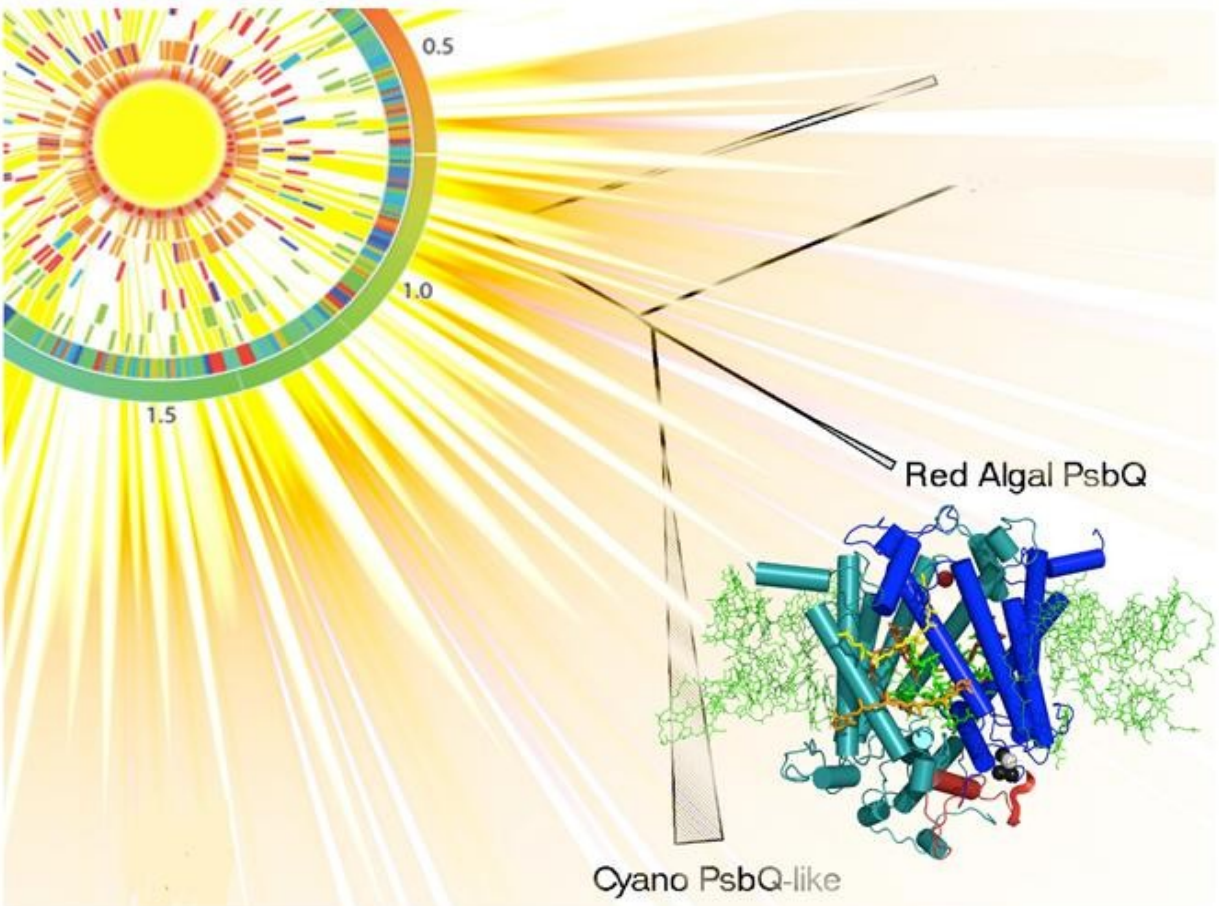
Advances in Photosynthesis and Respiration

Volume 33

Functional Genomics and Evolution of Photosynthetic Systems



Edited by

Robert L. Burnap

and

Willem F.J. Vermaas

Springer

Functional Genomics and Evolution of Photosynthetic Systems

Advances in Photosynthesis and Respiration

VOLUME 33

The scope of our series reflects the concept that photosynthesis and respiration are intertwined with respect to both the protein complexes involved and to the entire bioenergetic machinery of all life. *Advances in Photosynthesis and Respiration* is a book series that provides a comprehensive and state-of-the-art account of research in photosynthesis and respiration. Photosynthesis is the process by which higher plants, algae, and certain species of bacteria transform and store solar energy in the form of energy-rich organic molecules. These compounds are in turn used as the energy source for all growth and reproduction in these and almost all other organisms. As such, virtually all life on the planet ultimately depends on photosynthetic energy conversion. Respiration, which occurs in mitochondrial and bacterial membranes, utilizes energy present in organic molecules to fuel a wide range of metabolic reactions critical for cell growth and development. In addition, many photosynthetic organisms engage in energetically wasteful photorespiration that begins in the chloroplast with an oxygenation reaction catalyzed by the same enzyme responsible for capturing carbon dioxide in photosynthesis. This series of books spans topics from physics to agronomy and medicine, from femtosecond processes to season-long production, from the photophysics of reaction centers, through the electrochemistry of intermediate electron transfer, to the physiology of whole organisms, and from X-ray crystallography of proteins to the morphology of organelles and intact organisms. The goal of the series is to offer beginning researchers, advanced undergraduate students, graduate students, and even research specialists, a comprehensive, up-to-date picture of the remarkable advances across the full scope of research on photosynthesis, respiration and related processes.

For further volumes:
http://www.springer.com/series/5599

Functional Genomics and Evolution of Photosynthetic Systems

Edited by

Robert L. Burnap
Oklahoma State University
Stillwater
Oklahoma
USA

and

Willem F.J. Vermaas
Arizona State University
Tempe
Arizona
USA

Springer

Editors
Robert L. Burnap
Department of Microbiology & Molecular Genetics
307 Life Sciences East
Oklahoma State University
Stillwater, Oklahoma 74078
USA
rob.burnap@okstate.edu

Willem F.J. Vermaas
School of Life Sciences and Center for Bioenergy & Photosynthesis
Arizona State University
Box 874501
Tempe, Arizona 85287-4501
USA
wim@asu.edu

ISSN 1572-0233
ISBN 978-94-007-1532-5 (HB) ISBN 978-94-007-1533-2 (e-book)
DOI 10.1007/978-94-007-1533-2
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2011933555

Dedication

Teruo Ogawa

The editors dedicate this volume in the series *Advances in Photosynthesis and Respiration* to **Teruo Ogawa**, who has made numerous original contributions to the field of photosynthesis. Teruo began his graduate studies in the laboratory of Professor Kazuo Shibata in the early 1960s, at the Tokyo Institute of Technology. At that time, techniques for the biochemical separation of chlorophyll-protein complexes using detergents began to flourish. Teruo described the first electrophoretic separation of chlorophyll-protein complexes using sodium dodecyl sulfate. This and the concurrent work of Philip Thornber solidified the concept that chlorophyll is coordinated by protein in the photosynthetic membrane and marked a signal point in the path that has ultimately led to the high resolution crystal structures of the photosystems and light-harvesting complexes. Teruo continued this line of investigation joining the lab of Professor Leo Vernon as a postdoc at the Charles F. Kettering Research Laboratory in Yellow Springs, Ohio, working on the isolation of photosystem particles from cyanobacteria. Teruo rejoined Professor Shibata, who had become Head of the Plant Physiology Lab at RIKEN (The Institute of Physical and Chemical Research) to study the mechanism of stomatal movements. With the subsequent formation of a new unit at RIKEN, The Solar Energy Research Group, Teruo assumed a role as Associate Director and began what would become years of successful exploitation of cyanobacterial molecular genetics to understand photosynthetic function.

It was this effective utilization of molecular genetics, and the genomic sequence data of cyanobacteria that deserves special recognition in the context of the present volume. In 1994 Teruo was invited to become a Professor at Nagoya University, where he continued this genetics-based, soon to be genomics-based, understanding of photosynthetic physiology. An indicator of the prolific laboratory output, much of it personally executed, can be found at the Kazusa Institute's genomics website in the catalog of mutants registered by researchers around the world – the overwhelming majority were produced by Teruo and are archived as frozen stocks in the lab of Prof. Masahiko Ikeuchi at the University of Tokyo. The work was fruitful, especially in the area of defining the components of the inorganic carbon concentration mechanism. Overall, the research identified and functionally characterized the role of genes encoding (i) functionally distinct NADPH dehydrogenases, (ii) an iron transporter, (iii) a low-Mn sensing mechanism, (iv) two types of CO_2-uptake systems, and (v) genes encoding a sodium-dependent bicarbonate transporter. Finding functionally distinct NADPH dehydrogenases has led

to a new concept regarding their roles, which have been traditionally associated with respiration, yet now appear to exist in diverse structural variations that provide functions ranging from carbon dioxide hydration to sodium pumping. Teruo 'retired' from his Professorship at Nagoya in 2003, but this only marked a change in the experimental modus operarandi, at least from the outsiders' perspective. He continued research abroad and also in Ikeuchi's lab at Tokyo University. He has donned lab coats in Professor Himadri Pakrasi's lab at Washington University, Professor Eva-Mari Aro's lab at the University of Turku, Dr. Hualing Mi's lab at the Shanghai Institute of Plant Physiology and Ecology and in Professor Matthias Rögner's lab at the Ruhr University. Genetic and proteomic studies on the mutants of *Synechocystis* sp. PCC 6803 and *Thermosynechococcus elongatus* with the people in these laboratories advanced the understanding of the structure and function of NDH-1 complexes and the carbon concentrating mechanism. One can count more than sixteen peer-reviewed articles, as well as a number of articles such as book chapters and reviews, all emerging during this 8 year period of 'retired' status.

From the Series Editors

Advances in Photosynthesis and Respiration

Volume 33: Functional Genomics and Evolution of Photosynthetic Systems

We (Tom Sharkey and I) are delighted to announce the publication, in the Advances in Photosynthesis and Respiration (AIPH) Series, of *Functional Genomics and Evolution of Photosynthetic Systems*. Robert (Rob) L. Burnap (of the Department of Microbiology and Molecular Genetics of the Oklahoma State University) and Willem (Wim) F.J. Vermaas (of the School of Life Sciences and the Center for Bioenergy and Photosynthesis of Arizona State University), two international authorities of Molecular Biology of Photosynthesis, have edited this volume. I have personally known both Rob and Wim for many years as scientists of remarkable insight into the field of photosynthesis. Rob is a pioneer in the studies of the metabolic signals resulting from the light reactions and how these signals control the expression and activity of the cyanobacterial carbon-concentrating mechanism. Wim is a pioneer of functional genomics. His major research focus is to understand the physiology of cyanobacteria based not only on genomic, but also on biochemical and biophysical information. His research leads him from *genome* through *proteome* to *metabolome*.

Our Books: 32 Volumes

We list below information on all the 32 volumes that have been published thus far. Beginning with volume 31, Thomas D. Sharkey, who had earlier co-edited volume 9 (Photosynthesis: Physiology and Metabolism) of this Series, has joined me as a co-Series Editor. [Note: Another book 'Photosynthesis: Plastid Biology, Energy Conversion and Carbon Assimilation, edited by Julian J. Eaton-Rye, Baishnab C. Tripathy, and Thomas D. Sharkey', volume 34 in the series, is in the press.]

We are pleased to note that Springer, our publisher, is now producing complete tables of contents of these books and electronic copies of individual chapters of these books; their web sites include free downloadable front matter as well as indexes. The available and anticipated web sites of the books in the Series are listed below. This volume and volume 34 will be the last two with the familiar white cover. A green cover better suited to the increasing web presence will be used for volume 35, which will be published early 2012. The name of the series will also be updated to Advances in Photosynthesis and Bioenergy.

- **Volume 32 (2010): C4 Photosynthesis and Related CO_2 Concentrating Mechanisms**, edited by Agepati S. Raghavendra, and Rowan Sage, from India and Canada. Nineteen chapters, 410 pp, Hardcover, ISBN 978-90-481-9406-3 [http://www.springerlink.com/content/978-90-481-9406-3/]
- **Volume 31 (2010): The Chloroplast: Basics and Applications**, edited by Constantin Rebeiz (USA), Christoph Benning (USA), Hans J. Bohnert (USA), Henry Daniell (USA), J. Kenneth Hoober (USA), Hartmut K. Lichtenthaler (Germany), Archie R. Portis (USA), and Baishnab C. Tripathy (India). Twenty-five chapters, 421 pp, Hard cover, ISBN: 978-90-481-8530-6 [http://www.springerlink.com/content/978-90-481-8530-6/]
- **Volume 30 (2009): Lipids in Photosynthesis: Essential and Regulatory Functions**, edited by Hajime Wada and Norio Murata, both from Japan. Twenty chapters, 506 pp, Hardcover, ISBN: 978-90-481-2862-4;e-book,ISBN:978-90-481-2863-1[http://www.springerlink.com/content/978-90-481-2862-4/]
- **Volume 29 (2009): Photosynthesis in Silico: Understanding Complexity from Molecules**, edited by Agu Laisk, Ladislav Nedbal, and Govindjee, from Estonia, The Czech Republic, and

USA. Twenty chapters, 508 pp, Hardcover, ISBN: 978-1-4020-9236-7 [http://www.springerlink.com/content/978-1-4020-9236-7/]

- **Volume 28 (2009): The Purple Phototrophic Bacteria**, edited by C. Neil Hunter, Fevzi Daldal, Marion C. Thurnauer and J. Thomas Beatty, from UK, USA and Canada. Forty-eight chapters, 1014 pp, Hardcover, ISBN: 978-1-4020-8814-8 [http://www.springerlink.com/content/978-1-4020-8814-8/]
- **Volume 27 (2008): Sulfur Metabolism in Phototrophic Organisms**, edited by Christiane Dahl, Rüdiger Hell, David Knaff and Thomas Leustek, from Germany and USA. Twenty-four chapters, 551 pp, Hardcover, ISBN: 978-4020-6862-1 [http://www.springerlink.com/content/978-1-4020-6862-1/]
- **Volume 26 (2008): Biophysical Techniques in Photosynthesis**, Volume II, edited by Thijs Aartsma and Jörg Matysik, both from The Netherlands. Twenty-four chapters, 548 pp, Hardcover, ISBN: 978-1-4020-8249-8 [http://www.springerlink.com/content/978-1-4020-8249-8/]
- **Volume 25 (2006): Chlorophylls and Bacteriochlorophylls: Biochemistry, Biophysics, Functions and Applications**, edited by Bernhard Grimm, Robert J. Porra, Wolfhart Rüdiger, and Hugo Scheer, from Germany and Australia. Thirty-seven chapters, 603 pp, Hardcover, ISBN: 978-1-40204515-8 [http://www.springerlink.com/content/978-1-4020-4515-8]
- **Volume 24 (2006): Photosystem I: The Light-Driven Plastocyanin: Ferredoxin Oxidoreductase**, edited by John H. Golbeck, from USA. Forty chapters, 716 pp, Hardcover, ISBN: 978-1-40204255-3 [http://www.springerlink.com/content/978-1-4020-4255-3]
- **Volume 23 (2006): The Structure and Function of Plastids**, edited by Robert R. Wise and J. Kenneth Hoober, from USA. Twenty-seven chapters, 575 pp, Softcover, ISBN: 978-1-4020- 6570-6; Hardcover, ISBN: 978-1-4020-4060-3 [http://www.springerlink.com/content/978-1-4020-4060-3]
- **Volume 22 (2005): Photosystem II: The Light-Driven Water: Plastoquinone Oxidoreductase**, edited by Thomas J. Wydrzynski and Kimiyuki Satoh, from Australia and Japan. Thirty-four chapters, 786 pp, Hardcover, ISBN: 978-1-4020-4249-2 [http://www.springerlink.com/content/978-1-4020-4249-2]
- **Volume 21 (2005): Photoprotection, Photoinhibition, Gene Regulation, and Environment**, edited by Barbara Demmig-Adams, William W. Adams III and Autar K. Mattoo, from USA. Twenty-one chapters, 380 pp, Hardcover, ISBN: 978-14020-3564-7 [http://www.springerlink.com/content/978-1-4020-3564-7]
- **Volume 20 (2006): Discoveries in Photosynthesis**, edited by Govindjee, J. Thomas Beatty, Howard Gest and John F. Allen, from USA, Canada and UK. One hundred and eleven chapters, 1304 pp, Hardcover, ISBN: 978-1-4020-3323-0 [http://www.springerlink.com/content/978-1-4020-3323-0]
- **Volume 19 (2004): Chlorophyll *a* Fluorescence: A Signature of Photosynthesis**, edited by George C. Papageorgiou and Govindjee, from Greece and USA. Thirty-one chapters, 820 pp, Hardcover, ISBN: 978-1-4020-3217-2 [http://www.springerlink.com/content/978-1-4020-3217-2]
- **Volume 18 (2005): Plant Respiration: From Cell to Ecosystem**, edited by Hans Lambers and Miquel Ribas-Carbo, from Australia and Spain. Thirteen chapters, 250 pp, Hardcover, ISBN: 978-14020-3588-3 [http://www.springerlink.com/content/978-1-4020-3588-3]
- **Volume 17 (2004): Plant Mitochondria: From Genome to Function**, edited by David Day, A. Harvey Millar and James Whelan, from Australia. Fourteen chapters, 325 pp, Hardcover, ISBN: 978-1-4020-2399-6
- **Volume 16 (2004): Respiration in Archaea and Bacteria: Diversity of Prokaryotic Respiratory Systems**, edited by Davide Zannoni, from Italy. Thirteen chapters, 310 pp, Hardcover, ISBN: 978-14020-2002-5 [http://www.springerlink.com/content/978-1-4020-2002-5]
- **Volume 15 (2004): Respiration in Archaea and Bacteria: Diversity of Prokaryotic Electron Transport Carriers**, edited by Davide Zannoni, from Italy. Thirteen chapters, 350 pp, Hardcover, ISBN: 978-1-4020-2001-8 [http://www.springerlink.com/content/978-0-7923-2001-8/]
- **Volume 14 (2004): Photosynthesis in Algae**, edited by Anthony W. Larkum, Susan Douglas and John A. Raven, from Australia, Canada and UK. Nineteen chapters, 500 pp, Hardcover, ISBN: 978-0-7923-6333-0 [http://www.springerlink.com/content/978-0-7923-6333-0/]
- **Volume 13 (2003): Light-Harvesting Antennas in Photosynthesis**, edited by Beverley R. Green and William W. Parson, from Canada and USA. Seventeen chapters, 544 pp, Hardcover, ISBN: 978-07923-6335-4 [http://www.springerlink.com/content/978-0-7923-6335-4/]
- **Volume 12 (2003): Photosynthetic Nitrogen Assimilation and Associated Carbon and Respiratory Metabolism**, edited by Christine

H. Foyer and Graham Noctor, from UK and France. Sixteen chapters, 304 pp, Hardcover, ISBN: 978-07923-6336-1 [http://www.springer.com/life+sciences/plant+sciences/book/978-0-7923-6336-1]

- **Volume 11 (2001): Regulation of Photosynthesis**, edited by Eva-Mari Aro and Bertil Andersson, from Finland and Sweden. Thirty-two chapters, 640 pp, Hardcover, ISBN: 978-0-7923-6332-3 [http://www.springerlink.com/content/978-0-7923-6332-3]
- **Volume 10 (2001): Photosynthesis: Photobiochemistry and Photobiophysics**, authored by Bacon Ke, from USA. Thirty-six chapters, 792 pp, Softcover, ISBN: 978-0-7923-6791-8; Hardcover: ISBN: 978-0-7923-6334-7 [http://www.springerlink.com/content/978-0-7923-6334-7]
- **Volume 9 (2000): Photosynthesis: Physiology and Metabolism**, edited by Richard C. Leegood, Thomas D. Sharkey and Susanne von Caemmerer, from UK, USA and Australia. Twenty-four chapters, 644 pp, Hardcover, ISBN: 978-07923-6143-5 [http://www.springerlink.com/content/978-0-7923-6143-5]
- **Volume 8 (1999): The Photochemistry of Carotenoids**, edited by Harry A. Frank, Andrew J. Young, George Britton and Richard J. Cogdell, from UK and USA. Twenty chapters, 420 pp, Hardcover, ISBN: 978-0-7923-5942-5 [http://www.springerlink.com/content/978-0-7923-5942-5]
- **Volume 7 (1998): The Molecular Biology of Chloroplasts and Mitochondria in *Chlamydomonas***, edited by Jean David Rochaix, Michel Goldschmidt-Clermont and Sabeeha Merchant, from Switzerland and USA. Thirty-six chapters, 760 pp, Hardcover, ISBN: 978-0-7923-5174-0 [http://www.springerlink.com/content/978-0-7923-5174-0]
- **Volume 6 (1998): Lipids in Photosynthesis: Structure, Function and Genetics**, edited by Paul-André Siegenthaler and Norio Murata, from Switzerland and Japan. Fifteen chapters, 332 pp, Hardcover, ISBN: 978-0-7923-5173-3 [http://www.springerlink.com/content/978-0-7923-5173-3]
- **Volume 5 (1997): Photosynthesis and the Environment**, edited by Neil R. Baker, from UK. Twenty chapters, 508 pp, Hardcover, ISBN: 978- 07923-4316-5 [http://www.springerlink.com/content/978-0-7923-4316-5]
- **Volume 4 (1996): Oxygenic Photosynthesis: The Light Reactions**, edited by Donald R. Ort, and Charles F. Yocum, from USA. Thirty-four chapters, 696 pp, Softcover: ISBN: 978-0-7923-3684-6; Hardcover, ISBN: 978-0-7923-3683-9 [http://www.springerlink.com/content/978-0-7923-3683-9]
- **Volume 3 (1996): Biophysical Techniques in Photosynthesis**, edited by Jan Amesz and Arnold J. Hoff, from The Netherlands. Twenty-four chapters, 426 pp, Hardcover, ISBN: 978-0-7923-3642-6 [http://www.springerlink.com/content/978-0-7923-3642-6]
- **Volume 2 (1995): Anoxygenic Photosynthetic Bacteria**, edited by Robert E. Blankenship, Michael T. Madigan and Carl E. Bauer, from USA. Sixty-two chapters, 1331 pp, Hardcover, ISBN: 978-0-7923-3682-8 [http://www.springerlink.com/content/978-0-7923-3681-5]
- **Volume 1 (1994): The Molecular Biology of Cyanobacteria**, edited by Donald R. Bryant, from USA. Twenty-eight chapters, 916 pp, Hardcover, ISBN: 978-0-7923-3222-0 [http://www.springerlink.com/content/978-0-7923-3222-0/]

Further information on these books and ordering instructions can be found at http://www.springer.com/series/5599. Contents of volumes 1–29 can also be found at http://www.life.illinois.edu/govindjee/g/References.html.

Special 25% discounts are available to members of the International Society of Photosynthesis Research, ISPR http://www.photosynthesisresearch.org/: See http://www.springer.com/ispr

This Book

"*Functional Genomics and Evolution of Photosynthetic Systems*" is volume 33 of the Advances in Photosynthesis and Respiration Series. The preface of the book on pp. xxiii–xxv beautifully describes the context of this book, and the contents of this book on pp. xvii–xxi shows the breadth of this book.

According to the editors Rob Burnap and Wim Vermaas, "This book was inspired by the new possibilities brought about by the stunning number of genomic sequences that are currently, or will soon become, available for photosynthetic organisms. This new world of whole genome sequence data spans the phyla from photosynthetic microbes to algae to higher plants. These whole genome projects are intrinsically interesting, but also tell us about the variety of other molecular sequence databases including the recent 'meta-genomic' sequencing efforts that analyze entire communities of organisms. The fruits of these sequencing projects, as impressive as they are, are obviously

only the beginning of the effort to decipher the biological meaning encoded within them. This book highlights progress in this direction. It includes discussion of promising approaches in analyzing the wealth of sequence information and the resultant insights these analyses are providing regarding the function and evolution of photosynthesis. This book aims toward a genome-level understanding of the structure, function, and evolution of photosynthetic systems and the advantages accrued from the availability of diverse sets of gene sequences for the major components of the photosynthetic apparatus. It provides a good introduction to some of the many aspects of the genomics of photosynthetic systems in relation to the variety of photosynthetic mechanisms that have evolved. This book will serve both the established researchers and educators who wish to understand this rapidly developing area as well as young scientists starting their research career."

Authors

The current book contains 15 chapters written by 44 authors from seven countries (Canada; China; Denmark; Germany; Spain; Sweden and USA), with most authors from the USA. We thank all the authors for their valuable contribution to this book; their names (arranged alphabetically) are:

Iwona **A**damska (Germany; Chapter 11); Birgit E. Alber (USA; Chapter 9); Jens Appel (USA; Chapter 15); Shaun **B**ailey (USA; Chapter 6); J. Thomas Beatty (Canada; Chapter 10; co-editor of volumes 20 and 28 of this Series); Devaki Bhaya (USA; Chapter 2); Donald A. Bryant (USA; Chapter 1, Chapter 3; editor of volume 1 of this Series); Robert L. Burnap (USA; Chapter 13; co-editor of this volume); Fei **C**ai (USA; Chapter 14); You Chen (USA; Chapter 5); Wei Chi (China; Chapter 7); Frederick M. Cohan (USA; Chapter 1); Johannes **E**ngelken (Spain, Chapter 11); Niels-Ulrik **F**rigaard (Denmark; Chapter 3); Christiane Funk (Sweden; Chapter 11); Amaya M. **G**arcia-Costas (USA; Chapter 3); John H. Golbeck (USA; Chapter 12; editor of volume 24 of this Series); Susan S. Golden (USA; Chapter 5); David González-Ballester (USA; Chapter 6); Arthur R. Grossman (USA; Chapter 6); Thomas E. **H**anson (USA; Chapter 9); Caroline S. Harwood (USA; Chapter 10); C. Kay Holtman (USA; Chapter 5); Bharat **J**agannathan (USA; Chapter 12); Steven J. **K**arpowicz (USA; Chapter 6); Cheryl A. Kerfeld (USA; Chapter 14); Christian G. Klatt (USA; Chapter 1, Chapter 3); Andrew S. **L**ang (Canada; Chapter 10); Tao Li (USA; Chapter 3); Marc Linka (Germany; Chapter 8); Zhenfeng Liu (USA; Chapter 3); Sabeeha S. **M**erchant (USA; Chapter 6; coeditor of volume 7 of this Series); Aparna **N**agarajan (USA; Chapter 13); Jörg **O**vermann (Germany; Chapter 3); Brian **P**alenik (USA; Chapter 4); Gustaf **S**andh (USA; Chapter 14); Gaozhong Shen (USA; Chapter 12); F. Robert **T**abita (USA; Chapter 9); Arnaud Taton (USA; Chapter 5); David M. **W**ard (USA; Chapter 1, Chapter 3); Andreas P.M. Weber (Germany; Chapter 8); Jason Wood (USA; Chapter 1); Lixin **Z**hang (China; Chapter 7); and Fangqing Zhao (USA; Chapter 3).

Future Advances in Photosynthesis and Respiration and Other Related Books

The readers of the current series are encouraged to watch for the publication of the following books (not necessarily arranged in the order of appearance):

- The Bioenergetic Processes of Cyanobacteria: From Evolutionary Singularity to Ecological Diversity (Editors: Guenter A. Peschek, Christian Obinger, and Gernot Renger) [http://www.springer.com/life+sciences/book/978-94-007-0352-0]
- Chloroplast Biogenesis: During Leaf Development and Senescence (Editors: Basanti Biswal, Karin Krupinska and Udaya Chand Biswal)
- The Structural Basis of Biological Energy Generation (Editor: Martin Hohmann-Marriott)
- Genomics of Chloroplasts and Mitochondria (Editors: Ralph Bock and Volker Knoop)
- Photosynthesis in Bryophytes and Early Land Plants (Editors: David T. Hanson and Steven K. Rice)

In addition to the above contracted books, the following topics are under consideration:

- Algae, Cyanobacteria: Biofuel and Bioenergy
- Artificial Photosynthesis
- ATP Synthase and Proton Translocation
- Bacterial Respiration II
- Biohydrogen Production
- Canopy Photosynthesis

- Carotenoids II
- Cyanobacteria II
- The Cytochromes
- Ecophysiology
- Evolution of Photosynthesis
- FACE Experiments
- Global Aspects of Photosynthesis
- Green Bacteria and Heliobacteria
- Interactions between Photosynthesis and other Metabolic Processes
- Limits of Photosynthesis: Where do we go from here
- Photosynthesis, Biomass and Bioenergy
- Photosynthesis under Abiotic and Biotic Stress
- Plant Canopies and Photosynthesis
- Plant Respiration II

If you have any interest in editing/co-editing any of the above listed books, or being an author, please send an E-mail to Tom Sharkey (tsharkey@msu.edu) and/or to me at gov@illinois.edu. Suggestions for additional topics are also welcome.

In view of the interdisciplinary character of research in photosynthesis and respiration, it is our earnest hope that this series of books will be used in educating students and researchers not only in Plant Sciences, Molecular and Cell Biology, Integrative Biology, Biotechnology, Agricultural Sciences, Microbiology, Biochemistry, Chemical Biology, Biological Physics, and Biophysics, but also in Bioengineering, Chemistry, and Physics.

We take this opportunity to thank and congratulate Rob Burnap and Wim Vermaas for their outstanding editorial work; they have done a fantastic job not only in editing, but also in organizing this book for all of us, and for their highly professional dealing with the reviewing process. We thank all the 44 authors of this book (see the list above): without their authoritative chapters, there would be no such volume. We give special thanks to A. Lakshmi Praba of SPi Global, India for directing the typesetting of this book; her efficiency and politeness in dealing with the authors and the editors has been crucial in bringing this book to completion. We owe Jacco Flipsen, Ineke Ravesloot and André Tournois (of Springer) thanks for their friendly working relationship with us that led to the production of this book. Further, I thank Tom Sharkey, my co-editor of this Series since volume 31, for his constant involvement and support.

As always, I am indebted to the offices of the Department of Plant Biology (Head: Feng Sheng Hu) and of the Information Technology, Life Sciences (Director: Jeff Haas) of the University of Illinois at Urbana-Champaign, for their continued support. I remain highly indebted to my dear wife Rajni Govindjee for her everlasting support on all matters of my life.

August 15, 2011 **Govindjee**

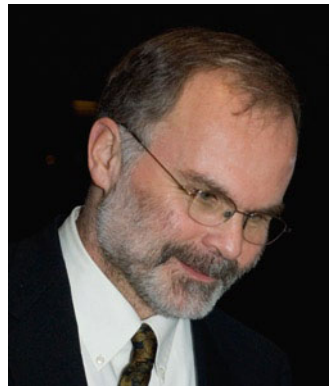
Founding Series Editor, Advances in Photosynthesis and Respiration
Department of Plant Biology, Department of Biochemistry and Center of Biopysics & Computational Biology
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA
e-mail: gov@life.illinois.edu

Series Editors

A photograph of Govindjee wearing Andy Benson's bow-tie (for details, see Govindjee (2010) Photosynth. Res. 105:201–208, Fig. 12).

Govindjee, who uses one name only, was born on October 24, 1932, in Allahabad, India. Since 1999, he has been Professor Emeritus of Biochemistry, Biophysics and Plant Biology at the University of Illinois at Urbana-Champaign (UIUC), Urbana, IL, USA. He obtained his B.Sc. (Chemistry and Biology) and M.Sc. (Botany; Plant Physiology) in 1952 and 1954, from the University of Allahabad. He studied 'Photosynthesis' at the UIUC, under two pioneers of photosynthesis Robert Emerson, and Eugene Rabinowitch, obtaining his Ph.D. in 1960, in Biophysics. He is best known for his research on the excitation energy transfer, light emission, the primary photochemistry and the electron transfer in "Photosystem II" (PS II, water-plastoquinone oxido-reductase). His research, with many collaborators, has included the discovery of a short-wavelength form of chlorophyll (Chl) *a* functioning in the Chl *b*-containing system, now called PS II; of the two-light effect in Chl *a* fluorescence; and, with his wife Rajni Govindjee, of the two-light effect (Emerson Enhancement) in NADP reduction in chloroplasts. His major achievements, together with several other researchers, include an understanding of the basic relationships between Chl *a* fluorescence and photosynthetic reactions; an unique role of bicarbonate/carbonate on the electron acceptor side of PS II, particularly in the protonation events involving the Q_B binding region; the theory of thermoluminescence in plants; the first picosecond measurements on the primary photochemistry of PS II; and the use of Fluorescence Lifetime Imaging Microscopy (FLIM) of Chl *a* fluorescence in understanding photoprotection, by plants, against excess light. His current focus is on the 'History of Photosynthesis Research', in 'Photosynthesis Education', and in the 'Possible Existence of Extraterrestrial Life'. He has served on the faculty of the UIUC for ~ 40 years. Govindjee's honors include: Fellow of the American Association of Advancement of Science (AAAS); Distinguished Lecturer of the School of Life Sciences, UIUC;

Fellow and Lifetime member of the National Academy of Sciences (India); President of the American Society for Photobiology (1980–1981); Fulbright Scholar and Fulbright Senior Lecturer; Honorary President of the 2004 International Photosynthesis Congress (Montréal, Canada); the first recipient of the Lifetime Achievement Award of the Rebeiz Foundation for Basic Biology, 2006; Recipient of the Communication Award of the International Society of Photosynthesis Research, 2007; and the Liberal Arts and Sciences Lifetime Achievement Award of the UIUC, 2008. Further, Govindjee was honored (1) in 2007, through two special volumes of Photosynthesis Research, celebrating his 75th birthday and for his 50-year dedicated research in 'Photosynthesis' (Guest Editor: Julian Eaton-Rye), and (2) in 2008, through a special International Symposium on 'Photosynthesis in a Global Perspective', held in November, 2008, at the University of Indore, India. Govindjee is coauthor of 'Photosynthesis' (John Wiley, 1969); and editor of many books, published by several publishers including Academic Press and Kluwer Academic Publishers (now Springer). Since 2007, each year a Govindjee and Rajni Govindjee Award is given to graduate students, by the Department of Plant Biology, at the UIUC, to recognize Excellence in Biological Sciences. For further information on Govindjee, see his web site at http://www.life.illinois.edu/govindjee.



Thomas D. (Tom) Sharkey obtained his Bachelor's degree in Biology in 1974 from Lyman Briggs College, a residential science college at Michigan State University, East Lansing, Michigan. After 2 years as a research technician, Tom entered a PhD program in the federally funded Plant Research Laboratory at Michigan State University under the mentorship of Klaus Raschke and finished in 1979 after just 3 years and 3 months. Post-doctoral research was carried out with Graham Farquhar at the Australian National University, in Canberra, Australia, where he coauthored a landmark review on photosynthesis and stomatal conductance that continues to get over 50 citations per year more than 25 years after its publication. For 5 years, he worked at the Desert Research Institute, Reno, Nevada. After Reno, Tom spent 20 years as Professor of Botany at the University of Wisconsin in Madison. In 2008, Tom became Professor and Chair of the Department of Biochemistry and Molecular Biology at Michigan State University. Tom's research interests center on the exchange of gases between plants and the atmosphere. The biochemistry and biophysics underlying carbon dioxide uptake and isoprene emission from plants form the two major research topics in his laboratory. Among his contributions are measurement of the carbon dioxide concentration inside leaves, an exhaustive study of short-term feedback effects in carbon metabolism, and a significant contribution to elucidation of the pathway by which leaf starch breaks down at night. In the isoprene research field, Tom is recognized as the leading advocate for thermotolerance of photosynthesis as the explanation for why plants emit isoprene. In addition, his laboratory has cloned many of the genes that underlie isoprene synthesis and published many papers on the biochemical regulation of isoprene synthesis. Tom has edited two books, the first on trace gas emissions from plants in 1991 and then volume 9 of this series (*Advances in Photosynthesis and Respiration*) on the physiology of carbon metabolism of photosynthesis in 2000. Tom is coeditor of volume 34, titled "*Photosynthesis: Plastid Biology, Energy Conversion and Carbon Assimilation*" of this series with Julian Eaton-Rye and Baishnab Tripathy. Tom is listed in Who's Who and is a "Highly Cited Researcher" according to the Thomson Reuters Institute for Scientific Information.

Contents

Preface

Recent advances in the genomics of photosynthetic organisms, coupled with increased understanding of the photosynthetic mechanisms, provide new perspectives on the function and evolution of the photosynthetic machinery. A central theme of this book '*Functional Genomics and Evolution of Photosynthetic Systems*' is how the explosive growth in biological sequence data is leading to a better understanding of the operation and evolution of photosynthetic mechanisms. The book illustrates the interplay between genomic information, bioinformatic analysis and molecular methods to obtain a better understanding of photosynthesis and its development. This interplay has further informed us on the evolution and functional diversification of well-known structures, such as the photosynthetic antennae and reaction centers. Since many photosynthetic structures and pathways are increasingly well understood, even at the atomic level, the large amount of genomic sequence information can now be integrated with structural and mechanistic knowledge of photosynthetic mechanisms and physiology. While not meant to be fully comprehensive in terms of the topics covered, the book provides detailed views of specific cases and thereby illustrates important new directions that are being taken in this fast-moving field – a field that involves the integration of bioinformatics, molecular biology, physiology, and ecology. The book is intended for a wide audience but it is designed specifically for advanced undergraduate and graduate students and for researchers in photosynthesis who are interested in functional genomics, bioinformatics, and evolution of photosynthesis.

The book includes two broad perspectives that have benefitted from the *genomics revolution* – first, the perspective examining whole genome and metagenomic information, and second, the perspective looking into the individual photosynthetic complexes and enzyme systems. The first perspective, introduced in **Chapter 1**, emphasizes the importance of placing genomic information within eco-physiological context in order to understand the functionalities, such as photosynthetic capacity, encoded within the genomic DNA sequences. What level of genomic diversity exists at the species/genus levels? How do various genotypes partition into different ecological niches? The authors of Chapter 1 provide a current analysis of hot springs mat communities in Yellowstone National Park; this habitat contains many photosynthetic bacteria, several of which are comprehensively discussed elsewhere in this book.

Chapter 2 describes a multi-pronged strategy that led to the total genomic sequencing of a pair of cyanobacterial strains (ecotypes) that play dominant roles in the photosynthetic ecology of the hot spring mat ecosystem. This has permitted genomic comparisons, for example, to evaluate the role of insertion and deletions, events leading to genomic divergence. And, when integrated with ecophysiological information, this kind of analysis leads to an understanding of how specific genotypes become associated with particular ecological niches. Further, by using the genomic sequences as reference 'scaffolds' for comparison with metagenomic sequence information, ideas on the origins of sequence diversity within closely related populations begin to emerge.

Chapter 3 summarizes insights gained from an analysis of a massive set of genomic and metagenomic sequence data for the three major taxa comprising the so-called 'green bacteria'. As with the cyanobacteria from the hot springs, discussed in the first two chapters, the synergy between whole-genome sequencing and the more fragmentary, but more ecologically representative metagenomic data of the various green bacteria permits researchers to draw detailed conclusions on the probable evolution of specific metabolic processes and, at the same time, to place these findings in a more meaningful ecological context. Together, these analyses bring us closer to the answers to the question of what constitutes a bacterial species and what are the molecular mechanisms of adaptation to specific ecological niches. Studies of different phototrophs that occupy very different ecological niches, likewise, move us towards answers to these crucial questions of biology. Recent findings on the marine cyanobacteria in the taxonomic group

Synechococcus are presented in **Chapter 4**. Here, the author describes how genomic information has led to advanced gene expression analysis to understand niche specialization in terms of the patterns of gene expression under changing environmental conditions. This again sheds light on fundamental biological questions into how various genotypes partition into different ecological niches within the marine environment.

Very much has been, and will continue to be, learned from photosynthetic model systems. Chapters 5–7 illustrate this by showing the latest application of cutting-edge functional genomics approaches to three well-studied photosynthetic organisms, the cyanobacterium *Synechococcus*, the green alga *Chlamydomonas*, and the angiosperm *Arabidopsis*. In **Chapter 5**, an elegant molecular genetic approach is described to define the function of potentially all genes in the chromosome of the *Synechococcus* sp PCC 7942. Chapters 6 and 7 lead us away from the prokaryotic phototrophs and introduce the reader to the latest developments regarding photosynthesis in the eukaryotes, *Chlamydomonas* and *Arabidopsis*. A comprehensive discussion of the new insights into the photosynthesis genes of the recently sequenced *Chlamydomonas* genome is presented in **Chapter 6**; this chapter describes a multipronged approach involving high-throughput methods and path-finding bioinformatic analysis to discover and find function of photosynthesis genes using, among other techniques, comparative genomics. High-throughput techniques also figure prominently into the description of the current status of understanding the photosynthetic mechanism of the very well studied higher plant, Arabidopsis, as presented in **Chapter 7**.

Photosynthetic carbon metabolism has turned out to be much more varied and complex than could have been imagined immediately following the studies that originally defined the individual steps of the Calvin-Benson-Bassham cycle. Furthermore, photosynthetic metabolism in eukaryotes involves cooperation of multiple subcellular compartments including the chloroplast. **Chapter 8** probes the question of the evolution of photosynthetic metabolism in organisms containing chloroplasts. As originally surmised by microscopists in the late nineteenth century, it is now recognized that the chloroplasts of algae and plants derive from the ancient endosymbiotic capture of once free-living cyanobacteria – these endosymbionts have evolved within the cytoplasm of their hosts to become present-day chloroplasts. The authors of Chapter 8 address the thought-provoking question of how two separate metabolic systems (endosymbiont and host) became integrated into the finely-tuned, compartmentalized metabolic systems found in algae and plants. The topic of photosynthetic metabolism is broadened in **Chapter 9** where an insightful presentation is given of the remarkable diversity of carbon-fixing mechanisms. The Chapter richly benefits from the growing wealth of genomic information on both the structural and regulatory components of carbon fixation systems in autotrophic prokaryotes and eukaryotes.

The final six chapters provide the other broad perspective of the book: how the ever-growing volume of genetic data has illuminated the study of individual photosynthetic complexes and enzyme systems. This section of the book considers 'genome-enabled' and bioinformatic approaches for the understanding of the structure, function, and evolution of photosynthetic mechanisms. The evolution of the light-harvesting complexes of purple phototrophic bacteria is analyzed in **Chapter 10**, where a careful and comprehensive phylogenetic analysis illustrates the fundamentals of gene family evolution in this ancient protein family. Evidence consistent with convergent evolution of function and examples of horizontal gene transfer are presented. Similarly, the evolution of the large and widespread light-harvesting chlorophyll protein superfamily is tackled in **Chapter 11**. Based upon their analysis, the authors discuss the taxonomy of this superfamily and present a plausible model how the ancestral protein duplicated and diverged to adopt the various functional roles played, now seen among the individual members of the superfamily. Insights on the structure, function, and evolution of individual complexes and pathways have developed, for example, from bioinformatic analyses through the identification of conserved amino acids, in turn reflecting functional constraints on replacement through mutational events. The discussion of the conservation features can be enriched by connecting the sequence/structure information with known functional roles when these roles are known from actual experiments or by a discussion of hypotheses where conserved features have

been identified, but experimental information is not yet established. The authors of **Chapter 12** examine our current knowledge on the evolution of Type I photosynthetic reaction centers and the selective forces that may have been at work during the evolution of structural asymmetry during the evolution of heterodimeric Photosystem I from the ancestral homodimeric versions that are presently represented in certain anaerobic phototrophs. This chapter also provides an excellent treatment of the literature on the current thinking about the evolution of photosynthesis on primordial Earth. In **Chapter 13**, the evolution of Photosystem II is discussed, including a description of techniques that allow the projection of amino acid residue conservation obtained from large sets of multiple sequence alignments on to the three-dimensional structure of the reaction center complex. The emerging topic of bacterial protein micro-compartments is discussed in **Chapter 14** with a presentation of the structural aspects of the carboxysome, which is part of the carbon concentrating mechanism in a wide range of prokaryotes, most notably the cyanobacteria. Finally, the function and evolution of the various hydrogenases found in Nature is considered in **Chapter 15**. A detailed presentation of structure, function, and genetics of the major hydrogenase protein families is given with an eye towards the ultimate biotechnological application of hydrogenases for solar energy production.

We emphasize that the impact of genomic and bioinformatic approaches on the field of photosynthesis is bidirectional: not only has the genomics revolution benefitted the field of photosynthesis, but the reciprocal is equally true. The photosynthetic community has the ability to make a unique contribution to the field of genomics and bioinformatics since a reasonably comprehensive knowledge-base of the various components (*e.g.*, reaction centers; see volumes 22, 24 and 28 in the Advances in Photosynthesis and Respiration Series) and pathways (*e.g.*, chlorophyll biosynthesis; see volume 31 in the same Series) of the different photosynthetic mechanisms has been established. It is often possible to add great depth to the 'annotations' of the sets of genes encoding photosynthetic structures. In this way, examples of how photosynthetic genomes function and evolve can provide more tangible insights into the broader issues of genome function, evolution, and speciation. Importantly, efforts to place genomic sequence data in the context of metagenomic community information are discussed in terms of long-standing questions of niche and speciation.

Finally, we thank the many people who have made this book project a reality. This includes the contributors of the chapters (see list of Contributors, pp. i-xxxiii) for their efforts and patience during the development of the book. Many thanks go to members of one of our groups (that of RLB), especially Hong Hwang and Anthony Kappell, for their contributions to the editing. RLB especially thanks his wife, Kathy (Ling), for her enduring support and patience during this and other projects. We also thank Jacco Flipsen, Ineke Ravesloot and Andre Tournois of Springer, and A. Lakshmi Praba of SPi Global for her friendly and valuable guidance during the typesetting and printing of this book.

We are grateful to Govindjee for inviting us to edit this work and for his constant guidance and support from its inception to the final release of this book. RLB acknowledges that his efforts on this book were supported by the US National Science Foundation under Grant MCB 0818371 and US Department of Energy grant DE-FG02-08ER15968 from Energy Biosciences Division.

Robert L. Burnap
Department of Microbiology and Molecular Genetics
307 Life Sciences East
Oklahoma State University
Stillwater, OK 74075, USA
e-mail: rob.burnap@okstate.edu

Willem F.J. Vermaas
School of Life Sciences and
Center for Bioenergy and Photosynthesis
Arizona State University
Box 874501
Tempe, AZ 85287-4501, USA
e-mail: wim@asu.edu

The Editors

Robert (Rob) L. Burnap first developed an interest in photosynthesis during a summer field course in marine biology at Woods Hole (Massachusetts, USA), while an undergraduate student at the University of Michigan from where he graduated in 1977. Entering the Master's program in biology at the University of California Los Angeles, he was influenced by the late J. Philip Thornber and by David Chapman, who team-taught a plant biochemistry course. Thornber infected him with an enthusiasm for experimental science and for studying chlorophyll-protein complexes. Chapman revealed to him that beneath the seemingly tedious nomenclature of taxonomic classification lay the workings of evolution and natural selection. Chapman also introduced him to what would later become the topic of his Ph.D. thesis: The taxonomically enigmatic and evolutionarily important photosynthetic protist, *Cyanophora paradoxa*. He joined the laboratory of Robert Trench, at the University of California Santa Barbara, a world expert on symbiosis and *Cyanophora paradoxa*. The result was a thesis, entitled "Biogenesis of the Endosymbiotic Cyanobacteria (Cyanelles) of *Cyanophora paradoxa*". That project began the life-long fascination with oxygenic photosynthesis as conducted by the originators of this earth-transforming process: the cyanobacteria. Following his PhD research, Rob joined the laboratory of Louis Sherman at Purdue University to focus on Photosystem II (PSII), the system that splits water. He studied the function of the manganese-stabilizing protein (PsbO) of PSII, finding that the protein was not essential for the function of the manganese cluster of the water oxidation complex, in contrast to what had been previously thought. That project ultimately led to a preoccupation with photoactivation; the light-driven assembly of the metal atoms of the water oxidation complex. As part of his postdoctoral research he spent time in Yorinao Inoue's lab at RIKEN in Japan, analyzing the PSII charge-separation properties of mutants using thermoluminescence and absorption spectroscopy with Hiroyuki Koike and sharing the lab with Teruo Ogawa, Masahiko Ikeuchi and others. During the time in the Sherman lab, Rob also discovered that the novel chlorophyll-protein complexes produced by cyanobacteria under iron-deficiency were related to the CP43 protein of PSII illustrating the evolutionary diversification of six transmembrane helix chlorophyll proteins beyond their canonical roles played in the reaction centers as proximal antennae. He joined the faculty of the Department of Microbiology and Molecular Genetics at the Oklahoma State University (OSU) in 1991. He has continued a study of the assembly and function of the water oxidation complex, which now takes on new mechanistic focus with the continued refinement

of the PSII crystal structure – thanks to others in the community of photosynthesis researchers. Ten years ago, his lab developed DNA microarrays for the cyanobacterium *Synechocystis*, and the global perspective afforded by using these arrays has returned Rob's attention to some of the more holistic ideas from his PhD thesis on how organisms and their parts are integrated into functional bioenergetic/metabolic units. Specifically, he has been studying the metabolic signals resulting from the light reactions and how these signals control the expression and activity of the cyanobacterial carbon-concentrating mechanism. Rob has served as a rotating Program Director at the United States National Science Foundation where, besides his regular duties in the Division of Cellular Molecular Biochemistry, he helped develop a special program for fielding and funding innovative proposals in photosynthesis research. For further information on Rob, see his web page at: http://microbiology.okstate.edu/faculty/burnap/.

Willem (Wim) F.J. Vermaas developed an interest in photosynthesis and cyanobacteria during his Master's research at the Agricultural University in Wageningen (The Netherlands) with Jack J.S. van Rensen, working on Photosystem II (PS II) herbicides. The University actively encouraged global research experience, and subsequent graduate work was done by spending about a year each at the University of Illinois at Urbana-Champaign, USA (with Govindjee, working on the bicarbonate effect in PSII), Michigan State University (with Charles Arntzen, working on the mode of action of PSII herbicides), and the Technical University in Berlin, Germany (with Gernot Renger, working on competitive interactions between the native plastoquinone and PSII herbicides). After defending his dissertation at the Agricultural University in Wageningen in 1984 (with Jack J.S. van Rensen, Wim Vredenberg and Gernot Renger), he joined the Arntzen group at the Du Pont Experimental Station in Wilmington, Delaware, learning the tricks of genetic manipulation of *Synechocystis* sp. PCC 6803 from John G.K. Williams. During his time at Du Pont, he published the first deletion and site-directed mutations leading to significant impairments in cyanobacterial photosynthesis. Wim joined Arizona State University in 1986, and before his lab was set up he spent time in Yorinao Inoue's lab at RIKEN in Japan, working on the PSII composition of mutants and sharing the lab with Teruo Ogawa, Masahiko Ikeuchi and others. At ASU, he was part of the founding of the Photosynthesis Center, and continued his work on discovery of protein residues and regions important for redox and chlorophyll-binding function in PSII. Over the years this work evolved into placing photosynthesis within the context of the molecular physiology and the evolutionary history of the cyanobacterium. New insights at the time included a common evolutionary ancestry of the two photosystems based on the primary structure of the heliobacterial reaction center, as well as the realization that photosynthesis and respiration share electrons and complexes in cyanobacteria. The sequencing of the *Synechocystis* genome by Satoshi Tabata's group at the Kazusa Institute in 1995 provided an excellent genomic framework. The genome sequence also facilitated the investigation of photosynthesis-related areas including the synthesis of photosynthetic pigments in relation to photosynthetic proteins, and of proteins aiding in chlorophyll reutilization. Structural insights complemented the functional and genomic research, with collaborative work on electron tomography and on hyperspectral imaging providing a detailed structural framework for the photosynthetic and physiological function of cyanobacteria. With the development of a detailed "toolbox" for modification of *Synechocystis* over the years, Wim's interests gradually evolved to include the use of *Synechocystis* as a photosynthetic platform for more applied purposes, such as production of petroleum substitutes, including fatty acids that are subsequently converted to alkanes, from sunlight, CO_2 and water. This emphasis necessitates a better understanding of the photosynthate metabolism in cyanobacteria, which also has become an active area of research. However, many of the basic processes in cyanobacteria, such as the formation of thylakoid membranes, remain poorly

understood, and research in these fundamental areas are likely to continue to be an important emphasis. Wim has received an NSF Presidential Young Investigator award, and is Fellow of the American Association for the Advancement of Science. He is on the Editorial Board of the Journal of Biological Chemistry, and has served as Associate Editor of Plant Molecular Biology and on the Editorial Boards of Photosynthesis Research and Plant Cell Physiology.

Contributors

Iwona Adamska
Department of Biology, University of Konstanz, Universitätsstrasse 10, D-78457 Constance, Germany
Iwona.Adamska@uni-konstanz.de

Birgit E. Alber
Department of Microbiology, The Ohio State University, 484 West 12th Avenue, Columbus, OH 43210-1292, USA
alber.8@osu.edu

Jens Appel
Botanisches Institut, Christian-Albrechts-Universität, Am Botanischen Garten 1-9, 24118 Kiel, Germany
jappel@bot.uni-kiel.de

Shaun Bailey
Aurora Biofuels, 301 Harbor Bay Parkway, Alameda, CA 94502, USA

J. Thomas Beatty
Department of Microbiology and Immunology, University of British Columbia, 2350 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada
jbeatty@interchange.ubc.ca

Devaki Bhaya
Department of Plant Biology, Carnegie Institution for Science, 260 Panama Street, Stanford, CA 94305, USA
dbhaya@stanford.edu

Donald A. Bryant
Department of Biochemistry and Molecular Biology, The Pennsylvania State University, S–235 Frear Building, University Park, PA 16802, USA
Department of Chemistry and Biochemistry, Montana State University, Bozeman, MT 59717, USA
dab14@psu.edu

Robert L. Burnap
Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK 74078, USA
rob.burnap@okstate.edu

Fei Cai
US Department of Energy, Joint Genome Institute, Walnut Creek, CA 94598, USA
Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA

You Chen
Synthetic Genomics Inc, 11149 North Torrey Pines Road, La Jolla, CA 92037, USA

Wei Chi
Photosynthesis Research Center, Key Laboratory of Photobiology, Institute of Botany, Chinese Academy of Sciences, Nanxincun 20, Xiangshan, Beijing100093, China

Frederick M. Cohan
Department of Biology, Wesleyan University, Middletown, CT 06459, USA

Johannes Engelken
Institute of Evolutionary Biology (CSIC-UPF), Pompeu Fabra University, Barcelona, Spain
Department of Biology, University of Konstanz, Universitätsstrasse 10, D-78457 Constance, Germany

Niels-Ulrik Frigaard
Department of Biology, University of Copenhagen, 3000 Helsingør, Denmark

Christiane Funk
Department of Chemistry and Umeå Plant Science Centre (UPSC), Umeå University, S-90187 Umeå, Sweden

Amaya M. Garcia Costas
Department of Biochemistry and Molecular Biology, The Pennsylvania State University, S-235 Frear Building, University Park, PA 16802, USA

John H. Golbeck
Department of Biochemistry and Molecular Biology and Department of Chemistry, The Pennsylvania State University, University Park, PA 16802, USA
jhg5@psu.edu

Susan S. Golden
Division of Biological Sciences, Center for Chronobiology, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
sgolden@ucsd.edu

David González-Ballester
Department of Plant Biology, Carnegie Institution for Science, 260 Panama Street, Stanford, CA 94305, USA

Arthur R. Grossman
Department of Plant Biology, Carnegie Institution for Science, 260 Panama Street, Stanford, CA 94305, USA
arthurg@stanford.edu

Thomas E. Hanson
College of Earth, Ocean and Environment, Department of Biological Sciences and Delaware Biotechnology Institute, University of Delaware, 127 DBI, 15 Innovation Way, Newark, DE 19711, USA
tehanson@udel.edu

Caroline S. Harwood
Department of Microbiology, University of Washington, Box 357242, Seattle, WA 98195-7242, USA

C. Kay Holtman
Department of Biology, Center for Biological Clocks Research, Texas A&M University, College Station, TX 77843-3258, USA

Bharat Jagannathan
Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA, USA

Steven J. Karpowicz
Department of Chemistry and Biochemistry, University of California – Los Angeles, Box 951569, Los Angeles, CA 90095-1569, USA

Cheryl A. Kerfeld
US Department of Energy, Joint Genome Institute, Walnut Creek, CA 94598, USA
Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA
ckerfeld@lbl.gov

Christian G. Klatt
Departments of Land Resources and Environmental Sciences and Chemistry, Montana State University, Bozeman, MT 59717, USA

Andrew S. Lang
Department of Biology, Memorial University of Newfoundland, 232 Elizabeth Ave, St. John's, NL A1B 3X9, Canada

Tao Li
Department of Biochemistry and Molecular Biology, The Pennsylvania State University, S-235 Frear Building, University Park, PA 16802, USA

Marc Linka
Institut für Biochemie der Pflanzen, Heinrich-Heine-Universität, Universitätsstraße 1, D-40225 Düsseldorf, Germany

Zhenfeng Liu
Department of Biochemistry and Molecular Biology, The Pennsylvania State University, S-235 Frear Building, University Park, PA 16802, USA

Sabeeha S. Merchant
Department of Chemistry and Biochemistry and Institute for Genomics and Proteomics, University of California – Los Angeles, Box 951569, Los Angeles, CA 90095-1569, USA

Aparna Nagarajan
Department of Microbiology and Molecular Genetics
Oklahoma State University
Stillwater, OK 74078, USA

Jörg Overmann
Bereich Mikrobiologie, Ludwig-Maximilians-Universität München, Großhadernerstraße 2–4, 82152 Planegg-Martinsried, Germany

Brian Palenik
Scripps Institution of Oceanography, University of California at San Diego (UCSD), 9500 Gilman Drive, La Jolla, CA 92093-0202, USA
bpalenik@ucsd.edu

Gustaf Sandh
US Department of Energy, Joint Genome Institute, Walnut Creek, CA 94598, USA

Gaozhong Shen
Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA, USA

F. Robert Tabita
Department of Microbiology and The Plant Molecular Biology Biology/Biotechnology Program, The Ohio State University, 484 West 12th Avenue, Columbus, OH 43210-1292, USA

Arnaud Taton
Center for the Study of Biological Complexity, Virginia Commonwealth University, 1000 W. Cary Street, Richmond, VA, 23284, USA

David M. Ward
Departments of Land Resources and Environmental Sciences and Chemistry, Montana State University, Bozeman, MT 59717, USA
umbdw@montana.edu

Andreas P.M. Weber
Institut für Biochemie der Pflanzen, Heinrich-Heine-Universität, Universitätsstraße 1, D-40225 Düsseldorf, Germany
andreas.weber@uni-duesseldorf.de

Jason Wood
Departments of Land Resources and Environmental Sciences and Chemistry, Montana State University, Bozeman, MT 59717, USA

Lixin Zhang
Photosynthesis Research Center, Key Laboratory of Photobiology, Institute of Botany, Chinese Academy of Sciences, Nanxincun 20, Xiangshan, Beijing 100093, China
zhanglixin@ibcas.ac.cn

Fangqing Zhao
Department of Biochemistry and Molecular Biology, The Pennsylvania State University, S-235 Frear Building, University Park, PA 16802, USA

Author Index

Chapter 1

Functional Genomics in an Ecological and Evolutionary Context: Maximizing the Value of Genomes in Systems Biology

David M. Ward*, Christian G. Klatt and Jason Wood
Land Resources and Environmental Sciences Department, Montana State University, Bozeman, MT 59717, USA

Frederick M. Cohan
Department of Biology, Wesleyan University, Middletown, CT 06459, USA

Donald A. Bryant
Department of Biochemistry and Molecular Biology, The Pennsylvania State University, S–235 Frear Building, University Park, PA 16802, USA; Department of Chemistry and Biochemistry, Montana State University, Bozeman, MT 59717, USA

Summary

The full power of functional genomics analyses comes from studying genomes of organisms that are known to be relevant to a system, as this permits a connection between metabolic networks within and across organisms within the systems these individuals inhabit. A genome is most effectively studied in an ecological and an evolutionary context, that is, with a view towards the relationships of the individual with other individuals within species and guilds that comprise a community and that mediate community function. We use examples from our own long-term studies of microbial mat communities found between ~50ºC and ~72ºC in alkaline siliceous hot springs of Yellowstone National Park, which are constructed by cyanobacteria

*Author for correspondence, e-mail: umbdw@montana.edu

R.L. Burnap and W.F.J. Vermaas (eds.), Functional Genomics and Evolution of Photosynthetic Systems, Advances in Photosynthesis and Respiration 33, pp. 1–16, DOI 10.1007/978-94-007-1533-2_1,

(*Synechococcus* spp.), filamentous (*Chloroflexus* spp., *Roseiflexus* spp.) and other newly discovered anoxygenic phototrophic bacteria (*Candidatus* Chloracidobacterium thermophilum, *Chlorobiales*), to illustrate that (i) isolates from systems are often numerically, genetically, and physiologically unrepresentative of the ecological diversity of predominant species, (ii) seemingly small genetic differences can be extremely important when attempting to infer system function from functions of an isolate, (iii) individuals group into ecologically distinct species populations that may or may not be typified by the characteristics of an individual and its genome, and (iv) co-evolution may be important for understanding metabolic networks among individuals of different species. Metagenomics, metatranscriptomics, metaproteomics and metabolomic/stable isotope technologies will soon enable global studies of the metabolic networking from the individual to the community level of biocomplexity. To interpret the results, we will need to understand the principles of networking at all biological levels and we will be better off if our knowledge of networking within individuals is based on the study of isolates relevant to ecological systems.

I. Introduction

Systems biology is an important emerging discipline fueled by genomic technologies (Buckley, 2004). Microbial physiologists are adopting an increasingly global view, attempting to understand gene content and how sets of genes comprise biochemical pathways, the regulation of which forms a dynamic intracellular network that enables an individual cell to maintain itself in a changing environment, and, if conditions are favorable, to grow and eventually reproduce. Similarly, microbial ecologists are beginning to understand the networking of individual cells into higher order ecological constructs, such as species, guilds[1] and communities. To be maximally informative across all levels of systems biology these two approaches need to be connected. Where do the organisms whose genomes we study come from? How did they come to be? What do their properties typify beyond those of the individual cultivated organisms (actually clonal populations of nearly identical individual descendants of single cells) containing these genomes? Functional genomic analyses facilitate an understanding of the metabolic network within such populations. However, beyond this, if you don't know whether an organism you are studying fits into a natural biological system, full stop! Ecologists have organized their thinking across a range of complexity that extends beyond the metabolic networking in an individual organism to networking among individuals within populations (e.g., species) and species within and among guilds, which, as an ensemble, perform various functions within a community (i.e., system) (Table 1.1). If you *do* know how the genomes you study connect with the ecological network in which they occur, you can better understand what the information you glean from functional genomic analyses of an individual means to the system from which it came. A natural view of the genome of a modern organism is that it represents the current state of the long evolutionary history of that organism. By studying a genome in an ecological context, you also have the potential to gain insight into its evolutionary context. You can learn what forces controlled this organism's evolutionary trajectory, in turn explaining what controls the organism's present ecological distribution, activity and contribution to system functions.

We will illustrate the points we wish to make by using examples from the studies we and other colleagues have performed over several decades on phototrophic microbial mats of alkaline siliceous hot springs (Ward et al., 1998, 2002, 2006) (Fig. 1.1). We consider these to be natural model communities from which we can learn fundamental principles of microbial community ecology. This

Abbreviations: SSU rRNA – small subunit ribosomal RNA; A/B-type *Synechococcus* – predominant unicellular cyanobacterial genotypes in 50–72°C alkaline siliceous hot spring microbial mats; FAP – filamentous anoxygenic phototroph; cDNA – complementary DNA; ITS – internal transcribed spacer (separating 16S rRNA and 23S rRNA genes)

[1]A guild is a hypothetical construct ecologists use to combine species that do the same kind of thing, but not in exactly the same way (e.g., sun- and shade-adapted plant species both do oxygenic photosynthesis).

Table 1.1. Ecological levels within a community, as exemplified by microbial mats occurring at 50–72°C in the effluents of alkaline siliceous hot springs

Ecological Level	Example
Ecosystem	Greater Yellowstone Ecosystem
Community	Alkaline Siliceous Hot Spring Mat
Guild 1	Oxygenic Photosynthesis
Species 1	Tens of unnamed A/B lineage *Synechococcus* species differently adapted to
Individuals	temperature
Species 2	light intensity
Individuals	nutrients (e.g., phosphonates?)
Species 3	other environmental features (e.g., Fe^{2+}?)
Individuals	
Guild 2	Bacteriochlorophyll *a*-based anoxygenic photoheterotrophy/photomixotrophy
Species 1	unknown number of unnamed *Roseiflexus* species differently adapted to
Individuals	temperature
Species 2	light intensity
Individuals	nutrients
Species 3	other environmental features
Individuals	
Guild 3	Bacteriochlorophyll *c*-based anoxygenic photoheterotrophy/photomixotrophy
Species 1	unknown number of unnamed *Chloroflexus*, *Candidatus* Chloracidobacterium
Individuals	and *Chlorobiales*[a] species differently adapted to
Species 2	temperature
Individuals	light intensity
Species 3	nutrients
Individuals	other environmental features
Guild 4	Bacteriochlorophyll *d*-based anoxygenic photoheterotrophy
Species 1	unknown number of unnamed *Chlorobiales* species differently adapted to
Individuals	temperature
Species 2	light intensity
Individuals	nutrients
Species 3	other environmental features
Individuals	
Other Guilds	e.g., other forms of phototrophy? aerobic chemoorganotrophy? fermentation,
Other Species	sulfate-reduction, methanogenesis?
	unknown numbers of species in other guilds

[a]At present it is not known whether these phylogenetically distinct taxa have different metabolic contributions to the community. Here, we consider them to be within one guild defined by the quality of light they use

will serve to provide ecological and evolutionary context for many organisms whose genomes we have studied and that will be discussed in other chapters of this book (e.g., *Synechococcus*, see Bhaya et al., Chapter 2; filamentous anoxygenic phototrophs (FAPs); a newly discovered phototrophic acidobacterium, *Candidatus* Chloracidobacterium thermophilum, and as yet uncultivated *Chlorobiales*; see Bryant et al., Chapter 3). Like microbial physiologists trying to understand the hierarchy from genes to operons to metabolic networks, we have studied how individuals group into populations that comprise guilds that in turn comprise the entire community. As ecologists, we are ultimately interested not just in community composition ("who's there") and structure ("who's where"), but also in community function ("who's doing what where and when"). We have studied various functions, such as oxygenic and anoxygenic photosynthesis (Revsbech and Ward, 1984; Nold and Ward, 1996; Ward et al., 2006; van der Meer et al., 2005) and photoheterotrophy (Anderson et al., 1987; Bateson and Ward, 1988) and, as methods have improved, we have been able to associate functions with the underlying guilds (van der Meer et al., 2000, 2003, 2005, 2007). With the advent of the genomics era, we have been able to begin to associate functions with populations within guilds (Steunou et al., 2006, 2008) and even with the underlying species responsible for specific processes (ED Becraft, FM Cohan, M Kühl, S Jensen, and DM Ward, unpublished). We have learned that for an understanding of the ecological networking in the hot spring mat system, it is essential that

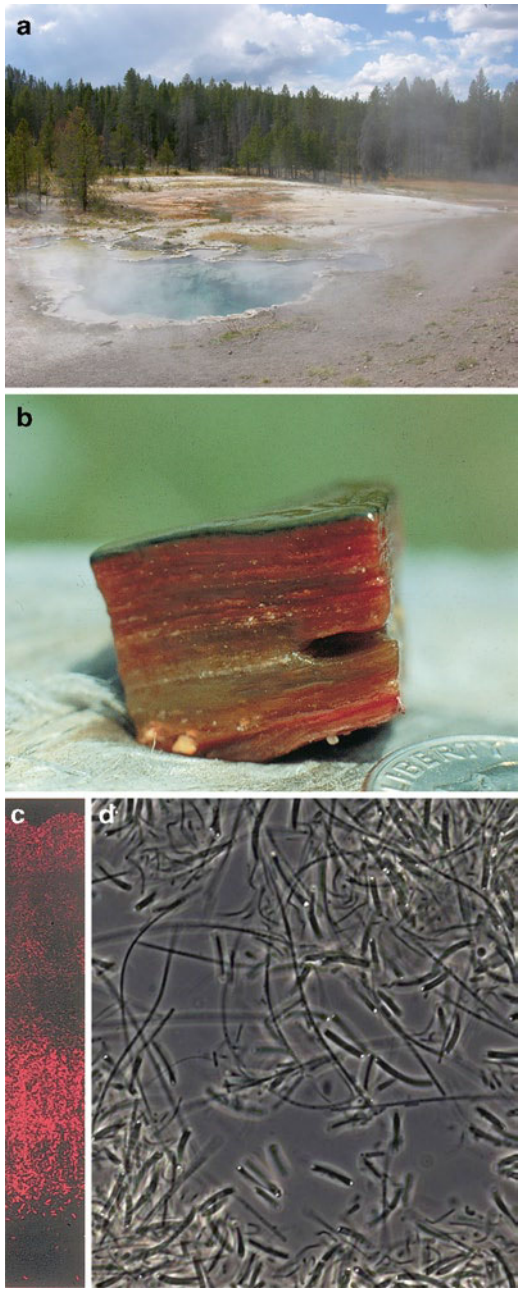


Fig. 1.1. Ecological networks in biology above the individual cell level, descending from the scale of (**a**) the Octopus Spring landscape within the Greater Yellowstone Ecosystem, to (**b**) ~60°C Octopus Spring *Synechococcus* mat sample (American dime for scale), to (**c**) a photomicrograph of the ~1 mm-thick top green mat layer showing chlorophyll *a* red autofluorescence within *Synechococcus* cells, to (**d**) a phase contrast photomicrograph of the same layer showing individual cells of sausage-shaped *Synechococcus* (~2 × 8 μm) and filamentous anoxygenic phototrophs.

functional genomics are done on organisms known to be relevant to the system. In other words, only when we know that the organisms whose genomes we study are closely related to members of native populations, can we associate the phenotypes we infer from these genomes with the system itself.

II. Ecological Context

In microbiology, especially microbial molecular biology and physiology, it has become a tradition to favor the pure culture, something that is manageable in the laboratory and whose molecular biology and physiology can be studied in a controlled laboratory setting. There is nothing wrong with this. We have certainly learned a lot from pure culture studies and genomic analyses of pure cultures have been no exception. Furthermore, some well-studied, rapidly growing pure cultures are especially useful for mutation-based studies of gene function and biochemical analyses—up to the level of processes that occur within an individual cell or among the nearly clonal individuals of a pure-cultured population.

However, if you wish for the observations you make on pure cultures to have meaning at higher levels of biological complexity, you must know how the organism whose genome you have obtained fits into the system from which it came. We fear that the approaches to systems biology taken by physiologists and ecologists are too often disconnected in this regard. For instance, in a recent workshop focused on microbial systems biology, when microbial physiologists performing studies of metabolic networks from genes to cells in pure culture were repeatedly asked "Where did your isolate come from?", the answers were alarmingly evasive.

Merely knowing what habitat an organism comes from (e.g., marine mud or a hypersaline pond) is not enough, as the enrichment and cultivation of microorganisms from Nature using artificial and highly unnatural laboratory-medium environments readily selects for "weedy" microorganisms that may be insignificant members of such communities. Winogradsky (1949), one of the inventors of the elective cultivation method[2]

[2] Also called the selective enrichment culture method.

raised the following concern regarding isolates obtained using this method: "…*are we allowed to invoke the ability of these products of our microbe gardens in order to explain with some precision the natural processes?*" His answer: "*For my part, I am fully prepared to respond negatively.*" and with this we agree. Compared to in situ niches, the culture-medium environments used to select organisms probably represent extreme disturbances, which challenge microorganisms so severely that only special and usually rare types of microorganisms can survive. Unless care is taken to mimic the natural niches of microbes and to prevent competitive exclusion (Santegoeds et al., 1996), cultivation typically favors organisms that have evolved for rapid growth (a phenotype that we often even select for) on unnatural substrates at unnatural concentrations in unnatural physiochemical settings. Recent molecular studies have revealed that the microorganisms we cultivate are often rare members of microbial communities (Santegoeds et al., 1996; Ferris et al., 1996). Perhaps they are microbes that are adapted for dispersal and persistence, merely a part of what we might think of as microbial "seed rain", waiting to respond quickly to an environmental disturbance, like fireweed blooming after a forest fire (Winogradsky's zymogenous organisms). In contrast, natural microbial communities appear to be dominated by diverse species that are finely adapted to a diversity of niches (Winogradsky's autochthonous organisms). It takes special care to cultivate them and it also takes knowledge of the system from which they came to judge their relevance to ecological networks.

Consider the following examples of biases of traditional microscopic and cultivation methods when applied to various kinds of phototrophic microorganisms inhabiting the Octopus and Mushroom Spring mat systems we study.

A. Cyanobacteria

The only cyanobacteria observed microscopically at these temperatures in the Octopus and Mushroom Spring mats are sausage-shaped unicells of *Synechococcus* (Fig. 1.1c and d). We had initially assumed that these were *Synechococcus lividus*, which had previously been cultivated from mats like these. On the grounds that several cultivated strains were genetically very similar (Ferris et al., 1996), we had the impression that this species was cosmopolitan, occurring in, and in many cases dominating in, hot spring microbial mats in the Western Hemisphere and East Asia (Castenholz, 1978, 1981; Ferris et al., 1996). Given possible oversimplification due to simple cell morphology and doubts expressed above about cultivation approaches, we developed molecular methods based on small subunit ribosomal RNA (SSU rRNA) cloning and sequencing (Ward et al., 1990) and used these methods to reveal that the *Synechococcus* in the mats we study (we call them A/B type *Synechococcus*) are genetically diverse and about as different from *S. lividus* as ferns are from angiosperms (Ward et al., 1998) (Fig. 1.2a). *Synechococcus* spp. strains closely related to *S. lividus* are indeed present in the mat and can be readily cultivated. In fact, if you don't dilute a sample enough before you enrich for native *Synechococcus* spp., *S. lividus* will outcompete *Synechococcus* strains that are many orders of magnitude more abundant (Ferris et al., 1996). *S. lividus*-like cyanobacteria appear to be dominant forms in Japanese hot springs, where A/B-type *Synechococcus* have not been detected; these lineages appear to have diverged due to geographic isolation (Papke et al., 2003) (Fig. 1.2a). By being aware of the genetic signatures of *Synechococcus* spp. that *are* predominant in the Yellowstone mats, we were able to know when we had cultivated representative *Synechococcus* strains (Allewalt et al., 2006), and thus that the phenotypic properties of these strains could be used to interpret their in situ distributions. For instance, we had noted that closely related *Synechococcus* SSU rRNA genotypes were uniquely distributed across the environmental landscape correlating with flow away from the hot spring source pool. We were able to demonstrate that at least part of the explanation for this patterning is that strains with different SSU rRNA genotypes have different temperature adaptations that correspond to their in situ distribution along the temperature gradient (Fig. 1.2a).

We have obtained the genome sequences of two of these *Synechococcus* strains (Bhaya et al., 2007, Chapter 2), *Synechococcus* spp. strains A and B′, which are adapted to different temperatures. These isolates represent the maximum

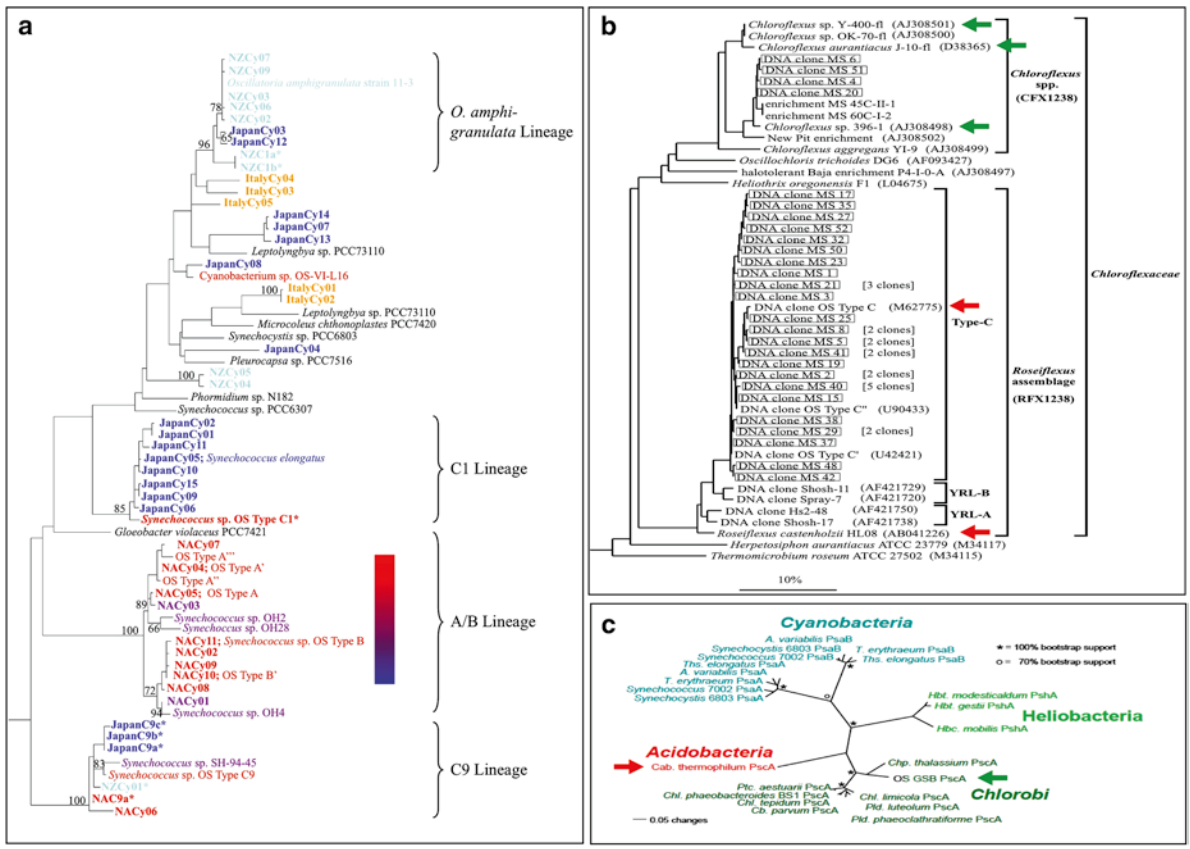


Fig. 1.2. Phylogenies of hot spring phototrophs. (**a**) Cyanobacteria. SSU rRNA phylogeny showing the correlation with geography (NA = North America; NZ = New Zealand) and temperature (bar ranging from blue = ~50°C to red = ~70°C within *Synechococcus* A/B lineage) (Modified from Papke et al., 2003) (**b**) Filamentous anoxygenic phototrophs. SSU rRNA phylogeny showing relationships among *Chloroflexus* and *Roseiflexus* isolates and mat clones (*boxed*); the *Roseiflexus* sp. RS1 isolate 16S rRNA sequences is ~99% related to type-C variants (Modified from Nübel et al., 2002) (**c**) Newly discovered anoxygenic phototrophs. PscA-like sequences of isolates, *Candidatus* Chloracidobacterium thermophilum (*red*) and an as-yet uncultivated green sulfur bacterium (OS GSB). (From Bryant et al., 2007). *Arrows* highlight specific strains mentioned in text.

divergence within the A/B *Synechococcus* clade; they share only ~83% of their genes and the average nucleotide identity of shared genes is 86%. When the proposal that funded the acquisition of these genomes was crafted, the authors (DMW, FMC and colleagues) anticipated that reviewers might wonder why new genomes for thermophilic *Synechococcus* spp. should be obtained when the genome of *Thermosynechococcus elongatus* was already available. We were careful to point out that the genome of *T. elongatus* strain BP1, a very close relative of *S. lividus* (Fig. 1.2a), but a very distant relative of predominant indigenous *Synechococcus* spp. strains, would not help us understand this mat community much, as this organism was a rare member of the community. Had we relied on the *T. elongatus* BP1 genome to make inferences about the mat community, we would have been using a genome that differs significantly in gene content from native *Synechococcus* populations, sharing only 57–58% of its genes with *Synechococcus* spp. A and B′.

