, 4819–4821.

Harmatha, J., Dinan, L., **1997**. Biological activity of natural and synthetic ecdysteroids in the BII bioassay. *Arch. Insect Biochem. Physiol. 35*, 219–225.

Harmatha, J., Dinan, L., Lafont, R., **2002**. Biological activities of a specific ecdysteroid dimer and of selected monomeric structural analogues in the B(II) bioassay. *Insect Biochem. Mol. Biol. 32*, 181–185.

Hegedus, D.D., Pfeifer, T.A., Hendry, J., Theilmann, D.A., Grigliatti, T.A., **1998**. A series of broad host range shuttle vectors for constitutive and inducible expression

of heterologous proteins in insect cells. *Gene 207*, 241–249.

Hegedus, D.D., Pfeifer, T.A., Theilmann, D.A., Kennard, M.L., Gabathuler, R., *et al.*, **1999**. Differences in the expression and localization of human melanotransferrin in lepidopteran and dipteran insect cell lines. *Protein Expr. Purif. 15*, 296–307.

Helgen, J.C., Fallon, A.M., **1990**. Polybrene-mediated transfection of cultured lepidopteran cells: induction of a *Drosophila* heat shock promoter. *In Vitro Cell Dev. Biol. 26*, 731–736.

Higashihashi, N., Arai, Y., Enjo, T., Horiuchi, T., Saeki, Y., *et al.*, **1991**. High-level expression and characterization of hepatitis B virus surface antigen in silkworm using a baculovirus vector. *J. Vir. Meth. 35*, 159–167.

Hill, C.A., Fox, A.N., Pitts, R.J., Kent, L.B., Tan, P.L., *et al.*, **2002**. G protein-coupled receptors in *Anopheles gambiae*. *Science 298*, 176–178.

Hill-Perkins, M.S., Possee, R.D., **1990**. A baculovirus expression vector derived from the basic protein promoter of *Autographa californica* nuclear polyhedrosis virus. *J. Gen. Virol. 71*, 971–976.

Ho, W.K., Meng, Z.Q., Lin, H.R., Poon, C.T., Leung, Y.K., *et al.*, **1998**. Expression of grass carp growth hormone by baculovirus in silkworm larvae. *Biochim. Biophys. Acta 1381*, 331–339.

Hoffmann, C.V., Sandig, G., Jennings, G.T., Rudolph, M., Schlag, P., *et al.*, **1995**. Efficient gene transfer into human hepatocytes by baculovirus vectors. *Proc. Natl Acad. Sci. USA 92*, 10088–10103.

Holcomb, T., Taylor, L., Trohkimoinen, J., Curthoys, N.P., **2000**. Isolation characterization and expression of a human brain mitochondrial glutaminase cDNA. *Molec. Brain Res. 76*, 56–63.

Hollenberg, M.D., Saifeddine, M., Al-Ani, B., **1996**. Proteinase-activated receptor-2 in rat aorta structural requirements for agonist activity of receptor-activating peptides. *Molec. Pharmacol. 49*, 229–233.

Hollister, J.R., Shaper, J.H., Jarvis, D.L., **1998**. Stable expression of mammalian beta 1,4-galactosyltransferase extends the N-glycosylation pathway in insect cells. *Glycobiology 8*, 473–480.

Homa, F.L., Baker, C.A., Thomsen, D.R, Elhammer, A.P., **1995**. Conversion of bovine UDP-GalNAcpolypeptide *N*-acetylgalactosaminyltransferase to a soluble secreted enzyme and expression in Sf9 cells. *Protein Expr. Purif. 6*, 141–148.

Hsu, T.-A., Eiden, J.J., Bourgarel, P., Meo, T., Betenbaugh, M.J., **1994**. Effects of co-expressing chaperone BiP on functional antibody production in the baculovirus system. *Protein Expr. Purif. 5*, 595–603.

Huang, Y.-W., Lu, M.-L., Qi, H., Lin, S.-X., **2000**. Membrane-bound human 3-hydroxysteroid dehydrogenase overexpression with his-tag using a baculovirus system and single step purification. *Protein Expr. Purif. 18*, 169–174.

Huynh, C.Q., Zieler, H., **1999**. Construction of modular and versatile plasmid vectors for the high-level expression of single or multiple genes in insects and insect cell lines. *J. Mol. Biol. 288*, 13–20.

Iatrou, K., Farrell, P., Hashimoto, Y., **2000**. Baculovirus Artificial Chromosomes and Methods of Use. U.S. Patent Number 6,090,584.

Iatrou, K., Meidinger, R.G., **1989**. Bombyx mori nuclear polyhedrosis virus-based vectors for expressing passenger genes in silk moth cells under viral or cellular promoter control. *Gene 75*, 59–71.

Ivey-Hoyle, M., Culp, J.S., Chaikin, M.A., Hellmig, B.D., Mathews, T.J., *et al.*, **1991**. Envelope glycoproteins from biologically diverse isolates of immunodeficiency viruses have widely different affinities for CD4. *Proc. Natl Acad. Sci. USA 88*, 512–516.

Jarvis, D.L., **2003**. Developing baculovirus – insect cell expression systems for humanized recombinant glycoprotein production. *Virology 310*, 1–7.

Jarvis, D.L., Fleming, J.-A.G.W., Kovacs, G.R., Summers, M.D., Guarino, L.A., **1990**. Use of early baculovirus promoters for continuous expression and efficient processing of foreign gene products in stably transformed lepidopteran cells. *Biotechnology 8*, 950–955.

Jarvis, D.L., Guarino, L.A., **1995**. Continuous foreign gene expression in stably transformed insect cells. In: "Methods in Molecular Biology Baculovirus Expression Protocols," vol. 39. Humana, Clifton, NJ, pp. 187–202.

Jarvis, D.L., Kawar, Z.S., Hollister, J.L., **1998**. Engineering N-glycosylation in the baculovirus-insect cell system. *Curr. Opin. Biotechnol. 9*, 528–533.

Jarvis, D.L., Summers, M.D., **1989**. Glycosylation and secretion of human tissue plasminogen activator in recombinant baculovirus-infected insect cells. *Mol. Cell. Biol. 9*, 214–223.

Jarvis, D.L., Summers, M.D., Garcia, A., Jr., Bohlmeyer, D.A., **1993**. Influence of different signal peptides and prosequences on expression and secretion of human tissue plasminogen activator in the baculovirus system. *J. Biol. Chem. 268*, 16754–16762.

Jarvis, D.L., Weinkauf, C., Guario, L.A., **1996**. Immediate-early baculovirus vectors for foreign gene expression in transformed or infected insect cells. *Protein Expr. Purif. 8*, 191–203.

Ji C., Wang Y., Ross J., Jiang, H., **2003**. Expression and *in vitro* activation of *Manduca sexta* prophenoloxidase-activating proteinase-2 precursor (proPAP-2) from baculovirus-infected insect cells. *Protein Expr. Purif. 29*, 235–243.