The degree to which a genome is representative of native populations can be determined by comparison to metagenomic sequences, using the genome in BLAST analyses to “recruit” homologous sequences and then determining the % nt identity of the matching alignments. Here, recruitment refers to the sequence alignment between

metagenomic and genomic homologs that is most statistically significant (i.e., lowest E-value) relative to other homologous sequences included in an analysis with many genomes. For instance, Fig. 1.3 shows the results of a recruitment experiment in which mat metagenomic sequences were aligned with 10 genomes of phototrophic microorganisms, most of which were cultivated from thermal environments. Metagenomic sequences were assigned to the genome that contains a homolog that aligns better than homologs in other genomes. An exception is that matches with E-values greater than 10^{-10} are placed in a "null bin", as the certainty of an association with any of the genomes used in the analysis is low (Klatt et al., 2011). The quality of recruited sequences (in terms of % nt identity) is reported by degree of whiteness or blackness in Fig. 1.3 and more detailed displays of the % nt identity patterns of metagenomic homologs and recruiting genomes are provided in Fig. 1.4. As shown in Figs. 1.3 and 1.4a, the *T. elongatus* genome recruits only a low number of distantly related (50–75% nt identity) homologs in metagenomes that we have obtained from these mats (Bhaya et al., 2007; Klatt et al., 2011), whereas the *Synechococcus* spp. strain A and B′ genomes recruit large numbers of sequences that are very closely related to genomic homologs. For reference, the degree of divergence of the *T. elongatus* genome from that of *Synechococcus* spp. strain A and B′ genomes is close to that found between cyanobacteria and organisms belonging to other noncyanobacterial kingdom lineages (Klatt et al., 2011). Had we relied upon the *T. elongatus* genome, we would have completely missed the fact that the predominant *Synechococcus* spp. in these mats contain genes for nutrient acquisition that are not present in the *T. elongatus* BP1 strain. For instance, the ability of these *Synechococcus* spp. strains to conduct nitrogen fixation (Steunou et al., 2006, 2008) and to metabolize phosphonates as a source of phosphorus (Adams et al., 2008) would have gone undiscovered. Genes for phosphonate utilization were present in the *Synechococcus* sp. strain B′, but not the *Synechococcus* sp. strain A genome. Thus, despite an average % nt identity of 86% between these two genomes, one cannot be used to infer the properties of the other. This is important, since A-like and B′-like *Synechococcus* predominate at different positions along the flow path (Figs. 1.2 and 1.3)

B. Filamentous Anoxygenic Phototrophs

Similarly, we had initially thought that the readily cultivated FAP *Chloroflexus aurantiacus,* which had been isolated from the Yellowstone mats, was a numerically important component of those mats. However, SSU rRNA analysis showed that *Roseiflexus* spp., belonging to a distantly related FAP clade, were more abundant except near the upper temperature limit of the mat (~70°C), under which conditions these two types of FAPs appear to be co-dominant (Nübel et al., 2002) (Fig. 1.2b). These two FAPs, whose genomic homologs exhibit only 50–75% nt identity, are adapted to utilize different regions of the light spectrum; *Chloroflexus* spp. use primarily bacteriochlorophyll *c* and *Roseiflexus* spp. use primarily bacteriochlorophyll *a*, so that inferences about FAP phototrophy from the available *C. aurantiacus* genome (strain J-10-fl) (NC_010175) would not have revealed the correct light utilization characteristics of the predominant mat FAPs. Fearing that the genome of the Japanese isolate *R. castenholzii* (NCBI Accession # NC_009767) might not inform us sufficiently about Yellowstone mat FAPs, we cultivated a *Roseiflexus* sp. strain (RS1) that is genetically relevant to Yellowstone mats (Madigan et al., 2005; van der Meer et al., 2010) and its genome sequence was obtained (NC_009523; see Bryant et al., Chapter 3 and Klatt et al., 2007). *R. castenholzii* shares only 82% of its genes with *Roseiflexus* sp. RS1 and recruits mat metagenomic homologs with a much lower % nt identity than does the RS1 isolate. (Figs. 1.3 and 1.4b). The *Roseiflexus* sp. RS1 genome, like the *Synechococcus* sp. B′ genome, contains genes for phosphonate utilization, whereas the *R. castenholzii* genome does not.

Likewise, we have obtained the genome of *Chloroflexus* sp. Y-400-fl (NC_012032) and the draft genome of *Chloroflexus* sp. 396–1 (http://www.jgi.doe.gov/sequencing/statusreporter/psr.php?projectid=400015), two Yellowstone isolates that are genetically similar to the Japanese *C. aurantiacus* J-10-fl isolate and to native *Chloroflexus* spp. of the Yellowstone mats. The SSU rRNA phylogeny (Fig. 1.2b) exhibits two *Chloroflexus* clades; the SSU rRNA gene of Yellowstone *Chloroflexus* sp. strain 396–1 is more closely related to Yellowstone mat homologs, while that of Yellowstone isolate Y-400-fl is more closely

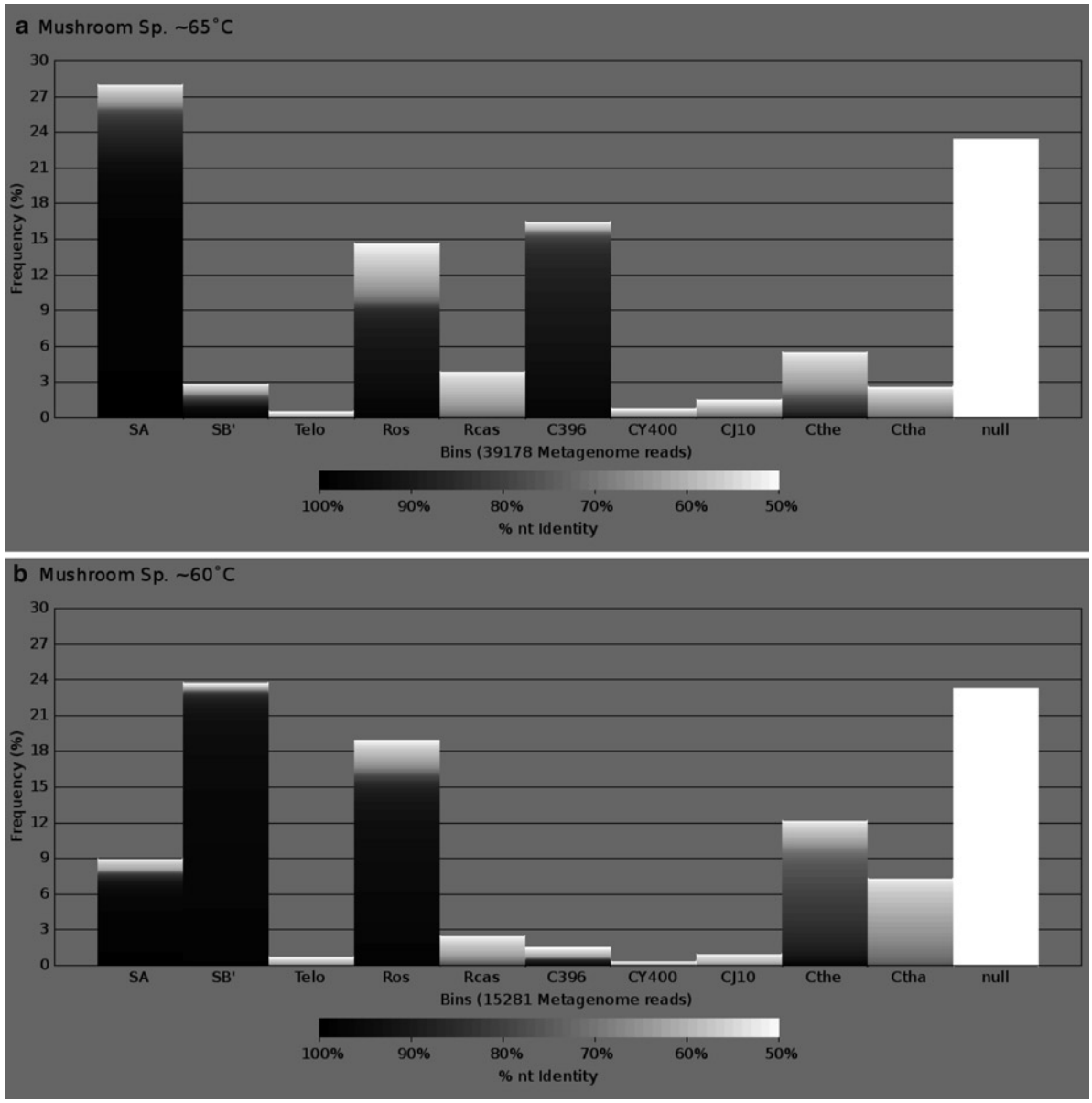


Fig. 1.3. BLASTN-based recruitment of metagenomic sequences from libraries prepared from top green (0–1 mm) mat layers from (**a**) Mushroom Spring ~65°C and (**b**) Mushroom Spring ~60°C by the genomes of 10 phototrophic microorganisms of possible relevance to these mats. SA, *Synechococcus* sp. A; SB′, *Synechococcus* sp. B′; Telo, *Thermosynechococcus elongatus*; Ros, *Roseiflexus* sp. RS-1; Rcas, *Roseiflexus castenholzii*; C396, *Chloroflexus aurantiacus* strain 396–1; CY400, *Chloroflexus aurantiacus* strain Y-400; CJ10, *Chloroflexus aurantiacus* strain J-10-fl; Cthe, *Candidatus* Chloracidobacterium thermophilum; Ctha, *Chloroherpeton thalassium*. Shading indicates % nt identity of sequences within bins from 50% (*white*) to 100% (*black*).

related to that of the Japanese isolate J-10-fl. Blast analyses of these genomes and mat metagenomic sequences reveal the relative contributions the three *Chloroflexus* genomes make to the mat community (Fig. 1.3) and the degrees to which these genomic sequences match with the metagenomic homologs they recruit (Fig. 1.4b). Not surprisingly, mat metagenomic homologs are much more closely related to genes in strain 396–1 (Figs. 1.3 and 1.4b); the genome of Yellowstone strain

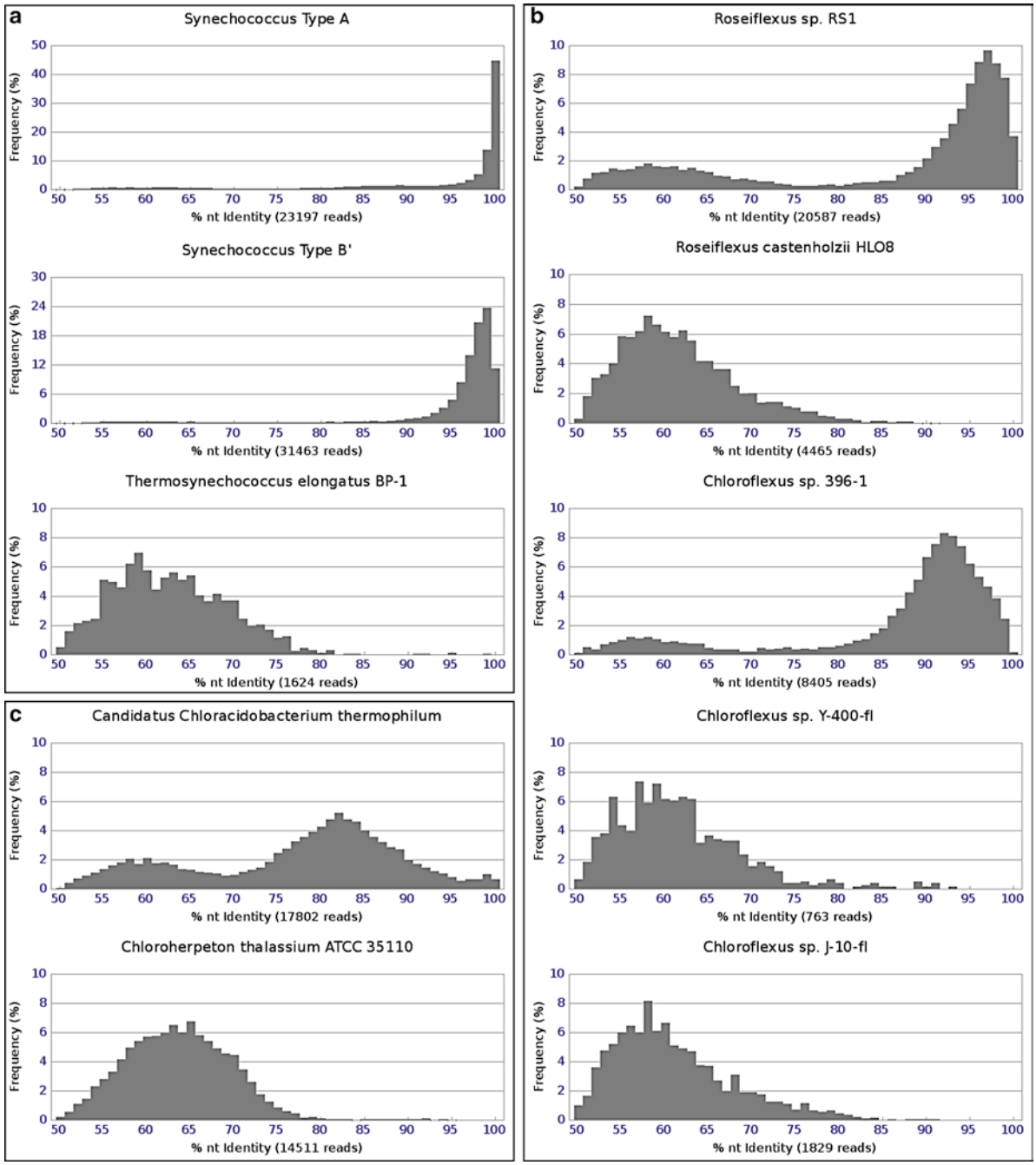


Fig. 1.4. Histograms showing the % nucleotide identity between genes in genomes from various phototrophic isolates from these or similar mats and homologous sequences in mat metagenomes recruited by BLASTN analyses. (**a**) Cyanobacteria, (**b**) Filamentous anoxygenic bacteria, (**c**) Newly discovered anoxygenic bacteria.

Y-400-fl, which is extremely closely related to that of Japanese strain J-10-fl (Bryant et al., Chapter 3), does not improve the recruitment of genes from the mat metagenomes. It definitely matters which Yellowstone *Chloroflexus* sp. strain/genome one assumes as the reference when one tries to do functional genomics studies relevant to this mat.

C. Newly Discovered Anoxygenic Phototrophs

Our analyses of mat metagenomes also revealed the presence of two new kinds of chlorophototrophic prokaryotes, *Candidatus* Chloracidobacterium thermophilum, the first known phototrophic member of the Kingdom *Acidobacteria*[3], and an as-yet uncultivated green sulfur bacterium distantly related to other *Chlorobiales* (Bryant et al., 2007) (Fig. 1.2c). We have recently obtained the complete genome for a *C.* Cab. thermophilum isolate (Garcia Costas et al., 2011; also see Bryant et al., Chapter 3) and used it to demonstrate that organisms related to this isolate are well represented in the mat community (Fig. 1.3). However, the divergence of the genes of this isolate from homologs of the predominant, indigenous organisms of this general type is quite significant (e.g., as great as that between *Synechococcus* spp. A and B′ or between Yellowstone and Japanese *Roseiflexus* spp. strains; Klatt et al., 2011) (Fig. 1.4c). Whether the properties of this isolate are typical of the properties of the more abundant population has not yet been determined, but comparisons of the genome of the isolate to the assembled metagenome will allow this to be assessed. The chlorophyll-based photoheterotrophic metabolism of the isolate is typical of predominant *C.* Cab. thermophilum relatives, as revealed by in situ transcription results (Liu et al., 2011) . It should be emphasized here that the ability to use next-generation sequencing methods to perform metatranscriptional profiling is highly dependent upon the availability of genetically relevant reference genomes and/or metagenomic data. Methods (e.g., SOLiD™) that produce short cDNA sequences (e.g., 50 bp) can only be matched with a high degree of confidence when the reference genomes are very similar to those found in the environment from which the RNA has been obtained. Specifically in the case of *C.* Cab. thermophilum, more complete transcription profiling was obtained by using the metagenomic assemblies, in spite of the fact that a completed genome was available for this organism (Liu et al., 2011). This is simply due to the fact that the isolated strain of *C.* Cab. thermophilum is a minor community member and its sequence has diverged significantly from that of the dominant *C.* Cab. thermophilum-like population(s) in the mats.

We have yet to cultivate an example of the *Chlorobiales* found in these mats; however, genomes of distantly related *Chlorobiales* isolates recruit many homologous genes from the mat metagenome (Fig. 1.3). The metagenome of this organism is most closely related in terms of gene content to the genomes of other *Chlorobiales* and, in terms of sequence relatedness to the genome of *Chloroherpeton thalassium*, a bona fide green sulfur bacterium (Gibson et al., 1984). The divergence of these predicted *Chlorobiales* genes from known *Chlorobiales* genomes is so great (Fig. 1.4c) that, were it not for the presence of functionally diagnostic genes in mat metagenomes (see Fig. 1.2c and Bryant et al., Chapter 3), we would be hard pressed to even consider these genes as being associated with populations that have the metabolic characteristics of green bacteria. However, we are able to able to infer many physiological and metabolic characteristics of mat *Chlorobi* from metagenomic assemblies (Klatt et al., 2011) and metatranscriptomic analyses (Liu et al., 2011).

III. Evolutionary Context

In a thoughtful introduction to their ecology text, Begon et al. (1990) wrote *"If, as T.H. Dobzansky said, 'nothing in biology makes sense, except in the light of evolution', then equally, very little in evolution makes sense except in the light of ecology."* In other words they offered that "*The* ultimate *explanation of the present distribution and abundance of this bird* [read genome or individual organism containing it] *lies in the ecological experiences of its ancestors"*. The point is that ecology and evolution are inseparable. To understand the current ecology of any organism it is necessary to think of the organism as being in the current state of a long evolutionary history in which the organism and its genome were molded through struggles for survival with variants like it in the face of environmental forces that act upon variation (Ward et al., 2008).

[3]While there is debate about what to call the major sublineages of Domains and we don't completely agree on this, the first author prefers the use of Kingdom, which is consistent with the tradition of calling Eukaryal lineages containing animals, plants and fungi Kingdoms.

A. What Do the Properties of an Individual Tell Us About the Individual's Relatives?

To address the question of what the genome of an individual means to the ecological network within a community, it is first necessary to consider whether individuals are grouped into species populations, in which all members of a species share certain properties. If not, the characteristics of an individual genome may represent only the characteristics of the individual cell that contains it, or, at best, those of other individuals within a clonal population of such organisms. If, however, individuals are grouped into species, at least some of the information contained in an individual genome may typify other individuals within the species. Here we encounter the problem that, as with plant and animal biologists (Mayden, 1997; de Quieroz, 2005; Rosenzweig, 1995), microbiologists have argued about what species are and have invented numerous species concepts (Fraser et al., 2009; Ward, 1998, 2006; Ward et al., 2008; Cohan and Perry, 2007; Cohan and Koeppel, 2008). In the mat community we study, there is strong evidence that species, at least *Synechococcus* species, which we have studied most to date, exist as ecologically distinct populations adapted uniquely to specific niches (i.e., ecological species; van Valen, 1976). The first evidence supporting this inference came from the unique distributions of closely related SSU rRNA sequence variants along flow (Ferris and Ward, 1997) and vertical (Ramsing et al., 2000) gradients, suggesting adaptations to temperature, nutrients and light. As mentioned above, we have been able to confirm temperature adaptations in *Synechococcus* isolates (Allewalt et al., 2006). Also, the existence of genetically distinct *Synechococcus* populations with differential depth distribution within the ~1 mm-thick photic zone (Ramsing et al., 2000) is consistent with the microscopic observation of differently pigmented *Synechococcus* populations at different depths (Fig. 1.1c) and suggests that this phenotypic difference has a genetic basis (i.e., differently adapted populations as opposed to one population differently acclimated under different environmental conditions). However, we discovered that SSU rRNA sequences, which are highly conserved, might offer insufficient molecular resolution to discover very closely related, yet ecologically distinct, populations. By examining the internal transcribed spacer (ITS) region separating SSU rRNA and large subunit rRNA genes, we achieved higher molecular resolution and could begin to see individual genetic variants that cluster together phylogenetically (Fig. 1.5); these clusters seemed to correlate with unique ecological distributions (Ferris et al., 2003).

These evolutionary (i.e., phylogenetic) and ecological (i.e., distribution) patterns are consistent with the predictions of the Stable Ecotype model of species and speciation (Ward and Cohan, 2005; Cohan and Perry, 2007), which predicts that a species is a long-lived group of individuals occupying a distinct niche, and that species rarely split to form new species. The individuals exhibit variation, but are ecologically similar enough to remain constrained by the same niche-determining parameters. From time to time, against the varying environment, a most-fit variant is selected via periodic selection to the exclusion of other less-fit individuals of the species and the descendants of this individual give rise to new variants with time; alternatively, variation within some species may be purged by genetic drift. In the Stable Ecotype model, a species is thus seen as a population of ecologically similar individuals that evolves through progressive selective sweeps of diversity. Occasionally, however, mutations (including recombination involving the movement of genes and gene cassettes from distantly or closely related organisms) may give a particular individual a different kind of ecological advantage, one that allows it to occupy a new niche. The next periodic selection will not affect such a variant, because it will have an alternative ecological niche. This kind of variant founds a new population that occupies this new niche and undergoes its own periodic selection events. The one-species population thus diverges into two-species populations. We may argue about when the distinction between populations is sufficient to claim they belong to distinct species (de Queiroz, 2005; Ward, 2006), but, in this model, they are ecologically distinct from the beginning. It is important to appreciate that taxa that have evolved under different circumstances may speciate in different ways (Cohan and Perry, 2007; Ward et al., 2008).

We developed an evolutionary simulation based on the Stable Ecotype Model (Ecotype

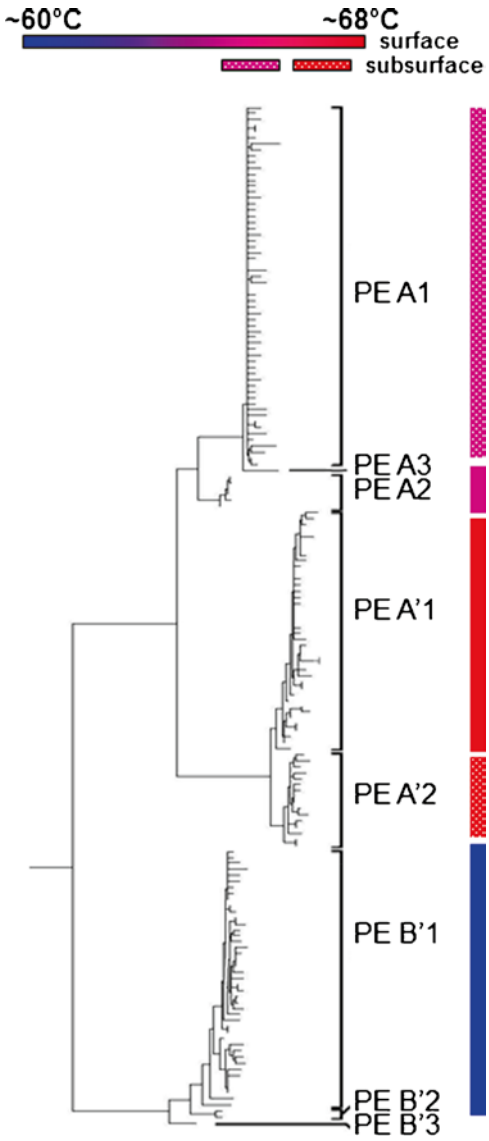


Fig. 1.5. Phylogenetic diversity of mat *Synechococcus* based on SSU-large subunit rRNA internal transcribed spacer sequences retrieved by PCR amplification. Putative ecotypes (PE) were demarcated by Ecotype Simulation analysis. A, A′ and B′ correspond to SSU rRNA sequence types. Highlighting indicates predominant association of different PE clades with temperature (*blue to red indicates* 60–68°C) and depth interval (solid for surface; stippled for subsurface) of collection site. (Modified from Ward et al., 2006).

Simulation), which predicts how genetic variants can be grouped into putative ecological species populations (Koeppel et al., 2008). We used the simulation to demonstrate how individual ITS variants are grouped into 2–3 A-like, A′-like and B′-like *Synechococcus* putative ecological species clades (Ward et al., 2006) (Fig. 1.5). Importantly, an ecological species should contain ecologically interchangeable individuals (i.e., each individual of a species has the same niche adaptations as any other member of the same species). In this regard it is particularly important to note that many of the members of these predicted species clades were obtained from samples collected at different temperatures or mat depths (i.e., these subclades appear to group ecologically similar individuals). Effectively, each SSU rRNA variant lumps several ITS-defined species populations. In our current work on genetic variation in protein-encoding genes, which offer yet higher molecular resolution, we have found evidence suggesting that each SSU rRNA-defined population contains up to ~15 clades predicted by the evolutionary simulation to be distinct species, which exhibit unique patterns of distribution along environmental gradients, unique population dynamics in response to environmental change and unique gene expression patterns (Melendrez et al., 2011; ED Becraft, FM Cohan, M Kühl, S Jensen, and DM Ward, unpublished). Instead of one cosmopolitan *Synechococcus* species there are apparently scores of *Synechococcus* species just in these Yellowstone mats and, undoubtedly, the same is true in other thermophilic *Synechococcus* lineages, whose members dominate in other geographical areas.

Interestingly, by comparing the genomic sequence of *Synechococcus* sp. strain B′ to mat metagenomic sequences, Bhaya et al. (2007) discovered evidence that the genome of this isolate is not representative of all native B′-like *Synechococcus* populations. Specifically, some native B′-like *Synechococcus* populations, unlike the *Synechococcus* sp. B′ isolate, possess genes that may confer the ability to transport ferrous iron. This was suggested as a possible niche-defining physiological capability (e.g., a population(s) that could exploit the availability of ferrous iron during anoxic periods and might correlate with one or more B′-like *Synechococcus* species). Given the existence of numerous very closely related species populations, as depicted in Fig. 1.6, we must revisit the issue of what a genome of an individual isolate actually represents. For instance, clearly the presence of phosphonate genes in *Synechococcus* sp. B′ and *Roseiflexus* sp. RS1 genomes, or the evidence of

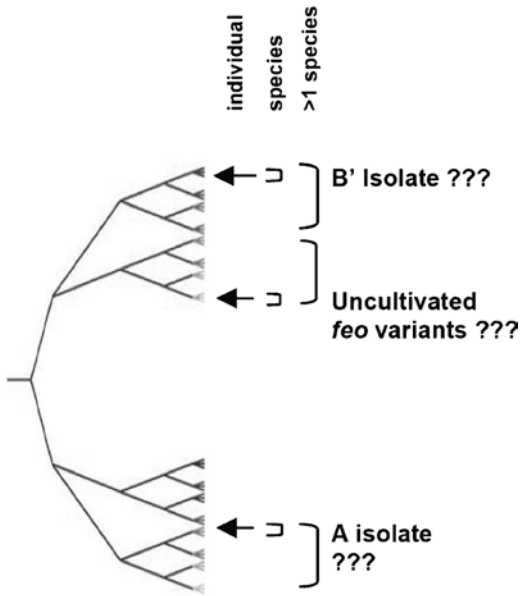


Fig. 1.6. Schematic phylogeny grouping individuals into many species populations in the *Synechococcus* B′-like and A-like lineages highlighting the unknown relationship between properties of individual isolates or environmental clones and (**a**) individuals (*arrows*), (**b**) species populations (*small brackets*) or (**c**) sets of species populations (*large brackets*).

ferrous iron-utilizing B′-like *Synechococcus* variants can be taken to indicate that these are properties of individuals. But can we assume that they are properties typical of all individuals of a species or perhaps of a set of species? We are only on the threshold of being able to discern ecologically meaningful genome features (e.g., gene content, copy number, adaptive substitutions) from those that might merely reflect the ecologically neutral and random coming and going of genes with horizontal gene transfer and/or mutation. It is therefore difficult to attribute adaptation to genomic features, but we can be more confident about inferring adaptation when we observe the same genomic property in many individuals of a species, many species in a lineage, or even across multiple guilds (suggesting that the trait is important in this particular community setting).

B. The Importance of Co-evolution to Ecological Networks

The elegant adaptations of fuchsia-colored monkey flowers to ensure pollination by hummingbirds (who get their sweet nectar reward in return) (Schemske and Bradshaw, 1999) remind us that coevolved relationships may influence the flow of energy and resources among phylogenetically diverged species within a community. The unique distributions of closely related SSU rRNA sequence variants along flow and vertical gradients observed for *Synechococcus* populations were accompanied by similar distributions of closely related *Roseiflexus* SSU rRNA sequence variants (Ferris and Ward, 1997). This suggests that *Roseiflexus* spp. in these mats are also likely to have speciated via adaptive radiation. When thinking about ecological networking within these mat communities, the co-occurrence of particular oxygenic and anoxygenic species in space (and gene expression in time) may foretell of the possible co-evolution among the phylogenetially diverged species, which partner naturally. There is good evidence, for instance, that a major portion of photosynthate produced by *Synechococcus* spp. is stored as polysaccharides that are fermented at night and transferred through subsequent photoheterotrophic uptake to *Roseiflexus* spp. (Konopka, 1992; Nold and Ward, 1996; van der Meer et al., 2005). Such cross feeding may occur between specific, co-adapted partners. The existence of coevolved species would also suggest that it might matter which species are combined in co-culture in attempts to model such interactions.

IV. Optimism for the Near Future

With the advances in genomics, metagenomics, metatranscriptomics, metaproteomics and metabolomic/stable isotope technologies and their declining costs, we will soon be in position to conduct comparative analyses of genomes from replicate members of different species, as well as global analyses of microbial communities. Once we have learned which genes typify different species, we should be able to better understand how individual genomes connect to natural systems. This, in turn, will allow us to soon be able to achieve a predictive knowledge of how microbial systems work, on a scale that ranges from genomes of individuals to the functions and functional networking of species and guilds within communities. We will then be

better able to predict how microbial systems will change when the environment changes and how to use microorganisms, alone or in combination, in artificial systems bioengineered for desired functions. The greatest challenge may be for those engaged in functional genomics and interested in the connectedness of their work to ecological networks to adopt a natural view, one based on the principles of ecology and evolution that gave rise to genomes they study and the relationship of these genomes to other individual genomes and therefore to ecological networking within a system. Brock (1987) suggested during the early days of molecular phylogeny "*How much more satisfying would be our understanding of microbial evolution if we could relate it to microbial ecology*". Equally, how much more satisfying would be our understanding of systems biology if we conducted studies across the full range of system biocomplexity (i.e., from genomes to communities) using principles of ecology and evolution. As you read the chapters of this book, ask yourself how much is known about the ecological and evolutionary context of the genomes under investigation.

Acknowledgments

This research was supported by the National Science Foundation (currently from the Frontiers in Integrative Biology Research Program, EF-0328698 and the IGERT Program in Geobiological Systems, DGE 0654336), the National Aeronautics and Space Administration (most recently from the Exobiology Program, NAG5-8824, -8807 and NX09AM87G) and the U.S. Department of Energy (BER), as part of BER's Genomic Science Program (GSP). This contribution originates from the GSP Foundational Scientific Focus Area (FSFA) at the Pacific Northwest National Laboratory (PNNL), contract no. 112443. We appreciate the support of the National Park Service personnel at Yellowstone National Park. DMW thanks Peter Gogarten (University of Connecticut) for helping him to understand that the value of a genome from an isolate depends on populations genetics (i.e., the relationship of the organism whose genome is under investigation to other related individuals (and their genomes) in a system). This seems obvious today, but was not known by DMW at the time because his traditional microbiology training did not acquaint him with principles of ecology and evolution. D. A. B. additionally and gratefully acknowledges support from the National Science Foundation (MCB-0523100), Dept. of Energy (DE-FG02-94ER20137), and the Joint Genome Institute for support in obtaining genomic and metatranscriptomic sequences mentioned herein.

References

Adams MM, Gomez-Garcia MR, Grossman AR and Bhaya D (2008) Phosphorus deprivation responses and phosphonate utilization in a thermophilic *Synechococcus* sp. from microbial mats. J Bacteriol 190: 8171–8184

Allewalt JP, Bateson MM, Revsbech NP, Slack K and Ward DM (2006) Effect of temperature and light on growth and photosynthesis of *Synechococcus* isolates typical of those predominating in the Octopus Spring microbial mat community. Appl Environ Microbiol 72: 544–550

Anderson KL, Tayne TA and Ward DM (1987) Formation and fate of fermentation products in hot spring cyanobacterial mats. Appl Environ Microbiol 53: 2343–2352

Bateson MM and Ward DM (1988) Photoexcretion and consumption of glycolate in a hot spring cyanobacterial mat. Appl Environ Microbiol 54: 1738–1743

Begon M, Harper JL and Townsend CR (1990) Ecology: Individuals, Populations and Communities. 2nd Ed. Blackwell Science Publishers, Cambridge, MA

Bhaya D, Grossman AR, Steunou A-S, Khuri N, Cohan FM, Hamamura N, Melendrez MC, Nelson, W, Bateson MM, Ward DM and Heidelberg JF (2007) Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. ISME J 1: 703–713

Brock TD (1987) The study of microorganisms in situ: progress and problems. Soc Gen Microbiol Symp 41: 1–17

Bryant DA, Garcia Costas AM, Maresca JA, Gomez Maqueo Chew A, Klatt CG, Bateson MM, Tallon LJ, Hostetler J, Nelson WC, Heidelberg JF and Ward DM (2007) "*Candidatus* Chloracidobacterium thermophilum": an aerobic phototrophic acidobacterium. Science 317: 523–526

Buckley MR (2004) Systems Microbiology: Beyond Microbial Genomics. Am Acad Microbiol. Washington, DC

Castenholz RW (1978) The biogeography of hot spring algae through enrichment cultures. Mitt Internat Verein Limnol 21: 296–315

Castenholz RW (1981) Isolation and cultivation of thermophilic cyanobacteria. In: Balows A, Trüper HG, Dworkin M, Harder W and Schleifer K-H (eds) The Prokaryotes, a Handbook on Habitats, Isolation and Identification of

Bacteria, Vol. I, pp 236–246. Springer-Verlag, Berlin/Heidelberg/New York

Cohan FM and Koeppel AF (2008) The origins of ecological diversity in prokaryotes. Curr Biol 18: R1024–R1034

Cohan FM and Perry EB (2007) A systematics for identifying the fundamental units of bacterial diversity. Current Biology 17: R373–R386

de Queiroz K (2005) A unified concept of species and its consequences for the future of taxonomy. Proc Cal Acad Sci 56(Suppl I): 196–215

Ferris MJ, and Ward DM (1997) Seasonal distributions of dominant 16S rRNA-defined populations in a hot spring microbial mat examined by denaturing gradient gel electrophoresis. Appl Environ Microbiol 63: 1375–1381

Ferris MJ, Ruff-Roberts AL, Kopczynski ED, Bateson MM and Ward DM (1996) Enrichment culture and microscopy conceal diverse thermophilic *Synechococcus* populations in a single hot spring microbial mat habitat. Appl Environ Microbiol 62: 1045–1050

Ferris MJ, Kühl M, Wieland A and Ward DM (2003) Different light-adapted ecotypes in a 68°C *Synechococcus* mat community revealed by analysis of 16S-23S intervening transcribed spacer variation. Appl Environ Microbiol 69: 2893–2898

Fraser C, Alm EJ, Polz MF, Spratt BG and Hanage WP (2009) The bacterial species challenge: making sense of genetic and ecological diversity. Science 323: 741–746

Garcia Costas AM, Liu Z, Tomsho LP, Schuster SC, Ward DM and Bryant DA (2011) Complete genome of *Candidatus* Chloracidobacterium thermophilum, a chlorophyll-based phototroph belonging to the phylum *Acidobacteria*. Environ Microbiol, accepted upon minor revision

Gibson J, Pfennig N and Waterbury JB (1984) *Chloroherpeton thalassium* gen. nov. et spec. nov., a non-filamentous, flexing and gliding green sulfur bacterium. Arch Microbiol 138: 96–101

Klatt CG, Bryant DA and Ward DM (2007) Comparative genomics provides evidence for the 3-hydroxypropionate autotrophic pathway in filamentous anoxygenic phototrophic bacteria and in hot spring microbial mats. Environ Microbiol 9: 2067–2078

Klatt CG, Wood JM, Rusch DB, Bateson MM, Hamamura N, Heidelberg JF, Grossman AR, Bhaya D, Cohan FM, Kühl M, Bryant DA and Ward DM (2011) Community ecology of hot spring cyanobacterial mats: predominant populations and their functional potential. ISME J, in press

Koeppel A, Perry EB, Sikorski J, Krizanc D, Warner A, Ward DM, Rooney AP, Brambilla E, Connor N, Ratcliff RM, Nevo E and Cohan FM (2008) Identifying the fundamental units of bacterial diversity: A paradigm shift to incorporate ecology into bacterial systematics. Proc Natl Acad Sci USA 105: 2504–2509

Konopka A (1992) Accumulation and utilization of polysaccharide by hot-spring phototrophs during a light-dark transition. FEMS Microb Ecol 102: 27–32

Liu Z, Klatt CG, Wood JM, Rusch DB, Ludwig M, Wittekindt N, Tomsho LP, Schuster SC, Ward DM and Bryant DA (2011) Metatranscriptomic analyses of chlorophototrophs of a hot-spring microbial mat. ISME J, in press

Madigan MT, Jung DO, Karr EA, Sattley WM, Achenbach LA and van der Meer MTJ (2005) Diversity of anoxygenic phototrophs in contrasting extreme environments. In: Inskeep WP and McDermott TR (eds) Geothermal Biology and Geochemistry in Yellowstone National Park, pp 203–219. Montana State University Publications, Bozeman MT

Mayden RL (1997) A hierarchy of species concepts: the denouement in the saga of the species problem. In: Claridge MF, Dawah HA and Wilson MR (eds) Species: the Units of Biodiversity, pp 381–424. Chapman & Hall, London, UK

Melendrez MC, Lange RK, Cohan FM and Ward DM (2011) Influence of molecular resolution on sequence-based discovery of ecological diversity among *Synechococcus* population in an alkaline siliceous hot spring microbial mat. Appl Environ Microbiol 77: 1359–1367

Nold SC and Ward DM (1996) Photosynthate partitioning and fermentation in hot spring microbial mat communities. Appl Environ Microbiol 62: 4598–4607

Nübel U, Bateson MM, Vandieken V, Kühl M and Ward DM (2002) Microscopic examination of distribution and phenotypic properties of phylogenetically diverse *Chloroflexaceae*-related bacteria in hot spring microbial mats. Appl Environ Microbiol 68: 4593–4603

Papke RT, Ramsing NB, Bateson MM and Ward DM (2003) Geographical isolation in hot spring cyanobacteria. Environ Microbiol 5: 650–659

Ramsing NB, Ferris MJ and Ward DM (2000) Highly ordered vertical structure of *Synechococcus* populations within the one-millimeter thick photic zone of a hot spring cyanobacterial mat. Appl Environ Microbiol 66: 1038–1049

Revsbech NP and Ward DM (1984) Microelectrode studies of interstitial water chemistry and photosynthetic activity in a hot spring microbial mat. Appl Environ Microbiol 48: 270–275

Rosenzweig ML (1995) Species diversity in space and time. University Press, Cambridge, MA

Santegoeds CM, Nold SC and Ward DM (1996) Denaturing gradient gel electrophoresis used to monitor the enrichment culture of aerobic chemoorganotrophic bacteria from a hot spring cyanobacterial mat. Appl Environ Microbiol 62: 3922–3928

Schemske DW and Bradshaw HD, Jr (1999) Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). Proc Natl Acad Sci USA 96: 11910–11915

Steunou A-S, Bhaya D, Bateson MM, Melendrez MC, Ward DM, Brecht E, Peters JW, Kühl M and Grossman AR (2006) In situ analysis of nitrogen fixation and metabolic switching in unicellular thermophilic cyanobacteria

inhabiting hot spring microbial mats. Proc Natl Acad Sci USA 103: 2398–2403

Steunou A-S, Jensen SI, Brecht E, Becraft ED, Bateson MM, Kilian O, Bhaya D, Ward DM, Peters JM, Grossman AR and Kühl M (2008) Regulation of *nif* gene expression and the energetics of cyanobacterial N_2 fixation over the diel cycle in a hot spring microbial mat. ISME J 2: 364–378

van der Meer MTJ, Schouten S, de Leeuw JW and Ward DM (2000) Autotrophy of green non-sulphur bacteria in hot spring microbial mats: biological explanations for isotopically heavy organic carbon in the geological record. Environ Microbiol 2: 428–435

van der Meer MTJ, Schouten S, Sinninghe-Damsté JS, de Leeuw JW and Ward DM (2003) Compound-specific isotopic fractionation patterns suggest different carbon metabolisms among *Chloroflexus*-like bacteria in hot spring microbial mats. Appl Environ Microbiol 69: 6000–6006

van der Meer MTJ, Schouten S, Bateson MM, Nübel U, Wieland A, Kühl M, de Leeuw JW, Sinninghe-Damsté JS and Ward DM (2005) Diel variations in carbon metabolism by green nonsulfur-like bacteria in alkaline silicious hot spring microbial mats from Yellowstone National Park, USA. Appl Environ Microbiol 71: 3978–3986

van der Meer MTJ, Schouten S, Sinninghe Damsté JS and Ward DM (2007) Impact of carbon metabolisms on ^{13}C signatures of cyanobacteria and green nonsulfur-like bacteria inhabiting a microbial mat from an alkaline siliceous hot spring in Yellowstone National Park (USA). Environ Microbiol 9: 482–491

van der Meer MTJ, Klatt CG, Wood J, Bryant DA, Bateson MM, Lammerts L, Schouten S, Sinninghe Damsté JS, Madigan MT and Ward DM (2010) Cultivation and genomic, nutritional and lipid biomarker characterization of *Roseiflexus* strains closely related to predominant *in situ* populations inhabiting Yellowstone hot spring microbial mats. J Bacteriol 192: 3033–3042

VanValen L (1976) Ecological species, multispecies, and oaks. Taxon 25: 233–239

Ward DM (1998) A natural species concept for prokaryotes. Curr Opinion Microbiol 1: 271–277

Ward DM (2006) A macrobiological perspective on microbial species. Microbe 1: 269–278

Ward DM and Cohan FM (2005) Microbial diversity in hot spring cyanobacterial mats: pattern and prediction. In: Inskeep B and McDermott T (eds) Geothermal Biology and Geochemistry in Yellowstone National Park, pp 185–201. Montana State University Publications, Bozeman MT

Ward DM, Weller R and Bateson MM (1990) 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature 345: 63–65

Ward DM, Bateson MM, Ferris MJ and Nold SC (1998) A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. Microbiol Mol Biol Rev 62: 1353–1370

Ward DM, Nübel U, Papke RT and McKitrick MC (2002) Natural history of microorganisms inhabiting hot spring microbial mats: clues to the origin of microbial diversity and implications for microbiology and macrobiology. In: Staley JT and Reysenbach A-L (eds) Biodiversity of Microbial Life: Foundation of Earth's Biosphere, pp 25–48. John Wiley and Sons, NY

Ward DM, Bateson MM, Ferris MJ, Kühl M, Wieland A, Koeppel A and Cohan FM (2006) Cyanobacterial ecotypes in the microbial mat community of Mushroom Spring (Yellowstone National Park, Wyoming) as species-like units linking microbial community composition, structure and function. Phil Trans Royal Soc London (Ser. B) 361: 1997–2008

Ward DM, Cohan FM, Bhaya D, Heidelberg JF, Kühl M and Grossman AR (2008) Genomics, environmental genomics and the issue of microbial species. Heredity 100: 207–219

Winogradsky S (1949) Microbiologie du sol. Problemes et methodes. Barneoud Freres, France

Chapter 2

Probing Functional Diversity of Thermophilic Cyanobacteria in Microbial Mats

Devaki Bhaya*
Department of Plant Biology, Carnegie Institution for Science, 260 Panama Street, Stanford, CA 94305, USA

Summary

In recent years there has been growing appreciation of the unexpected genetic diversity of microbes in the environment. This diversity has important implications for our understanding of photosynthesis in populations and in the environment. Conventional methodologies often cannot effectively capture this aspect.

*Author for correspondence, e-mail: dbhaya@stanford.edu

R.L. Burnap and W.F.J. Vermaas (eds.), Functional Genomics and Evolution of Photosynthetic Systems, Advances in Photosynthesis and Respiration 33, pp. 17–46, DOI 10.1007/978-94-007-1533-2_2,

This chapter describes new approaches including comparative genomic and metagenomic approaches combined with a more detailed understanding of the metabolism and functionalities of cyanobacteria. This approach, which can be defined broadly as "functional ecogenomics" is partly motivated by the availability of high-throughput sequence data, which are steadily being acquired. The focus is on unicellular cyanobacteria in the hot spring microbial mats of Yellowstone National Park, which are primary producers in this prokaryotic community. We took a three-pronged approach, in which we (a) acquired complete genome sequences from two dominant *Synechococcus* sp. and carried out a comparative genomic analysis to understand the functional differences between these temperature adapted isolates; (b) produced a metagenome dataset that allows us to place genomic information in the context of the community within which these cyanobacteria grow and evolve; and (c) obtained pure isolates of some dominant organisms that allow us to manipulate them in a well-defined laboratory setting. *In situ* transcriptomics has allowed quantification of transcripts during the diel cycle. These diverse approaches and the ability to measure environmental parameters such as light and O_2 levels allow us detailed insight into the microbial mat system. Such an approach could be used to study a wide array of photosynthetic microorganisms as populations and interacting communities. As sequencing capacity, single cell capture techniques, proteomics and imaging techniques become more widely accessible we expect to obtain ever more detailed information about natural communities.

I. Introduction

Fundamental insights into the functional components of oxygenic photosynthesis have been derived from experiments conducted with model cyanobacteria. Cyanobacteria are the progenitors of plastids in all vascular plants and much of the photosynthetic machinery of these vastly separate lineages still function in similar ways (Howe et al., 2008; Archibald, 2009). Thus, the cyanobacterial photosynthetic apparatus is considered to be a useful proxy for the study of photosynthetic processes in all oxygenic photoautotrophs. Crystal structures of photosynthetic proteins and complexes, elucidation of pathways of photosynthetic electron flow, and biophysical features of photosynthetic reactions have used model cyanobacteria (such as *Synechococcus elongatus* PCC 7942 {also known as *Anacystis nidulans* R2}, *Synechocystis* sp. strain PCC 6803 and *Thermosynechococcus elongatus*). The major advantages in using cyanobacteria for studying photosynthesis and other processes are: (i) they have a much shorter life cycle than most plants and algae, (ii) unicellular cyanobacteria in particular represent a uniform population of cells that all may respond similarly within the population, (iii) some cyanobacteria are able to grow heterotrophically, so mutants can survive if photosynthesis has been eliminated, (iv) it is easy to grow large volumes of cells for the isolation of cell fractions or for biochemical analyses, (v) many cyanobacteria are easily manipulated genetically (e.g., targeted disruption of specific genes), (vi) they have relatively small genomes, without introns, facilitating the identification of coding regions in comparison with eukaryotic photosynthetic organisms.

In recent years there has been growing appreciation of the unexpected genetic diversity of microbes in the environment (see Ward et al., Chapter 1). This molecular diversity has important implications for functional genomics and particularly for our understanding of photosynthesis in populations and in the environment. Conventional methodologies that rely on isogenic populations of model organisms and gene-by-gene analysis often cannot effectively capture this aspect. Thus, the goal of this chapter will be to describe some new approaches and to emphasize their potential. We have attempted to bring together perspectives from a comparative genomic and metagenomic approach, combined with those generated by an increasingly detailed understanding of

Abbreviations: DGGE – denaturing gradient gel electrophoresis; GNSLB – green non-sulfur bacteria; Hlips – high-light-inducible proteins; ITS – internal transcribed spacer; MePhn – methyl phosphonate; Phn – phosphonate; ROS – reactive oxygen species; TMHs – transmembrane helices; YNP – Yellowstone National Park

the metabolism and functionalities of cyanobacteria in the hot spring microbial mats. This approach, which can be defined broadly as "functional ecogenomics" is motivated by the availability of the vast wealth of information derived from genomics and metagenomics projects and focuses on the dynamic aspects of gene expression.

A. Cyanobacteria: Ubiquitous and Important Members of Communities and Environments

The revolution brought about by using high-throughput genomic sequencing has made an enormous impact on the field of microbiology. Traditionally, microbiology relied on the use of axenic cultures maintained in the laboratory coupled to a variety of approaches ranging from biochemistry to gene expression studies. However, the vast majority of bacteria in the environment cannot be easily cultivated and therefore are not amenable to traditional microbiological methodologies. This is a serious drawback if one wishes to understand microbial function in the context of the environment. Thus, the ability to leapfrog over the arduous and difficult step of getting axenic cultures by obtaining DNA sequence information directly from environmental samples has provided major insights. One important new insight is that the microbial universe is almost unbelievably diverse and that we know almost nothing about the range of diversity that exists for microbes in nature. Another important insight is that microbes can be found in significant numbers in almost any environment from the most mundane to the most extreme (Whitman et al., 1998). Thirdly, as microbial diversity is being explored with these powerful new genomic tools it is also becoming clear that most microbes in the environment do not "hack it alone" but exist and thrive as members of communities or consortia (Schloss and Handelsman, 2007; Cardenas and Tiedje, 2008; Wilmes et al., 2009).

Cyanobacteria, in particular, are ubiquitous (Whitton and Potts, 2000) and have the ability to thrive in a range of extreme environmental conditions including thermophilic (Brock, 1978; Ley et al., 2006) and psychrophilic environments (Christner et al., 2003) as well as in desert crusts where they can withstand extreme desiccation (Nagy et al., 2005). Cyanobacteria also form critical associations with other microbes in metabolically integrated consortia, and function as symbionts supplying carbon to fungi, plants and animals (Adams, 2000; Adams and Duggan, 2008). These features make cyanobacteria one of the most versatile and enduring groups of photosynthetic organisms on the Earth. Although freshwater organisms such as *Synechocystis* sp. strain PCC 6803, *Synechococcus elongatus* PCC 7942 and filamentous cyanobacteria such as *Nostoc* sp. (*Anabaena* sp.) strain PCC 7120 have been studied extensively as model organisms for various processes, we have very limited knowledge of genomic diversity and its relationship to photosynthetic potential among terrestrial cyanobacteria.

In environments such as the oceans it has been estimated that cell counts in surface water may exceed 10^5 cells per ml, which would amount to a total of ~3.6×10^{29} microbial cells with a total cellular carbon content of $\approx 3 \times 10^{17}$ g (Whitman et al., 1998). In the oceans, cyanobacteria are important and abundant primary producers; as a consequence, they play an important role in primary productivity and the cycling of inorganic carbon in diverse environments (Falkowski et al., 2008). They are also important players in the global biogeochemical cycling of nutrients since they can fix inorganic carbon, reduce molecular nitrogen, ferment sugars, and can alter the redox state of iron and sulfur compounds (Cohen et al., 1975; Partensky et al., 1999; Paerl et al., 2000; Guerrero et al., 2002; Teske and Stahl, 2002; Decker et al., 2005). These activities impact other microbes that associate with the cyanobacteria, and may be important driving forces that shape microbial interactions. Recent genomic sequencing efforts with various ecotypes of marine *Synechococcus* sp. and *Prochlorococcus* sp. have greatly expanded our knowledge of cyanobacterial diversity in the marine environment and demonstrated the potential of a comparative genomic approach. Several recent articles and reviews cover this active area of research (Suzuki and DeLong, 2002; DeLong and Karl, 2005; Coleman et al., 2006; Kettler et al., 2007; Frias-Lopez et al., 2008; Haverkamp et al., 2008).

The focus of this chapter will be on the thermophilic cyanobacterial populations in the microbial mats of Yellowstone hot springs, which have been a test case for the development of some relevant methodologies. Progress in the use of comparative genomic and metagenomic tools to

understand evolving functions in environmentally relevant photosynthetic communities will be discussed. This is a relatively young field that has the potential to provide unique insights that are not accessible through the study of isogenic populations of model organisms. It also holds the promise of understanding evolutionary processes in bacteria within the context of an environment that has both predictable (e.g., day/night light levels) and unpredictable (e.g., cloud cover or nutrient fluxes) dynamic fluctuations.

B. Cyanobacteria in Microbial Mats and Biofilms

Microbial mats are considered modern-day analogs of ancient ecosystems represented by stromatolites, in which mineral depositions within a copious exopolysaccharide matrix have preserved the stratified cyanobacterial community structure (Schopf, 2000; Stal, 2000; Teske and Stahl, 2002). The oxygenic, photoautotrophic cyanobacteria are believed to have pioneered the formation of these early Earth communities ~2.5 billion years ago, and may have contributed to the oxygenation of the early Earth's atmosphere (Hoehler et al., 2001). Photosynthetic microbial mats occur in many terrestrial and aquatic environments such as hypersaline coastal pools of Guererro Negro (Ley et al., 2006), freshwater ponds, geothermal hot springs of Yellowstone National Park (Ward et al., 2002), cold dry valleys of Antarctica (Vincent, 2000; Vincent et al., 2004; Jungblut et al., 2006), and alkaline and low sulfide hot springs in Russia (Namsaraev et al., 2003). Photosynthetic microbial mats also form crusts on rocks (Stal, 2000; Wynn-Williams, 2000; Arakawa et al., 2006).

Some of these types of mats have extremely high calculated ratios of carbon assimilation to standing biomass, suggesting highly efficient carbon utilization. It has been suggested that microbial mats in marine environments may be much more efficient in nutrient acquisition and utilization than mixed planktonic populations (Paerl, 2000; Guerrero et al., 2002; Decker et al., 2005). Microbial mats can flourish in regions where predation is low, and stratified microbial mats tend to proliferate in diverse environments that are often inhospitable, so they provide a perfect paradigm for studying how moderately complex communities of microbes develop and optimize the utilization of scarce resources. Most but not all mat communities are comprised of a limited number of dominant prokaryotic genera, and the system is less complex than soil or marine ecosystems (Nubel et al., 1999, 2002; Paerl et al., 2000; Guerrero et al., 2002; Ley et al., 2006). In photosynthetic microbial mats, cyanobacteria and other prokaryotes can position themselves at various interfaces (such as the sediment/liquid interface or the air/surface interface) as a function of light, chemical and gas gradients. This leads to the formation and stabilization of biologically stratified layers, within which diverse metabolic processes can occur on a temporal scale (e.g., a diurnal or seasonal cycle). This partitioning of nutrient cycling, niche differentiation and homeostasis within the mat community may promote biological control of the microenvironment. This in turn can support higher survival rates than may be possible for individual species growing alone (Paerl et al., 2000). Finally, this biological partitioning can also influence trapped or underlying sediments and mineral precipitation so that the mats can take on a number of different morphological forms such as hard lithified crusts, laminated structures or loose biofilms (Reid et al., 2000).

C. Microbial Mat Communities in the Hot Springs of Yellowstone National Park

Yellowstone National Park (YNP) is unique in that it contains a vast array of geothermal features, several of which have been studied extensively from a geochemical, biological and historical perspective over many decades. There is an extensive literature pertaining to the geology and biology of YNP (for example, Allen and Day, 1935; Brock, 1978; Keiter and Boyce, 1991; Ward et al., 1998; Reysenbach and Cady, 2001; Reysenbach and Shock, 2002; Teske and Stahl, 2002; Inskeep and McDermott, 2005; Sheehan et al., 2005) (see Ward et al., Chapter 1). Yellowstone National Park, which was established in 1872 and was the first national park in the USA, is a protected environment, so carrying out experiments requires permission and co-ordination with the Park authorities. At any one time, many experiments are being carried out within the Park (http://www.nps.gov/yell/naturescience/ynpconditions.htm). Experimental sites in the Park are carefully maintained, which

enables return site visits over several years. This is particularly valuable for any experiments that may require data from a time series, although the springs and channels themselves have been known to change course or exhibit other alterations over time (Brock, 1978).

The alkaline siliceous springs such as Octopus and Mushroom Springs (Fig. 2.1a) are located in the White Creek drainage area of the Lower Geyser Basin and have been studied over many decades by geochemists, ecologists and microbiologists. Therefore, there is extensive literature on aspects of physiology, biogeochemistry and the identification of species in these environments. Indeed the identification, purification and use of thermostable DNA polymerase from *Thermus aquaticus* (a species first identified and studied from these springs by microbiologist Thomas Brock) was instrumental in improving the efficiency of the polymerase chain reaction, which is a routine and widely used technique in molecular biology for the amplification of DNA. The varied geothermal features and their associated microbes are actively studied by many groups aimed at finding uncharacterized bacterial phyla and enzymes with particularly valuable qualities (Brock, 1997; Botero et al., 2004; Podar and Reysenbach, 2006; Sato et al., 2007; Wheeler et al., 2008). These environments also provide a unique opportunity to study the interactions among different prokaryotes and following the pioneering research of Thomas Brock (Brock, 1978), many of his students and other investigators have made significant contributions to our understanding of photosynthetic communities in these extreme environments.

Variously colored (orange to brownish green), often dense microbial mats are formed in the alkaline siliceous hot spring effluent channels (Fig. 2.1b) and these have been extensively documented, examined and categorized (Brock, 1978; Inskeep and McDermott, 2005). The heated water from the source pool gradually cools in the effluent channels, so stable temperature and flow gradients are formed and a variety of microorganisms flourish between particular temperature ranges. Typically, these mats are formed by stable and simple communities of microorganisms such as cyanobacteria (predominantly *Synechococcus* sp.) and green non-sulfur bacteria (GNSLB), such as *Roseiflexus* and *Chloroflexus* sp. There are also less well-characterized heterotrophic anaerobic and aerobic bacteria that are found in these particular mats (Pierson and Castenholz, 1974; Brock, 1978; Ruff-Roberts et al., 1994; Ward et al., 1998; Ward and Castenholz, 2000). The metabolic activities of these mat-forming organisms create stratified layers, within which steep and fluctuating gradients of light, oxygen and nutrients can exist (Stal, 2000; Ferris et al., 2003). Molecular, microsensor and biochemical approaches have been used to measure the metabolism and diversity of bacteria within the microbial mat communities (Revsbech and Ward, 1984; Stal and Caumette, 1994; Kuhl et al., 1998; Kuhl, 2005).

Fig. 2.1a. Octopus Spring in Yellowstone National Park, Summer 2007 (Photograph courtesy of Sheila Ingraham Jensen and Melissa Adams).

Fig. 2.1b. Close up of an effluent channel of an alkaline siliceous hot spring showing green/yellow green mats (Photograph courtesy of Ilina Bhaya-Grossman, Summer 2006).

The hot spring mats also provide an ideal setting for understanding how thermophilic microbes have adapted to a particular temperature range as well as delving into the question of what sets the upper temperature limits of life. Questions such as why certain phyla or ecotypes within a bacterial species are better adapted to a particular temperature or niches and how these processes are evolving are also of wide interest to ecologists, evolutionary biologists and microbiologists. Despite the growing interest in, and concern with, the effects of global warming on different environments and the macrofauna and flora, surprisingly little attention has been paid to the possible effects on microbes. These hot spring environments, which have stable temperature gradients and where microbes have evolved to deal with high temperatures, offer an opportunity to witness and investigate some of these effects.

In summary, the advantages of using the hot spring microbial communities to further understand the dynamics and function of photosynthetic microbes are: (i) these mats are stable, simple, stratified prokaryotic communities, (ii) there is a rich diversity of micro-organisms present at the temperature range from ~40°C to 70°C although it is not as complex as certain environments such as soil, (iii) there are steep and fluctuating gradients of various environmental parameters such as light, oxygen etc. and microsensor data of these parameters can be acquired over a diel cycle or over different seasons, (iv) there is extensive 16S RNA-based phylogenetic analysis of cyanobacteria in the mats, and (v) the ecophysiology of the mats has been studied extensively for decades. These studies set the stage for a deeper exploration of functional genomics in photosynthetic microbes within the context of the environment to which they have successfully adapted. These attractive features can also be used to understand and integrate across different levels of organization, from regulation at the level of single cells to community dynamics.

II. Cyanobacteria as Primary Producers in Hot Spring Mats

Knowledge of microbial mat physiology is crucial for the development of the functional genomics of thermophilic cyanobacteria. Microbial physiology, in turn, is linked to knowledge about the geochemistry, the major 'players' in the community, and their energy requirements and interactions. As mentioned above, mat communities (at Yellowstone Park and elsewhere) have been studied in great detail by many groups, so it is impossible to provide a comprehensive bibliography within the scope of this review. In this section only a brief summary is provided. For descriptions of microbial biodiversity and the techniques initially used to probe diversity in hot springs the reader is referred to Brock (1978), Stal and Caumette (1994), Ward et al. (1998), Whitton and Potts (2000), Teske and Stahl (2002), Krumbein et al. (2003), and Inskeep and McDermott (2005). The focus of this review will be on selected new techniques and their use to probe function and diversity in the context of mat physiology, structure and diel changes.

A. Mat Structure and Community Members

The dense, high biomass microbial communities in the effluent channels of the hot springs of YNP contain a diversity of microorganisms that range from phototrophs such as cyanobacteria and green non-sulfur bacteria, to heterotrophic anaerobic and aerobic bacteria (Brock, 1978; Ward and Castenholz, 2000; Ferris et al., 2001). At the lower temperature range (40–50°C) mats may be dominated by *Phormidium* or *Plectonema* species, which are filamentous cyanobacteria (Ward et al., 1998; Ward and Castenholz, 2000; Lau et al., 2005). At higher temperatures these are replaced by the unicellular cyanobacterium *Synechococcus* sp. These highly fluorescent rod-shaped cyanobacteria (~1 μm wide and 4–6 μm long) are typically found in the top green layer of mats where they can carry out photosynthesis, while the lower orange-pigmented layers of the mat contain members of the filamentous anoxygenic phototrophs or green non-sulfur bacteria as well as other less well-identified heterotrophs. Above ~72°C, cyanobacteria cannot survive and the archaea become common, while below 45°C the cyanobacterial populations are grazed extensively by larger copepods, etc. (Brock, 1978). Thus, between 45°C and 72°C, there is a stable temperature gradient and cyanobacteria, predominantly *Synechococcus* spp., are found along this gradient. No evidence of other dominant filamentous cyanobacteria was noted at these higher temperatures

although in other microbial mats it is quite common to have filamentous cyanobacteria as major mat constituents.

Different groups of microbes have been identified in microbial mat consortia, including (i) oxygenic phototrophs, such as the cyanobacteria, (ii) anoxygenic phototrophs, primarily purple and green bacteria that can use hydrogen sulfide as an electron donor (or potentially H_2), (iii) aerobic heterotrophs that generate energy by respiring organic carbon thus consuming O_2, (iv) fermenters that use organic carbon or sulfur compounds as electron donors and acceptors, (v) anaerobic heterotrophs, predominantly sulfate-reducing bacteria, that respire organic carbon using SO_4^{2-} as an electron acceptor and producing H_2S, and (vi) sulfide-oxidizing bacteria, many of which are chemolithoautotrophs that oxidize reduced sulfur compounds with O_2 or nitrate as electron acceptors, while fixing CO_2 (Stal and Caumette, 1994; Guerrero et al., 2002; Dupraz and Visscher, 2005). So far, a detailed molecular analysis of the functional pathways in the context of carbon cycling in the mat community has not yet been established. The identification of enzymes that are unique to a pathway or key in an organism is the first step towards developing a thorough understanding of critical pathways within microbial communities; such understanding can be tested and further expanded by techniques such as global microarray analysis (Gentry et al., 2006).

B. Metabolism and Diel Cycling Events in the Mat

The chemical composition and pH of hot springs vary substantially by spring location, making generalizations difficult. Furthermore, fluctuations over diel and seasonal cycles are not easily captured. However, the environment in many mat communities is nutrient-poor and especially deficient in nitrogen (N) and phosphorus (P); levels of iron (Fe) and sulfur (S) compounds vary (Brock, 1978; Stal, 2000; Papke et al., 2003; Ludwig et al., 2006). Mats undergo dramatic changes in metabolic processes over the diel cycle, so the organisms in the mat may have evolved a temporally complementary set of metabolisms (van der Meer et al., 2005). During the day, under conditions of high light, the mat becomes highly oxic, with O_2 levels within the matrix of the mat reaching up to 8 times air saturation. The cyanobacteria fix CO_2 via the reductive pentose-phosphate pathway and export a considerable portion of the fixed carbon that was generated (see below). This sustains other members of the microbial community (including the photoheterotrophic GNSLB, such as *Chloroflexus* and *Roseiflexus*). They also secrete exopolysaccharides that form part of the dense matrix, within which the microbes survive. This exopolysaccharide matrix plays multiple roles, such as serving as a diffusion barrier, providing substrates for growth of heterotrophs, binding certain heavy metals, and controlling calcium carbonate precipitation or lithification in mats (Paerl et al., 2000; Dupraz and Visscher, 2005).

Fixation of CO_2 in mats can also be achieved by organisms other than the cyanobacteria, and by processes other than the reductive pentose phosphate pathway (see Chapters 3 and 9 in this volume). For example, based on carbon fractionation data, the GNSLB appear to fix CO_2 via the novel cyclic 3-hydroxypropionate pathway (Holo, 1989; van der Meer et al., 2005). This pathway, first discovered in *Chloroflexus*, proceeds in a cyclic manner from acetyl-CoA to 3-hydroxypropionate, which may be released under certain conditions (Strauss and Fuchs, 1993; Herter et al., 2001; Ishii et al., 2004). While the ATP that drives the pathway is likely to come from anoxygenic photosynthesis when light levels are low (the mat is anoxic with photosynthesis being driven by low levels of excitation energy), the source of electrons used in the reductive steps are not known. The dominant carbon compound exported by cyanobacteria under conditions of photoautotrophic growth is acetate, while at night they export fermentation products (mostly ethanol and formate) generated by the breakdown of polyglucose (Staley, 1997; Teske and Stahl, 2002; van der Meer et al., 2005). The mat reaches different characteristic carbon/energy/redox states during the night and day. As the light intensity declines during the late afternoon, the relative ratio of O_2 evolution by photosynthesis to O_2 consumption by respiration begins to decline. Once O_2 uptake exceeds O_2 evolution, the mat becomes anoxic, so that, over a period of minutes, the mat must transition from a consortium functioning under oxic conditions to one that operates anoxically.

C. Phototaxis, Vertical Migration and Positioning in the Mat Environment

In the mat environment, which is stratified in the vertical dimension and is densely packed with phototrophic organisms such as *Synechococcus* sp. (oxygenic phototrophs), GNSLB (such as *Chloroflexus* and *Roseiflexus* sp.) and green sulfur bacteria (such as *Chlorobium* sp.), access to optimal light conditions is of prime importance. Although this aspect has not been studied intensively, a few interesting concepts and measurements are worth noting regarding ordered structure in the mat and phototaxis. Ramsing et al. (2000) used micro-sensors and observed that oxygenic photosynthesis peaked in the uppermost 100–200 μm region of the mat in the morning, but by afternoon this peak had shifted into the deeper layers. Interestingly this correlated with a vertical shift in discrete bands of auto-fluorescence emanating from the *Synechococcus* populations. They were unable to find strong evidence of any diel migration of particular species within the mat and analysis of cells from vertical thin sections revealed that the rod shaped *Synechococcus* sp. were randomly oriented in various parts of the mat with one exception. Around noon, when light levels are high and may be damaging to the photosynthetic apparatus, there was a narrow band of cells that assumed an upright position about 400–800 μm below the surface. This could suggest that cells have the ability to change their orientation in response to light levels and raises the question of how this is perceived by the cell and how the light signal is transduced to the cell surface to cause a change in cell orientation. These results also correlate with related work of Ramsing et al. (1997) showing that particular *Synechococcus* sp. isolates show light-dependent motility. This phototactic motility was a complex phenotype, in which cells showed different rates of motility depending on the light intensity, and movement was observed both toward and away from the light source. Different strains may show differing motility phenotypes, although all strains appeared to produce copious exopolysaccharide trails as they moved along the surface. This preliminary report did not describe motility as a function of light quality, but our unpublished results show that *Synechococcus* sp. isolates from the mat can move rapidly towards white or red light (D. Bhaya, unpublished) (Fig. 2.2). Motility within the mat environment may be quite advantageous in terms of optimizing photosynthesis, since light is strongly attenuated by the mat, and so gliding up into more lighted areas of the mat may be an advantage; conversely, when light is damaging at certain times of the day, it may be optimal to move deeper into the mat and avoid damage to the photosynthetic apparatus. Thus, one might expect these mat cyanobacteria to be able to sense light direction as well as light quality, but this has not yet been examined.

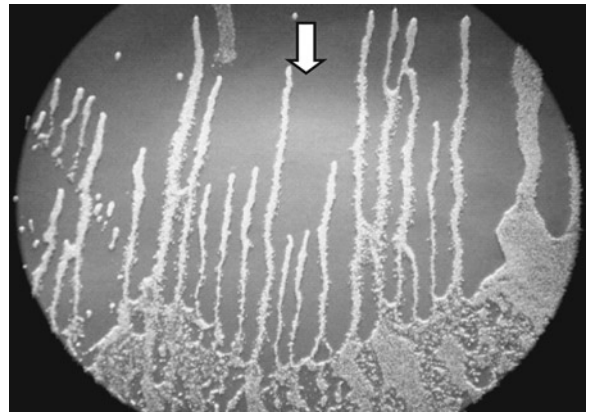


Fig. 2.2. Phototactic motility of *Synechococcus* OS-B′ cells on low concentration agarose plates with a white light source positioned directionally (*arrow*). Note thin fingers of cells that have moved towards the light. (Photograph taken after 48 h of placing on agarose surface).

Our preliminary results show that the *Synechococcus* isolates contain all of the genes required for pilus-dependent motility and thus one might expect that motility is a surface-dependent phenomenon in these cyanobacteria. Light-dependent motility has been characterized in unicellular model cyanobacteria and photoreceptors and other components of the motility apparatus have been identified (Bhaya, 2004; Yoshihara and Ikeuchi, 2004). Extending these studies to environmentally relevant cyanobacteria that have evolved to cope with fluctuating light conditions in a mat community is likely to uncover novel features of photo-movement and perhaps of social communication.

Gliding motility, which is strongly influenced by light, has also been characterized in filamentous thermophilic cyanobacteria such as *Oscillatoria* and *Phormidium*, which are components of hot

spring microbial mats and show vertical migration patterns. Other filamentous bacteria that can be found in hot spring mats (such as *Chloroflexus*, *Heliothrix* and the purple sulfur bacterium, *Chromatium*) also show light-dependent motility but the molecular mechanisms for these movements are still not well understood. Some of these photo-movements may be UV-A irradiation dependent suggesting that complex photo sensory mechanisms may have evolved to protect the phototrophs in mats (Richardson and Castenholz, 1987; Bebout and Garcia-Pichel, 1995; Castenholz and Garcia-Pichel, 2000). Indeed, some cyanobacteria produce UV-protective pigments such as scytonemin, for which part of the biosynthesis pathway has been elucidated (Castenholz and Garcia-Pichel, 2000; Soule et al., 2007).

D. Molecular Techniques to Study Community Structure and Composition

Analysis and examination of the microorganisms in the mat were initially based on classic microbiological tools such as attempts to cultivate microbes under specific enrichment conditions and/or microscopy (Brock, 1978; Ward et al., 1998). These methodologies are used by many laboratories and are still considered to be important but not necessarily comprehensive for identification. Thus, early on, the use of these techniques led to the identification of cyanobacterial species in the mat that were categorized as unicellular *Synechococcus* sp. Similarly, other major microbial components of the mat were identified to be filamentous *Chloroflexus* sp. and the related *Roseiflexus* sp. by Castenholz and others (Pierson and Castenholz, 1974). However, with the development of simple but powerful molecular tools, namely 16S RNA clone sequencing and denaturing gradient gel electrophoresis (or DGGE) to identify bacterial species, the picture that emerged of the microbial populations in the mat became considerably more complex. Both of these techniques are still widely used for the identification and classification of bacterial populations in the environment because of the ease of use (in the case of DGGE) and the ability to use 16S RNA and 16S–23S internal transcribed spacer (ITS) sequences to build phylogenetic relationships (Stahl et al., 1985; Ward et al., 1998). Use of these techniques in the mat environment has shown that the mat contains a very large number and diversity of unique 16S RNA sequences (Ward et al., 1990; Ferris and Ward, 1997). Furthermore, these 16S RNA sequences did not match those of the most readily cultivated isolates from the mat. DGGE studies showed that the distribution of 16S RNA gene variants from cyanobacteria varied along the temperature gradients of the mat. An example of this type of distribution was the characterization of closely related 16S RNA sequences, designated A″, A′, A, B″ and B, that were detected along a temperature gradient ranging from ~70°C to ~50°C (Ferris and Ward, 1997; Ward et al., 1998). Interestingly, 16S RNA sequences of the *Synechococcus* populations were also found to vary predictably along the vertical transect of the mat and appeared to be correlated with the presence of differentially fluorescent *Synechococcus* populations (Ramsing et al., 2000). These detailed molecular studies coupled with other measurements (such as light availability within the depth of the mat) substantiate the view that cyanobacterial (*Synechococcus* sp.) communities within alkaline siliceous mats have well-defined distributions of 16S rRNA sequences and this diversity appears to correlate with established environmental gradients (Ward and Cohan, 2005; Ward 2006a). Questions, some of which will be addressed in subsequent sections (also see Chapter 1), that follow from these results include:

1. Do these *Synechococcus* populations with different 16S sequences have measurable functional differences that can be probed by molecular approaches, and if so, how?
2. Have certain populations "adapted" to a particular niche, and what functional differences have developed? How does this correlate with the regulation of photosynthesis and related processes in the mat?
3. Does this microbial mat system provide a good model system to understand how populations evolve and adapt to environmental fluctuations?
4. Can one begin to understand how different members of a community are interacting metabolically and are sharing resources at a molecular and metabolic level?
5. Can the use of functional genomics and metagenomics clarify notions on the still evolving concept of a "bacterial species" (Konstantinidis and Tiedje, 2004; Gevers et al., 2005; Bhaya et al., 2007; Achtman and Wagner, 2008; Doolittle and Zhaxybayeva, 2009)?

III. Comparative Genomics and Transcriptomics to Probe Function of Closely Related *Synechococcus* spp.

A. Genomic Content and Architecture of Two Related Synechococcus Isolates

To address some of the questions posed above, we took advantage of two available *Synechococcus* isolates (see Ward et al., Chapter 1 in this volume). *Synechococcus* OS-A (*Syn* OS-A) was isolated by dilution culturing (filter cultivation approach) from samples derived from a high-temperature region of the mat (58–65°C), while *Synechococcus* OS-B′ (*Syn* OS-B′) was isolated from low-temperature mat samples (51–55°C) (Allewalt et al., 2006). Allewalt et al. (2006) demonstrated that *Syn* OS-A also showed the highest temperature optimum and upper limit for photosynthesis. This provided partial support to the concept that isolates from different temperature ranges of the mat had growth and photosynthetic characteristics consistent with their location and suggested that these gross measures of "adaptation" may also be reflected at the gene/genomic level.

We acquired complete genome sequences of *Syn* OS-A and OS-B′ by means of shotgun sequencing (Bhaya et al., 2007). Both isolates came from the Octopus Spring mats and contained circular genomes of approximately the same size (~3.0 Mb), exhibited a relatively high G+C% content of 60.3 and 58.5, and included 2,892 and 2,933 predicted coding sequences, respectively (Fig. 2.3). Comparative analysis of these two closely related and mat-dominant cyanobacteria revealed that *Syn* OS-A and *Syn* OS-B′ each have two identical copies of the rRNA genes in their respective genomes. A comparison of the *Syn* OS-A and *Syn* OS-B′ 16S rRNA sequences showed 96.4% identity (i.e., 3.6% differences in sequence). Previous research by Ward and colleagues had demonstrated that (i) *Synechococcus* sp. dominating the mat were substantially different from the readily cultivated *Synechococcus lividus* strains with as much as 8–10% difference in 16S rRNA sequence (Ferris et al., 1996; Ward et al., 1998 and references therein), whereas (ii) five of the predominant *Synechococcus* genotypes (A, A′, A″, B and B′) were more closely related (with < 3% difference

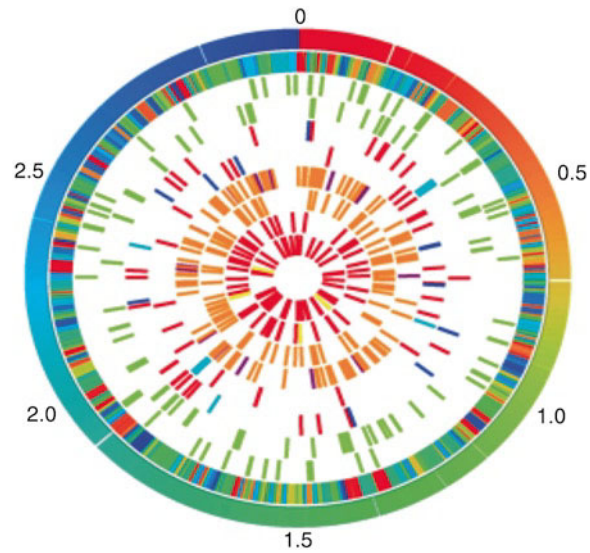


Fig. 2.3. Genomes of *Synechococcus* OS-A and OS-B′. Circle 1 (outermost): *Syn* OS-B′ genes assigned pseudo-spectrum colors based on relative position along the length of the genome, with genes nearest to the putative origin of replication colored red and genes most distal from this origin colored blue. Circle 2: putative homologs in *Syn* OS-A are assigned the same color as is presented in circle 1 for the *Syn* OS-B′ genes; the scrambling of colors in circle 2 is a reflection of the striking lack of relative conserved large-scale gene order between the genomes of Syn OS-A and OS-B′. Circles 3 and 4: photosynthesis genes (*green*). Circles 5 and 6: nitrogen fixation genes (*light blue*), urease genes (*dark blue*) and fermentation genes (*red*). Circles 7 and 8: insertion elements (*orange*) and ISSoc13 transposases (*purple*). Circles 9 and 10: tRNAs (*red*) and rRNAs (*yellow*). Circles 3, 5, 7 and 9 show the relative gene positions on the genome of *Synechococcus* OS-B′. Circles 4, 6, 8 and 10 show the relative gene positions on the genome of *Synechococcus* OS-A. (Modified from Bhaya et al., 2007).

at the 16S rRNA sequence level). The sequences also matched previously identified 16S rRNA sequences from the mat and confirmed that we were working with isolates that were dominant in the mat. Similar comparative analyses regarding phylogenetic relationships based on ribosomal sequences of related marine cyanobacterial ecotypes (*Prochlorococcus* and *Synechococcus* sp.) that have been sequenced have also been carried out (West et al., 2001; Rocap et al., 2002; Ernst et al., 2003; Ahlgren and Rocap, 2006).

A comparison of the coding sequences of the two *Synechococcus* genomes showed that they share a large fraction (~83% based on bidirectional

best BLAST scores) of their coding sequences with a high identity between putative orthologs (~87% amino acid identity on average) (Bhaya et al., 2007). This is consistent with the information that these cyanobacterial isolates are morphologically identical and closely related at the 16S RNA sequence level (which is a well-accepted robust marker for phylogenetic relationships). Despite this, at the level of whole-genome architecture a comparison of *Syn* OS-A and OS-B′ genomes shows a marked lack of synteny or conserved, large-scale gene order, indicating an extensive history of rearrangement events (Fig. 2.3). Comparison of complete genomes of closely related bacteria usually reveals extensive synteny, so this is a surprising deviation that requires further analysis. Regions of co-linearity between the *Syn* OS-A and OS-B′ genomes were short with the largest region of conserved gene order between the two isolates being ~32 kbp (containing the genes required for nitrogen fixation).

Genome rearrangements and recombination events are often mediated by transposons or phage and could play an important role in evolution (Rocha, 2004, 2008). Both *Syn* OS-A and OS-B′ genomes contain many transposon-like or insertion sequence (IS) elements (~100 intact IS genes on each genome as well as many IS gene fragments (Nelson, W., Heidelberg, J., and Bhaya, D, unpublished). *Syn* OS-B′ contains 17 identical copies of an IS4 family of transposase genes (ISSoc13 or Interpro ID 002559) that are completely absent in the *Syn* OS-A genome and that may still be active in *Syn* OS-B′ (Fig. 2.3, Circle 7). We found that the IS elements are not always located at the borders of re-arranged regions, so their role in the large-scale gene rearrangements cannot be easily assigned (Parkhill et al., 2003). These thermophilic *Syn* isolates contain a higher percentage of transposable elements than expected based on their genome size (*Prochlorococcus*, which has a small or "minimal" genome, does not contain any transposable elements) (Zhou et al., 2008). We have examined insertion element content and location in natural populations of *Synechococcus* by comparing metagenomic data to the genomes of the sequenced cultured isolates, to explore the mechanism of IS acquisition within natural populations and survival in the face of high IS abundance (Nelson et al., 2011).

B. Functional Categories and Unique Genes in Genomes and Their Roles in Adaptation

The *Syn* OS-A and OS-B′ genomes contain genes encoding complete sets of proteins required for, among others, photosynthesis, the biosynthesis of glycolate (Bateson and Ward, 1988), glycogen (Bateson and Ward, 1988; Konopka, 1992), sulfolipids (Ward et al., 1994), and fermentative and respiratory metabolisms (Nold and Ward, 1996). We also identified genes required for the biosynthesis of Type IV pili and photoreceptors associated with phototaxis, which fits with earlier reports of motility of *Syn* cells in the mats (Ramsing et al., 1997; Bhaya, 2004). However, the presence and activity of a functional pathway for nitrogen fixation in *Syn* OS-A and OS-B′ were unexpected, as previous attempts had failed to measure nitrogen fixation in the mats at higher temperatures (see Section III.D).

To identify functional differences between *Syn* OS-A and OS-B′, we examined subsets of genes unique to each of these isolates. There are ~400 and ~500 isolate-specific genes in *Syn* OS-A and OS-B′, respectively, of which about half are annotated as either "hypothetical" or "conserved hypothetical". Within this set we identified examples of genes encoding proteins with known functions that are present on one but not on the other genome. Only the *Syn* OS-B′ genome contains genes for the synthesis and metabolism of cyanophycin, a N storage compound. Cyanophycin synthetase is the enzyme that synthesizes cyanophycin non-ribosomally from aspartate and arginine, and cyanophycinase can degrade the polymer to provide the cell with a source of N when needed (Simon, 1987). Cyanophycin levels vary with growth conditions, but can be high in stationary-phase cultures or under conditions in which the growth potential of the cell declines because of a limitation for other nutrients such as sulfate or phosphate. Cyanophycin has also been implicated in the integration of carbon and nitrogen metabolism in unicellular and filamentous cyanobacteria (Mackerras et al., 1990). The ability to store N in the form of cyanophycin granules by *Syn* OS-B′ suggests that it may experience fluctuating nitrogen levels. Since *Syn* OS-A does not appear to have either of these genes it would unable to

store nitrogen as effectively as *Syn* OS-B′ but the significance of this in the context of its environment or 'niche' is not clear.

Another interesting example of genome-specific functionality is the presence of a large 8 kbp region on the *Syn* OS-B′ genome, which contains ten genes (*phn* genes) responsible for the transport and metabolism of phosphonates. This could enable the organism to utilize phosphonate (compounds in which a carbon-oxygen-phosphorus bond is replaced by a direct carbon-phosphorus linkage) as a source of phosphorus in addition to phosphate. Phosphonates are relatively inert, stable compounds and may have preceded phosphates in the early atmosphere when oxygen levels were low. Although the importance of biogenic phosphonates in the terrestrial biosphere has not been established, phosphonate levels are high in the marine environment (Quinn et al., 2007). The entire *phn* gene cluster is missing in *Syn* OS-A, but the region flanking the *phn* cluster is syntenic between the *Syn* OS-A and OS-B′ genomes, indicating that the *phn* gene cluster was either recently acquired by *Syn* OS-B′ or lost in the *Syn* OS-A lineage. There is evidence suggesting that operons required for phosphonate uptake and utilization may be acquired through lateral gene transfer events in prokaryotes (Huang et al., 2005). Recently, genes for phosphonate utilization have been identified in metagenomic studies of marine, oxygenic photosynthetic prokaryotes, but the *phn* operon is not universally found in cyanobacteria, perhaps reflecting the different availability of phosphonates in various environments (Palenik et al., 2003; Dyhrman et al., 2006).

One approach to explore these differential abilities between two closely related cyanobacteria is to use axenic cultures of both isolates under defined laboratory conditions (e.g., low and high N conditions) to explore the benefits of storing cyanophycin or the advantages of being able to use phosphonate as a P source. We have pioneered such an approach with *Syn* OS-B′ with the rationale that some questions are more powerfully addressed with axenic isolates while other questions are much better explored with *in situ* techniques (both will be addressed later in this section).

C. Axenic Cultures to Study Ecologically Important Questions

With the advent of high-throughput genome sequencing and metagenomics there is a flood of information about the genetic repertoire of various bacteria. Although this is a powerful information database that has been exploited in many ways (see Section IV), it still is a big leap to advance from a dictionary of genes in the genome (or the environment in the case of a metagenomics approach) to an understanding of the biology of dominant players in any particular environment. One way to achieve this deeper insight is to be able to work with axenic isolates from the environment of interest. This is not always feasible since only a very small percentage of bacteria can be axenically grown in the laboratory. However, for strains where axenic growth is possible, it opens the door for a number of exciting new areas for research since one can combine *in situ* approaches with more detailed experiments under controlled conditions.

The initial experiments to check if isolates derived from different temperature regions of the mat would show different physiological characteristics consistent with their location were carried out with isolates that were uni-cyanobacterial but not axenic (Allewalt et al., 2006). Although this may not have significantly impacted the interpretation of results, axenic cultures are preferable. Towards that end we repeatedly streaked the enriched cultures on plates at low agarose concentration and placed these plates in directional light (Fig. 2.2). Since *Syn* OS-B′ cells are phototactic, we were able to separate them away from non-motile heterotrophs. We used 16S ribosomal sequencing, growth on nutrient-rich plates (to test for slow-growing contaminants), and phase-contrast microscopy to ensure that the culture was axenic. Two examples of approaches with axenic cultures are described to demonstrate how it has provided insight into the physiology and acclimation of phototrophs to light and nutrients.

1. Acclimation to High and Fluctuating Light Levels

To understand how thermophilic cyanobacteria in microbial mats can respond to fluctuating environmental parameters such as light, we used

axenic isolates of *Syn* OS-B′. The ability to monitor the growth and other key parameters of this isolate under environmentally relevant temperature conditions, as well as prior knowledge about how cyanobacterial cells respond to high-light conditions at the biochemical and gene regulation levels, allowed us to assess the physiological state of the cells. Surprisingly, even though microbial mats may contend with very high irradiances during the day, *Syn* OS-B′ did not appear to cope well with continuous high-light conditions. Axenic cultures of *Syn* OS-B′ grew optimally at relatively low light-fluence rates of between 75 and 130 µmol photons m^{-2} s^{-1} as shown by their blue-green color and characteristic absorption spectrum, while cells grown at higher irradiances were chlorotic and lost phycobiliproteins (Fig. 2.4). Cells grown in continuous light at an irradiance of 400 µmol photons m^{-2} s^{-1} stopped growing after 3 days and died, but it is important to note that these culture conditions do not replicate the mat conditions. Within the mat, cells are extensively packed, there may be protective pigments present, and light is strongly attenuated, particularly in the blue and red regions of the spectrum (Kuhl et al., 1997), so cells may be experiencing a very different light regime under these conditions.

Photosynthetic organisms acclimate to the damaging consequences of the absorption of excess light energy in a number of ways, including by a marked decline in light-harvesting pigments, changes in the level and composition of photosynthetic reaction centers, the development of sinks to efficiently remove electrons from the electron transport chain, the establishment of mechanisms to eliminate reactive oxygen species (ROS) that might accumulate, and the ability to repair damaged cellular components. We attempted to measure some of these parameters in the axenic isolates. At 200 µmol photons m^{-2} s^{-1}, we noted several responses that had previously been associated with acclimation of cyanobacteria to high light levels such as strong bleaching of cells, reduced levels of phycobilisomes and chlorophyll, and elevated levels of carotenoids. Quantification of the abundance of transcripts encoding the polypeptides that make up the PBS was consistent with this observation. These results suggest that at higher light irradiances there is a reduction in the absorbance cross section of the light-harvesting antenna. Other parameters tested also suggested that the cells are acclimating to high light in a number of ways. Interestingly, 77 K fluorescence emission spectra suggest that *Syn* OS-B′ accumulates very small amounts of photosystem II relative to that of photosystem I. This ratio was further decreased at higher growth irradiances, which may reflect potential photodamage following exposure to high light intensity. High light intensity also reduced levels of transcripts encoding phycobilisome components, particularly for CpcH, which is a 20.5-kDa rod linker polypeptide. There was enhanced transcript abundance

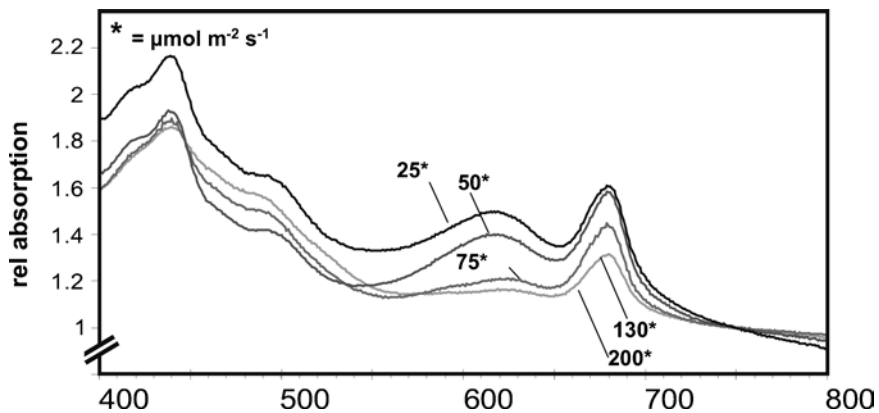


Fig. 2.4. Whole-cell absorption spectra of *Synechococcus* OS-B′ absorption at different light irradiances. Wavelength is on the X axis and absorbance on the Y axis. (Modified from Kilian et al., 2007).

of genes encoding terminal oxidases, superoxide dismutase, tocopherol cyclase, and phytoene desaturase. Genes encoding the photosystem II D1:1 and D1:2 isoforms (*psbAI* and *psbAII*/*psbAIII*, respectively) were also regulated according to the light regimen (Kilian et al., 2007).

2. Acclimation to Nutrient Limitation

Since the *Syn* OS-B′ but not the OS-A genome harbors the *phn* gene cluster, which might allow OS-B′ to grow on phosphonate (Phn) as a sole phosphorus source, we grew axenic cultures in medium lacking P_i as well as on different Phn sources (Adams et al., 2008; Gomez-Garcia et al., 2011). Cells continued to progress through 3–4 cell divisions after P_i was removed from the growth medium suggesting that the storage and metabolism of intracellular poly P may be one mechanism that enables the organism to cope with low exogenous P_i. Consistent with this possibility, we found large poly-P pools in *Syn* OS-B′ (M. R. Gomez-Garcia, A. Grossman, and D. Bhaya, unpublished). Pi limitation of *Syn* OS-B′ was found to elicit the accumulation of extracellular alkaline phosphatase activity and increased levels of transcripts encoding several putative phosphatases. The gene encoding PhoX of *Syn* OS-B′ was most highly induced (based on Q-PCR measurements) and PhoX may be responsible for most of the extracellular phosphatase activity assayed during P deprivation. In addition, transcripts encoding the high-affinity ABC-type Pst transport systems as well as the genes in the *phn* operon were induced. Many Pho regulon genes that are present in the hot spring cyanobacteria, including *phoX* and the *phn* gene cluster, are only present on a few other cyanobacterial genomes, which suggest environmental or niche-specific adaptation of P metabolism (Adams et al., 2008).

Even though *phn* transcripts of *Syn* OS-B′ accumulated rapidly in response to P starvation, *Syn* OS-B′ was unable to effectively use methyl Phn (MePhn) as a sole P-source until the cells had acclimated for approximately 3 weeks. Although this is a somewhat unexpected result, it is possible that the long and variable acclimation phase during which cells grow very slowly (Phn is supplied as a sole source of P) may be the consequence of steps that are limiting in Phn degradation. This long acclimation period has been noted in various other bacteria, such as *E. coli*, as they acclimate to different phosphonate sources (Wanner, 1994). For instance, induction of transport systems may take varying amounts of time to enable the transport of different Phn compounds, or Phn may be toxic to certain cellular processes. After ~20 days, the cells started to grow more rapidly and once the cells had acclimated, they initiated growth immediately upon transfer into fresh medium containing MePhn as a sole source of P and attained a doubling time similar to that of cells using P_i as their sole P-source. Currently we do not know if the MePhn acclimation phenomenon represents a genetically-based mechanism or whether a small subpopulation becomes responsive during the lag phase and ultimately outgrows the cells that were unable to acclimate. The *phn* gene cluster has been found in many microorganisms isolated from a variety of ecosystems, including marine ecosystems, in which Phns constitute a substantial fraction of dissolved organic P of the total P pool (Clark et al., 1998). The capacity to utilize Phns when other sources of P are limiting could confer an adaptive advantage to the *Syn* OS-B′ cells in an environment where P_i is scarce. However, it is not known whether the available P_i fluctuates on a temporal or spatial scale. P_i starvation on a daily or seasonal basis may allow *Syn* OS-B′ to acclimate to low P conditions, which includes an increased capability for utilizing phosphonates to satisfy the P demand. If fluctuations in the P_i concentration are frequent, the cells may remain in the acclimated state even when availability is elevated over a short time interval.

D. In Situ Transcriptomics to Probe Diel Cycles

In the course of our comparative analysis of the *Syn* OS-A and OS-B′ genomes, we identified a 30 kbp region that harbored genes required for nitrogen fixation (Rubio and Ludden, 2005; Bhaya et al., 2007). This was surprising since most reports of nitrogen fixation (N_2-fixation) in hot springs suggested that it occurred only at low-temperature regions of the mat, possibly catalyzed by filamentous heterocystous cyanobacteria (Stewart, 1970; Belay et al., 1984). To provide strong experimental evidence that these genes were indeed functional and that N fixation was occurring in the mats, we developed an *in situ* transcriptomics approach: *nif* gene-specific primers were used for quantitative RT-PCR (qPCR) on

RNA samples isolated from the mats at different times of the diel cycle (Steunou et al., 2006); the *nif* genes were expressed *in situ* in the mat (Steunou et al., 2006, 2008) but were only detected at night and into the early morning (i.e., the period of time when the mat was anoxic) (Fig. 2.5). During the diel cycle collections, light and oxygen levels were also measured. Nitrogenase activity (monitored by acetylene reduction assay) and nitrogenase subunits (monitored by Western blot analyses) in mat samples were also detected in the evening and early morning. This suggests that nitrogenase activity is restricted to certain parts of the diel cycle. Since nitrogenase activity is irreversibly inactivated by oxygen and N fixation is energetically expensive (requiring at a minimum 16 ATP molecules per N fixed), it is important to determine how nitrogenase activity is regulated and how energy is made available for this process. This required us to accurately monitor other genes of interest, e.g., genes involved in photosynthesis, fermentation and respiration. Thus, our approach was to monitor mat metabolism and gene regulation, and to correlate these data with environmental parameters such as light, pH and nutrient availability to build a model of how various energetic processes vary over a diel cycle (Fig. 2.5).

A conceptual model showing the factors that influence nitrogenase activity over the diel cycle in hot spring mats has now been developed (Steunou et al., 2008). During the day, the upper few millimetres of the mat are supersaturated with O_2 because of cyanobacterial oxygenic photosynthesis. Under these conditions, the *nif* genes are not expressed. As irradiance falls towards the end of the day, the O_2 concentration in the mat also drops because of (i) a decline in photosynthetic O_2 evolution and (ii) a sustained or increased respiratory consumption of O_2 by cyanobacteria and other microbes in the community. At the same time, both *nif* and specific fermentation transcripts increase, corresponding polypeptides are synthesized and assembled into active complexes, and N_2 fixation can be initiated. By the time the level of nitrogenase becomes maximal and its activity is fully established, photosynthetic energy production has decreased substantially due to the absence of light. Oxygen is largely depleted in the upper 0.1–0.2 mm of the mat due to respiration and re-oxidation of reduced compounds. Thus, the only source of energy for cyanobacterial

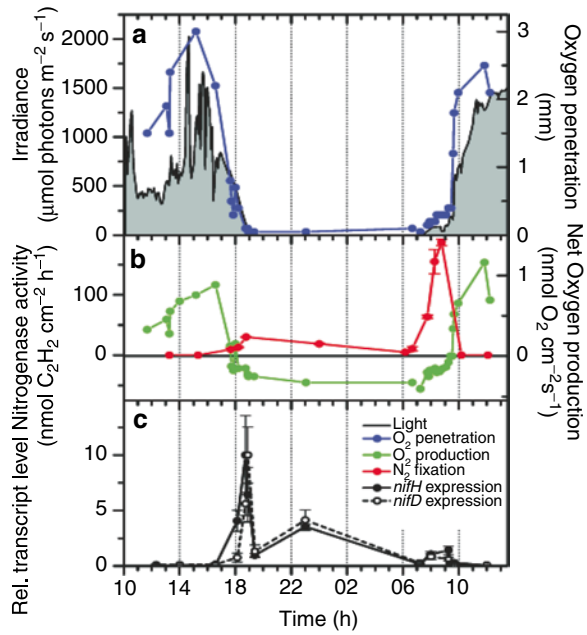


Fig. 2.5. In situ nitrogenase activity, levels of NifH subunit and transcripts encoding NifH and NifD, and oxygen penetration and net production over the diel cycle in the microbial mat of Mushroom Spring in September 2005. (**a**) Incident down welling irradiance (micromol photons m^{-2} s^{-1}) and O_2 penetration (mm) in the hot spring mat over the diel cycle. (**b**) In situ nitrogenase activity (nmol ethylene per cm^2 per h) and net O_2 production (nmol O_2 cm^{-2} s^{-1}) over the diel cycle. (**c**) Relative abundance of transcripts encoding the nitrogenase subunits NifH and NifD over the diel cycle. Curves and colors are defined in the inset of panel C. Nitrogenase activity and gene expression data points represent means ± sd (N = 3). (Modified from Steunou et al., 2008).

N_2-fixation in the anoxic part of the mat is derived from fermentation of organic carbon accumulated during the preceding day. A similar scenario has been proposed for hypersaline mats (Bebout et al., 1993).

In the morning, as light levels increase, nitrogenase activity increases in parallel with photosynthetic activity. The increased nitrogenase activity is not accompanied by increased nitrogenase transcript or protein levels, but reflects elevated production of ATP and reductant. The mat remains largely anoxic with net O_2 consumption during the early hours of the day period, until increasing irradiance drives the rate of photosynthetic O_2 evolution above the rate of respiratory O_2 consumption. As O_2 begins to accumulate in the mat, the nitrogenase activity is strongly inhibited. Since these processes are interlinked, other

factors such as day length, light intensity and accumulation of stored carbon compounds such as polysaccharides (Steunou et al., 2008) may affect the dynamics of the metabolism over a diel cycle.

Transcriptomics together with integrating eco-physiological methods to quantify the mat microenvironment and nitrogenase activity *in situ* give new insights into the complex dynamics of N_2 fixation in hot spring cyanobacterial mats. N_2 fixation in thermophilic *Synechococcus* sp. is closely linked to their energy metabolism and photosynthesis, which also shows pronounced shifts during a diel cycle. The next challenge with this system will be to develop an understanding of the functional interactions with other microbes in the mat community, as well as the variants in the populations of different cyanobacteria in the mat.

The integrative approach presented here attempts to tie together three methodologies: eco-physiology, *in situ* microsensor measurements of activity, and transcriptomics. This approach requires, as a pre-requisite, a detailed knowledge of the physiology of the microbial mat system and the genomic content of some of the dominant players in the microbial mat. This approach sheds light on the metabolic dynamics of the cyanobacterial populations, which are well-adapted to the diurnal fluctuations in light and oxygen.

E. Recent Acquisition or Loss of Nutrient Utilization Pathways

As the *phn* cluster and the cyanophycin synthesis/metabolism pathway are only found in one of the two *Synechococcus* isolates (OS-B′) we further explored the possibility that nutrient utilization functionality can be gained and lost in related lineages. In this context, we noted significant differences between the *Syn* OS-A and OS-B′ isolates for genes required for the putative uptake and utilization of urea. Some cyanobacteria can utilize urea as a source of nitrogen, which requires the enzyme urease and a dedicated transport system (Luque et al., 1994; Collier et al., 1999). *Syn* OS-A has one genomic region encoding urease (*ureA1B1C*) and accessory factors (*ureEFG1D1*) (Cluster 1 urease) (Fig. 2.6). The genes of Cluster 1 urease are flanked by transposons. This cluster is lacking in *Syn* OS-B′ genome but the synteny between *Syn* OS-A and OS-B′ in the region flanking Cluster 1 is maintained, suggestive of a relatively recent gain of genes by *Syn* OS-A. Alternatively, it might represent a loss of genes from *Syn* OS-B′ if the shared ancestor of *Syn* OS-A and OS-B′ contained these genes. Without further analysis it is not possible to distinguish between these possibilities, but the presence of flanking transposons suggests a gain of function in OS-A. *Syn* OS-B′ contains a second set of

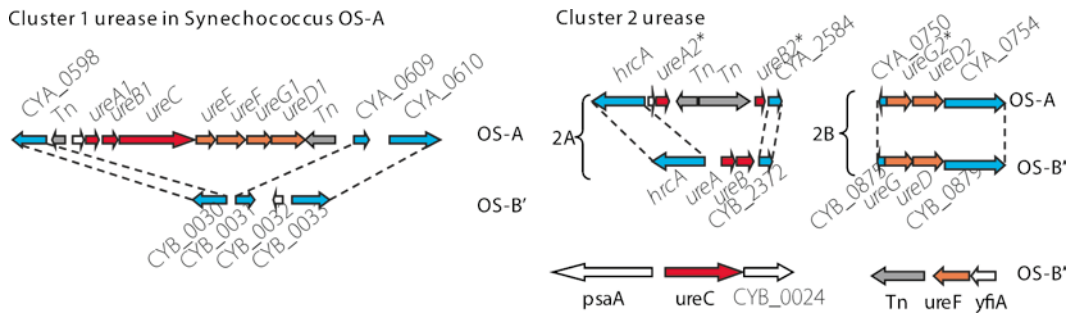


Fig. 2.6. Urease genes in *Syn* OS-A and OS-B′. *Left Panel*: *top row* represents the Cluster 1 urease and flanking genes of *Syn* OS-A; the bottom row shows homologs of the flanking genes in *Syn* OS-B′. *Right panel* shows: locations of the Cluster 2 urease genes in *Syn* OS-A and OS-B″. In *Syn* OS-A, *ureB2* is inactivated by a transposon; *ureA* and *ureG2* contain frameshift mutations (inactive genes are designated by * symbols). Genes encoding urease (*ureA*, *–B* and *–C*) are shown in *red*; accessory factor genes are shown in *orange*; putative transposons are shown in *gray*; flanking gene homologs in *blue* and other genes in *white*. Syntenic regions are indicated by *dashed lines*.

urease genes, which are scattered at five different regions of the genome, and remnants or transposon-mediated interruptions of these genes are found in *Syn* OS-A. In *Syn* OS-B′ these Cluster 2 urease genes are intact. These obvious differences in gene content underscore the possibility that both gene gain and loss are ongoing events in the *Synechococcus* populations. We envisage a possible scenario, in which there was a relatively recent acquisition of Cluster 1 urease by *Syn* OS-A as suggested by the transposons flanking this region. Subsequently there may have been a progressive loss of functionality of the original urease cluster (i.e., the Cluster 2 urease), which is still maintained by and likely to be functional in *Syn* OS-B′. These observations raise obvious questions to be addressed in the future:

(a) If *Syn* OS-A acquired this cluster by "recent" lateral gene transfer, can we date this process and possibly identify the donor organism?
(b) Could further metagenomic analyses of the populations reveal other variants, with different gene arrangements, suggestive of fluid functionalities that are evolving in this environment?
(c) The more recently acquired Cluster 1 may provide selective advantage to *Syn* OS-A, which the Cluster 2 urease did not provide. Can the use of axenic strains of these organisms allow us to test some of these hypotheses?
(d) Are there other examples of such loss/acquisition of nutrient utilization capacity? If so, is this a general theme of phototrophs that are evolving in fluctuating environments?

F. Genomic Rearrangements/Fusions in the Context of Photosynthesis

The high-light-inducible proteins (Hlips) of cyanobacteria are members of an extended CAB/ELIP/Hlip superfamily of distantly related polypeptides, which have between one and four transmembrane helices (TMHs) (Green, 1995; Jansson, 1999). The best-studied members of this superfamily are the membrane-integral light-harvesting chlorophyll *a/b* binding proteins (LHCs or also known as CABs), which are abundant in the chloroplast (see Chapter 11, this volume). They bind chlorophylls and carotenoids, associate with both photosystem I and II, and are regulated by various environmental factors including light levels and nutrients. The LHC proteins are well-conserved polypeptides with three TMHs and they may have evolved by gene duplication of an ancestral gene encoding a single TMH (Kuhlbrandt et al., 1994).

The Hlips are small membrane proteins consisting of a single TMH, similar to those in the LHC proteins that are induced by various stress conditions including illumination with UV-A and high light intensity, and are considered ancestral to the LHC proteins (Green and Pichersky, 1994; Dolganov et al., 1995; Funk and Vermaas, 1999; He et al., 2001). A *Synechocystis* sp. PCC 6803 mutant in which all four *hli* genes were inactivated died under high light conditions, suggesting that Hlips are crucial for acclimation to high light (He et al., 2001). It has also been suggested that certain members of this family may regulate an early step of tetrapyrrole biosynthesis and act as regulators that function based on chlorophyll availability, such that they activate chlorophyll biosynthesis steps when their pigment binding sites are unoccupied (Xu et al., 2002). One of the Hlips is associated with the periphery of photosystem II following exposure to high light intensity, suggesting that it may be required for photosystem II assembly or repair (Promnares et al., 2006; Yao et al., 2007).

In *Syn* OS-A and *Syn* OS-B′ there are five *hli* genes comprising a small gene family (Kilian et al., 2007); *hli4* and *hli5* are tandemly arranged on the genome in both organisms. Such an arrangement is often taken as evidence of a recent duplication event (Koonin, 2005). The high-light acclimated ecotype of marine *Prochlorococcus* MED4 has 22 *hli* gene copies including four tandemly arranged, almost identical ones (*hli6–hli10*); this entire region is also duplicated (*hli16–hli19*) elsewhere in the genome (Bhaya et al., 2002). Transcripts from all of these genes are significantly up-regulated at high light intensity, although other members of the *hli* gene family are constitutively expressed (Steglich et al., 2006). In *Synechococcus* OS-B′ Hlips1–4 are between 50 and 73 amino acids in length, as is the case in most other cyanobacteria (Dolganov et al., 1995; Kilian et al., 2007). However, Hlip5 is considerably larger (102 amino acids) than the other Hlips and appears to represent a novel fusion

event between another small membrane protein of unknown function (Coh1) at the N-terminus (called TMH1) and the Hlip TMH domain (called TMH2) (Fig. 2.7).

The small Coh1 (cyanobacterial one helix) protein is highly conserved and its gene is found in most other cyanobacterial genomes with the exception of the marine lineage cyanobacteria *Synechococcus* sp. and *Prochlorococcus* sp. It is annotated as "conserved hypothetical". It has been suggested that new multi-domain proteins are formed during the course of evolution by the process of gene duplication followed by gene fusion. Such a fusion may also result in tight co-regulation of these genes and serve to expand the functional role of fusion proteins (Yanai et al., 2000; Bashton and Chothia, 2007). If *hli4* and *hli5* represent a recent duplication event, then the gene fusion between the *coh1* homolog (represented by TMH1 in Hlip5) and a *hli*-like gene (represented by TMH2 in Hlip5) may also have occurred recently. Hlip5, which has two TMH domains, may have increased stability in the membrane and/or additional partner proteins may alter its ability to bind chlorophyll or change its association with complexes of the photosynthetic apparatus (Kuhlbrandt et al., 1994; Funk and Vermaas, 1999; Standfuss et al., 2005; Kilian et al., 2007). Since *Syn* OS-B′ is unable to survive when grown at photon fluence levels of 400 μmol photons m^{-2} s^{-1} (Kilian et al., 2007), the role of Hlips, and in particular the novel two-TMH Hlip5, is of interest. Membrane topology algorithms predict that the loop between TMH1 and TMH2 is more likely to be located in the lumen; thus the conserved aspartate/glutamate residues that are positioned between TMH1 and TMH2 would be present in the lumen and could serve a regulatory function. Molecular analysis of non-photochemical quenching in *A. thaliana* suggests that PsbS senses hyper-excitation of photosynthetic electron transport via the development of a large ΔpH across the thylakoid membranes, which, in turn, leads to protonation of conserved, lumenal glutamate residues in PsbS (Li et al., 2004). PsbS may bind carotenoids that may quench chlorophyll excitation through the formation of a carotenoid radical (Holt et al., 2005; Standfuss et al., 2005).

Hlips may represent one of the single TMH progenitors of the CAB/ELIP/Hlip superfamily (Hoffman et al., 1987; Green and Pichersky, 1994; Kuhlbrandt et al., 1994). An Hlip sequence-like may have fused with another TMH to create a two-TMH structure. A four-helix protein (e.g., PsbS) could represent an event, in which a gene for a two-TMH protein experienced a tandem duplication, followed by fusion of the duplicated genes. The evolution of a three-TMH LHC from a PsbS-type, four-TMH protein would involve loss of the

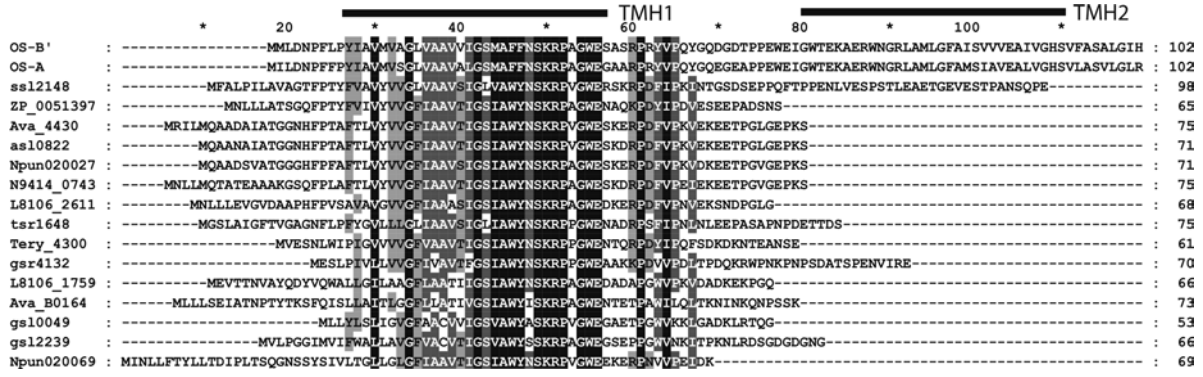


Fig. 2.7. ClustalW alignment of Hlips from *Synechococcus* OS-B′ and OS-A with Coh1 found in other cyanobacteria. These are *Synechocystis* PCC 6803 (ssl2148), *Crocosphaera watsoni* (ZP_00051397), *Nodularia spumigena* (N9414_0743), *Trichodesmium erythraeum* (Tery_4300), *Nostoc* sp. PCC 7120 (asl0822), *Thermosynechococcus elongatus* (tsr1648), *Nostoc punctiforme* (Npun02002710 and Npun02006958), *Anabaena variabilis* (Ava_4430 and Ava_B0164), and *Lyngbya* sp. PCC8106 (L8106_2611 and L8106_1759) and *Gloeobacter violaceus* PCC 7421(gsr4132, gsl0049, and gsl2239). *Black*, *dark gray* and *light gray boxes* indicate completely conserved (100%), highly conserved (80%) or moderately conserved (60%) residues, respectively. The positions of TMH1 and TMH2 (of Hlip5) are shown as *black bars* above (Modified from Kilian et al., 2008).

last TMH (Green et al., 1991; Green and Pichersky, 1994; Montane and Kloppstech, 2000). The TMH1 and TMH3 of the LHCs are similar to each other and are also similar to TMHs of the distantly related CAB/ELIP/Hlip polypeptides, but the TMH2 of the LHCs is not well conserved (Heddad and Adamska, 2000; Montane and Kloppstech, 2000). The discovery of Hlip5, which represents a novel protein containing an Hlip domain and a domain representing a conserved protein in cyanobacteria, suggests that scrutiny of the extensive genomic databases might reveal evidence of other novel fusion events. The extensive and ongoing rearrangements experienced by the genomes of *Syn* OS-A and OS-B′ (Bhaya et al., 2007) could accelerate the rate of gene fusion events. The large majority of these events would be lethal or lead to a non-competitive phenotype, but if a gene fusion event conferred a selective advantage it may be selected for during evolution.

IV. Metagenomic Analysis

The term metagenomics (also known as "community or environmental genomics") was first used in 1998 (Handelsman et al., 1998), but its impact on research areas such as ecology, evolution and bacterial diversity has really begun to grow exponentially over the last 4 years. Metagenomics can be broadly defined as a field, in which genomic approaches are taken to analyze microbial communities. Because this approach requires neither prior knowledge of the microbial members of the community nor purification of particular species, it has the potential to provide an unbiased view of the functionalities in a community (Streit and Schmitz, 2004; Konstantinidis et al., 2006; Handelsman, 2008; Vieites et al., 2009; Wilmes et al., 2009). Thus, metagenomic analyses are particularly attractive for environments where a large fraction of microbes remain uncultured or are not amenable to standard microbiological analyses (Woyke et al., 2006; Eisen, 2007; Kowalchuk et al., 2007; Raes and Bork, 2008). Sequencing projects have been initiated for a number of different purposes including (i) identification of dominant species (Yooseph et al., 2007), (ii) characterization of the community structure of complex consortia ('phylogenomics') (Strous et al., 2006; Woyke et al., 2006), (iii) identification of novel genes and operons in the population (Beja et al., 2001; Entcheva et al., 2001), and (iv) 'bioprospecting' or searching for desirable gene functions, such as cellulases or chitinases (Lorenz et al., 2002; Riesenfeld et al., 2004; Allen et al., 2007; Gabor et al., 2007; Ferrer et al., 2009). The focus of the first Sargasso Sea metagenome sequence analysis was an attempt to recreate complete genomes of surface-water organisms, notably *Prochlorococcus*. Subsequent metagenomic, genomic and comparative analyses of ocean microbes has led to insights on community genomics in stratified environments in the oceans (Coleman et al., 2006). Metagenome analysis of acid mine-drainage biofilms led to the complete or partial reconstruction of five genomes because of the domination of a small number of relatively homogenous species (*Leptospirillum* and *Ferroplasma*) (Tyson et al., 2004; Tyson and Banfield, 2005). Several groups have focused on the analysis of specific marker genes, such as the *nif* or 16S RNA genes, in metagenomic samples (Johnston et al., 2005; Sogin et al., 2006).

Methods are being developed to use comparative metagenomics and metaproteomic analysis to probe the diversity of bacterial populations and in various environments without relying on the assembly of genomes (Ram et al., 2005; Tringe et al., 2005; DeLong et al., 2006; Ward 2006b; Kowalchuk et al., 2007; Raes et al., 2007; von Mering et al., 2007). Tringe et al. (2005) applied an automated annotation process to establish "environmental gene tags" based on predicted genes on DNA fragments from the metagenome. To derive a measure of the functional profile of different microbial communities, they developed a method to "bin" similar sequences. This analysis allows one to compare the abundance of binned sequences belonging to specific metabolic pathways across communities. This approach rests on the assumption that a few defining habitats may determine genomic profiles in the microbes, since the habitats provide the context for the physiology and the pathways/processes critical for survival of the organism, and place less importance on relationships of specific genes to specific organisms. For instance, photosynthesis genes were found to be highly represented in the Sargasso Sea metagenomic study, but were not nearly as prevalent in other environments.

These results demonstrate that an aggregative approach to understand major functional metabolic pathways may be used without the need to link these functions to particular organisms. Likewise, DeLong et al. (2006) identified protein categories (based on KEGG and COG databases) and subsequently used cluster analysis to identify specific genes that were differentially distributed in the water column.

There are now well over a hundred metagenome projects underway and this number may rise further as sequencing costs fall (Edwards et al., 2006) and high throughput platforms become accessible (Havre et al., 2005; Huson et al., 2007). However, with the exception of studies that target low-complexity environments (Tyson et al., 2004) or where sequence coverage is very high, it is a major technical challenge to assemble complete or even partial microbial genomes from metagenomic data since the depth of sequencing required is usually far greater than most projects permit (Chen and Pachter, 2005; Rusch et al., 2007; Raes and Bork, 2008).

A. Microbial Mat Metagenomics

We have recently obtained an environmental genomic dataset from random shotgun sequencing of total DNA collected from the top green layer of microbial mats at two different temperature sites of Octopus Spring (the spring from which *Syn* OS-A and *Syn* OS-B′ were originally isolated) and from Mushroom Spring, a nearby spring with similar physicochemical characteristics (Fig. 2.8) (Bhaya et al., 2007). Recombinant libraries containing small inserts (2–6 kbp) or large inserts (10–12 kbp) from total DNA isolated from the top green layer were created using standard sequencing vectors. Sequencing was carried out on ~200,000 clones and "paired end" sequences were derived; this represented ~200 Mbp of sequence data from the four collection sites. By way of comparison, the read coverage from other metagenome sequencing projects range from 76.2 Mbp from acid-mine biofilms (Tyson et al., 2004) to 1.6 Gbp (surface water, Sargasso Sea) in the first global ocean sampling trip in 2004 (Venter et al., 2004). These datasets are a significant community asset and an exploitable addition to the toolbox available for this well-studied microbial ecosystem. The microbial-mat metagenomic data set is being used to better understand population structure in the mats as well as to gain an understanding of diversity within the *Synechococcus* populations at the different temperatures (Bhaya et al., 2007).

We were able to create large environmental genome scaffolds such that a "virtual" or "composite" genome spanned almost the entire reference genomes of *Syn* OS-A and OS-B′. This suggests that the organisms from which the reference genomes were derived are abundant members of the community (Bhaya et al., 2007). This approach, in which metagenome sequences can be "pasted" back onto an "anchor" or reference genome, is somewhat different from many other ongoing metagenomic projects where no such reference genomes are available. In fact, lack of these

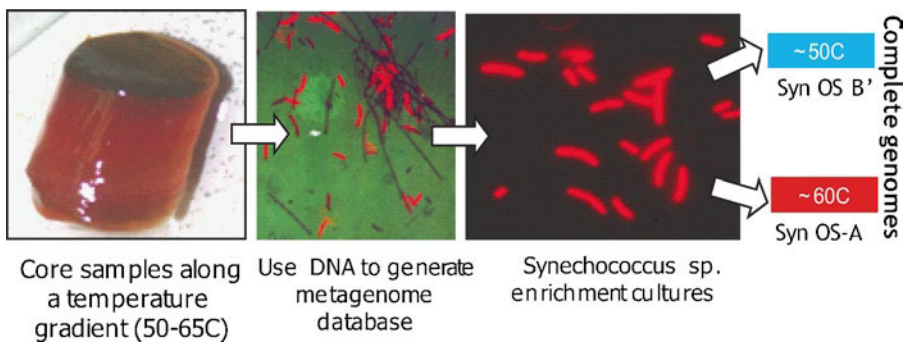


Fig. 2.8. A simplified schematic showing the major steps going from initial collection of a core mat sample (which can be harvested from different regions) (*left panel*), via selection of a part of the core sample, where total DNA can be extracted from the top green layer of the core. This layer, which contains a diversity of microbes, yields the metagenomic database of sequences. Specific *Synechococcus* isolates can be obtained from this community by dilution enrichment (*middle panels*) and the entire genome of specific isolates can be sequenced (*right panel*).

reference genomes makes it difficult or impossible to create any large contiguous sequences or "contigs" and even more difficult to accurately identify the species of microorganism from which it is derived. The metagenome dataset provided further evidence that there is a rich population of *Synechococcus* spp. in the mats.

Interestingly, a significant fraction (~40%) of the metagenomic sequence reads were quite similar to the *Syn* OS-A and OS-B′ sequences, but could not be confidently assembled onto the anchor genome because they failed to meet the stringent criteria that we set for assembling the sequences. These clones, which we dubbed "illegal clones", might represent regions of recombination, transversion or gene gain/loss ('indels'). As significant gene rearrangements between populations of *Synechococcus* sp. along the temperature gradient (caused by transposon activity, recombination events etc.) have occurred and as there is significant diversity at the gene level, it is clear that this model system can be used for the analysis of genetic variation and of its origins and causes.

B. Functional Diversity in the Metagenome

The variations in nutrient acquisition and utilization capabilities suggested by differences in genome content between the two sequenced *Synechococcus* isolates prompted us to carry out a detailed examination of *Synechococcus*-like sequences in the metagenome dataset for further examples of functionally specialized populations. We investigated a category of "illegal clones", in which one end of the clone had high sequence identity to a specific region on the *Synechococcus* OS-B′ genome, while the paired end sequence did not match any sequences in the *Syn* OS-B′ genome. These clones could represent a *Synechococcus* population closely related to *Syn* OS-B′ that had acquired additional sequences that are absent in the 'anchor' genome. Indeed, we found an example in which a sequenced clone matched the *Syn* OS-B′ genome at both ends but contained an extra 5.5 kbp region relative to the *Syn* OS-B′ anchor genome (Fig. 2.9). This clone contained seven additional genes, and is flanked by genes that are almost identical to genes in the *Syn* OS-B′ genome (CYB_0562 (99.72% NAID) on the left and CYB_0565 (99.76% NAID) on the right). Of the genes present in the 5.5 kbp region, two exhibited significant identity to the *feoA* and *feoB* genes of the unicellular thermophilic cyanobacterium, *Thermosynechococcus elongatus.*

The *feoA* and *feoB* genes encode proteins required for ferrous ion transport in several bacteria (Andrews et al., 2003). Iron in the hot-spring environment exists in both ferrous and ferric forms, with the ferrous form predominating under conditions of low oxygen. The *feoA*/*feoB*-like sequences are absent on the *Syn* OS-A and OS-B′ anchor genomes, although they contain several genes required for ferric ion uptake and assimilation. The presence of *feo* genes in an organism that appears to be closely related to *Syn* OS-B′ is interesting because it suggests the presence of functionally specialized populations capable of using Fe^{2+}. Based on Q-PCR results, we hypothesize that the ferrous transport system may accumulate as the mat becomes anoxic during the night, allowing some *Syn* OS-B′-like organisms to scavenge the reduced ferrous ions that increase

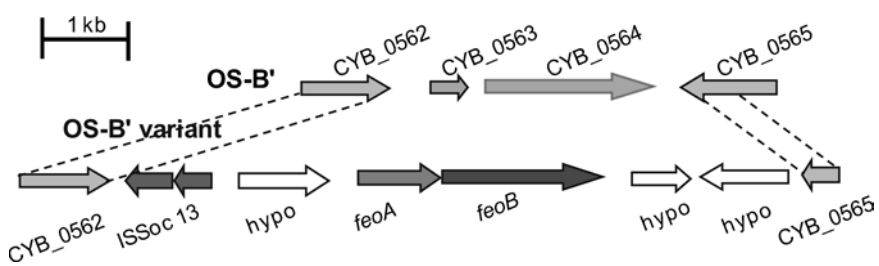


Fig. 2.9. Feo genes containing variant of *Syn* OS-B′. Region containing CYB_0562, CYB_0563, CYB_0564 and CYB_0565 from *Syn* OS-B′. Below is the variant clone containing the *feo* genes plus five additional open reading frame, including OrfB and OrfA of ISSoc13 transposase, which appears to be *Syn* OS-B′-specific. The similarity of this variant to *Syn* OS-B′ beyond this region is not known.

during anoxic mat conditions. This is an example of the advantages of using metagenomic information to identify *Synechococcus* variants in the mat and raises questions relating to functional genomics and technology development in the context of the environment.

C. Lateral Gene Transfer, Transposons and Viruses in the Creation of Variant Populations in Microbial Mats

The dense microbial mats, which harbor a diversity of organisms, are likely to be fertile territory for the study of how genes are transferred between organisms. Although this aspect has not yet been studied in detail among the phototrophs of the mat, it has the potential of revealing the importance of gene transfer processes in communities. Transposons are an important part of the genetic repertoire of the cyanobacteria and in some cases they may be quite active (Bhaya et al., 2007; Zhou et al., 2008). A consequence of this activity would be an accelerated pace of recombination events, which may be responsible to a larger or lesser extent for the large number of "variants" that we see in the populations of cyanobacteria. Some transpositions by IS elements may have occurred very recently (i.e., after/during the process in which the isolates were brought into culture from the original mat environment) and lateral gene transfer may be occurring in the mat cyanobacterial populations (Nelson et al., 2011). This suggests that these genomes may be quite fluid and that genetic change in natural populations is an ongoing occurrence. Lateral gene transfer has been extensively documented in numerous bacterial lineages and is considered to play a significant role in genome evolution (Boucher et al., 2003; Gogarten and Townsend, 2005; Lerat et al., 2005). Hot spring mats have been examined for the presence of viruses using both a viral metagenomic approach (Schoenfeld et al., 2008) as well as by other methods (Rice et al., 2001; Ortmann et al., 2006; Snyder et al., 2007) but cyanophages have not yet been carefully studied. A recent study concluded that the newly discovered "viral immunity" systems mediated by the CRISPR/Cas systems are active in the mat *Synechococcus* populations (Heidelberg et al., 2009). This would imply that a better understanding of virology in the context of cyanobacterial populations in the microbial mat is warranted. Marine cyanophages are capable of carrying specific cyanobacterial (or host genes, such as *psbA*) within their genomes and these genes, when functional, may confer some advantage to the viruses. Terrestrial cyanophages may also have similar strategies but this has not yet been shown for cyanobacteria in the hot-spring microbial mats.

V. Future Directions

I will briefly mention three directions that are likely to be directly relevant to many of the issues and concepts discussed here in the future.

A. From a 'Wild' Cyanobacterium to a Model Organism

One of the unique and powerful aspects of our approach was the ability to combine microbial ecology with sophisticated molecular tools to probe function and to correlate it with important and variable environmental parameters. Furthermore, the ability to grow axenic strains of environmentally relevant and dominant thermophilic cyanobacteria adds another important tool to our arsenal. However, one of the crucial requirements for any microorganism to be used as a "model organism" is for it to be effectively genetically manipulated. Some cyanobacteria are naturally transformable and are able to take up foreign DNA and integrate it into the genome by recombination. So it has been relatively easy to develop systems for gene inactivation and the creation of targeted mutants (Koksharova and Wolk, 2002). We have recently demonstrated that we can transform *Syn* OS-B′ and that it appears to be stably transformed (Bhaya et al., unpublished). This opens the door for the next generation of experiments to be performed and to develop a powerful model platform that extends from the *in situ* experiments and environmental measurements in the mat to the testing and refinement of hypotheses under controlled laboratory conditions. This development combined with microarray analysis makes the system much more powerful for examining questions about how environmental parameters affect gene expression.

B. Single Cells and Their Applications

Single-cell analysis is a new technology that has provided new insights into the concept that there is marked phenotypic heterogeneity even in genetically homogenous microbial cultures (Avery, 2006). Single-cell analysis allows one to observe cells as individuals, and to manipulate them in many ways (Breslauer et al., 2006; El-Ali et al., 2006). Microfluidic platforms, which allow for the capture of single cells, can be coupled with numerous downstream analytical manipulations including capillary electrophoresis to examine protein content of a single cell as well as counting individual fluorescent molecules (Huang et al., 2007; Kim et al., 2007). It can also be combined with techniques that include whole genome amplification (Lasken, 2007). In the context of the environment the ability to capture and sequence the entire genome of a single cell opens up untold possibilities (Ottesen et al., 2006; Ishoey et al., 2008; Woyke et al., 2009), since most bacteria cannot be cultivated and the genetic diversity of the microbial world is still to be detailed (Warnecke and Hugenholtz, 2007). This type of analysis provides a large "bank" of sequences relevant to a particular environment and therefore is likely to improve the ability to build larger scaffolds from metagenomic sequence data.

C. Community Proteomics

Metagenomics and transcriptomics provide a strong basis for assessing the inherent encoded capabilities of a community. However, information about the "business end" of a cell, i.e., the proteins, is perhaps the most technically challenging (Ram et al., 2005). Proteins are made, modified and degraded over the diel cycle or during the life cycle of a cell and new technology, which uses mass spectrometry and 2-D gel electrophoresis, may help identify these proteins and monitor changes and modifications. In the context of the hot spring mats, responses of phototrophic microbes to dynamic environmental conditions, which include temperature, light, lack of oxygen and nutrient levels, would be particularly important to examine. A challenge offered by such an undertaking is to generate concepts and principles from large datasets and to reveal physiological and metabolic changes underlying adaptation to extreme and fluctuating environmental conditions. Modeling and computation analyses will be key components of such an endeavor and are likely to provide essential insights.

Acknowledgments

I would like to acknowledge generous research support from the Frontiers in Integrative Biology Program of the National Science Foundation (Grant EF-0328698), which allowed us to develop various aspects of this multifaceted project. I would particularly like to thank members of the FIBR group (Fred Cohan, Arthur Grossman, John Heidelberg, Michael Kuhl, and Dave Ward) for their individual intellectual contributions to the success of the project. This chapter represents a synthesis of several joint publications, although the emphasis on specific topics is my perspective. Many of the post-docs and students at the Carnegie Institution and elsewhere put in heroic efforts at various stages. Funding and partial support from the Carnegie Institution and from other NSF grants (including Grant 0758374 to Doron Levy, University of Maryland and D.B.) is also gratefully acknowledged. The extended cooperation of Christie Hendrix and members of the Parks Department of Yellowstone National Park is also greatly appreciated.

References

Achtman M and Wagner M (2008) Microbial diversity and the genetic nature of microbial species. Nat Rev Microbiol 6: 431–440

Adams DG (2000) Symbiotic interactions. In: Whitton BA and Potts M (eds) The Ecology of Cyanobacteria, pp 37–59. Kluwer Academic Publishers, Norwell

Adams DG and Duggan PS (2008) Cyanobacteria-bryophyte symbioses. J Exp Bot 59: 1047–1058

Adams MM, Gomez-Garcia MR, Grossman AR and Bhaya D (2008) Phosphorus deprivation responses and phosphonate utilization in a thermophilic *Synechococcus* sp. from microbial mats. J Bacteriol 190: 8171–8184

Ahlgren NA and Rocap G (2006) Culture isolation and culture-independent clone libraries reveal new marine *Synechococcus* ecotypes with distinctive light and N physiologies. Appl Environ Microbiol 72: 7193–7204

Allen RT and Day AL (1935) Hot Springs of the Yellowstone National Park. Carnegie Institute of Washington

Allen EE, Tyson GW, Whitaker RJ, Detter JC, Richardson PM and Banfield JF (2007) Genome dynamics in a natural archaeal population. Proc Natl Acad Sci USA 104: 1883–1888

Allewalt JP, Bateson MM, Revsbech NP, Slack K and Ward DM (2006) Effect of temperature and light on growth of and photosynthesis by *Synechococcus* isolates typical of those predominating in the octopus spring microbial mat community of Yellowstone National Park. Appl Environ Microbiol 72: 544–550

Andrews SC, Robinson AK and Rodriguez-Quinones F (2003) Bacterial iron homeostasis. FEMS Microbiol Rev 27: 215–237

Arakawa S, Sato T, Sato R, Zhang J, Gamo T, Tsunogai U et al (2006) Molecular phylogenetic and chemical analyses of the microbial mats in deep-sea cold seep sediments at the northeastern Japan Sea. Extremophiles 10: 311–319

Archibald JM (2009) The puzzle of plastid evolution. Curr Biol 19: R81–88

Avery SV (2006) Microbial cell individuality and the underlying sources of heterogeneity. Nat Rev Microbiol 4: 577–587

Bashton M and Chothia C (2007) The generation of new protein functions by the combination of domains. Structure 15: 85–99

Bateson MM and Ward DM (1988) Photoexcretion and consumption of glycolate in a hot spring cyanobacterial mat. Appl Environ Microbiol 54: 1738–1743

Bebout BM and Garcia-Pichel F (1995) UV B-induced vertical migrations of cyanobacteria in a microbial mat. Appl Environ Microbiol 61: 4215–4222

Bebout BM, Fitzpatrick MW and Paerl HW (1993) Identification of the sources of energy for nitrogen fixation and physiological characterization of nitrogen-fixing members of a marine microbial mat community. Appl Environ Microbiol 59: 1495–1503

Beja O, Spudich EN, Spudich JL, Leclerc M and DeLong EF (2001) Proteorhodopsin phototrophy in the ocean. Nature 411: 786–789

Belay N, Sparling R and Daniels L (1984) Dinitrogen fixation by a thermophilic methanogenic bacterium. Nature 312: 286–288

Bhaya D (2004) Light matters: phototaxis and signal transduction in unicellular cyanobacteria. Mol Microbiol 53: 745–754

Bhaya D, Dufresne A, Vaulot D and Grossman A (2002) Analysis of the *hli* gene family in marine and freshwater cyanobacteria. FEMS Microbiol Lett 215: 209–219

Bhaya D, Grossman AR, Steunou AS, Khuri N, Cohan FM, Hamamura N et al (2007) Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. ISME 1: 703–713

Botero LM, Brown KB, Brumefield S, Burr M, Castenholz RW, Young M and McDermott TR (2004) Thermobaculum terrenum gen. nov., sp. nov.: a non-phototrophic gram-positive thermophile representing an environmental clone group related to the Chloroflexi (green non-sulfur bacteria) and Thermomicrobia. Arch Microbiol 181: 269–277

Boucher Y, Douady CJ, Papke RT, Walsh DA, Boudreau ME, Nesbo CL et al (2003) Lateral gene transfer and the origins of prokaryotic groups. Annu Rev Genet 37: 283–328

Breslauer DN, Lee PJ and Lee LP (2006) Microfluidics-based systems biology. Mol Biosyst 2: 97–112

Brock TD (1978) Thermophilic microorganisms and life at high temperatures. Berlin: Springer Verlag

Brock TD (1997) The value of basic research: discovery of Thermus aquaticus and other extreme thermophiles. Genetics 146: 1207–1210

Cardenas E and Tiedje JM (2008) New tools for discovering and characterizing microbial diversity. Curr Opin Biotechnol 19: 544–549

Castenholz RW and Garcia-Pichel F (2000) Cyanobacterial responses to UV-radiation. In: Whitton BA and Potts M (eds) The Ecology of Cyanobacteria, pp 591–611. Kluwer Academic Publishers, Norwell

Chen K and Pachter L (2005) Bioinformatics for whole-genome shotgun sequencing of microbial communities. PLoS Comput Biol 1: 106–112

Christner BC, Mosley-Thompson E, Thompson LG, and Reeve JN (2003) Bacterial recovery from ancient glacial ice. Environ Microbiol 5: 433–436

Clark LL, Ingall ED and Benner R (1998) Marine phosphorus is selectively mineralized. Nature 393: 426

Cohen Y, Padan E and Shilo M (1975) Facultative anoxygenic photosynthesis in the cyanobacterium *Oscillatoria limnetica*. J Bacteriol 123: 855–861

Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, DeLong EF and Chisholm SW (2006) Genomic islands and the ecology and evolution of *Prochlorococcus*. Science 311: 1768–1770

Collier JL, Brahamsha B and Palenik B (1999) The marine cyanobacterium *Synechococcus* sp. WH7805 requires urease (urea amidohydrolase, EC 3.5.1.5) to utilize urea as a nitrogen source: molecular-genetic and biochemical analysis of the enzyme. Microbiology 145: 447–459

Decker KLM, Potter CS, Bebout BM, Marais DJD, Carpenter S, Discipulo M et al (2005) Mathematical simulation of the diel O, S, and C biogeochemistry of a hypersaline microbial mat. FEMS Microbiol Ecol 52: 377–395

DeLong EF and Karl DM (2005) Genomic perspectives in microbial oceanography. Nature 437: 336–342

DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard N.-U et al (2006) Community genomics among stratified microbial assemblages in the ocean's interior. Science 311: 496–503

Dolganov NA, Bhaya D and Grossman AR (1995) Cyanobacterial protein with similarity to the chlorophyll

a/b binding proteins of higher plants: evolution and regulation. Proc Natl Acad Sci USA 92: 636–640

Doolittle WF and Zhaxybayeva O (2009) On the origin of prokaryotic species. Genome Res 19: 744–756

Dupraz C and Visscher PT (2005) Microbial lithification in marine stromatolites and hypersaline mats. Trends Microbiol 13: 429–438

Dyhrman ST, Chappell PD, Haley ST, Moffett JW, Orchard ED, Waterbury JB and Webb EA (2006) Phosphonate utilization by the globally important marine diazotroph *Trichodesmium*. Nature 439: 68–71

Edwards RA, Rodriguez-Brito B, Wegley L, Haynes M, Breitbart M, Peterson DM et al (2006) Using pyrosequencing to shed light on deep mine microbial ecology. BMC Genomics 7: 57

Eisen JA (2007) Environmental shotgun sequencing: its potential and challenges for studying the hidden world of microbes. PLoS Biol 5: e82

El-Ali J, Sorger PK and Jensen KF (2006) Cells on chips. Nature 442: 403–411

Entcheva P, Liebl W, Johann A, Hartsch T and Streit WR (2001) Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. Appl Environ Microbiol 67: 89–99

Ernst A, Becker S, Wollenzien UI and Postius C (2003) Ecosystem-dependent adaptive radiations of picocyanobacteria inferred from 16S rRNA and ITS-1 sequence analysis. Microbiology 149: 217–228

Falkowski PG, Fenchel T and Delong EF (2008) The microbial engines that drive Earth's biogeochemical cycles. Science 320: 1034–1039

Ferrer M, Beloqui A, Timmis KN and Golyshin PN (2009) Metagenomics for mining new genetic resources of microbial communities. J Mol Microbiol Biotechnol 16: 109–123

Ferris MJ and Ward DM (1997) Seasonal distributions of dominant 16S rRNA-defined populations in a hot spring microbial mat examined by denaturing gradient gel electrophoresis. Appl Environ Microbiol 63: 1375–1381

Ferris MJ, Ruff-Roberts AL, Kopczynski ED, Bateson MM and Ward DM (1996) Enrichment culture and microscopy conceal diverse thermophilic *Synechococcus* populations in a single hot spring microbial mat habitat. Appl Environ Microbiol 62: 1045–1050

Ferris MJ, Nold SC, Santegoeds CM and Ward DM (2001) Examining bacterial population diversity within Octopus Springs microbial mat community. In: Reysenbach AL, Voytek M, Mancinelli R (eds) Thermophiles, pp 51–64. Kluwer, New York

Ferris MJ, Kuhl M, Wieland A and Ward DM (2003) Cyanobacterial ecotypes in different optical microenvironments of a 68 degrees C hot spring mat community revealed by 16S-23S rRNA internal transcribed spacer region variation. Appl Environ Microbiol 69: 2893–2898

Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW and DeLong EF (2008) Microbial community gene expression in ocean surface waters. Proc Natl Acad Sci USA 105: 3805–3810

Funk C and Vermaas W (1999) A cyanobacterial gene family coding for single-helix proteins resembling part of the light-harvesting proteins from higher plants. Biochemistry 38: 9397–9404

Gabor E, Liebeton K, Niehaus F, Eck J and Lorenz P (2007) Updating the metagenomics toolbox. Biotechnol J 2: 201–206

Gentry TJ, Wickham GS, Schadt CW, He Z and Zhou J (2006) Microarray applications in microbial ecology research. Microbial Ecol 52: 159–175

Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ et al (2005) Opinion: Re-evaluating prokaryotic species. Nat Rev Microbiol 3: 733–739

Gogarten JP and Townsend JP (2005) Horizontal gene transfer, genome innovation and evolution. Nat Rev Microbiol 3: 679–687

Gomez-Garcia MR, Davison M, Hartnung MB, Grossman AR and Bhaya D (2011) Alternative pathways for phosphonate metabolism in thermophilic cyanobacteria from microbial mats. ISME J 5: 141–149

Green BR (1995) Sequence conservation of light-harvesting and stress response proteins in relation to the molecular structure of LHCII. Photosynth Res 44: 139–148

Green BR and Pichersky E (1994) Hypothesis for the evolution of three helix Chl*a/b* and Chl*a/c* light harvesting antenna proteins from two-helix and four helix ancestors. Photosynth Res 39: 149–162

Green BR, Pichersky E and Kloppstech K (1991) Chlorophyll *a/b* binding proteins: an extended family. Trends Biochem Sci 16: 181–186

Guerrero R, Piqueras M and Berlanga M (2002) Microbial mats and the search for minimal ecosystems. Int Microbiol 5: 177–188

Handelsman J (2008) Metagenomics is not enough. DNA Cell Biol 27: 219–221

Handelsman J, Rondon MR, Brady SF, Clardy J and Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol 5: R245–249

Haverkamp T, Acinas SG, Doeleman M, Stomp M, Huisman J and Stal LJ (2008) Diversity and phylogeny of Baltic Sea picocyanobacteria inferred from their ITS and phycobiliprotein operons. Environ Microbiol 10: 174–188

Havre SL, Webb-Robertson BJ, Shah A, Posse C, Gopalan B and Brockman FJ (2005) Bioinformatic insights from metagenomics through visualization. Proc IEEE Comput Syst Bioinform Conf: 341–350

He Q, Dolganov N, Bjorkman O and Grossman AR (2001) The high light-inducible polypeptides in *Synechocystis* PCC6803. Expression and function in high light. J Biol Chem 276: 306–314

Heddad M and Adamska I (2000) Light stress-regulated two-helix proteins in Arabidopsis thaliana related to the chlorophyll *a*/*b*-binding gene family. Proc Natl Acad Sci USA 97: 3741–3746

Heidelberg JF, Nelson WC, Schoenfeld T and Bhaya D (2009) Germ warfare in a microbial mat community: CRISPRs provide insights into the co-evolution of host and viral genomes. PLoS ONE 4: e4169

Herter S, Farfsing J, Gad'On N, Rieder C, Eisenreich W, Bacher A and Fuchs G (2001) Autotrophic CO_2 fixation by *Chloroflexus aurantiacus*: study of glyoxylate formation and assimilation via the 3-hydroxypropionate cycle. J Bacteriol 183: 4305–4316

Hoehler TM, Bebout BM and Des Marais DJ (2001) The role of microbial mats in the production of reduced gases on the early Earth. Nature 412: 324–327

Hoffman NE, Pichersky E, Malik VS, Castresano C, Ko K, Darr SC and Cashmore AR (1987) A cDNA encoding a photosystem I protein with homology to photosystem II chlorophyll *a/b*-binding proteins. Proc Natl Acad Sci USA 84: 8844–8848

Holo H (1989) *Chloroflexus aurantiacus* secretes 3-hydroxypropionate, a possible intermediate in the assimiliation of CO_2 and acetate. Arch Microbiol 151: 252–256

Holt NE, Zigmantas D, Valkunas L, Li XP, Niyogi KK and Fleming GR (2005) Carotenoid cation formation and the regulation of photosynthetic light harvesting. Science 307: 433–436

Howe CJ, Barbrook AC, Nisbet RE, Lockhart PJ and Larkum AW (2008) The origin of plastids. Philos Trans R Soc Lond B Biol Sci 363: 2675–2685

Huang J, Su Z and Xu Y (2005) The evolution of microbial phosphonate degradative pathways. J Mol Evol 61: 682–690

Huang B, Wu H, Bhaya D, Grossman A, Granier S, Kobilka BK and Zare RN (2007) Counting low-copy number proteins in a single cell. Science 315: 81–84

Huson DH, Auch AF, Qi J and Schuster SC (2007) MEGAN analysis of metagenomic data. Genome Res 17: 377–386

Inskeep WP and McDermott TR (2005) Geothermal Biology and Geochemistry in Yellowstone National Park. Thermal Biology Institute, Montana State University Publications, Bozeman MT

Ishii M, Chuakrut S, Arai H and Igarishi Y. (2004) Occurence, biochemistry and possible biotechnological application of the 3-hydroxypropionate cycle. Appl Microbiol Biotechnol 64: 605–610

Ishoey T, Woyke T, Stepanauskas R, Novotny M and Lasken RS (2008) Genomic sequencing of single microbial cells from environmental samples. Curr Opin Microbiol 11: 198–204

Jansson S (1999) A guide to the Lhc genes and their relatives in Arabidopsis. Trends Plant Sci 4: 236–240

Johnston AW, Li Y and Ogilvie L (2005) Metagenomic marine nitrogen fixation-feast or famine? Trends Microbiol 13: 416–420

Jungblut A.-D, Hoeger SJ, Mountfort D, Hitzfeld BC, Dietrich DR and Neilan BA (2006) Characterization of microcystin production in an Antarctic cyanobacterial mat community. Toxicon 47: 271–278

Keiter RB and Boyce MS (1991) The Greater Yellowstone Ecosystem. Yale University Press, New Haven

Kettler GC, Martiny AC, Huang K, Zucker J, Coleman ML, Rodrigue S et al (2007) Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. PLoS Genet 3: e231

Kilian O, Steunou AS, Fazeli F, Bailey S, Bhaya D and Grossman AR (2007) Responses of a thermophilic *Synechococcus* isolate from the microbial mat of Octopus Spring to light. Appl Environ Microbiol 73: 4268–4278

Kilian O, Steunou AS, Grossman AR, Bhaya D. (2008) A novel two domain-fusion protein in cyanobacteria with similarity to the CAB/ELIP/HLIP superfamily: evolutionary implications and regulation. Mol Plant 1: 155–166

Kim S, Huang B and Zare RN (2007) Microfluidic separation and capture of analytes for single-molecule spectroscopy. Lab Chip 7: 1663–1665

Koksharova OA and Wolk CP (2002) Genetic tools for cyanobacteria. Appl Microbiol Biotechnol 58: 123–137

Konopka A (1992) Accumulation and utilization of polysaccharide by hot-spring phototrophs during a light-dark transition. FEMS Microbiol Lett 102: 27–32

Konstantinidis KT and Tiedje JM (2004) Trends between gene content and genome size in prokaryotic species with larger genomes. Proc Natl Acad Sci USA 101: 3160–3165

Konstantinidis KT, Ramette A and Tiedje JM (2006) The bacterial species definition in the genomic era. Philos Trans R Soc Lond B Biol Sci 361: 1929–1940

Koonin EV (2005) Orthologs, paralogs, and evolutionary genomics. Annu Rev Genet 39: 309–338

Kowalchuk GA, Speksnijder AGCL, Zhang K, Goodman RM and van Veen JA (2007) Finding the needles in the metagenome haystack. Microb Ecol 53: 475–485

Krumbein WE, Paterson DM and Zavarin GA (2003) Fossil and Recent Biofilms: A Natural History of Life on Earth. Kluwer Academic Publishers, Dordrecht

Kuhl M (2005) Optical microsensors for analysis of microbial communities. Methods Enzymol 397: 166–199

Kuhl M, Lassen C and Revsbech NP (1997) A simple light meter for measurements of PAR (400–700 nm) with fiber-optic microprobes. Aquatic Microbial Ecol 13: 197–207

Kuhl M, Steukart C, Eikert G and Jeroschewski P. (1998) A H_2S sensor for profiling biofilms and sediments: application in an acidic lake sediment. Aquatic Microbial Ecol 15: 201–209

Kuhlbrandt W, Wang DN and Fujiyoshi Y (1994) Atomic model of plant light harvesting complex by electron crystallography. Nature 367: 614–621

Lasken RS (2007) Single-cell genomic sequencing using Multiple Displacement Amplification. Current Opin Microbiol 10: 510–516

Lau E, Nash CZ, Vogler DR and Cullings KW (2005) Molecular diversity of cyanobacteria inhabiting coniform structures and surrounding mat in a Yellowstone hot spring. Astrobiol 5: 83–92

Lerat E, Daubin V, Ochman H and Moran NA (2005) Evolutionary origins of genomic repertoires in bacteria. PLoS Biol 3: e130

Ley RE, Harris JK, Wilcox J, Spear JR, Miller SR, Bebout BM et al (2006) Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. Appl Environ Microbiol 72: 3685–3695

Li XP, Gilmore AM, Caffarri S, Bassi R, Golan T, Kramer D and Niyogi KK (2004) Regulation of photosynthetic light harvesting involves intrathylakoid lumen pH sensing by the PsbS protein. J Biol Chem 279: 22866–22874

Lorenz P, Liebeton K, Niehaus F and Eck J (2002) Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. Curr Opin Biotechnol 13: 572–577

Ludwig R, Pringault O, de Wit R, de Beer D and Jonkers HM (2006) Limitation of oxygenic photosynthesis and oxygen consumption by phosphate and organic nitrogen in a hypersaline microbial mat: a microsensor study. FEMS Microbiol Ecol 57: 9–17

Luque I, Flores E and Herrero A (1994) Nitrate and nitrite transport in the cyanobacterium *Synechococcus* sp. PCC 7942 are mediated by the same permease. Biochim Biophys Acta 1184: 296–298

Mackerras AH, Youens BN, Wier RC and Smith GD (1990) Is cyanophycin involved in the integration of nitrogen and carbon metabolism in the cyanobacteria *Anabaena cylindrica* and *Gleothece* grown on light/dark cycles? J Gen Microbiol 136: 2049–2056

Montane M.-H and Kloppstech K (2000) The family of light-harvesting-related proteins (LHCs, ELIPs, HLIPs): was the harvesting of light their primary function? Gene 258: 1–8

Nagy ML, Perez A and Garcia-Pichel F (2005) The prokaryotic diversity of biological soil crusts in the Sonoran Desert (Organ Pipe Cactus National Monument, AZ). FEMS Microbiol Ecol 54: 233–245

Namsaraev ZB, Gorlenko VM, Namsaraev BB, Buriukhaev SP and Iurkov VV (2003) [The structure and biogeochemical activity of the phototrophic communities from the Bol'sherechenskii alkaline hot spring]. Mikrobiologiia 72: 228–238

Nelson WC, Wollerman L, Bhaya D and Heidelberg JF (2011) Analysis of insertion sequences in thermophilic cyanobacteria: Exploring the mechanisms of establishing, maintaining and withstanding high insertion sequence abundance. Appl Environ Microbiol *in press*

Nold SC and Ward DM (1996) Photosynthate partitioning and fermentation in hot spring microbial mat communities. Appl Environ Microbiol 62: 4598–4607

Nubel U, Garcia-Pichel F, Kuhl M and Muyzer G (1999) Quantifying microbial diversity: morphotypes, 16S rRNA genes, and carotenoids of oxygenic phototrophs in microbial mats. Appl Environ Microbiol 65: 422–430

Nubel U, Bateson MM, Vandieken V, Wieland A, Kuhl M and Ward DM (2002) Microscopic examination of distribution and phenotypic properties of phylogenetically diverse Chloroflexaceae-related bacteria in hot spring microbial mats. Appl Environ Microbiol 68: 4593–4603

Ortmann AC, Wiedenheft B, Douglas T and Young M (2006) Hot crenarchaeal viruses reveal deep evolutionary connections. Nat Rev Microbiol 4: 520–528

Ottesen EA, Hong JW, Quake SR and Leadbetter JR (2006) Microfluidic digital PCR enables multigene analysis of individual environmental bacteria. Science 314: 1464–1467

Paerl HW (2000) Marine plankton. In: Whitton BA and Potts M (eds) The Ecology of Cyanobacteria, pp 121–148. Kluwer Academic Publishers, Norwell

Paerl HW, Pinckney JL and Steppe TF (2000) Cyanobacterial-bacterial mat consortia: examining the functional unit of microbial survival and growth in extreme environments. Environ Microbiol 2: 11–26

Palenik B, Brahamsha B, Larimer FW, Land M, Hauser L, Chain P et al (2003) The genome of a motile marine *Synechococcus*. Nature 424: 1037–1042

Papke RT, Ramsing NB, Bateson MM and Ward DM (2003) Geographical isolation in hot spring cyanobacteria. Environ Microbiol 5: 650–659

Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N, Harris DE et al (2003) Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. Nat Genet 35: 32–40

Partensky F, Hess WR and Vaulot D (1999) *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. Microbiol Mol Biol Reviews 63: 106–127

Pierson BK and Castenholz RW (1974) A phototrophic gliding filamentous bacterium of hot springs, *Chloroflexus aurantiacus*, gen. and sp. nov. Arch Microbiol 100: 5–24

Podar M and Reysenbach AL (2006) New opportunities revealed by biotechnological explorations of extremophiles. Curr Opin Biotechnol 17: 250–255

Promnares K, Komenda J, Bumba L, Nebesarova J, Vacha F and Tichy M (2006) Cyanobacterial small chlorophyll-binding protein ScpD (HliB) is located on the periphery of photosystem II in the vicinity of PsbH and CP47 subunits. J Biol Chem 281: 32705–32713

Quinn JP, Kulakova AN, Cooley NA and McGrath JW (2007) New ways to break an old bond: the bacterial

carbon-phosphorus hydrolases and their role in biogeochemical phosphorus cycling. Environ Microbiol 9: 2392–2400

Raes J and Bork P (2008) Molecular eco-systems biology: towards an understanding of community function. Nat Rev Microbiol 6: 693–699

Raes J, Foerstner KU and Bork P (2007) Get the most out of your metagenome: computational analysis of environmental sequence data. Curr Opin Microbiol 10: 490–498

Ram RJ, Verberkmoes NC, Thelen MP, Tyson GW, Baker BJ, Blake RC, 2nd et al (2005) Community proteomics of a natural microbial biofilm. Science 308: 1915–1920

Ramsing NB, Ferris MJ and Ward DM (1997) Light-induced motility of thermophilic *Synechococcus* isolates from Octopus Spring, Yellowstone National Park. Appl Environ Microbiol 63: 2347–2354

Ramsing NB, Ferris MJ and Ward DM (2000) Highly ordered vertical structure of *Synechococcus* populations within the one-millimeter-thick photic zone of a hot spring cyanobacterial mat. Appl Environ Microbiol 66: 1038–1049

Reid RP, Visscher PT, Decho AW, Stolz JF, Bebout BM, Dupraz C et al (2000) The role of microbes in accretion, lamination and early lithification of modern marine stromatolites. Nature 406: 989–992

Revsbech NP and Ward DM (1984) Microelectrode studies of interstitial water chemistry and photosynthetic activity in a hot spring microbial mat. Appl Environ Microbiol 48: 270–275

Reysenbach AL and Cady SL (2001) Microbiology of ancient and modern hydrothermal systems. Trends Microbiol 9: 79–86

Reysenbach AL and Shock E (2002) Merging genomes with geochemistry in hydrothermal ecosystems. Science 296: 1077–1082

Rice G, Stedman K, Snyder J, Wiedenheft B, Willits D, Brumfield S et al (2001) Viruses from extreme thermal environments. Proc Natl Acad Sci USA 98: 13341–13345

Richardson LL and Castenholz RW (1987) Diel vertical movements of the cyanobacterium *Oscillatoria terebriformis* in a sulfide-rich hot spring microbial mat. Appl Environ Microbiol 53: 2142–2150

Riesenfeld CS, Schloss PD and Handelsman J (2004) Metagenomics: genomic analysis of microbial communities. Annu Rev Genet 38: 525–552

Rocap G, Distel DL, Waterbury JB and Chisholm SW (2002) Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. Appl Environ Microbiol 68: 1180–1191

Rocha EP (2004) Order and disorder in bacterial genomes. Curr Opin Microbiol 7: 519–527

Rocha EP (2008) The organization of the bacterial genome. Annu Rev Genet 42: 211–233

Rubio LM and Ludden PW (2005) Maturation of nitrogenase: a biochemical puzzle. J Bacteriol 187: 405–414

Ruff-Roberts AL, Kuenen JG and Ward DM (1994) Distribution of cultivated and uncultivated cyanobacteria and *Chloroflexus*-like bacteria in hot spring microbial mats. Appl Environ Microbiol 60: 697–704

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S et al (2007) The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. PLoS Biol 5: e77

Sato M, Masuda Y, Kirimura K and Kino K (2007) Thermostable ATP regeneration system using polyphosphate kinase from *Thermosynechococcus elongatus* BP-1 for D-amino acid dipeptide synthesis. J Biosci Bioeng 103: 179–184

Schloss PD and Handelsman J (2007) The last word: books as a statistical metaphor for microbial communities. Annu Rev Microbiol 61: 23–34

Schoenfeld T, Patterson M, Richardson PM, Wommack KE, Young M and Mead D (2008) Assembly of viral metagenomes from Yellowstone hot springs. Appl Environ Microbiol 74: 4164–4174

Schopf JW (2000) The fossil record: tracing the roots of the cyanobacterial lineage. In: Whitton BA and Potts M (eds) The Ecology of Cyanobacteria, pp 13–35. Kluwer Academic Publishers, Norwell

Sheehan KB, Patterson DJ, Dicks BL and Henson JM (2005) Seen and Unseen, Discovering the Microbes of Yellowstone. Globe Press and Montana State University

Simon RD (1987) Inclusion bodies in the cyanobacteria: cyanophycin, polyphosphate, polyhedral bodies. In: Fay P and Van Baalen C (eds) The Cyanobacteria, pp 199–225. Elsevier Science Publishers BV, Amsterdam

Snyder JC, Wiedenheft B, Lavin M, Roberto FF, Spuhler J, Ortmann AC et al (2007) Virus movement maintains local virus population diversity. Proc Natl Acad Sci USA 104: 19102–19107

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR et al (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc Natl Acad Sci USA 103: 12115–12120

Soule T, Stout V, Swingley WD, Meeks JC and Garcia-Pichel F (2007) Molecular genetics and genomic analysis of scytonemin biosynthesis in *Nostoc punctiforme* ATCC 29133. J Bacteriol 189: 4465–4472

Stahl DA, Lane DJ, Olsen GJ and Pace NR (1985) Characterization of a Yellowstone hot spring microbial community by 5 S rRNA sequences. Appl Environ Microbiol 49: 1379–1384

Stal LJ (2000) Cyanobacterial mats and stromatolites. In: Whitton BA and Potts M (eds) The Ecology of Cyanobacteria, pp 61–120. Kluwer Academic Publishers, Norwell

Stal LJ and Caumette P (1994) Microbial Mats: Structure, Development and Environmental Significance. Springer Verlag, Berlin

Staley JT (1997) Biodiversity: are microbial species threatened? Curr Opin Biotechnol 8: 340–345

Standfuss J, Terwisscha van Scheltinga AC, Lamborghini M and Kuhlbrandt W (2005) Mechanisms of photoprotection and nonphotochemical quenching in pea light-

harvesting complex at 2.5 A resolution. EMBO J 24: 919–928

Steglich C, Futschik M, Rector T, Steen R and Chisholm SW (2006) Genome-wide analysis of light sensing in *Prochlorococcus*. J Bacteriol 188: 7796–7806

Steunou AS, Bhaya D, Bateson MM, Melendrez MC, Ward DM, Brecht E et al (2006) In situ analysis of nitrogen fixation and metabolic switching in unicellular thermophilic cyanobacteria inhabiting hot spring microbial mats. Proc Natl Acad Sci USA 103: 2398–2403

Steunou AS, Jensen SI, Brecht E, Becraft ED, Bateson MM, Kilian O et al (2008) Regulation of *nif* gene expression and the energetics of N_2 fixation over the diel cycle in a hot spring microbial mat. ISME J 2: 364–378

Stewart WDP (1970) Nitrogen fixation by blue-green algae in Yellowstone thermal areas. Phycologia 9: 261–268

Strauss G and Fuchs G (1993) Enzymes of a novel autotrophic CO_2 fixation pathway in the phototrophic bacterium *Chloroflexus aurantiacus*, the 3-hydroxypropionate cycle. Eur J Biochem 215: 633–643

Streit WR and Schmitz RA (2004) Metagenomics - the key to the uncultured microbes. Curr Opin Microbiol 7: 492–498

Strous M, Pelletier E, Mangenot S, Rattei T, Lehner A, Taylor MW et al (2006) Deciphering the evolution and metabolism of an anammox bacterium from a community genome. Nature 440: 790–794

Suzuki MT and DeLong EF (2002) Marine prokaryotic diversity. In: Staley JT and Reysenbach AL (eds) Biodiversity of Microbial Life, pp 209–234. Wiley-Liss, New York

Teske A and Stahl D (2002) Microbial mats and biofilms:evolution, structure, and function of fixed microbial communities. In: Staley JT and Reysenbach AL (eds) Biodiversity of Microbial Life, pp 49–100. Wiley-Liss, New York

Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW et al (2005) Comparative metagenomics of microbial communities. Science 308: 554–557

Tyson GW and Banfield JF (2005) Cultivating the uncultivated: a community genomics perspective. Trends Microbiol 13: 411–415

Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM et al (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428: 37–43

van der Meer MT, Schouten S, Bateson MM, Nubel U, Wieland A, Kuhl M et al (2005) Diel variations in carbon metabolism by green nonsulfur-like bacteria in alkaline siliceous hot spring microbial mats from Yellowstone National Park. Appl Environ Microbiol 71: 3978–3986

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA et al (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304: 66–74

Vieites JM, Guazzaroni ME, Beloqui A, Golyshin PN and Ferrer M (2009) Metagenomics approaches in systems microbiology. FEMS Microbiol Rev 33: 236–255

Vincent WF (2000) Cyanobacterial dominance in the polar regions. In: Whitton BA and Potts M (eds) The Ecology of Cyanobacteria, pp 321–340. Kluwer Academic Publishers, Norwell

Vincent WF, Mueller DR and Bonilla S (2004) Ecosystems on ice: the microbial ecology of Markham Ice Shelf in the high Arctic. Cryobiology 48: 103–112

von Mering C, Hugenholtz P, Raes J, Tringe SG, Doerks T, Jensen LJ et al (2007) Quantitative phylogenetic assessment of microbial communities in diverse environments. Science 315: 1126–1130

Wanner BL (1994) Phosphate-regulated genes for the utilization of phosphonates in members of the family *Enterobacteriaceae*. In: Torriani-Gorrini A, Yagil E and Silver S (eds) Phosphate in Microorganisms: Cellular and Molecular Biology, pp 215–222. ASM Press, Washington

Ward DM (2006a) Microbial diversity in natural environments: focusing on fundamental questions. Antonie Van Leeuwenhoek 90: 309–324

Ward N (2006b) New directions and interactions in metagenomics research. FEMS Microbiol Ecol 55: 331–338

Ward DM and Castenholz RW (2000) Cyanobacteria in geothermal habitats. In: Whitton BA and Potts M (eds) The Ecology of Cyanobacteria, pp 37–59. Kluwer Academic Publishers, Norwell

Ward DM and Cohan FM (2005) Microbial diversity in hot spring cyanobacterial mats: pattern and prediction. In: Inskeep WT and McDermott T (eds) Geothermal Biology and Geochemistry in Yellowstone National Park, pp 185–201. Thermal Biology Institute, Montana State University, Bozeman

Ward DM, Weller R and Bateson MM (1990) 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature 345: 63–65

Ward DM, Panke S, Kloppel KD, Christ R and Fredrickson H (1994) Complex polar lipids of a hot spring cyanobacterial mat and its cultivated inhabitants. Appl Environ Microbiol 60: 3358–3367

Ward DM, Ferris MJ, Nold SC and Bateson MM (1998) A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. Microbiol Mol Biol Rev 62: 1353–1370

Ward DW, Papke RT, Nubel U and McKitrick MC (2002) Natural history of microorganisms inhabiting hot spring microbial mat communities: Clues to the origin of microbial diversity and implications for microbiology and macrobiology. In: Staley JT and Reysenbach AL (eds) Biodiversity of Microbial Life, pp 25–48. Wiley-Liss, New York

Warnecke F and Hugenholtz P (2007) Building on basic metagenomics with complementary technologies. Genome Biol 8: 231

West NJ, Schonhuber WA, Fuller NJ, Amann RI, Rippka R, Post AF and Scanlan DJ (2001) Closely related

Prochlorococcus genotypes show remarkably different depth distributions in two oceanic regions as revealed by in situ hybridization using 16S rRNA-targeted oligonucleotides. Microbiology 147: 1731–1744

Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V et al (2008) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 36: D13–21

Whitman WB, Coleman DC and Wiebe WJ (1998) Prokaryotes: the unseen majority. Proc Natl Acad Sci USA 95: 6578–6583

Whitton BA and Potts M (2000) Introduction to the cyanobacteria. In: Whitton BA and Potts M (eds) The Ecology of Cyanobacteria, pp 1–11. Kluwer Academic Publishers, Norwell, MA

Wilmes P, Simmons SL, Denef VJ and Banfield JF (2009) The dynamic genetic repertoire of microbial communities. FEMS Microbiol Rev 33: 109–132

Woyke T, Teeling H, Ivanova NN, Huntemann M, Richter M, Gloeckner FO et al (2006) Symbiosis insights through metagenomic analysis of a microbial consortium. Nature 443: 950–955

Woyke T, Xie G, Copeland A, Gonzalez JM, Han C, Kiss H et al (2009) Assembling the marine metagenome, one cell at a time. PLoS ONE 4: e5299

Wynn-Williams DD (2000) Cyanobacteria in deserts- Life at the limit? In: Whitton BA and Potts M (eds) The Ecology of Cyanobacteria, pp 37–59. Kluwer Academic Publishers, Norwell

Xu H, Vavilin D, Funk C and Vermaas W (2002) Small Cab-like proteins regulating tetrapyrrole biosynthesis in the cyanobacterium *Synechocystis* sp. PCC6803. Plant Mol Biol 49: 149–160

Yanai I, Camacho CJ and DeLisi C (2000) Predictions of gene family distributions in microbial genomes: evolution by gene duplication and modification. Phys Rev Lett 85: 2641–2644

Yao D, Kieselbach T, Komenda J, Promnares K, Prieto MA, Tichy M et al (2007) Localization of the small CAB-like proteins in photosystem II. J Biol Chem 282: 267–276

Yooseph S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, Remington K et al (2007) The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. PLoS Biol 5: e16

Yoshihara S and Ikeuchi M (2004) Phototactic motility in the unicellular cyanobacterium *Synechocystis* sp. PCC 6803. Photochem Photobiol Sci 3: 512–518

Zhou F, Olman V and Xu Y (2008) Insertion sequences show diverse recent activities in Cyanobacteria and Archaea. BMC Genomics 9: 36

Chapter 3

Comparative and Functional Genomics of Anoxygenic Green Bacteria from the Taxa *Chlorobi*, *Chloroflexi*, and *Acidobacteria*

Donald A. Bryant*, Zhenfeng Liu, Tao Li, Fangqing Zhao and Amaya M. Garcia Costas
Department of Biochemistry and Molecular Biology, The Pennsylvania State University, S-235 Frear Building, University Park, PA 16802, USA

Christian G. Klatt and David M. Ward
Departments of Land Resources and Environmental Sciences and Chemistry, Montana State University, Bozeman, MT 59717, USA

Niels-Ulrik Frigaard
Department of Biology, University of Copenhagen, 3000 Helsingør, Denmark

Jörg Overmann
Leibniz-Institut DSMZ - Deutsche Sammlung von Mikroorganismen, und Zellkulturen GmbH, Inhoffenstraße 7B, 38124 Braunschweig, Germany

*Author for correspondence, e-mail: dab14@psu.edu

R.L. Burnap and W.F.J. Vermaas (eds.), Functional Genomics and Evolution of Photosynthetic Systems, Advances in Photosynthesis and Respiration 33, pp. 47–102, DOI 10.1007/978-94-007-1533-2_3,

Summary

Green bacteria are a diverse group of chlorophototrophic organisms belonging to three major taxa within the domain *Bacteria*: *Chlorobi*, *Chloroflexi*, and *Acidobacteria*. Most, although not all, of these organisms synthesize bacteriochlorophylls *c*, *d* or *e* and utilize chlorosomes for light harvesting. The pace of discoveries concerning the metabolism and physiology of these bacteria has accelerated rapidly since completion of the sequencing of the genomes of the green sulfur bacterium *Chlorobaculum tepidum* and the filamentous anoxygenic phototroph, *Chloroflexus aurantiacus*. This chapter summarizes insights gained from the extensive genome sequence data for members of these three taxa. The discovery of the first chlorophototrophic member of the phylum *Acidobacteria*, *Candidatus* Chloracidobacterium thermophilum, is also described, and recent insights into the physiology and metabolism of this unique, aerobic photoheterotroph are presented. Based upon phylogenetic inferences derived from analyses of sequences for reaction centers and enzymes of (bacterio)chlorophyll biosynthesis, some implications concerning the evolutionary origins of photosynthesis are discussed.