Jiang, X., Wang, M., Graham, D.Y., Estes, M.K., **1992**. Expression self-assembly and antigenicity of the Norwalk virus capsid protein. *J. Virol. 66*, 6527–6532.

Johansen, H., van der Straten, A., Sweet, R., Otto, E., Maroni, G., *et al.*, **1989**. Regulated expression of high copy number allows production of a growth inhibitory oncogene product in Drosophila Schneider cells. *Genes Devel. 3*, 882–889.

Johansen, K., Appelbaum, E., Doyle, M., Hensley, P., Zhao, B., *et al.*, **1995**. Binding interactions of human interleukin 5 with its receptor alpha subunit. Large scale production structural and functional studies of

Drosophila-expressed recombinant proteins. *J. Biol. Chem. 270*, 9459–9471.

Johnson, E.C., Garczynski, S.F., Park, D., Crim, J.W., Nässel, D.R., *et al.*, **2003**. Identification and characterization of a G protein-coupled receptor for the neuropeptide proctolin in *Drosophila melanogaster. Proc. Natl Acad. Sci. USA 100*, 6198–6203.

Johnson, R., Meidinger, R.G., Iatrou, K., **1992**. A cellular promoter-based expression cassette for generating recombinant baculoviruses directing rapid expression of passenger genes in infected insects. *Virology 190*, 815–823.

Joyce, K.A., Atkinson, A.E., Bermudez, I., Beadle, D.J., King, L.A., **1993**. Synthesis of functional GABA-A receptors in stable insect cell lines. *FEBS Lett. 335*, 61–64.

Juarbe-Osorio, L.G., Gorziglia, M., Betenbaugh, M.J., **1993**. Recovery and characterization of a rotavirus outer capsid protein expressed in a recombinant insect cell system. *Protein Expr. Purif. 4*, 232–239.

Kapitskaya, M.Z., Dittmer, N.T., Deitsch, K.W., Cho, W.-L., Taylor, D.G., *et al.*, **1998**. Three isoforms of a hepatocyte nuclear factor-4 transcription factor with tissue- and stage-specific expression in the adult mosquito. *J. Biol. Chem. 273*, 29801–29810.

Kang, X.-Q., Wiggins, J., Mallick, S., Grant, S.R., **1992**. Production purification and characterization of human recombinant IL-8 from the eukaryotic expression system baculovirus. *Protein Expr. Purif. 3*, 313–321.

Keith, M.B.A., Farrell, P.J., Iatrou, K., Behie, L.A., **1999**. Screening of transformed insect cell lines for recombinant protein production. *Biotechnol. Prog. 15*, 1046–1052.

Kirkpatrick, R.B., Ganguly, S., Angelichio, M., Griego, S., Shatzmann, A., *et al.*, **1995**. Heavy chain dimers as well as complete antibodies are efficiently formed and secreted from *Drosophila* via a BiP-mediated pathway. *J. Biol. Chem. 270*, 19800–19805.

Kitts, P.A., Ayres, M.D., Possee, R.D., **1990**. Linearization of baculovirus DNA enhances the recovery of recombinant virus expression vectors. *Nucleic Acids Res. 18*, 5667–5672.

Kitts, P.A., Possee, R.D., **1993**. A method for producing recombinant baculovirus expression vectors at high frequency. *Biotechniques 14*, 810–817.

Kleymann, G., Boege, F., Hahn, M., Hampe, W., Vasudevan, S., *et al.*, **1993**. Human β_2-adrenergic receptor produced in stably transformed insect cells is functionally coupled via endogenous GTP-binding protein to adenylyl cyclase. *Eur. J. Biochem. 213*, 797–804.

Knight, M.R., Campbell, A.K., Smith, S.M., Trewavas, A.J., **1991**. Recombinant aequorin as a probe for cytosolic free Ca^{2+} in *Escherichia coli. FEBS Lett. 282*, 405–408.

Koelle, M.R., Talbot, W.S., Segraves, W.A., Bender, M.T., Cherbas, P., *et al.*, **1991**. The *Drosophila EcR* gene encodes an ecdysone receptor, a new member of the steroid receptor superfamily. *Cell 67*, 59–77.

Kostenis, E., **2001**. Is $G\alpha_{16}$ the optimal tool for fishing ligands of orphan G-protein-coupled receptors? *Trends Pharmacol. Sci. 22*, 560–564.

Kovach, M.J., Carlson, J.O., Beaty, B.J., **1992**. A *Drosophila* metallothionein promoter is inducible in mosquito cells. *Insect Mol. Biol. 1*, 37–43.

Kurkela, R., Herrala, A., Henttu, P., Nal, H., Vihk, P., **1995**. Expression of active secreted human prostrate-specific antigen by recombinant baculovirus-infected insect cells on a pilot-scale. *Biotechnology 13*, 1230–1234.

Kuzio, J., Jaques, R., Faulkner, P., **1989**. Identification of p74, a gene essential for virulence of baculovirus occlusion bodies. *Virology 173*, 759–763.

Lan, Q., Riddiford, L.M., **1997**. DNA transfection in the ecdysteroid-responsive GV1 cell line from the tobacco hornworm, *Manduca sexta. In Vitro Cell Dev. Anim. 33*, 615–621.

Laukkanen, M.-L., Oker-Blom, C., Keinanen, K., **1996**. Secretion of green fluorescence protein from recombinant baculovirus-infected cells. *Biochem. Biophys. Res. Commun. 266*, 755–781.

Lee, D.F., Chen, C.-C., Hsu, T.-A., Juang, J.-L., **2000**. A baculovirus superinfection system: efficient vehicle for gene transfer into *Drosophila* S2 cells. *J. Virol. 74*, 11873–11880.

Lenhard, T., Reilander, H., **1997**. Engineering the folding pathway of insect cells generation of a stably transformed insect cell line showing improved folding of a recombinant membraneprotein. *Biochem. Biophys. Res. Commun. 238*, 823–830.

Li, E., Brown, S.L., Dolman, C.S., Brown, G.B., Nemerow, G.R., **2001**. Production of functional antibodies generated in a nonlytic insect cell expression system. *Protein Expr. Purif. 21*, 121–128.

Li, Z., Smith, C.D., Smolley, J.R., Bridge, J.H.B., Frank, K.S., *et al.*, **1992**. Expression of the cardiac Na^+–Ca^{2+} exchanger in insect cells using a baculovirus vector. *J. Biol. Chem. 267*, 7828–7833.

Lindley, K.M., Su, J.L., Hodges, P.K., Wisely, G.B., Bledsoe, R.K., *et al.*, **2000**. Production of monoclonal antibodies using recombinant baculovirus displaying gp64-fusion proteins. *J. Immunol. Methods 234*, 123–135.

Loisel, T.P., Ansanay, H., St-Onge, S., Gay, B., Boulanger, P., *et al.*, **1997**. Recovery of homogeneous and functional β_2-andrenergic receptors from extracellular baculovirus particles. *Nature Biotechnol. 15*, 1300–1304.