I. Introduction

Thanks to the pioneering efforts of Dr. Satoshi Tabata and colleagues at the Kazusza DNA Research Institute, the photosynthesis research community was one of the first scientific groups to enter the genomic era. Remarkably, when the description of the 3.4-Mb *Synechocystis* sp. PCC 6803 genome was published in June 1996, it was only the third genome to be completely sequenced (Kaneko et al., 1996). Only the much smaller genomes of *Haemophilis influenzae* (1.83 Mb; Fleischmann et al., 1995) and *Mycoplasma genitalium* (0.58 Mb; Fraser et al., 1995), which were determined by The Institute for Genomic Research (TIGR), had been completed before that of *Synechocystis* sp. PCC 6803, and the 4.64-Mb genome of *Escherichia coli* K-12 would not published until more than 1 year later (Blattner et al., 1997). In addition to their contributions to the sequencing of the *Arabidopsis thaliana* genome and systems biology research, Tabata and coworkers subsequently completed three additional cyanobacterial genomes: *Nostoc* (formerly *Anabaena*) sp. PCC 7120 (Kaneko et al., 2001), *Thermosynechococcus elongatus* BP-1 (Nakamura et al., 2002), and *Gloeobacter violaceus* PCC 7421 (Nakamura et al., 2003).

Abbreviations: APS – adenosine-5′-phosphosulfate; BChl – bacteriochlorophyll; BChlide – bacteriochlorophyllide; BPhe – bacteriopheophytin; *Cab.* – *Chloracidobacterium*; *Cba.* – *Chlorobaculum*; *Cfx.* – *Chloroflexus*; *Chl.* – *Chlorobium*; Chl – chlorophyll; Chlide – chlorophyllide; *Chp.* – *Chloroherpeton*; COG – cluster of orthologous genes; COR – chlorophyllide oxidoreductase; Cyt – cytochrome; DPOR – dark-operative protochlorophyllide oxidoreductase; DSR – dissimilatory sulfite reductase; FAP – filamentous anoxygenic phototroph; Fd – ferredoxin; FMO – Fenna-Matthews-Olson protein; FNR – ferredoxin:$NADP^+$ oxidoreductase; GSB – green sulfur bacterium/bacteria; *H.* – *Herpetosiphon*; *I.* – *Ignavibacterium*; *O.* – *Oscillochloris*; ORF – open reading frame; PAPS – 3′-phosphoadenosine-5′-phosphosulfate; PChlide – protochlorophyllide; Phe – pheophytin; Proto IX – protoporphyrin IX; *Pld.* – *Pelodictyon*; PSB – purple sulfur bacteria; *Ptc.* – *Prosthecochloris*; *R.* – *Roseiflexus*; ROS – reactive oxygen species; SAM – S-adenosyl-L-methionine; SQR – sulfide:quinone oxidoreductase; *T.* – *Thermomicrobium*

Since the first bacterial genome sequences were determined in the mid-nineties, genome sequencing has become far less expensive, much faster, and thus much more commonplace - to the extent that now there really are only two types of bacteria: those with sequenced genomes and those without. As of July 2010, the Genomes Online database v. 2.0 (http://www.genomesonline.org/cgi-bin/GOLD/bin/gold.cgi) lists 1,100 completed bacterial genomes and 4,617 ongoing bacterial genome projects, and because of the increasing number of sequences determined by individual researchers and industrial research laboratories, this is not likely to be a complete listing. Chlorophototrophs, organisms that can derive some or all of their energy from chlorophyll (Chl)-based phototrophy, presently represent about 200 (4.4%) of these genome projects. In order of their discovery, there are currently six bacterial taxa (either kingdoms or phyla, depending on one's view of taxonomy) that contain chlorophototrophic members: *Cyanobacteria*, *Proteobacteria*, *Chlorobi*, *Chloroflexi*, *Firmicutes*, and *Acidobacteria* (Bryant and Frigaard, 2006; Bryant et al., 2007). The completed and ongoing sequencing projects are not evenly distributed across these six taxonomic groupings. *Cyanobacteria* (~120 projects; 60%) and *Proteobacteria* (~60 projects; 30%) account for ~90% of the sequencing projects. *Chlorobi* (15 strains), *Chloroflexi* (~12 strains), *Firmicutes*/heliobacteria (two strains), and *Acidobacteria* (one strain) make up the remaining 10% of chlorophototroph genome projects. This chapter will summarize some of the results derived from sequencing the genomes of "green bacteria," which will here be defined as the chlorophototrophic members of three kingdoms/phyla: *Chlorobi*, *Chloroflexi*, and *Acidobacteria*.

II. *Chlorobiales*: Green Sulfur Bacteria

A. Brief Description of Chlorobi

Nadson (1906) described the first green sulfur bacterium (GSB), *Chlorobium limicola*, in 1906. For most of the ensuing century, it was widely accepted that organisms of the kingdom/phylum *Chlorobi* were phylogenetically coherent, and that all of the organisms of the taxon *Chlorobi* were physiologically and metabolically quite similar. This situation abruptly changed in 2009 because of the discovery and description of *Ignavibacterium album*, the first non-chlorophototrophic member of the *Chlorobi* (Iino et al., 2010). *I. album* (phylum *Chlorobi*, class *Ignavibacteria*, order *Ignavibacteriales*, family *Ignavibacteriaceae*) is a moderately thermophilic chemoheterotroph that does not synthesize chlorophylls and was isolated by growing the organism fermentatively on glucose, pyruvate, and a few other sugars. It has an exceptionally low mol-% G+C value of 33.5%. However, the 3.75-Mb draft genome sequence of *I. album* suggests this organism is an aerobic heterotroph, which has flagella for swimming motility, lacks the ability to reduce sulfate and nitrate, and probably depends on amino acids as its nitrogen source (Z. Liu, T. Iino, N-U. Frigaard, J. Overmann, Y. Kosako, and D. A. Bryant, unpublished results). The discovery of *I. album* has made it necessary to place the GSB into a separate class, *Chlorobea*, order *Chlorobiales,* and family (*Chlorobacteriaceae*) (Iino et al., 2010).

Members of the order *Chlorobiales* are chlorophototrophs that are commonly found in stratified, anoxic sediments and aquatic environments that contain substantial amounts of H_2S and that receive some light (van Gemerden and Mas, 1995; Overmann, 2001). Some GSB are also found in mat communities associated with sulfide-rich, thermal features (Wahlund et al., 1991). Because of the unique properties of their light-harvesting antennae, i.e., chlorosomes (Frigaard and Bryant, 2006; Ganapathy et al., 2009), these organisms are capable of phototrophic growth in astonishingly low-light environments (Beatty et al., 2005; Manske et al., 2006; Overmann, 2006). Although GSB are frequently found in association with purple sulfur bacteria (PSB) in natural ecosystems, they can account for up to 83% of the total annual productivity in some aquatic environments. Thus, it is clear that these organisms can be important contributors of fixed carbon in certain ecological niches (Overmann, 1997). Although only five to seven genera and only about 20 distinct species have been described (Overmann, 2001; Imhoff, 2003), the members of the *Chlorobiales* are nevertheless morphologically diverse; they include non-motile, occasionally gas-vacuolated organisms, which may be coccoid, ovoid, rod- or vibrio-shaped, filamentous or even prosthecate (Imhoff, 1995; Overmann, 2001).

Physiological and biochemical studies have suggested that GSB/*Chlorobiales* are metabolic

specialists, although this perception might be skewed by a recent radiation of species within the group. This has been suggested to derive from a proposed, recent lateral acquisition of genes that are related to those encoding a dissimilatory sulfite reductase (see Section II.D.2 below). This gain of function would have allowed sulfide to be oxidized to sulfite (and eventually to sulfate) and might have significantly expanded the niches available to these organisms. All described species are strictly anaerobic and obligately photolithoautotrophic in growth mode. No strain that is capable of dark respiratory or strictly fermentative metabolism has yet been reported. Although all characterized GSB can grow with CO_2 as the sole C-source, most isolates can also assimilate varying amounts of simple organic acids such as acetate and pyruvate under photomixotrophic conditions. Most if not all GSB can synthesize nitrogenase and thus can grow with N_2 as sole N-source. Although some strains can oxidize thiosulfate, sulfide is almost universally used as an electron donor for CO_2 reduction as well as the S-source (Frigaard and Bryant, 2008a, b; Frigaard and Dahl, 2009). Both compounds are oxidized to sulfate with the intermediate, extracellular deposition of elemental sulfur globules. *Chlorobium ferrooxidans*, which obtains the electrons for CO_2 fixation from the oxidation of ferrous iron (Fe^{2+}; Heising et al., 1999; Hegler et al., 2008), is the only characterized strain that does not oxidize thiosulfate or sulfide to provide electrons for CO_2 fixation. *Chl. ferrooxidans* is correspondingly the only known GSB that obtains its sulfur for growth by assimilatory sulfate reduction. Many strains can additionally use H_2 as an electron donor for photoautotrophic growth. Although GSB share many biochemical and metabolic properties, as judged by 16S rRNA comparisons, the *Chlorobiales* appear to be only very distantly related to other chlorophototrophic bacteria (*Cyanobacteria*, purple bacteria (*Proteobacteria*), *Chloroflexi* and *Firmicutes,* and *Acidobacteria*) (Overmann and Tuschak, 1997; Overmann, 2001; Iino et al., 2010). However, based upon phylogenetic analyses of rRNA genes and housekeeping proteins, members of the *Chlorobi* share a common origin with the *Bacteriodetes* and *Flavobacteria* (Ludwig and Klenk, 2001).

Carbon fixation in GSB differs significantly from the Calvin-Benson-Bassham (CBB) cycle that occurs in plants, cyanobacteria, and purple bacteria. In the CBB cycle, the key carboxylating enzyme, ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO), carboxylates and then cleaves ribulose 1,5-bisphosphate to produce two molecules of 3-phosphoglycerate. In contrast, GSB fix CO_2 by the reductive carboxylation of succinyl-CoA, α-ketoglutarate, and acetyl-CoA to form α-ketoglutarate, isocitrate, and pyruvate, respectively (Sirevåg, 1995) (see Chapter 9). Thus, CO_2 is fixed by reversing the flow of carbon through the TCA cycle in a set of reactions collectively referred to as the reductive (or reverse) TCA cycle (Buchanan and Arnon, 1990). In this cycle, three CO_2 molecules can be converted to triose-P in a reaction that requires ten hydrogen equivalents (*i.e.*, 10 electrons + 10 protons) and 5 mol of ATP per mole of triose-P produced. A critical requirement for this process is the availability of strongly reducing ferredoxins that provide the electrons required to catalyze the carboxylation reactions. GSB have been found to contain large amounts of water-soluble, 8Fe-8S ferredoxins (the *Chlorobaculum* (*Cba.*) *tepidum* genome encodes three such proteins), and two of these ferredoxins have been shown to function in the key carboxylation reactions (Yoon et al., 2001).

B. Genome Sequences of Green Sulfur Bacteria: Chlorobiales

The model GSB *Cba. tepidum* (formerly *Chl. tepidum*; Imhoff, 2003) was the first chlorophototrophic bacterium other than a cyanobacterium to have its genome sequenced (Eisen et al., 2002). Since that time, 11 additional GSB strains have been completely sequenced, one strain has been nearly completely sequenced, and two additional genomes have recently been determined in draft form by pyrosequencing (Table 3.1). Although the gene sequences surrounding the only remaining gap in the genome of *Chl. ferrooxidans* are similar to those for a region of the *Cba. tepidum* genome, repeated attempts to close this last gap have failed, and no further attempts at closing are planned (Liu and Bryant, 2011). Based upon previous experience, the most likely cause for this failure is the presence of a very large inverted repeat in this intergenic region.

Table 3.1 summarizes some general information about the genomes of these 15 *Chlorobiales* strains. The chromosomes of all GSB strains

Table 3.1. Properties of the genomes of sequenced *Chlorobiales* (GSB) strains

Strain	Genome size (bp)[a]	Mol-% GC content	ORFs	Ave. ORF length (bp)	rRNA operons	tRNAs	ORFs in COGs	ORFs not in COGs	Insertion sequences	Unique ORFs[f]
Chlorobium phaeovibriodes DSM 265	1,966,858	52.99	1,753	1,021	1	45	1,501	252	22	119
Chlorobaculum tepidum ATCC 49652^T	2,154,946	56.53	2,245	926	2	50	1,607	638	19	140
Chlorobaculum parvum NCBI 8327	2,289,249	55.81	2,043	964	2	49	1,678	365	2	216
Chlorobium luteolum DSM 273^T	2,364,842	57.33	2,083	1,010	2	48	1,733	350	8	214
Chlorobium ferrooxidans DSM 13031^T	~2,558,373[b]	50.11	2,349	941	2	44	1,645	328	5	328
Chlorobium chlorochromatii CaD3[c]	2,572,079	44.28	2,002	1,131	1	45	1,586	416	25	307
Prosthecochloris aestuarii DSM 271^T	2,579,695[d]	50.05	2,327	966	1	47	1,845	482	43	303
Chlorobium phaeobacteroides BS-1	2,736,403	48.92	2,469	908	2	46	1,886	645	45	371
Chlorobium limicola DSM 245^T	2,763,181	51.33	2,434	976	2	48	1,794	523	56	349
Chlorobium clathratiforme DSM 5477^T	3,018,238	48.05	2,707	948	3	49	2,062	528	48	528
Chlorobium phaeobacteroides DSM 266^T	3,133,902	48.35	2,650	993	2	47	2,033	617	88	643
Chloroherpeton thalassium ATCC 35110^T	3,293,448	45.06	2,710	975	1	45	2,126	584	10	1,139
Chlorobium bathyomarinum GSB1, "TY Vent"	2,468,479	55.97	2,447	~950	2	48	~1,600	~500	N. D.	~300
Pelodictyon phaeum CIB2401	~2,385,000[e]	51.9	~2,200	~950	2	~45	~1,600	~300	N. D.	~300
Chlorobaculum limnaeum DSM 1677^T	~2,700,000[e]	56.3	~2,400	~950	2	~45	~1,800	~600	N. D.	~300

[a] All genomes are circular, double-stranded DNA molecules

[b] The *Chl. ferrooxidans* genome is complete except for one gap that is believed to contain a long, inverted repeat that could not be amplified by PCR. Consistent with this interpretation, the genes on either side of this gap in this genome are syntenous with those in other GSB genomes

[c] *Chl. chlorochromatii* is the epibiont from the phototrophic consortium *'Chlorochromatium aggregatum,'* which was isolated from Lake Dagow (Vogl et al., 2006, 2008)

[d] These values include those for circular plasmid pPaes01, which is 66,772 bp

[e] Only a draft genome from pyrosequencing is currently available

[f] Genes not found in other GSB genomes

appear to be circular, double-stranded DNA molecules with sizes ranging from slightly less than 2.0 Mb up to about 3.3 Mb, and they have mol-% GC contents extending from 44% to 57%. Although plasmids have been reported to occur in some GSB strains (Méndez-Alvarez et al., 1994, 1995), among the sequenced strains only *Prosthecochloris* (*Ptc.*) *aestuarii* DSM 271^T harbors a plasmid (pPaes1: 66,772 bp). Although the *Chl. clathratiforme* genome encodes three sets of rRNA genes, most of the *Chlorobiales* genomes contain only one or two copies of the rRNA genes. This observation is generally consistent with the comparatively slow growth rates of *Chlorobiales* strains and their occurrence in environments that do not change rapidly (Lee et al., 2009b). The number of genes encoding tRNAs similarly varied over a small range (44–50); the strain with the fastest growth rate, *Cba. tepidum*, had the largest number of tRNAs. As predicted from the number of open reading frames (ORFs), the protein-coding potential differs by about 1,000 ORFs, from a minimum of 1,753 ORFs in *Chl. phaeovibrioides* to a maximum of 2,710 ORFs in *Chloroherpeton* (*Chp.*) *thalassium*. Approximately 15–25% of the genes in any given genome could not be assigned to known clusters of orthologous genes (COGs). As the genome size increases, the number of unique genes, *i.e.*, those genes for which orthologs identified by reciprocal best-Blast hit are not found in another GSB, increases about tenfold, from a minimum value of 119 genes in *Chl. phaeovibrioides* to a maximum of 1,139 genes in *Chp. thalassium*. The latter value is nearly a factor of two larger than that for the similarly sized genome of *Chl. clathratiforme* (643 unique ORFs). In general these observations are consistent with phylogenetic analyses, which show that *Chp. thalassium* is not closely related to other GSB strains (see Fig. 3.1 and discussion below).

Comparative analyses using blastp (Altschul et al., 1990) to determine the reciprocal best hits in each genome pair show that *Chlorobiales* strains contain a large core set of genes that are shared by all strains. Consistent with its divergent relationship to other GSB as predicted by 16S rRNA phylogenetic analyses, *Chp. thalassium* shares the smallest number of genes, about 1,300 genes, in pairwise comparisons with other strains, and thus it currently establishes the minimal core set of shared genes for the class. This will certainly change when the genome of *I. album* is

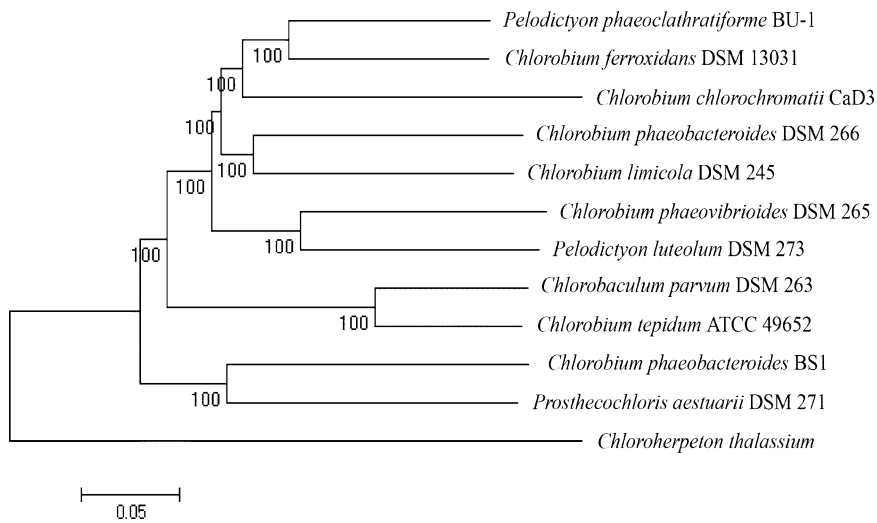


Fig. 3.1. A maximum likelihood phylogenetic tree of 813 concatenated proteins conserved in the genomes of 12 *Chlorobiales* strains. OrthoMCL (Li et al., 2003) was used to detect and group the orthologous proteins in 12 strains of *Chlorobiales.* The program builds the list of orthologs by doing an all-against-all BLASTP search, and the orthologs are clustered using the Markov clustering algorithm. 813 clusters containing one protein from each strain were detected. The protein sequences from each strain were concatenated and used to generate phylogenetic trees using the PhyML package. The tree shown in the figure was generated by maximum likelihood with 200 bootstrap resamplings. Similar results with 100% bootstrap support were obtained using the neighbor joining and maximum parsimony methods and when the collection of shared proteins was randomly sampled and analyzed.

completely sequenced. Nevertheless, the presence of this large set of core genes, representing nearly half of its genome, clearly establishes that *Chp. thalassium* is a member of the phylum *Chlorobi*. Although there is some variation in shared gene content, the average number of shared genes for all other pairwise comparisons of sequenced strains ranges from about 1,400–1,700 genes. These observations indicate that any given *Chlorobiales* strain shares a significant percentage of its total gene content (~50–90%) with those of all other strains of this phylum—the consequence being that all *Chlorobiales* obviously share a large number of biochemical and physiological traits. However, a closer analysis also shows that some pairs of strains share a significantly higher than average number of these genes in syntenous arrangements (T. Li, F. Zhao, Z. Liu, and D. A. Bryant, unpublished results). For example, not only do *Chl. luteolum* and *Chl. phaeovibrioides* have very similar gene contents, but ~85% of the genes in *Chl. phaeovibrioides* are in syntenous gene neighborhoods with their counterparts in *Chl. luteolum*. Other pairs of organisms showing elevated gene synteny are *Cba. tepidum* and *Cba. parvum*; *Chl. limicola* and *Chl. phaeobacteroides*; and *Chl. phaeobacteroides* BS-1 and *Ptc. aestuarii* (it is clear that *Chl. phaeobacteroides* BS-1 and *Chl. bathyomarinum* TyVent will have to be renamed as *Prosthecochloris* spp.). These genome synteny results are indicative of closer phylogenetic relationships for these pairs of organisms. Consistent with the divergent relationship of *Chp. thalassium* to other members of the *Chlorobiales*, less than half of the genes that it shares with all other *Chlorobiales* strains are in syntenous arrangements. These observations emphasize how divergent this unusual GSB strain is in comparison to all other strains that have been characterized to present. The quantitative assessment of gene synteny was determined using custom software (T. Li, F. Zhao, and Q. Ji, unpublished results), although other methods have been described that produce similar information (*e.g.*, Horimoto et al., 2001).

Figure 3.2 shows a global analysis of genome synteny for ten *Chlorobiales* strains. The X-shaped patterns seen in these comparisons indicate that gene inversions have frequently occurred relative to the origin of replication. Such patterns have been previously observed in analyses of other closely related bacterial genomes, and replication-directed translocations are believed to be the underlying cause of this pattern (Tiller and Collins, 2000). Large blocks of genes occur in similar positions in most of these ten genomes, and in some instances, only a few large inversions, insertions, and deletions have occurred (*e.g.*, compare *Chl. phaeovibrioides* and *Chl. luteolum* or *Cba. tepidum* and *Cba. parvum*). In contrast, although there are some conserved transcription units in *Chl. chlorochromatii*, relatively few genes and transcription units occur in similar positions or even in inverted positions relative to the origin of replication. This difference suggests that the *Chl. chlorochromatii* genome is evolving much faster and perhaps by a different mechanism than the other *Chlorobiales* genomes. This is perhaps not surprising when one considers that this genome has one of the lowest GC values observed for the class *Chlorobea* and that this organism lives symbiotically with a *Betaproteobacterium* (Fröstl and Overmann, 1998a, b; Vogl et al., 2006, 2008; Wanner et al., 2008). The patterns of closer relatedness noted above for local genome synteny are also observed in this overall visualization of global synteny.

The *Chlorobiales* have been the subject of several taxonomic studies, and based on sequence information for the 16S rRNA gene and the *fmoA* gene, which encodes the apoprotein for the Fenna-Matthews-Olson bacteriochlorophyll *a*-binding protein, Imhoff (2003) recently proposed a revision of the taxonomy of the *Chlorobiales*. Although he did not include *Chp. thalassium* in his analyses, Imhoff suggested that most GSB strains should be assigned to one of only three genera: *Chlorobium*, *Chlorobaculum*, and *Prosthecochloris.* He further suggested that the genus *Pelodictyon* should be eliminated and its strains transferred to genus *Chlorobium* (Imhoff, 2003). Figure 3.1 shows a maximum likelihood phylogenetic tree constructed from an analysis of 813 concatenated proteins shared by 12 *Chlorobiales* strains (F. Zhao, Z. Liu, and T. Li, unpublished results). It should be noted that this tree has very strong statistical support and that all of the bootstrap values are 100%. Essentially identical results were obtained when other phylogenetic methods (*e.g.*, neighbor joining, maximum parsimony) were used, or when the sampling

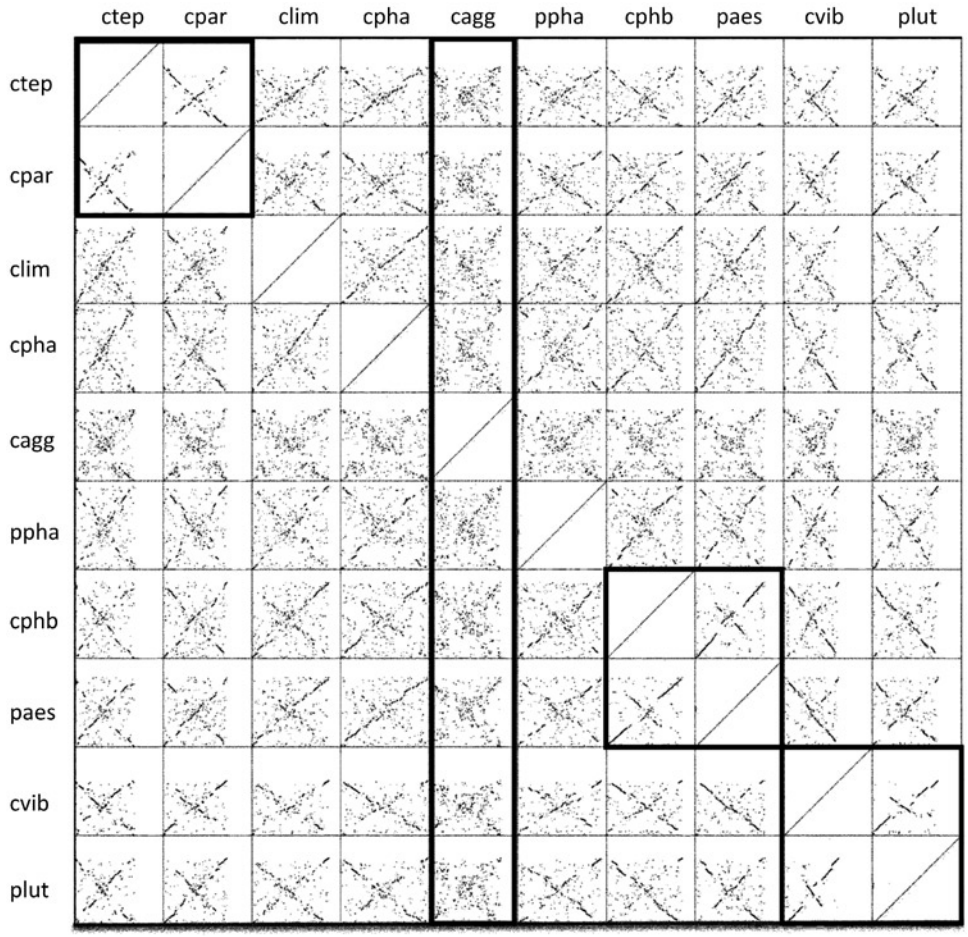


Fig. 3.2. Pairwise genomic synteny comparisons for ten *Chlorobiales* genomes. Pairwise BLASTN comparisons (cutoff: $1 \times e^{-4}$) of ten complete *Chlorobiales* genomes were used to identify conserved, orthologous gene blocks. Custom PERL scripts (developed by F. Z.) were used to visualize the genomic synteny. The box in the upper left corner highlights the higher degree of synteny for *Cba. tepidum* (ctep) and *Cba. parvum* (cpar). The box in the lower right corner emphasizes the higher degree of synteny for *Chl. luteolum* (plut) and *Chl. phaeovibrioides* (cvib). The upper box at the lower right indicates the higher degree of synteny for *Ptc. aestuarii* (paes) and *Chl. phaeobacteroides BS1* (cphb). The elongated box in the middle indicates the very low degree of synteny exhibited by *Chl. chlorochromatii* (cagg) with nine other *Chlorobiales* strains. The other abbreviations are *Chl. limicola* (clim); *Chl. phaeobacteroides* DSM 266 (cpha); and *Chl. clathratiforme* or *Pelodictyon phaeoclathratiforme* (ppha).

procedure for the sequences was changed (*e.g.*, when random sets of proteins were selected and compared from the set of 813 shared proteins). As can be seen in Fig. 3.1, the 12 *Chlorobiales* strains are distributed into four clades, a result that agrees well with the genera proposed by Imhoff (2003): *Chlorobium, Chlorobaculum, Prosthecochloris,* and *Chloroherpeton*. The closer relationships of some pairs of strains noted above are clearly reflected by the phylogenetic analyses. Consistent with the differences noted above, *Chp. thalassium* is obviously not closely related to the other sequenced *Chlorobiales* strains. It should be noted, however, that an organism that is phylogenetically related, albeit very distantly, to *Chp. thalassium* occurs in the mats of Octopus and Mushroom Springs in Yellowstone National Park (Bryant et al., 2007; see Fig. 2C of Ward et al., Chapter 1 of this volume). The very large phylogenetic distance between *Chp. thalassium* and other *Chlorobiales* strains suggests that it may be appropriate to divide the *Chlorobea* into two families: the α-*Chlorobiales* (including *Chp. thalassium* and the Yellowstone *Chlorobiales* strain) and β-*Chlorobiales* (strains of the genera *Chlorobium*, *Chlorobaculum*, and *Prosthecochloris*).

C. Genes for Photosynthesis

1. Photosynthetic Apparatus and Electron Transport

Prior to the discovery of *Cab. thermophilum* (see Section IV), the photosynthetic apparatus of GSB was unique among chlorophototrophs. The GSB light-harvesting and energy conversion system universally comprises a homodimeric type-1 reaction center characteristically comprising four subunits (PscA, PscB, PscC, and PscD) and with BChl *a* and Chl *a*; the BChl *a*-binding Fenna-Mattthews-Olson (FMO) protein (FmoA); and light-harvesting chlorosomes containing BChl *c*, *d* or *e* and smaller amounts (~1% w/w) of BChl *a* (for reviews, see Hauska et al., 2001; Frigaard and Bryant, 2004, 2006; Tronrud et al., 2009). All type-1 reaction centers have a homodimeric core in anoxygenic chlorophototrophs or a heterodimeric core in oxygenic chlorophototrophs; these core subunits are large polypeptides with 11 transmembrane α-helices (Hauska et al., 2001; Heinnickel and Golbeck, 2007; Bryant et al., 2007; Fromme et al., 2001). These large polypeptides bind most of the electron transfer cofactors and (B)Chls that occur in these reaction centers. Additionally, all type-1 reaction centers have a second, membrane-extrinsic subunit, which is an 8Fe-8S ferredoxin that carries the two terminal [4Fe-4S] clusters that participate in the transfer of electrons from the special pair to a water-soluble electron acceptor that is usually another ferredoxin (Jagannathan and Golbeck, 2008, 2009).

The reaction centers of all *Chlorobiales* strains differ from that of *Cab. thermophilum* by having two subunits, PscC and PscD, which are not encoded in the *Cab. thermophilum* genome (A. Garcia Costas, Z. Liu, and D. A. Bryant, unpublished). The *pscC* gene encodes an essential, membrane-associated *c*-type cytochrome, known as Cyt c_Z (also known as Cyt c_{551}), which donates electrons to the oxidized special pair, $P840^+$ (Okkels et al., 1992; Higuchi et al., 2009) (The special pair of the reaction center is located near the periplasmic surface of the reaction center.). Multi-flash optical kinetic studies suggest that there are two PscC/Cyt c_Z cytochromes per PscA homodimer (Oh-Oka et al., 1995, 1997, 1998). PscD is a non-essential, 16-kDa protein that is bound to the stromal surface of the reaction center (Hager-Braun et al., 1995). A *Cba. tepidum* mutant lacking PscD had less efficient energy transfer from BChl *c* to reaction centers and also exhibited slightly less ferredoxin-mediated $NADP^+$ reduction than the wild type (Tsukatani et al., 2004). Orthologs of the *pscC* and *pscD* genes are not found in the genomes of any other chlorophototrophs.

PscC/Cyt c_Z acts as a shuttle for electrons derived from sulfide between the menaquinol:Cyt *c* oxidoreductase and $P840^+$, but PscC/Cyt c_Z can also receive electrons derived from thiosulfate oxidation and transfer them to $P840^+$ (Oh-Oka et al., 1998; Tsukatani et al., 2008; Azai et al., 2009). During the oxidation of sulfide by sulfide:quinone oxidoreductase (SQR), another small membrane-associated, mono-heme *c*-type Cyt (Cyt c_{556}) with one transmembrane α-helix, apparently shuttles electrons derived from the oxidation of menaquinol between the Cyt *b*/Rieske complex and PscC/Cyt c_Z (Tsukatani et al., 2008; Azai et al., 2009). Cyt c_{556} is probably the product of ORF CT0073 in the *Cba. tepidum* genome. When *Cba. tepidum* cells are oxidizing thiosulfate, an alternative electron transfer pathway, which utilizes a soluble, periplasmic *c*-Cyt, Cyt $c_{554/555}$, appears to participate in the electron transfer chain. Cyt $c_{554/555}$ is the product of ORF CT0075 in *Cba. tepidum*, and it donates electrons directly to Cyt c_Z and does not act as a soluble shuttle between Cyt c_{556} and Cyt c_Z (Tsukatani et al., 2008; Azai et al., 2009). The predicted monoheme-binding domains of CT0073 and CT0075 are very similar in amino acid sequence, and both genes are found in all GSB strains except *Chp. thalassium*, even though only a few of these strains can oxidize thiosulfate. The *Chp. thalassium* genome does encode a membrane-associated, *c*-type Cyt that is similar to that encoded by CT0073 of *Cba. tepidum*; however, this genome does not encode any small, periplasmic *c*-type Cyt equivalent to that encoded by CT0075. *Chp. thalassium* ORF Ctha_1874 encodes a protein with strong sequence similarity to both PscC and Cyt $c_{554/555}$. This protein appears to be the result of a gene fusion event: the N-terminus is similar to PscC (Cyt c_Z) and the C-terminal domain is similar to Cyt $c_{554/555}$. Whether the C-terminal monoheme *c* domain remains covalently bound to PscC or is released by proteolysis after export to the periplasm is unknown.

The remainder of the light-harvesting apparatus in members of the *Chlorobiales* is formed by

the FMO protein and chlorosomes. A high-resolution, 1.3-Å X-ray structure of the FMO protein of *Ptc. aestuarii* 2K has recently been determined, and this structure confirmed the sub-stoichiometric presence of an eighth BChl *a* molecule (Tronrud et al., 2009). This additional BChl *a* molecule also occurs in the FMO protein of *Cba. tepidum* (Ben Shem et al., 2004), and a rationale to explain the spectroscopic differences of various FMO proteins on the basis of the mode of ligation of this extra BChl *a* molecule to the FmoA apoprotein was recently proposed (Tronrud et al., 2009). Chemical labeling and mass spectrometry were used to map the orientation of the FMO protein relative to the reaction centers and chlorosomes (Wen et al., 2009), and this information places the additional BChl *a* binding site near the chlorosome baseplate (*i.e.*, near CsmA). Whether this newly discovered binding site for BChl *a* is fully occupied in the FMO protein *in vivo* is unknown, but irrespective of its occupancy, this eighth BChl molecule could play an important role in transferring energy from the chlorosome baseplate to the reaction center. Another interesting finding by Hohmann-Marriott and Blankenship (2007) is that GSB exhibit variable fluorescence. The mechanism responsible for this modulation of energy transfer efficiency is not yet known, but it was suggested that this regulation occurred at the level of FMO.

Because chlorosome structure, function, and biogenesis have recently been reviewed (Frigaard and Bryant, 2006; Hohmann-Marriott and Blankenship, 2007; Oostergetel et al., 2010), only a few recent findings will be mentioned here. Cryo-electron microscopy (Cryo-EM) and X-ray diffraction studies by Pšenčík and coworkers strongly suggested that the BChl *c* and *e* in chlorosomes are arranged in lamellae with an interlamellar spacing of ~2.1 nm (Arellano et al., 2008; Pšenčík et al., 2004, 2006, 2009; Ikonen et al., 2007). Other cryo-EM images (Oostergetel et al., 2007; Ganapathy et al., 2009) of *Cba. tepidum* chlorosomes showed that the BChl *c* molecules are arranged in multi-layered, coaxial nanotubes. Importantly, from end-on views of wild-type chlorosomes it could be seen that there are multiple BChl nanotube domains in a single chlorosome and that no two chlorosomes have the same arrangement (Oostergetel et al., 2007). These results are in good agreement with observations of Staehelin et al. (1980) and more recent results of Saga and Tamiaki (2006) by freeze-fracture electron microscopy. The latter authors observed that chlorosomes contain bundles of rod-shaped, nanotubular structures with diameters of ~10 nm for cells synthesizing either BChl *c* or BChl *d*. By combining results from cryo-EM, solid-state magic-angle-spinning NMR, and computational modeling for chlorosomes from a *bchQ bchR bchU* mutant of *Cba. tepidum* (Gomez Maqueo Chew et al., 2007), the structure of the [8-ethyl, 12-methyl]-BChl d_F molecules in the mutant chlorosomes could be determined (Ganapathy et al., 2009; Oostergetel et al., 2010). The BChl *d* molecules are arranged in a previously unknown packing mode: *syn-anti* monomer stacks with the hydrophobic tails of the two BChls pointing outward in opposite directions (Ganapathy et al., 2009). The *syn-anti* stacks of BChl *d* are arranged in rings with a shallow helical pitch around the central axis of the nanotube, with several layers of BChl *c* or *d* forming coaxial cylinders of increasing diameter. The chlorosomes of wild-type *Cba. tepidum* form similar nanotubular structures (Oostergetel et al., 2007), but the *syn-anti* monomer stacks have a different orientation relative to the long axis of the chlorosome (Ganapathy et al., 2009). This difference in monomer stack orientation may occur because the BChl *c* molecules found in the wild-type chlorosomes have a much greater proportion of *S*- versus *R*-stereochemistry at the C-3^1 chiral carbon and much bulkier side chains at the C-8^2 and C-12^1 carbon atoms (Ganapathy et al., 2009).

Protein co-fractionation studies using immunological detection methods and targeted gene inactivation studies show that the chlorosome envelopes of the best-studied model organism, *Cba. tepidum*, contain ten proteins that belong to four structural motif families: CsmA/CsmE, CsmB/CsmF, CsmC/CsmD/CsmH and CsmI/CsmJ/CsmX (Bryant et al., 2002; Vassilieva et al., 2002; Frigaard et al., 2004a, b; Frigaard and Bryant, 2006). Crosslinking studies showed that CsmA, which most likely binds a single BChl *a* molecule and 1–2 carotenoid molecules (Bryant et al., 2002), forms paracrystalline arrays in the chlorosome baseplate (Li et al., 2006). CsmA is an essential gene in *Cba. tepidum*, and all attempts to construct mutants lacking CsmA have failed. Recently, the solution structure of the

CsmA apoprotein was determined by NMR, and reconstitution of the CsmA apoprotein with BChl *a* was achieved (Pedersen et al., 2008a, b). Cross-linking studies as well as surface plasmon resonance studies have shown that CsmA interacts directly with FmoA (Li et al., 2006; Pedersen et al., 2007). Protease susceptibility studies, coupled with mass spectrometry, show that the more hydrophobic N-terminal region of CsmA is protected from proteolysis in the chlorosome baseplate (Milks et al., 2005). All of these data have been combined to produce a model for the packing of CsmA in chlorosome baseplates (Pedersen et al., 2008b, 2010). Lipids and wax esters were identified and quantified in the chlorosomes of *Cba. tepidum*, and the conclusion is that only a small portion of the chlorosome surface would be covered by lipids and that most of the chlorosome surface is therefore covered by proteins (Sørensen et al., 2008).

The roles of the other nine proteins in the chlorosome envelope have been more difficult to establish. At least one Fe/S protein is universally present in the envelopes of chlorosomes from *Chlorobiales*, *Chloroflexales,* and *Acidobacteria* strains. These Fe/S proteins have N-terminal, adrenodoxin-like domains that are predicted or have been experimentally verified to harbor a [2Fe-2S] cluster (Vassilieva et al., 2001, 2002; see below). Studies with isolated chlorosomes and cells of mutants lacking these Fe/S proteins have shown that these proteins play important roles in the reduction of the quinone energy transfer quencher after chlorosomes or cells have been exposed to oxygen (Frigaard and Matsuura, 1999). Thus, these Fe/S proteins participate directly in reactivating energy transfer from BChl *c* to reaction centers after oxygen exposure (Li, 2006; H. Li and D. A. Bryant, manuscript in preparation).

To account for the failure to detect phenotypes for some of single mutants eliminating chlorosome envelope proteins, Frigaard et al. (2004a) suggested that the proteins of the CsmC/CsmD and CsmB/CsmF motif classes might have redundant functions. To test this hypothesis, multi-locus mutants completely lacking proteins of a motif class (*e.g.*, *csmC csmD csmH* and *csmB csmF csmH* triple mutants) were constructed (Li and Bryant, 2009). These mutants can still assemble functional chlorosomes, but their size, shape (*i.e.*, length:width ratio) and biogenesis was affected. Mutants lacking proteins of the CsmC/CsmD motif family produced chlorosomes that were much smaller than those of the wild type, and the Q_y absorbance maximum for the BChl *c* aggregates in these chlorosomes was strongly blue-shifted. Conversely, the chlorosomes of mutants lacking proteins of the CsmB/CsmF motif family were much larger than wild-type chlorosomes, and the Q_y absorption for their BChl *c* aggregates was slightly red-shifted (Li and Bryant, 2009).

From a comparative genomic perspective, only three (CsmA, CsmC, and CsmX) of the ten chlorosome envelope proteins found in *Cba. tepidum* are universally encoded in the genomes of *Chlorobiales* strains. However, the *csmJ* gene is found in most genomes and is only missing in two strains, *Chl. chlorochromatii* and possibly *Cba. limnaeum*; three genes (*csmB*, *csmE*, and *csmH*) are missing in *Chp. thalassium*. However, it should also be noted that CsmH is highly variable in both sequence and length. The *csmD* gene only is only found in three strains (*Cba. tepidum*, *Cba. parvum*, and *Ptc. aestuarii* DSM 271), and thus CsmD may have arisen from a recent gene duplication event and lateral gene transfer. The *csmF* gene is found in all of the *Chlorobiales* genomes except those of *Chp. thalassium*, *Chl. chlorochromatii*, *Chl. luteolum*, and *Chl. phaeovibrioides* DSM 265. The *csmF* gene might have arisen by duplication of *csmB* after the ancestor of strains related to *Chp. thalassium* diverged from the lineage leading to other *Chlorobiales* strains, and *csmF* may have subsequently been lost from the genomes of the other three strains, which have relatively small genomes. *Cba. limnaeum* and *Chl. bathyomarinum* appear to have recent duplications of *csmE,* and the former organism has a duplication of *csmB* as well. Finally, *csmI, csmJ, and csmX* clearly arose by two duplication events, and *csmI* appears to be missing from the genomes of *Chp. thalassium*, *Chl. phaeovibrioides*, and *Chl. ferrooxidans*. Collectively, these observations strongly suggest that lineage-specific (and perhaps even strain-specific) gene duplications have occurred, which have significantly increased the complexity of the complement of chlorosome envelope proteins in some but not all members of the *Chlorobiales*. Given the mosaic distribution of these genes and the abundant evidence for horizontal gene transfer identified for other functions such as sulfur oxidation (see Section II.D), it is

likely that lateral gene transfer has also played a role in the current distribution of these genes among the *Chlorobiales*. The gene distribution data are in excellent agreement with the hypothesis that the proteins of the major chlorosome motif families in *Cba. tepidum* are likely to play redundant roles in chlorosome biogenesis and function as proposed (Frigaard et al., 2004a; Li and Bryant, 2009).

All *Chlorobiales* strains fix CO_2 by the reverse TCA cycle (Buchanan and Arnon, 1990; Sirevåg, 1995), a pathway for carbon fixation that has also recently been shown to function in sulfide-oxidizing *Epsilonproteobacteria* (Hügler et al., 2005). Three enzymes are highly diagnostic for this pathway: ATP citrate lyase (EC 2.3.3.8), 2-oxoglutarate:ferredoxin oxidoreductase (EC 1.2.7.3), and pyruvate:ferredoxin oxidoreductase (EC 1.2.7.1). Two of the three carboxylation reactions require reductants that are much stronger than pyridine nucleotides; these reductants are provided by abundant, water-soluble bacterial ferredoxins, which harbor two [4Fe-4S] clusters (Yoon et al., 2001). *In vitro* studies with pyruvate:ferredoxin oxidoreductase from *Cba. tepidum* showed that, when functioning in the decarboxylative direction of acetyl-CoA synthesis from pyruvate, this enzyme uses rubredoxin ($E_0' = -87$ mV) as the electron acceptor (Yoon et al., 1999). The *Cba. tepidum* genome encodes three rubredoxins (CT1100, CT1101, and CT2024), two of which are adjacent on the genome and which are apparently the result of a recent duplication. Most other *Chlorobiales* genomes encode two rubredoxins. An exception is the genome of *Chp. thalassium*, which lacks the genes for both rubredoxin and the oxygen-reducing protein rubredoxin:oxygen oxidoreductase (see Section II.E below).

Reduced rubredoxin is unable to provide electrons to pyruvate:ferredoxin oxidoreductase when the enzyme was operating in the carboxylative direction as a pyruvate synthase (Yoon et al., 1999). Under these conditions two distinct ferredoxins, denoted Fd I and Fd II with midpoint potentials of −514 and −584 mV, respectively, drive this reaction *in vitro* (Yoon et al., 2001). The *Cba. tepidum* genome encodes seven putative ferredoxins: two are predicted to bind one [2Fe-2S] cluster, and five are predicted to bind two [4Fe-4S] clusters (Eisen et al., 2002). One of the [2Fe-2S] proteins (CT1541) is clustered with *nif* genes in all sequenced *Chlorobiales* strains, and this ferredoxin probably plays a role in nitrogenase biogenesis, or it may act as an electron donor during dinitrogen reduction. Three of the latter group of ferredoxin genes (*CT1260*, *CT1261*, and *CT1736*) encode small, acidic proteins of 62 amino acids, which are closely related in sequence. The products of all three genes are expressed in *Cba. tepidum*, and each can function in photoreduction of $NADP^+$ in the presence of purified reaction centers (Seo et al., 2001). Because all GSB strains have this small multi-gene family, the gene duplication events that led to its formation probably occurred prior to the divergence of *Chp. thalassium* from other *Chlorobiales* strains. The ferredoxin:$NADP^+$ oxidoreductase (FNR) of *Cba. tepidum* has also been isolated and characterized. This enzyme, which is the product of ORF *CT1512*, is more similar to thioredoxin reductase than to the FNRs of plants and cyanobacteria (Seo and Sakurai, 2002). The enzyme has recently been crystallized, and diffraction data have been collected to 2.4 Å, but structural details are not yet available (Muraki et al., 2008). The gene encoding this enzyme occurs in a similar gene context in all sequenced *Chlorobiales* strains.

The principal pathway for electron transfer in sulfide- or thiosulfate-oxidizing GSB strains occurs from thiosulfate or sulfide oxidation to strongly reducing, soluble [8Fe-8S] ferredoxins in the cytoplasm (Yoon et al., 2001; Seo et al., 2001). Most of the electrons derived from sulfur oxidation (see Section II.D) are expected to enter the electron transport chain at the level of menaquinone/menaquinol, although as noted above, thiosulfate electrons may be transferred directly to Cyt c_Z by soluble Cyt $c_{554/555}$. To produce the strong reductants required for the carboxylation reactions of the reverse TCA cycle, the electrons derived from sulfide or thiosulfate must be driven to much lower redox potentials, which is the principal function of light and the type-1 reaction centers. However, there must also be a mechanism to return electrons from reduced ferredoxin to the menaquinone pool, or directly to the menaquinol:Cyt *c* oxidoreductase, to allow for the production of ATP through cyclic photophosphorylation. The pathway(s) by which electrons return to $P840^+$ is presently unknown, but the mechanism must differ from that in cyanobacteria

and plants, which probably return electrons to the quinone pool via heme c_n bound to the Cyt b_6 subunit of the plastoquinol:Cyt *c* oxidoreductase (Baniulis et al., 2008). Electrons in that case are subsequently returned to $P700^+$ via Cyt *f* (Iwai et al., 2010). The PetB sequences of *Chlorobiales* strains do not have the Cys residues that participate in ligating heme c_n in the PetB proteins of cyanobacteria, algae and plants.

All sequenced *Chlorobiales* genomes encode a cluster of 11 *nuo/ndh* genes that encode the subunits of a type-1, NADH dehydrogenase (similar to Complex I). This oxidoreductase might play a role in cyclic electron transfer by accepting electrons from reduced ferredoxin, as has been suggested for cyanobacteria (Yu et al., 1993; Mi et al., 1995). As noted previously (Frigaard et al., 2003), the NuoE, NuoF, and NuoG subunits are not encoded by the *Cba. tepidum* genome, and this is also true for all other *Chlorobiales* strains except *Chl. phaeobacteroides* BS-1. Therefore, this oxidoreductase is unlikely to oxidize or reduce pyridine nucleotides. In some but not all strains, an uptake hydrogenase operon occurs immediately downstream from the 11-gene *ndh/nuo* operon. It is not known whether the proximity of these genes implies a functional relationship in hydrogen metabolism or reverse or cyclic electron transport.

The menaquinol:Cyt *c* oxidoreductase in GSB has proven difficult to isolate and characterize, and considerable controversy still surrounds its precise subunit composition and cofactor content. All *Chlorobiales* strains contain a dicistronic operon containing the *petC* and *petB* genes, encoding a Rieske iron-sulfur protein and the membrane-intrinsic Cyt *b*, respectively. However, no genes for *c*-type cytochromes are linked to any of these *petCB* operons. As noted above, it is possible that orthologs of CT0073, which encode the membrane-bound Cyt, Cyt c_{556} of *Cba. tepidum*, participate in electron transfer from menaquinol to PscC/Cyt c_Z in reaction centers (Tsukatani et al., 2008).

Eight of the sequenced *Chlorobiales* strains in Table 3.1 have an RnfABCDGE complex (*e.g.*, Paes_1309–1314), which has been proposed in *Acetobacterium woodii* to couple exergonic electron transfer from ferredoxin to NAD^+ to vectoral transport of Na^+ across the cytoplasmic membrane (Schmidt et al., 2009). Consistent with this possibility, several of the strains that have this electron transfer complex (*Ptc. aestuarii*, *Chl. bathyomarinum*, *Pld. phaeum*, *Chl. phaeobacteroides* BS-1, and *Chp. thalassium*) are marine strains. Assuming that the properties ascribed to this potential energy-conserving oxidoreductase are similar to those of other bacteria, this complex could couple electrons derived from hydrogen or reduced sulfur compounds oxidized in the light to the formation of a sodium- or proton-motive force. Further suggesting that sodium ion gradients might be important in some *Chlorobiales* strains, four strains (*Ptc. aestuarii*, *Chl. bathyomarinum*, *Pld. phaeum*, and *Chl. phaeobacteroides* BS-1) also have a six-gene operon (*nqrABCDEF*; *e.g.*, Paes_2133–2,138) encoding a Na^+-transporting NADH:quinone oxidoreductase. This enzyme couples sodium export to the periplasm to oxidation of NADH and reduction of a quinone (Steuber, 2001). These same four strains, together with *Cba. parvum*, *Cba. limnaeum*, *Chl. limicola*, and *Chl. phaeovibrioides*, have an operon encoding a multisubunit Na^+:H^+ Mnh/Mrp antiporter (*mnhABCDEFG*; *e.g.*, Paes_0094–0100), which also has some similarity to the membrane-intrinsic portion of the type-1 NADH dehydrogenase (Swartz et al., 2005). *Chlorobiales* genomes do not appear to encode Na^+-dependent ATP synthases. The occurrence of these Na^+-dependent processes is generally well correlated with the requirement for elevated salt (NaCl) in the growth media for these strains (see Imhoff, 2003). Together with a menaquinol:Cyt oxidoreductase and the type-1 reaction center, these three complexes provide two coupling sites for the transfer of ferredoxin electrons to the quinone pool, with the formation of a sodium-motive force, which could be converted into a protonmotive force by the Mnh/Mrp Na^+:H^+ antiporter.

2. Chlorophyll Biosynthesis

All GSB strains synthesize three types of (B) Chls: BChl *c*, *d*, or *e*, which is found in chlorosomes; BChl *a*, which is found in small amounts in chlorosomes (bound to CsmA), in the FMO protein, and as the principal pigment of the type-1 reaction centers; and Chl *a*, which is exclusively found in the reaction centers (Frigaard and Bryant, 2006; Gomez Maqueo

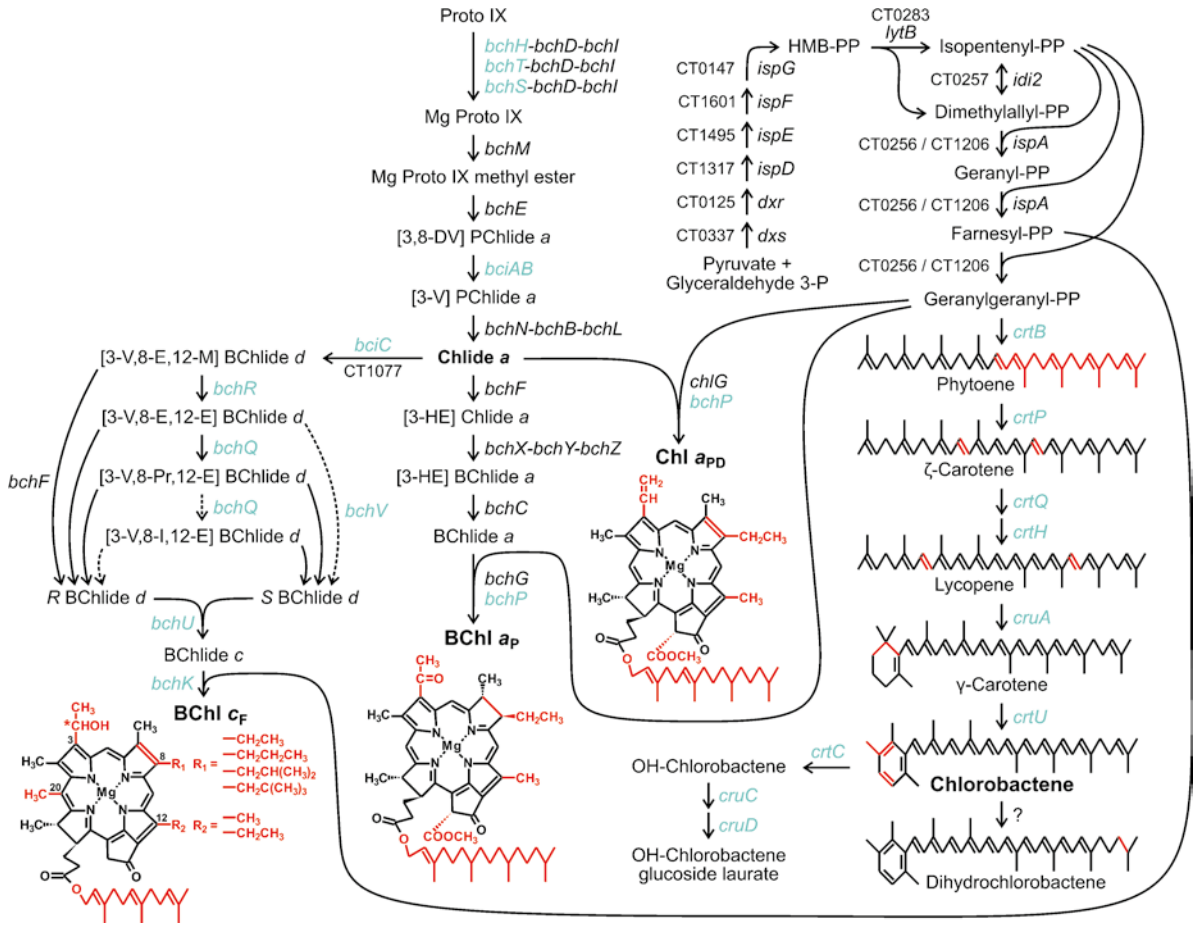


Fig. 3.3. Biosynthetic pathways for BChl *c*, BChl *a*, and Chl *a* and isoprenoid precursors and carotenoids. Chemical moieties indicated in red illustrate key structural differences between adjacent compounds or end products. Gene loci indicated in blue indicate that null mutant strains have been produced and biochemically and physiologically characterized. Subscripts identify the principal esterifying alcohol: P, phytol; PD, Δ2,6-phytadienol; and F, farnesol. Thus, BChl a_p indicates Bchl *a* esterified with phytol.

Chew and Bryant 2007a). Prior to the completion of the *Cba. tepidum* genome sequence, little was known about the biosynthetic pathway for BChl *a* and Chl *a* biosynthesis in GSB, and essentially the only thing that was known about the synthesis of BChl *c* was that the methyl groups added to the C-8^2, C-12^1, and C-20 carbons were derived from S-adenosyl-L-methionine (Huster and Smith, 1990). Figure 3.3 shows the branched pathway leading from protoporphyrin (Proto) IX to Chl *a*, BChl *a*, and BChl *c*, as it is currently understood. Because detailed reviews of BChl synthesis in bacteria have been published recently (Willows, 2003; Bollivar, 2006; Frigaard et al., 2006; Gomez Maqueo Chew and Bryant, 2007a; Tanaka and Tanaka, 2006, 2007; Willows and Kriegel, 2009), this section will focus on recent results.

As shown in Fig. 3.3, orthologs of most of the genes that are known to be required for BChl *a* synthesis in *Rhodobacter sphaeroides* and *Rhodobacter capsulatus* also occur in the *Cba. tepidum* genome as well as those of the 14 other *Chlorobiales* strains in Table 3.1 (Frigaard et al., 2003, 2004a, 2006; Gomez Maqueo Chew and Bryant, 2007a, b; Gomez Maqueo Chew et al., 2007, 2009). These include the genes for Proto IX Mg-chelatase (*bchH/bchS/bchT, bchI, bchD;* see below); SAM:Mg-Proto IX methyltransferase (*bchM*); Mg-Proto IX monomethylester oxygen-independent oxidative ring cyclase (*bchE*); 8-vinyl-protochlorophyllide (PChlide)

reductase (*bciA* or *bciB*; see below); PChlide reductase (*bchB, bchL, bchN*); Chl *a* synthase (*chlG*); chlorophyllide *a* (Chlide) reductase (*bchX, bchY, bchZ*); 3-vinyl-BChlide hydratase (*bchF*); 3-hydroxyethyl BChlide *a* dehydrogenase (*bchC*); and BChl *a* synthase (*bchG*). Additionally, as will be described below, an ortholog of *bchJ,* which was erroneously believed to encode 8-vinyl-PChlide reductase (Suzuki and Bauer, 1995), is also present. However, BchJ has no apparent enzymatic activity (Gomez Maqueo Chew and Bryant, 2007b), and this protein may function as a tetrapyrrole-binding protein and/or could act as a regulatory protein, similar to Gun4 of plants and cyanobacteria (Davison et al., 2005; Verdecia et al., 2005; Sobotka et al., 2008; Adhikari et al., 2009).

The first committed step in the biosynthesis of (B)Chl is the insertion of Mg^{2+} into Proto IX by Mg-chelatase. In all (B)Chl-synthesizing organisms, this enzyme has three types of subunits: BchH/ChlH, BchD/ChlD, and BchI/ChlI (Tanaka and Tanaka, 2007; Gomez Maqueo Chew and Bryant, 2007b; Gomez Maqueo Chew et al., 2009; Willows and Kriegel, 2009). *Cba. tepidum* has three paralogous genes for the largest, BchH-type subunit, which have been named *bchH*, *bchS*, and *bchT*, but it only has single copies of the genes for *bchD* and *bchI* (Eisen et al., 2002; Frigaard et al., 2003; Gomez Maqueo Chew et al., 2009). A recent study showed that each of these large subunits can be combined with BchD and BchI to form active enzymes, but the specific activities of the resulting Mg-chelatases *in vitro* varied over five orders of magnitude (Johnson and Schmidt-Dannert, 2008). The different large subunits of Mg-chelatase also affected the activity of BchM, which catalyzes the next step in the biosynthetic pathway, in different ways and to different extents; however, BchM had no effect on the activity of any of the forms of Mg-chelatase (Johnson and Schmidt-Dannert, 2008). Mutants of *Cba. tepidum* lacking only BchH or BchT had relatively little phenotype, but a mutant lacking BchS produced much less BChl *c* than the wild type (Gomez Maqueo Chew et al., 2009). Furthermore, double mutants that could only synthesize BchS or BchH were viable, but it was not possible to isolate mutants in which both *bchS* and *bchH* had been inactivated (*i.e.,* when BchT was the only large subunit in the cells). These results suggest that the Mg-chelatase activity of the enzyme formed from the BchT subunit is unable to meet the minimum Mg-chelatase activity required for cell viability. This result agrees with the enzymological results, which showed that the Mg-chelatase formed from BchT had the lowest activity *in vitro* (Johnson and Schmidt-Dannert, 2008). A *bchS* mutant, which only produced ~10% of the BChl *c* of the wild type, excreted very large amounts of Proto IX into the growth medium (Gomez Maqueo Chew et al., 2009). The mutational studies supported the hypothesis that these three BchH homologs function in end-product regulation and/or substrate channeling of intermediates in the trunk (shared) portion of the branched BChl biosynthesis pathway. Genomic comparisons show that nearly all *Chlorobiales* strains listed in Table 3.1 have all three *bchH* homologs. The only exception is *Chl. chlorochromatii*, which does not have a *bchH* homolog (Gomez Maqueo Chew et al., 2009).

Mutational studies showed that *Cba. tepidum* has an 8-vinyl PChlide reductase that is highly similar to the enzyme found in higher plants, green algae, and some marine cyanobacteria (Nagata et al., 2005, 2007; Gomez Maqueo Chew and Bryant, 2007b). Plants apparently use 3,8-divinyl-Chlide *a* as the principal substrate for Chl biosynthesis (Nagata et al., 2007). However, *in vitro* studies with recombinant *Cba. tepidum* BciA produced in *Escherichia coli* have shown that this enzyme is highly active with 3,8-divinyl-PChlide and NADPH and that it uses an ordered sequential substrate binding mechanism (Gomez Maqueo Chew and Bryant, 2007b; F. Shen and D. A. Bryant, unpublished results). Furthermore, characterization of *bciA* and *bchJ* mutants of *Cba. tepidum* have clearly established that BchJ has no 3,8-divinyl PChlide reductase activity (Gomez Maqueo Chew and Bryant, 2007b); however, the absence of BchJ causes cells to excrete very large amounts of 3,8-divinyl-PChlide into the growth medium. Conveniently, this material can be isolated, purified, and used as a substrate for *in vitro* characterization of BciA, which is the actual 3,8-divinyl PChlide reductase (Gomez Maqueo Chew and Bryant, 2007b; F. Shen and D. A. Bryant, unpublished).

Comparative analyses show that 11 other *Chlorobiales* strains listed in Table 3.1 have the *bciA* gene. However, the genomes of three com-

pletely sequenced strains (*Chl. limicola*, *Chl. phaeobacteroides* DSM 266, and *Chp. thalassium*) do not encode any gene similar to *bciA*. Each of these three strains, and most but not all *Chloroflexi* strains (see Section III.C.2), instead have an 8-vinyl-PChlide reductase that is homologous to the product of ORF *slr1923* of the cyanobacterium *Synechocystis* sp. PCC 6803 (Islam et al., 2008; Ito et al., 2008). This gene is usually annotated as coenzyme F420 hydrogenase/dehydrogenase, beta subunit. Null mutants of *slr1923* in *Synechocystis* sp. PCC 6803 are viable, are unable to reduce the 8-vinyl side chain of 3,8-divinyl-PChlide, and thus accumulate 3,8-divinyl-Chl *a*. Although it was known that the *slr1923* product was necessary for this reaction, it was not known if the *slr1923* product was sufficient (Islam et al., 2008; Ito et al., 2008). When the *bciA* gene of *Cba. tepidum* was replaced by the unique *slr1923* homolog (Ctha_1208) from the *Chp. thalassium* genome, the resulting transformants retained their ability to synthesize BChl *c* (Liu and Bryant, 2011). This result establishes that the product of this *slr1923* ortholog, which we here designate as BciB, is sufficient to reduce the 8-vinyl group of 3,8-divinyl-PChlide (*i.e.,* no additional protein subunits or cofactors other than those normally present in *Cba. tepidum* are additionally required for the activity). Four *Chlorobiales* strains (*Chl. ferrooxidans*, *Chl. clathratiforme*, *Ptc. aestuarii*, and *Prosthecochloris* sp. Ty Vent) putatively encode both BciA- and BciB-types of 8-vinyl reductases, and the *bciB* homologs in *Ptc. aestuarii*, *Chl. phaeobacteroides* and *Chl. limicola* could also complement a *bciA* deletion mutant of *Cba. tepidum* (Liu and Bryant, 2011). *Chloroflexus* spp. strains have homologs of BciB but not BciA. Although *Roseiflexus* spp. synthesize BChl *a*, their genomes do not encode homologs of BciA or BciB. This observation implies that *Roseiflexus* spp. must have a third type of 8-vinyl PChlide reductase (see Section III.C.2).

In anoxygenic chlorophototrophs, the C-17/C-18 double bond of the D-ring of PChlide is stereospecifically reduced to form Chlide *a* by the dark-operative PChlide oxidoreductase (DPOR), an enzyme comprising three subunits: BchL, BchB, and BchN (Coomber et al., 1990; Burke et al., 1993b; Bollivar et al., 1994; Fujita and Bauer, 2000). Similarly, Chlide *a* oxidoreductase (COR) comprises three subunits (BchX, BchY, and BchZ), and this enzyme stereospecifically reduces the C-7/C-8 double bond in the B-ring of Chlide *a* to form 3-vinyl-BChlide *a* (Burke et al., 1993a; Nomata et al., 2006). These two oxygen-sensitive enzymes are distantly related to nitrogenase (NifH, NifD, and NifK) (Burke et al., 1993a, b; Fujita and Bauer, 2000; Nomata et al., 2005). Most of the earlier studies on these enzymes were performed with proteins derived from *Rhodobacter capsulatus*, and it was found that BchN and BchB form a heterotetramer that carries two oxygen-tolerant [4Fe-4S] clusters as well as the PChlide-binding active site (Fujita and Bauer, 2000; Nomata et al., 2005, 2006, 2008). Reduction of PChlide to Chlide *a* required the addition of reduced BchL and Mg-ATP. Similar studies with Chlide *a* oxidoreductase showed that BchYZ and BchX can be functionally reconstituted to produce an active enzyme capable of reducing Chlide *a* to 3-vinyl-BChlide *a* in the presence of dithionite and ATP (Nomata et al., 2006).

The PChlide oxidoreductase subunits from *Cba. tepidum* have also been heterologously expressed, purified, and reconstituted to produce an active enzyme (Bröcker et al., 2008a). Site-directed mutagenesis studies with this enzyme suggested that three Cys residues from BchN (Cys 21, 46, and 103) and one from BchB (Cys 94) coordinate an intersubunit [4Fe-4S] cluster. No Mo-containing cofactor was present, and other site-directed mutagenesis studies identified residues potentially important in making protein-protein contacts (Bröcker et al., 2008a; Wätzlich et al., 2009). Catalytically active DPOR could be produced with subunits derived from both oxygenic and anoxygenic chlorophototrophs, even when BchL was replaced with BchX (Wätzlich et al., 2009). The substrate recognition properties of one member of this family of enzymes were also studied (Bröcker et al., 2008b). These studies show that the DPOR and COR enzyme components are functionally and structurally very highly conserved, but specific protein contact surfaces probably have evolved to prevent chimera formation in nitrogen-fixing chlorophototrophs like *Cba. tepidum*, which could have up to three related proteins expressed at one time. This conclusion is strongly supported by recent X-ray

crystallographic studies, in which the structure of the *R. capsulatus* BchL dimer with bound MgADP has been solved at a resolution of 1.6 Å (Sarma et al., 2008). The BchL dimer has a structure very similar to that of the NifH dimer, but the known or predicted surface charges in the areas where these two reductase proteins interact with NifD/NifK or BchB/BchN differ markedly. Finally, the structure of the *R. capsulatus* BchN/BchB heterotetramer has recently been solved at a resolution of 2.3 Å (Muraki et al., 2008, 2010) and that of *Thermosynechococcus elongatus* BP-1 at a resolution of 2.4 Å (Bröcker et al., 2010). Each BchN/BchB dimer binds one PChlide without axial ligation, and each dimer additionally ligates one [4Fe-4S] cluster through three Cys residues and one Asp residue.

Based upon an analysis of genes in *Cba. tepidum,* which are similar to those involved in BChl *a* biosynthesis in *Rhodobacter* spp. (Coomber et al., 1990; Bollivar et al., 1994; Suzuki et al., 1997) and their paralogs, possible pathways for the synthesis of BChl *c* were proposed (Eisen et al., 2002; Frigaard et al., 2003). Luckily, *Cba. tepidum* is genetically amenable and BChl *c* is not essential for viability (Frigaard et al., 2002a), because for the most part these "hypotheses" generated *in silico* proved to be incorrect when rigorously tested by the reality of biochemical genetics. Figure 3.3 indicates the seven genes (*bciC*, *bchR*, *bchQ*, *bchF*, *bchV*, *bchU* and *bchK*) that encode the enzymes that participate in the synthesis of BChl *c*. These seven genes occur in all GSB genomes, although the *bchU* gene in *Cba. parvum* naturally carries a frameshift mutation, which causes the "wild-type" strain to synthesize BChl *d* rather than BChl *c* (Maresca et al., 2004). A key observation was that only small amounts of BChl *c* accumulated in a *bchJ* mutant, and that this strain correspondingly excreted large amounts of 3,8-divinyl-PChlide into the growth medium (Gomez Maqueo Chew and Bryant 2007a, b). This observation is most consistent with only one of the three previously postulated pathways (Fig. 3.3).

In the current scheme, the pathways leading to the three principal products diverge from the hub intermediate, Chlide *a*. The first committed step in BChl *c* biosynthesis is the removal of the methylcarboxyl group at C-13^2 from Chlide *a* (Fig. 3.3). Studies of Chl degradation in senescent plant tissues have shown that removal of the methyl group by a methylesterase can be followed by spontaneous decarboxylation (Shioi et al., 1996; Hörtensteiner, 2006). Using phylogenetic profiling and gene neighborhood analyses together with targeted gene inactivation, repeated attempts to identify the gene for such an enzyme failed. However, when the draft genome of *Cab. thermophilum* became available (see Section IV), a gene cluster was identified that encoded several enzymes involved in tetrapyrrole and BChl biosynthesis (*bchH-hemA-bchL-bciB-bchF-bciC-bchQ-bchR*). The gene identified as *bciC* is a homolog of ORF *CT1077* in *Cba. tepidum* (Z. Liu, A. Garcia Costas, D. A. Bryant, unpublished), and the presence of this ORF in this cluster suggested that this gene might have a role in BChl *c* biosynthesis. Furthermore, this single-copy gene occurs in all GSB genomes, is present in all *Chloroflexus* spp. genomes, is missing in the *Roseiflexus* spp. and *H. aurantiacus* genomes, and is not present in any organism that does not synthesize BChl *c, d,* or *e*. When CT1077 was insertionally inactivated in *Cba. tepidum*, an orange mutant unable to synthesize BChl *c* was obtained (Liu and Bryant, 2011). This mutant could still synthesize BChl a_P and Chl a_{PD}, and interestingly, this mutant produced substantial amounts of Chl a_P, which is never observed in wild-type *Cba. tepidum* cells (The subscripts P, PD, F and S denote the principal esterifying alchohols phytol, Δ2,6-phytadienol, farnesol, and octadecanol/stearol, respectively.). All of these results are expected if CT1077 catalyzes the first committed step in BChl *c* biosynthesis beyond Chlide *a* (see Fig. 3.3). Studies to determine whether BciC is a methylesterase, or if this protein acts by some other mechanism to remove the methylcarboxyl group, have not yet established the chemistry of the reaction catalyzed by BciC (Liu and Bryant, 2011). It should be noted that the *bciC* gene is not a member of any COG group and is co-localized with other genes for BChl biosynthesis in only one of the 19 genomes in which it is found.

Biochemical characterization of null mutants was also used to identify the genes encoding the other enzymes for the remaining steps of the BChl *c* biosynthetic pathway in *Cba. tepidum* (Fig. 3.3). The first gene identified, *bchK,* encodes the BChl *c* synthase that esterifies BChlide *c* with farnesol

(Frigaard et al., 2002a). Mutants lacking BchK are unable to synthesize BChl *c*, are orange in color instead dark green, and have a doubling time seven times longer than the wild type under limiting light. The genes encoding three methyltransferases, *bchU*, *bchQ*, and *bchR*, were tentatively identified by comparative bioinformatics and validated by mutagenesis (Maresca et al., 2004; Gomez Maqueo Chew et al., 2007). Although it was first believed that BchU might act early in the BChl *c* pathway (see Frigaard et al., 2003), *in vitro* activity studies (Harada et al., 2005) and the X-ray structure of BchU (Wada et al., 2006) showed that the C-3^1 hydroxyl group is important in substrate recognition and binding. These results imply that BchU acts after BchF and BchV have hydrated the C-3^1–C-3^2 double bond. Because the stereochemistry of the chiral carbon at C-3^1 is determined by the activities of BchF and BchV (Gomez Maqueo Chew and Bryant, 2007a), and because the relative proportions of R- and S-stereoisomers of BChl *c* varies as a function of C-8^2 and C-12^1 methylation (and growth rate), the methylation reactions catalyzed by BchQ and BchR must precede hydration of the C-3^1–C-3^2 double bond (Gomez Maqueo Chew, 2007). The C-3 vinyl hydratase, BchF, is the only enzyme of the BChl *c* biosynthetic pathway that is required for cellular viability in *Cba. tepidum*. Attempts to inactivate this gene were unsuccessful (A. Gomez Maqueo Chew and D. A. Bryant, unpublished results), and this result suggests that BchV cannot substitute for BchF in the synthesis of BChl *a*. Considering that mutants lacking BchQ can still normally methylate the C-12^1 position, and that mutants lacking BchR can still normally methylate the C-8^2 position, the order of these two methylation reactions is probably not absolutely fixed (Gomez Maqueo Chew et al., 2007).

One can reasonably ask why organisms evolved the capability to extend the Chl biosynthetic pathway by three distinctive methylation reactions and what the functional consequences of these methylations might be. A *bchQ bchR bchU* triple mutant, which is unable to perform these methylation reactions, still assembles functional chlorosomes (Oostergetel et al., 2007), and the [8-ethyl, 12-methyl]-BChl *d* molecules in these chlorosomes had >95% *R*-chirality at C-3^1 (Ganapathy et al., 2009). The BChl *d* molecules in the mutant chlorosomes were much more highly ordered than those of wild-type chlorosomes, and as discussed above, this allowed the structure of these BChls to be determined by solid-state NMR (Ganapathy et al., 2009). Physiological studies have additionally shown that these methyl groups: (1) tune the absorption maximum of the aggregated BChl molecules; (2) increase the half-bandwidth of the Q_y absorption band of the BChl aggregates in chlorosomes (thus increasing the optical cross-section by inhomogeneous broadening); and (3) modulate the metabolic flux through the BChl biosynthesis pathway, perhaps by differential feedback effects on Mg-chelatase or other enzymes of the biosynthetic pathway (Gomez Maqueo Chew et al., 2007). It is important to emphasize that mutants lacking the ability to methylate BChl *d* are viable and produce functional chlorosomes; however, these mutants grow more slowly at all light intensities than the wild type (Gomez Maqueo Chew et al., 2007). The biosynthesis pathway for BChl *c* thus provides strong support for the Granick hypothesis (1957, 1965), which in one formulation states that pathways evolve as organisms evolve. The addition of three methyltransferases to a pathway that could already synthesize [8-Et, 12-Me]-BChl d_F and assemble functional chlorosomes led to a more disordered BChl structure in chlorosomes, but this was correlated with the modification of the BChl suprastructure to assemble an antenna with improved light-harvesting characteristics. In this particular case, one can directly recapitulate the evolutionary process by assessing the improvements in light harvesting that resulted from the extension of the BChl *c* biosynthetic pathway (for more information on evolutionary aspects, see Section V).

3. Carotenoid Biosynthesis

Although the ability to synthesize aromatic carotenoids is not a unique trait among chlorophototrophs (*e.g.*, *Streptomyces griseus;* see Krügel et al., 1999), the majority of *Chlorobiales* strains can synthesize aromatic carotenoids (*e.g.*, chlorobactene; Fig. 3.3) (Takaichi, 2001; Graham and Bryant, 2008; Maresca et al., 2008a). Virtually all strains that synthesize BChl *e* produce at least some isorenieratene, while strains that synthesize BChl *c* or *d* usually produce chlorobactene as a major carotenoid. Prior to the availability of the *Cba. tepidum* genome, no genes involved in carotenoid biosynthesis had been identified in any

Chlorobiales strain. As summarized in Fig. 3.3, nearly all of the genes required to produce the enzymes of carotenogenesis in *Chlorobiales* have now been identified (for a review, see Maresca et al., 2008b). The one exception is the gene product that is responsible for the reduction of the 1′,2′ double bond in 1′,2′-dihydro-chlorobactene and 1′,2′-dihydro-γ-carotene, which has not yet been identified. Attempts to demonstrate a role for the product of ORF *CT0180*, which has sequence similarity to carotenoid isomerases and some other enzymes of carotenogenesis (*e.g.*, CrtD), have not yet been successful (K. Vogl, J.E. Graham, and D.A. Bryant, unpublished).

The carotenoid biosynthetic pathway shown in Fig. 3.3 was established by targeted gene inactivation and studies of enzyme functions in *Escherichia coli* strains that express various combinations of carotenogenesis genes (Frigaard et al., 2004b; Maresca and Bryant, 2006; Maresca et al., 2007; 2008a, b). The first step in carotenogenesis is the synthesis of a C-40 compound, phytoene, which is synthesized by CrtB from two molecules of geranylgeranyl pyrophosphate. As in most cyanobacteria, algae, and plants (Sandmann, 2009), lycopene synthesis in all sequenced *Chlorobiales* strains occurs by a three-enzyme pathway that includes two dehydrogenases (CrtP and CrtQ) and one isomerase (CrtH) (Frigaard et al., 2004b). A fourth family of lycopene cyclases was first identified in *Cba. tepidum*, and this previously unknown type of lycopene cyclase was also found in many cyanobacteria (Maresca et al., 2007). Unlike cyanobacteria, which use the product of the *cruF* gene to hydrate the 1′,2′ double bond during myxoxanthophyll biosynthesis (Graham and Bryant, 2009), *Chlorobiales* strains use CrtC (homolog of neurosporene 1′,2′- hydratase) to hydrate the 1′,2′ double bond (Frigaard et al., 2004b). A type-2 glycosyl transferase (CruC) and an acyl-transferase (CruD) catalyze the terminal reactions required for synthesis of carotenoid glycosyl esters (Maresca and Bryant, 2006). All strains that synthesize isorenieratene have two paralogous enzymes, CruA and CruB, which sequentially introduce two β-rings into lycopene to produce β-carotene. The β-rings are subsequently desaturated and isomerized by CrtU (carotene desaturase/methyltransferase) to produce the φ–rings of chlorobactene or isorenieratene (Maresca et al., 2008a, b).

With only a few notable exceptions, all of the carotenogenesis genes shown in the pathway in Fig. 3.3 occur in all sequenced *Chlorobiales* strains. *Chl. chlorochromatii* synthesizes γ-carotene and derivatives, but this strain does not synthesize chlorobactene (Vogl et al., 2006), and as expected, its genome does not encode a homolog of *crtU*. Only the five sequenced strains (*i.e., Chl. clathratiforme*, *Chl. phaeobacteroides* DSM 266 *Chl. phaeobacteroides* BS-1, *Cba. limnaeum*; and *Pld. phaeum*) that are capable of synthesizing isorenieratene and BChl *e* have both CruA (lycopene cyclase) and CruB (γ-carotene cyclase) (Maresca et al., 2008a, b).

D. Oxidation of Sulfur Compounds

GSB are nearly universally dependent on reduced sulfur compounds for growth (Frigaard and Dahl, 2009; Gregersen et al., 2011; for an overview, see Fig. 3.4). Sulfide is readily oxidized by almost all strains (Imhoff, 1995; Overmann, 2001). Elemental sulfur may form as the result of incomplete sulfide and thiosulfate oxidation, and if so, it occurs as globules that are deposited outside the cells. Most strains will consume and oxidize these sulfur globules to sulfate once other sources of reduced sulfur compounds are exhausted. One notable exception is *Chp. thalassium*, which only very slowly oxidizes sulfur globules formed by sulfide oxidation (Gibson et al., 1984). Thiosulfate utilization is only observed in a few strains and tetrathionate utilization in even fewer strains (Frigaard and Dahl, 2009). *Chl. ferrooxidans* is the only described GSB that cannot grow on sulfide and only Fe^{2+} and H_2 have been described as electron donors for its growth (Heising et al., 1999). Although one or two sulfide:quinone reductases are present (see Section II.D.1 below), this strain (DSM 13031; Table 3.1) appears to have lost most genes related to thiotrophic metabolism.

1. Oxidation of Sulfide

The best-characterized, sulfide-oxidizing enzyme is sulfide:quinone oxidoreductase (SQR) (Griesbeck et al., 2000; Chan et al., 2009; Marcia et al., 2009). SQR is a flavoprotein that consists of only one subunit; it donates electrons to isoprenoid quinones and thus is associated with energy-transforming membranes (Fig. 3.4). This enzyme is widespread among both chemotrophic and

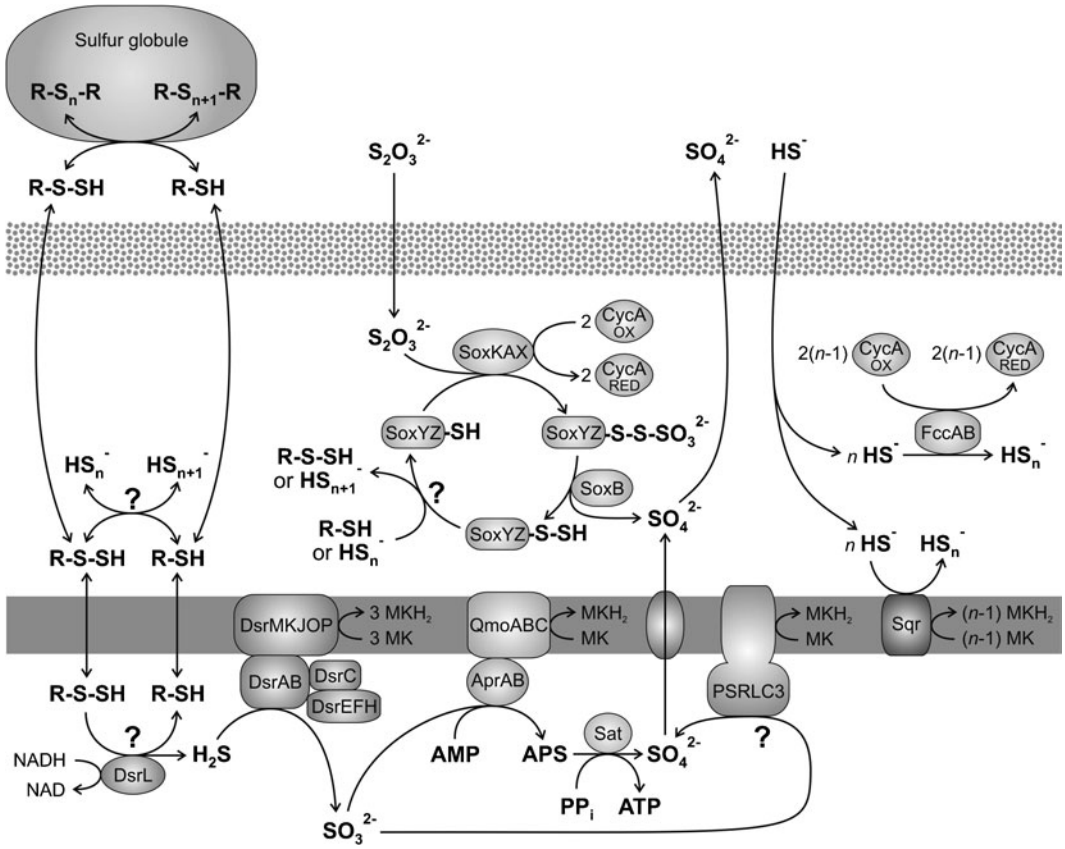


Fig. 3.4. Overview of proposed pathways of thiotrophic sulfur transformations in green sulfur bacteria (*Chlorobiales*). No single organism probably has all of the reactions shown here. In the periplasm, the polysulfur chains are probably very short (n = ~3–4), whereas the polysulfur chains in the sulfur globules can be very long (n > 3 and possibly >10^5; Frigaard and Dahl, 2009; Gregersen et al., 2011). For additional details, see text.

chlorophototrophic bacteria, and it apparently functions both in energy conservation and in detoxification. The structure of SQR from *Aquifex aeolicus* has recently been solved, and this structure provides insights into how this enzyme is partly embedded in membranes in order to interact with both hydrophobic and hydrophilic substrates (Marcia et al., 2009). Phylogenetic analyses have identified several types of distantly related SQRs (Theissen et al., 2003; Gregersen et al., 2011). All GSB have at least one type IV SQR homolog and some have additional SQR homologs. In *Cba. tepidum* the type IV SQR (CT0117/NP_661023) and a distantly related SQR (CT1087/NP_661978; type VI) have been shown through mutational analyses to constitute the SQR activity in this organism (Chan et al., 2009). Because of a gene duplication event, type IV SQRs occur in two forms, IVa and IVb, in GSB. Type IIIa SQR also commonly occurs in some *Proteobacteria*, including phototrophic sulfur bacteria, whereas type IVb occurs almost exclusively in GSB. Some strains have type IVa SQR and not type IVb SQR (*e.g.*, *Cba. tepidum*), while other strains have type IIIb SQR and not type IVa SQR (*e.g.*, *Chl. limicola* DSM 245). Still other strains have both type IVa and IVb SQRs (*e.g.*, *Chl. phaeovibrioides* DSM 265). Type VI SQR is found in *Chp. thalassium*, *Cba. tepidum*, and only a few other GSB strains. Because *Chl. ferrooxidans* cannot use sulfide for growth and its growth is in fact inhibited by low concentrations of sulfide, the two SQRs in this organism (one type IVa [ZP_01385816] and one type VI [ZP_01385814]) appear to be functionally restricted to sulfide detoxification.

2. Oxidation of Sulfur Globules

Globules of elemental sulfur are formed by incomplete oxidation of sulfide and thiosulfate by an unknown mechanism, probably involving polysulfides as intermediates (Frigaard and Dahl, 2009). The redox state of the deposited sulfur is similar to that of elemental sulfur, although the molecular structure is not known in detail. In some PSB these globules are located inside the cell in the periplasm and are encapsulated by the "sulfur globule proteins" SgpA, SgpB, and SgpC. These proteins do not occur in *Chlorobiales* or *Chloroflexales* strains, for which the sulfur globules always are located outside the cell (Fig. 3.4). Complete oxidation of the sulfur globules also occurs by an unknown mechanism, probably via sulfide formed by reductive activation of elemental sulfur. Both GSB and PSB encode an enzyme system with high sequence similarity to the dissimilatory sulfite reductase (DSR) enzyme system of sulfate-reducing bacteria. It is assumed that the DSR systems in GSB and PSB function in the oxidative direction from sulfide to sulfite. Because mutants of *Cba. tepidum* (Holkenbrink et al., 2011) and the PSB *Allochromatium vinosum* (Dahl et al., 2005, 2008) lacking *dsr* genes accumulate elemental sulfur globules when grown with sulfide or thiosulfate, it appears that DSR is essential for complete oxidation of reduced sulfur compounds to sulfate in these organisms. All GSB except *Chp. thalassium* and *Chl. ferrooxidans* have the conserved *dsrNCABLEFHTMKJOP* cluster. Interestingly, the DsrAB proteins and other extrinsic Dsr proteins are related to the corresponding Dsr proteins in PSB, whereas the membrane-intrinsic DsrMKJOP proteins are most clearly related to the corresponding Dsr proteins in sulfate-reducing bacteria. These observations suggest that the DSR system of GSB has a chimeric origin and probably arose by horizontal gene transfer events. It is possible that the "explosive" and recent radiation of the GSB was associated with the acquisition of these genes (see Section II.A), which would have conferred on the recipient organism a much more efficient metabolism due to the ability to oxidize sulfide completely to sulfate (Frigaard and Bryant, 2008b). Finally, it should be noted that, because of a recent duplication and recombination event, the *dsr* cluster in *Cba. tepidum* is split into two gene clusters (Eisen et al., 2002).

There appear to be two mutually exclusive mechanisms for the oxidation of the sulfite putatively produced by the DSR system in GSB (Frigaard and Bryant, 2008a, b). One proposed mechanism relies on the products of the gene cluster, *sat-aprBA-qmoABC* (CT0862–CT0868 in *Cba. tepidum*) (Fig. 3.4). The products of this cluster have high sequence similarity with proteins from sulfate-reducing bacteria, and they only occur in some GSB. In the proposed mechanism, sulfite would be oxidized to APS by membrane-bound adenosine-5′-phosphosulfate (APS) reductase, encoded by the *aprAB* genes. The electrons would be transferred to menaquinone by the membrane-bound QmoABC complex, and sulfate would be released from APS by the cytoplasmic ATP sulfurylase (Sat) enzyme while generating ATP by substrate-level phosphorylation. The alternative proposed mechanism for the oxidation of sulfite involves a putative molybdopterin-containing, membrane-bound complex called sulfide:quinone-reductase-like protein 3 (PSRLC3), which is encoded by Cpha266_0119 (YP_910613) and Cpha266_0120 (YP_910614) in *Chl. phaeobacteroides* (Fig. 3.4). This protein is homologous to the polysulfide reductase of *Wolinella succinogenes* and is proposed to oxidize sulfite located in the cytoplasm while reducing menaquinone. The phylogeny of the genes encoding PSRLC3 is congruent with that for most proteins in GSB (Fig. 3.1). PSRLC3 is only present in about half the sequenced GSB strains but is present in *Chp. thalassium*, which as noted above only slow oxidizes sulfur globules to sulfate. PSRLC3 could represent an ancestral mechanism for sulfite oxidation in the GSB, which was subsequently replaced with the more efficient, energy-conserving Sat-Apr-Qmo system by horizontal gene transfer into selected strains. Interestingly, the PSRLC3 is also found in the genomes of all sequenced *Chloroflexus* spp. and *Roseiflexus* spp., but this enzyme rarely occurs in other bacteria.

3. Oxidation of Thiosulfate

The ability of some GSB to grow on thiosulfate can be accounted for by the presence of the sulfur-oxidizing (SOX) system in these organisms

(Frigaard and Dahl, 2009). All GSB strains that encode this system contain a conserved gene cluster of eight genes, *soxJXYZAKBW* (CT1015–CT1023 in *Cba. tepidum*). The *soxCD* genes found in many other thiosulfate-utilizing organisms are not found in GSB. Thiosulfate is linked to the carrier complex SoxYZ by oxidation by the complex SoxAXK (Ogawa et al., 2008). Hydrolytic cleavage of the sulfone moiety by SoxB releases sulfate, after which SoxYZ is presumably regenerated by transfer of the sulfane moiety to soluble polysulfides in the periplasm (Fig. 3.4). In organisms like *Paracoccus denitrificans* that have the *soxCD* genes, the sulfane moiety is oxidized by SoxCD to sulfone, which subsequently is hydrolytically released as sulfate by SoxB thereby regenerating SoxYZ. The *sox* gene cluster occurs sporadically in all genera of GSB (genera *Chlorobium*, *Chlorobaculum*, and *Prosthecochloris* as defined by Imhoff (2003)) except *Chloroherpeton*. The *sox* gene cluster of *Chl. phaeovibrioides* DSM 265 is located on an 11-kbp island that is flanked by inverted repeats and that also contains genes indicative of a mobile element (including a transposase and an integrase) (Frigaard and Bryant, 2008a; Gregersen et al., 2011).

4. Other Putative Thiotrophic Enzymes

Flavocytochrome *c* is a soluble, periplasmic enzyme (FccAB), which *in vitro* catalyzes the oxidation of sulfide by Cyt *c* (Frigaard and Dahl, 2009) (Fig. 3.4). However, its role *in vivo* is uncertain, and the enyme may not be principally involved in sulfide oxidation. Knock-out mutants of the *fccAB* genes in *Cba. tepidum* (S. O. Barbas and N.-U. Frigaard, unpublished results) and in the PSB *Alc. vinosum* (Reinartz et al., 1998) oxidize sulfide at the same rates as the corresponding wild-type strains. Additionally, the *fccAB* genes only occur in some strains of GSB, and their distribution pattern does not match any recognizable pattern of substrate utilization (Frigaard and Bryant, 2008a, b; Gregersen et al., 2011).

Some GSB strains have a modified SoxYZ complex, which has been called SoyYZ (Frigaard and Bryant, 2008a, b). The most notable difference is that the C-terminal sequence motif containing the putative substrate-carrying cysteine is GGCGG in SoxY whereas it is VXAQAC in SoyY. The function of SoyYZ is not known, and its existence is exclusively based on bioinformatic predictions. Similar to the *fccAB* genes described above, the occurrence of the *soyYZ* genes does not correlate with the ability to utilize thiosulfate or any other known sulfur compound. Although *fccAB* genes occur in some strains that do not have *soy* genes, when both occur in GSB, the *soy* genes are always co-localized with the *fccAB* genes encoding flavocytochrome *c*.

5. Assimilatory Sulfate Reduction

Until the discovery of *Chl. ferrooxidans* that grows with sulfate as the sole sulfur source (Heising et al., 1999), the GSB had been considered to be incapable of assimilatory sulfate reduction (Lippert and Pfennig, 1969). However, subsequent genome analyses have revealed that *Chl. ferrooxidans,* and unexpectedly, another GSB strain, *Chl. luteolum* DSM 273, contain a *cys* gene cluster (*cysIHDNCG-cysPTWA*) for assimilatory sulfate reduction (Frigaard and Bryant, 2008a).

E. Oxidative Stress

During initial analyses of the *Cba. tepidum* genome, a surprising finding was the presence of a group of genes predicted to encode enzymes that could confer protection against oxygen by consuming oxygen or oxygen reduction products (reactive oxygen species, ROS) as substrates (Eisen et al., 2002). The presence of these genes suggested that *Cba. tepidum* probably encounters oxygen periodically in its natural environment and that it likely has redundant mechanisms to protect against the oxygen/ROS toxicity. The relevant genes include those predicted to encode a Cyt *bd* (ubiquinol) oxidase (*cydAB*), NADH oxidase (*nox*), rubredoxin:oxygen oxidoreductase (*roo*), bacterioferritin comigratory protein (or thiol peroxidases) (*bcp-1, bcp-2*), thiol peroxidase-1 (*tpx-1*), thioredoxin peroxidase-2 (*tpx-2*), alkyl hydroperoxide reductase (or peroxiredoxin) (*ahpC*), superoxide dismutase (*sodB*), and methionine sulfoxide reductase (*msrA*) (Eisen et al., 2002; Li et al., 2009). Although *Cba. tepidum* cannot produce catalase, it does have a gene for rubrerythrin (*rbr*, CT1327), which is often associated with oxidative stress responses in other bacteria (Kurtz, 2006; Hillmann et al., 2008).

Although a few sequenced strains lack one or two members of the thiol peroxidase superfamily, the genes for these ROS-protective proteins are nearly universally present in *Chlorobiales* genomes. All of the sequenced strains either have a Cyt *bd* oxidase or a Cyt bb_3 oxidase, both of which have very high affinity for oxygen. Several strains (*e.g.*, *Chl. limicola*, *Cba. parvum*, *Chl. ferrooxidans*, and *Pld. phaeum*) potentially produce both of these terminal oxidases (Li et al., 2009). *Cba. tepidum* is surprisingly tolerant of oxygen, especially when the cells are not exposed to strong illumination. However, when the genes encoding the putative ROS-protective proteins were inactivated, most of the resulting mutants had significantly enhanced sensitivity to oxygen or ROS (Li et al., 2009). The most important of the genes for protection from ROS were superoxide dismutase, rubredoxin:oxygen oxidoreductase, and the Cyt *bd* (quinol) oxidase. The mRNA levels for these genes did not change after exposure to oxygen, which shows that oxygen protection is not induced or enhanced after exposure to oxygen. This could be because oxygen rapidly inactivates energy transfer from chlorosomes to reaction centers (see Frigaard and Bryant, 2006), and in such an energy-limited state, cells might not be able to mount a robust, inducible response to oxygen/ROS exposure.

III. *Chloroflexales*: Filamentous Anoxygenic Phototrophs

A. Brief Description of Chloroflexales (FAPs)

The well described chlorophototrophic organisms in the taxon *Chloroflexi* currently make up a monophyletic lineage within the family *Chloroflexales*. Although they have previously been called green gliding bacteria or green nonsulfur bacteria, the chlorophototrophic organisms of the phylum *Chloroflexi* (Order: *Chloroflexales*, Family: *Chloroflexaceae*) are currently known as Filamentous Anoxygenic Phototrophs (FAPs). FAPs are enigmatic, and they defy concrete inferences of their evolutionary origin with respect to other chlorophototrophic phyla (Blankenship, 1992; Raymond et al., 2002). Based upon phylogenetic analyses of 16S rRNA sequences, the *Chloroflexi* represent a very early diverging branch on the Tree of Life, and this clade is most closely related to the *Deinococcus-Thermus* and *Thermotoga* lineages by distance methods (Garrity and Holt, 2001). However, based upon traits relating to their phototrophic metabolism (*e.g.*, reaction centers, chlorophyll biosynthesis, carotenoid biosynthesis), the picture is less clear.

The chlorophototrophic *Chloroflexales* have heterodimeric, type-2 (quinone-type) photochemical reaction centers that are similar to those found in anoxygenic chlorophototrophic *Proteobacteria*. Because it now seems likely that homodimeric type-1 reaction centers predated heterodimeric type-2 reaction centers (Sadekar et al., 2006), this trait is not consistent with the phylogenetic position of these organisms based on 16S rRNA sequences. Although all FAPs can synthesize BChl *a*, the ability to synthesize BChl *c* is limited to those organisms that use chlorosomes for light harvesting. These include strains of the genera *Chloroflexus, Chlorothrix, Oscillochloris,* and *Chloronema* (Pierson and Castenholz, 1974; Keppen et al., 2000; Gich et al., 2003; Klappenbach and Pierson, 2004). FAPs that lack chlorosomes and BChl *c,* such as strains of the genera *Roseiflexus* and *Heliothrix,* are thought to have either lost the ability to make these pigments and antennae structures, or they represent lineages that never obtained the genes necessary to synthesize BChl *c* and chlorosomes, if these genes were acquired laterally (Frigaard and Bryant, 2004; Olson and Blankenship, 2004; however, see Section V). This ambiguity illustrates an important point when considering the evolutionary history of chlorophototrophy in FAPs. Although these organisms are described as "mosaic," in that they contain individual properties that are similar to those in multiple (phototrophic) bacterial lineages, and while it is apparent that horizontal gene transfer has been instrumental in the evolutionary history of the current assemblage of chlorophototroph diversity, it is not yet generally clear which organisms were the donors and recipients in such horizontal gene transfer events, nor is there unanimous agreement on the properties of organisms that today most closely resemble the ancestral type of chlorophototrophy. By increasing both the numbers and the diversity of sequenced genomes for chlorophototrophic prokaryotes, it is hoped that one will gain insights into such important issues (see discussion in Section V).

Table 3.2. Properties of the genomes of sequenced *Chloroflexi* strains

Strain	Genome size (bp)[a]	Mol-% GC content	ORFs	Ave. ORF length (bp)	rRNA operons	tRNAs	ORFs in COGs	Not in COGs	Insertion sequences	Unique ORFs[b]
Chloroflexus sp. Y-400-fl[c]	5,268,950	56.7	4,159	1,030	3	49	3,019	1,140	15	290
Chloroflexus aurantiacus J-10-fl[T, d]	5,258,541	56.7	3,853	1,095	3	49	3,030	823	13	17
Chloroflexus aggregans DSM 9485[T, d]	4,684,931	56.4	3,730	1,084	3	48	2,869	861	25	213
Chloroflexus sp. 396-1[c]	~5,072,000[e]	~53	~4,000	ND	ND	ND	ND	ND	ND	ND
Roseiflexus castenholzii DSM 13941[T, d]	5,723,298	60.7	4,330	1,084	2	48	3,363	967	4	276
Roseiflexus sp. RS-1[c]	5,801,598	60.4	4,517	1,090	2	48	3,452	1,065	54	366
Herpetosiphon aurantiacus ATCC 23779[T]	6,785,430[f]	50.9	5,278	1,121	5	51	3,782	1,496	13	1,759
Thermomicrobium roseum DSM 5159[T, c]	2,920,744[g]	64.3	2,854	911	1	44	2,075	779	1	954
Sphaerobacter thermophilus DSM 20745[T]	~3,989,000[e]	~68	~3,550	~950	1	50	ND	ND	ND	71 188
Dehalococcoides ethenogenes 195	1,469,720	48.9	1,580	833	1	46	1,195	385	2	61
Dehalococcoides sp. BAV1	1,341,892	47.2	1,371	875	1	46	1,117	254	4	ND
Dehalococcoides sp. VS	~2,288,000[e]	~55	~2,100	~900	1	60	ND	ND	ND	ND

ND not determined

[a] All genomes are circular, double-stranded DNA molecules

[b] Genes not found in other complete FAP genomes

[c] These strains were isolated from alkaline siliceous hot springs in Yellowstone National Park

[d] These strains were isolated from hot springs in Japan

[e] Only a draft genome is currently available

[f] This value includes the chromosome (6,346,587 bp) and two circular plasmids, pHAU01 and pHAU02, which are 339,639 and 99,204 bp, respectively

[g] This value includes the genome (2,003,006 bp) and a circular, unnamed megaplasmid (or a second chromosome), which is 917,738 bp

The genomes of six FAPs belonging to two genera, *Chloroflexus* and *Roseiflexus*, have been sequenced thus far (Table 3.2). The draft genome of *Oscillochloris trichoides* DG-6 has recently been completed (Kuznetsov et al., 2011). The genome of *Herpetosiphon aurantiacus,* a heterotrophic member of the Order *Herpetosiphonales* (Family: *Herpetosiphonaceae*), has also been sequenced (Table 3.2). This aerobic, filamentous, mesophilic and heterotrophic organism was isolated from the slime coat of a green alga (*Chara* sp.) from Birch Lake, MN (Holt and Lewin, 1968). *H. aurantiacus* serves as an important outgroup for comparative genomic studies to understand FAP ecology and evolution.

B. Genome Sequences of Chloroflexales

The kingdom/phylum *Chloroflexi* is a deep branching lineage of the domain *Bacteria*, and as noted above, on the basis of 16S rRNA sequences it is the earliest branching lineage that contains chlorophototrophs. Currently, the taxon *Chloroflexi* contains five class-level lineages, only four of which have cultured representatives. These include the *Chloroflexi*, *Anaerolinea*, *Dehalococcoidetes*, *Thermomicrobia*, and a clade of 16S rRNA sequences from mostly marine organisms (Hugenholtz and Stackebrandt, 2004). Recently, Sekiguchi et al. (2003) described the first cultured members of the *Anaerolinea*. *Anaerolinea thermophila* is a moderately thermophilic, filamentous anaerobe isolated from a sludge reactor treating wastewater from a tofu manufacturing plant. *Caldilinea aerophila* is a moderately thermophilic, facultatively aerobic, filamentous, heterotrophic organism isolated from the sulfur turf of Nakao hot spring (Sekiguchi et al., 2003). No genomic sequence information is yet available for any strain of *Anaerolinea* or *Caldilinea* spp. However, recent analyses of metagenomic and metatranscriptomic data suggest that there may be organisms from the same class in the phototrophic mats of alkaline siliceous hot springs in Yellowstone National Park (Klatt et al., 2011). The metagenomic data predicts that this organism(s) is a chlorophototroph, can synthesize BChl *a,* has type-2 reaction centers, and is quite divergent from both *Chloroflexus* and *Roseiflexus* spp. Further analyses will be required to determine if this organism is indeed a new type of chlorophototroph within the taxon *Chloroflexi*.

The class *Dehalococcoidetes* includes organisms that have the ability to reductively dehalogenate a variety of environmental pollutants, including trichloroethene, tetrachloroethene and some related halogenated compounds including vinyl chloride in some cases. As shown in Table 3.2, genome sequences (or a draft) are currently available for three *Dehalococcoides* spp. These bacteria generally have characteristically small genomes (1.3–2.3 Mb), and the data have shown that these genomes contain multiple paralogous genes encoding reductive dehalogenases and hydrogenases (Seshadri et al., 2005). The available genomes for these organisms do not appear to contain any obvious vestiges of chlorophyll biosynthesis and photosynthesis.

Genome sequence data are currently available for two members of the *Thermomicrobia*. *Thermomicrobium roseum* was isolated from Toadstool Spring (Jackson et al., 1973), a small thermal feature associated with the much larger, alkaline siliceous hot spring, Mushroom Spring, in the Lower Geyser Basin of Yellowstone National Park. This carotenoid-rich, reddish-colored, aerobic, and Gram-negative bacterium grows optimally at 70–75°C. Its genome comprises two circular DNA molecules of 2.0 and 0.92 Mb, and although the isolated organism has never been reported to have flagella or swimming motility, the megaplasmid encoded all of the proteins required to produce a flagellum (Wu et al., 2009). Interestingly, the genome revealed that *T. roseum* has the capacity to oxidize carbon monoxide aerobically, and laboratory studies with *T. roseum* confirmed that it is the first thermophile known to perform this metabolism (Wu et al., 2009). Finally, *Sphaerobacter thermophilus,* which was isolated from an aerobic, thermophilic sludge, is a Gram-positive, non-spore-forming bacterium that was recently transferred from the *Actinobacteria* to the family *Thermomicrobia* (Hugenholtz and Stackebrandt, 2004). A draft of this genome is also currently available in the databases.

The Class *Chloroflexi* currently includes all of the well-characterized genera of chlorophototrophs (*i.e.*, FAPs) found in the phylum *Chloroflexi* as well as some non-phototrophic organisms,

e.g., *H. aurantiacus*. Unlike FAPs, *H. aurantiacus* is a saprophytic bacterium that facultatively preys on other microbes. Like myxobacteria, this predatory "wolfpack" bacterium can use its gliding motility and chemotaxis capabilities to "hunt," and its large arsenal of digestive enzymes to feed on, other bacteria (Jurkevitch, 2007). The *H. aurantiacus* genome comprises three circular DNA molecules: the chromosome (6,346,587 bp), and two plasmids, pHAU01 (339,639 bp) and pHAU02 (99,204 bp). Although this genome contains no remnant genes for proteins associated with photosynthesis, unlike the genomes of other *Chloroflexi* it contains 13 large gene clusters, accounting for nearly 400 kb and ~6.3% of the genome, that are postulated to encode secondary metabolites (M. Nett and D. A. Bryant, unpublished observations). The presence of numerous genes encoding putative polyketide synthases and non-ribosomal peptide synthetases is similar to other specialist producers of secondary metabolites (*e.g.*, *Streptomyces* spp. and *Myxococcus xanthus*).

Complete genome sequences are available for three *Chloroflexus* spp. and for two *Roseiflexus* spp., and a draft sequence has also been determined for *Chloroflexus* sp. 396–1 (Table 3.2). The type strain for the phylum *Chloroflexi* is *Cfx. aurantiacus* J-10-fl, which was isolated in 1970 from a hot spring near Sokokura, Hakone district, Japan (Pierson and Castenholz, 1974). Two of the *Chloroflexus* spp. strains, Y-400-fl (Octopus Spring) and 396–1 (Conophyton Pool, Fairy Creek Meadow), were isolated from the Lower Geyser Basin of Yellowstone National Park (Pierson and Castenholz, 1974; Walter et al., 1972; Madigan et al., 1974; Madigan and Brock, 1977a, b). Surprisingly, the Y-400-fl and J-10-fl strains of *Cfx. aurantiacus* are remarkably similar. The genome of strain J-10-fl contains four genes (Caur_1381–1384) that are not found in the strain Y-400-fl genome, while the Y-400-fl genome contains only one gene, Chy400_1091 that is not found in strain J-10-fl. In the shared regions of these two genomes, only 687 nt differences are found (*i.e.*, the genomes are 99.98% identical). It is very difficult to rationalize this level of sequence similarity for strains that were supposedly isolated from such chemically distinct and geographically isolated environments. Although the incomplete genome of strain 396–1 has not yet been fully analyzed, it is already clear that the differences between strains Y-400-fl and 396–1, which were both isolated from chemically similar environments, are far greater than the differences between the geographically distant isolates, J-10-fl and Y-400-fl. Moreover, strain 396–1 is much more similar to the predominant *Chloroflexus* spp. strains in the phototrophic mats of alkaline siliceous hot springs in Yellowstone National Park (see Ward et al., Chapter 1 of this volume).

In contrast to the situation for *Cfx. aurantiacus* strains Y-400-fl and J-10-fl, *Roseiflexus* sp. RS-1 and *R. castenholzii* differ significantly. The type strain of the genus *Roseiflexus, R. castenholzii*, was also isolated in Japan from Nakabusa hot spring (Hanada et al., 2002), while strain RS-1 was isolated from Octopus Spring in Yellowstone National Park (van der Meer et al., 2010). Reciprocal best-hit analyses by BLASTP show that strain RS-1 has 824 ORFs (~18%) not found in *R. castenholzii*, which in turn has 637 ORFs (15%) not found in *Roseiflexus* sp. RS-1. These genomic differences, and important accompanying metabolic differences (*e.g.*, phosphonate utilization by strain RS-1 but not *R. castenholzii*), indicate that these very different *Roseiflexus* spp. isolates represent different ecological species. These very large differences dramatically illustrate why it is important to know the ecological origins of organisms, if one is interested in studying community-level properties of ecosystems through systems biology approaches (see Ward et al., Chapter 1 of this volume). Similarly, knowledge of the ecological origins of isolates can also help one to gain insights into the functions of genes that are not important when cells are grown in non-natural environments and in isolation from other microorganisms.

C. Genes for Photosynthesis in Chloroflexales

1. Photosynthetic Apparatus

Reaction centers have been isolated and biochemically characterized from two FAPs: *Cfx. aurantiacus* (Bruce et al., 1982; Blankenship et al., 1983; Shiozawa et al., 1987, 1989; Feick et al., 1996) and *R. castenholzii* (Yamada et al., 2005). These type-2 reaction centers have three subunits, L, M, and the tetraheme Cyt c_{554}, but

both of these reaction centers characteristically lack the H subunit that functions in Q_B protonation and stabilization of the L and M heterodimers in *Proteobacteria*. An additional difference is that, unlike purple bacterial reaction centers, which bind four BChl *a* molecules and two BPheo *a* molecules, the reaction centers of both FAPs bind three BChl *a* and three Bpheo *a* molecules (Pierson and Thornber, 1983; Yamada et al., 2005). As expected from these biochemical analyses, the six FAP genomes sequenced thus far contain homologs of the *pufL* and *pufM* genes, which encode the L and M subunits of type-2 bacterial reaction centers, and likewise, all six lack the *puhA* gene that encodes the H subunit of bacterial reaction centers. Unlike the situation in most purple bacteria, which have a *pufABLM(C)* operon (Youvan et al., 1984; Kiley et al., 1987), the *pufLM* operon of *Cfx. aurantiacus* is physically distant from the *pufBA-pufC* operon, which encodes the beta and alpha subunits of light-harvesting B806-866 complex (Xiong et al., 2000; Xin et al., 2005), respectively, and the membrane-bound, multi-heme Cyt c_{554} associated with the reaction center (Watanabe et al., 1995). In contrast, the genes of these two operons occur at a single locus with a different gene order, *pufBA(L-M)C,* in *R. castenholzii* and *Roseiflexus* sp. RS1 and may form a single transcription unit (Yamada et al., 2005). An important difference, which distinguishes *Roseiflexus* spp. from chlorophototrophs of the phylum *Proteobacteria* and the *Chloroflexus* spp., is that the *pufL* and *pufM* genes are fused into a single ORF that presumably is posttranslationally cleaved to produce the L and M polypeptides that are observed in purified reaction centers (Yamada et al., 2005).

Chloroflexus spp. synthesize large amounts of BChl *c* and correspondingly produce chlorosomes under anoxic growth conditions (Madigan and Brock, 1977b; Sprague et al., 1981). Early studies suggested that the chlorosomes of *Cfx. aurantiacus* only contained three proteins: CsmA, CsmM, and CsmN (Feick et al., 1982; Feick and Fuller, 1984; Niedermeier et al., 1994). The 5.7-kDa CsmA polypeptide is typically the most abundant chlorosome protein; it is structurally and functionally orthologous to CsmA in *Chlorobiales* and is similarly synthesized as a precursor with a carboxyl-terminal extension of 27 amino acids (Theroux et al., 1990; Chung et al., 1994; Bryant et al., 2002). In *Cfx. aurantiacus* CsmA has been implicated in BChl *a* binding and almost certainly forms the paracrystalline baseplate that is also observed in the chlorosomes of this organism (Sakuragi et al., 1999; Montaño et al., 2003; Frigaard and Bryant, 2006; Pšenčík et al., 2009). CsmM (11 kDa) and CsmN (14 kDa) are paralogous and are possibly more distantly related in sequence to the CsmC and CsmD proteins of *Cba. tepidum* (Vassilieva et al., 2000, 2002). Magnesium-Proto IX monomethyl ester aerobic oxidative cyclase (encoded by *acsF*) was recently suggested to be a component of the chlorosome envelope in *Cfx. aurantiacus* (Tang et al., 2009). AcsF is characteristically required for oxygen-dependent BChl *a* biosynthesis in other organisms (see Gomez Maqueo Chew and Bryant, 2007b), and BChl *c* synthesis and chlorosome biogenesis typically only occur under microoxic or anoxic conditions. However, Tang et al., (2009) presented no evidence showing that AcsF uniquely co-localizes with other chlorosome envelope proteins. Moreover, no other enzymes of BChl biosynthesis have been detected in similar studies with chlorosomes (Frigaard et al., 2002b, 2004b; Vassilieva et al., 2002).

At least three additional proteins have been reported to occur in the chlorosome envelope of *Cfx. aurantiacus*. CsmO (Caur_1311) is a member of the CsmB/CsmF motif family (Vassilieva et al., 2002). Genes homologous to *csmO* are found in all *Chloroflexus* spp., and the predicted amino acid sequence matches the N-terminal sequence of a protein found in chlorosomes (Frigaard et al., 2002b; Lehmann et al., 1994). Frigaard et al., (2002a) additionally found CsmP (Caur_0142; 17.2 kDa) and CsmY (Caur_0356; 22 kDa) in chlorosomes prepared from *Cfx. aurantiacus* J-10-fl. The *csmP* gene lies just downstream from the genes encoding CsmM and CsmN (and may be cotranscribed with them) and is found in the genomes of all chlorosome-producing *Chloroflexus* spp. strains, but not in the genomes of *Chlorobiales* or *Cab. thermophilum*. CsmY is predicted to be an Fe/S protein and has sequence similarity to CsmI, CsmJ, and CsmX (Vassilieva et al., 2000, 2002). Orthologs are found in all chlorosome-producing *Chloroflexus* spp. strains, but, as expected, no homologs are found in *Roseiflexus* spp. or *H. aurantiacus*. Thus, the chlorosome envelope of *Cfx. auraunticus*

contains proteins of the same four motif families that are found in *Cba. tepidum* (Vassilieva et al., 2002). This suggests that these two types of chlorosomes are more similar than has generally been acknowledged in the literature (Li and Bryant, 2009).

Consistent with biochemical evidence that *Roseiflexus* spp. do not synthesize Bchl *c* and ultrastructural studies indicating that chlorosomes are absent, the *Roseiflexus* spp. genomes lack the *csmA*, *csmM*, and *csmN* genes for chlorosome envelope proteins (van der Meer et al., 2010). However, homologs of *acsF* occur in the genomes of *Roseiflexus* sp. RS-1 and *R. castenholzii*, both of which lack chlorosomes. Conversely, homologs of *acsF* are not found in the genomes of any sequenced *Chlorobiales* strain, which produce chlorosomes but only grow and synthesize (B) Chls under anoxic conditions. The ubiquitous presence of both *acsF* and *bchE* genes in FAP genomes, and the absence of *acsF* from the genomes of the strictly anaerobic *Chlorobiales* strains, which exclusively use BchE, the oxygen-independent Mg-Proto IX methyl ester oxidative ring cyclase, suggest that the ability to synthesize BChl *a* and/or BChl *c* under oxic or microoxic conditions may be an important aspect of FAP physiology. These observations also demonstrate that the postulated interaction between AcsF and chlorosome envelopes in *Cfx. aurantiacus* cannot be an obligate one.

2. Chlorophyll Biosynthesis

Based upon predictions from sequence homology to experimentally characterized genes in PSB (*Rhodobacter sphaeroides* and *Rhodobacter capsulatus*) and the GSB *Cba. tepidum*, genes required for the biosynthesis of BChl *a* occur in each of the currently sequenced FAP genomes (for a review, see Gomez Maqueo Chew and Bryant, 2007b). The only exception to this statement concerns the apparent absence of a recognizable gene encoding 3,8-divinyl PChlide 8-vinyl reductase in *Roseiflexus* spp. (see below). Phylogenetic analyses suggest that these genes are much more similar to the BChl *a* biosynthesis genes in the *Chlorobiales* than to those of the chlorophototrophic *Proteobacteria* (Xiong et al., 2000; also see Section V below). The genes required for BChl *c* biosynthesis have also been identified in the genomes of *Chloroflexus* spp. by sequence similarity to experimentally characterized genes in *Cba. tepidum* (Frigaard and Bryant, 2004; Gomez Maqueo Chew and Bryant, 2007b; Liu and Bryant, 2011).

One similarity between the BChl biosynthetic pathways of FAPs and GSB is that organisms of both taxa have a small multigene family encoding the large subunit of Proto IX Mg-chelatase (BchH; Eisen et al., 2002; Frigaard et al., 2003; Frigaard and Bryant, 2006; Johnson and Schmidt-Dannert, 2008; Gomez Maqueo Chew et al., 2009). A possible reason for this might be that different H subunits are expressed under oxic and anoxic (or microoxic) growth conditions to accomplish differential regulation of the pathway under these two growth conditions (Frigaard and Bryant, 2004). Unlike *Chlorobiales* strains that only grow under anoxic conditions and that only have BchE (oxygen-independent Mg-Proto IX monomethyl ester oxidative cyclase; Gomez Maqueo Chew and Bryant, 2007b), *Chloroflexus* and *Roseiflexus* spp. can synthesize BChl under both anoxic and microoxic conditions. Correspondingly, their genomes encode both AcsF (oxygen-dependent Mg-Proto IX monomethyl ester oxidative cyclase; Pinta et al., 2002; Ouchane et al., 2004; Tang et al., 2009) and BchE. Enzymes of the radical SAM superfamily, which includes BchE, are usually rapidly inactivated upon exposure to even low levels of oxygen (Ouchane et al., 2004; Layer et al., 2005). It is interesting to note that *Cfx. aurantiacus* synthesizes some BChls even when grown under microoxic conditions (Foster et al., 1986; Oelze, 1992).

A few important differences occur in the BChl *c* biosynthetic pathways of the *Chlorobiales* and chlorophototrophic *Chloroflexales*. For example, the genomes of all sequenced *Chloroflexus* spp. lack the *bchQ* and *bchR* genes, which encode the methyltransferases that add methyl groups to the C-8^2 and C-12^1 positions of the BChl *c* homologs synthesized by all *Chlorobiales* (Gomez Maqueo Chew and Bryant, 2007b; Gomez Maqueo Chew et al., 2007; see Section II.C.2). The absence of these two gene products is consistent with the observation that BChl *c* molecules in *Chloroflexus* spp. have only ethyl and methyl groups at the C-8 and C-12 positions, respectively. Although there is not yet a genome sequence available for

Chloronema giganteum, this organism has been reported to methylate BChl *c* at the C-8[2] and C-12[1] positions (Gich et al., 2003), and thus the genome of *Chloronema giganteum* should encode *bchQ* and *bchR,* as is the case in *Cab. thermophilum* (see Section IV below). *Oscillochloris* spp. also synthesize BChl *a* and BChl *c* (Keppen et al., 2000), but it has not yet been reported whether side-chain methylation of BChl *c* occurs (Taisova et al., 2002). Therefore, it is not yet known how the biosynthetic pathway in this organism might compare to those in *Chloroflexus* spp. and *Chlorobiales* strains.

Another important difference in BChl synthesis in members of the *Chlorobiales* and *Chloroflexales* concerns the reduction of the 8-vinyl side chain of 3,8-divinyl PChlide. As discussed above in Section II.C.2, most of the *Chlorobiales* strains listed in Table 3.1 have orthologs of *bciA*, which encodes an 8-vinyl PChlide reductase that is similar to the enzyme found in *Arabidopsis thaliana* (Gomez Maqueo Chew and Bryant, 2007a, b; Nagata et al., 2005). However, at least three *Chlorobiales* strains (*Chl. limicola* DSM 245, *Chl. phaeobacteroides* DSM 266, and *Chp. thalassium*), and all *Chloroflexus spp.* strains, instead apparently use an 8-vinyl PChlide reductase that is orthologous to that encoded by ORF *slr1923* of the cyanobacterium *Synechocystis* sp. PCC 6803 (Ito et al., 2008; Islam et al., 2008). These genes are usually annotated as coenzyme F420 hydrogenase/dehydrogenase beta subunit, but functional studies in cyanobacteria and *Cba. tepidum* (Liu and Bryant, 2011) confirm that these genes have the expected activity. Surprisingly, although *Roseiflexus* spp. synthesize BChl *a*, neither type of 8-vinyl PChlide reductase is encoded in the genomes of these organisms. This observation implies that a third type of 8-vinyl PChlide reductase must occur in *Roseiflexus* spp.

3. Carotenoid Biosynthesis

Carotenoids are thought to play an important role in photoprotection and light harvesting in FAPs, but none of the genes involved in carotenoid biosynthesis in these organisms have been characterized in biochemical detail. However, functional predictions based on sequence similarities to characterized genes from *Cba. tepidum* and cyanobacteria allow inferences to be made about these pathways (Takaichi, 2001; Frigaard et al., 2004b; Maresca et al., 2008b). *Chloroflexus* spp., *Roseiflexus* spp., and *H. aurantiacus* contain homologs of the *crtB*, *crtI*, *crtP*, *crtO*, and *crtY/crtL/cruA* genes, which should allow these organisms to produce the precursors to the β-carotene and γ-carotene derivatives that are observed in chlorosomes and/or cell membranes of these organisms (Takaichi, 2001; Maresca et al., 2008b). Enzymes belonging to three of the four known families of lycopene cyclases occur in various members of the *Chloroflexi* (Maresca et al., 2007, 2008b). *Chloroflexus* spp. have CrtY-type lycopene cyclases, *Roseiflexus* spp. have CrtL-type lycopene cyclases, and *H. aurantiacus* has a CruA-type lycopene cyclase. The presence of *crtO* homologs in all of these organisms explains the ability of these organisms to synthesize carotenoids with keto groups at the 4 and 4′ ring positions. In addition to these genes for carotenoid biosynthesis, the *Chloroflexus* spp., *Roseiflexus* spp., and *H. aurantiacus* genomes also encode homologs of *cruF* (γ-carotene 1′,2′-hydratase), *cruC* (1′-OH glycosyltransferase), and *cruD*, (acyl transferase) (Maresca et al., 2006, 2008b; Graham and Bryant, 2009). The presence of these genes allows the synthesis of monocyclic carotenoids carrying glycosyl or glycosyl-fatty acyl ester moieties attached to a 1′-OH group at the ψ–end of the carotenoid. Although *R. castenholzii* does not synthesize β-carotene and accumulates only a trace amount of γ-carotene, this strain produces significant amounts of 4-keto-γ-carotene derivatives, including methoxy-4-keto-myxocoxanthin, 4-keto-myxocoxanthin-glucoside, and 4-keto-myxocoxanthin-glucoside fatty acid esters (Takaichi 2001; Maresca et al., 2008b). Because none of the *Chloroflexales* genomes encodes an ortholog of *crtF*, there must be an alternative and presently unidentified methyltransferase for the synthesis of the methoxy derivatives of 4-keto-myxocoxanthin. Although the genome of *Heliothrix oregonensis* is not available, this FAP is also known to synthesize keto-carotenoid derivatives of γ-carotene (Pierson et al., 1985). Nothing is known at this time about the carotenoid complements of FAPs such as *Chloronema* and *Oscillochloris* spp., organisms that inhabit anoxic environments and whose genomes are not yet sequenced.

D. Electron Transport Chains in Chloroflexus and Roseiflexus spp.

Possibly due to their long history of co-existence with oxygen-evolving cyanobacteria, FAPs contain genes that putatively encode electron-transport enzymes capable of functioning under both oxic and anoxic conditions, with or without the use of light energy. Some (*e.g.*, strain OK-70-fl) but not all *Chloroflexus* spp. are able to use sulfide as an electron donor for photoautotrophy (Madigan et al., 1974; Hanada et al., 1995). Similar to many *Cyanobacteria*, the *Chloroflexus* and *Roseiflexus* spp. genomes contain genes for a bidirectional [Ni-Fe]-hydrogenase, which gives some of these organisms (*e.g.*, strain OK-70-fl) the ability to use hydrogen as a source of electrons. All of the FAPs except *Roseiflexus* sp. RS-1 also have genes encoding an uptake-type (HupL) [NiFe]-hydrogenase. Finally, the two *Roseiflexus* spp., but not the *Chloroflexus* spp., have a *b*-Cyt-coupled [NiFe]-hydrogenase similar to that found in *Chlorobiales* strains (van der Meer et al., 2010).

The six FAPs and the non-phototrophic *H. aurantiacus* all have the ability to respire aerobically using various organic compounds, which would presumably be oxidized through the tricarboxylic acid cycle and serve as the source of electrons for respiratory electron transport. All have similar electron transport chains comprising a Type-I NADH dehydrogenase (NADH:quinone oxidoreductase); a succinate dehydrogenase/fumarate reductase (succinate:quinone oxidoreductase), which has been purified, characterized and crystallized (Xin et al., 2009); and Cyt oxidase. The succinate dehydrogenase/fumarate reductase can function in either direction *in vitro*, and probably does so as well *in vivo*, because it could be used for both chemotrophic and phototrophic electron transport. This enzyme also functions in autotrophic metabolism as a component of the 3-hydroxypropionate pathway (see Section III.F).

Cyanobacteria have a Cyt b_6f complex that acts as a plastoquinol:Cyt c_6 (plastocyanin) oxidoreductase (Baniulis et al., 2008), and heliobacteria have a similar electron transfer complex (Ducluzeau et al., 2008). Chlorophototrophic *Proteobacteria* have a Cyt bc_1 complex that acts as a ubiquinol:Cyt *c* oxidoreductase, which is structurally similar to the core of mitochondrial complex III and to the Cyt b_6f complex (Berry et al., 2004; Mulkidjanian, 2007). *Chlorobiales* have genes encoding Cyt *b* and Rieske Fe/S proteins (Schütz et al., 1994; Brugna et al., 1998; Schütz et al., 2000), but these organisms do not have a *c*-type Cyt similar to Cyt *f* or c_1. As noted above, Tsukatani et al., (2008) have suggested that Cyt c_{556} is the membrane-associated product of *Cba. tepidum* ORF CT0073 and that it functions in place of Cyt *f* or c_1. FAPs do not have any electron transport complexes equivalent to the ones found in these other chlorophototrophs. Moreover, *Chloroflexus* spp. possess an arsenite oxidase that accounts for their only Rieske-like protein (Lebrun et al., 2003). Instead, FAPs have so-called alternative complex III that acts as a menaquinol:auracyanin (acceptor) oxidoreductase (Yanyushin et al., 2005; Pereira et al., 2007; Gao et al., 2009). Biochemical characterization of alternative complex III from *Rhodothermus marinus* indicates that this enzyme can oxidize quinol and reduce either Cyt *c* or a high-potential iron-sulfur protein (Pereira et al., 2007). Alternative complex III typically comprises six polypeptides: two subunits are *c*-type cytochromes, with one being a monoheme-type and one being a pentaheme-type; one subunit has sequence similarity to molybdopterin oxidoreductases but does not contain molybdenum and instead harbors three [4Fe-4S] clusters and one [3Fe-4S] cluster; and three subunits are integral membrane proteins. *Chloroflexus* spp. have two paralogous forms of alternative complex III (Yanyushin et al., 2005). One form was found in cells grown under aerobic chemotrophic conditions (denoted complex Cr for respiration), and one form was found under anoxic phototrophic conditions (denoted complex Cp for photosynthesis). Thus, Yanyushin et al., (2005) hypothesized that *Cfx. aurantiacus* uses different electron transport chains under these different growth conditions. *Roseiflexus* spp. genomes only encode homologs of the photosynthetic, Cp-type complex. Electrons from the Cp form of alternative complex III are postulated to reduce auracyanin A, a water-soluble, copper-containing, redox protein that can potentially reduce the photosynthetic reaction center by donation of elections to the tetraheme Cyt (Gao et al., 2009; Lee et al., 2009a). The electrons from the Cr form of alternative complex III are hypothesized to reduce auracyanin B, a

membrane-tethered, copper-containing redox protein (Lee et al., 2009a, b). Auracyanin B is structurally similar to auracyanin A and both have very similar Cu-site geometries and redox potentials, but the proteins differ significantly in amino acid sequence and have very different surface properties. Auracyanin A is only found in cells growing photosynthetically, but auracyanin B is found in cells grown photosynthetically and non-photosynthetically. Auracyanin B is postulated to donate electrons preferentially to the terminal oxidase (Lee et al., 2009a). Although *Roseiflexus* spp. can grow heterotrophically or photoheterotrophically (Hanada et al., 2002; van der Meer et al., 2010), these organisms only have auracyanin A and the photosynthetic, Cp-type of alternative complex III. Interestingly, the heterotroph *H. aurantiacus* does not produce alternative complex III, but this organism also has genes for Cyt oxidase. Instead, *H. aurantiacus* has genes encoding proteins related to the PetB and the Rieske Fe/S proteins of *Chlorobiales*. *H. aurantiacus* does not have homologs of the small membrane associated, mono-heme *c*-type Cyt of GSB (see Section II.C.1 above), but the *petCB* genes are found immediately upstream from a gene encoding a diheme *c*-type Cyt similar to the CcoP subunit of Cyt bb_3-type oxidases.

E. Sulfur Metabolism

Although the chlorophototrophic *Chloroflexales* were once referred to as green non-sulfur bacteria, it is clear that some FAPs found in sulfide-containing hot springs and microbial mats utilize sulfide as an electron donor for photoautotrophic growth (Hanada and Pierson, 2006). However, most FAP isolates are not dependent on reduced sulfur compounds for growth, and autotrophic growth with accompanying sulfide oxidation is usually not the preferred growth mode under laboratory conditions. Nevertheless, autotrophic growth using sulfide has convincingly been demonstrated for *Cfx. aurantiacus* OK-70-fl (Madigan et al., 1974; Madigan and Brock, 1977a) and several *O. trichoides* strains (Keppen et al., 2000), and also occurs in *Chloronema giganteum* (Gich et al., 2003) and *Candidatus* Chlorothrix halophila (Klappenbach and Pierson, 2004). *Chloroflexus spp.* strains J-10-fl, Y-400-fl, and 396–1 (Madigan et al., 1974), and *Cfx. aggregans* strains MD-66 (DSM 9485) and YI-9 (DSM 9486) (Hanada et al., 1995), lack the ability to grow autotrophically on sulfide. *Roseiflexus* spp. are not yet known to occur in sulfide-rich environments, and the characterized isolates do not require sulfide. When sulfide is oxidized by *Chloroflexus* and *Oscillochloris* spp., globules of elemental sulfur are deposited outside the cells. One study found that thiosulfate was used as a source of sulfur under photoheterotrophic growth of *Chloroflexus* spp. but not under photoautotrophic conditions (Krasil'nikova, 1987; Krasil'nikova and Kondrat'eva, 1987; Kondrat'eva and Krasil'nikova, 1988).

In contrast to the situation in the *Chlorobiales* strains that often have more than one type of sulfide: quinone oxidoreductase (SQR), all sequenced *Chloroflexus and Roseiflexus* spp. strains have only a single type of SQR, which belongs to the type II family (Caur_3894 in *Cfx. aurantiacus* J-10-fl). The *H. aurantiacus* genome does not encode an SQR. This type II SQR is also found in some cyanobacteria, some Firmicutes, the mitochondria of ascomycetes (Vande Weghe and Ow, 1999), and other eukaryotes. The SQRs of FAP strains and fungal mitochondria have about 40% sequence identity. However, because none of the sequenced FAP strains in Table 3.2 can grow on sulfide as electron source, this type II SQR apparently does not support autotrophic growth on sulfide and thus may function in sulfide detoxification. It seems likely, therefore, that true sulfidotrophic strains like OK-70-fl will contain additional SQR homologs. None of the sequenced strains of *Chloroflexi* encode *dsr* or *sox* genes, but it is possible that these genes may also occur in strains such as *Cfx. aurantiacus* OK-70-fl that are capable of autotrophic growth on sulfide or thiosulfate.

Sulfate uptake in FAPs and *H. aurantiacus* presumably involves a homolog of the high-affinity Sul1 sulfate transporter system similar to that found in yeast (Frigaard and Dahl, 2009). Activation of sulfate for biosynthetic purposes in bacteria usually is initiated by formation of adenosine-5′-phosphosulfate (APS), a reaction catalyzed by a sulfurylase encoded either by *sat* or *cysDN*. APS is then further phosphorylated to 3′-phosphoadenosine-5′-phosphosulfate (PAPS) by an APS kinase encoded by *cysC*. The enterobacterial CysDN-type ATP-sulfurylase is not present in FAPs or *H. aurantiacus*. Instead, all

sequenced *Chloroflexus* and *Roseiflexus* spp. contain a fusion of a Sat-type ATP-sulfurylase and a CysC-type APS kinase (Caur_2113 in *Cfx. aurantiacus* J-10-fl), which could transform sulfate directly to PAPS. However, this enzyme is not found in *H. aurantiacus*. PAPS is probably used in all *Chloroflexus* and *Roseiflexus* spp. as a substrate in the biosynthesis of sulfonated compounds such as sulfolipids, which do not require reduction of the sulfate. A complete pathway for assimilatory sulfate reduction is only present in some strains. In such strains a gene cluster, which does not involve the genes putatively involved in the sulfonation reactions mentioned previously, encodes the enzymes required for activation and reduction of sulfate via PAPS. In *Chloroflexus* spp. strains J-10-fl, Y-400-fl, and Y-396-1 this gene cluster has the structure *cysCH-sat-cysGI* (Caur_0692 to Caur_0686 in *Cfx. aurantiacus* J-10-fl) and in *H. aurantiacus* it has the structure *cysH-sat-cysCIG* (Haur_1749–Haur_1744). *Cfx. aggregans* and *Roseiflexus* spp. strains apparently cannot perform assimilatory sulfate reduction.

F. Autotrophy and Central Carbon Metabolism

FAPs are metabolically versatile organisms: some are capable of growth under anoxic conditions as photolithoautotrophs; all can grow as photoorganoheterotrophs or photomixotrophs with a variety of substrates under anoxic conditions; and many can grow as chemoorganoheterotrophs on a various substrates under oxic conditions. Strikingly, two pathways for carbon dioxide fixation are known among chlorophototrophic members of the *Chloroflexales*. *O. trichoides*, a sulfide-oxidizing, anaerobic FAP, has been shown to have the enzymes of the reductive pentose phosphate pathway (*i.e.*, the Calvin-Benson-Bassham cycle) (Ivanovsky et al., 1999). Portions of the *cbbL* gene, encoding the large subunit of Rubisco, and *nifH* gene, encoding the iron protein of nitrogenase, have been amplified by PCR and sequenced (Turova et al., 2006). Interestingly, the *O. trichoides* Rubisco appears to be closely similar to that of *Sulfobacillus acidophilus*, a facultatively autotrophic, sulfide-oxidizing acidophile (Caldwell et al., 2007). No additional genomic details are currently available for *O. trichoides*, although genome sequencing of this organism is reportedly in progress (R. Ivanovsky and M. Kuznetsov, personal communication). All other characterized FAPs utilize the 3-hydroxypropionate pathway for CO_2 fixation.

The 3-hydroxypropionate pathway for autotrophic CO_2 fixation was discovered in *Cfx. aurantiacus* (Strauss and Fuchs, 1993), and all of the enzymes postulated to function in the proposed pathway, and the genes that encode them, have now been described (Alber and Fuchs, 2002; Herter et al., 2001, 2002a, b; Hügler et al., 2002; Friedmann et al., 2006a, b, 2007; Zarzycki et al., 2008, 2009; for a brief summary, see Tabita (2009) and Chapter 9). This autotrophic pathway is conveniently described as operating as two interconnected cycles (Herter et al., 2002b; Zarzycki et al., 2009). In the first cycle two CO_2 molecules are combined with acetyl-CoA at the expense of four ATP equivalents and three NADPH equivalents to produce glyoxylate; in the second cycle glyoxylate is combined with propionyl-CoA to produce pyruvate, which can enter central metabolism. Three additional ATP equivalents and three more NADPH are consumed (Fig. 3.5), and the acetyl-CoA is regenerated (Zarzycki et al., 2008, 2009). The enzymes that function in some of these reactions are also potentially involved in acetate assimilation, fatty-acid biosynthesis, and the reactions of the tricarboxylic acid cycle. Among the currently available genome sequences, all components of the 3-hydroxypropionate pathway can be found in each of the genomes of *Chloroflexus* and *Roseiflexus* spp., but the genome of the heterotrophic *H. aurantiacus* does not encode any of the enzymes specific to this pathway (Klatt et al., 2007; unpublished data).

In the first step of the pathway, an acyl-CoA carboxylase catalyzes the carboxylation of acetyl-CoA to form malonyl-CoA (Fig. 3.5). This reaction is similar to the carboxylation of propionyl-CoA to form methylmalonyl-CoA in step four, and many acyl-CoA carboxylases can utilize both acetyl-CoA and propionyl-CoA as substrates (Hügler et al., 2003). The FAP genomes encode multiple paralogous genes that are predicted to encode subunits of various acyl-CoA carboxylases. The substrate specificities of the enzymes for acetyl-CoA or propionyl-CoA were predicted *in silico* for all carboxyltransferase subunit paralogs (Klatt et al., 2007). Previous studies had shown that sequence motifs in this subunit can accurately predict the substrate specificities of acyl-CoA carboxylases of high G+C Gram-positive bacteria (Diacovich et al., 2004).

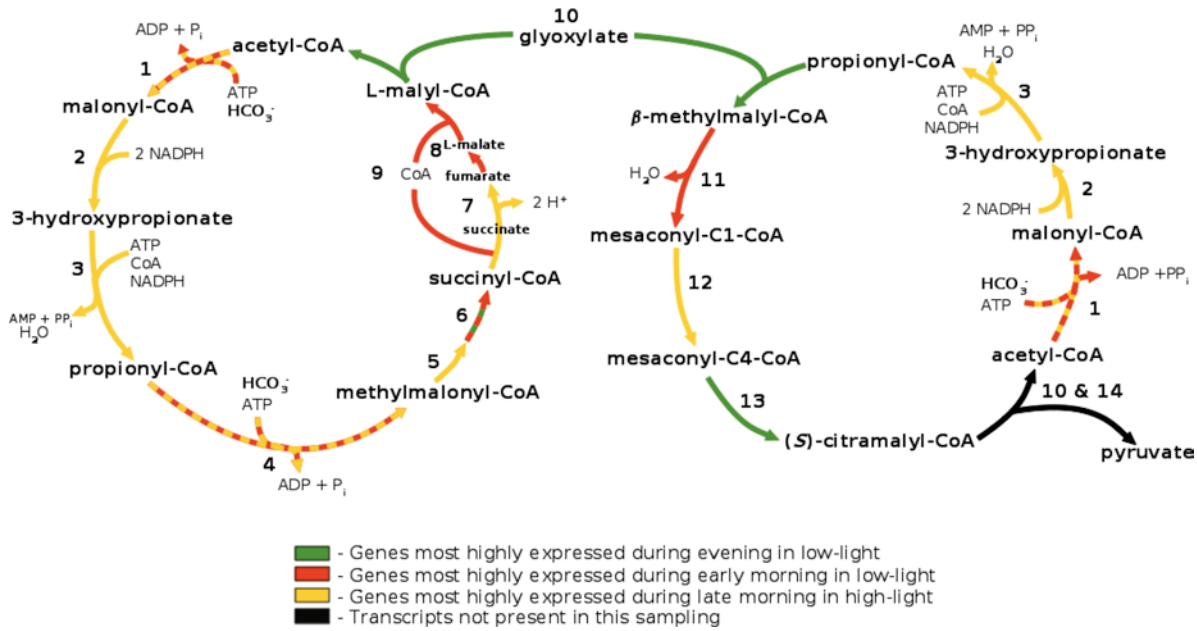


Fig. 3.5. Transcription of genes from uncultivated *Roseiflexus* spp. from Mushroom Spring, WY involved in the 3-hydroxypropionate autotrophic pathway. Reaction steps are numbered and catalyzed by the following enzymes: *1*, acetyl-CoA carboxylase; *2*, malonyl-CoA reductase; *3*, propionyl-CoA synthase; *4*, propionyl-CoA carboxylase; *5*, methylmalonyl-CoA epimerase; *6*, methylmalonyl-CoA mutase; *7*, succinate dehydrogenase; *8*, fumarate hydratase; *9*, succinyl-CoA:(S)-malate-CoA transferase; *10*, (S)-malyl-CoA/(beta)-methylmalyl-CoA/(S)-citramalyl-CoA; *11*, mesaconyl-C1-CoA hydratase; *12*, mesaconyl-CoA C1-C4 CoA transferase; *13*, mesaconyl-C4-CoA hydratase.

As an aid in predicting which carboxylase enzymes are involved in step 1 or 4 (or both), it was helpful to examine the genomic context of the various genes. Genes encoding subunits of one of the acyl-CoA carboxylases are co-localized with the genes encoding malonyl-CoA reductase (*mcr*) and propionyl-CoA synthase (*pcs*), which catalyze steps 2 and 3, on the genomes of *Roseiflexus* sp. RS1 and *R. castenholzii* (Klatt et al., 2007) (Fig. 3.6a). Because this carboxyltransferase subunit is predicted to have dual substrate specificity, it is not known with certainty whether this acyl-CoA carboxylase is specific for acetyl- or propionyl-CoA. However, these genes are coordinately transcribed with the *mcr* and *pcs* genes in *Roseiflexus* spp., which implies a common regulatory control for expression of these genes and therefore probably implies related functionality as well. This gene order does not occur in *Cfx. aurantiacus* or *Cfx. aggregans*, in which only the carboxyltransferase subunits *accD* and *accA* are adjacent. It is unclear whether the different gene arrangements in *Chloroflexus* spp. have any implications for the regulation of the genes encoding these steps of the 3-hydroxypropionate pathway.

The next steps (5–8) in the pathway converge with a portion of the tricarboxylic acid cycle, and genes encoding these enzymes also in the heterotrophic *H. aurantiacus*. These enzymes include methylmalonyl-CoA mutase, methylmalonyl-CoA epimerase, menaquinol:fumarate (succinate dehydrogenase) oxidoreductase, and fumarate hydratase. The resulting L-malate recombines with the Coenzyme A transferred from the earlier intermediate succinyl-CoA via another key enzyme of the pathway, succinyl-CoA:L-malate-CoA transferase (*smtAB*, step 9) (Friedmann et al., 2006a). The genes encoding this enzyme are only found in the FAP genomes, and these genes are co-localized with other key genes of the pathway including L-malyl-CoA lyase:β-methylmalonyl-CoA lyase (*mcl, s*tep 10) (Herter et al., 2002a), and mesaconyl-CoA hydratase (*mch*, step 11) (Zarzycki et al., 2008) (Fig. 3.6b). This gene cluster encodes a conserved ORF that belongs to

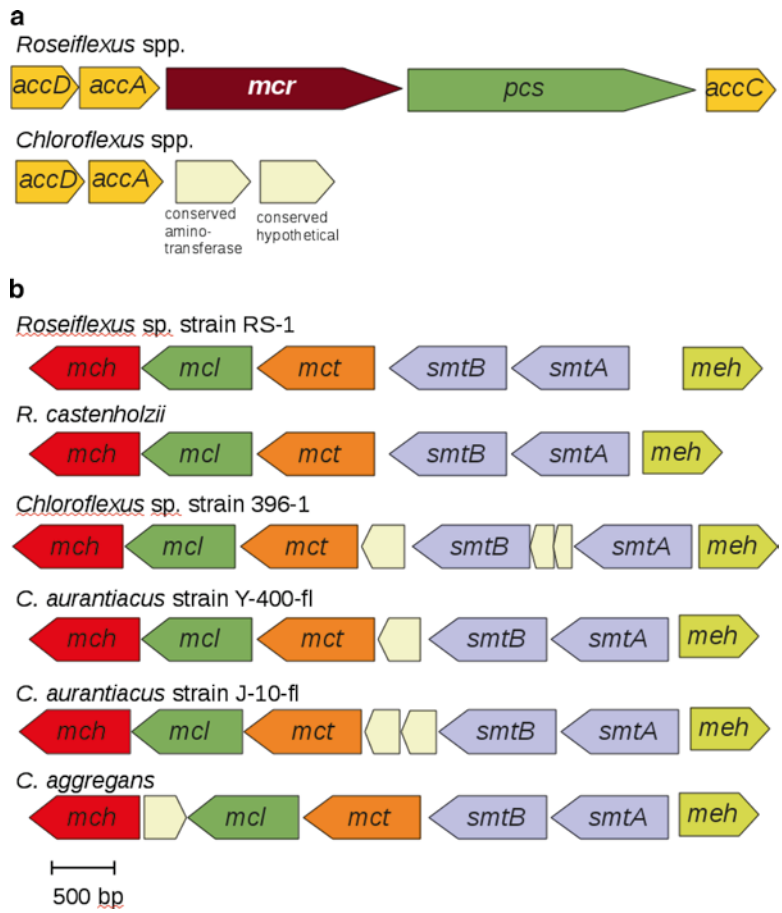


Fig. 3.6 Genomic context of genes involved in the 3-hydroxypropionate pathway. (**a**). Genes encoding subunits of acetyl-CoA carboxylase (*accDA*) are co-localized with malonyl-CoA reductase (*mcr*) and propionyl-CoA synthase (*pcs*) in *Roseiflexus* spp. but not in *Chloroflexus* spp. (**b**). All *Chloroflexales* genomes have a locus in which the mesaconyl-C1-CoA hydratase (*mch*), (S)-malyl-CoA/(beta)-methylmalyl-CoA/(S)-citramalyl-CoA lyase (*mcl*), mesaconyl-CoA C1-C4 transferase (*mct*), succinyl-CoA:(S)-malate-CoA transferase (*smtAB*), and mesaconyl-C4-CoA hydratase (*meh*) are co-localized.

COG 1,840 and is predicted to have CoA transferase activity, but it not yet clear whether the product of this ORF plays a role in cleaving the thioester bond of mesaconyl-CoA (step 12). The *smtA* and *smtB* genes are co-transcribed in *Cfx. aurantiacus* (Friedmann et al., 2006a), and they are coordinately transcribed for *Roseiflexus* spp. *in situ* (see Section III.G). The last steps in the pathway convert R-citramalate to pyruvate, and these steps are catalyzed by succinyl-CoA:D-citramalate CoA-transferase (*sct*, step 13) and R-citramalyl-CoA lyase (*ccl*, step 14) (Friedmann et al., 2006b). These enzymes are stereospecific for the R-conformation, whereas the *smtAB* and *mcl* gene products can also catalyze reactions involving S-citramalate and S-citramalyl-CoA, respectively (Friedmann et al., 2007).

It is currently unknown if other carboxylating enzymes encoded by the FAP genomes, such as phosphoenolpyruvate carboxylase, pyruvate carboxylase, and pyruvate ferredoxin oxidoreductase, play significant roles in inorganic carbon incorporation in these organisms. It is certainly possible that CO_2 fixation via the acyl-CoA carboxylases or these other "anapleurotic" carboxylation reactions occur simultaneously with organic carbon incorporation, analogous to the mixotrophy that has been hypothesized to occur in aerobic anoxygenic phototrophs (Swingley et al., 2007). Additionally, the role of a putative CO dehydrogenase encoded by the FAP genomes, but not found in *H. aurantiacus*, is currently unknown. None of these genomes encode homologs of acetyl-CoA synthase, which

could be coupled with the CO dehydrogenase to allow autotrophic acetate synthesis by the Wood-Ljungdahl pathway. It presently seems more likely that the CO dehydrogenase allows these FAPs to use carbon monoxide as an electron source under oxic or microoxic conditions; the resulting CO_2 would be fixed by the 3-hydroxypropionate pathway. The possibility that C1-metabolism occurs simultaneously with organic carbon utilization, or that these two metabolic processes might occur in one organism in a temporally separated manner, blurs the boundaries for autotrophy, mixotrophy and heterotrophy. In order to understand the roles FAPs play in the carbon-cycling dynamics of the communities in which they occur, it is absolutely critical to understand the actual metabolic capabilities of the organisms in question and the extents to which both autotrophic and/or mixotrophic metabolism occurs.

Finally, each of the genomes discussed here contains genes for the enzymes of the oxidative tricarboxylic acid cycle as well as those for isocitrate-CoA lyase and malyl-CoA synthase, which are required for the glyoxylate shunt. Some FAPs have been shown to assimilate glycolate from their environment (Bateson and Ward, 1988), and glycolate can be converted to glyoxylate by the enzyme glycolate oxidase, which is encoded by the *glcDEF* genes found in all six FAP genomes. These genomes, as well as that of *H. aurantiacus*, also contain homologs of *phbABC* genes. Thus, these organisms should also produce polyhydroxyalkanoic acid, which would likely be an important storage polymer for these organisms (Pierson and Castenholz, 1974; Sirevåg and Castenholz, 1979). Complete Embden-Meyerhoff, gluconeogenesis, and oxidative pentose phosphate pathways are encoded in the genomes of all the *Chloroflexales* strains sequenced thus far. Therefore, to understand the relative contributions of the various branches of autotrophic and heterotrophic carbon metabolism discussed here to the overall carbon metabolism of any given organism will require that one understands the primary carbon sources present in the environment where the organism lives, the availability of electron donors and light for photoautotrophic metabolism, and the temporal gene expression and nutrient acquisition strategies of the organism for each of these resources. All of these considerations illustrate the importance of studying the physiology of these organisms within the natural contexts of their *in situ* microbial communities (see Ward et al., Chapter 1).

G. Ecological Functional Genomics

Although FAP genome sequences have facilitated the characterization of enzymes, electron transfer processes, and biochemical pathways in laboratory isolates, they also provide the basis for application of systems biology approaches to understanding the ecological roles of these organisms in their natural habitats (see Ward et al., Chapter 1 of this volume; Ward, 2006; Ward et al., 1998, 2006). A preliminary analysis of the metatranscriptome of the phototrophic microbial mat community of Mushroom Spring, an alkaline siliceous thermal feature in the Lower Geyser Basin of Yellowstone National Park, illustrates the power of this approach (Liu et al., 2011). The data were obtained from total RNA fractions extracted from mat samples at ~60–62°C collected on July 10–11, 2008, at four time points during a diel cycle: Condition 1, 9:00 PM (sunset); Condition 2, 5:15 AM (pre-dawn); Condition 3, 6:40 (low light, post-dawn); and Condition 4, 8:40 AM (full sunlight). Metagenomics analyses of these mats (Klatt et al., 2011) show that they are less complex than soil or whalefall communities (Tringe et al., 2005) but more complex than acid-mine drainage communities (Tyson et al., 2004). At temperatures above 50°C, these mats are dominated by prokaryotes that belong to four kingdoms/phyla: *Cyanobacteria*, *Chloroflexi*, *Acidobacteria*, and *Chlorobi* (Ward, 2006; Ward et al., 1998, 2006). Based upon the ribosomal rRNA sequences from the metatranscriptome, organisms belonging to these phyla account for about 84% of the stable RNA recovered. One of the heterotrophs in the community, *T. roseum*, accounted for less than 0.5% of the stable RNAs that could be identified, and rRNAs from ~1,000 organisms were detected at levels far below 1% of the total stable RNA fraction.

Past studies have suggested that the FAPs in these mats are predominantly photoheterotrophs, which assimilate acetate and glycolate excreted by the co-existing cyanobacteria (Anderson et al., 1987; Bateson and Ward, 1988). These FAPs might also assimilate carbon dioxide at particular times during the diel cycle, when their environment

provides them with sufficient energy, reductant and inorganic carbon to do so. Isotopically labelled bicarbonate is incorporated at low light levels into the lipid biomarkers of FAPs in the early morning hours and to a lesser extent in the evenings (van der Meer et al., 2000, 2005). Although pure cultures of *R. castenholzii* are unable to grow photoautotrophically (Hanada et al., 2002; van der Meer et al., 2010), these observations suggested that the FAPs in this community might use the 3-hydroxypropionate pathway for uptake and assimilation of some CO_2 *in situ*. We have investigated this question by enumerating the mRNA transcripts of the genes involved in the 3-hydroxypropionate pathway at different time points in the diel cycle to determine if community gene expression patterns, and inferences regarding accompanying metabolism, match hypotheses concerning the timing of carbon fixation by the FAPs in these mats.

The transcripts with highest sequence similarity to 3-hydroxypropionate genes were mostly related to *Roseiflexus* sp. RS-1, which is an organism isolated from the nearby, chemically similar, thermal feature, Octopus Spring (van der Meer et al., 2010), and which has proven to be a relevant reference strain to understand the genomic composition of the FAP populations in these mats (Klatt et al., 2011; Liu et al., 2011). From the metatranscriptomic data, it is apparent that the *Roseiflexus* spp. in the mats transcribe genes involved in the 3-hydroxypropionate cycle during both the day and the night, and that the coordinate transcription of genes encoding enzymes for certain steps in the pathway recapitulate the genomic context of these genes on the *Roseiflexus* sp. RS-1 genome (Fig. 3.6). For example, transcripts were detected for the acyl-CoA carboxylase, malonyl-CoA reductase, and propionyl-CoA synthase putatively encoding the enzymes responsible for the first four steps of the pathway (Fig. 3.5), and the transcript abundances for these genes were highest in the morning under high light. Keeping in mind the uncertain substrate specificity of enzymes putatively involved in steps 1 and 4, different carboxyltransferase subunits were most highly expressed at different times, and in particular the gene encoding the carboxyltransferase-3 subunit, which is predicted to have a preference for propionyl-CoA (step 4) (Klatt et al., 2007), was most highly expressed in the early morning. The genes adjacent to the *mcr* and *pcs* genes cannot be differentiated with respect to substrate preference, and it is possible they are bifunctional and able to catalyze both steps 1 and 4. The regulation patterns of genes encoding enzymes unique to this pathway and to CO_2 assimilation support the hypothesis that *in situ* CO_2 fixation by FAPs mainly occurs in the morning.

Another interesting feature of the regulation patterns for genes involved in this pathway concerns malyl-CoA lyase (step 9 in Fig. 3.5), which, unlike most other genes in the pathway, is up-regulated in the evening. Because this enzyme not only replenishes acetyl-CoA in the first cycle but also connects the two cycles, its different regulation pattern could provide important insights into understanding how this pathway specifically, and carbon metabolism more generally, functions in *Roseiflexus* spp. *in situ* (assuming transcription regulation patterns reflect changes in enzymatic activity). Firstly, the data imply that there should be an abundant source of acetyl-CoA available for the first cycle in the early morning. Given that FAPs assimilate acetate during the evening (van der Meer et al., 2005), it is likely some of these acetate units are subsequently used in the morning as substrates for the carboxylation reactions of the first cycle. Secondly, there might be alternate metabolic outlets from the first cycle other than compounds serving as inputs of carbon for the second cycle. For example, the malyl-CoA produced in the first cycle could be used to build cell materials through conversion to oxaloacetate and pyruvate, or this carbon might be stored in some form. Therefore, it is likely that *Roseiflexus* spp. in the mat performs a complex type of mixotrophy that consumes both acetate and CO_2 rather than the strict autotrophy suggested by the net reactions of the bicyclic 3-hydroxypropionate pathway as shown in Fig. 3.5.

The power of having ecologically relevant genomic data for understanding systems biology and functional genomics in the context of the environment cannot be overstated. The reductionism of modern molecular biology and genomics, which was founded upon studies with pure cultures of model organisms, has found an unlikely bedfellow with the sublime complexity of microbial community ecology. It is through the marriage

of these very different fields that one may eventually unlock the secrets of the evolutionary origins and environmental roles of chlorophototrophic organisms.

IV. Chlorophototrophic Acidobacteria: *Candidatus* Chloracidobacterium thermophilum

Candidatus Chloracidobacterium (Cab.) thermophilum is the first, and currently the only, chlorophototrophic member of the kingdom/phylum *Acidobacteria*, the sixth bacterial phylum shown to contain members capable of chlorophyll-based phototrophy. Phylogenetic analyses suggest that the *Acidobacteria* are a sister group to the *Deltaproteobacteria*, the earliest diverging subdivision of the phylum *Proteobacteria* (Ciccarelli et al., 2006). *Acidobacteria* are a large and highly diverse group of bacteria that are commonly found in soils, sediments, freshwater and marine ecosystems, hot springs, and polluted environments worldwide. Soil isolates do not grow on standard media or grow very slowly; they require several days to weeks to produce visible colonies on complex, low-nutrient media. Because of this, strains of *Acidobacteria* have proven to be very difficult to isolate and to cultivate under laboratory conditions. As a result, cultured representatives currently exist for only five of the 26 subdivisions of this highly diverse phylum. However, complete genome sequences are now available from *Acidobacterium capsulatum*, "*Candidatus* Koribacter versatilis Ellin345" and "*Candidatus* Solibacter usitatus Ellin6076," and *Cab. thermophilum* (Ward et al., 2009; Garcia Costas et al., 2011).

Cab. thermophilum is a poster-child for the genomics era because the discovery of its existence, as well as much of what we now know about its metabolic and biochemical properties, were directly gained through metagenomic, metatranscriptomic, and complete genomic sequence analyses. Initial evidence for the existence of *Cab. thermophilum* came in 2005 from tblastx queries of a large (~200 Mbp), metagenomic sequence database for the chlorophototrophic mat communities of Octopus and Mushroom Springs in Yellowstone National Park. David M. Ward and collaborators had generated this database to study cyanobacterial speciation within these microbial mat communities (for a description of this community see Ward, 2006; Ward et al., 1998, 2006; Klatt et al., 2011). When this database was queried with *Cba. tepidum* PscA, which is the large, apoprotein of its homodimeric type-1 reaction center, three distinctive classes of reaction center sequences were detected: (1) PsaA and PsaB subunits of Photosystem I from cyanobacteria (*Synechococcus* spp.); (2) PscA-like sequences from a GSB that is highly divergent from other characterized GSB; and (3) a third group of sequences that were obviously not derived from cyanobacteria, GSB, or heliobacteria. A nearly full-length example of this latter sequence provided strong evidence for the existence of an organism that produced a type-1 reaction center polypeptide that was much larger than those of other type-1 reaction centers due to the insertion of ~165-amino acids between helices 7 and 8 in the electron transfer domain of the protein. The remainder of the C-terminal electron transfer domain was clearly similar to the electron transfer domains of other type-1 reaction centers, especially the homodimeric ones. Finally, the N-terminal antenna domain was quite divergent from those of the other three types of type-1 reaction centers.

Several overlapping plasmids that collectively accounted for the sequences encoding the *pscA* gene found in the metagenome were recovered and were completely sequenced. The resulting data revealed a putative operon comprising three genes: *pscA*, *pscB* and *fmoA*. The *pscAB* genes encode the two subunits of a homodimeric type-1 reaction center and the *fmoA* gene encodes the BChl *a*-binding Fenna-Matthews-Olson (FMO) antenna protein. Phylogenetic analyses of the genes in the flanking sequences of this operon strongly suggested that these genes came from a member of the phylum *Deltaproteobacteria* or *Acidobacteria*. By mapping a library of end-sequenced, bacterial artificial chromosome (BAC) inserts onto an assembly of the metagenome data, a BAC clone was identified that was predicted to encode the *pscAB-fmoA* operon, *recA* gene as well as a rRNA cluster. This BAC clone was recovered, and its insert was completely sequenced (GenBank Accession = EF531339). The data obtained unequivocally demonstrated that the

pscAB-fmoA genes came from an organism belonging to the kingdom/phylum *Acidobacteria* (Bryant et al., 2007).