Lord, K.A., Wang, X.-M., Simmons, S.J., Bruckner, R.C., Loscig, J., *et al.*, **1997**. Variant cDNA sequences of human ATPcitrate lyase cloning expression and purification from baculovirus-infected insect cells. *Protein Expr. Purif. 9*, 133–141.

Lowe, M.E., **1994**. Human pancreatic procolipase expressed in insect cells purification and characterization. *Protein Expr. Purif. 5*, 583–586.

Lu, M., Farrell, P.J., Johnson, R., Iatrou, K., **1997**. A baculovirus (BmNPV) repeat element functions as a powerful constitutive enhancer in transfected insect cells. *J. Biol. Chem. 272*, 30724–30728.

Lu, M., Johnson, R.R., Iatrou, K., **1996**. *Trans*-activation of a cell housekeeping gene promoter by the IE1 gene product of baculoviruses. *Virology 218*, 103–113.

Lu, A., Miller, L.K., **1995**. The roles of eighteen baculovirus late gene expression factor genes in transcription and DNA replication. *J. Virol. 69*, 975–982.

Lu, A., Miller, L.K., **1996**. Generation of recombinant baculoviruses by direct cloning. *Biotechniques 21*, 63–68.

Luckow, V.A., **1991**. Cloning and expression of heterologous genes in insect cells with baculovirus vectors. In: Prokop, A., Bajpai, R.K., Ho, S.C. (Eds.), Recombinant DNA Technology and Applications. McGraw Hill, New York, pp. 97–125.

Luckow, V.A., Lee, S.C., Barry, G.F., Olins, P.O., **1993**. Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in Escherischia coli. *J. Virol. 67*, 4566–4579.

Luckow, V.A., Summers, M.D., **1988**. Signals important for high-level expression of foreign genes in *Autographa californica* nuclear polyhedrosis virus. *Virology 167*, 56–71.

Luckow, V.A., Summers, M.D., **1989**. High level expression of nonfused foreign genes with *Autographa californica* nuclear polyhedrosis virus expression vectors. *Virology 170*, 31–39.

Lycett, G.J., Crampton, J.M., **1993**. Stable transformation of mosquito cell lines using a hsp70:neo fusion gene. *Gene 136*, 129–136.

Maeda, S., **1989**. *Bombyx mori* nuclear polyhedrosis virus and their use for expression of foreign genes in insect cells. In: Mitshuashi, J. (Ed.), Invertebrate Cell Systems and Applications, vol. 1. CRC, Boca Raton, London, pp. 167–181.

Maeda, S., Kawai, T., Obinata, M., Fujiwara, H., Horiuchi, T., *et al.*, **1995**. Production of human α-interferon in silkworm using a baculovirus vector. *Nature 315*, 592–594.

Marchal, I., Jarvis, D.L., Cacan, R., Verbert, A., **2001**. Glycoproteins from insect cells sialylated or not? *Biol. Chem. 382*, 151–159.

Marinissen, M.J., Gudkind, S.J., **2001**. G protein-coupled receptors and signalling networks: emerging paradigms. *Trends Pharmacol. Sci. 22*, 368–376.

Massotte, D., Baroche, L., Simonin, F., Yu, L., Kieffer, B., *et al.*, **1997**. Characterization of δ, κ and μ human opoid receptors overexpressed in baculovirus-infected insect cells. *J. Biol. Chem. 272*, 19987–19992.

Matsuura, Y., Possee, R.D., Overton, H.A., Bishop, D.H.L., **1987**. Baculovirus expression vectors the requirements for high level expression of proteins including glycoproteins. *J. Gen. Virol. 68*, 1233–1250.

Mathews, S., Dobeli, H., Pruschy, M., Bosser, R., D'Arcy, A., *et al.*, **1996**. Recombinant human rennin produced in different expression systems biochemical properties and 3D structure. *Protein Expr. Purif. 7*, 81–92.

Matthews, R.E.F., **1982**. Classification and nomenclature of viruses. *Intervirology 17*, 1–199.

McLachlin, J.R., Miller, L.K., **1997**. Stable transformation of insect cells to coexpress a rapidly selectable marker gene and an inhibitor of apoptosis. *In vitro Cell. Devel. Virol. Animal 33*, 575–579.

McQueney, M.S., Feild, J., Hanning, C.R., Brun, K., Ramachandran, K., *et al.*, **1998**. Cynomolgus monkey (*Macaca fascicularis*) cathespin K cloning expression purification and activation. *Protein Expr. Purif. 14*, 387–394.

Meeusen, T., Mertens, I., Clynen, E., Baggerman, G., Nichols, R., *et al.*, **2002**. Identification in *Drosophila melanogaster* of the invertebrate G protein-coupled FMRFamide receptor. *Proc. Natl Acad. Sci. USA 99*, 15363–15368.

Meij, P., Vervoort, M.B., de Gooijer, K., Bloemena, E., Meijer, C.J., *et al.*, **2000**. Bioreactor-scale production and one-step purification of Epstein-Barr nuclear antigen 1 expressed in baculovirus-infected insect cells. *Protein Expr. Purif. 20*, 324–333.

Merrihew, R.V., Clay, W.C., Condreay, J.P., Witherspoon, S.M., Dallas, W.S., *et al.*, **2001**. Chromosomal integration of transduced recombinant baculovirus DNA in mammalian cells. *J. Virol. 75*, 903–909.

Millar, N.S., Baylis, H.A., Reaper, C., Bunting, R., Mason, W.T., *et al.*, **1995**. Functional expression of a cloned *Drosophila* muscarinic acetylcholine receptor in a stable *Drosophila* cell line. *J. Exp. Biol. 198*, 1843–1850.

Millar, N.S., Buckingham, S.D., Sattelle, D.B., **1994**. Stable expression of a functional homo-oligomeric *Drosophila* GABA receptor in a *Drosophila* cell line. *Proc. Roy. Soc. Lond. 258*, 307–314.

Miller, L.H., Sakai, R.K., Romans, P., Gwadz, R.W., Coon, H.G., **1987**. Stable integration and expression of a bacterial gene in the mosquito *Anopheles gambiae*. *Science 237*, 779–781.

Minev, B.R., Chaves, F.L., Mitchell, M.S., **1999**. Cancer vaccines novel approaches and a new promise. *Pharmacol. Ther. 81*, 121–139.

Miras, R., Cuillel, M., Catty, P., Guillain, F., Mintz, E., **2001**. Purification of heterologous sarcoplasmic reticulum Ca^{2+}-ATPase Serca1a allowing phosphoenzyme and Ca^{2+}-affinity measurements. *Protein Expr. Purif. 22*, 299–306.

Miyajima, A., Schreurs, J., Otsu, K., Kondo, A., Arai, K., *et al.*, **1987**. Use of the silkworm *Bombyx mori* and an insect baculovirus vector for high-level expression and secretion of biologically active mouse interleukin-3. *Gene 58*, 273–281.