These sequence data further established criteria that could be used to identify and bin the assembled metagenomic sequences that were likely to be associated with this unknown bacterium. The mole% G+C content of the BAC clone insert was ~60%, and as noted above, the housekeeping genes grouped phylogenetically with those of *Acidobacteria* or *Deltaproteobacteria*. Metagenomic assemblies meeting these criteria were identified and binned together, and the annotation information for these sequences was examined to predict physiological properties of the new chlorophototroph, which was named *Candidatus* Chloracidobacterium (Cab.) thermophilum. The presence of a *bchU* gene implied that this organism would synthesize BChl *c* (Maresca et al., 2004), and the presence of *pscA* and *fmoA* further suggested that the organism should synthesize BChl *a*. All organisms that can synthesize BChl *c* also produce chlorosomes, and the identification of the gene encoding the BChl *a*-binding baseplate protein, CsmA, of chlorosomes validated this predication. Although at the time of these analyses all chlorosome-containing organisms grew phototrophically under anoxic conditions, the presence of an *acsF* gene, encoding an oxygen-dependent, Mg-Proto IX monomethyl ester oxidative ring cyclase, strongly implied that *Cab. thermophilum* was a microaerophile, or possibly even a full aerobe.

Armed with these physiological and biochemical clues, and a healthy dose of serendipity, "forensic PCR analyses" with enrichment cultures revealed that *Cab. thermophilum* did not grow under anaerobic conditions and that it indeed grows under oxic (and perhaps micro-oxic) conditions. This discovery prompted a survey of existing cyanobacterial enrichment cultures, which were known to contain single strains of cyanobacteria but which also carried uncharacterized bacterial contaminants from the microbial mats (Allewalt et al., 2006). One of these *Synechococcus* sp. enrichment cultures (for strain JA-2-3B′a), which had been stably maintained in the Ward laboratory for ~3.5 years and whose genome has been sequenced (Bhaya et al., 2007), was also found to contain a population of *Cab. thermophilum*. By replacing the nitrate in the medium with ammonium and by adding an equimolar mixture of six carbon sources (pyruvate, lactate, acetate, glycolate, butyrate, and succinate) and the potent Photosystem II inhibitor atrazine to cultures, *Cab. thermophilum* was isolated in co-culture enrichment with two heterotrophic bacteria, an *Anoxybacillus* sp. and *Meiothermus* sp. Although pure cultures of the latter two organisms can be isolated, *Cab. thermophilum* does not form isolated colonies on plates and does not survive dilution to extinction. However, the co-culture of these three organisms has now been stably maintained for more than 3 years. Unfortunately, *Cab. thermophilum* never grows to high cell densities and usually undergoes only a few (~4–5) doublings after transfer to fresh medium.

Because *Cab. thermophilum* is the only BChl-containing member of the enrichment culture, it was possible to demonstrate conclusively that *Cab. thermophilum* is a phototroph. When serial subcultures were grown in light or darkness, *Cab. thermophilum* grew much faster in the light and at best grew very slowly, if at all, in the dark (Bryant et al., 2007). As predicted, *Cab. thermophilum* synthesized BChl *c* as its major light-harvesting pigment and contained smaller amounts of BChl a_P and Chl a_{PD}. Like chlorosome-producing members of the *Chloroflexales*, the BChl *c* is esterified with several alcohols, the most abundant of which is stearol (octadecanol) and with smaller amounts of hexadecanol and phytol (A. Garcia Costas, Y. Tsukatani, G. Oostergetel G, E. J. Boekema, and D. A. Bryant, in preparation). Like the BChl *c* found in strains of *Chlorobiales*, the C-8 and C-12 side chains are methylated. The most abundant BChl *c* homolog under normal laboratory growth conditions was shown by HPLC analyses and mass spectrometry to be [8-isobutyl, 12-methyl]-BChl c_S esterified with stearol. In thin sections viewed by transmission electron microscopy, chlorosomes are only observed on the inner surfaces of the lateral-wall cytoplasmic membrane and are never observed at the poles of cells.

Chlorosomes (Bryant et al., 2007; A. Garcia Costas and D. A. Bryant, unpublished results) and FMO protein (Tsukatani et al., 2010) have been isolated from cells and characterized. *Cab. thermophilum* chlorosomes are very large (100–300 nm in length; diameter about 30–40 nm). As observed with chlorosomes of GSB, the

fluorescence emission is strongly quenched under oxidizing conditions, and this is correlated with the presence of menaquinone-8 in the chlorosomes. Surprisingly, two keto-carotenoids, echinenone and canthaxanthin, are the major carotenoids in *Cab. thermophilum* chlorosomes. Consistent with the notion that *Cab. thermophilum* is an aerobe and based upon its complete genome sequence, an oxygen-dependent, flavoprotein ketolase (CrtO) is predicted to ketolate β-carotene to produce these two carotenoids. Chlorosome envelopes of *Cab. thermophilum* contain ~7 polypeptides, including CsmA, an Fe/S protein that is distantly related to a protein encoded in the genome of *Chp. thalassium*, peptidyl-proline *cis-trans* isomerase of the cyclophilin family. The other chlorosome envelope proteins do not appear to have any significant similarity to other proteins in databases.

The FMO protein of *Cab. thermophilum* was recently purified to homogeneity and initially characterized (Tsukatani et al., 2010). Based upon comparisons of the amino acid sequence of FmoA with the proteins from *Chlorobiales*, the *Cab. thermophilum* FMO protein should bind the same number of BChl *a* molecules as the proteins previously characterized from GSB strains. The purified protein only bound BChl *a*, and its absorption maximum occurred at 797 nm, with shoulders at 810 nm and 830 nm. At room temperature, this FMO protein exhibited fluorescence emission maxima at 815 and 832 nm, but there was a single emission peak at 834 nm at 77 K. This behavior suggests this FMO protein has two pools of BChls that are not strongly coupled.

Preliminary analyses of the binned metagenomic sequence data did not provide any evidence for carbon fixation or obvious electron transport proteins that might provide an inorganic source of electrons to support photoautotrophic growth. Thus, we had tentatively concluded that *Cab. thermophilum* was probably a photoheterotroph. Recently, the sequencing of *Cab. thermophilum* genome was completed (Garcia Costas et al., 2011). The *Cab. thermophilum* genome comprises two circular chromosomes of 2.6 Mb and 1.0 Mb. The smaller chromosome encodes essential genes, *e.g.*, genes encoding enzymes of BChl biosynthesis, so it is clear that this smaller DNA molecule is a chromosome and not a megaplasmid. Only a single set of ribosomal RNA genes is encoded in the genome, and these genes occur on the larger chromosome. The genome encodes a complete Type 1 NADH dehydrogenase and Cyt oxidase. However, *Cab. thermophilum* does not have a typical Cyt bc_1 or b_6f complex to oxidize menaquinol. Instead, as is the case in *Cfx. aurantiacus* (Yanyushin et al., 2005; see Section III.D above), the *Cab. thermophilum* genome encodes alternative complex III. In agreement with the initial observations from the metagenome, no genes for any of the known pathways for carbon dioxide fixation are encoded in this genome. Moreover, key genes involved in sulfur oxidation (*sqr*, *sox*, *dsr*) and assimilatory sulfate reduction (*cysHIG*) are absent from the genome. These observations strongly suggest that *Cab. thermophilum* obtains its reduced sulfur (or sulfur-containing amino acids) from the metabolic activities of either *Anoxybacillus* sp. or *Meiothermus* sp., or both. Finally, preliminary analyses of the *Cab. thermophilum* genome show that the enzymes required for the synthesis of leucine, isoleucine, and valine are missing (other than transaminases and tRNA synthetases), but that the genome encodes an ABC-type transporter for branched-chain amino acids and complete degradation pathways for these same three amino acids are present. These observations provide important clues for the apparent metabolic co-dependence of *Cab. theromophilum* with *Anoxybacillus* and *Meiothermus* spp. in the enrichment culture.

One might question whether *Cab. thermophilum* is truly a phototroph *in situ*. Results from a preliminary metatranscriptome study by pyrosequencing showed that *pscA* is the most highly transcribed gene in *Cab. thermophilum* cells in the microbial mat (Liu et al., 2011). Other genes relating to the formation of chlorosomes and biosynthesis of chlorophylls were also among the most highly expressed genes near sunset, the time when *Cab. thermophilum* was most active transcriptionally of the four diel conditions that were sampled (sunset, pre-dawn, post-dawn/low light; morning/high light). These results strongly suggest that *Cab. thermophilum* is a phototroph *in situ* and that it may be most metabolically active during periods when cyanobacterial oxygen evolution rates are not maximal and when the mat is microoxic or anoxic. Further in-depth transcription profiling *in situ* will be required to place more of the metabolic activities of this organism into an ecological context.

V. Evolution of Chlorophyll Biosynthesis and Photosynthetic Apparatus

The evolution of photosynthesis was one of the most important events in the history of biology, because this process allowed biological energy production to be coupled to an inexhaustible energy source. By coupling a solar energy module with a biochemical module for carbon dioxide (and often dinitrogen) reduction, ancestral organisms eventually gained both energy and nutritional independence. The emergence of primary producers, and the eventual development of oxygenic photosynthesis, inevitably changed the trajectory of the evolution of life on Earth.

Given the central importance of photosynthesis in biology, it is not surprising that much attention and speculation has focused on when and how photosynthesis might have evolved. Phototrophy, *i.e.*, the ability to utilize light as an energy source, has evolved at least twice and occurs by two fundamentally different mechanisms. Retinalophototrophs, organisms that synthesize bacteriorhodopsin or proteorhodopsin, occur in the domains *Archaea* and *Bacteria* and can directly couple proton translocation to light absorption through light-induced conformational changes driven by retinal photo-isomerization. No currently known retinalophototroph couples this type of phototrophy to carbon dioxide fixation (Bryant and Frigaard, 2006). One reason for this might be the low absorption cross-section of retinal-based systems, as very high light fluxes would be required in order to support energy-intensive, autotrophic metabolism by this mechanism. Alternatively, the inability of retinal-based systems to catalyze electron transfer reactions and generate low-potential reductants might explain the absence of autotrophs among the retinalophototrophs. Chlorophototrophs, organisms that have photochemical reaction centers and that perform Chl-based phototrophy (Bryant and Frigaard, 2006), evolved to overcome both of these limitations. Recent developments, including the discovery of *Cab. thermophilum* and the extensive sequencing of chlorophototroph genomes, provides new insights into the evolution of photosynthesis.

Considering that photochemical reaction centers are the essential engines of Chl-based phototrophy, the evolutionary origins of Chl and these reaction centers are inexorably linked. Several rationales for how and why reaction centers could have evolved have been proposed. Perhaps the most plausible possibility is that the primitive reaction center was a porphyrin-containing membrane protein that catalyzed light-induced electron transfer from some abundant donor, *e.g.*, ferrous iron, to a suitable biological acceptor (*e.g.*, an Fe/S cluster) across a biological membrane (Olson and Pierson, 1987; Olson, 1999). Consistent with this possibility, examples of phototrophic organisms that can oxidize ferrous iron (photoferrotrophy) still exist on Earth today (Widdel et al., 1993; Heising et al., 1999). Another possibility is that reaction centers evolved from proteins used for photo-sensing of near-infrared radiation from thermal vents (Nisbet et al., 1995). In support of this concept, there are biliverdin-containing bacteriophytochromes that are used to sense near-infrared light (Giraud and Verméglio, 2008). Alternatively, Mulkidjanian and Junge (1997) suggested that reaction centers evolved from proteins that provided photoprotection from ultraviolet radiation by absorption and dissipation of this harmful light, although there is no direct experimental evidence to support this claim.

Although the evolutionary ancestor(s) of reaction centers is not known, Meyer (1994) and Xiong and Bauer (2002a, b) have suggested that reaction centers (and light-harvesting proteins) are descendents from Cyt *b*, a broadly distributed, heme-binding protein, which could provide a connection between respiratory and light-driven electron transport. Two reaction center families, designated type-1 and type-2, are currently recognized (Golbeck, 1993). Type-1 reaction centers produce weak oxidants and strong reductants; their electron transport chains terminate with three, strongly reducing [4Fe-4S] clusters, denoted F_X, F_A, and F_B. Type-2 reaction centers produce a strong oxidant and a weak reductant; their electron transport chains terminate with a (bacterio) pheophytin ((B)Phe) and two quinones, denoted Q_A and Q_B. All reaction centers contain a bifurcated chain of six (B)Chls and/or (B)Phe, which initiate with a "special pair" of (B)Chls and which are ligated at the interface between two polypeptide chains, which can be homodimers or heterodimers. The structural similarities of the cofactors and polypeptides in the two reaction center types strongly suggest that photochemical reaction centers had a single evolutionary origin and that this family of proteins expanded by gene duplication and divergence (Schubert et al., 1998;

Mix et al., 2005; Sadekar et al., 2006). Although some arguments can be made to support either scenario, several lines of evidence suggest that homodimeric type-1 reaction centers, derived from a protein with 11 transmembrane α-helices, may be the ancestral type of reaction center. An ancestral homodimeric reaction center could have given rise to heterodimeric type-1 reaction centers, *e.g.*, cyanobacterial Photosystem I, by gene duplication and divergence. Homodimeric reaction centers could also have given rise to heterodimeric type-2 reaction centers by gene duplication and fission of the reaction center polypeptide into polypeptides with six and five transmembrane α-helices. This scenario is supported by a structure-based phylogenetic analysis by Sadekar et al. (2006), who suggested that homodimeric type-1 reaction centers were the earliest bacterial reaction centers. Similarly, Mix et al. (2005) concluded that horizontal gene transfer was not required to explain the current distribution of type-1 reaction centers among chlorophototrophs and that homodimeric type-1 reaction centers were the ancestral type. The discovery of *Cab. thermophilum*, which has homodimeric type-1 reaction centers, and whose PscA sequence is highly divergent from other PscA of *Chlorobiales* and heliobacteria, provides additional support for the idea that type-1 reaction centers are the ancestral type of reaction centers (Bryant et al., 2007).

Numerous studies have attempted to use phylogenetic reconstruction methods to gain insights into the origins of photosynthesis. Although there is negligible sequence similarity between the electron transfer domains of type-1 and type-2 reaction centers, structural similarities nevertheless exist (Schubert et al., 1998), and Sadekar et al. (2006) took advantage of this to build phylogenetic trees. Attempts to use whole-genome methods have met with mixed success, principally because the number of shared genes across all groups of chlorophototrophs is very small, and these genes tend to reflect the phylogenetic relationships of the traits of these organisms other than photosynthesis (Raymond et al., 2002; Mulkidjanian et al., 2006). Photochemical reaction centers contain two types of pigments: carotenoids and (B)Chls. Carotenoids are non-essential components of reaction centers, particularly under anoxic conditions, and the genes for their synthesis are broadly distributed among all domains of life and appear to have been subject to extensive horizontal transfer (Phadwal, 2005). Given this situation, the enzymes of (B)Chl biosynthesis, have been extensively studied by phylogenetic methods to gain insights into the origins of photosynthesis (*e.g.,* Xiong et al., 2000; Xiong and Bauer, 2002a, b; see below).

A previous discussion concerning the evolution of (B)Chl biosynthesis posed the question, "what color was photosynthesis originally?" (Xiong, 2006). Although this question is certainly an important one, in some ways it overlooks an obvious point: no matter how inefficient the original photochemical reaction center might have been mechanistically, as soon as functioning reaction centers evolved, those organisms that could use light energy would have gained a large selective advantage over non-phototrophic organisms, but just as surely they would have immediately begun to compete with their neighbors for available photons. This selection pressure would have inevitably led to a rapid chemical diversification of the light-absorbing molecules within the reaction centers, through extension and branching of the tetrapyrrole biosynthesis pathway. As noted in Section II.C.2, at least in the terminal stages of the (B)Chl biosynthetic pathway, the Granick hypothesis that pathways evolve as organisms evolve seems very likely to be correct (Granick, 1957, 1965). Furthermore, this competition for light would provide the selection mechanism for the evolution of antenna complexes to trap and deliver excitation energy to the reaction centers. Although details of the pathways leading to some types of (B)Chls are still not fully elucidated, pathways leading from Chlide to BChl *a* (and then to BChl *b* and *g*), to BChl *c*, *d* and *e*; to Chl *b*; and Chl *d* are known or proposed (Gomez Maqueo Chew and Bryant, 2007b; Tanaka and Tanaka, 2007). In a manner analogous to the position of pyruvate in fermentative metabolism, Chlide *a,* and not BChlide *a*, is the obvious hub compound of (B)Chl biosynthesis (Gomez Maqueo Chew and Bryant, 2007b; see Fig. 3.3). There is no particular reason to assert that reaction centers, whether simple or complex, came into existence with BChl *a* rather than Chl *a*, as the initial and essential light-absorbing cofactor. The presence of Chl *a* or derivatives thereof in extant homodimeric type-1 reaction centers is also particularly informative. It is easy to imagine that the extension of the biosynthetic pathway for Chlide *a* synthesis to produce BChlide *a* led to

reaction centers with different absorption behavior, in much the same way that Chl *d* is now known to produce variant forms of Photosystem I and II in *Acaryochloris marina* (Ohashi et al., 2008). Similarly, the homodimeric type-1 reaction centers of *Chlorobiales*, heliobacteria, and probably *Cab. thermophilum* still employ Chl a_{PD} or 8-OH-Chl *a* as primary electron acceptors, although the bulk BChls in these reaction centers have been replaced by BChl *a* or BChl *g* (Hauska et al., 2001; Heinnickel and Golbeck, 2007). There is no *a priori* reason to assume that these reaction centers came into existence this way, but it seems much more likely that they originally may have contained only Chl *a* (or one of its biosynthetic precursors).

Some of the enzymes of (B)Chl biosynthesis are very highly conserved, and phylogenetic analyses of these sequences were performed long before the recent explosion of genomic information (*e.g.*, see Xiong et al., 2000; Xiong and Bauer, 2002a, b; Xiong, 2006). The evolution of (B)Chl biosynthesis should be reexamined in light of the wealth of genomic sequence information that has been obtained for chlorophototrophs in recent years. As the number of chlorophototroph genomes has increased, statistical support for phylogenetic trees has correspondingly increased. Figure 3.7 shows a phylogenetic tree for the concatenated sequences of Mg-Proto IX chelatase, ChlHID/BchHID, which catalyzes the first committed step in (B)Chl biosynthesis. This tree provides very strong statistical support for three groups of chlorophototrophs, each containing members of two major taxa containing chlorophototrophs. However, the branching order for these three subgroups cannot be determined on the basis of these data. The three groups are: (1) *Chloroflexales* and *Chlorobiales*; (2) *Acidobacteria* and *Proteobacteria*; and (3) *Firmicutes* (heliobacteria) and *Cyanobacteria*. Within each of these three subgroups of chlorophototrophs, organisms that have homodimeric type-1 reaction centers diverge early from organisms that have heterodimeric reaction centers in the other phylum. This observation is in agreement with those of others, who have suggested that homodimeric type-1 reaction centers were the earliest bacterial reaction centers (Sadekar et al., 2006; Mix et al., 2005; see above). Given the strong similarity of the type-2 reaction centers of *Chloroflexales* to those of purple bacteria (*Proteobacteria*), it seems possible that *Chloroflexales* might have originally had homodimeric type-1 reaction centers like members of the *Chlorobiales*. These type-1 reaction centers might have been displaced later by the horizontal acquisition of *pufLM* and *pufBAC*-like genes (for B806–866) at some later time. This could partly account for why phylogenetic trees based on 16S rRNA and enzymes of (B)Chl biosynthesis are not congruent and contribute to the notion that the *Chloroflexi* have mosaic physiological and metabolic properties (see Section III.A). An interesting alternative viewpoint is that heterodimeric type-2 reaction centers and LH-1-like (B806–866) light-harvesting genes evolved in *Chloroflexi* and were subsequently laterally transferred to the *Proteobacteria*. The position of *Cab. thermophilum* as the earliest diverging organism of the *Acidobacteria-Proteobacteria* clade agrees very well with the sister-group relationship of these two phyla (Ciccarelli et al., 2006) and is also in agreement with the "type-1 reaction centers early" hypothesis. However, it will be necessary to obtain sequence information from more examples of acidobacterial chlorophototrophs before this relationship can be strongly supported.

Figure 3.8 shows a similar phylogenetic analysis of concentrated sequences encoding light-independent PChlide reductase (BchL, BchN, BchB; DPOR) and Chlide reductase (BchX, BchY, BchZ; COR), using nitrogenase (NifH, NifD, NifK) sequences as outgroup. An obvious horizontal gene transfer event has occurred for the marine *Synechococcus* and *Prochlorococcus* spp., which have a PChlide reductase that is clearly derived from a purple bacterium (*Proteobacteria*). The upper portion of the tree for PChlide reductase contains three subgroups. The earliest diverging lineage is that containing the *Chloroflexales* and *Chlorobiales,* and the *Firmicutes* (heliobacteria) and *Cyanobacteria* diverge next. *Cab. thermophilum* (*Acidobacteria*) and *Proteobacteria* diverge last, with *Cab. thermophilum* being the earliest diverging member of this clade. The lower portion of the tree for Chlide reductase is generally very similar to the upper part. The exception is the position of

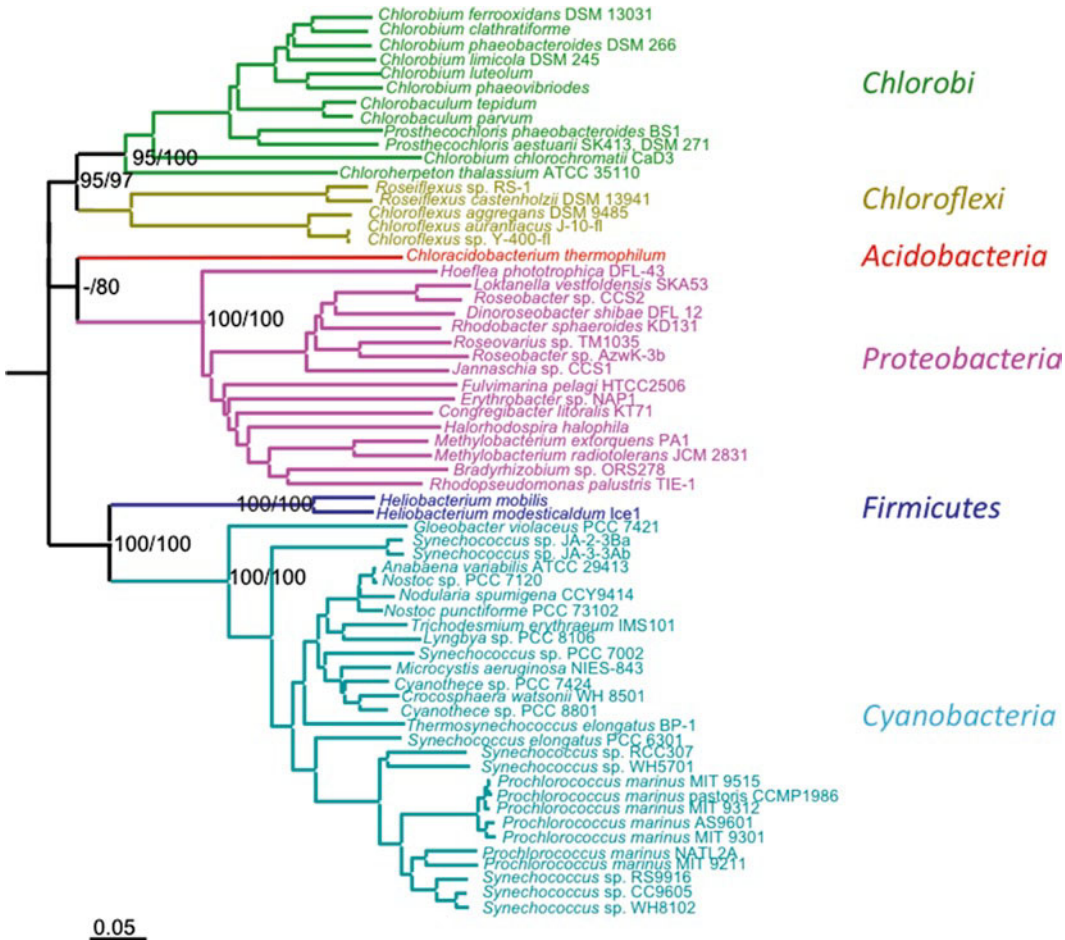


Fig. 3.7. Phylogenetic relationships among photosynthetic bacteria based on the concatenated BchIDH/ChlIDH protein sequences. Mg-chelatase family proteins (NP_614634, NP_613479) from *Methanopyrus kandleri* AV19 (NC_003551) were used as outgroup. Sets of homologous protein sequences were aligned using MUSCLE (Edgar, 2004) and conserved regions of the resulting multiple sequence alignments were identified with GBlocks (Castresana, 2000). The NJ method implemented in MEGA 4.0 (Tamura et al., 2007) and the ML method implemented in PhyML 2.4.4 (Guindon and Gascuel, 2003) with 500 replicates were used to construct phylogenetic tree. Numbers at nodes indicate the NJ and ML bootstrap values, respectively; a dash indicates an unsupported branch. The six phyla/kingdoms containing chlorophototrophic members are indicated at the right and the names are color-coded.

Heliobacterium modesticaldum, the position of which is questionable due to possible long-branch attraction in the absence of any additional heliobacterial or cyanobacterial sequences. Cyanobacteria obviously do not appear in the lower portion of the tree because they do not synthesize BChl *a*. As for the analysis shown in Fig. 3.7, each of the three major subgroups contains early-diverging members with type-1 reaction centers. It is also worth noting that early diverging members of two of the three subdivisions, and three of the major taxa containing phototrophs, utilize chlorosomes for light harvesting. This observation suggests that chlorosomes might be a much more ancient type of antenna system than had previously been appreciated. The synthesis of the simplest self-assembling BChl, BChl *d*, requires only two enzymes beyond those required for the synthesis of Chl *a* (Gomez Maqueo Chew and Bryant, 2007b). Because one of these enzymes, BchF, also participates in the synthesis of BChl *a*, only a single enzyme,

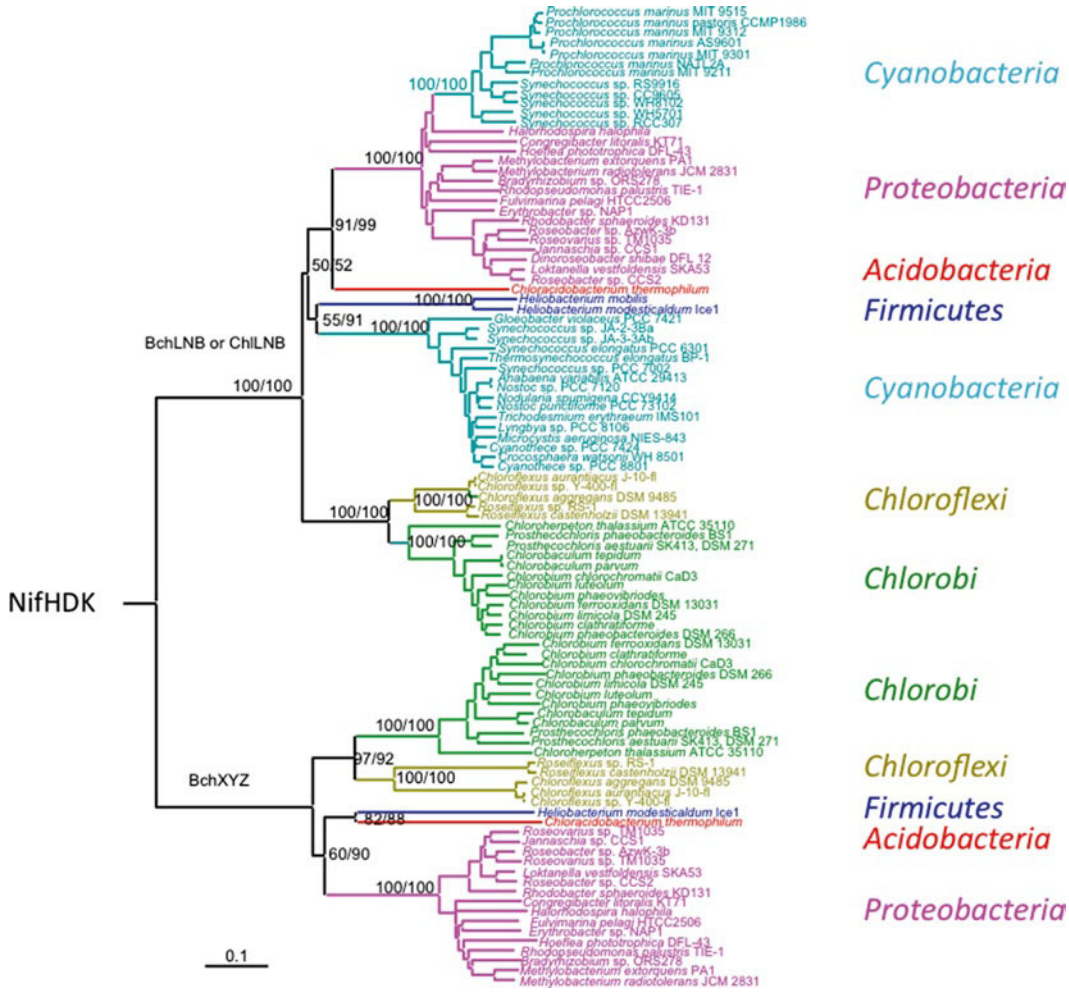


Fig. 3.8. Phylogenetic relationships among photosynthetic bacteria based on the concatenated BchLNB/ChlLNB/BchXYZ protein sequences. Nitrogenase proteins (NifHDK) from *Methanococcus maripaludis* C5 (NC_009135) were used as outgroup. Sets of homologous protein sequences were aligned using MUSCLE (Edgar, 2004) and conserved regions of the resulting multiple sequence alignments were identified with GBlocks (Castresana, 2000). The NJ method implemented in MEGA 4.0 (Tamura et al., 2007) and the ML method implemented in PhyML 2.4.4 (Guindon and Gascuel, 2003) with 500 replicates were used to construct phylogenetic tree. Numbers at nodes indicate the NJ and ML bootstrap values, respectively. The six phyla/kingdoms containing chlorophototrophic members are indicated at the right and the names are color-coded.

BciC (CT1077) is actually required to produce BChl *d* if the organism could already produce BChl *a* (see Fig. 3.3). Of course, additional antenna proteins, including the BChl *a*-binding proteins FMO and CsmA, as well as other components of the chlorosome envelope, would be required to produce a hydrophobic compartment that favors BChl *d* aggregate formation. The selective pressure to develop such antenna systems might have come in response to the photosynthetic activities of cyanobacteria, as green bacteria retreated into low light, sulfidic environments because of rising oxygen levels in the atmosphere.

Acknowledgments

Original research described in this chapter was supported by grants to D. A. B. by the U. S. Department of Energy (DE-FG02-94ER20137), the U. S. National Science Foundation (MCB-0523100), and National Aeronautics and Space Administration: Astrobiology (NNX09AM87G;

D. M. W., P. I.). D. A. B. and D. M. W. additionally and gratefully acknowledge support from the U.S. Department of Energy (BER), as part of BER's Genomic Science Program (GSP). This contribution originates from the GSP Foundational Scientific Focus Area (FSFA) at the Pacific Northwest National Laboratory (PNNL), contract no. 112443. N.-U. F. acknowledges support from the Danish Science Research Council. D. A. B. and J. O. thank the Joint Genome Institute for their assistance in generating draft genomes of some *Chlorobiales* and completed genomes of *Chloroflexales* described here. Research of J. O. was supported by grants from the Deutsche Forschungsgemeinschaft (DFG Ov20/8-1 through 8–3, and Ov20/10-1 and 10–2). D. M. W. acknowledges support from NASA (NAG5-8824, 13468 and NX09AM87G), NASA's support of the Montanta State Thermal Biology Institute (NAG5-8807), and the U. S. National Science Foundation (EF-0328698).

References

Adhikari ND, Orler R, Chory J, Froehlich JE and Larkin RM (2009) Porphyrins promote the association of GENOMES UNCOUPLED 4 and a Mg-chelatase subunit with chloroplast membranes. J Biol Chem 284: 24783–24796

Alber BE and Fuchs G (2002) Propionyl-coenzyme A synthase from *Chloroflexus aurantiacus*, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO_2 fixation. J Biol Chem 277: 12137–12143

Allewalt JP, Bateson MM, Revsbech NP, Slack K and Ward DM (2006) Effect of temperature and light on the growth of and photosynthesis by *Synechococcus* isolates typical of those predominating in the Octopus Spring microbial mat community of Yellowstone National Park. Environ Microbiol 72: 544–550

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410

Anderson KL, Tayne TA and Ward DM (1987) Formation and fate of fermentation products in hot spring cyanobacterial mats. Appl Environ Microbiol 53: 2343–2352

Arellano JB, Torkkeli M, Tuma R, Laurinmäki P, Melø TB, Ikonen TP, Butcher SJ, Serimaa RE and Pšenčík J (2008) Hexanol-induced order-disorder transitions in lamellar self-assembling aggregates of bacteriochlorophyll *c* in *Chlorobium tepidum* chlorosomes. Langmuir 24: 2035–2041

Azai C, Tsukatani Y, Harada J and Oh-oka H (2009) Sulfur oxidation in the mutants of the photosynthetic green sulfur bacterium *Chlorobium tepidum* devoid of cytochrome *c*-554 and SoxB. Photosynth Res 100: 57–65

Baniulis D, Yamashita E, Zhang H, Hasan SS and Cramer WA (2008) Structure-function of the cytochrome b_6f complex. Photochem Photobiol 84: 1349–1358

Bateson MM and Ward DW (1988) Photoexcretion and fate of glycolate in a hot spring cyanobacterial mat. Appl Environ Microbiol 54: 1738–1743

Beatty JT, Overmann J, Lince MT, Manske AK, Lang AS, Blankenship RE, Van Dover CL, Martinson TA and Plumley FG (2005) An obligately photosynthetic bacterial anaerobe from a deep-sea hydrothermal vent. Proc Natl Acad Sci USA 102: 9306–9310

Ben Shem A, Frolow F and Nelson N (2004) Evolution of Photosystem I—from symmetry through pseudo-symmetry to asymmetry. FEBS Lett 564: 274–280

Berry EA, Huang S, Saechao LK, Pon NG, Valkova-Valchanova M and Daldal F (2004) X-ray structure of *Rhodobacter capsulatus* cytochrome bc_1: comparison with its mitochondrial and chloroplast counterparts. Photosynth Res 81: 251–275

Bhaya D, Grossman AR, Steunou AS, Khuri N, Cohan FM, Hamamura N, Melendrez MC, Bateson MM, Ward DM and Heidelberg JF (2007) Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. ISME J 1: 703–713

Blankenship RE (1992) Origin and early evolution of photosynthesis. Photosynth Res 33: 91–111

Blankenship RE, Feick R, Bruce BD, Kirmaier C, Holten D and Fuller RC (1983) Primary photochemistry in the facultative green photosynthetic bacterium *Chloroflexus aurantiacus*. J Cell Biochem 22: 251–261

Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B and Shao Y (1997) The complete genome sequence of *Escherichia coli* K-12. Science 277: 1453–1462

Bollivar DW (2006) Recent advances in chlorophyll biosynthesis. Photosynth Res 90: 173–194

Bollivar DW, Suzuki JY, Beatty JT, Dobrowolski JM and Bauer CE (1994) Directed mutational analysis of bacteriochlorophyll *a* biosynthesis in *Rhodobacter capsulatus*. J Mol Biol 237: 622–640

Bröcker MJ, Virus S, Ganskow S, Heathcote P, Heinz DW, Schubert WD, Jahn D and Moser J (2008a) ATP-driven reduction by dark-operative protochlorophyllide oxidoreductase from *Chlorobium tepidum* mechanistically resembles nitrogenase catalysis. J Biol Chem 283: 10559–10567

Bröcker MJ, Wätzlich D, Uliczka F, Virus S, Saggu M, Lendzian F, Scheer H, Rüdiger W, Moser J and Jahn D (2008b) Substrate recognition of nitrogenase-like dark operative protochlorophyllide oxidoreductase from *Prochlorococcus marinus*. J Biol Chem 283: 29873–29881

Bröcker MJ, Schomburg S, Heinz DW, Jahn D, Schubert W-D and Moser J (2010) Crystal structure of the nitrogenase-like dark operative protochlorophyllide oxidoreductase catalytic complex $(ChlN/ChlB)_2$. J Biol Chem 285: 27336–27345

Bruce BD, Fuller RC and Blankenship RE (1982) Primary photochemistry in the facultatively aerobic green photosynthetic bacterium *Chloroflexus aurantiacus*. Proc Natl Acad Sci USA 79: 6532–6536

Brugna M, Albouy D and Nitschke W (1998) Diversity of cytochrome *bc* complexes: example of the Rieske protein in green sulfur bacteria. J Bacteriol 180: 3719–3723

Bryant DA and Frigaard N-U (2006) Prokaryotic photosynthesis and phototrophy illuminated. Trends Microbiol 14: 488–496

Bryant DA, Vassilieva EV and Frigaard N-U (2002) Selective protein extraction from *Chlorobium tepidum* chlorosomes using detergents. Evidence that CsmA forms multimers and binds bacteriochlorophyll *a*. Biochemistry 41: 14403–14411

Bryant DA, Garcia Costas AM, Maresca JA, Chew AG, Klatt CG, Bateson MM, Tallon LJ, Hostetler J, Nelson WC, Heidelberg JF and Ward DM (2007) *Candidatus* Chloracidobacterium thermophilum: an aerobic phototrophic *Acidobacterium*. Science 317: 523–526

Buchanan BB and Arnon DI (1990) A reverse Krebs cycle in photosynthesis: consensus at last. Photosynth Res 24: 47–53

Burke DH, Alberti M and Hearst JE (1993a) The *Rhodobacter capsulatus* chlorin reductase-encoding locus, *bchA*, consists of three genes, *bchX*, *bchY*, and *bchZ*. J Bacteriol 175: 2407–2413

Burke DH, Alberti M and Hearst JE (1993b) *bchFNBH* bacteriochlorophyll synthesis genes of *Rhodobacter capsulatus* and identification of the third subunit of light-independent protochlorophyllide reductase in bacteria and plants. J Bacteriol 175: 2414–2422

Caldwell PE, MacLean MR and Norris PR (2007) Ribulose bisphosphate carboxylase activity and a Calvin cycle gene cluster in *Sulfobacillus* species. Microbiology 153: 243–252

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17: 540–552

Chan LK, Morgan-Kiss RM and Hanson TE (2009) Functional analysis of three sulfide-quinone oxidoreductase homologs in *Chlorobaculum tepidum*. J Bacteriol 191: 1026–1034

Chung S, Frank G, Zuber H and Bryant DA (1994) Genes encoding two chlorosome components from the green sulfur bacterium *Chlorobium vibrioforme* strain 8327D and *Chlorobium tepidum*. Photosynth Res 41: 261–275

Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B and Bork P (2006) Toward automatic reconstruction of a highly resolved tree of life. Science 311: 1283–1287

Coomber SA, Chaudhri M, Connor A, Britton G and Hunter CN (1990) Localized transposon Tn5 mutagenesis of the photosynthetic gene cluster of *Rhodobacter sphaeroides*. Mol Microbiol 4: 977–989

Dahl C, Engels S, Pott-Sperling AS, Schulte A, Sander J, Lübbe Y, Deuster O and Brune DC (2005) Novel genes of the *dsr* gene cluster and evidence for close interaction of Dsr proteins during sulfur oxidation in the phototrophic sulfur bacterium *Allochromatium vinosum*. J Bacteriol 187: 1392–1404

Dahl C, Schulte A, Stockdreher Y, Hong C, Grimm F, Sander J, Kim R, Kim S-H and Shin DH (2008) Structural and molecular genetic insight into a widespread sulfur oxidation pathway. J Mol Biol 31: 1287–1300

Davison PA, Schubert HL, Reid JD, Iorg CD, Heroux A, Hill CP and Hunter CN (2005) Structural and biochemical characterization of Gun4 suggests a mechanism for its role in chlorophyll biosynthesis. Biochemistry 44: 7603–7612

Diacovich L, Mitchell DL, Pham H, Gago G, Melgar MM, Khosla C, Gramajo H and Tsai SC (2004) Crystal structure of the beta-subunit of acyl-CoA carboxylase: structure-based engineering of substrate specificity. Biochemistry 43: 14027–14036

Ducluzeau AL, Chenu E, Capowiez L and Baymann F (2008) The Rieske/cytochrome *b* complex of Heliobacteria. Biochim Biophys Acta 1777: 1140–1146

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl Acids Res 32: 1792–1797

Eisen JA, Nelson KE, Paulsen IT, Heidelberg JF, Wu M, Dodson RJ, Deboy R, Gwinn ML, Nelson WC, Haft DH, Hickey EK, Peterson JD, Durkin AS, Kolonay JL, Yang F, Holt I, Umayam LA, Mason T, Brenner M, Shea TP, Parksey D, Nierman WC, Feldblyum TV, Hansen CL, Craven MB, Radune D, Vamathevan J, Khouri H, White O, Gruber TM, Ketchum KA, Venter JC, Tettelin H, Bryant DA and Fraser CM (2002) The complete genome sequence of *Chlorobium tepidum* TLS, a photosynthetic, anaerobic, green-sulfur bacterium. Proc Natl Acad Sci USA 99: 9509–9514

Feick RG and Fuller RC (1984) Topography of the photosynthetic apparatus of *Chloroflexus aurantiacus*. Biochemistry 23: 3693–3700

Feick RG, Fitzpatrick M and Fuller RC (1982) Isolation and characterization of cytoplasmic membranes and chlorosomes from the green bacterium *Chloroflexus aurantiacus*. J Bacteriol 150: 905–915

Feick R, Ertlmaier A and Ermler U (1996) Crystallization and X-ray analysis of the reaction center from the thermophilic green bacterium *Chloroflexus aurantiacus*. FEBS Lett 396: 161–164

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA Merrick JM, McKenney K, Sutton G, FitzHugh W, Fields C, Gocayne JD, Scott J, Shirley R, Liu L-I, Glodek A, Kelley JM, Weidman JF, Phillips CA, Spriggs T, Hedblom E, Cotton MD, Utterback TR, Hanna MC, Nguyen DT, Saudek DM, Brandon RC, Fine LD,

Fritchman JL, Fuhrmann JL, Geoghagen NSM, Gnehm, CL, McDonald LA, Small KV, Fraser CM, Smith HO and Venter JC (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269: 496–512

Foster JM, Redlinger TE, Blankenship RE and Fuller RC (1986) Oxygen regulation of development of the photosynthetic membrane system in *Chloroflexus aurantiacus*. J Bacteriol 167: 655–659

Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, Fritchman RD, Weidman JF, Small KV, Sandusky M, Fuhrmann J, Nguyen D, Utterback TR, Saudek DM, Phillips CA, Merrick JM, Tomb JF, Dougherty BA, Bott KF, Hu PC, Lucier TS, Peterson SN, Smith HO, Hutchison CA 3rd and Venter JC (1995) The minimal gene complement of *Mycoplasma genitalium*. Science 270: 397–403

Friedmann S, Alber BE and Fuchs G (2006a) Properties of succinyl-coenzyme A:L-malate coenzyme A transferase and its role in the autotrophic 3-hydroxypropionate cycle of *Chloroflexus aurantiacus*. J Bacteriol 188: 2646–2655

Friedmann S, Steindorf A, Alber BE and Fuchs G (2006b) Properties of succinyl-coenzyme A:D-citramalate coenzyme A transferase and its role in the autotrophic 3-hydroxypropionate cycle of *Chloroflexus aurantiacus*. J Bacteriol 188: 6460–6468

Friedmann S, Alber BE and Fuchs G (2007) Properties of *R*-citramalyl-coenzyme A lyase and its role in the autotrophic 3-hydroxypropionate cycle of *Chloroflexus aurantiacus*. J Bacteriol 189: 2906–2914

Frigaard N-U and Bryant DA (2004) Seeing green bacteria in a new light: genomics-enabled studies of the photosynthetic apparatus in green sulfur bacteria and filamentous anoxygenic phototrophic bacteria. Arch Microbiol 182: 265–276

Frigaard N-U and Bryant DA (2006) Chlorosomes: antenna organelles in green photosynthetic bacteria. In: Shively JM (ed) Complex Intracellular Structures in Prokaryotes, Microbiology Monographs, Vol 2, pp 79–114. Springer, Berlin

Frigaard N-U and Bryant DA (2008a) Genomic insights into the sulfur metabolism of phototrophic sulfur bacteria. In: Hell R, Dahl C, Knaff DB and Leustek T (eds) Sulfur Metabolism in Phototrophic Organisms, pp 343–361. Springer, Dordrecht

Frigaard N-U and Bryant DA (2008b) Genomic and evolutionary perspectives on sulfur metabolism in green sulfur bacteria. In: Friedrich CG and Dahl C (eds) Microbial Sulfur Metabolism (Proceedings of the International Symposium on Microbial Sulfur Metabolism), pp 60–76. Springer, Dordrecht

Frigaard N-U and Dahl C (2009) Sulfur metabolism in phototrophic sulfur bacteria. Adv Microb Physiol 54: 103–200

Frigaard N-U and Matsuura K (1999) Oxygen uncouples light absorption by the chlorosome antenna and photosynthetic electron transfer in the green sulfur bacterium *Chlorobium tepidum*. Biochim Biophys Acta 1412: 108–117

Frigaard N-U, Voigt GD and Bryant DA (2002a) *Chlorobium tepidum* mutant lacking bacteriochlorophyll *c* made by inactivation of the *bchK* gene, encoding bacteriochlorophyll *c* synthase. J Bacteriol 184: 4358–4370

Frigaard N-U, Vassilieva EV, Li H, Milks KJ, Zhao J and Bryant DA (2002b) The remarkable chlorosome. PS2001 Proceedings, Proceedings of the 12th International Congress on Photosynthesis, Brisbane, Australia. Article S1-003, CSIRO Publishing, Canberra, Australia

Frigaard N-U, Li H, Gomez Maqueo Chew A, Maresca JA and Bryant DA (2003) *Chlorobium tepidum*: insights into the physiology and biochemistry of green sulfur bacteria from the complete genome sequence. Photosynth Res 78: 93–117

Frigaard N-U, Li H, Milks KJ and Bryant DA (2004a) Nine mutants of *Chlorobium tepidum* each unable to synthesize a different chlorosome protein still assemble functional chlorosomes. J Bacteriol 186: 646–653

Frigaard N-U, Maresca JA, Yunker CE, Jones AD and Bryant DA (2004b) Genetic manipulation of carotenoid biosynthesis in the green sulfur bacterium *Chlorobium tepidum*. J Bacteriol 186: 5210–5220

Frigaard N-U, Gomez Maqueo Chew A, Maresca JA and Bryant DA (2006) Bacteriochlorophyll biosynthesis in green bacteria. In: Grimm B, Porra RJ, Rüdiger W and Scheer H (eds) Chlorophylls and Bacteriochlorophylls: Biochemistry, Biophysics, Functions and Applications, pp 201–221. Springer, Dordrecht

Fromme P, Jordan P and Krauß N (2001) Structure of photosystem I. Biochim Biophys Acta 1507: 5–31

Fröstl JM and Overmann J (1998a) Physiology and tactic response of the phototrophic consortium "*Chlorochromatium aggregatum*." Arch Microbiol 169: 129–135

Fröstl JM and Overmann J (1998b) Phylogenetic affiliation of the bacteria that constitute phototrophic consortia. Arch Microbiol 174: 50–58

Fujita Y and Bauer CE (2000) Reconstitution of light-independent protochlorophyllide reductase from purified BchL and BchN-BchB subunits. *In vitro* confirmation of nitrogenase-like features of a bacteriochlorophyll biosynthesis enzyme. J Biol Chem 275: 23583–23588

Ganapathy S, Oostergetel GT, Wawrzyniak PK, Reus M, Gomez Maqueo Chew A, Buda F, Boekema EJ, Bryant DA, Holzwarth AR and de Groot HM (2009) Alternating *syn-anti* bacteriochlorophylls form concentric helical nanotubes in chlorosomes. Proc Natl Acad Sci USA 106: 8525–8530

Gao X, Xin Y and Blankenship RE (2009) Enzymatic activity of the alternative complex III as a menaquinol:auracyanin oxidoreductase in the electron transfer chain of *Chloroflexus aurantiacus*. FEBS Lett 583: 3275–3279

Garcia Costas AM, Liu Z, Tomsho LP, Schuster SC, Ward DM and Bryant DA (2011) Complete genome of *Candidatus* Chloracidobacterium thermophilum, a

chlorophyll-based phototroph belonging to the phylum *Acidobacteria*. Environ Microbiol, accepted upon minor revision

Garrity GM and Holt JG (2001) The road map to the Manual. In: Boone DR and Castenholz RW (eds) Bergey's Manual of Systematic Bacteriology, 2nd Ed, Vol 1, pp 119–166. Springer, Berlin

van Gemerden H and Mas J (1995) Ecology of phototrophic sulfur bacteria. In: Blankenship RE, Madigan MT and Bauer CE (eds) Anoxygenic Photosynthetic Bacteria, pp 49–85. Kluwer Academic Publishers, Dordrecht

Gibson J, Pfennig N and Waterbury JB (1984) *Chloroherpton thalassium* gen. nov. et spec. nov., a non-filamentous, flexing and gliding green sulfur bacterium. Arch Microbiol 138: 96–101

Gich F, Airs RL, Danielsen M, Keely BJ, Abella CA, Garcia-Gil J, Miller M and Borrego CM (2003) Characterization of the chlorosome antenna of the filamentous anoxygenic phototrophic bacterium *Chloronema* sp. strain UdG9001. Arch Microbiol 180: 417–426

Giraud E and Verméglio A (2008) Bacteriophytochromes in anoxygenic photosynthetic bacteria. Photosynth Res 97: 141–153

Golbeck JH (1993) Shared thematic elements in photochemical reaction centers. Proc Natl Acad Sci USA 90: 1642–1646

Gomez Maqueo Chew A (2007) Elucidation of the bacteriochlorophyll *c* biosynthesis pathway in green sulfur bacterium *Chlorobium tepidum*. PhD thesis, The Pennsylvania State University, University Park, PA

Gomez Maqueo Chew A and Bryant DA (2007a) Chlorophyll biosynthesis in bacteria: the origins of structural and functional diversity. Annu Rev Microbiol 61: 113–129

Gomez Maqueo Chew A and Bryant DA (2007b) Characterization of a plant-like protochlorophyllide *a* divinyl reductase in green sulfur bacteria. J Biol Chem 282: 2967–2975

Gomez Maqueo Chew A, Frigaard N-U and Bryant DA (2007) Bacteriochlorophyllide c C-8^2 and C-12^1 methyltransferases are essential for adaptation to low light in *Chlorobaculum tepidum*. J Bacteriol 189: 6176–6184

Gomez Maqueo Chew A, Frigaard N-U and Bryant DA (2009) Mutational analysis of three *bchH* paralogs in (bacterio-)chlorophyll biosynthesis in *Chlorobaculum tepidum*. Photosynth Res 101: 21–34

Graham JE and Bryant DA (2008) The biosynthetic pathway for synechoxanthin, an aromatic carotenoid synthesized by the euryhaline, unicellular cyanobacterium *Synechococcus* sp. strain PCC 7002. J Bacteriol 190: 7966–7974

Graham JE and Bryant DA (2009) The biosynthetic pathway for myxol-2′-fucoside (myxoxanthophyll) in the cyanobacterium *Synechococcus* sp. strain PCC 7002. J Bacteriol 191: 3292–3300

Granick S (1957) Speculations on the origins and evolution of photosynthesis. Ann NY Acad Sci 69: 292–308

Granick S (1965) Evolution of heme and chlorophyll. In: Bryson V and Vogel HG (eds), Evolving Genes and Proteins, pp 67–68. Academic Press, New York

Gregersen LH, Bryant DA and Frigaard N-U (2011) Mechanisms and evolution of oxidative sulfur metabolism in green sulfur bacteria. Front Microbio 2: 116

Griesbeck C, Hauska G and Schütz M (2000) Biological sulfide oxidation: sulfide:quinone reductase (SQR), the primary reaction. In: Pandalai SG (ed) Recent Research Developments in Microbiology, Vol 4, pp 179–204. Research Signpost, Trivandrum

Guindon S and Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. System Biol 52: 696–704

Hager-Braun C, Xie DL, Jarosch U, Herold E, Büttner M, Zimmermann R, Deutzmann R, Hauska G and Nelson N (1995) Stable photobleaching of P840 in *Chlorobium* reaction center preparations: presence of the 42-kDa bacteriochlorophyll *a* protein and a 17-kDa polypeptide. Biochemistry 34: 9617–9624

Hanada S and Pierson BK (2006) The family *Chloroflexaceae*. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H and Stackebrandt E (eds) The Prokaryotes, Vol 7, pp 815–842. Springer, Berlin

Hanada S, Hiraishi A, Shimada K and Matsuura K (1995) *Chloroflexus aggregans* sp. nov., a filamentous phototrophic bacterium which forms dense cell aggregates by active gliding movement. Int J Syst Bacteriol 45: 676–681

Hanada S, Takaichi S, Matsuura K and Nakamura K (2002) *Roseiflexus castenholzii* gen. nov., sp. nov., a thermophilic, filamentous, photosynthetic bacterium that lacks chlorosomes. Int J Syst Evol Microbiol 52: 187–193

Harada J, Saga Y, Yaeda Y, Oh-oka H and Tamiaki H (2005) *In vitro* activity of C-20 methyltransferase, BchU, involved in bacteriochlorophyll *c* biosynthetic pathway in green sulfur bacteria. FEBS Lett 579: 1983–1987

Hauska G, Schoedle T, Remigy H and Tsiotis G (2001) The reaction center of green sulfur bacteria. Biochim Biophys Acta 1507: 260–277

Hegler F, Posth NR, Jiang J and Kappler A (2008) Physiology of phototrophic iron(II)-oxidizing bacteria: implications for modern and ancient environments. FEMS Microbiol Ecol 66: 250–260

Heinnickel M and Golbeck JH (2007) Heliobacterial photosynthesis. Photosynth Res 46: 2530–2536

Heising S, Richter L, Ludwig W and Schink B (1999) *Chlorobium ferrooxidans* sp. nov., a phototrophic green sulfur bacterium that oxidizes iron in coculture with a "*Geospirillum*" sp. strain. Arch Microbiol 172: 116–124

Herter S, Farfsing J, Gad'On N, Rieder C, Eisenreich W, Bacher A and Fuchs G (2001) Autotrophic CO_2 fixation by *Chloroflexus aurantiacus*: study of glyoxylate formation and assimilation via the 3-hydroxypropionate cycle. J Bacteriol 183: 4305–4316

Herter S, Busch A and Fuchs G (2002a) L-Malyl-coenzyme A lyase/beta-methylmalyl-coenzyme A lyase from *Chloroflexus aurantiacus*, a bifunctional enzyme involved in autotrophic CO_2 fixation. J Bacteriol 184: 5999–6006

Herter S, Fuchs G, Bacher A and Eisenreith W (2002b) A bicyclic autotrophic CO_2 fixation pathway in *Chloroflexus aurantiacus*. J Biol Chem 277: 1501–1509

Higuchi M, Hirano Y, Kimura Y, Oh-oka H, Miki K and Wang Z-Y (2009) Overexpression, characterization, and crystallization of the functional domain of cytochrome c_z from *Chlorobium tepidum*. Photosynth Res 102: 77–84

Hillmann F, Fischer R-J, Saint-Prix F, Girbal L and Bahl H (2008) PerR acts as a switch for oxygen tolerance in the strict anaerobe *Clostridium acetobutylicum*. Mol Microbiol 68: 848–860

Hohmann-Marriott MF and Blankenship RE (2007) Variable fluorescence in green sulfur bacteria. Biochim Biophys Acta 1767: 106–113

Holkenbrink C, Ocón Barbas S, Mellerup A, Otaki H and Frigaard N-U (2011) Sulfur globule oxidation in green sulfur bacteria is dependent on the dissimilatory sulfite reductase system. Microbiology 157: 1229–1239

Holt JG and Lewin RA (1968) *Herpetosiphon aurantiacus* gen. et sp. n., a new filamentous gliding organism. J Bacteriol 95: 2407–2408

Horimoto K, Fukuchi S and Mori K (2001) Comprehensive comparison between locations of orthologous genes on archaeal and bacterial genomes. Bioinformatics 17: 791–802

Hörtensteiner S (2006) Chlorophyll degradation during senescence. Annu Rev Plant Biol 57: 55–77

Hugenholtz P and Stackebrandt E (2004) Reclassification of *Sphaerobacter thermophilis* from the subclass *Sphaerobacteridae* in the phylum *Actinobacteria* to the class *Thermomicrobia* (emended description) in the phylum *Chloroflexi* (emended description). Int J Syst Evol Microbiol 54: 2049–2051

Hügler M, Menendez C, Schägger H and Fuchs G (2002) Malonyl-coenzyme A reductase from *Chloroflexus aurantiacus*, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO_2 fixation. J Bacteriol 184: 2404–2410

Hügler M, Krieger RS, Jahn M and Fuchs G (2003) Characterization of acetyl-CoA/propioinyl-CoA carboxylase in *Metallosphaera sedula*. Carboxylating enzyme in the 3-hydroxypropionate cycle for autotrophic carbon fixation. Eur J Biochem 270: 736–744

Hügler M, Wirsen, CO, Fuchs G and Taylor CD, Sievert SM (2005) Evidence for autotrophic CO_2 fixation via the reductive tricarboxylic acid cycle by members of the ε subdivision of Proteobacteria. J Bacteriol 187: 3020–3027

Huster MS and Smith KM (1990) Biosynthetic studies of substituent homologation in bacteriochlorophylls *c* and *d*. Biochemistry 29: 4348–4355

Iino T, Mori K, Uchino Y, Nakagawa T, Harayama S and Suzuki KI (2010) *Ignavibacterium album* gen. nov., sp. nov., a moderately thermophilic anaerobic bacterium isolated from microbial mats at a terrestial hot spring, and proposal of *Ignavibacteria* classis nov. for a novel lineage at the periphery of the green sulfur bacteria. Int J Syst Evol Microbiol 60: 1376–1382

Ikonen TP, Li H, Pšenčík J, Laurinmaki PA, Butcher SJ Frigaard N-U, Serimaa RE, Bryant DA and Tuma R (2007) X-ray scattering and electron cryomicroscopy study on the effect of carotenoid biosynthesis to the structure of *Chlorobium tepidum* chlorosomes. Biophys J 93: 620–628

Imhoff JF (1995) Taxonomy and physiology of phototrophic purple bacteria and green sulfur bacteria. In: Blankenship RE, Madigan MT and Bauer CE (eds) Anoxygenic Photosynthetic Bacteria, pp 1–15. Kluwer Academic Publishers, Dordrecht

Imhoff JF (2003) Phylogenetic taxonomy of the family *Chlorobiaceae* on the basis of 16S rRNA and FMO (Fenna-Matthews-Olson protein) gene sequences. Int J Syst Evol Microbiol 53: 941–951

Islam MR, Aikawa S, Midorikawa T, Kashino Y, Satoh K and Koike H (2008) Slr1923 of *Synechocystis* sp. PCC6803 is essential for conversion of 3,8-divinyl(protochlorophyll (ide) to 3-monovinyl(proto)chlorophyll(ide). Plant Physiol 148: 1068–1081

Ito H, Yokono M, Tanaka R and Tanaka A (2008) Identification of a novel vinyl reductase gene essential for the biosynthesis of monovinyl chlorophyll in *Synechocystis* sp. PCC 6803. J Biol Chem 283: 9002–9011

Ivanovsky RN, Fal YI, Berg IA, Ugolkova NV, Karsilnikova EN, Keppen OI, Zakharchuc LM and Zyakun AM (1999) Evidence for the presence of the reductive pentose phosphate cycle in a filamentous anoxygenic photosynthetic bacterium *Oscillochloris triochoides* strain DG-6. Microbiology 145: 1743–1748

Iwai M, Takizawa K, Tokutsu R, Okamuro A, Takahashi Y and Minagawa J (2010) Isolation of the elusive supercomplex that drives cyclic electron flow in photosynthesis. Nature 464: 1210–1213

Jackson TJ, Ramaley RF and Meinschein WG (1973) *Thermomicrobium*, a new genus of extremely thermophilic bacteria. Int J Syst Bacteriol 23: 28–36

Jagannathan B and Golbeck JH (2008) Unifying principles in homodimeric type I photosynthetic reaction centers: properties of PscB and the F_A, F_B and F_X iron clusters in green sulfur bacteria. Biochim Biophys Acta 1777: 1371–1380

Jagannathan B and Golbeck JH (2009) Understanding of the binding interface between PsaC and the PsaA/PsaB heterodimer in Photosystem I. Biochemistry 48: 5405–5416

Johnson ET and Schmidt-Dannert C (2008) Characterization of three homologs of the large subunit of the magnesium chelatase from *Chlorobaculum tepidum* and interaction with the magnesium protoporphyrin IX methyltransferase. J Biol Chem 283: 27776–27784

Jurkevitch E (2007) Predatory behaviors in bacteria: diversity and transitions. Microbe 2: 67–73

Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hisosawa M, Sugiura M, Sasamoto S, Kimura T, Hosouchi T, Matsuno A, Muraki A, Nakazaki N Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M and Tabata S (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. DNA Res 3: 109–136 and 185–209

Kaneko T, Nakamura Y, Wolk CP, Kuritz T, Sasamoto S, Watanabe A, Iriguchi M, Ishikawa A, Kawashima K, Kimura T, Kishida Y, Kohara M, Matsumoto M, Matsuno A, Muraki A, Nakazaki N, Shimpo S, Sugimoto M, Takazawa M, Yamada M, Yasuda M and Tabata S (2001) Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. DNA Res 8: 205–213, 227–253

Keppen OI, Tourova TP, Kuznetsov BB, Ivanovsky RN and Gorlenko VM (2000) Proposal of *Oscillochloridaceae* fam. Nov. on the basis of a phylogenetic analysis of the filamentous anoxygenic phototrophic bacteria, and emended description of *Oscillochloris* and *Oscillochloris trichoides* in comparison with further new isolates. Int J Syst Evol Microbiol 50: 1529–1537

Kiley PJ, Donohue TJ, Havelka WA and Kaplan S (1987) DNA sequence and in vitro expression of the B875 light-harvesting polypeptides of Rhodobacter sphaeroides. J Bacteriol 169: 742–750

Klappenbach JA and Pierson BK (2004) Phylogentic and physiological characterization of a filamentous anoxygenic photoautotrophic bacterium '*Candidatus* Chlorothrix halophila' gen. nov., sp. nov., recovered from hypersaline microbial mats. Arch Microbiol 181: 17–25

Klatt CG, Bryant DA and Ward DM (2007) Comparative genomics provides evidence for the 3-hydroxypropionate autotrophic pathway in filamentous anoxygenic phototrophic bacteria and in hot spring microbial mats. Environ Microbiol 9: 2067–2078

Klatt CG, Wood JM, Rusch DB, Bateson MM, Hamamura N, Heidelberg JF, Grossman AR, Bhaya D, Cohan FM, Kühl M, Bryant DA and Ward DM (2011) Community ecology of hot spring cyanobacterial mats: predominant populations and their functional potential. ISME J, in press

Kondrat'eva EN and Krasil'nikova EN (1988) Utilization of thiosulfate by *Chloroflexus aurantiacus*. Mikrobiologiya (Eng transl) 57: 291–294

Krasil'nikova EN (1987) ATP sulfurylase activity in *Chloroflexus aurantiacus* and other photosynthesizing bacteria as a function of temperature. Mikrobiologiya (Eng transl) 55: 418–421

Krasil'nikova EN and Kondrat'eva EN (1987) Growth of *Chloroflexus aurantiacus* under anaerobic conditions in the dark and the metabolism of organic substrates. Mikrobiologiya (Eng transl) 56: 281–285

Krügel H, Krubasik P, Weber K, Saluz HP and Sandmann G (1999) Functional analysis of genes from *Streptomyces griseus* involved in the synthesis of isorenieratene, a carotenoid with aromatic end groups, revealed a novel type of carotenoid desaturase. Biochim Biophys Acta 1439: 57–64

Kurtz DM Jr (2006) Avoiding high valent iron intermediates: superoxide reductase and rubrerythrin. J Inorg Biochem 100: 679–693

Kuznetsov BB, Ivanovsky RN, Keppen OI, Sukhacheva MV, Bumazhkin BK, Patutina EO, Beletsky AV, Mardanov AV, Baselerov RV, Panteleeva AN, Kolganova TV, Ravin NV and Skryabin KG (2011) Draft genome sequence of the anoxygenic filamentous phototrophic bacterium *Oscillochloris trichoides* DG-6. J Bacteriol 193: 321–322

Layer G, Kervio E, Morlock G, Heinz DW, Jahn D, Retey J and Schubert WD (2005) Structural and functional comparison of HemN to other radical SAM enzymes. Biol Chem 386: 971–980

Lebrun E, Brugna M, Baymann F, Muller D, Lièvremont D, Lett M-C and Nitschke W (2003) Arsenite oxidase, an ancient bioenergetic enzyme. Mol Biol Evol 20: 686–693

Lee M, del Rosario MC, Harris HH, Blankenship RE, Guss JM and Freeman HC (2009a) The crystal structure of auracyanin A at 1.85 Å resolution: the structures and functions of auracyanins A and B, two almost identical "blue" copper proteins, in the photosynthetic bacterium *Chloroflexus aurantiacus*. J Biol Inorg Chem 14: 329–345

Lee ZM, Bussema C 3rd and Schmidt TM (2009b) rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. Nucl Acids Res 37: D489–493

Lehmann RP, Brunisholz RA and Zuber H (1994) Structural differences in chlorosomes of *Chloroflexus aurantiacus* grown under different conditions support the BChl *c*-binding function of the 5.7 kDa protein. FEBS Lett 342: 319–324

Li H (2006) Organization and function of chlorosome proteins in the green sulfur bacterium *Chlorobium tepidum*. PhD thesis, The Pennsylvania State University, University Park, PA

Li H and Bryant DA (2009) Envelope proteins of the CsmB/CsmF and CsmC/CsmD motif families influence the size, shape, and composition of chlorosomes in *Chlorobaculum tepidum*. J Bacteriol 191: 7109–7120

Li H, Frigaard N-U and Bryant DA (2006) Molecular contacts for chlorosome envelope proteins revealed by cross-linking studies with chlorosomes from *Chlorobium tepidum*. Biochemistry 45: 9095–9103

Li H, Jubelirer S, Garcia Costas AM, Frigaard N-U and Bryant DA (2009) Multiple antioxidant proteins protect *Chlorobaculum tepidum* against oxygen and reactive oxygen species. Arch Microbiol 191: 853–867

Li L, Stoeckert CJ Jr and Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13: 2178–2189

Lippert K-D and Pfennig N (1969) Die Verwertung von molecularem Wasserstoff durch *Chlorobium thiosulfatophilum*. Arch Microbiol 65: 29–47

Liu Z and Bryant DA (2011) Identification of a gene essential for the first committed step in the biosynthesis of bacteriochlorophyll *c*. J Biol Chem 286: 22393–22402

Liu Z, Klatt CG, Wood JM, Rusch DB, Ludwig M, Wittekindt N, Tomsho LP, Schuster SC, Ward DM and Bryant DA (2011) Metatranscriptomic analyses of chlorophototrophs of a hot-spring microbial mat. ISME J, in press

Ludwig W and Klenk H-P (2001) Overview: a phylogenetic backbone and taxonomic framework for procaryotic systematics. In: Boone DR and Castenholz RW (eds) Bergey's Manual of Systematic Bacteriology, 2nd Ed, Vol 1, pp 49–65. Springer, Berlin

Madigan MT, Petersen SR and Brock TD (1974) Nutritional studies on *Chloroflexus*, a filamentous photosynthetic, gliding bacterium. Arch Microbiol 100: 97–103

Madigan MT and Brock T (1977a) Photosynthetic sulfide oxidation by *Chloroflexus aurantiacus*, a filamentous, photosynthetic, gliding bacterium. J Bacteriol 122: 782–784

Madigan MT and Brock TD (1977b) 'Chlorobium-type' vesicles of photosynthetically-grown *Chloroflexus aurantiacus* observed using negative staining techniques. J Gen Microbiol 102: 279–285

Manske AK, Glaeser J, Kuypers MM and Overmann J (2006) Physiology and phylogeny of green sulfur bacteria forming a monospecific phototrophic assemblage at a depth of 100 meters in the Black Sea. Appl Environ Microbiol 71: 8049–8060

Marcia M, Ermler U, Peng G and Michel H (2009) The structure of *Aquifex aeolicus* sulfide:quinone oxidoreductase, a basis to understand sulfide detoxification and respiration. Proc Natl Acad Sci USA 106: 9625–9630

Maresca JA and Bryant DA (2006) Two genes encoding new carotenoid-modifying enzymes in the green sulfur bacterium *Chlorobium tepidum*. J Bacteriol 188: 6217–6223

Maresca JA, Gomez Maqueo Chew A, Ponsatí MR, Frigaard N-U, Ormerod JG and Bryant DA (2004) The *bchU* gene of *Chlorobium tepidum* encodes the C-20 methyltransferase in bacteriochlorophyll *c*-biosynthesis. J Bacteriol 186: 2558–2566

Maresca JA, Graham JE, Wu M, Eisen JA and Bryant DA (2007) Identification of a fourth family of lycopene cyclases in photosynthetic bacteria. Proc Natl Acad Sci USA 104: 11784–11789

Maresca JA, Romberger SP and Bryant DA (2008a) Isorenieratene biosynthesis in green sulfur bacteria requires the cooperative actions of two carotenoid cyclases. J Bacteriol 190: 6384–6391

Maresca JA, Graham JE and Bryant DA (2008b) The biochemical basis for structural diversity in the carotenoids of chlorophototrophic bacteria. Photosynth Res 97: 121–140

van der Meer MTJ, Schouten S, de Leeuw JW and Ward DM (2000) Autotrophy of green non-sulphur bacteria in hot spring microbial mats: biological explanations for isotopically heavy organic carbon in the geological record. Environ Microbiol 2: 428–435

van der Meer MTJ, Schouten S, Bateson MM, Nübel U, Wieland A, Kühl M, de Leeuw JW, Sinninghe Damsté JS and Ward DM (2005) Diel variations in carbon metabolism by green nonsulfur-like bacteria in alkaline siliceous hot spring microbial mats from Yellowstone National Park. Appl Environ Microbiol 71: 3978–3986

van der Meer MTJ, Klatt CG, Wood J, Bryant DA, Bateson MM, Lammerts L, Schouten S, Sinninghe Damsté JS, Madigan MT and Ward DM (2010) Cultivation and genomic, nutritional and lipid biomarker characterization of *Roseiflexus* sp. strains closely related to predominant *in situ* populations inhabiting Yellowstone hot spring microbial mats. J Bacteriol 192: 3033–3042

Méndez-Alvarez S, Pavón V, Esteve I, Guerrero R and Gaju N (1994) Transformation of *Chlorobium limicola* by a plasmid that confers the ability to utilize thiosulfate. J Bacteriol 176: 7395–7397

Méndez-Alvarez S, Pavón V, Esteve I, Guerrero R and Gaju N (1995) Genomic heterogeneity in *Chlorobium limicola*: chromosomic and plasmidic differences among strains. FEMS Microbiol Lett 134: 279–285

Meyer TE (1994) Evolution of photosynthetic reaction centers and light harvesting chlorophyll proteins. BioSystems 33: 165–175

Mi H, Endo T, Ogawa T and Asada K (1995) Thylakoid membrane-bound NADPH-specific pyridine nucleotide dehydrogenase complex mediates cyclic electron transport in the cyanobacterium *Synechocystis* sp. PCC 6803. Plant Cell Physiol 36: 661–668

Milks KJ, Danielsen M, Persson S, Jensen ON, Cox RP and Miller M (2005) Chlorosome proteins studied by MALDI-TOF-MS: topology of CsmA in *Chlorobium tepidum*. Photosynth Res 86: 113–121

Mix LJ, Haig D and Cavanaugh CM (2005) Phylogenetic analyses of the core antenna domain: investigating the origin of photosystem I. J Mol Evol 60: 153–163

Montaño GA, Wu HM, Lin S, Brune DC and Blankenship RE (2003) Isolation and characterization of the B798 light-harvesting baseplate from the chlorosomes of *Chloroflexus aurantiacus*. Biochemistry 42: 10246–10251

Mulkidjanian AY (2007) Proton translocation by the cytochrome bc_1 complexes of phototrophic bacteria: introducing the activated Q-cycle. Photochem Photobiol Sci 6: 19–34

Mulkidjanian AY and Junge W (1997) On the origin of photosynthesis as inferred from sequence analysis. A primordial UV-protector as common ancestor of reaction centers and antenna proteins. Photosynth Res 51: 27–42

Mulkidjanian AY, Koonin EV, Makarova KS, Mekhedov SL, Sorokin A, Wolf YI, Dufresne A, Partensky F, Burd H, Kaznadzey D, Haselkorn R and Galperin MY (2006) The cyanobacterial genome core and the origin of photosynthesis. Proc Natl Acad Sci USA 103: 13126–13131

Muraki N, Seo D, Shiba T, Sakurai T and Kurisu G (2008) Crystallization and preliminary X-ray studies of

ferredoxin-NAD(P)$^+$ reductase from *Chlorobium tepidum*. Acta Crystallogr Sect F Struct Biol Cryst Commun 64: 186–189

Muraki N, Nomata J, Ebata K, Mizoguchi T, Shiba T, Tamiaki H, Kurisu G and Fujita Y (2010) X-ray crystal structure of the light-independent protochlorophyllide reductase. Nature 465: 110–114

Nadson GA (1906) The morphology of inferior algae. III. *Chlorobium limicola* Nads., the green chlorophyll bearing microbe. Bull Jard Bot St Petersb 6: 190

Nagata N, Tanaka R, Satoh S and Tanaka A (2005) Identification of a vinyl reductase gene for chlorophyll biosynthesis in *Arabidopsis thaliana* and implications for the evolution of *Prochlorococcus* species. Plant Cell 17: 233–240

Nagata N, Tanaka R and Tanaka A (2007) The major route for chlorophyll synthesis include [3,8-divinyl]-chlorophyllide *a* reduction in *Arabidopsis thaliana*. Plant Cell Physiol 48: 1803–1808

Nakamura Y, Kaneko T, Sato S, Ikeuchi M, Katoh H, Sasamoto S, Watanabe A, Iriguchi M, Kawashima K, Kimura T, Kishida Y, Kiyokawa C, Kohara M, Matsumoto M, Matsuno A, Nakazaki N, Shimpo S, Sugimoto M, Takeuchi C, Yamada M and Tabata S (2002) Complete genome structure of the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1. DNA Res 9: 123–130

Nakamura Y, Kaneko T, Sato S, Mimuro M, Miyashita H, Tsuchiya T, Sasamoto S, Watanabe A, Kawashima K, Kishida Y, Kiyokawa C, Kohara M, Matsumoto M, Matsuno A, Nakazaki N, Shimpo S, Takeuchi C, Yamada M and Tabata S (2003) Complete genome structure of *Gloeobacter violaceus* PCC 7421, a cyanobacterium that lacks thylakoids. DNA Res 10: 137–145

Niedermeier G, Shiozawa JA, Lottspeich F and Feick RG (1994) The primary structure of two chlorosome proteins from *Chloroflexus aurantiacus*. FEBS Lett 342: 61–65

Nisbet EG, Cann JR and Van Dover CL (1995) Did photosynthesis begin from thermotaxis? Nature 373: 479–480

Nomata J, Swem LR, Bauer CE and Fujita Y (2005) Overexpression and characterization of dark-operative protochlorophyllide reductase from *Rhodobacter capsulatus*. Biochim Biophys Acta 1708: 229–237

Nomata J, Mizoguchi T, Tamiaki H and Fujita Y (2006) A second nitrogenase-like enzyme for bacteriochlorophyll biosynthesis: reconstitution of chlorophyllide *a* reductase with purified X-protein (BchX) and YZ-protein (BchY-BchZ) from *Rhodobacter capsulatus*. J Biol Chem 281: 15021–15028

Nomata J, Ogawa T, Kitashima M, Inoue K and Fujita Y (2008) NB-protein (BchN-BchB) of dark-operative protochlorophyllide reductase is the catalytic component containing oxygen-tolerant Fe-S clusters. FEBS Lett 582: 1346–1350

Oelze J (1992) Light and oxygen regulation of the synthesis of bacteriochlorophylls *a* and *c* in *Chloroflexus aurantiacus*. J Bacteriol 172: 5021–5026

Ogawa T, Furusawa T, Nomura R, Seo D, Hosoya-Matsuda N, Sakurai H and Inoue K (2008) SoxAX binding protein, a novel component of the thiosulfate oxidizing multienzyme system in the green sulfur bacterium *Chlorobium tepidum*. J Bacteriol 190: 6097–6110

Ohashi S, Miyashita H, Okada N, Iemura T, Watanabe T and Kobayashi M (2008) Unique photosystems in *Acaryochloris marina*. Photosynth Res 98: 141–149

Oh-oka H, Kamei S, Matsubara H, Iwaki M and Itoh S (1995) Two molecules of cytochrome *c* function as the electron donors to P840 in the reaction center complex isolated from a green sulfur bacterium *Chlorobium tepidum*. FEBS Lett 365: 30–24

Oh-oka H, Iwaki M and Itoh S (1997) Vicosity dependence of the electron transfer rate from bound cytochrome *c* to P840 in the photosynthetic reaction center of the green sulfur bacterium *Chlorobium tepidum*. Biochemistry 36: 9267–9272

Oh-oka H, Iwaki M and Itoh S (1998) Membrane-bound cytochrome c_z couples quinol oxidoreductase to the P840 reaction center complex in isolated membranes of the green sulfur bacterium *Chlorobium tepidum*. Biochemistry 37: 12293–12300

Okkels JS, Kjær B, Hansson O, Svendsen I, Møller BL and Scheller HV (1992) A membrane-bound monoheme cytochrome c_{551} of a novel type is the immediate electron donor to P840 of the *Chlorobium vibrioforme* photosynthetic reaction center complex. J Biol Chem 267: 21139–21145

Olson JM (1999) Early evolution of chlorophyll-based photosystems. Biochem Molec Biol 12: 468–482

Olson JM and Blankenship RE (2004) Thinking about the evolution of photosynthesis. Photosynth Res 80: 373–386

Olson JM and Pierson BK (1987) Evolution of reaction centers in photosynthetic prokaryotes. Int Rev Cytol 108: 209–248

Oostergetel GT, Reus M, Gomez Maqueo Chew A, Bryant DA, Boekema EJ and Holzwarth AR (2007) Long-range organization of bacteriochlorophyll in chlorosomes of *Chlorobium tepidum* investigated by cryo-electron microscopy. FEBS Lett 581: 5435–5439

Oostergetel GT, van Amerongen H and Boekema EJ (2010) The chlorosome: a prototype for efficient light harvesting in photosynthesis. Photosynth Res 104: 245–255

Ouchane S, Stenou AS, Picaud M and Astier C (2004) Aerobic and anaerobic Mg-protoporphyrin monomethyl ester cyclases in purple bacteria: a strategy adopted to bypass the repressive oxygen control system. J Biol Chem 279: 6385–6394

Overmann J (1997) Mahoney Lake: a case study of the ecological significance of phototrophic sulfur bacteria. Adv Microbial Ecol 15: 251–288

Overmann J (2001) Green sulfur bacteria. In: Boone DR and Castenholz RW (eds) Bergey's Manual of Systematic Bacteriology, 2nd Ed, Vol 1, pp 601–605. Springer, Berlin

Overmann J (2006) Green sulfur bacteria. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H and Stackebrandt E (eds) The Prokaryotes, Vol 7, pp 359–378. Springer, Berlin

Overmann J and Tuschak C (1997) Phylogeny and molecular fingerprinting of green sulfur bacteria. Arch Microbiol 167: 302–309

Pedersen MØ, Borch J, Højrup P and Cox RP, Miller M (2007) The light-harvesting antenna of *Chlorobium tepidum*: interactions between the FMO protein and the major chlorosome protein CsmA studied by surface plasmon resonance. Photosynth Res 89: 63–69

Pedersen MØ, Pham, L, Steensgaard DB and Miller M (2008a) A reconstituted light-harvesting complex from the green sulfur bacterium *Chlorobium tepidum* containing CsmA and bacteriochlorophyll *a*. Biochemistry 47: 1435–1441

Pedersen MØ, Underhaug J, Dittmer J, Miller M and Nielsen NC (2008b) The three dimensional structure of CsmA: a small antenna protein from the green sulfur bacterium *Chlorobium tepidum*. FEBS Lett 582: 2869–2874

Pedersen MØ, Linnanto J, Frigaard N-U, Nielsen NC and Miller M (2010) A model of the protein-pigment baseplate complex in chlorosomes. Photosynth Res 104: 233–243

Pereira MM, Refojo PN, Hreggvidsson GO, Hjorleifsdottir S and Teixeira M (2007) The alternative complex III from *Rhodothermus marinus* – a prototype of a new family of quinol:electron acceptor oxidoreductases. FEBS Lett 581: 4831–4835

Phadwal K (2005) Carotenoid biosynthetic pathway: molecular phylogenies and evolutionary behavior of *crt* genes in eubacteria. Gene 345: 35–43

Pierson BK and Castenholz RW (1974) A phototrophic gliding filamentous bacterium of hot springs, *Chloroflexus aurantiacus*, gen. and sp. nov. Arch Microbiol 100: 5–24

Pierson BK and Thornber JP (1983) Isolation and spectral characterization of photochemical reaction centers from the thermophilic green bacterium *Chloroflexus aurantiacus* strain J-10-fl. Proc Natl Acad Sci USA 80: 80–84

Pierson BK, Giovannoni SJ, Stahl DA and Castenholz RW (1985) *Heliothrix oregonensis*, gen. nov., sp. nov., a phototrophic filamentous bacterium containing bacteriochlorophyll *a*. Arch Microbiol 142: 164–167

Pinta V, Picaud M, Reiss-Husson F and Astier C (2002) *Rubrivivax gelatinosus acsF* (previously orf358) codes for a conserved, putative binuclear-iron-cluster-containing protein involved in aerobic oxidative cyclization of Mg-protoporphyrin IX monomethylester. J Bacteriol 184: 746–753

Pšenčík J, Ikonen TP, Laurinmäki P, Merckel MC, Butcher SJ, Serimaa RE and Tuma R (2004) Lamellar organization of pigments in chlorosomes, the light harvesting complexes of green photosynthetic bacteria. Biophys J 87: 1165–1172

Pšenčík J, Arellano JB, Ikonen TP, Borrego CM, Laurinmäki PA, Butcher SJ, Serimaa RE and Tuma R (2006) Internal structure of chlorosomes from brown-colored *Chlorobium* species and the role of carotenoids in their assembly. Biophys J 91: 1433–1440

Pšenčík J, Collins AM, Liljeroos L, Torkkeli M, Laurinmäki P, Ansink HM, Ikonen TP, Serimaa RE, Blankenship RE, Tuma R and Butcher SJ (2009) Structure of chlorosomes from the green filamentous bacterium *Chloroflexus aurantiacus*. J Bacteriol 191: 6701–6708

Raymond J, Zhaxybayeva O, Gogarten JP, Gerdes SY and Blankenship RE (2002) Whole-genome analysis of photosynthetic prokaryotes. Science 298: 1616–1620

Reinartz M, Tschäpe J, Brüser T, Trüper HG and Dahl C (1998) Sulfide oxidation in the phototrophic sulfur bacterium *Chromatium vinosum*. Arch Microbiol 170: 59–68

Sadekar S, Raymond J and Blankenship RE (2006) Conservation of distantly related membrane proteins: photosynthetic reaction centers share a common structural core. Mol Biol Evol 23: 2001–2007

Saga Y and Tamiaki H (2006) Transmission electron microscopic study on supramolecular nanostructures of bacteriochlorophyll self-aggregates in chlorosomes of green photosynthetic bacteria. J Biosci Bioeng 102: 118–123

Sakuragi Y, Frigaard N-U, Shimada K and Matsuura K (1999) Association of bacteriochlorophyll *a* with the CsmA protein in chlorosomes of the photosynthetic green filamentous bacterium *Chloroflexus aurantiacus*. Biochim Biophys Acta 1413: 172–180

Sandmann G (2009) Evolution of carotene desaturation: the complication of a simple pathway. Arch Biochem Biophys 483: 169–174

Sarma R, Barney BM, Hamilton TL, Jones A, Seefeldt LC and Peters JW (2008) Crystal structure of the L protein of *Rhodobacter sphaeroides* light-independent protochlorophyllide reductase with MgADP bound: a homologue of the nitrogenase Fe protein. Biochemistry 47: 13004–13015

Schmidt S, Biegel E and Müller V (2009) The ins and outs of Na^+ bioenergetics in *Acetobacterium woodii*. Biochim Biophys Acta 1787: 691–696

Schubert W-D, Klukas O, Saenger W, Witt HT, Fromme P and Krauß N (1998) A common ancestor for oxygenic and anoxygenic photosynthetic systems: a comparison based on the structural model of photosystem I. J Mol Biol 280: 297–314

Schütz M, Zirnbigl S, le Coutre J, Büttner M, Xie D-L, Nelson N, Deutzmann R and Hauska G (1994) A transcription unit for the Rieske FeS-protein and cytochrome *b* in *Chlorobium limicola*. Photosynth Res 39: 163–174

Schütz M, Brugna M, Lebrun E, Baymann F, Huber R, Stetter KO, Hauska G, Toci R, Lemesle-Meunier D, Tron P, Schmidt C and Nitschke W (2000) Early evolution of cytochrome *bc* complexes. J Mol Biol 300: 663–675

Sekiguchi Y, Yamada T, Hanada S, Ohashi A, Harada H and Kamagata Y (2003) *Anaerolinea thermophila* gen. nov., sp. nov. and *Caldilinea aerophila* gen. nov., sp. nov., novel filamentous thermophiles that represent a previously uncultured lineage of the domain Bacteria at the subphylum level. Int J Syst Evol Microbiol 53: 1843–1851

Seo D and Sakurai H (2002) Purification and characterization of ferredoxin-$NAD(P)^+$ reductase from the green sulfur bacterium *Chlorobium tepidum*. Biochim Biophys Acta 1597: 123–132

Seo D, Tomioka A, Kusumoto N, Kamo M, Enami I and Sakurai H (2001) Purification of ferredoxins and their reaction with purified reaction center complex from the green sulfur bacterium *Chlorobium tepidum*. Biochim Biophys Acta 1503: 377–384

Seshadri R, Adrian L, Fouts DE, Eisen JA, Phillippy AM, Methe BA, Ward NL, Nelson WC, Deboy RT, Khouri HM, Kolonay JF, Dodson RJ, Daugherty SC, Brinkac LM, Sullivan SA, Madupu R, Nelson KE, Kang KH, Impraim M, Tran K, Robinson JM, Forberger HA, Fraser CM, Zinder SH and Heidelberg JF (2005) Genome sequence of the PCE-dechlorinating bacterium *Dehalococcoides ethenogenes*. Science 307: 105–108

Shioi Y, Watanabe K and Takamiya K (1996) Enzymatic conversion of pheophorbide *a* to a precursor of pyropheophorbide *a* in leaves of *Chenopodium album*. Plant Cell Physiol 37: 1143–1149

Shiozawa JA, Lottspeich F and Feick R (1987) The photochemical reaction center of *Chloroflexus aurantiacus* is composed of two structurally similar polypeptides. Eur J Biochem 167: 595–600

Shiozawa JA, Lottspeich F and Feick R (1989) The primary structure of the *Chloroflexus aurantiacus* reaction-center polypeptides. Eur J Biochem 180: 75–84

Sirevåg R (1995) Carbon metabolism in green bacteria. In: Blankenship RE, Madigan MT and Bauer CE (eds) Anoxygenic Photosynthetic Bacteria, pp 871–883. Kluwer Academic Publishers, Dordrecht

Sirevåg R and Castenholz RW (1979) Aspects of carbon metabolism in *Chloroflexus*. Arch Microbiol 120: 151–153

Sobotka R, Dühring U, Komenda J, Peter E, Gardian Z, Tichy M, Grimm B and Wilde A (2008) Importance of the cyanobacterial Gun4 protein for chlorophyll metabolism and assembly of photosynthetic complexes. J Biol Chem 283: 25794–25802

Sørensen PG, Cox RP and Miller M (2008) Chlorosome lipids from *Chlorobium tepidum*: characterization and quantification of polar lipids and wax esters. Photosynth Res 95: 191–196

Staehelin LA, Golecki JR and Drews G (1980) Supermolecular Organization of chlorosomes (Chlorobium vesicles) and of their membrane attachment sites. Biochim Biophys Acta 589: 30–45

Sprague SG, Staehelin LA, DiBartolomeis MJ and Fuller RC (1981) Isolation and development of chlorosomes in the green bacterium *Chloroflexus aurantiacus*. J Bacteriol 147: 1021–1031

Steuber J (2001) Na^+ translocation by bacterial NADH:quinone oxidoreductases: an extension to the complex-I family of primary redox pumps. Biochim Biophys Acta 1505: 45–56

Strauss G and Fuchs G (1993) Enzymes of a novel autotrophic CO_2 fixation pathway in the phototrophic bacterium *Chloroflexus aurantiacus*, the 3-hydroxypropionate cycle. Eur J Biochem 215: 633–643

Suzuki JY and Bauer CE (1995) Altered monovinyl and divinyl protochlorophyllide pools in *bchJ* mutants of *Rhodobacter capsulatus*. Possible monovinyl substrate discrimination of light-independent protochlorophyllide reductase. J Biol Chem 270: 3732–3740

Suzuki JY, Bollivar DW and Bauer CE (1997) Genetic analysis of chlorophyll biosynthesis. Annu Rev Genet 31: 61–89

Swartz TH, Ikewada S, Ishikawa O, Ito M and Krulwich TA (2005) The Mrp system: a giant among monovalent cation/proton antiporters? Extremophiles 9: 345–354

Swingley WD, Sadekar S, Mastrian SD, Matthies JH, Hao J, Ramos H, Acharya CR, Conrad AL, Taylor HL, Dejesa LC, Shah MK, O'huallachain ME, Lince MT, Blankenship RE, Beatty JT and Touchman JW (2007) The complete genome sequence of *Roseobacter denitrificans* reveals a mixotrophic rather than photosynthetic metabolism. J Bacteriol 189: 683–690

Tabita FR (2009) The hydroxypropionate pathway of CO_2 fixation: *fait accompli*. Proc Natl Acad Sci USA 106: 21015–21016

Taisova AS, Keppen OI, Lukashev EP, Arutyunyan AM and Fetisova Z (2002) Study of the chlorosomal antenna of the green mesophilic filamentous bacterium *Oscillochloris trichoides*. Photosynth Res 74: 73–85

Takaichi S (2001) Carotenoids and carotenogenesis in anoxygenic photosynthetic bacteria. In: Frank HA, Young AJ, Britton G and Cogdell RJ (eds) The Photochemistry of Carotenoids, pp 39–69. Kluwer Academic Publishers, Dordrecht

Tamura K, Dudley J, Nei M and Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596–1599

Tanaka R and Tanaka A (2006) Chlorophyll metabolism. Curr Opin Plant Biol 9: 248–255

Tanaka R and Tanaka A (2007) Tetrapyrrole biosynthesis in higher plants. Annu Rev Plant Biol 58: 321–346

Tang KH, Wen J, Li X and Blankenship RE (2009) Role of the AcsF protein in *Chloroflexus aurantiacus*. J Bacteriol 191: 3580–3587

Theissen U, Hoffmeister M, Grieshaber M and Martin W (2003) Single eubacterial origin of eukaryotic sulfide:quinone oxidoreductase, a mitochondrial enzyme conserved from the early evolution of eukaryotes during anoxic and sulfidic times. Mol Biol Evol 20: 1564–1574

Theroux SJ, Redlinger TE, Fuller RC and Robinson SJ (1990) Gene encoding the 5.7 kilodalton chlorosome protein of *Chloroflexus aurantiacus*: Regulated message levels and a predicted carboxy-terminal protein extension. J Bacteriol 172: 4497–4504

Tiller ERM and Collins RA (2000) Genome rearrangement by replication-directed translocation. Nat Genet 26: 195–197

Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mather EJ, Detter JC, Bork P, Hugenholtz P and Rubin EM (2005) Comparative metagenomics of microbial communities. Science 308: 554–557

Tronrud DE, Wen J, Gay L and Blankenship RE (2009) The structural difference in absorption spectra for the FMO protein from green sulfur bacteria. Photosynth Res 100: 79–87

Tsukatani Y, Miyamoto R, Itoh S and Oh-oka H (2004) Function of a PscD subunit in the homodimeric reaction center complex of the photosynthetic green sulfur bacterium *Chlorobium tepidum* studied by insertional gene inactivation. Regulation of energy transfer and ferredoxin-mediated $NADP^+$ reduction on the cytoplasmic side. J Biol Chem 279: 51122–51130

Tsukatani Y, Azai C, Kondo T, Itoh S and Oh-oka H (2008) Parallel electron donation pathways to cytochrome c_z in the type I homodimeric photosynthetic reaction center complex of *Chlorobium tepidum*. Biochim Biophys Acta 1777: 1211–1217

Tsukatani Y, Wen J, Blankenship RE and Bryant DA (2010) Characterization of the bacteriochlorophyll *a*-binding, Fenna-Matthews-Olson protein from *Candidatus* Chloracidobacterium thermophilum. Photosynth Res 104: 201–209

Turova TP, Spiridonova EM, Slobodova NV, Bulygina ES, Keppen OI, Kuznetsov BB and Ivanovskiĭ RN (2006) Phylogeny of anoxygenic filamentous phototrophic bacteria of the family *Oscillochloridaceae* as inferred from comparative analyses of the *rrs*, *cbbL*, and *nifH* genes. Mikrobiologia 75: 235–244

Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS and Banfield JF (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428: 37–43

Vande Weghe JG and Ow DW (1999) A fission yeast gene for mitochondrial sulfide oxidation. J Biol Chem 274: 13250–13257

Vassilieva EV, Frigaard N-U and Bryant DA (2000) Chlorosomes: the light-harvesting complexes of the green bacteria. The Spectrum 13: 7–13

Vassilieva EV, Antonkine ML, Zybailov BL, Yang F, Jakobs CU, Golbeck JH and Bryant DA (2001) Electron transfer may occur in the chlorosome envelope: the CsmI and CsmJ proteins of chlorosomes are 2Fe-2S ferredoxins. Biochemistry 40: 564–473

Vassilieva EV, Stirewalt VL, Jakobs CU, Frigaard N-U, Inoue-Sakamoto K, Baker MA, Sotak A and Bryant DA (2002) Subcellular localization of chlorosome proteins in *Chlorobium tepidum* and characterization of three new chlorosome proteins: CsmF, CsmH, and CsmX. Biochemistry 41: 4358–4370

Verdecia MA, Larkin RM, Ferrer JL, Riek R, Chory J and Noel JP (2005) Structure of the Mg-chelatase cofactor GUN4 reveals a novel hand-shaped fold for porphyrin binding. PLoS Biol 3:e151

Vogl K, Glaeser J, Pfannes KR, Wanner G and Overmann J (2006) *Chlorobium chlorochromatii* sp. nov., a symbiotic green sulfur bacterium isolated from the phototrophic consortium "*Chlorochromatium aggregatum*." Arch Microbiol 185: 363–372

Vogl K, Wenter R, Dressen M, Schlickenrieder M, Plöscher M, Eichacker L and Overmann J (2008) Identification and analysis of four candidate symbiosis genes from '*Chlorochromatium aggregatum*,' a highly developed bacterial symbiosis. Environ Microbiol 10: 2842–2856

Wada K, Yamaguchi H, Harada J, Niimi K, Osumi S, Saga Y, Oh-oka H, Tamiaki H and Fukuyama K (2006) Crystal structures of BchU, a methyltransferase involved in bacteriochlorophyll *c* biosynthesis, and its complex with S-adenosylhomocysteine: implications for reaction mechanism. J Mol Biol 360: 839–849

Wahlund TM, Woese CR, Castenholz RW and Madigan MT (1991) A thermophilic green sulfur bacterium from New Zealand hot springs, *Chlorobium tepidum* sp. nov. Arch Microbiol 156: 81–90

Walter MR, Bauld J and Brock TD (1972) Siliceous algal and bacterial stromatolites in hot spring and geyser effluents of Yellowstone National Park. Science 178: 402–405

Wanner G, Vogl K and Overmann J (2008) Ultrastructural characterization of the prokaryotic symbiosis in "*Chlorochromatium aggregatum*." J Bacteriol 190: 3721–3730

Ward DM (2006) Microbial diversity in natural environments: focusing on fundamental questions. Antonie Van Leeuwenhoek 90: 309–324

Ward DM, Ferris MJ, Nold SC and Bateson MM (1998) A natural view of microbial diversity within hot spring cyanobacterial mat communities. Microbiol Mol Biol Rev 62: 1353–1370

Ward DM, Bateson MM, Ferris MJ, Kühl M, Wieland A, Koeppel A and Cohan FM (2006) Cyanobacterial ecotypes in the microbial mat community of Mushroom Spring (Yellowstone National Park, Wyoming) as species-like units linking microbial community composition, |structure and function. Phils Trans R Soc Lond B Biol Sci 361: 1997–2008

Ward NL, Challacombe JF, Janssen PH, Henrissat B, Coutinho PM, Wu M, Xie G, Haft DH, Sait M, Badger J, Barabote RD, Bradley B, Brettin TS, Brinkac LM, Bruce D, Creasy T, Daugherty SC, Davidsen TM, DeBoy RT, Detter JC, Dodson RJ, Durkin AS, Ganapathy A, Gwinn-Giglio M, Han CS, Khouri H, Kiss H, Kothari SP, Madupu R, Nelson KE, Nelson WC, Paulsen I, Penn K, Ren Q, Rosovitz MJ, Selengut JD, Shrivastava S, Sullivan SA, Tapia R, Thompson LS, Watkins KL, Yang Q, Yu C, Zafar N, Zhou L and Kuske CR (2009) Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. Appl Environ Microbiol 75: 2046–2056

Watanabe Y, Feick RG and Shiozawa JA (1995) Cloning and sequencing of the genes encoding the light-harvesting B806-866 polypeptides and initial studies on the transcriptional organization of *puf2B*, *puf2A* and *puf2C* in *Chloroflexus aurantiacus*. Arch Microbiol 63: 124–30

Wätzlich D, Bröcker MJ, Uliczka F, Ribbe M, Virus S, Jahn D and Moser J (2009) Chimeric nitrogenase-like enzymes of (bacterio)chlorophyll biosynthesis. J Biol Chem 284: 15530–15540

Wen J, Zhang H, Gross ML and Blankenship RE (2009) Membrane orientation of the FMO antenna protein from *Chlorobaculum tepidum* as determined by mass spectrometry-based footprinting. Proc Natl Acad Sci USA 106: 6134–6139

Widdel F, Schnell S, Heising S, Ehrenreich A, Assmus B and Schink B (1993) Ferrous iron oxidation by anoxygenic phototrophic bacteria. Nature 362: 834–836

Willows RD (2003) Biosynthesis of chlorophylls from protoporphyrin IX. Nat Prod Rep 20: 327–341

Willows RD and Kriegel AM (2009) Biosynthesis of bacteriochlorophylls in purple bacteria. In: Hunter CN, Daldal F, Thurnauer MC and Beatty JT (eds) Advances in Photosynthesis and Respiration, Vol 28, pp 57–79. Springer, Dordrecht

Wu D, Raymond J, Wu M, Chatterji S, Ren Q, Graham JE, Bryant DA, Robb F, Colman A, Tallon LJ, Badger JH, Madupu R, Ward NL and Eisen JA (2009) Complete genome sequence of the aerobic CO-oxidizing thermophile *Thermomicrobium roseum*. PLoS One 4: e4207

Xin Y, Lin S, Montaño GA and Blankenship RE (2005) Purification and characterization of the B808-866 light-harvesting complex from green filamentous bacterium *Chloroflexus aurantiacus*. Photosynth Res 86: 155–163

Xin Y, Lu YK, Fromme R, Fromme P and Blankenship RE (2009) Purification, characterization and crystallization of the menaquinol:fumarate oxidoreductase from the green filamentous photosynthetic bacterium *Chloroflexus aurantiacus*. Biochim Biophys Acta 1787: 86–96

Xiong J (2006) Photosynthesis: what color was its origin? Genome Biol 7: article 245

Xiong J and Bauer CE (2002a) Complex evolution of photosynthesis. Annu Rev Plant Biol 53: 503–521

Xiong J and Bauer CE (2002b) A cytochrome *b* origin of photosynthetic reaction centers: an evolutionary link between respiration and photosynthesis. J Mol Biol 322: 1025–1037

Xiong J, Fischer WM, Inoue K, Nakahara M and Bauer CE (2000) Molecular evidence for the early evolution of photosynthesis. Science 289: 1724–1730

Yamada M, Zhang H, Hanada S, Nagashima KV, Shimada K and Matsuura K (2005) Structural and spectroscopic properties of a reaction center complex from the chlorosome-lacking filamentous anoxygenic phototrophic bacterium *Roseiflexus castenholzii*. J Bacteriol 187: 1702–1709

Yanyushin MF, del Rosario MC, Brune DC and Blankenship RE (2005) New class of bacterial membrane oxidoreductases. Biochemistry 44: 10037–10045

Yoon K-S, Hille R, Hemann C and Tabita FR (1999) Rubredoxin from the green sulfur bacterium *Chlorobium tepidum* functions as an electron acceptor for pyruvate ferredoxin oxidoreductase. J Biol Chem 274: 29772–29778

Yoon K-S, Bobst C, Hemann CF, Hille R and Tabita FR (2001) Spectroscopic and functional properties of novel 2[4Fe-4S] cluster-containing ferredoxins from the green sulfur bacterium *Chlorobium tepidum*. J Biol Chem 276: 44027–44036

Youvan DC, Bylina EJ, Alberti M, Begusch H and Hearst JE (1984) Nucleotide and deduced polypeptide sequences of the photosynthetic reaction-center, B870 antenna, and flanking polypeptides from *R. capsulata*. Cell 37: 949–957

Yu L, Zhao J, Mühlenhoff U, Chitnis PR, Bryant DA and Golbeck JH 1993) PsaE is required for cyclic electron flow around photosystem I in the cyanobacterium *Synechococcus* sp. PCC 7002 and *Synechocystis* sp. PCC 6803. Plant Physiol 103: 171–180

Zarzycki J, Schlichting A, Strychalsky N, Müller M, Alber BE and Fuchs G (2008) Mesaconyl-coenzyme A hydratase, a new enzyme of two central carbon metabolic pathways in bacteria. J Bacteriol 290: 1366–1374

Zarzycki J, Brecht V, Müller M and Fuchs G (2009) Final steps of an autotrophic pathway: closing the 3-hydroxypropionate bicycle in *Chloroflexus aurantiacus*. Proc Natl Acad Sci USA 106: 21317–21322

Chapter 4

Recent Functional Genomics Studies in Marine *Synechococcus*

Brian Palenik*
Scripps Institution of Oceanography, University of California, San Diego (UCSD), 9500 Gilman Drive, La Jolla 92093-0202, CA, USA

Summary

Marine *Synechoccocus* are major contributors to global primary productivity. Genomics and metagenomics have revealed high levels of gene content diversity in these cyanobacteria, partly due to horizontal gene transfer. These differences would be extremely important for ecological niche adaptation. Functional genomics studies using microarrays are now revealing how gene expression in marine cyanobacteria is responding to common environmental stresses such as nutrient deprivation, metal stress, and even the presence of other microbes. Many genes highly expressed under environmental stresses seem to be clade – or even strain-specific, which may change our view of how microbes adapt to new environmental conditions.

*Author for correspondence, e-mail: bpalenik@ucsd.edu

R.L. Burnap and W.F.J. Vermaas (eds.), Functional Genomics and Evolution of Photosynthetic Systems, Advances in Photosynthesis and Respiration 33, pp. 103–118, DOI 10.1007/978-94-007-1533-2_4,

I. Introduction

The strikingly high abundance of cyanobacteria in marine waters was discovered only in 1979 (Waterbury et al., 1979; Johnson and Sieburth, 1979) and subsequent studies documented the importance of marine *Synechococcus* to primary productivity and their near global distribution (Waterbury et al., 1986; Olson et al., 1990). Although considered ecologically as one functional group, it soon became apparent that *Synechococcus* has significant genetic diversity as seen in %GC studies and in some phenotypic characteristics (Waterbury et al., 1986). Molecular sequence evidence from field samples showed the presence of distinct genetic clusters (Palenik, 1994; Toledo and Palenik 1997; Ferris and Palenik, 1998) and that these sometimes had distinct distributions in the water column suggesting distinct species with different ecological niches. Due to the ongoing debate on the nature of microbial species, the marine cyanobacterial field has used the term "clades" or sometimes "ecotypes" and in *Synechococcus* there are approximately nine of these that have been given number or roman numeral designations, although new clades likely remain to be described (Scanlan et al., 2009). The understanding and classification of *Prochlorococcus*, a phylogenetically related "sister" group of *Synechococcus* lacking phycobilisomes, has followed a similar trajectory after its discovery in 1988 (Chisholm et al., 1988). Indeed the first genomes of these organisms were obtained and published together (Dufresne et al., 2003; Palenik et al., 2003; Rocap et al., 2003). In the brief span of a few decades we have gone from near complete ignorance to complete genomes for these abundant photosynthetic microbes.

Additional marine *Synechococcus* (and *Prochlorococcus*) genomes have been obtained from divergent clades and different ecosystems (Palenik et al., 2006) and a recent analysis (Dufresne et al., 2008) and a review (Scanlan et al., 2009) of eleven *Synechococcus* genomes have been published. It should be noted that many cyanobacteria described as *Synechococcus* (*e.g.*, the important cyanobacterial model *Synechococcus* PCC 7002) are completely unrelated and not found in marine waters. Recently, metagenomic studies are also adding to our picture of marine *Synechococcus* diversity, suggesting they have a "core" genome (shared by all *Synechococcus*), potentially a clade-specific genome, and a variable, possibly strain-specific, genome dominated by horizontal gene transfer (Palenik et al., 2009). For example, when comparing four *Synechococcus* genomes from four different *Synechococcus* clades from coastal and open ocean niches, the core genome is estimated at around 1,500 genes while each genome has hundreds of "unique" genes, including niche-specific and completely "unique" ones (Fig. 4.1). Including more genomes, as in Dufresne et al. (2008), confirms this view.

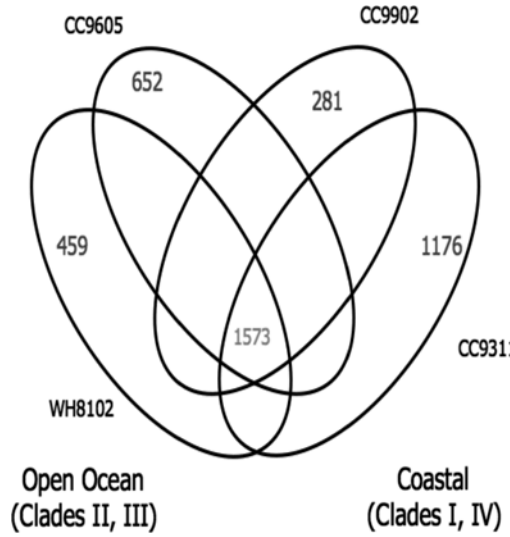


Fig. 4.1. Clusters of strains seen in environmental DNA samples using single gene markers such as *rpoC* (RNA polymerase subunit) have guided strain isolation and ultimately genome sequencing by the Joint Genome Institute (strains WH8102, CC9605, CC9902) and JCVI (CC9311). These initial strains were chosen to also represent coastal and open ocean strains. These results show an approximately 1,500-gene *Synechococcus* core with "unique" genes in the several-hundred-gene range.

Abbreviations: *amt1* – ammonium transporter; *apcF* – gene encoding light-harvesting phycobilisome core component-allophycocyanin subunit; *chlB* – gene encoding light-independent protochlorophyllide reductase subunit; Fur – Ferric uptake regulator, a transcription factor controlling iron uptake; *gltS* – gene encoding monocomponent sodium-dependent glutamate permease; *nirA* – gene encoding nitrite reductase; NtcA – transcription factor that is key element for nitrogen control; *petF2* – gene encoding ferredoxin

The availability of these whole and metagenomes has transformed the ability of researchers to do functional studies to better understand metabolic networks in cyanobacteria, their regulatory networks, the processes of genome evolution, and the relationships between genome and environment. The goal of this chapter is to summarize some very recent functional genomics studies in marine *Synechococcus* carried out using whole genome microarrays and to possibly compare the results to those in other model cyanobacteria.

II. Functional Genomics/Microarray Studies

Little is known about how marine cyanobacteria regulate gene expression in response to various significant environmental parameters changing in their environment such as nitrogen, phosphate, or micronutrient concentrations. Marine cyanobacteria may regulate many processes similarly to all cyanobacteria or there may be adaptations to the marine environment that have necessitated specific regulatory adaptations. Certainly one of the early insights from genomics was that marine *Synechococcus* has a drastically reduced number of two component regulatory systems (typically five histidine sensor kinases) compared to the major freshwater model systems *Synechococcus* PCC7942, *Synechocystis* PCC 6803, or *Anabaena/Nostoc* PCC 7120 suggesting that gene regulation may be greatly simplified or less specifically tailored to specific environmental changes (Palenik et al., 2003). In addition, different cyanobacterial species (or clades in the *Synechococcus* literature) may also show significant differences in gene expression, a topic that has been largely unexplored in the literature on cyanobacteria. A coastal *Synechococcus* strain sp. CC9311 had nearly twice the number of two-component systems (still low compared to other cyanobacteria) suggesting that the ecological niche may influence the evolution of regulatory systems (Palenik et al., 2006). The coastal to open ocean transition shows dramatic decreases in major nutrients from micromolar to nanomolar as well as changes in light quality, mixing regimes, etc., all of which could affect the regulatory systems need by species in these different types of environments. To investigate gene regulation in two significantly different *Synechococcus* strains, *Synechococcus* sp. WH8102 (clade III, open ocean) and *Synechococcus* sp. 9311 (clade I, coastal), full genome microarrays have been used, consisting of a mix of PCR amplicons and 70-mer oligonucleotides (Tetu et al., 2009; Stuart et al., 2009; Su et al., 2006; Tai et al., 2009; Thomas et al., 2009). The major findings from these studies are described below.

III. Nitrogen

Nitrogen is one of the major nutrients needed by cyanobacteria and other photoautotrophs (followed by phosphorus and then trace metals in smaller amounts). Nitrogen metabolism is likely somewhat different in different cyanobacterial species, at least as seen in the differential utilization of common nitrogen sources (Moore et al., 2002) and the presence of different nitrogen transporters found in cyanobacterial genomes (http://www.membranetransport.org/; Ren and Paulsen, 2007). Also, some cyanobacteria are able to fix N_2. Whole-genome microarray gene expression data were collected under ammonium-rich and nitrate-rich conditions. Su et al. (2006) observed that 247 genes were down-regulated, while 91 genes were up-regulated by growth on ammonium based on the D-values from the SAM package (Su et al., 2006). It is highly likely that these 338 genes, either down- or up-regulated on ammonium, are involved in the nitrogen assimilation process or related global responses as nitrogen metabolism would likely be highly integrated with carbon metabolism.

Some well-known genes and processes were down-regulated by growth on ammonia, including the nitrogen regulator NtcA (SYNW0275), GlnA (SYNW1073) involved in nitrogen assimilation, and transporters for urea, nitrate, and cyanate. Similar results have been found for *Prochlorococcus* (Tolonen et al., 2006) and a general overview of nitrogen metabolism in cyanobacteria provides a useful context for these results (Flores and Herrero, 2005). Interestingly the genes involved in global stress responses were also down-regulated. This suggests growth on nitrate might be more "stressful" than growth on ammonia, which would be consistent with the use of additional electron flow to reduce nitrate to ammonia.

These microarray analyses were done in the context of an effort to computationally model the nitrogen metabolism network of *Synechococcus* using diverse approaches including the bioinformatic prediction of the nitrogen regulator NtcA binding sites. Interestingly, although the predictions were highly statistically significant, they verified only 31.6% (42/133) of the genes that were predicted to be parts of the nitrogen assimilation network using the microarray data. Several possible reasons might account for this relatively low number. First, the microarray experimental conditions (ammonium versus nitrate as the sole nitrogen source) might represent a mild perturbation to the whole nitrogen assimilation network. This might be one of the reasons that some predicted genes of the NtcA regulon were not differentially expressed when the cells were grown on ammonium versus nitrate, including *amt1* (SYNW0253), *nirA* (SYNW2477), *gltS* (SYNW0882) *apcF* (SYNW1074) and *petF2* (SYNW1274). These genes might be constitutively expressed at similar (high) levels of ammonium or nitrate. We expect that more genes will be found to be up- or down-regulated under stronger perturbations. Starvation for nitrogen clearly needs active research using time course microarray studies.

Coordination between the nitrogen assimilation and photosynthesis processes was apparent in the nitrogen microarray data set. In agreement with the prediction that the photosynthetic gene *psbA3* (SYNW2151) is a member of the NtcA regulon, the microarray data indicate that *psbA3* was down-regulated by ammonium. In addition, an NtcA binding site was found for the putative operon SYNW2124-SYNW2123, which encodes the photosystem I P700 chlorophyll *a* apoprotein subunits Ia (PsaA, SYNW2124) and Ib (PsaB, SYNW2123). Both SYNW2124 and SYNW2123 were up-regulated by ammonium, suggesting that this operon might be negatively regulated by NtcA through the predicted NtcA binding site. Further analyses of the connections between NtcA and photosynthesis can be found in Su et al. (2005).

The role of secondary rather than direct effects of NtcA is likely also important in nitrogen regulation. Intriguingly, many photosynthetic genes lacking a canonical NtcA promoter were down-regulated by ammonium, and these included the ferredoxin-thioredoxin reductase catalytic ß chain gene (SYNW0318), the cytochrome b_6f complex subunit (Rieske iron-sulfur protein) gene (SYNW1841), the photosystem II chlorophyll-binding protein CP43 gene (SYNW0676), the photosystem II D1 protein form II gene (SYNW0983), the 3Fe-4S ferredoxin gene (SYNW0624), and the high-light-inducible protein genes (SYNW2403 and SYNW0330). On the other hand, some other photosynthesis-related genes were up-regulated by growth on ammonium that also do not bear canonical NtcA binding sites in their regulatory regions, including the C-phycoerythrin class II chain genes (SYNW2008 and SYNW2010), the C-phycoerythrin class I alpha chain gene (SYNW2016), and the light-independent protochlorophyllide reductase *chlBLN* genes (SYNW1723, SYNW1724 and SYNW1725). These results may be due to the storage of fixed nitrogen under high ammonium in the production of phycobilisomes, as these are some of the first components of the cell that are degraded under nitrogen stress (Yamanaka and Glazer, 1980). Since some photosynthetic genes can be either down- or up-regulated by ammonium, but lack NtcA promoters/binding sites, they are probably indirectly regulated by the downstream events of NtcA or carbon balancing, possibly by some of the eight putative transcription regulators (SYNW0926, SYNW2401, SYNW1592, SYNW1875, SYNW0808, SYNW2289, SYNW1462 and SYNW0549) that were down- or up-regulated by ammonium. In addition, it is also possible that other transcriptional regulators are not altered in their own transcription, but have their activity modulated by N-dependent changes in metabolic co-inducers or co-repressors and thus serve an important role of in nitrogen-carbon balancing. Taken together, these results unequivocally demonstrate the strong interactions of nitrogen metabolism and carbon fixation networks.

IV. Phosphorus

Phosphorus is thought to be a limiting nutrient in some areas of the ocean, with the Sargasso Sea in the North Atlantic being an area of substantial study (Palenik and Dyhrman, 1998). *Synechococcus* and *Prochlorococcus* strains show

variations in phosphorus metabolism (Moore et al., 2005) and coastal *Synechococcus* genomes seem to have lost regulatory components associated with phosphorus likely because phosphorus is relatively abundant in coastal environments (Palenik et al., 2006). Gene expression in *Synechococcus* sp. WH8102 has been compared between cells grown in phosphate-replete and -depleted conditions (Tetu et al., 2009). For P-deficient cultures, cells were sampled at early P-stress, when alkaline phosphatase was first detectable, or at late P-stress, corresponding to late-log phase and greatly reduced growth. This approach allowed the identification of genes that were affected by the onset of P-stress or by extended P-stress.

In response to early P-stress a total of 617 genes were found to be significantly up-regulated, while the expression of 649 genes was negatively affected. Genes whose early expression was induced or repressed more than 2-fold under P-stress conditions included a set of 36 up-regulated genes and 23 down-regulated genes, a somewhat more manageable list. Most of the 36 genes observed to be strongly induced under early P-stress conditions have not been directly studied in this organism (with the exception of the motility genes *swmA* and *swmB* (Brahamsha, 1996a; McCarren and Brahamsha, 2007). However, bioinformatic analyses provided clues to the function of 24 genes of these, while 12 genes encoded hypothetical (3) or conserved hypothetical (9) proteins. A high proportion of the strongly repressed genes, 14 of 23, are annotated as hypothetical (10) or conserved hypothetical (4) proteins.

A. Genes Induced Under Conditions of P-Stress

Among the 24 strongly induced genes for which a function was predicted, one third are hypothesized to play a role in transport. Two of these, SYNW2223 and SYNW2224, encode putative outer membrane porins, which we hypothesize to be involved in P transport. The putative proteins encoded by these genes show similarity to each other and to outer membrane porins, *somA* and *somB*, which have been characterized in *Synechococcus* sp. PCC 6301 (Hansel and Tadros, 1998) and to porins observed to be up-regulated under depleted nitrogen conditions in *Synechococcus* sp. PCC7942 (Sauer et al., 2001). The remaining six are predicted to encode components of P-specific ABC transporters. All three genes of the ABC transporter consisting of *pstC*, *pstA* and *pstB* homologs (SYNW1270, SYNW1271 and SYNW1272) were significantly up-regulated. Of the four putative P-specific solute-binding proteins encoded within this genome, three were up-regulated by >2 fold (SYNW1018, SYNW1286, SYNW1815) while the fourth, SYNW2507 was also up-regulated significantly, but fell just below the >2-fold cutoff (1.9-fold). Under the conditions utilized in this study, expression of genes comprising a putative phosphonate ABC transporter (SYNW1169, SYNW1168 and SYNW1170) was not significantly affected.

Also prevalent among the strongly induced genes were putative alkaline phosphatases. Contiguous genes SYNW2390 and SYNW2391 show homology to domains within an atypical alkaline phosphatase, PhoA (Moore et al., 2005), previously characterized in *Synechococcus* sp. PCC7942 (Ray et al., 1991) and in *Synechocystis* sp. PCC6803 (*sll0654*) (Hirani et al., 2001), where they were also found to be highly expressed under P-stress conditions. The other induced putative phosphatase, SYNW0196, has no homolog in any other sequenced *Synechococcus* genome. However, BLASTP searches indicate that related phosphatases are present in other cyanobacterial genera, such as *Gloeobacter, Nostoc* and *Anabaena*. The conserved hypothetical protein SYNW0762 also contained a region with similarity to putative alkaline phosphatase genes. However, other matches are to phytase genes, which would specifically remove phosphate from phosphorylated inositols. The presence of a phytase-like insertion in a homologous atypical alkaline phosphatase, SYNWH7803_0111, in *Synechococcus* sp. WH7803 was previously noted (Moore et al., 2005).

The remaining 16 significantly up-regulated genes for which a functional prediction is available are involved in a range of cellular processes. SYNW1019, which lies directly downstream of the phosphate binding protein SYNW1018 (*pstS*), was previously designated *ptrA* (putative transcriptional regulator) and speculated to play a role in global P regulation, based on its proximity to *pstS* in *Synechococcus* sp. WH7803 and the

similarity to members of the CRP family of regulators (Scanlan et al., 1996). The regulation of *ptrA* in response to phosphate limitation supports the possibility that this regulator is involved in a P-specific response.

A number of genes that are involved in aspects of carbohydrate metabolism were up-regulated under low P conditions. This includes genes SYNW1119 (*gnd*) and SYNW1120 (*pgl*), which catalyze steps in the pentose phosphate pathway, and SYNW0799 (*gap3*), which catalyses the conversion of D-glycerate 1,3-bisphosphate to glyceraldehyde-3-phosphate (and the reverse reaction) during glycolysis/glyconeogenesis. SYNW0156, which encodes a putative starch phosphorylase, was also among the set of significantly up-regulated genes. Additionally, one of the up-regulated conserved hypothetical genes, SYNW0160, contains a conserved domain, GalM, indicating possible galactose mutarotase activity. Previous work in other microorganisms has likewise indicated that aspects of carbohydrate metabolism are affected by P-starvation. In *E. coli* the *gnd* homolog of SYNW1119 was similarly observed to be significantly up-regulated under conditions of P-limitation (Van Bogelen et al., 1996) while *gap1* was observed to be one of only two genes up-regulated in both MED4 and MIT9313 strains of *Prochlorococcus* following P-stress (Martiny et al., 2006). This *gap1* appears to be the likely ortholog of SYNW0799 *gap3*, also up-regulated by P-stress. The metabolic role of *gap3* in the context of P-stress is unknown.

Other aspects of cellular metabolism may also be affected by P-stress. Significant up-regulation of SYNW1529, cytochrome *c* oxidase subunit I, was observed in this work, indicating that aspects of energy metabolism may alter under this stress condition. Interestingly, SYNW1213, encoding a thioredoxin peroxidase was also significantly up-regulated. In *Synechocystis* sp. PCC 6803, this enzyme has been shown to be coupled to photosynthetic electron transport (Yamamoto et al., 1999) and it is unclear why this enzyme appears responsive to P-stress.

Among the most interesting observations to emerge from this examination of the Pho regulon in *Synechococcus* is that the genes for three known cell-surface proteins are among this set of significantly up-regulated genes. SYNW0953 (*swmB*) and SYNW0085 (*swmA*) encode polypeptides involved in the unique swimming motility observed in this genus (Brahamsha, 1996a; McCarren and Brahamsha, 2007), and SYNW0406 encodes the heavily glycosylated 70-kDa integral outer-membrane polypeptide found in *Synechococcus* sp. WH8102 (Brahamsha, 1996a) (B. Brahamsha, unpublished). One explanation for this may be that phosphate stress results in a reorganization of the cell envelope, or that increased expression of the motility genes facilitates movement of cells (chemotaxis) towards regions of higher P concentration, a possibility that has not been studied to date.

Only five of the genes strongly (>2-fold) up-regulated under early P-stress were also strongly up-regulated at late P stress, indicating that the transcriptional response to early P-stress is, in most cases, transitory. This may indicate that transitory increases in gene expression are sufficient to generate a protein response, which can support phosphate utilization over more extended time periods. In addition, very few of the genes strongly up-regulated only after extended P-stress play an obvious role in phosphate metabolism. Many of the ones that are most affected are conserved hypotheticals or of unknown function. One exception is phosphoglycerate mutase (SYNW0519), involved in transfer of phosphate groups within glycerate molecules during glycolysis, which was strongly up-regulated after prolonged P-limitation. Several photosystem I genes were up-regulated while various photosystem II genes were down-regulated under late P-stress. This might be consistent with an enhanced use of photosystem I-dependent cyclic photophosphorylation to generate ATP. Interestingly, two genes with putative regulatory roles were among the ones that were strongly up-regulated at this time point, including SYNW2037, encoding a possible *pex* (*pe*riod *ex*tender) gene, thought to play a role in regulation of the circadian clock in these cyanobacteria (Kutsuna et al., 2007) and SYNW1660, also a putative transcriptional regulator.

B. Genes Repressed Under Conditions of P-Stress

Among the nine functionally characterized genes whose expression was significantly repressed under early P-stress conditions, four play a role in

translation. SYNW1149, peptide-chain-release factor RF-3, involved in termination of protein synthesis, displayed the greatest degree of down-regulation. However, SYNW2340, SYNW2341 and SYNW2082, encoding 50S ribosomal proteins, were also significantly repressed under early P-stress. The suppression of genes involved in protein synthesis has been observed in other microorganisms in response to both P- (Ishige et al., 2003; Martiny et al., 2006) and N-limitation (Silberbach et al., 2005) and has been suggested to be a result of a general reduction in the metabolic rate of the cells. After extended P-stress, genes SYNW1717 and SYNW1718, encoding ribulose bisphosphate carboxylase (both large and small subunits), were among the most strongly repressed indicating that such nutrient limitation, if experienced for extended periods, may be extremely detrimental to fundamental carbon fixation processes in these cells.

It is clear from these studies and those on nitrogen metabolism that future microarray studies of nutrient stress would greatly benefit from time courses, but the logistics of obtaining sufficient RNA from marine *Synechococcus* as well as the still high cost of microarray studies has made this difficult to achieve.

C. Gene Knockouts of the Putative Phosphate Two Component System

Because of the development of genetics in marine *Synechococcus* (in contrast to its sister *Prochlorococcus*) it is possible to study genetic regulation in more detail using knockout mutants in combination with microrrays (Thomas et al., 2009). In other well-studied bacteria, including cyanobacteria, a two-component regulator, comprised of a sensory kinase (*phoB*/*sphR*) and response regulator gene (*phoR*/*sphS*), are commonly observed to play a key role in the cellular response to P-limitation (Aiba et al., 1993; Suzuki et al., 2004). Based on sequence similarity, the most apparent candidate genes for controlling phosphate regulation in *Synechococcus* sp. WH8102 are SYNW0947 and SYNW0948, encoding a probable phosphate sensor kinase and response regulator (Scanlan and West, 2002; Palenik et al., 2003). To investigate the role of these genes in this organism, knockouts were generated in each of these genes by insertional inactivation (Brahamsha, 1996b). Comparison of the transcript levels in the knockout mutants of SYNW0947 and SYNW0948 with their isogenic parent (all grown under P-replete conditions) supported the hypothesis that the two-component regulatory system encoded by these genes is important to the regulation of P-metabolism of *Synechococcus* sp. WH8102.

D. Regulation of the P-Stress Response

There is considerable overlap in the set of genes whose expression was significantly affected by inactivation of SYNW0947 and SYNW0948 (even under P-replete conditions), and those were observed to be responsive to P-stress. A cluster analysis of the microarray data from each set of experiments provides statistical support for the observation that similar sets of genes were affected by low (compared to replete) P and loss of SYNW0947 or SYNW0948. Two clusters of interest emerged. The first included genes encoding alkaline phosphatases and the second cluster contained genes involved in aspects of P-transport and metabolism as well as the possible transcriptional regulator SYNW1019, which may also play a role in the P-stress response in this organism. All members of this first cluster were strongly affected by P-stress and in the sensor kinase/response regulator knockout mutants, with observed $\log_2$ fold changes of 2.0–4.1 under P-stress conditions at the early time point.

The second cluster was observed to be less strongly affected, with members showing $\log_2$ fold changes of 0.6–1.8 under P-stress conditions (at the early time point). This set of genes included members of the pentose phosphate pathway and others linked to P metabolism such as the NUDIX protein, SYNW1334, as well as many expected to play a role in the transport of P. Interestingly, genes found in this cluster also encode key enzymes in oxidative phosphorylation and amino acid metabolism pathways. The upregulation of both cytochrome *c* oxidase subunit I and II may function to ensure that sufficient ATP is generated from the available P in the cell.

The response regulator, PhoB, affects transcription by binding to a conserved sequence motif, known as a Pho box, located upstream of genes in the Pho regulon. The Pho box motif has been well characterized in enteric bacteria (Wanner, 1993)

and *Bacillus subtilis* (Qi et al., 1997). Work on cyanobacteria has likewise shown that the response regulator, SphR, functions by binding to conserved Pho box sequences (Suzuki et al., 2004). Comparison of these regions indicated that the consensus sequence of the Pho box in cyanobacteria differed from that of *E. coli*, in this case being comprised of repetitive PyTTAAPyPy(T/A)-like sequences (Suzuki et al., 2004). More recently, computational prediction was used to generate a profile for Pho boxes in 16 sequenced cyanobacterial genomes in order to explore the Pho regulons of these organisms (Su et al., 2007). This work predicted Pho boxes for 38 independent transcriptional units in *Synechococcus* sp. WH8102, only five of which correlate to genes observed in this study to be up-regulated by more than 2-fold under early P-stress (SYNW2390, SYNW2391, SYNW1019, SYNW0085 and SYNW0165). None of the ≥2-fold down-regulated genes were observed to be part of these Pho box associated transcriptional units.

The long-term value of linking computational predictions and experimental validation is that we will eventually gain a solid basis for predicting metabolic outcomes of environmental perturbations in these ecologically important microbes. However, this goal still seems frustratingly distant.

V. Copper

Despite being found at low concentrations, copper's roles as both a micronutrient and a toxicant have lead to laboratory and field studies showing its likely importance influencing cyanobacterial abundance and diversity (Moffett et al., 1997; Mann et al., 2002). A microarray analysis of copper metabolism (at excess concentrations) in *Synechococcus* has recently been published (Stuart et al., 2009). A total of four sets of global expression microarray experiments were conducted, two for the coastal strain, CC9311 (clade I), and two for the open ocean strain, WH8102 (clade III) and this is one of the first studies to show differential gene expression between different *Synechococcus* clades. For each strain both a moderate and high copper shock was applied, and sets of significantly regulated genes were determined. For CC9311 the moderate shock yielded 120 up-regulated and 78 down-regulated genes, of a possible 2,893 total protein-coding genes, meaning that 6.8% of the protein-coding genes were differentially regulated. The high-copper shock yielded 317 up-regulated and 528 down-regulated genes (29.2% differentially regulated). The overlap of significant genes between these two treatments was fairly high, with 122 genes differentially regulated under both shocks of 198 possible genes (this being the total number of significant genes in the moderate shock), which indicates a physiologically consistent response between the two treatments. For WH8102 the moderate shock yielded 129 up-regulated and 138 down-regulated genes of a possible 2,528 total protein-coding genes, with 10.3% of all protein-coding genes differentially regulated. The high-copper shock had 45 up-regulated and 98 down-regulated genes (5.5% differentially regulated). Again, the overlap of significant genes between the two treatments was high, with 54 genes of 143 possible genes regulated under both treatments.

The overlap of orthologous genes between the two strains was modest, with 24 orthologous genes significantly regulated under high-copper shock in both strains and 16 orthologous genes overlapping under moderate-copper shock. However, there were also common types of responses, not involving directly orthologous genes. For example, the putative osmoregulatory two-component response regulator was up-regulated in both strains, but there were also accompanying osmoregulatory responses in each strain not involving orthologs. The open ocean strain up-regulated an osmolyte transporter, glycine betaine (SYNW0229), and the coastal strain up-regulated a gene involved in osmolyte synthesis (glucosylglycerol-phosphate phosphatase, sync_1171). Furthermore, both strains highly down-regulated several porins each under both shock types, although only one pair, sync_1542/SYNW2224, is orthologous.

Aside from the shared osmoregulatory response there was also a common photosynthetic response. Based on KEGG classifications, photosynthesis-related genes are grouped into several categories including PSI, PSII, cytochrome b_6f complex, photosynthetic electron transport, F-type ATPase and antenna proteins. There are 87 of these genes in WH8102 and 81 in CC9311 and excluding the antenna proteins, which vary between the two strains, the other categories contain orthologous genes between strains except for one extra gene

in the strain WH8102 in the electron transport category. Of the 37 total photosystem genes (PSI and PSII), 14 are down-regulated in the coastal strain and 10 in the open ocean strain, under high-copper shock. Neither strain up-regulated any of the PSI, PSII, photosynthetic electron transport or F-type ATPases under the high-copper shock. However, under moderate-copper shock, both strains up-regulated F-type ATPase genes as well as genes involved in carbon fixation. The coastal strain upregulated photosynthetic antenna proteins (including phycoerythrin, phycocyanin and allophycocyanin) and down-regulated cytochrome b_6f complex genes whereas WH8102 down-regulated photosynthetic antenna proteins.

Finally, there were also several conserved hypothetical proteins that were up-regulated in both strains including one, SYNW0921/sync_1050, with a signaling peptide and four conserved cysteines as some of the only conserved residues, creating a possible metal-binding domain. Other conserved hypothetical proteins including SYNW1510/sync_1903 appear to be stress-related and come up under other shock conditions.

There were several clear differences in response between the two strains, the first being the type of stress response that was activated. The coastal strain activated genes that have been linked to oxidative stress response in bacteria (Imlay, 2008) including up-regulation of ferritin to sequester iron and keep it from redox cycling, iron-sulfur (FeS) assembly proteins to repair damaged FeS clusters that are especially sensitive to oxidative stress, methionine sulfoxide reductase and ferredoxin thioredoxin reductase to reduce unwanted disulfide bonds created by oxidative stress, and components of the pentose phosphate pathway to generate reducing power. Conversely, the open-ocean strain mostly up-regulated general stress response genes, that is, protein folding genes, chaperones, and proteases such as heat shock proteins, GroEL/GroES, Clp proteases and DnaJ/DnaK. Excepting one, HtpG, these genes were not up-regulated at all in CC9311, and under high-copper shock many of these genes actually were significantly down-regulated. However, CC9311 is known to up-regulate these genes under other shock conditions, such as mitomycin C shock (unpublished).

Another intriguing difference was the coastal strain's activation of many genes that are not conserved in the other 11 sequenced marine *Synechococcus* strains, as compared to the open ocean strain, which activated mostly conserved genes. Of particular interest among the CC9311 "unique" genes were two operons activated under both shock types. The first, sync_1213–1217, has genes for an ATP-binding cassette (ABC) transporter, an efflux mediator and a large protein, likely effluxed, with a divalent metal cation binding domain and a protein-binding domain. The entire operon is predicted to be horizontally acquired based on atypical trinucleotide usage, and the closest gene ortholog appears in another marine cyanobacterium, *Cyanobium* sp. PCC7001. The second operon, sync_1491–1494, includes genes for a sigma-70 factor, a thioredoxin-like protein, a biogenesis protein and a possible anti-sigma factor. The genes in this operon had the highest fold changes of the whole dataset in both the high and moderate shock, and have orthologs in only one other *Synechococcus* strain, WH5701, but they do not have atypical trinucleotide usage. This operon is also of interest as it represents a difference in regulatory response between the two strains. CC9311 activated several sigma factors whereas WH8102 did not activate any. Of the other "unique" CC9311 genes activated under copper stress there were several others with atypical trinucleotide content, some of which have orthologs in marine bacteria. These results suggest that putative horizontally transferred genes can be highly induced by stress conditions and likely improve the fitness of *Synechococcus* strains that possess them. These are often induced by stress (*e.g.*, excess copper, phosphate depletion, etc.) and this fact may be significant in understanding their acquisition and maintenance by the cells.

Other differences between WH8102 and CC9311 under copper stress include the expression of transport and membrane proteins, but for a coverage of these aspects the reader is referred to Stuart et al. (2009).

VI. Interactions with Heterotrophic Bacteria

Cyanobacteria do not normally live in axenic culture, but instead are found with eukaryotic phytoplankton, heterotrophic bacteria, heterotrophic eukaryotic grazers, and phages. All of these could affect their gene expression as they respond to different co-occurring microbes. This has been

the subject of a fascinating microarray study of the changes in gene expression in *Prochlorococcus* during phage infection (Lindell et al., 2007). In order to study a simple co-culture of *Synechococcus* and a heterotrophic bacterium, a co-culture with *Vibrio parahaemolyticus* was analyzed. *V. parahaemolyticus* can be found in marine ecosystems, grows well with marine *Synechococcus*, and has a large genome allowing potentially diverse signals and responses to be expressed (Makino et al., 2003).

A. Gene Expression Changes in WH8102 When Grown with V. Parahaemolyticus

A statistical analysis of replicated microarray experiments revealed that 285 and 295 genes were significantly up- and down-regulated, respectively, in WH8102 when grown with *V. parahaemolyticus* (Tai et al., 2009). Genes from several major functional categories changed in expression level. Transporter and photosynthesis genes comprised the highest number of up-regulated genes. In addition, genes involved in cell wall biosynthesis, glycosyltransferases, and nucleotide biosynthesis were more prominently up-regulated rather than down-regulated. The most prominent categories of down-regulated genes also included transporter and photosynthesis genes. However, particularly genes for amino acid and co-factor biosynthesis, proteases, post-translational modification, and DNA replication, recombination, and repair were typically down-regulated. Also down-regulated were 4 out of 8 cytochrome b_6f genes and 9 out of 16 NADH dehydrogenase genes. In regards to photosynthesis, genes responsible for the synthesis of chlorophyll and phycobilins increased in expression, while genes primarily related to photosystem II were down-regulated. As the co-cultures received adequate amounts of light and the cells were not self-shaded, light limitation does not explain the increase in expression of chlorophyll and phycobilin synthesis genes.

Interesting patterns have emerged from the microarray data, but the reasons for these gene expression changes are complex. As opposed to microarray experiments testing a single factor such as a change in phosphate concentration, the presence of *V. parahaemolyticus* has caused or produced numerous chemical, nutritive, or enzymatic changes that were responsible for the altered gene expression. By categorizing the changes in the gene expression patterns, the possible causal factors can be enumerated and especially appear to involve changes in phosphate and nitrogen metabolism.

B. Phosphate Stress in WH8102 Co-cultures

In cyanobacteria, many of the genes related to phosphate stress are regulated by the same mechanism. For example, low phosphate availability typically activates a transcription factor through a two-component sensor kinase system (Hirani et al., 2001) and the phosphate binding protein, PstS, was found to be specifically expressed when the availability of inorganic phosphate was low (Scanlan et al., 1993). WH8102 generally follows this pattern as described above and in more detail in Tetu et al. (2009).

Many genes involved in phosphate acquisition and the phosphate stress response were significantly up-regulated in WH8102 when grown with *V. parahaemolyticus* and none were down-regulated. The phosphate-binding protein of an ABC phosphate transporter homologous to PstS (SYNW1018) and a phosphate transcriptional regulator (SYNW1019) were among the most significantly up-regulated genes. Genes for two additional phosphate transporters and other genes involved in phosphate acquisition also increased in expression including alkaline phosphatases and a possible phytase. Based on qRT-PCR, phosphate acquisition genes also showed the greatest increase in transcript level, ranging from 2.7 to 5.3 $\log_2$ fold increases.

These increases in the expression of phosphate stress genes occurred despite the availability of inorganic phosphate. Typically, phosphate must be nearly depleted (< 0.05 μM) to induce increased expression and synthesis of phosphate stress genes in cyanobacteria (Ray et al., 1991; Scanlan et al., 1997). The artificial seawater media used initially contained 100 μM phosphate and was never completely depleted.

C. Expression of Cell Wall Modification Genes in Co-cultured WH8102

A major response of WH8102 to the presence of *V. parahaemolyticus* was the increased expression of cell wall biosynthesis genes. In addition,

9 glycosyltransferases were also up-regulated. Glycosyltransferases have many functions in the cell, but they are prominent in the formation of cell walls and exopolysaccharide layers. Ten of these cell-wall biosynthesis and glycosyltransferase genes are located together on the genome, transcribed in the same direction, and may be function as an operon. Based on these results, it is likely that WH8102 has modified the composition of its cell wall due to the presence of *V. parahaemolyticus*.

D. Metal Regulation and Oxidative Stress May Be Linked Through a Ferric Uptake Regulation (Fur)-family Gene

Genes mediating photosystem II, oxidative stress, and genes requiring metal co-factors were often down-regulated in WH8102 when grown with *V. parahaemolyticus.* These genes are connected physiologically because photosynthesis is a process requiring many metalloenzymes, particularly iron, and generates reactive oxygen species (Latifi et al., 2009).

The expression of these genes may be regulated by a ferric uptake regulation (Fur)-family gene (SYNW2401). This gene decreased in expression when WH8102 was grown with *V. parahaemolyticus*. In *Escherichia coli,* Fur was first recognized as a transcriptional repressor that requires iron as a cofactor and regulates iron acquisition in bacteria by repressing iron transporters and siderophores when iron is sufficient (Bagg and Neilands, 1987; Escolar et al., 1999). Fur-family genes also regulate genes related to oxidative stress such as superoxide dismutases (Niederhoffer et al., 1990) in addition to genes involved in acid shock response, chemotaxis, metabolism, and virulence (Escolar et al., 1999).

Several genes typically induced by Fur were down-regulated in WH8102 when grown with *V. parahaemolyticus* including the cytochrome b_6f complex, ferrodoxins, superoxide dismutase and other anti-oxidant enzymes. This result may be due to iron-limitation in WH8102 when it was co-cultured. A recent study with *E. coli* reports that Fur regulates the expression of iron-containing respiratory proteins and that this may be a method for bacteria to mediate iron-protein levels based on iron availability (McHugh et al., 2003). A similar homeostatic process may be occurring in WH8102 co-cultures.

Alternatively, SYNW2401 may be zinc-related or a general metal-regulator since zinc-transporters showed increased expression. Iron transporters, on the other hand, were not up-regulated as would be expected if SYNW2401 was a classical Fe-Fur-type transcription factor. Often, in addition to an iron-regulated Fur, bacteria have multiple Fur-family genes that regulate gene expression for many functions, in particular zinc homeostasis and oxidative stress (Hantke, 2001). Most cyanobacteria, including all of the fully sequenced marine *Synechococcus* strains, have three Fur-family genes, but their exact functions are not known (Hernandez et al., 2004).

E. Co-cultured WH8102 May Use Ammonia Produced from V. parahaemolyticus

Interestingly, when WH8102 was grown in monoculture with ammonia as opposed to nitrate as the nitrogen source, some similar patterns emerge as when grown with *Vibrio* species. In comparison with a microarray analysis of WH8102 cells grown with ammonia (Su et al., 2006), the expression of 75% and 58% of the up- and down-regulated genes, respectively, changed similarly when WH8102 was grown with *V. parahaemolyticus.* In particular, ammonia-grown WH8102 and WH8102+*V. parahaemolyticus* co-cultures shared decreased expression of a nitrate transporter (SYNW2463) and urease genes (SYNW2447 and SYNW2449) compared to nitrate-grown WH8102 monocultures. However, the global nitrogen regulatory protein, NtcA, did not change in expression level in WH8102 when grown with *V. parahaemolyticus* whereas it decreased in expression when grown with ammonia as the sole nitrogen source (Su et al., 2006).

These similarities in gene expression patterns could be due to the use of ammonia by WH8102 that was generated by the heterotroph, *V. parahaemolyticus*. Heterotrophic marine bacteria are known to regenerate ammonia particularly if they are carbon-limited (Goldman and Dennett, 1991). Ammonia concentrations measured from the media when the cells were harvested for RNA extraction were approximately 1 μM in both mono- and co-cultures. Although differences in the ammonia concentrations between the mono- and

co-cultures were not observed, rapid transport and turn-over of the ammonia produced by *V. parahaemolyticus* would explain the lack of an accumulation of ammonia in the co-cultures.

F. Mechanisms by Which Synechococcus Gene Expression Is Changed in Co-culture

The mechanism by which *V. parahaemolyticus* affected expression of WH8102 phosphate-related genes, nitrogen metabolism, and cell wall biosynthesis is unknown, but a few possible mechanisms are: (1) *V. parahaemolyticus* may have secreted a protease that degrades phosphate uptake proteins, phosphate or other sensors, and thus WH8102 metabolism may have become unbalanced. (2) Pathogenic strains of *V. parahaemolyticus* are known to secrete hemolysins that lyse erythrocytes by forming pores or by phospholipase activity (Zhang and Austin, 2005) and the expression of hemolysins increases under phosphate limitation (McCarter and Silverman, 1987). Their effect on bacterial cells is not known but they might have interfered with nutrient sensing or acquisition and they could induce cell wall changes seen in WH8102. (3) *V. parahaemolyticus* may have secreted a molecule that affected the regulation of phosphate and/or nitrogen utilization genes in WH8102. For example, *Vibrio* species are known to excrete quorum sensing molecules, such as acyl homoserine lactones (Henke and Bassler, 2004). In particular, *V. parahaemolyticus* can secrete molecules that induce gene expression in other *Vibrio* species (Bassler et al., 1997) and this could also occur across more divergent bacteria. While no one mechanism has been proven to be responsible, these mechanisms are rather general and could be expected to been found in future studies of cyanobacteria/heterotrophic bacteria interactions.

VII. Concluding Remarks

Microarray analyses characterizing diverse physiological conditions have helped to elucidate regulatory networks in marine *Synechococcus*. Each physiological condition had some unique aspects due to the specific nature of that stress, *e.g.*, nitrogen source affects the nitrogen transporters that are part of the nitrogen metabolism network. However, the microarray analyses reveal the interconnectedness of regulation in *Synechococcus*. Any one condition also seems to lead to significant changes in gene expression in other systems, especially for photosynthesis and general stress responses, which were often highly differentially expressed (up or down) across the diverse stresses studied to date. An extreme example was the growth of *Synechococcus* under non-axenic conditions, where multiple cellular processes were affected relative to axenic culture. At the same time, and not elaborated on in the above sections, a large fraction of genes have not shown differential regulation under the conditions tested to date and seem either relatively constant or not expressed at all. It is not yet clear if this is specific to *Synechococcus*, marine cyanobacteria, or cyanobacteria in general, or is an artifact due to our inability to detect small changes in relative gene expression using microarrays.

Despite the range of conditions examined to date, many more studies of *Synechococcus* gene expression would be valuable in elucidating their regulation under common environmental stresses. In particular light quantity and quality effects on gene expression remain to be determined. Such studies would be valuable on strains from different clades and could contrast chromatic adapters versus clades that do not seem to possess this capability (Palenik, 2001). These studies have been carried out on *Prochlorococcus* with interesting results (Steglich et al., 2006). Iron limitation, as an ecologically important stress, also would likely yield interesting insights into *Synechococcus* regulatory networks.

Unfortunately, microarray studies have revealed our ignorance of cyanobacteria genomes as well. One difficult problem is that we do not know which hypothetical genes are miscalled ORFs and which are metabolically significant. Fortunately, microarrays can help in solving this problem as microarray experiments over diverse conditions will accumulate information about each predicted gene in the cell. Presumably hypothetical genes showing no expression over many experiments become candidates for misannotated ORFs.

Many hypothetical genes and conserved hypothetical genes are differentially expressed, but we do not yet know their functions in the cell. Microarrays may help reveal these functions, but

genetic and biochemical studies will be needed to confirm microarray-based predictions.

Current technology using PCR or oligonucleotides spotted on microarrays typically represents genes and ORFs annotated in the genome analysis. It has become clear that small non-coding RNAs are present in marine cyanobacteria and could play roles in regulation (Voß et al., 2009). New tiling arrays where all the genome is represented are needed to begin to examine these RNAs in the context of the changes in expression in protein-coding genes.

As metagenomes reveal the conserved and mosaic aspects of *Synechococcus* genomes, it is clear from the microarray studies that putative horizontally transferred genes, even those found in only some strains in the same clade, are differentially expressed and thus are likely to contribute to the fitness of that strain. Some of these horizontally transferred genes now seem to be related to stress responses in some way that is still a mystery. In other words and perhaps not surprising, many horizontally transferred genes may not be part of conserved cyanobacterial metabolism but part of new strategies a *Synechococcus* strain may have acquired that can be used to survive environmental stresses.

The availability of whole genomes of *Synechococcus* has made possible the development of whole genome microarrays and laboratory studies to probe cyanobacterial metabolic and regulatory networks. This is still ongoing, but it is possible that new technologies such as high-throughput sequencing of mRNA will change the way these experiments are carried out. In addition, an important next step will be to study changes in gene expression in field conditions where the cells must integrate diverse environmental parameters. However, both new technologies and new types of studies will likely rely on the model *Synechococcus* strain genomes that became available as part of the genomics revolution.

Acknowledgments

Much of the microarray work summarized here was a collaboration with a number of scientists including Ian Paulsen, Bianca Brahamsha, Qinhu Ren, Sasha Tetu, Kathy Phillipy, Aaron Johnson, Chris Dupont, Rhona Stuart, Vera Tai, Rob Herman, Lori Crumbliss, Ed Thomas, Zhengchang Su, and Ying Xu. This chapter borrows liberally from their work and collective insights into *Synechococcus* biology. In addition I would like to thank DOE and NSF for providing funding that led to the sequencing and analysis of the genomes of strains WH8102, CC9902, CC9605 (DOE-JGI) and CC9311 (JCVI) and their use in microarray studies.

References

Aiba H, Nagaya M and Mizuno T (1993) Sensor and regulator proteins from the cyanobacterium *Synechococcus* species PCC7942 that belong to the bacterial signal-transduction protein families: implication in the adaptive response to phosphate limitation. Mol Microbiol 8: 81–91

Bagg A and Neilands JB (1987) Molecular mechanism of regulation of siderophore-mediated iron assimilation. Microbial Rev 51: 509–518

Bassler BL, Greenberg EP and Stevens AM (1997) Cross species induction of luminescence in the quorum sensing bacterium *Vibrio harveyi*. J Bacteriol 179: 4043–4045

Brahamsha B (1996a) An abundant cell-surface polypeptide is required for swimming by the nonflagellated marine cyanobacterium *Synechococcus*. Proc Natl Acad Sci USA 93: 6504–6509

Brahamsha B (1996b) A genetic manipulation system for oceanic cyanobacteria of the genus *Synechococcus*. Appl Environ Microbiol 62: 1747–1751

Chisholm SW, Olson RJ, Zettler ER, Goericke R, Waterbury JB and Welschmeyer NA (1988) A novel free living prochlorophyte abundant in the oceanic euphotic zone. Nature 334: 340–343

Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP, Paulsen IT, de Marsac NT, Wincker P, Dossat C, Ferriera S, Johnson J, Post AF, Hess WR and Partensky F (2008) Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. Genome Biology 9: R90

Dufresne A, Salanoubat M, Partensky F, Artiguenave F, Axmann IM, Barbe V, Duprat S, Galperin MY, Koonin EV, Le Gall F, Makarova KS, Ostrowski M, Oztas S, Robert C, Rogozin IB, Scanlan DJ, de Marsac NT, Weissenbach J, Wincker P, Wolf YI and Hess WR (2003) Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. Proc Natl Acad Sci USA 100: 10020–10025

Escolar L, Perez-Martin J and De Lorenzo V (1999) Opening the iron box: transcriptional metalloregulation by the fur protein. J Bacteriol 181: 6223–6229

Ferris MJ and Palenik B (1998) Niche adaptation in ocean cyanobacteria. Nature 396: 226–228

Flores E and Herrero A (2005) Nitrogen assimilation and nitrogen control in cyanobacteria. Biochem Soc Trans 33: 164–167

Goldman JC and Dennett MR (1991) Ammonium regeneration and carbon utilization by marine bacteria grown on mixed substrates. Mar Biol 109: 369–378

Hansel A and Tadros MH (1998) Characterization of two pore-forming proteins isolated from the outer membrane of *Synechococcus* PCC 6301. Curr Microbiol 36: 321–326

Hantke K (2001) Iron and metal regulation in bacteria. Curr Opin Microbiol 4: 172–177

Henke JM and Bassler BL (2004) Bacterial social engagements. Trends Cell Biol 14: 648–656

Hernandez JA, Lopez-Gomollon S, Bes MT, Fillat MF and Peleato ML (2004) Three *fur* homologues from *Anabaena* sp PCC7120: exploring reciprocal protein-promoter recognition. FEMS Microbiol Lett 236: 275–282

Hirani TA, Suzuki I, Murata N, Hayashi H and Eaton-Rye JJ (2001) Characterization of a two-component signal transduction system involved in the induction of alkaline phosphatase under phosphate-limiting conditions in *Synechocystis* sp. PCC 6803. Plant Mol Biol 45: 133–144

Imlay JA (2008) Cellular defenses against superoxide and hydrogen peroxide. Ann Rev Biochem 77: 755–776

Ishige T, Krause M, Bott M, Wendisch VF and Sahm H (2003) The phosphate starvation stimulon of *Corynebacterium glutamicum* determined by DNA microarray analyses. J Bacteriol 185: 4519–4529

Johnson PW and Sieburth JM (1979) Chroococcoid cyanobacteria in the sea: a ubiquitous and diverse phototrophic biomass. Limnol Oceanogr 24: 928–935

Kutsuna S, Kondo T, Ikegami H, Uzumaki T, Katayama M and Ishiura M (2007) The circadian clock-related gene *pex* regulates a negative cis element in the *kaiA* promoter region. J Bacteriol 189: 769–7696

Latifi A, Ruiz M and Zhang CC (2009) Oxidative stress in cyanobacteria. FEMS Microbiol Rev 33: 258–278

Lindell D, Jaffe JD, Coleman ML, Futschik ME, Axmann IM, Rector T, Kettler G, Sullivan MB, Steen R, Hess WR, Church GM and Chisholm SW (2007) Genome wide expression dynamics of a marine virus and host reveal features of co-evolution. Nature 449: 83–86

Makino K, Oshima K, Kurokawa K, et al (2003) Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae.* Lancet 361: 743–749

Mann EL, Ahlgren N, Moffett JW and Chisholm SW (2002) Copper toxicity and cyanobacteria ecology in the Sargasso Sea. Limnol Oceanogr 47: 976–988

Martiny AC, Coleman ML and Chisholm SW (2006) Phosphate acquistion genes in *Prochlorococcus* ecotypes: Evidence for genome wide adaptation. Proc Natl Acad Sci USA 103: 12552–12557

McCarren J and Brahamsha B (2007) SwmB, a 1.12-megadalton protein that is required for nonflagellar swimming motility in *Synechococcus*. J Bacteriol 189: 1158–1162

McCarter LL and Silverman M (1987) Phosphate regulation of gene expression in *Vibrio parahaemolyticus.* J Bacteriol 169: 3441–3449

McHugh JP, Rodriguez-Quinones F, Abdul-Tehrani H, Svistunenko DA, Poole RK, Cooper CE and Andrews SC (2003) Global iron-dependent gene regulation in *Escherichia coli*- A new mechanism for iron homeostasis. J Biol Chem 278: 29478–29486

Moffett JW, Brand LE, Croot PL and Barbeau KA (1997) Cu speciation and cyanobacterial distribution in harbors subject to anthropogenic Cu inputs. Limnol Oceanogr 42: 789–799

Moore LR, Ostrowski M, Scanlan DJ, Feren K and Sweetsir T (2005) Ecotypic variation in phosphorus acquisition mechanisms within marine picocyanobacteria. Aquatic Microbial Ecol 39: 257–269

Moore LR, Post AF, Rocap G and Chisholm SW (2002) Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. Limnol Oceanog 47: 989–996

Niederhoffer EC, Naranjo CM, Bradley KL and Fee JA (1990) Control of *Escherichia coli* superoxide dismutase (SodA and SodB) genes by ferric uptake regulation (Fur) locus. J Bacteriol 172: 1930–1938

Olson RJ, Chisholm SW, Zettler ER and Armbrust EV (1990) Pigments, size and distribution of *Synechococcus* in the North Atlantic and Pacific Oceans. Limnol Oceanog 35: 45–58

Palenik B (1994) Cyanobacterial community structure as seen from RNA-polymerase gene sequence analysis. Appl Environ Microbiol 60: 3212–3219

Palenik B (2001) Chromatic adaptation in marine *Synechococcus* strains. Appl Environ Microbiol 67: 991–994

Palenik B, Brahamsha B, Larimer FW, Land M, Hauser L, Chain P, Lamerdin J, Regala W, Allen EE, McCarren J, Paulsen I, Dufresne A, Partensky F, Webb EA and Waterbury J (2003) The genome of a motile marine *Synechococcus*. Nature 424: 1037–1042

Palenik B and Dyhrman ST (1998). Recent progress in understanding the regulation of marine primary productivity by phosphorus. In: Lynch JP and Deikman J (eds) Phosphorus in Plant Biology: Regulatory Roles in Molecular, Cellular, Organismic and Ecosystem Processes, pp 26–38. American Society of Plant Physiologists Rockville, MD

Palenik B, Ren Q, Dupont CL, Myers GS, Heidelberg JF, Badger JH, Madupu R, Nelson WC, Brinkac LM, Dodson RJ, Durkin AS, Daugherty SC, Sullivan SA, Khouri H, Mohamoud Y, Halpin R and Paulsen IT (2006) The genome of *Synechococcus* CC9311: Insights into adaptation

to a coastal environment. Proc Natl Acad Sci USA 103: 13555–13559

Palenik B, Ren Q, Tai V and Paulsen IT (2009) Coastal *Synechococcus* metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity. Environ Microbiol 11: 349–359

Qi Y, Kobayashi Y and Hulett FM (1997) The *pst* operon of *Bacillus subtilis* has a phosphate--regulated promoter and is involved in phosphate transport but not in regulation of the pho regulon. J Bacteriol 179: 2534–2539

Ray JM, Bhaya D, Block MA and Grossman AR (1991) Isolation transcription and inactivation of the gene for an atypical alkaline phosphatase of *Synechococcus* sp. strain PCC 7942. J Bacteriol 173: 4297–4309

Ren Q and Paulsen IT (2007) Large-scale comparative genomic analyses of cytoplasmic membrane transport systems in prokaryotes. J Mol Microbiol Biotechnol 12: 165–179

Rocap G, Larimer F, Lamerdin J, Malfatti S, Chain P, Ahlgren N, Arellano A, Coleman ML, Hauser L, Hess W, Johnson Z, Land M, Lindell D, Post A, Regala W, Shah M, Shaw S, Steglich C, Sullivan MB, Ting C, Tolonen A, Webb E, Zinser E and Chisholm SW (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. Nature 424: 1042–1047

Sauer J, Schreiber U, Schmid R, Volker U and Forchhammer K (2001) Nitrogen starvation-induced chlorosis in *Synechococcus* PCC 7942. Low-level photosynthesis as a mechanism of long-term survival. Plant Physiol 126: 233–243

Scanlan D, Bourne J and Mann N (1996) A putative transcriptional activator of the Crp/Fnr family from the marine cyanobacterium *Synechococcus* sp. WH7803. J Appl Phycol 8: 565–567

Scanlan DJ, Mann NH and Carr NG (1993) The response of the picoplankton marine cyanobacterium *Synechococcus* species WH7803 to phosphate starvation involves a protein homologous to the periplasmic phosphate-binding protein of *Escherichia coli*. Mol Microbiol 10: 181–191

Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR, Post AF, Hagemann M, Paulsen I and Partensky F (2009) Ecological genomics of marine picocyanobacteria. Microbiol Mol Biol Rev 73: 249–299

Scanlan DJ, Silman NJ, Donald KM, Wilson WH, Carr NG, Joint I and Mann NH (1997) An immunological approach to detect phosphate stress in populations and single cells of photosynthetic picoplankton. Appl Environ Microbiol 63: 2411–2420

Scanlan DJ and West NJ (2002) Molecular ecology of the marine cyanobacterial genera *Prochlorococcus* and *Synechococcus*. FEMS Microbiol Ecol 40: 1–12

Silberbach M, Huser A, Kalinowski J, Puhler A, Walter B, Kramer R et al (2005) DNA microarray analysis of the nitrogen starvation response of *Corynebacterium glutamicum*. J Biotechnol 119: 357–367

Steglich C, Futschik M, Rector T, Steen R and Chisholm SW (2006) Genome-wide analysis of light sensing in *Prochlorococcus*. J Bacteriol 188: 7796–7806

Stuart RK, Dupont CL, Johnson DA, Paulsen IT and Palenik B (2009) Coastal strains of marine *Synechococcus* species exhibit increased tolerance to copper shock and a distinctive transcriptional response relative to those of open-ocean strains. Appl Environ Micro 75: 5047–5057

Su Z, Olman V, Mao F and Xu Y (2005) Comparative genomics analysis of NtcA regulons in cyanobacteria: regulation of nitrogen assimilation and its coupling to photosynthesis. Nucl Acids Res 33: 5156–5171

Su Z, Olman V and Xu Y (2007) Computational prediction of Pho regulons in cyanobacteria. BMC Genomics 8: Article 156

Su ZC, Mao FL, Dam P, Wu HW, Olman V, Paulsen IT, Palenik B and Xu Y (2006) Computational inference and experimental validation of the nitrogen assimilation regulatory network in cyanobacterium *Synechococcus* sp WH 8102. Nucl Acids Res 34: 1050–1065

Suzuki S, Ferjani A, Suzuki I and Murata N (2004) The SphS-SphR two component system is the exclusive sensor for the induction of gene expression in response to phosphate limitation in *Synechocystis*. J Biol Chem 279: 13234–13240

Tai V, Paulsen IT, Phillipy K, Johnson DA and Palenik B (2009) Whole-genome microarray analyses of *Synechococcus-Vibrio* interactions. Environ Microbiol 11: 2698–2709

Tetu SG, Brahamsha B, Johnson DA, Tai V, Phillipy K, Palenik B and Paulsen IT (2009) Microarray analysis of phosphate regulation in the marine cyanobacterium *Synechococcus* sp. WH8102. ISME J: 1–15

Thomas EV, Phillippy KH, Brahamsha B, Haaland DM, Timlin JA, Elbourne LDH, Palenik B and Paulsen IT (2009) Statistical analysis of microarray data with replicated spots: a case study with *Synechococcus* WH8102. Comp Funct Genom Article 950171

Toledo G and Palenik B (1997) *Synechococcus* diversity in the California Current as seen by RNA polymerase (*rpoC1*) gene sequences of isolated strains. Appl Environ Microbiol 63: 4298–4303

Tolonen AC, Aach J, Lindell D, Johnson ZI, Rector T, Steen R, Church GM and Chisholm SW (2006) Global gene expression of *Prochlorococcus* ecotypes in response to changes in nitrogen availability. Molec Syst Biol 2: 53

Van Bogelen RA, Olson ER, Wanner BL and Neidhardt FC (1996) Global analysis of proteins synthesized during phosphorus restriction in *Escherichia coli*. J Bacteriol 178: 4344–4366

Voß B, Georg J, Schön V, Ude S and Hess WR (2009) Biocomputational prediction of non--coding RNAs in model cyanobacteria. BMC Genomics 10: 123

Wanner BL (1993) Gene regulation by phosphate in enteric bacteria. J Cell Biochem 51: 47–54

Waterbury JB, Watson FW, Valois FW and Franks DG (1986) Biological and ecological characterization of the marine unicellular cyanobacterium *Synechococcus*. In: Platt T and Li WKW (eds) Photosynthetic Picoplankton, pp 71–120. Canadian Dept. of Fisheries and Oceans, Ottawa

Waterbury JB, Watson SW, Guillard RR and Brand LE (1979) Widespread occurrence of a unicellular marine planktonic cyanobacterium. Nature 227: 293–294

Yamamoto H, Miyake C, Dietz K-J, Tomizawa K-I, Murata N and Yokota A (1999) Thioredoxin peroxidase in the cyanobacterium *Synechocystis* sp. PCC 6803. FEBS Lett 447: 269–273

Yamanaka G and Glazer AN (1980) Dynamic aspects of phycobilisome structure - phycobilisome turnover during nitrogen starvation in *Synechococcus* sp. Archives of Microbiol 124: 39–47

Zhang XH and Austin B (2005) Haemolysins in *Vibrio* species. J Appl Microbiol 98: 1011–1019

Chapter 5

Functional Analysis of the *Synechococcus elongatus* PCC 7942 Genome

You Chen
Synthetic Genomics Inc, 11149 North Torrey Pines Road, La Jolla, CA 92037, USA

C. Kay Holtman
Department of Biology, Center for Biological Clocks Research, Texas A&M University, College Station, TX 77843-3258, USA

Arnaud Taton
Center for the Study of Biological Complexity, Virginia Commonwealth University, 1000 W. Cary Street, Richmond, VA 23284, USA

Susan S. Golden*
Division of Biological Sciences, Center for Chronobiology, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

*Author for correspondence, e-mail: sgolden@ucsd.edu

R.L. Burnap and W.F.J. Vermaas (eds.), Functional Genomics and Evolution of Photosynthetic Systems, Advances in Photosynthesis and Respiration 33, pp. 119–137, DOI 10.1007/978-94-007-1533-2_5,

Summary

Many genes related to the circadian clock have been discovered and studied in *Synechococcus elongatus* PCC 7942, the model organism for cyanobacterial circadian rhythms. However, the partners of some known clock components are still unidentified, and undiscovered pathways are predicted to exist that connect the central clock to other cellular functions. Identification of all clock components in *S. elongatus* is necessary for fully elucidating molecular mechanisms of the cyanobacterial circadian clock, as well as the relationship of the circadian clock to metabolism and other essential cellular activities. We adopted a transposon-mediated in vitro mutagenesis and sequencing strategy to disrupt essentially every locus in the genome and screen each insertional mutant for altered circadian phenotypes in *S. elongatus*. The completion of the genome sequence by the Department of Energy Joint Genome Institute greatly facilitated our functional genomics project, which is very close to the finish line with 88% of the genome mutagenized and more than 75% of loci screened for circadian function. Among the first 700 genes surveyed, 70 new clock loci were discovered that represent an array of functional categories.