Miyasaka, T., Kaminogawa, S., Shimizu, M., Hisatsune, T., Reinach, P.S., *et al.*, **2001**. Characterization of human taurine transporter expressed in insect cells using a recombinant baculovirus. *Protein Expr. Purif. 23*, 389–397.

Monsma, S.A., Oomens, A.G., Blissard, G.W., **1996**. The gp64 envelope fusion protein is an essential baculovirus protein required for cell-to-cell transmission of an infection. *J. Virol. 70*, 4607–4616.

Mori, H., Yamao, M., Nakazawa, H., Sugahara, Y., Shirai, N., *et al.*, **1995**. Transovarian transmission

of a foreign gene in the silkworm *Bombyx mori* by *Autographa californica* nuclear polyhedrosis virus. *Biotechnology 13*, 1005–1007.

Morris, T.D., Miller, L.K., **1992**. Promoter influence on baculovirus-mediated gene expression in permissive and nonpermissive insect cell lines. *J. Virol. 66*, 7397–7405.

Moto, K., Kojima, H., Kurihara, M., Iwami, M., Matsumoto, S., **2003**. Cell-specific expression of enhanced green fluorescence protein under the control of neuropeptide gene promoters in the brain of the silkworm, *Bombyx mori* nucleopolyhedrosis-derived vectors. *Insect Biochem. Molec. Biol. 33*, 7–12.

Mounier, N., Prudhomme, J.C., **1986**. Isolation of actin genes in *Bombyx mori*. The coding sequence of a cytoplasmic actin gene expressed in the silk gland is interrupted by a single intron in an unusual position. *Biochemie 68*, 1053–1061.

Mroczkowski, B.S., Huvar, A., Lernhardt, W., Misono, K., Nielson, K., *et al.*, **1994**. Secretion of thermostable DNA polymerase using a novel baculovirus vector. *J. Biol. Chem. 269*, 13522–13528.

Murphy, C.A., McIntire, J.R., Davis, D.R., Hodgdon, H., Seals, J.R., *et al.*, **1993**. Enhanced expression and large-scale purification of recombinant HIV-1 gp120 in insect cells using the baculovirus egt and p67 signal peptides. *Protein Expr. Purif. 4*, 349–357.

Myamoto, C., Smith, G.E., Farrell-Towt, J., Chizzonite, R., Summers, M.D., *et al.*, **1985**. Production of human c-*myc* protein in insect cells infected with a baculovirus expression vector. *Mol. Cell. Biol. 5*, 2860–2865.

Nagarajan, S., Selvaraj, P., **1999**. Expression and characterization of glycolipid-anchored B7 (CD80) from baculovirus-infected insect cells protein transfer onto tumor cells. *Protein Expr. Purif. 17*, 273–281.

Newcomb, W.W., Homa, F.L., Thomsen, D.R., Trus, B.L., Cheng, N., *et al.*, **1999**. Assembly of the herpes simplex virus procapsid from purified components and identification of small complexes containing the major capsid and scaffolding proteins. *J. Virol. 73*, 4239–4250.

Ng, G.Y.K., Trogadis, J., Stevens, J., Bouvier, M., O'Dowd, B.F., *et al.*, **1995**. Agonist-induced desensitization of dopamine D1 receptor-stimulated adenylyl cyclase activity is temporally and biochemically separated from D1 receptor internalization. *Proc. Natl Acad. Sci. USA 92*, 10157–10161.

Ng, G.Y.K., Varghese, G., Chung, H.T., Trogadis, J., Seeman, P., *et al.*, **1997**. Resistance of the dopamine $D2_L$ receptor to desensitization accompanies the upregulation of receptors on to the surface of Sf9 cells. *Endocrinology 138*, 4199–4206.

Nilsen, S.L., Castellino, F.J., **1999**. Expression of human plasminogen in *Drosophila* Schneider S2 cells. *Protein Expr. Purif. 16*, 136–143.

Ohtaki, T., Ogi, K., Masuda, Y., Mitsuoka, K., Fujiyoshi, Y., *et al.*, **1998**. Expression, purification, and reconstitution of receptor for pituitary adenylate cyclase-activating polypeptide. *J. Biol. Chem. 273*, 15464–15473.

Oppenheimer, D.I., MacNicol, A.M., Patel, N.H., **1999**. Functional conservation of the *wingless-engrailed* interaction as shown by a widely applicable baculovirus misexpression system. *Curr. Biol. 9*, 1288–1296.

O'Reilly, D.R., Miller, L.K., Luckow, V.A., **1992**. Baculovirus Expression Vectors A Laboratory Manual. W.H. Freeman and Co, New York.

Otto, E., Allen, J.M., Young, J.E., Palmiter, R.D., Maroni, G., **1987**. A DNA segment controlling metal-regulated expression of the *Drosophila melanogaster* metallothionein gene. *Mol. Cell. Biol. 7*, 1710–1715.

Page, M.J., **1989**. p36C an improved baculovirus expression vector for producing high levels of mature recombinant proteins. *Nucleic Acids Res. 17*, 454.

Pajot-Augy, E., Couture, L., Bozon, V., Remy, J.J., Biache, G., *et al.*, **1995**. High-level expression of recombinant porcine LH receptor in baculovirus-infected insect cells or caterpillars. *J. Mol. Endocrinol. 14*, 1–66.

Palombo, F., Monciotti, A., Recchia, A., Cortese, R., Ciliberto, G., *et al.*, **1998**. Site-specific integration in mammalian cells mediated by a new hybrid baculovirus-adeno-associated virus vector. *J. Virol. 72*, 5025–5034.

Passarelli, A.L., Todd, J.W., Miller, L.K., **1994**. A baculovirus gene involved in late gene expression predicts a large polypeptide with a conserved motif of RNA polymerases. *J. Virol. 68*, 4673–4678.

Patel, G., Nasmyth, K., Jones, N., **1992**. A new method for the isolation of recombinant baculovirus. *Nucleic Acids Res. 20*, 97–104.

Pennock, G.D., Shoemaker, C., Miller, L.K., **1984**. Strong and regulated expression of *Escherichia coli* β-galactosidase in insect cells with a baculovirus vector. *Mol. Cell. Biol. 4*, 399–406.

Percival, M.D., Bastien, L., Griffin, P.R., Kargman, S., Ouellet, M., *et al.*, **1997**. Investigation of human cyclooxygenase-2 glycosylation heterogeneity and protein expression in insect and mammalian cell expression systems. *Protein Expr. Purif. 9*, 388–397.

Pereira, C.A., Pouliquen, Y., Rodas, V., Massotte, D., Mortensen, C., *et al.*, **2001**. Optimized insect cell culture for the production of recombinant heterologous proteins and baculovirus particles. *Biotechniques 6*, 1263–1268.

Perret, B.G., Wagner, R., Lecat, S., Brillet, K., Rabut, G., *et al.*, **2003**. Expression of EGFP-amino-tagged human mu opioid receptor in *Drosophila* Schneider 2 cells: a potential expression system for large-scale production of G-protein coupled receptors. *Protein Expr. Purif. 31*, 123–132.

Pfeifer, T.A., Guarna, M.M., Kwan, E.M., Lesnicki, G., Theilmann, D.A., *et al.*, **2001**. Expression analysis of a modified FactorX in stable insect cell lines. *Protein Expr. Purif. 23*, 233–241.

Pfeifer, T.A., Hegedus, D.D., Grigliatti, T.A., Theilmann, D.A., **1997**. Baculovirus immediate-early promoter-mediated expression of the Zeocin resistance gene for use as a dominant selectable marker in dipteran and lepidopteran insect cell lines. *Gene 188*, 183–190.

Phan, D., Han, E., Birrell, G., Bonnal, S., Duggan, L., *et al.*, **2001**. Purification and characterization of human

cell–cell adhesion molecule 1 (C-CAM1) expressed in insect cells. *Protein Expr. Purif. 21*, 343–351.

Pittelkow, S., Lindner, H., Rohm, K.-L., **1998**. Human and porcine aminoacylase I overproduced in a baculovirus expression vector system: evidence for structural and functional identity with enzymes isolated from kidney. *Protein Expr. Purif. 12*, 269–276.

Possee, R.D., Howard, S.C., **1987**. Analysis of the polyhedrin gene promoter of the *Autographa californica* nuclear polyhedrosis virus. *Nucleic Acids Res. 15*, 10233–10248.

Possee, R.D., Sun, T.P., Howard, S.C., Ayres, M.D., Hill-Perkins, M., *et al.*, **1991**. Nucleotide sequence of the *Autographa californica* nuclear polyhedrosis 9.4 kbp EcoRI-I and -R (polyhedrin gene) region. *Virology 185*, 229–241.

Pullen, S.S., Friesen, P.D., **1995**. Early transcription of the *ie-1* transregulator gene of *Autographa californica* nuclear polyhedrosis virus is regulated by DNA sequences within its 5′ noncoding leader region. *J. Virol. 69*, 156–165.

Ragunath, C., Sundar, K., Ramasubbu, N., **2002**. Expression characterization and biochemical properties of recombinant human salivary amylase. *Protein Expr. Purif. 24*, 202–211.

Rankin, C., Ooi, B.G., Miller, L.K., **1988**. Eight base pairs encompassing the transcriptional start point are the major determinant for baculovirus polyhedrin gene expression. *Gene 70*, 39–49.

Rankl, N.B., Rice, J.W., Gurganus, T.M., Barbee, J.L., Burns, D.J., **1994**. The production of an active protein kinase C-delta in insect cells is greatly enhanced by the use of the basic protein promoter. *Protein Expr. Purif. 5*, 346–356.

Rapp, J.C., Wilson, J.A., Miller, L.K., **1998**. Nineteen baculovirus open reading frames including LEF-12, support late gene expression. *J. Virol. 72*, 10197–10206.

Ren, Y., Busch, R., Durban, E., Taylor, C., Gustafson, W.C., *et al.*, **1996**. Overexpression of human nucleolar proteins in insect cells: characterization of nucleolar protein p120. *Protein Expr. Purif. 7*, 212–219.

Rodems, S.M., Friesen, P.D., **1993**. The *hr5* transcriptional enhancer stimulates early expression from the *Autographica californica* nuclear polyhedrosis virus genome but is not required for virus replication. *J. Virol. 67*, 5776–5785.

Roy, P., Mikhailov, M., Bishop, D.H.L., **1997**. Baculovirus multigene expression vectors and their use for understanding the assembly process of architecturally complex virus particles. *Gene 190*, 119–127.

Royer, C., Bossin, H., Romane, C., Bergoin, M., Couble, P., **2001**. High amplification of a densovirus-derived vector in larval and adult tissues of *Drosophila*. *Insect Mol. Biol. 10*, 275–280.

Russell, R.L., Rohrmann, G.F., **1993**. Nucleotide sequence of the ubiquitin-39K gene region from the *Orgyia pseudotsugata* multinucleocapsid nuclear polyhedrosis virus genome. *J. Gen. Virol. 74*, 1191–1195.

Sadatmansoori, S., MacDougall, J., Khademi, S., Cooke, L.S., Guarino, L., *et al.*, **2001**. Construction expression and characterization of a baculovirally expressed catalytic domain of human matrix metalloproteinase-9. *Protein Expr. Purif. 23*, 447–452.

Sarvari, M., Csikos, G., Sass, M., Gal, P., Schumaker, V.N., *et al.*, **1990**. Ecdysteroids increase the yield of recombinant protein produced in baculovirus insect cell expression system. *Biochem. Biophys. Res. Commun. 30*, 1154–1161.

Schneider, I., **1972**. Cell lines derived fromlate embryonic stages of *Drosophila melanogaster*. *J. Embryol. Exp. Morphol. 27*, 353–356.

Sehgal, D., Malik, P.S., Jameel, S., **2003**. Purification and diagnostic utility of a recombinant hepatitis E virus capsid protein expressed in insect larvae. *Protein Expr. Purif. 27*, 27–34.

Seo, N.-S., Hollister, J.R., Jarvis, D.L., **2001**. Mammalian glycotransferases expression allows sialoglycoprotein production by baculovirus-infected insect cells. *Protein Expr. Purif. 22*, 234–241.

Shikata, M., Sano, Y., Hashimoto, Y., Matsumoto, T., **1998**. Isolation and characterization of a temperature-sensitive mutant of *Bombyx mori* nucleopolyhedrovirus for a putative RNA polymerase gene. *J. Gen. Virol. 79*, 2071–2078.

Shin, H.S., Cha, H.J., **2003**. Statistical optimization for immobilized metal affinity purification of secreted human erythropoietin from Drosophila S2 cells. *Protein Expr. Purif. 28*, 331–339.

Shoji, I., Aizaki, H., Tani, H., Ishii, K., Chiba, T., *et al.*, **1997**. Efficient gene transfer into various mammalian cells including non-hepatic cells by baculovirus vectors. *J. Gen. Virol. 78*, 2657–2664.

Shotkoski, F.A., Fallon, A.M., **1993**. The musquito dihydrofolate reductase gene functions as a dominant selectable marker in transfected cells. *Insect Biochem. Mol. Biol. 23*, 883–893.

Shotkoski, F., Zhang, H.-G., Jackson, M.B., ffrench-Constant, R.H., **1996**. Stable expression of insect GABA receptors in insect cell lines: promoters for efficient expression of *Drosophila* and mosquito *Rdl* GABA receptors in stably transformed mosquito cell lines. *FEBS Lett. 380*, 257–262.

Smagghe, G., Braeckman, B.P., Huys, N., Tirry, L., Raes, H., **2003**. Cultured mosquito cells *Aedes albopictus* C6/36 (Dip., Culicidae) responsive to 20-hydroxyecdysone and non-steroidal ecdysone agonist. *J. Appl. Ent. 127*, 1–7.

Smith, C.D., Hirayama, B.A., Wright, E.M., **1992**. Baculovirus-mediated expression of the Na^+/glucose cotransporter in Sf9 cells. *Biochim. Biophys. Acta 1104*, 151–159.

Smith, G.E., Fraser, M.J., Summers, M.D., **1983c**. Molecular engineering of the *Autographa californica* nuclear polyhedrosis virus genome deletion mutations within the polyhedrin gene. *J. Virol. 46*, 584–593.

Smith, G.E., Ju, G., Ericson, B.L., Moschera, J., Lahm, H.-W., *et al.*, **1985**. Modification and secretion of

human interleukin 2 produced in insect cells by a baculovirus expression vector. *Proc. Natl Acad. Sci. USA 82*, 8404–8408.

Smith, G.E., Summers, M.D., Fraser, M.J., **1983b**. Production of human beta interferon in insect cells infected with a baculovirus expression vector. *Mol. Cell. Biol. 3*, 2156–2165.

Smith, G.E., Vlak, J.M., Summers, M.D., **1983a**. Physical analysis of *Autographa californica* nuclear polyhedrosis virus transcripts for polyhedrin and 10,000-molecular-weight protein. *J. Virol. 45*, 215–225.

Smith, L.A., Amar, M., Harvey, R.J., Darlison, M.G., Earley, F.G.P., *et al.*, **1995**. The production of a stably transformed insect cell line expressing an invertebrate GABAA receptor β-subunit. *J. Recep. Signal Transduct. Res. 15*, 33–42.

Soares, M., Hanisch, F.-G., Finn, O.J., Ciborowski, P., **2001**. Recombinant human tumor antigen MUC1 expressed in insect cells structure and immunogenicity. *Protein Expr. Purif. 22*, 92–100.

Sohi, S.S., Palli, S.R., Cook, B.J., Retnakaran, A., **1995**. Forest insect cell lines responsive to 20-hydroxyecdysone and two nonsteroidal ecdysone agonists, RH-5849 and RH-5992. *J. Insect Physiol. 41*, 457–464.

Sollerbrant, K., Elmen, J., Wahlestedt, C., Acker, J., Leblois-Prehaud, H., *et al.*, **2001**. A novel method using baculovirus-mediated gene transfer for production of recombinant adeno-associated virus vectors. *J. Gen. Virol. 82*, 2051–2060.

Sorbrier, M.L., Chapel, S., Couderc, J.L., Micard, D., Lecher, P., *et al.*, **1989**. 20-OH-ecdysone regulates 60 C beta tubulin gene expression in Kc cells and during *Drosophila* development. *Exp. Cell Res. 184*, 241–249.

Sorci-Thomas, M.G., Parks, J.S., Kearns, M.W., Pate, G.N., Zhang, C., *et al.*, **1996**. High level secretion of wild-type and mutant forms of human proapoA-I using baculovirus-mediated Sf-9 cell expression. *J. Lipid Res. 37*, 673–683.

Spradling, A.C., **1986**. P element-mediated transformation. In: Roberts, D.B. (Ed.), *Drosophila*: A Practical Approach. IRL Press, Oxford, pp. 175–197.

Stables, J., Green, A., Marshall, F., Fraser, N., Knight, E., *et al.*, **1997**. A bioluminescent assay for agonist activity at potentially any G-protein-coupled receptor. *Anal. Biochem. 252*, 115–126.

Staubli, F., Jørgensen, T.J.D., Cazzamali, G., Williamson, M., Lenz, C., *et al.*, **2002**. Molecular identification of the insect adipokinetic hormone receptors. *Proc. Natl Acad. Sci. USA 99*, 3446–3451.

Steed, P.M., Lasala, D., Liebman, J., Wigg, A., Clark, K., *et al.*, **1998**. Characterization of recombinant human cathespin B expressed at high levels in baculovirus. *Protein Science 7*, 2033–2037.

Steiner, H., Pohl, G., Gunne, H., Hellers, M., Elhammer, A, *et al.*, **1988**. Human tissue-type plasminogen activator synthesized by using a baculovirus vector in insect cells compared with human plasminogen activator produced in mouse cells. *Gene 73*, 449–457.

Steller, H., Pirrotta, V., **1984**. Regulated expression of genes injected into early *Drosophila* embryos. *EMBO J. 3*, 165–173.

Stoltenberg, J.K., Straney, R.A., Tritch, R.J., Mackin, W.M., George, H.J., **1993**. Expression and functional characterization of recombinant human vascular cell adhesion molecule-1 (VCAM1) synthesized by baculovirus-infected insect cells. *Protein Expr. Purif. 4*, 585–593.

Strosberg, D., Guillaume, J.-L., **2000**. Expression of β-andrenergic receptors in recombinant baculovirus-infected insect cells. *Methods Mol. Biol. 126*, 207–214.

Sugiura, T., Takamatsu, H., Kudo, A., Amann, E., **1995**. Expression and characterization of murine osteoblast-specific factor 2 (OSF-2) in a baculovirus expression system. *Protein Expr. Purif. 6*, 305–311.

Summers, M.D., Smith, G.G., **1987**. A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures. Texas Agricultural Experiment Station Bulletin 155.

Sun, S., Albright, C.F., Fish, B.H., George, H.J., Selling, B.H., *et al.*, **2002**. Expression purification and kinetic characterization of full-length human fibroblast activation protein. *Protein Expr. Purif. 24*, 274–281.

Suzuki, T., Kawasaki, H., Yu, R.T., Ueda, H., Umesono, K., **2001**. Segmentation gene product Fushi tarazu is an LXXLL motif-dependent coactivator for orphan receptor FTZ-F1. *Proc. Natl Acad. Sci. USA 98*, 12403–12408.

Sviridov, D., Luong, A., Pyle, L., Fidge, N., **1999**. Effectivity of expression of mature forms of mutant human apolipoprotein A-I. *Protein Expr. Purif. 17*, 231–238.

Swevers, L., Cherbas, L., Cherbas, P., Iatrou, K., **1996**. *Bombyx* EcR (BmEcR) and *Bombyx* USP (BmCF1) combine to form a functional ecdysone receptor. *Insect Biochem. Molec. Biol. 26*, 217–221.

Swevers, L., Farrell, P.J., Kravariti, L., Xenou-Kokoletsi, M., Sdralia, N., *et al.*, **2003a**. Transformed insect cells as high throughput screening tools for the discovery of new bioactive compounds. *Comm. Agric. Appl. Biol. Sci. 68/2(b)*, 333–341.

Swevers, L., Ito, K., Iatrou, K., **2002**. The BmE75 nuclear receptors function as dominant repressors of the nuclear receptor BmHR3A. *J. Biol. Chem. 277*, 41637–41644.

Swevers, L., Kravariti, L., Ciolfi, S., Xenou-Kokoletsi, M., Ragoussis, N., *et al.*, **2003b**. A cell-based high-throughput screening system for detecting ecdysteroid agonists and antagonists in plant extracts and libraries of synthetic compounds. *FASEB J.* Express Article 10.1096/fj.03-0627fje.

Szerenscei, R.T., Tucker, J.E., Cooper, C.C., Winkfein, R.J., Farrell, P.J., *et al.*, **2000**. Minimal domain requirement for cation transport by the potassium-dependant Na–Ca+K exchanger: comparison with a NCKX homologue from *Caenorhabditis elegans*. *J. Biol. Chem. 275*, 669–676.

Tate, C.G., **1998**. Baculovirus-mediated expression of neurotransmitter transporters. *Methods Enzymol. 296*, 443–455.

Tate, C.G., **1999**. Molecular chaperones stimulate the functional expression of the cocaine-sensitive serotonin transporter. *J. Biol. Chem. 274*, 17551–17558.

Tessier, D.C., Thomas, D.Y., Khouri, H.E., Laliberte, F., Vernet, T., **1991**. Enhanced secretion from insect cells of a foreign protein fused to the honeybee melittin signal peptide. *Gene 98*, 177–183.

Theilmann, D.A., Stewart, S., **1992**. Tandemly repeated sequence at the 3′ end of the IE-2 gene of the baculovirus *Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus is an enhancer element. *Virology 187*, 97–106.

Thomas, H.E., Stunnenberg, H.G., Stewart, A.F., **1993**. Heterodimerization of the *Drosophila* ecdysone receptor with retinoid X receptor and ultraspiracle. *Nature 362*, 471–475.

Thordarson, G., Wu, K., Talamantes, F., **1996**. Purification and characterization or recombinant mouse growth hormone binding protein produced in the baculovirus expression system. *Protein Expr. Purif. 7*, 74–80.

Thummel, C.S., Pirrotta, V., **1992**. New pCaSpeR P-element vectors. *Drosopila Inf. Serv. 71*, 150.

Tomita, S., Kawai, Y., Woo, S.D., Kamimura, M., Iwabuichi, K., *et al.*, **2001**. Ecdysone-inducible foreign gene expression in stably-transformed lepidopteran insect cells. *In Vitro Cell Dev. Biol. Animal 37*, 564–571.

Tomlinson, S., Ueda, E., Maruniak, J.E., Garcia-Canedo, A., Bjes, E.S., *et al.*, **1993**. The expression of hemolytically active human complement protein C9 in mammalian insect and yeast cells. *Protein Expr. Purif. 4*, 141–148.

Torfs, H., Poels, J., Detheux, M., Dupriez, V., Van Loy, T., *et al.*, **2002**. Recombinant aequorin as a reporter for receptor-mediated changes of intracellular Ca^{2+}-levels in *Drosophila* S2 cells. Invert. *Neurosci. 4*, 119–124.

Torfs, H., Shariatmadari, R., Guerrero, F., Parmentier, M., Poels, J., *et al.*, **2000**. Characterization of a receptor for insect tachykinin-like peptide agonists by functional expression in a stable *Drosophila Schneider* 2 cell line. *J. Neurochem. 74*, 2182–2189.

Tota, M.R., Sirotina, A., Strader, C.D., Graziano, M.P., **1995**. Interaction of [fluorescein-Trp^{25}]glucagons with the human glucagons receptor expressed in *Drosophila* Schneider 2 cells. *J. Biol. Chem. 270*, 26466–26472.

Tsai, C.C., Kao, H.Y., Yao, T.P., McKeown, M., Evans, R.M., **1999**. SMRTER, a *Drosophila* nuclear receptor coregulator, reveals that EcR-mediated repression is critical for development. *Mol Cell. 4*, 175–186.

Tschantz, W.R., Zhang, L., Casey, P.J., **1999**. Cloning, expression, and cellular localization of a human prenylcysteine lyase. *J. Biol. Chem. 274*, 35802–35808.

Urakawa, T., Ferguson, M., Minor, P.D., Cooper, J., Sullivan, M., *et al.*, **1989**. Synthesis of immunogenic but non-infectious poliovirus particles in insect cells by a baculovirus expression vector. *J. Gen. Virol. 70*, 1453–1463.

Vanden Broeck, J., **2001**. Insect G protein-coupled receptors and signal transduction. *Arch. Insect Biochem. Physiol. 48*, 1–12.

Vanden Broeck, J., Vulsteke, V., Huybrechts, R., De Loof, A., **1995**. Characterization of a cloned locust tyramine receptor cDNA by functional expression in permanently transformed *Drosophila* S2 cells. *J. Neurochem. 64*, 2387–2395.

van Oers, M.M., Flipsen, J.T., Reusken, C.B., Sliwinsky, E.L., Goldbach, R.W., *et al.*, **1993**. Functional domains of the p10 protein of *Autographa californica* nuclear polyhedrosis virus. *J. Gen. Virol. 74*, 563–574.

Vasudevan, S., Hulme, E.C., Bach, M., Haase, W., Pavia, J., *et al.*, **1995**. Characterization of the rat m3 muscarinic acetylcholine receptor produced in insect cells infected with recombinant baculovirus. *Eur. J. Biochem. 227*, 466–475.

Vasudevan, S., Premkumar, L., Stowe, S., Gage, P.W., Reilander, H., *et al.*, **1992**. Muscarinic acetylcholine receptor produced in recombinant baculovirus infected Sf9 insect cells couples with endogenous G-proteins to activate ion channels. *FEBS Lett. 311*, 7–11.

Vaughn, J.L., Goodwin, R.H., Tompkins, G.J., McCawley, P., **1977**. The establishment of two cell lines from the insect *Spodoptera frugiperda* (Lepidoptera: Noctuidae). *In Vitro 13*, 213–217.

Vialard, J.E., Arif, B.M., Richardson, C.D., **1995**. Introduction to the molecular biology of baculoviruses. *Methods Mol. Biol. 39*, 1–24.

Vlak, J.M., Klinkenberg, F.A., Zaal, K.J.M., Usmany, N., Klinge-Roode, E.C., *et al.*, **1988**. Functional studies on the p10 gene of *Autographa californica* nuclear polyhedrosis virus using a recombinant expression of a p10-B-galactosidase fusion gene. *J. Gen. Virol. 69*, 765–776.

Vogtli, M., Elke, C., Imhof, M.O., Lezzi, M., **1998**. High level transactivation by the ecdysone receptor complex at the core recognition motif. *Nucleic Acids Res. 26*, 2407–2414.

Volkman, L.E., Goldsmith, P.A., **1983**. In-vitro survey of *Autographa-californica* nuclear polyhedrosis virus interaction with nontarget vertebrate host cells. *Appl. Environ. Microbiol. 45*, 1085–1093.

Volkman, L.E., Summers, M.D., **1977**. Autographa californica nuclear polyhedrosis virus: comparative infectivity of the occluded, alkali-liberated, and nonoccluded forms. *J Invertebr Pathol. 30*, 102–103.

Wagner, R., Liedtke, S., Kretzschmar, E., Geyer, H., Geyer, R., *et al.*, **1996**. Elongation of the N-glycans of fowl plague virus hemagglutinin expressed in *Spodoptera frugiperda* (Sf9) cells by coexpression of human beta 1,2-N-acetylglucosaminyltransferase I. *Glycobiology 6*, 165–175.

Wang, Y., Jiang, H., Kanost, M.R., **2001**. Expression and purification of *Manduca sexta* prophenoloxidase-activating proteinase precursor (proPAP) from baculovirus-infected insect cells. 2001. *Protein Expr. Purif. 22*, 328–337.

Wang, S.F., Miura, K., Miksicek, R.J., Segraves, W.A., Raikhel, A.S., **1998**. DNA binding and transactivation characteristics of the mosquito ecdysone receptor-Ultraspiracle complex. *J. Biol. Chem. 273*, 27531–27540.

Wang, Y., Kaguni, L.S., **1999**. Baculovirus expression reconstitutes *Drosophila* mitochondrial DNA polymerase. *J. Biol. Chem. 274*, 28972–28977.

Wehmeyer, A., Schulz, R., **1997**. Overexpression of δ-opioid receptors in recombinant baculovirus-infected *Trichoplusia ni* "High 5" insect cells. *J. Neurochem. 68*, 1361–1371.

Wester, L., Johansson, M.U., Alerstrom, B., **1997**. Physiochemical and biochemical characterization of human α_1-microglobulin expressed in baculovirus-infected insect cells. *Protein Expr. Purif. 11*, 95–103.

Weyer, U., Knight, S., Possee, R.D., **1990**. Analysis of very late gene expression by *Autographa californica* nuclear polyhedrosis virus and the further development of multiple expression vectros. *J. Gen. Virol. 71*, 1525–1534.

Whitford, M., Faulkner, P., **1992**. Nucleotide sequence and transcriptional analysis of a gene encoding p41, a structural glycoprotein of the baculovirus *Autographa californica* nuclear polyhedrosis virus. *J. Virol. 66*, 4763–4768.

Whitford, M., Stewart, S., Kuzio, J., Faulkner, P., **1989**. Identification and sequence analysis of a gene encoding gp67 and abundant envelope glycoprotein of the baculovirus *Autographa californica* nuclear polyhedrosis virus. *J. Virol. 63*, 1393–1399.

Wickham, T.J., Davis, T., Granados, R.R., Shuler, M., Wood, H.A., **1992**. Screening of insect cell lines for the production of recombinant proteins and infectious virus in the baculovirus expression system. *Biotechnol. Progr. 8*, 391–396.

Willard, D., Chen, W.-J., Barrett, G., Blackburn, K., Bynum, J., *et al.*, **2000**. Expression purification and characterization of the human receptor activator of NF-κB ligand (RANKL) extracellular domain. *Protein Expr. Purif. 20*, 48–57.

Williams, G.V., Rohel, D.Z., Kuzio, J., Faulkner, P., **1989**. A cytopathological investigation of *Autographa californica* nuclear polyhedrosis virus p10 gene function using insertion/deletion mutants. *J. Gen. Virol. 70*, 187–202.

Winzerling, J.J., Pham, D.Q.-D., Kunz, S., Samaraweera, P., Law, J.H., *et al.*, **1996**. Purification of recombinant insect transferrin from large volumes of cell culture medium using high capacity Ni^{2+}-dipicolylamine gel. *Protein Expr. Purif. 7*, 137–142.

Withers, B.E., Keller, P.R., Fry, D.W., **1996**. Expression, purification and characterization of focal adhesion kinase using a baculovirus system. *Protein Expr. Purif. 7*, 12–18.

Wu, X., Kamei, K., Sato, H., Sato, S.I., Takano, R., *et al.*, **2001**. High-level expression of human acidic fibroblast growth factor and basic fibroblast growth factor in silkworm (*Bombyx mori* L.) using recombinant baculovirus. *Protein Expr. Purif. 21*, 192–200.

Xiao, Z., Jiang, X., Beckett, M.L., Wright, G.L., Jr., **2000**. Generation of a baculovirus recombinant prostrate-specific membrane antigen and its use in the development of a novel protein biochip quantitative immunoassay. *Protein Expr. Purif. 19*, 12–21.

Yamanaka, T., Kuroki, M., Kinugasa, T., Matsuo, Y., Matsuoka, Y., **1996**. Preparation and characterization of two human carcinoembryonic antigen family proteins of neutrophils CD66b and c in silkworm larvae. *Protein Expr. Purif. 7*, 438–446.

Yamao, M., Katayama, N., Nakazawa, H., Yamakawa, M., Hayashi, Y., *et al.*, **1999**. Gene targeting in the silkworm by use of a baculovirus. *Genes Develop. 13*, 511–516.

Yang, S., Miller, L.K., **1999**. Activation of baculovirus very late promoters by interaction with very late factor 1. *J. Virol. 73*, 3404–3409.

Yao, T.-P., Forman, B.M., Jiang, Z., Cherbas, L., Chen, J.-D., *et al.*, **1993**. Functional ecdysone receptor is the product of EcR and Ultraspiracle genes. *Nature 366*, 476–479.

Yap, C.C., Ishii, K., Aoki, Y., Aizaki, H., Tani, H., *et al.*, **1997**. A hybrid baculovirus-T7 RNA polymerase system for recovery of an infectious virus from cDNA. *Virology 231*, 192–200.

Yet, F.-S., Moon, Y.K., Sul, H.S., **1995**. Purification and reconstitution of murine mitochondrial glycerol-3-phosphate acyltransferase. Functional expression in baculovirus-infected insect cells. *Biochemistry 34*, 7303–7310.

Zdobnov, E.M., von Mering, C., Letunic, I., Torrents, D., Suyama, M., *et al.*, **2002**. Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster. Science 298*, 149–159.

Zhang, B., Egli, D., Georgiev, O., Schaffner, W., **2001**. The *Drosophila* homolog of mammalian zinc finger factor MTF-1 activates transcription in response to heavy metals. *Mol. Cell. Biol. 21*, 4505–4514.

Zhang, J., Alfonso, P., Thotakura, N.R., Su, J., Buergin, M., *et al.*, **1998**. Expression purification and bioassay of human stanniocalcin from baculovirus-infected insect cells and recombinant CHO cells. *Protein Expr. Purif. 12*, 390–398.

Zhang, J., Kalogerakis, N., Behie, L.A., Iatrou, K., **1993**. A two-stage bioreactor system for the production of recombinant proteins using a genetically engineered baculovirus/insect cell system. *Biotechnol. Bioeng. 42*, 357–366.

Zhang, R., Cai, H., Fatima, N., Buczko, E., Dufau, M.L., **1995**. Functional glycosylation sites of the rat luteinizing hormone receptor required for ligand binding. *J. Biol. Chem. 270*, 21722–21728.

Zhao, Y.-G., Eggleston, P., **1999**. Comparative analysis of promoters for transient gene expression in cultured mosquito cells. *Insect Mol. Biol. 8*, 31–38.

zu Putlitz, J., Kubasek, W.L., Duchene, M., Marget, M., von Specht, B.U., *et al.*, **1990**. Antibody production in baculovirus-infected insect cells. *Biotechnology. 8*, 651–654.