I. Introduction

A. Synechococcus elongatus PCC 7942, a Genetic Model Organism for Cyanobacteria

The unicellular cyanobacterium *S. elongatus* PCC 7942 (hereafter *S. elongatus*), formerly called *Anacystis nidulans* R2, is a freshwater obligate photoautotroph. Its genome consists of a circular chromosome (~2.7 Mb, GenBank accession No. CP000100), a large endogenous plasmid (pANL, 46.3 kb, GenBank accession No. AF441790) (Chen et al., 2008), and a small plasmid, pANS (8.6 kb, GenBank accession No. S89470). The chromosome is 2,695,903 bp in length with a G+C content of ~55.47%. There are 2,612 predicted protein-coding genes and 51 RNA coding loci, including 6 rRNA genes in 2 operons and 45 tRNA genes. The coding regions represent ~88% of the genome and the average length of protein coding genes is 908 bp. Around 60% of protein-coding genes have been assigned with predicted functions based on their closest homologs or conserved domains. Others encode either conserved hypothetical proteins or hypothetical proteins without significant homologs in the available databases. The first nucleotide for numbering of the genome sequence is assigned near the *dnaN* locus encoding the *beta* subunit of DNA polymerase III (synpcc7942_0001). The replication origin of the chromosome has been suggested to be located in this region, which contains 11 DnaA boxes (TTTTCCACA) (Liu and Tsinoremas, 1996).

The genome sequence of a closely-related strain, *S. elongatus* PCC 6301, has also been determined (GenBank accession No. AP008231) (Sugita et al., 2007). Its chromosome is 2,696,255 bp long with 99.93% nucleotide sequence identity to that of *S. elongatus* PCC 7942. The major differences between these two strains are: (***1***) a single inversion, 188.6 kb in length and flanked by genes for two porin-like proteins and a pair of 20-bp inverted repeats, accounting for known RFLPs between the two genomes (Golden et al., 1989); and (***2***) *S. elongatus* PCC 7942 is naturally transformable, while *S. elongatus* PCC 6301 is not.

The facile transformation property of *S. elongatus* (Golden et al., 1987; Shestakov and Khyen, 1970) has made it a model organism for studies of photosynthesis and light-regulated gene expression (Capuano et al., 1993; Schaefer and Golden, 1989; Sippola et al., 1998; Soitamo et al., 1994; Spiegel and Bader, 2003), signal transduction (Aldehni et al., 2003; Hirani et al., 2001), transcription and its regulation (Bovy et al., 1993; Sugimoto et al., 1997), response to nutrient deprivation (Collier and Grossman, 1994; van Waasbergen et al., 2002), iron, sulfur, nitrogen, and carbon metabolism (Bovy et al., 1993; Bradley and Reddy, 1997; Durham et al., 2003; Garcia-Dominguez and Florencio, 1997; Green et al.,

Abbreviations: *S. elongatus* – *Synechococcus elongatus* PCC 7942; Cm – chloramphenicol; Km – kanamycin; ORF – open reading frame; PCR – polymerase chain reaction; RFLP – restriction fragment length polymorphism

1989; Hirani et al., 2001; Richaud et al., 2001; Vazquez-Bermudez et al., 2003), protein degradation (Stanne et al., 2007), and other biological questions for decades. In particular, *S. elongatus* is the only developed model organism for exploring the mechanism of a prokaryotic circadian clock (Dong and Golden, 2008; Golden, 2003).

B. The Cyanobacterial Circadian Clock

As an autonomous intrinsic biological timing mechanism, a circadian clock is distinguished by three major characteristics: a ~24 h free-running period of persistent daily rhythms under constant conditions; resetting of the phase of the rhythm by environmental cues, as needed for synchronization with the external cycles; and temperature compensation of the period over a range of physiologically relevant ambient temperatures (Dunlap et al., 2003). Most eukaryotic organisms, such as fungi (bread molds), insects (fruit flies), plants (rock cresses), birds (chicken), and mammals (mice, humans, etc.), have evolved an endogenous circadian timer to optimize their daily activities in a cyclic environment (Bell-Pedersen et al., 2005; Young and Kay, 2001). The fitness advantage of the circadian clock has been demonstrated not only in model eukaryotic organisms, such as *Arabidopsis* (Dodd et al., 2005; Michael et al., 2003) and *Drosophila* (Beaver et al., 2002), but also in cyanobacteria, the only prokaryotes known to possess a circadian clock (Johnson et al., 1998; Ouyang et al., 1998; Woelfle et al., 2004).

Studies on nitrogen fixation in some unicellular or non-heterocystous filamentous cyanobacteria revealed the earliest evidence for circadian rhythms in cyanobacteria (Huang and Chow, 1986; Mitsui et al., 1986; Stal and Krumbein, 1985); Sweeney and Borgese, 1989). Since then, more cyanobacterial species have been established to have endogenous circadian clocks, such as *Cyanothece* ATCC 51142 (Schneegurt et al., 1994), *Synechocystis* sp. PCC 6803 (Aoki et al., 1995), *Trichodesmium* IMS 101 (Berman-Frank et al., 2001; Chen et al., 1998) and *S. elongatus* (Kondo et al., 1993; Kondo and Ishiura, 1994).

It takes at least several days of continuous monitoring to establish a circadian phenotype, and requires the ability to screen hundreds or thousands of clones to use the phenotype for a forward genetic screen. For this purpose, a high-throughput, high-precision, non-invasive screening method for continuous recording of circadian rhythms was developed more than 15 years ago for *S. elongatus*. This method is based on artificial bioluminescent reporters constructed by fusing the promoterless *luxAB* gene set, which encodes the luciferase enzyme from the marine bacterium *Vibrio harveyi*, to the promoter of cyanobacterial genes of interest (Kondo and Ishiura, 1994; Kondo et al., 1993; Liu et al., 1995a). Alternatively, the firefly luciferase gene, *luc*, can also serve as a reporter (Andersson et al., 2000). Bioluminescence from cyanobacterial cells in 96-well microplates or individual colonies on agar plates can be counted automatically and continuously by a luminometer or cooled CCD camera. The resulting bioluminescence output over time provides an easily assay for mutated circadian phenotypes (Fig. 5.1).

The majority of the studies of cyanobacterial circadian rhythms have been done with *S. elongatus* because of the development of the luciferase-based automated bioluminescence reporter system (Golden and Canales, 2003; Mori and Johnson, 2001). The alternative model system for studying cyanobacterial clock function would be the unicellular facultative photoheterotroph *Synechocystis sp*. PCC 6803, whose 3.6 Mb genome was sequenced earlier (Ikeuchi, 1996). However, the circadian rhythms generated from reporter genes in *S. elongatus* are more robust than those from *Synechocystis*, and the smaller *S. elongatus* genome has less functional redundancy; for example, there is one *kaiC* gene, which encodes a central circadian oscillator component, in *S. elongatus* compared to three paralogs of *kaiC* in *Synechocystis*. The relative simplicity of the circadian clock in *S. elongatus* makes it a better system for elucidating the clock mechanism in cyanobacteria.

A circadian clock system consists of three major parts: input pathways – mechanisms for sensing environment cues, such as light and temperature, and setting the phase of the clock; a central oscillator – the timekeeper itself; and output pathways – a means of relaying clock phasing to the various behaviors controlled by the clock (Johnson and Hastings, 1986). Several genes necessary for clock function have been identified in *S. elongatus*, including *kaiA* and *kaiBC* (Ishiura et al., 1998), a locus that encodes the components

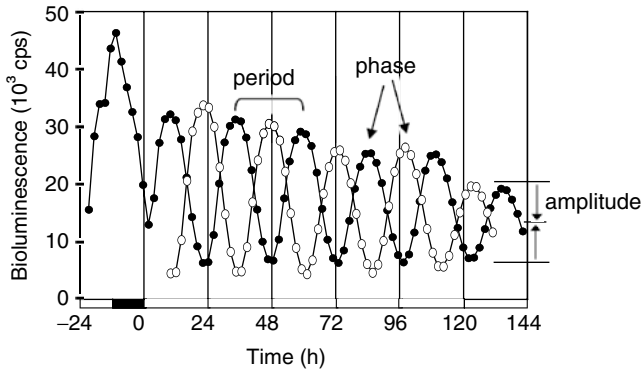


Fig. 5.1. Representative bioluminescence traces from reporter strains produce characteristic circadian patterns that depend on the promoter used to drive luciferase expression. Various mutations affect circadian period (length of cycle, *e.g.,* peak-to-peak), relative phasing (peak time relative to a reference point, such as lights on), or amplitude of the rhythm (deviation of peak and trough from the mean of the oscillation). X axis, circadian time in hours; closed circles, cells were released to constant light (0–144 h) after light/dark entrainment for one day (−24–0 h); open circles, cells were entrained in the opposite phase (the early time points were removed to simplify the figure). Y axis, bioluminescence plotted as 10^3 counts per second (cps).

of the central oscillator; *cikA* (Schmitz et al., 2000) and *ldpA* (Katayama et al., 2003), which encode input pathway components; the sensor kinase gene *sasA* (Iwasaki et al., 2000) and its cognate response regulator gene *rpaA* (Takai et al., 2006), which encode key output pathway components; and other clock-related loci such as *pex* (Kutsuna et al., 1998), *cpmA* (Katayama et al., 1999), *labA* (Taniguchi et al., 2007), the *clpP-2clpX* operon (Holtman et al., 2005), and group two sigma factor genes (*rpoD2*, *rpoD3*, *rpoD4* and *sigC*) (Nair et al., 2002; Tsinoremas et al., 1996; Fig. 5.2).

In addition to identification of these clock genes, tremendous progress has been achieved in elucidating the molecular mechanism of the central oscillator, as well as the input and output pathways of the cyanobacterial circadian clock (reviewed in Dong and Golden, 2008). The cyanobacterial clock tick-tocks using a post-translational oscillator, as robust 24-h KaiC phosphorylation cycles have been reconstituted in a test tube with only the purified KaiA, KaiB, and KaiC proteins and ATP (Nakajima et al., 2005). Details of the phosphorylation and dephosphorylation steps that occur during the cycle have been described by several labs (Kim et al., 2008; Kitayama et al., 2008; Pattanayek et al., 2008; Rust et al., 2007). The expression of the whole genome of *S. elongatus* is under the control of the circadian clock, which requires both SasA and RpaA, a two-component system pair, for strong oscillations (Iwasaki et al., 2000; Liu et al., 1995b; Takai et al., 2006). Temporal information transmitted from the central oscillator is also involved in a chromosomal compaction rhythm, which is temperature-compensated and Kai dependent, but SasA independent (Smith and Williams, 2006; Fig. 5.2).

Despite recent achievements in elucidating the molecular mechanism of the KaiC-based oscillator, many details of the basic steps in generating circadian rhythms of biological processes remain unsolved. For example, phase determination is still a mystery. Apparently, no specific cis element is responsible for determination the timing of peak gene expression (Min et al., 2004). The cyclic pattern of KaiC phosphorylation might not be relevant to phase (Chen et al., 2009; Kiyohara et al., 2005). It is likely that DNA topology or chromosome compaction is involved in phase determination and phase resetting (Min et al., 2004; Smith and Williams, 2006). We also do not understand how the circadian clock is embedded in cellular physiology in cyanobacteria. Many clock components are still missing, such as the input pathway protein(s) that connect CikA and KaiA (Schmitz et al., 2000).

To fully understand how the circadian clock in cyanobacteria functions at the molecular level and how is it entrained by environmental signals, and to further understand the physiological significance of circadian rhythms in cyanobacteria, it is necessary to identify all of the components that are required for circadian clock function. The most thorough approach to achieve this goal is to

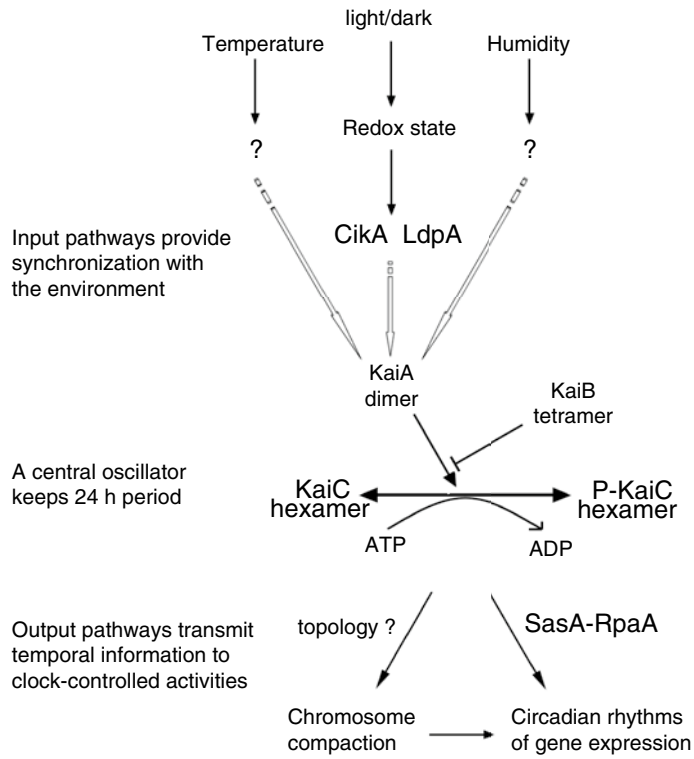


Fig. 5.2. Simplified depiction of the circadian clock in *S. elongatus*: The circadian clock includes a central oscillator, input pathways, and output pathways; key components of each of these divisions have been identified for *S. elongatus*, some of which are indicated here.

query each individual locus in the genome by mutagenesis and screen every resulting mutant for altered circadian clock phenotypes.

C. Transposon-Mediated Functional Genomics in Cyanobacteria

Transposon-based mutagenesis is widely used in functional genomic analysis of many organisms, including bacteria (Gehring et al., 2000; Mills, 2001; Pajunen et al., 2005), fungi (Michielse et al. 2005), plants (Alonso et al., 2003; Tadege et al., 2005), worms (Granger et al., 2004), mammals (Carlson and Largaespada, 2005; Kitada et al., 2007), and various pathogens (Balu and Adams, 2006; Bourhy et al., 2005; Su et al., 2007). Transposons are mobile DNA elements that are ubiquitous in prokaryotic and eukaryotic genomes and have become a powerful tool for insertional mutagenesis. Compared to other in vivo mutagenesis methods, such as chemical (Singer and Kusmierek, 1982) and UV mutagenesis (Lawrence et al., 1993), in which characterization of isolated point mutations and cloning of the mutated genes are often time-consuming, the main advantage of transposon-mediated mutagenesis is that transposons carry selectable markers, such as an antibiotic resistance gene that tags the affected locus and facilitates cloning (Maloy, 2007). In addition, transposons can be engineered to contain sequencing primer binding sites, signature tags, or even promoters, which greatly facilitate the subsequent molecular identification and genetic analysis of affected loci (Hayes, 2003). The feature that transposons can randomly insert into any locus in a genome, usually as a single insertion, makes it very convenient to construct pools of mutants in vivo or in vitro for global phenotypic analysis (Pobigaylo et al., 2006; Vidan and Snyder, 2001).

More recently, in vitro transposon mutagenesis techniques have been developed. The *in vitro* transposition systems offer many advantages over in vivo systems in functional genomics because of higher efficiency and less specificity with respect to the insertion site (Hayes, 2003; Mills, 2001). In addition, the in vitro reactions are generally more reliable and convenient for large-scale mutagenesis. Although commercial kits require

an investment, in vitro systems reap a savings in labor in analysis of insertion sites. Usually, cosmid or plasmid libraries of genomic DNA are mutated in vitro and then reintroduced into a target organism that favors double recombination (allelic substitution), through transformation or conjugation to disrupt target genes. Highly efficient in vitro mutagenesis systems have been developed for many well-studied transposons, such as Tn*5* (Goryshin and Reznikoff, 1998; Kirby, 2007), Tn*7* (Biery et al., 2000), and *Mu* (Haapa et al., 1999a), which are also commercially available with various antibiotic resistance markers: Tn*5* (EZ-Tn5™ Transposon Tools, Epicentre), Tn*7* (GPS™ Mutagenesis System, New England Biolabs) and *Mu* (GeneJumper™ Kit, Invitrogen).

Both in vivo and in vitro transposon mutagenesis, mainly based on Tn*5*, have been widely used in cyanobacteria. A Tn*5* derivative (Tn*5*-1063) was introduced into *Nostoc sp*. ATCC 29133 (*Nostoc punctiforme* PCC 73102) to generate mutants with defective phenotypes in nitrogen fixation (Cohen et al., 1994). Several cell division genes were identified in *S. elongatus* PCC 7942 using another derivative of transposon Tn*5* (Tn*5*-692) (Koksharova and Wolk, 2002; Miyagishima et al., 2005). Tn*5* derivatives have been used to isolate swimming motility mutants in a marine *Synechococcus* strain (McCarren and Brahamsha, 2005). All above are in vivo examples. In vitro Tn*5*-mediated mutagenesis has also been used in *Synechocystis sp*. PCC 6803 for mutants that are defective in optimal photoautotrophic growth (Zhang et al., 2004) or in motility (Bhaya et al., 2001).

Traditional in vivo mutagenesis has contributed to identification of most known clock-related loci, which are non-redundant and non-essential, in *S. elongatus*, such as chemical mutagenesis (EMS, ethyl methanesulfonate) to discover the *kaiABC* locus (Ishiura et al., 1998; Kondo et al., 1994) and Tn*5* transposon-mediated insertional mutagenesis in identification of *cikA* (Andersson et al., 2000; Schmitz et al., 2000). Error-prone PCR, an in vitro mutagenesis strategy, was used to produce hundreds of point mutations in *kaiA* (Nishimura et al., 2002). As a strain featuring natural transformation and efficient double-recombination, *S. elongatus* is amenable to large-scale in vitro mutagenesis.

D. Transposon-Mediated Functional Genomics of S. elongatus

To identify all of the clock components in *S. elongatus*, we are carrying out a functional genomics project, aiming to inactivate each locus of the genome and screen those insertional mutants for changes in circadian rhythms. The immediate goal is to identify all open reading frames (ORFs) that contribute to circadian timing. An additional and broadly useful result will be a complete archived set of mutagenesis templates that allows inactivation of any given locus in the genome of *S. elongatus*.

The overall functional genomics project was conducted in essentially two parallel tracks: determination of the genomic sequence and production of a library of mutations for the majority of genes of the genome. For global mutagenesis of the *S. elongatus* genome, which was begun before the genomic sequence was available for this organism, we initiated a transposon-mediated mutagenesis and sequencing strategy to determine the sequences surrounding transposon insertions in essentially every *S. elongatus* gene (Holtman et al., 2005). Two transposon systems were used: Tn*5* from Epicentre and *Mu* from Invitrogen. Bacteriophage *Mu* (Lavoie and Chaconas, 1996; Mizuuchi, 1992) has the least target-specificity among known transposons (Haapa et al., 1999a, b) with a broad consensus target site NY(G/C)RN or CY(G/C)RG (Haapa-Paananen et al., 2002; Mizuuchi and Mizuuchi, 1993), which is relatively frequent in GC-rich *S. elongatus*. The base composition in ORFs of *S. elongatus* averages about 60% in the third codon position, whereas intergenic regions tend to be more AT-rich. Thus, there are more chances for *Mu* to hop into coding regions. Another option, the hyperactive in vitro Tn*5* transposition system (Goryshin and Reznikoff, 1998) modified with reduced site specificity, is also sufficient for inactivating essentially any gene (Kirby, 2007).

As previously illustrated (Holtman et al., 2005), in vitro transposition reactions, using *Mu* as an example, were performed on cosmids that carry ~30–40 kb genomic DNA. A collection of insertionally inactivated cosmid alleles was then introduced into cyanobacterial reporter strains via transformation. Transposons were integrated

into corresponding chromosome loci through double recombination. Insertional mutants were screened for circadian phenotypes by checking altered bioluminescence traces from the reporter genes. The primer binding sites at the ends of *Mu* transposon were used to amplify flanking genomic sequences for localization the insertion site. Because the transposons we use insert almost randomly into the genome, the sequence surrounding the insertion sites, which can be determined by sequencing outward from transposon end primers, should provide widespread coverage of the whole genome.

II. Transposon-Mediated Mutagenesis and Sequencing of the *Synechococcus elongatus* Genome

Transposon-mediated in vitro mutagenesis and sequencing started with a 960-cosmid genomic library, which had been constructed previously by inserting *Sau*3AI partially digested PCC 7942 genomic DNA into the *Bam*HI site of the SuperCos I cosmid vector (Holtman et al., 2005). For each cosmid, carrying ~30–40 kb of genomic DNA, an in vitro transposition assay was performed with a commercial derivative of bacteriophage/transposon *Mu* (GeneJumper™ Primer Insertion Kit for Sequencing). The kit uses a minimal *Mu*Cm, which contains only the inverted repeats of *Mu* right end sequences at both of its ends and a chloramphenicol (Cm) resistance marker, *cat*, as well as purified MuA transposase proteins. In the reactions, MuA assembles onto the MuA binding sites R1 and R2 at the ends of *Mu*Cm to form a functional transposition complex, which then inserts the *Mu*Cm into the target cosmid randomly (Haapa et al., 1999a, b). *Mu* is one of the transposons with the least target specificity. With a proper molar ratio of the transposon to the target DNA, only one copy of the *Mu*Cm will be inserted into a random position of the cosmid in most cases. After transforming an *E. coli* host strain with the transposition reaction and selecting with both chloramphenicol (Cm) and kanamycin (Km) (resistance encoded by SuperCos I), a population of colonies that each carries usually one *Mu*Cm inserted into a different position on the cosmid will be obtained. Restriction digestion of the cosmid DNA extracted from the colonies was performed to identify *Mu* insertions located only on the genomic sequence, not the cosmid vector. The resulting clone set, in 96-well format, was used for both generating *S. elongatus* insertional mutants and sequencing.

Mu-mediated sequencing was carried out as reported (Holtman et al., 2005). When the whole sequence of a cosmid was completed, the precise position of each *Mu*Cm was annotated. Normally, 96 × 4 sequences were produced for each cosmid. These sequences were then analyzed using the ContigExpress program of Vector NTI bioinformatics software suite (Invitrogen). Each sequence was polished by trimming the 5′-*Mu*Cm end sequence and 3′-unreadable sequence. The trimmed sequences were assembled using modified standard parameters of the program. The assembled contigs were then oriented and connected using a "primer walking" method. In total, nine cosmids were completely sequenced in this manner and manually annotated, and eight of them, including a cosmid that carries the large endogenous plasmid pANL, were deposited into GenBank (Table 5.1).

A "cosmid walking" strategy was adopted to choose new cosmids whose end sequences overlap with sequenced cosmids to proceed around the genome. Approximately 70 tiled cosmids would be required to cover the entire genome. Due to temporary technical problems with the

Table 5.1. Current progress[a] of the functional genomics project of *S. elongatus*

	Transposon		Insertion sites sequenced	Total kb sequenced
	Mu	Tn*5*		
Published genomic cosmids	8		~2,300	~268
Completed genomic cosmids	28	9	~9,500	~1,201
Completed JGI plasmid sets		44	~8,400	~903
Total	89		~20,200	~2,372
	Cosmids or plasmid sets		Mutants	ORFs
Screened in *S. elongatus*	70		~2,550	~1,850

[a]Modified and updated from Holtman et al. (2005)

GeneJumper™ Primer Insertion Kit, we used the EZ::TN <Kan-2> transposon (Epicentre) for part of the project; this approach required us to subclone fragments from cosmid inserts into plasmids because of incompatibility of the Tn*5* (Km^R) with the cosmid. In total, 45 cosmids carrying chromosomal sequences were completely mutagenized and sequenced, and nine of them were subcloned and mutagenized with Tn*5* (Table 5.1).

Our laboratory collaborated with the DOE Joint Genome Institute (JGI) for determining the genome sequence of *S. elongatus,* providing the genomic DNA that was analyzed. As collaborators we received the genomic libraries that JGI used for shotgun sequencing when that project was completed. JGI developed 3-kb and 8-kb plasmid libraries, as well as a fosmid library for large (~40 kb) inserts. All shotgun sequences used for assembling the whole genome also were sent to us as output files from the Phred/Phrap program (Ewing and Green, 1998; Ewing et al., 1998). The coordinates of each sequence in the assembly can be accessed in Consed, a Unix-based graphical editor and automated finishing program for Phrap sequence assemblies (Gordon et al., 1998; 2001). Because plasmids are much more convenient for sequencing and transformation than cosmids, we switched to using JGI 8-kb plasmids as templates for mutagenesis for the remainder of the functional genomics project. For economy and convenience, we used sets of five 8-kb plasmids (termed JGI 8-kb plasmid sets, or plasmid sets, hereafter) combined in one in vitro mutagenesis reaction with Tn*5*. Each plasmid set carries ~40 kb genomic sequence, comparable to the size of a cosmid insert. A total of 44 JGI 8-kb plasmid sets (designated 8S1–8S44) that cover most gaps between sequenced cosmids were identified using Consed. All of these plasmid sets have been fully mutagenized and sequenced (Table 5.1). Sequencing of the sites of insertion for more than half of the clones was outsourced to High-Throughput Sequencing Solutions, a non-profit facility of the Department of Genome Sciences, University of Washington.

For the first few thousand insertions, annotation of each insertion site was performed with the assembling tool, ContigExpress, of the Vector NTI software package (Invitrogen). First, transposon-mediated sequences were assembled with the complete genome sequence. The exact site of each *Mu*/Tn*5* insertion was then localized to its footprint (duplicated host sequence due to insertion of transposons, 5-bp for *Mu* and 9-bp for Tn*5* (Mizuuchi and Craigie, 1986; Reznikoff, 1993), which is determined as the genomic sequence immediately upstream or downstream of transposon end sequences. Originally at least 384 (4 × 96-well plates) individual insertion-mediated sequences were used to assemble contigs or the main scaffold of a cosmid. This usually caused a cosmid to be over-saturated by transposon insertions as shown in cosmid 7G3 (Fig. 5.3). With the complete genome sequence available, only 192 (2 × 96-well plates) insertions are required for saturation of a ~40 kb cosmid or plasmid set.

In summary, we have performed saturation mutagenesis on over 88% of the chromosome (approximately 2,372 kb) from 89 cosmids or plasmid sets. Some 20,200 positions of insertions have been sequenced and annotated (see Table 5.1).

III. Creating a Unigene Set

The in vitro mutagenesis effort resulted in multiple individual insertion alleles for many *S. elongatus* ORFs. To facilitate the *S. elongatus* mutagenesis and screening process, we chose a single insertion

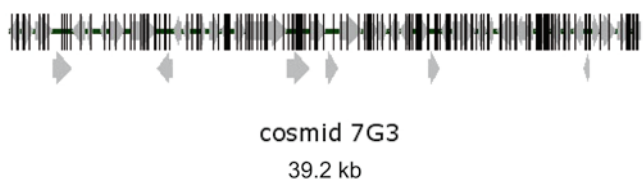


Fig. 5.3. Saturation of cosmid 7G3 with *Mu* transposon insertions. The 39.2 kb *S. elongatus* DNA segment in cosmid 7G3 is depicted. Gray arrows represent ORFs and those below the line overlap other ORFs. Black vertical lines indicate the positions of *Mu* insertions.

that maps near the N-terminal coding region of each ORF and created a unigene set, a subset of the library with one insertion per ORF, arranged in order in 96-well plates. This streamlined set contains 2451 mutagenesis substrates for inactivation of the respective gene in the chromosome of an *S. elongatus* WT or reporter strain by homologous recombination.

Annotation of the insertions and choice of insertions for the unigene set were carried out using CyanoBIKE – a Web-based, programmable, integrated biological knowledge base (http://biobike.csbc.vcu.edu; Elhai et al., 2009). The procedure used the insertion flanking sequences and a number of scripts written with CyanoBIKE to localize each insertion precisely in the chromosome of *S. elongatus* and to determine the context of each insertion relative to the mutagenized locus or neighbor loci for intergenic insertions.

From a FASTA file including 14,587 genomic sequences flanking the transposons, the first 50 nucleotides of each sequence (of *circa* 200 nucleotides each) were extracted and searched, allowing 5 mismatches, over the chromosome sequence. The chromosome coordinates of the 5 or 9 first nucleotides of the sequences, depending on the mutation type (*Mu* or Tn*5*), were returned. Using these coordinates for the insertion footprint, the context of each insertion was determined, rendering the following information: (1) the mutagenized locus or the surrounding loci when the insertion is found to be in an intergenic region, (2) the strand of the mutagenized locus or the strands of the surrounding loci, (3) the number of nucleotides between the insertion footprint and the start and stop codons of a mutagenized ORF, or between the insertion footprint and both surrounding ORFs. In cases for which the script could not ensure an automatic annotation of the insertions (e.g., more than one hit or no hit), the insertions were not annotated and were flagged to be reviewed manually. This outcome represented *circa* 2.5% of the total number of all insertions produced during the functional genomics project.

Although experimental results for 700 ORFs did not show differences in phenotypes between mutants that carry insertions in the N-terminal vs. C-terminal coding regions, the information listed above permitted us to select the insertion mutation that would be the most likely to knock out the locus. Another CyanoBIKE script using the annotation previously obtained was written for this purpose and one insertion per locus was selected according to the following parameters (Fig. 5.4). When multiple insertions are available for an ORF, the script chose the insertion located nearest the start codon of each ORF, up to a minimum distance of 50 nucleotides; this minimum was applied to avoid straying outside the ORF if the start codon of the locus had been called incorrectly in the automated annotation. On the other hand, for the loci without an insertion, the first insertion within 100 nucleotides upstream from the start codon of the locus was chosen. This group represented 76 mutagenized substrates that were included in the unigene set.

IV. Making the Unigene Set of Mutagenesis Substrates Searchable Online by the Scientific Community

A web application including a form to search for insertions from a locus, or vice versa, and a scrollable/zoomable chromosome map on which the insertions were superimposed (Fig. 5.4) was designed for the purpose of sharing the mutagenized substrates with the community of researchers working with *S. elongatus* and is available at: http://golden.ucsd.edu/S7942/.

The web application uses the relational database management system MySQL (http://www.mysql.com/) to store and manage the insertion data: e.g., chromosome coordinates of the insertion footprint, context of the insertion, mutagenized ORF or surrounding ORFs, the number of nucleotides between the insertion footprint and the start and stop codons of a mutagenized ORF, or between the insertion footprint and both surrounding ORFs, the mutation type, the insertion footprint sequence, and the 96 well plate number; the web interface was written in PHP (http://www.php.net). The creation of the chromosome map was made with CGView (http://wishart.biology.ualberta.ca/cgview/index.html; Stothard and Wishart, 2005), a Java package for generating maps of circular genomes. All of the data were generated in the proper format for either MySQL or the CGView applet by CyanoBIKE scripts using public data available within CyanoBIKE and data obtained by the annotation and selection of the insertions.

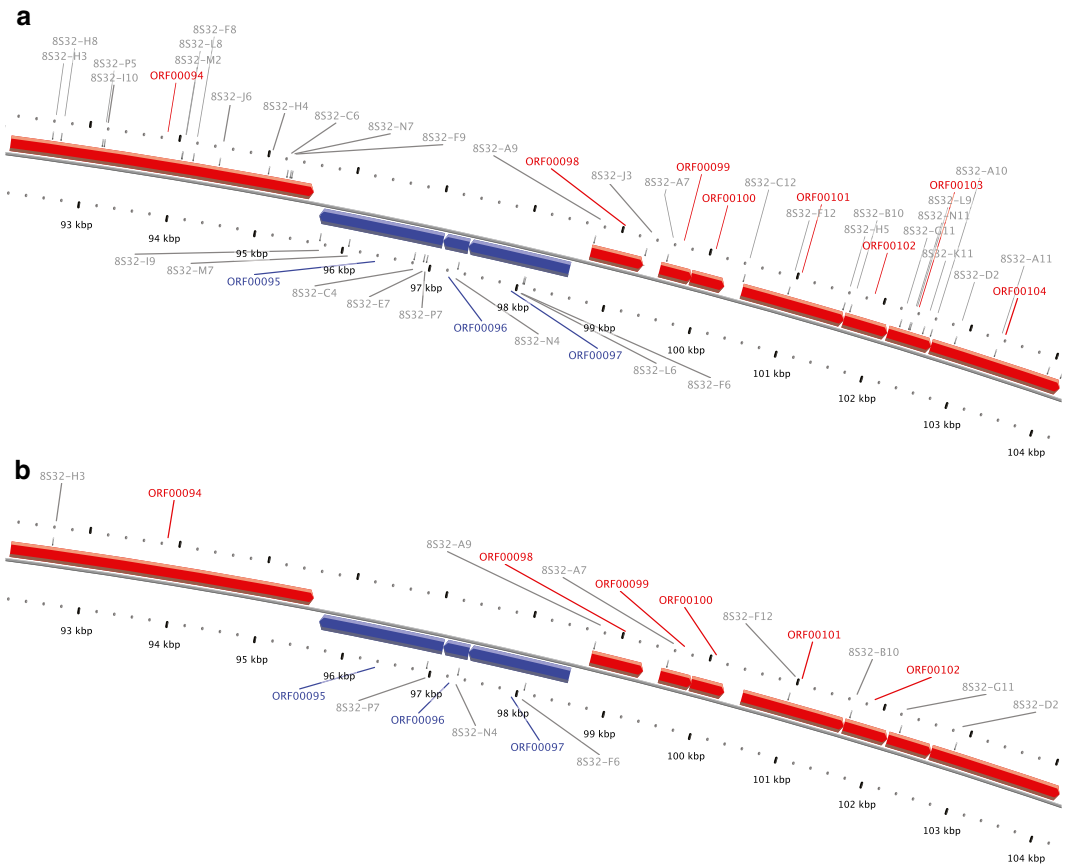


Fig. 5.4. Screenshots of a portion of the chromosome map of *S. elongatus* on which the insertions were superimposed (**a**) before and (**b**) after selection of the insertions for the unigene set.

V. Screening for Circadian Clock Phenotypes of Transposon Insertion Mutants

Transposon-inserted cosmid/plasmid DNA can be used directly to transform *S. elongatus* bioluminescent reporter strains. Because these cosmids or plasmids cannot replicate autonomously in cyanobacterial cells, the only surviving cells under the selection pressure of antibiotics to which the transposons encode resistance are those that harbor transposons integrated via homologous recombination into the corresponding region of the chromosome. In most cases, transformants are apparent double-recombinants, such that the vector backbone is lost (Clerico et al., 2007). If the disrupted region is essential, a wild-type copy and a mutant allele are both maintained. This may happen through a single recombination event, resulting in both alleles in the same chromosome, or the maintenance of both wild-type and mutant (double recombination event) chromosomes (Clerico et al., 2007). Otherwise, no transformants will be obtained.

Transformation of *S. elongatus* PCC 7942 strains and screening of circadian clock phenotypes were performed as described previously (Andersson et al., 2000; Holtman et al., 2005; Mackey et al., 2007). All cyanobacterial wild-type reporter and mutant strains were created in *S. elongatus* PCC 7942. Cyanobacterial strains were grown in BG-11 medium (Rippka et al., 1979) under continuous light conditions (~70 μmol photons/m^2s) at 30°C with appropriate antibiotics (Holtman et al., 2005). High-throughput transformation of *S. elongatus* strains was conducted in 24-well microplates using liquid BG-11 medium, instead of agar plates (Holtman et al., 2005). This method greatly

increased the efficiency of mutant screening, but has a disadvantage in that no single colonies are isolated. Thus, a mixture of several or many independent clones is used for further phenotypic analysis. Single recombinants and double recombinants can coexist in the same population. Even wild-type alleles may be present if segregation has not been completed. If disruption of a gene by insertion has a growth phenotype, single recombinants probably will outgrow double recombinants during the course of selection. Thus, it might be more difficult to detect the attenuated phenotype. Despite this drawback, we have been able to identify mutant phenotypes with this high-throughput approach. Clones that show a mutant phenotype by this method are re-tested by plating for individual colonies.

We are focusing mainly on circadian clock phenotypes, using the automated bioluminescence reporter system, in which *S. elongatus* promoters are fused to promoterless *luxAB* (*Vibrio harvey* luciferase genes) or *luc* (firefly luciferase gene) and then integrated into one of the Neutral Sites, which mediate homologous recombination with the *S. elongatus* chromosome and cause no apparent phenotypes when they carry insertions (Andersson et al., 2000), on the chromosome through homologous recombination. Bioluminescence of cyanobacterial cells produced by the luciferase reporter gene in 96-well black microplates is counted automatically and continuously using a Packard TopCount luminometer (Mackey et al., 2007). The reporter strains currently used for the functional genomics project are AMC1020 and AMC1300 for *Mu* mutants, as well as AMC462 for Tn*5* mutants. AMC1020 carries the promoter for *psbAI* fused to *luxAB* (P*psbAI*::*luxAB*) in NS1 and P*psbAI*::*luxCD_E* in NS2. The operon *luxCD_E* is from *Xenorhabdus luminescens* and encodes production of the aldehyde substrate for the luciferase encoded by *luxAB* (Andersson et al., 2000). The *psbAI* gene is one of three *psbA* paralogs that encode a critical photosystem II reaction center protein, D1 (Golden et al., 1986; Nair et al., 2001). The *psbAI* promoter was chosen for *luxCD_E* expression because it is strong relative to most other *S. elongatus* promoters, and seemed likely to keep substrate levels high. Northern blot results have shown that the expression of *psbAI* is rhythmic at the mRNA level (Liu et al., 1995a), and peaks at dusk as do most *S. elongatus* genes (designated as class 1) identified by a whole-genome screening of *luxAB* fusions using a cooled-CCD camera monitoring system (Liu et al., 1995b). AMC1300, almost identical to AMC1020 except for the promoter that drives *luxAB*, is a P*kaiB* (also class 1) reporter strain (Holtman et al., 2005). AMC462 is another P*kaiB* reporter strain, with P*kaiB::luxAB* in NS1 and P*psbAI::luxCDE* in NS2 (Katayama et al., 1999). KaiB is one of the three central oscillator components of cyanobacterial circadian clock.

We are currently screening for mutants with altered circadian periods or phasing, or arrhythmic expression under constant light conditions. Additional screening will be done to identify mutants with defects in entrainment or phase resetting. The latter mutants would be those that can be entrained by 12-h/12-h light/dark cycles, but cannot reset the relative timing of peak bioluminescence in response to a 5-h dark pulse applied at certain circadian time points (Kondo et al., 1994). A P*purF* (encodes glutamine PRPP amidotransferase and peaks at dawn)-based class two reporter gene may also be used (Liu et al., 1996; Min et al., 2004) to detect phenotypes of mutants that affect different aspects of the circadian clock.

Initially, mutants representing two insertions per ORF were picked for transformation and screening, choosing one near the N- and one near the C-terminal coding region. However, no phenotypic differences were identified based on transposon position, and the process was streamlined to 1 N-terminal insertion per ORF after the first 256 loci were screened. The initial pass does not include insertions located in the intergenic regions other than those upstream of and proximal to genes for which no insertion was recovered in the ORF; those mutants may be analyzed in the future to determine the influence of unidentified small RNAs or small un-annotated protein-coding ORFs. Around 95 cosmids or plasmid sets have been screened for the first-round screening, accounting for 1850 ORFs.

VI. Novel Circadian Clock-Related Loci in the Genome

Preliminary results show that among the first ~700 ORFs screened, several known clock genes were confirmed, including the *kaiABC* locus. In addition, mutations in around 70 additional loci,

not previously connected to the circadian clock, showed altered circadian phenotypes, primarily period changes (Holtman et al., 2005). Some of these novel loci are clustered, for which polar effects are likely. The current mutants more likely represent loss (or decrease in function) of approximately 40 unrelated loci, which belong to several different functional categories (Table 5.2). The mutant phenotype has been re-tested, after recombination into a different reporter strain, for 22 of the original 71 mutants, and all but one were confirmed in the second test. A representative circadian clock mutant, 2E8-LLL11, is shown in Fig. 5.5. The putative ORF synpcc7942_2475 that was disrupted in this mutant encodes a small hypothetical protein (53 amino acids) with no known function. The absence of a circadian phenotype following recombination of upstream and downstream interruption alleles suggests that these genes carry their own promoters and do not comprise an operon.

Many of the mutants that affect circadian rhythmicity were recovered in merodiploid cells (functionally 'heterozygous') in which the mutant alleles do not fully segregate. In *S. elongatus*, complete segregation of a selectable marker that does not affect viability is very rapid; recovery of merodiploids is evidence of interruption of an essential cellular function (Golden et al., 1986; 1987). Typically, apparent double-crossover events occur 10^3 more frequently than apparent single crossovers (Tsinoremas et al., 1994), which preserve the WT allele and incorporate into the chromosome the mutant allele along with the vector backbone. In 11 of the first 71 mutants we detected the presence of the vector in the gene-knock-out transformants, which indicates selection for a rare single crossover; we also sometimes detect a mixed population of WT and mutant chromosomes in the cell. Either of these outcomes suggests that the interrupted locus is essential for viability.

Merodiploids that exhibit circadian defects include those with mutations in genes that encode the ClpP2ClpX protease (Holtman et al., 2005). A lengthened circadian period was observed in *Mu*-inserted mutants of the *clpP2clpX* operon, the first two genes in the first cosmid (7H1) that were screened. The *clp* operon encodes an ATP-dependent Clp protease complex that is ubiquitous in bacteria, plants, and animals. ClpP is a serine-type protease (Maurizi et al., 1990), while ClpX functions as an ATPase subunit, now known as a member of the Clp/Hsp100 family of chaperones (Schelin et al., 2002). This protease complex may be involved in degradation of clock proteins, leading to the period phenotype in the mutants (Holtman et al., 2005). These two *clp* genes are the first clock-related genes in the cyanobacterium found to be essential for cell viability.

In addition, we identified a novel allele of *kaiA* that produces a short-period phenotype in the insertion mutant (Chen et al., 2009). The *kaiA* gene encodes KaiA, a protein of the circadian oscillator, and mutations in it usually cause arrhythmia or a long-period phenotype (Nishimura et al., 2002). The insertion mutation truncates the protein by three amino acid residues by creation of a stop codon (Fig. 5.6a). By a combination of genetic, biochemical, and biophysical methods we were able to show that: the truncation of KaiA, rather than an increase in KaiB and KaiC that results from a cis effect of the insertion mutation (Kutsuna et al., 2005), is responsible for the short-period phenotype; and, the truncation of KaiA strengthens its interaction with KaiC (Chen et al., 2009).

Mu insertions in two other ORFs of cosmid 4G8 show slightly long period circadian phenotypes (~0.5–1.0 h longer than wild type) in reporter strains AMC1020 and AMC1300 (Holtman et al., 2005). 4G8-L6 and 4G8-EEE10 disrupt the N-terminus of a putative inosine monophosphate dehydrogenase (IMPDH, Synpcc7942_1831) gene (Fig. 5.6b). Two other *Mu* insertions are related to the downstream thioredoxin gene, *trxM* (synpcc7942_1830): 4G8-D6, which sits ~40 bp upstream of the start codon, and 4G8-EEE5, which is in the C-terminal coding region of the ORF. The *Mu* insertions in the IMPDH mutant strains could not be fully segregated, while segregation in *trxM* mutants was complete. This result suggests that the IMPDH gene is essential for viability, whereas the *trxM* gene is not. Because the IMPDH gene is immediately upstream of *trxM* and in the same orientation, we could not rule out a possible polar effect such that insertions in IMPDH gene affect function of the downstream thioredoxin gene. The *trxM* gene was reported previously to be

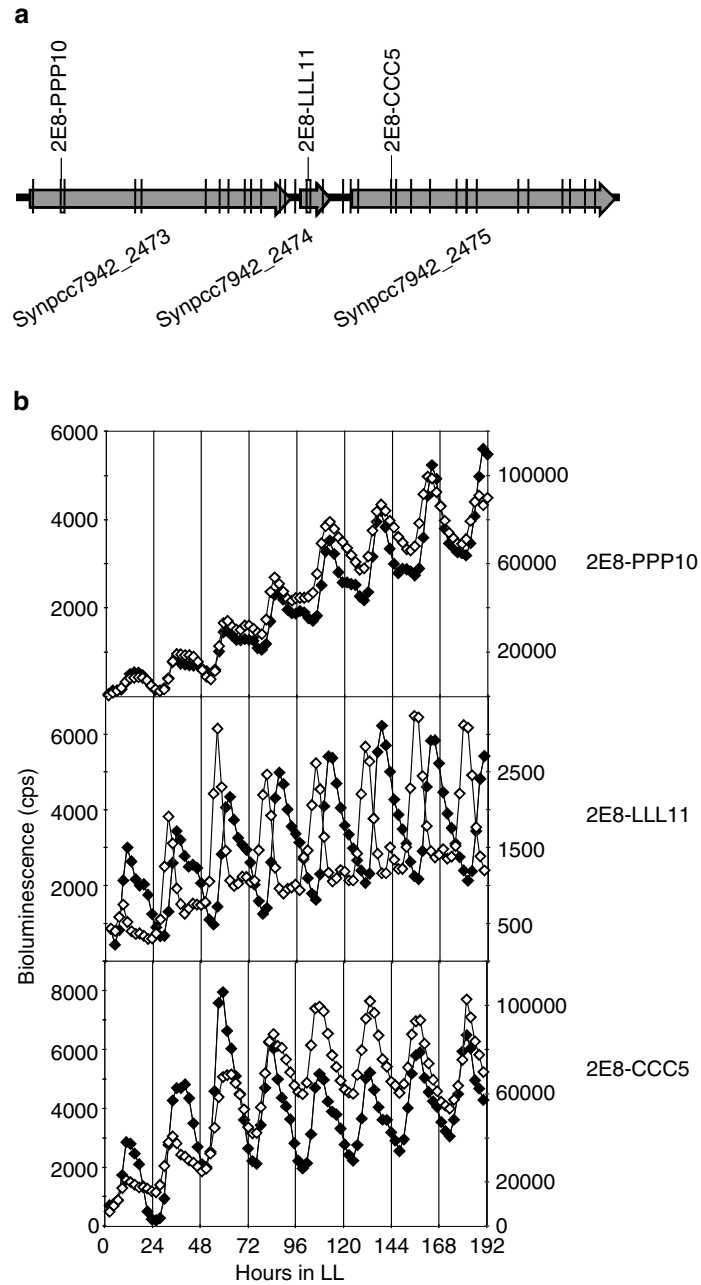


Fig. 5.5. Insertions in synpcc7942_2474 cause a circadian phenotype but insertions into the upstream and downstream ORFs do not. The map in (**a**) shows 3 ORFs, synpcc7942_2473, synpcc7942_2474 and synpcc7942_2475, with the positions of *Mu* insertions shown as *black vertical lines*. Shown in (**b**) are circadian traces of mutants, in *open diamonds*, that carry insertions 2E8-PPP10 (synpcc7942_2473, *top panel*), 2E8-LLL11 (synpcc7942_2474, *middle panel*) or 2E8-CCC5 (synpcc7942_2475, *bottom panel*). Each is compared with a WT control strain (*close diamonds*), positioned similarly on the assay plate to control for normal minor variations in circadian period as a function of ambient light intensity. Mutation 2E8-LLL11 causes a slightly shorter circadian period and a change in the phasing of circadian peak time of the reporter. The other two mutations (like more than 90% of mutations tested) have no effect. Absolute bioluminescence levels depend on cell number and were not matched for the traces shown here; thus, their differences do not reflect phenotypes.

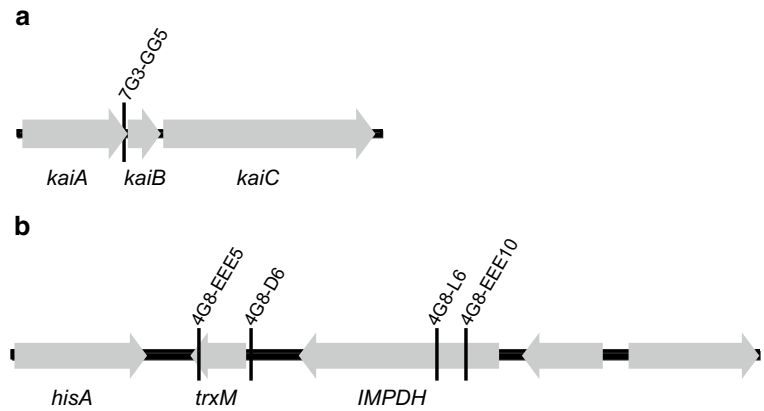


Fig. 5.6. Graphical representation of ORFs for insertional analysis. (**a**) *kaiABC* locus in cosmid 7G3. (**b**) Thioredoxin (*trxM*) locus in cosmid 4G8. *Gray arrows* are ORFs and *black vertical bars* are *Mu* insertion sites.

Table 5.2. Categories of novel clock ORFs

Functional Category	# of ORFs
Transporters	5
Metabolic (synthases, isomerases, etc.)	11
Light harvesting antenna	10
Membrane-related	3
Redox-related	4
Cell division-related	2
Hypothetical (conserved or unique)	21
Intergenic regions	3
Proteases	2
Regulatory proteins	3
Other	7
Total	71

essential in our organism, but not in *E. coli* (Muller and Buchanan, 1989). There are at least three thioredoxin genes (synpcc7942_1793, synpcc7942_1830, and synpcc7942_1978) in *S. elongatus* genome and they are similar in their size and protein sequence. It is possible that their functions are overlapping or redundant.

Other novel clock-related loci are still under second-round testing to confirm their phenotypes. Many are likely to be also involved in essential cellular functions, such as cell division and metabolism (Table 5.2). These genes, together with the *clpP2clpX* genes mentioned above are, for the first time, linking circadian clock functions to essential housekeeping pathways in cyanobacteria, which indicates a more significant role of the circadian clock in cellular functions than previously demonstrated.

VII. Conclusions

Above all, this functional genomics project aims to assay the function of each locus of the *S. elongatus* genome through insertional gene inactivation, with a focus on the phenotypes associated with the circadian rhythmicity of gene expression. More than 88% of genes in the *S. elongatus* genome were mutagenized by *Mu* or Tn*5* transposons. Among ~700 putative ORFs screened for circadian clock phenotypes, over 70 novel clock loci were discovered by the project team. Preliminary functional analysis has been performed for the Clp protease, KaiA, and thioredoxin loci. This project has resulted in an unprecedented cyanobacterial genomic resource that holds great promise for achieving a systems-level understanding of not only circadian clocks, but also photosynthetic energy production, oxygen generation, carbon fixation, nutrient utilization, stress response, and biofuels production.

Acknowledgments

We thank the following undergraduate students who contributed to the *S. elongatus* functional genomics project: Kimberly Baker, Carla Blumentritt, Audra Boettcher-Herring, Zack Bussey, Krista Cole, Priyanka Desai, Robert Dover, Ramiro Fernandez, Randall Gil, Judith Kwarteng-Amaning, Shaun McMurtry, Minh Nguyen, Ogonna (Kinney) Nwawka, Victoria

Omishakin, Cassie Reyna, Nancy Reyes, Johnathan Siefert, Lindsay Stautzenberger, Ola Tokunbo. Alejandra Gonzales, Brandi Mohler, Isabel Salinas, and Pamela Sandoval contributed as technicians to the early stages of this work and Drs. Lisa Campbell and Sunil Pabbi participated as sabbatical visitors. Dr. Phil Beremand of the Laboratory for Functional Genomics and Larry Harris-Haller of the Gene Technologies Laboratory provided technical advice and support. JGI provided the complete *S. elongatus* PCC 7942 sequence. This project was initiated with grants to Philip A Youderian and SSG by the National Science Foundation (MCB-0196144 to PY and MCB-9818031 to SSG) and continued with support by the US Department of Energy (DE-FG03-00ER15055 to SSG) and the National Institutes of Health (R01 GM59336 and P01 NS39546 to SSG). AT and development of CyanoBIKE were supported by NSF grant DBI- DBI-0516378 to Jeff Elhai.

References

Aldehni MF, Sauer J, Spielhaupter C, Schmid R and Forchhammer K (2003) Signal transduction protein P(II) is required for NtcA-regulated gene expression during nitrogen deprivation in the cyanobacterium *Synechococcus elongatus* strain PCC 7942. J Bacteriol 185: 2582–2591

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, *et al* (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science 301: 653–657

Andersson CR, Tsinoremas NF, Shelton J, Lebedeva NV, Yarrow J, Min H and Golden SS (2000) Application of bioluminescence to the study of circadian rhythms in cyanobacteria. Methods Enzymol 305: 527–542

Aoki S, Kondo T and Ishiura M (1995) Circadian expression of the *dnaK* gene in the cyanobacterium *Synechocystis* sp. strain PCC 6803. J Bacteriol 177: 5606–5611

Balu B and Adams JH (2006) Functional genomics of *Plasmodium falciparum* through transposon-mediated mutagenesis. Cell Microbiol 8: 1529–1536

Beaver LM, Gvakharia BO, Vollintine TS, Hege DM, Stanewsky R and Giebultowicz JM (2002) Loss of circadian clock function decreases reproductive fitness in males of *Drosophila melanogaster*. Proc Natl Acad Sci USA 99: 2134–2139

Bell-Pedersen D, Cassone VM, Earnest DJ, Golden SS, Hardin PE, Thomas TL and Zoran MJ (2005) Circadian rhythms from multiple oscillators: lessons from diverse organisms. Nat Rev Genet 6: 544–556

Berman-Frank I, Lundgren P, Chen YB, Kupper H, Kolber Z, Bergman B and Falkowski P (2001) Segregation of nitrogen fixation and oxygenic photosynthesis in the marine cyanobacterium *Trichodesmium*. Science 294: 1534–1537

Bhaya D, Takahashi A, Shahi P and Grossman AR (2001) Novel motility mutants of *Synechocystis* strain PCC 6803 generated by *in vitro* transposon mutagenesis. J Bacteriol 183: 6140–6143

Biery MC, Stewart FJ, Stellwagen AE, Raleigh EA and Craig NL (2000) A simple *in vitro* Tn7-based transposition system with low target site selectivity for genome and gene analysis. Nucleic Acids Res 28: 1067–1077

Bourhy P, Louvel H, Saint Girons I and Picardeau M (2005) Random insertional mutagenesis of *Leptospira interrogans* the agent of leptospirosis using a mariner transposon. J Bacteriol 187: 3255–3258

Bovy A, de Kruif J, de Vrieze G, Borrias M and Weisbeek P (1993) Iron-dependent protection of the *Synechococcus* ferredoxin I transcript against nucleolytic degradation requires *cis*-regulatory sequences in the 5′ part of the messenger RNA. Plant Mol Biol 22: 1047–1065

Bradley RL and Reddy KJ (1997) Cloning sequencing and regulation of the global nitrogen regulator gene *ntcA* in the unicellular diazotrophic cyanobacterium *Cyanothece* sp. strain BH68K. J Bacteriol 179: 4407–4410

Capuano V, Thomas JC, Tandeau de Marsac N and Houmard J (1993) An *in vivo* approach to define the role of the LCM the key polypeptide of cyanobacterial phycobilisomes. J Biol Chem 268: 8277–8283

Carlson CM and Largaespada DA (2005) Insertional mutagenesis in mice: new perspectives and tools. Nat Rev Genet 6: 568–580

Chen Y, Holtman CK, Magnuson RD, Youderian PA and Golden SS (2008) The complete sequence and functional analysis of pANL the large plasmid of the unicellular freshwater cyanobacterium *Synechococcus elongatus* PCC 7942. Plasmid 59: 176–192

Chen Y, Kim YI, Mackey SR, Holtman CK, Liwang A and Golden SS (2009) A novel allele of *kaiA* shortens the circadian period and strengthens interaction of oscillator components in the cyanobacterium *Synechococcus elongatus* PCC 7942. J Bacteriol 191: 4392–4400

Chen YB, Dominic B, Mellon MT and Zehr JP (1998) Circadian rhythm of nitrogenase gene expression in the diazotrophic filamentous nonheterocystous cyanobacterium *Trichodesmium* sp. strain IMS 101. J Bacteriol 180: 3598–3605

Chen YB, Dominic B, Zani S, Mellon MT and Zehr JP (1999) Expression of photosynthesis genes in relation to nitrogen fixation in the diazotrophic filamentous nonheterocystous cyanobacterium *Trichodesmium* sp. IMS 101. Plant Mol Biol 41: 89–104

Clerico EM, Ditty JL and Golden SS (2007) Specialized techniques for site-directed mutagenesis in cyanobacteria. Methods Mol Biol 362: 155–171

Cohen MF, Wallis JG, Campbell EL and Meeks JC (1994) Transposon mutagenesis of *Nostoc* sp. strain ATCC 29133 a filamentous cyanobacterium with multiple cellular differentiation alternatives. Microbiology 140: 3233–3240

Collier JL and Grossman AR (1994) A small polypeptide triggers complete degradation of light-harvesting phycobiliproteins in nutrient-deprived cyanobacteria EMBO J 13: 1039–1047

Dodd AN, Salathia N, Hall A, Kevei E, Toth R, Nagy F, Hibberd JM, Millar AJ and Webb AA (2005) Plant circadian clocks increase photosynthesis growth survival and competitive advantage. Science 309: 630–633

Dong G and Golden SS (2008) How a cyanobacterium tells time. Curr Opin Microbiol 11: 541–546

Dunlap JC, Loros JJ and DeCoursey P, (eds) (2003) Chronobiology: Biological Timekeeping. Sinauer Associates Inc, Sunderland

Durham KA, Porta D, McKay RM and Bullerjahn GS (2003) Expression of the iron-responsive *irpA* gene from the cyanobacterium *Synechococcus* sp. strain PCC 7942. Arch Microbiol 179: 131–134

Elhai J, Taton A, Massar JP, Myers JK, Travers M, Casey J, Slupesky M and Shrager J (2009) BioBIKE: a Web-based programmable integrated biological knowledge base. Nucleic Acids Res 37: W28-32

Ewing B and Green P (1998) Base-calling of automated sequencer traces using phred II Error probabilities. Genome Res 8: 186–194

Ewing B, Hillier L, Wendl MC and Green P (1998) Base-calling of automated sequencer traces using phred I. Accuracy assessment. Genome Res 8: 175–185

Garcia-Dominguez M and Florencio FJ (1997) Nitrogen availability and electron transport control the expression of *glnB* gene (encoding PII protein) in the cyanobacterium *Synechocystis* sp. PCC 6803. Plant Mol Biol 35: 723–734

Gehring AM, Nodwell JR, Beverley SM and Losick R (2000) Genomewide insertional mutagenesis in *Streptomyces coelicolor* reveals additional genes involved in morphological differentiation. Proc Natl Acad Sci USA 97: 9642–9647

Golden SS (2003) Timekeeping in bacteria: the cyanobacterial circadian clock. Curr Opin Microbiol 6: 535–540

Golden SS, Brusslan J and Haselkorn R (1986) Expression of a family of *psbA* genes encoding a photosystem II polypeptide in the cyanobacterium *Anacystis nidulans* R2. EMBO J 5: 2789–2798

Golden SS, Brusslan J and Haselkorn R (1987) Genetic engineering of the cyanobacterial chromosome. Methods Enzymol 153: 215–231

Golden SS and Canales SR (2003) Cyanobacterial circadian clocks--timing is everything. Nat Rev Microbiol 1: 191–199

Golden SS, Nalty MS and Cho DS (1989) Genetic relationship of two highly studied *Synechococcus* strains designated *Anacystis nidulans*. J Bacteriol 171: 24–29

Gordon D, Abajian C and Green P (1998) Consed: a graphical tool for sequence finishing. Genome Res 8: 195–202

Gordon D, Desmarais C and Green P (2001) Automated finishing with autofinish. Genome Res 11: 614–625

Goryshin IY and Reznikoff WS (1998) Tn5 *in vitro* transposition. J Biol Chem 273: 7367–7374

Granger L, Martin E and Segalat L (2004) *Mos* as a tool for genome-wide insertional mutagenesis in *Caenorhabditis elegans*: results of a pilot study. Nucleic Acids Res 32: e117

Green LS, Laudenbach DE and Grossman AR (1989) A region of a cyanobacterial genome required for sulfate transport. Proc Natl Acad Sci USA 86: 1949–1953

Haapa S, Taira S, Heikkinen E and Savilahti H (1999a) An efficient and accurate integration of mini-Mu transposons *in vitro*: a general methodology for functional genetic analysis and molecular biology applications. Nucleic Acids Res 27: 2777–2784

Haapa S, Suomalainen S, Eerikainen S, Airaksinen M, Paulin L and Savilahti H. (1999b). An efficient DNA sequencing strategy based on the bacteriophage *Mu* in vitro DNA transposition reaction. Genome Res 9: 308–315

Haapa-Paananen S, Rita H and Savilahti H (2002) DNA transposition of bacteriophage Mu. A quantitative analysis of target site selection *in vitro*. J Biol Chem 277: 2843–2851

Hayes F (2003) Transposon-based strategies for microbial functional genomics and proteomics. Annu Rev Genet 37: 3–29

Hirani TA, Suzuki I, Murata N, Hayashi H and Eaton-Rye JJ (2001) Characterization of a two-component signal transduction system involved in the induction of alkaline phosphatase under phosphate-limiting conditions in *Synechocystis* sp. PCC 6803. Plant Mol Biol 45: 133–144

Holtman CK, Chen Y, Sandoval P, Gonzales A, Nalty MS, Thomas TL, Youderian P and Golden SS (2005) High-throughput functional analysis of the *Synechococcus elongatus* PCC 7942 genome. DNA Res 12: 103–115

Huang T-C and Chow T-J (1986) New type of N2-fixing unicellular cyanobacterium (blue-green alga). FEMS Microbiol Lett 36: 109–110

Ikeuchi M (1996) [Complete genome sequence of a cyanobacterium *Synechocystis* sp. PCC 6803 the oxygenic photosynthetic prokaryote] Tanpakushitsu Kakusan Koso 41: 2579–2583

Ishiura M, Kutsuna S, Aoki S, Iwasaki H, Andersson CR, Tanabe A, Golden SS, Johnson CH and Kondo T (1998) Expression of a gene cluster *kaiABC* as a circadian feedback process in cyanobacteria. Science 281: 1519–1523

Iwasaki H, Williams SB, Kitayama Y, Ishiura M, Golden SS and Kondo T (2000) A *kaiC*-interacting sensory histidine kinase SasA necessary to sustain robust circadian oscillation in cyanobacteria. Cell 101: 223–233

Johnson CH, Golden SS and Kondo T (1998) Adaptive significance of circadian programs in cyanobacteria. Trends Microbiol 6: 407–410

Johnson CH, Hastings J W (1986) The elusive mechanism of the circadian clock. Am Sci 74: 29–36

Katayama M, Kondo T, Xiong J and Golden SS (2003) *ldpA* encodes an iron-sulfur protein involved in light-dependent modulation of the circadian period in the cyanobacterium *Synechococcus elongatus* PCC 7942. J Bacteriol 185: 1415–1422

Katayama M, Tsinoremas NF, Kondo T and Golden SS (1999) *cpmA* a gene involved in an output pathway of the cyanobacterial circadian system. J Bacteriol 181: 3516–3524

Kim YI, Dong G, Carruthers CW, Jr, Golden SS and LiWang A (2008) The day/night switch in KaiC a central oscillator component of the circadian clock of cyanobacteria. Proc Natl Acad Sci USA 105: 12825–12830

Kirby JR (2007) *In vivo* mutagenesis using EZ-Tn5. Methods Enzymol 421: 17–21

Kitada K, Ishishita S, Tosaka K, Takahashi R, Ueda M, Keng VW, Horie K and Takeda J (2007) Transposon-tagged mutagenesis in the rat. Nat Methods 4: 131–133

Kitayama Y, Nishiwaki T, Terauchi K and Kondo T (2008) Dual KaiC-based oscillations constitute the circadian system of cyanobacteria. Genes Dev 22: 1513–1521

Kiyohara YB, Katayama M and Kondo T (2005) A novel mutation in *kaiC* affects resetting of the cyanobacterial circadian clock. J Bacteriol 187: 2559–2564

Koksharova OA and Wolk CP (2002) Genetic tools for cyanobacteria. Appl Microbiol Biotechnol 58: 123–137

Kondo T and Ishiura M (1994) Circadian rhythms of cyanobacteria: monitoring the biological clocks of individual colonies by bioluminescence. J Bacteriol 176: 1881–1885

Kondo T, Strayer CA, Kulkarni RD, Taylor W, Ishiura M, Golden SS and Johnson CH (1993) Circadian rhythms in prokaryotes: luciferase as a reporter of circadian gene expression in cyanobacteria. Proc Natl Acad Sci USA 90: 5672–5676

Kondo T, Tsinoremas NF, Golden SS, Johnson CH, Kutsuna S and Ishiura M (1994) Circadian clock mutants of cyanobacteria. Science 266: 1233–1236

Kutsuna S, Kondo T, Aoki S and Ishiura M (1998) A period-extender gene *pex* that extends the period of the circadian clock in the cyanobacterium *Synechococcus* sp. strain PCC 7942. J Bacteriol 180: 2167–2174

Kutsuna S, Nakahira Y, Katayama M, Ishiura M and Kondo T (2005) Transcriptional regulation of the circadian clock operon *kaiBC* by upstream regions in cyanobacteria. Mol Microbiol 57: 1474–1484

Lavoie BD and Chaconas G (1996) Transposition of phage Mu DNA. Curr Top Microbiol Immunol 204: 83–102

Lawrence CW, Gibbs PE, Borden A, Horsfall MJ and Kilbey BJ (1993) Mutagenesis induced by single UV photoproducts in *E. coli* and yeast. Mutat Res 299: 157–163

Liu Y, Golden SS, Kondo T, Ishiura M and Johnson CH (1995a) Bacterial luciferase as a reporter of circadian gene expression in cyanobacteria. J Bacteriol 177: 2080–2086

Liu Y and Tsinoremas NF (1996) An unusual gene arrangement for the putative chromosome replication origin and circadian expression of *dnaN* in *Synechococcus* sp. strain PCC 7942. Gene 172: 105–109

Liu Y, Tsinoremas NF, Golden SS, Kondo T and Johnson CH (1996) Circadian expression of genes involved in the purine biosynthetic pathway of the cyanobacterium *Synechococcus* sp. strain PCC 7942. Mol Microbiol 20: 1071–1081

Liu Y, Tsinoremas NF, Johnson CH, Lebedeva NV, Golden SS, Ishiura M and Kondo T (1995b) Circadian orchestration of gene expression in cyanobacteria. Genes Dev 9: 1469–1478

Mackey SR, Ditty JL, Clerico EM and Golden SS (2007) Detection of rhythmic bioluminescence from luciferase reporters in cyanobacteria. Methods Mol Biol 362: 115–129

Maloy SR (2007) Use of antibiotic-resistant transposons for mutagenesis. Methods Enzymol 421: 11–17

Maurizi MR, Clark WP, Kim SH and Gottesman S (1990) ClpP represents a unique family of serine proteases. J Biol Chem 265: 12546–12552

McCarren J and Brahamsha B (2005) Transposon mutagenesis in a marine *Synechococcus* strain: isolation of swimming motility mutants. J Bacteriol 187: 4457–4462

Michael TP, Salome PA, Yu HJ, Spencer TR, Sharp EL, McPeek MA, Alonso JM, Ecker JR and McClung CR (2003) Enhanced fitness conferred by naturally occurring variation in the circadian clock. Science 302: 1049–1053

Michielse CB, Hooykaas PJ, van den Hondel CA and Ram AF (2005) *Agrobacterium*-mediated transformation as a tool for functional genomics in fungi. Curr Genet 48: 1–17

Mills DA (2001) Mutagenesis in the post genomics era: tools for generating insertional mutations in the lactic acid bacteria. Curr Opin Biotechnol 12: 503–509

Min H, Liu Y, Johnson CH and Golden SS (2004) Phase determination of circadian gene expression in *Synechococcus elongatus* PCC 7942. J Biol Rhythms 19: 103–112

Mitsui A, Kumazawa S, Takahashi A, Ikemoto H, Cao S and Arai T (1986) Strategy by which nitrogen-fixing unicellular cyanobacteria grow photoautotrophically. Nature 323: 720–722

Miyagishima SY, Wolk CP and Osteryoung KW (2005) Identification of cyanobacterial cell division genes by comparative and mutational analyses. Mol Microbiol 56: 126–143

Mizuuchi K (1992) Transpositional recombination: mechanistic insights from studies of mu and other elements. Annu Rev Biochem 61: 1011–1051

Mizuuchi K and Craigie R (1986) Mechanism of bacteriophage mu transposition. Annu Rev Genet 20: 385–429

Mizuuchi M and Mizuuchi K (1993) Target site selection in transposition of phage Mu. Cold Spring Harb Symp Quant Biol 58: 515–523

Mori T and Johnson CH (2001) Circadian programming in cyanobacteria. Semin Cell Dev Biol 12: 271–278

Muller EG and Buchanan BB (1989) Thioredoxin is essential for photosynthetic growth. The thioredoxin m gene of *Anacystis nidulans*. J Biol Chem 264: 4008–4014

Nair U, Ditty JL, Min H and Golden SS (2002) Roles for sigma factors in global circadian regulation of the cyanobacterial genome. J Bacteriol 184: 3530–3538

Nair U, Thomas C and Golden SS (2001) Functional elements of the strong *psbAI* promoter of *Synechococcus elongatus* PCC 7942. J Bacteriol 183: 1740–1747

Nakajima M, Imai K, Ito H, Nishiwaki T, Murayama Y, Iwasaki H, Oyama T and Kondo T (2005) Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation *in vitro*. Science 308: 414–415

Nishimura H, Nakahira Y, Imai K, Tsuruhara A, Kondo H, Hayashi H, Hirai M, Saito H and Kondo T (2002) Mutations in KaiA a clock protein extend the period of circadian rhythm in the cyanobacterium *Synechococcus elongatus* PCC 7942. Microbiology 148: 2903–2909

Ouyang Y, Andersson CR, Kondo T, Golden SS and Johnson CH (1998) Resonating circadian clocks enhance fitness in cyanobacteria. Proc Natl Acad Sci USA 95: 8660–8664

Pajunen MI, Pulliainen AT, Finne J and Savilahti H (2005) Generation of transposon insertion mutant libraries for Gram-positive bacteria by electroporation of phage Mu DNA transposition complexes. Microbiology 151: 1209–1218

Pattanayek R, Williams DR, Pattanayek S, Mori T, Johnson CH, Stewart PL and Egli M (2008) Structural model of the circadian clock KaiB-KaiC complex and mechanism for modulation of KaiC phosphorylation. EMBO J 27: 1767–1778

Pobigaylo N, Wetter D, Szymczak S, Schiller U, Kurtz S, Meyer F, Nattkemper TW and Becker A (2006) Construction of a large signature-tagged mini-Tn5 transposon library and its application to mutagenesis of *Sinorhizobium meliloti*. Appl Environ Microbiol 72: 4329–4337

Reznikoff WS (1993) The Tn5 transposon. Annu Rev Microbiol 47: 945–963

Richaud C, Zabulon G, Joder A and Thomas JC (2001) Nitrogen or sulfur starvation differentially affects phycobilisome degradation and expression of the *nblA* gene in *Synechocystis* strain PCC 6803. J Bacteriol 183: 2989–2994

Rippka R, Deruelles J, Waterbury J, Herdman M and Stanier R (1979) Generic assignments strain histories and properties of pure cultures of cyanobacteria. J Gen Microbiol 111: 1–61

Rust MJ, Markson JS, Lane WS, Fisher DS and O'Shea EK (2007) Ordered phosphorylation governs oscillation of a three-protein circadian clock. Science 318: 809–812

Schaefer MR and Golden SS (1989) Light availability influences the ratio of two forms of D1 in cyanobacterial thylakoids. J Biol Chem 264: 7412–7417

Schelin J, Lindmark F and Clarke AK (2002) The *clpP* multigene family for the ATP-dependent Clp protease in the cyanobacterium *Synechococcus*. Microbiology 148: 2255–2265

Schmitz O, Katayama M, Williams SB, Kondo T and Golden SS (2000) CikA, a bacteriophytochrome that resets the cyanobacterial circadian clock. Science 289: 765–768

Schneegurt MA, Sherman DM, Nayar S and Sherman LA (1994) Oscillating behavior of carbohydrate granule formation and dinitrogen fixation in the cyanobacterium *Cyanothece* sp. strain ATCC 51142. J Bacteriol 176: 1586–1597

Shestakov SV and Khyen NT (1970) Evidence for genetic transformation in blue-green alga *Anacystis nidulans*. Mol Gen Genet 107: 372–375

Singer B and Kusmierek JT (1982) Chemical mutagenesis. Annu Rev Biochem 51: 655–693

Sippola K, Kanervo E, Murata N and Aro EM (1998) A genetically engineered increase in fatty acid unsaturation in *Synechococcus* sp. PCC 7942 allows exchange of D1 protein forms and sustenance of photosystem II activity at low temperature. Eur J Biochem 251: 641–648

Smith RM and Williams SB (2006) Circadian rhythms in gene transcription imparted by chromosome compaction in the cyanobacterium *Synechococcus elongatus*. Proc Natl Acad Sci USA 103: 8564–8569

Soitamo AJ, Zhou G, Clarke AK, Oquist G, Aro EM and Gustafsson P (1994) Over-production of the D1 protein of photosystem II reaction centre in the cyanobacterium *Synechococcus* sp. PCC 7942. Plant Mol Biol 26: 709–721

Spiegel S and Bader KP (2003) Dependence of the flash-induced oxygen evolution pattern on the chemically and far red light-modulated redox condition in cyanobacterial photosynthetic electron transport. Z Naturforsch [C] 58: 93–102

Stal LJ and Krumbein WE (1985) Nitrogenase activity in the nonheterocystous cyanobacterium *Oscillatoria* sp. grown under alternating light-dark cycles. Arch Microbiol 143: 67–71

Stanne TM, Pojidaeva E, Andersson FI and Clarke AK (2007) Distinctive types of ATP-dependent Clp proteases in cyanobacteria. J Biol Chem 282: 14394–14402

Stothard P and Wishart DS (2005) Circular genome visualization and exploration using CGView. Bioinformatics 21: 537–539

Su J, Yang J, Zhao D, Kawula TH, Banas JA and Zhang JR (2007) Genome-wide identification of *Francisella tularensis* virulence determinants. Infect Immun 75: 3089–3101

Sugimoto Y, Tanaka K, Masuda S and Takahashi H (1997) The *rpoD*1: gene of *Synechococcus* sp. strain PCC 7942 encodes the principal sigma factor of RNA polymerase. J Gen Appl Microbiol 43: 17–21

Sugita C, Ogata K, Shikata M, Jikuya H, Takano J, Furumichi M, Kanehisa M, Omata T, Sugiura M and Sugita M (2007)

Complete nucleotide sequence of the freshwater unicellular cyanobacterium *Synechococcus elongatus* PCC 6301 chromosome: gene content and organization. Photosynth Res 93: 55–67

Sweeney BM and Borgese MB (1989) A circadian rhythm in cell division in a prokaryote the cyanobacterium *Synechococcus* WH7803. J Phycol 25: 183–186

Tadege M, Ratet P and Mysore KS (2005) Insertional mutagenesis: a Swiss Army knife for functional genomics of *Medicago truncatula*. Trends Plant Sci 10: 229–235

Takai N, Nakajima M, Oyama T, Kito R, Sugita C, Sugita M, Kondo T and Iwasaki H (2006) A KaiC-associating SasA-RpaA two-component regulatory system as a major circadian timing mediator in cyanobacteria. Proc Natl Acad Sci USA 103: 12109–12114

Taniguchi Y, Katayama M, Ito R, Takai N, Kondo T and Oyama T (2007) *labA*: a novel gene required for negative feedback regulation of the cyanobacterial circadian clock protein KaiC. Genes Dev 21: 60–70

Tsinoremas NF, Ishiura M, Kondo T, Andersson CR, Tanaka K, Takahashi H, Johnson CH and Golden SS (1996) A sigma factor that modifies the circadian expression of a subset of genes in cyanobacteria. EMBO J 15: 2488–2495

Tsinoremas NF, Kutach AK, Strayer CA and Golden SS (1994) Efficient gene transfer in *Synechococcus* sp. strains PCC 7942 and PCC 6301 by interspecies conjugation and chromosomal recombination. J Bacteriol 176: 6764–6768

Vazquez-Bermudez MF, Herrero A and Flores E (2003) Carbon supply and 2-oxoglutarate effects on expression of nitrate reductase and nitrogen-regulated genes in *Synechococcus* sp. strain PCC 7942. FEMS Microbiol Lett 221: 155–159

Vidan S and Snyder M (2001) Large-scale mutagenesis: yeast genetics in the genome era. Curr Opin Biotechnol 12: 28–34

van Waasbergen LG, Dolganov N and Grossman AR (2002) *nblS* a gene involved in controlling photosynthesis-related gene expression during high light and nutrient stress in *Synechococcus elongatus* PCC 7942. J Bacteriol 184: 2481–2490

Woelfle MA, Ouyang Y, Phanvijhitsiri K and Johnson CH (2004) The adaptive value of circadian clocks: an experimental assessment in cyanobacteria. Curr Biol 14: 1481–1486

Young MW and Kay SA (2001) Time zones: a comparative genetics of circadian clocks. Nat Rev Genet 2: 702–715

Zhang S, Laborde SM, Frankel LK and Bricker TM (2004) Four novel genes required for optimal photoautotrophic growth of the cyanobacterium *Synechocystis* sp. strain PCC 6803 identified by *in vitro* transposon mutagenesis. J Bacteriol 186: 875–879

Chapter 6

Understanding Photosynthetic Electron Transport Using Chlamydomonas: The Path from Classical Genetics to High Throughput Genomics

Arthur R. Grossman* and David González-Ballester
Department of Plant Biology, Carnegie Institution for Science, 260 Panama Street, Stanford, CA 94305, USA

Shaun Bailey
Aurora Biofuels, 301 Harbor Bay Parkway, Alameda, CA 94502, USA

Steven J. Karpowicz
Department of Chemistry and Biochemistry, University of California – Los Angeles, Box 951569, Los Angeles, CA 90095-1569, USA

Sabeeha S. Merchant
Department of Chemistry and Biochemistry and Institute for Genomics and Proteomics, University of California – Los Angeles, Box 951569, Los Angeles, CA 90095-1569, USA

*Author for correspondence, e-mail: arthurg@stanford.edu

R.L. Burnap and W.F.J. Vermaas (eds.), Functional Genomics and Evolution of Photosynthetic Systems, Advances in Photosynthesis and Respiration 33, pp. 139–176, DOI 10.1007/978-94-007-1533-2_6,

Summary

The Volvocales, an order of the green algal class Chlorophyceae, and the Streptophyte algae, the lineage that evolved into land plants, shared a common ancestor about one billion years ago. *Chlamydomonas reinhardtii* (Chlamydomonas throughout), a unicellular member of the Volvocales, has traditionally been considered a strong model organism that has been probed with sophisticated tools and methodologies to

Abbreviations: A – antheroxanthin; ALAD – δ-aminolevulinic acid dehydratase; BAC – bacterial artificial chromosome; CES – control of epistasis of synthesis; CGL – refers to proteins of the GreenCut conserved in the green lineage organisms; CGLD – refers to proteins of the GreenCut conserved in the green lineage organisms and the diatoms; Chl – chlorophyll; Chl^+ – chlorophyll cation; CHLH1 – H subunit of Mg-chelatase; CHLI1 – I subunit of the Mg-chelatase; CTH1 – catalyzes the cyclase reaction in chlorophyll biosynthesis; CP43 – chlorophyll binding antenna protein tightly bound to photosystem II; CPXI – coproporphyrinogen oxidase; CRD – an iron requiring cyclase involve in chlorophyll synthesis; D1 – the rapidly turning over reaction center protein of photosystem II; D2 – reaction center protein of photosystem II that associates with D1; DEG – protease involved in chloroplast biogenesis; Fd – ferredoxin; Fe – iron; FNR – ferredoxin NADP oxido-reductase; FQR – ferredoxin-quinone reductase; FLU – protein that regulates chlorophyll biosynthesis; EST – expressed sequence tag; FTSH – protease involved in the turnover of proteins of the photosynthetic apparatus; GSA – glutamate 1-semialdehyde (GSA) aminotransferase; GLK – golden-like kinase transcriptional regulator; HEMA – glutamyl tRNA reductase; IsiA – antenna chlorophyll binding proteins synthesized during iron deprivation (similar to CP43); IsiB – flavodoxin synthesized during iron deprivation; JGI – Joint Genome Institute; LHC – light harvesting complex; LHCII – light harvesting complex II; LHCA – light harvesting complex of photosystem I; LHCB – light harvesting proteins associated with photosystem II; LHCSR – protein in the light harvesting complex family that may be involved in photoprotection in Chlamydomonas; MCA1 – protein required for stable accumulation of *petA* RNA; Mg-ProtoIX – Mg-protoporphyrin IX; Mg-ProtoIXMe – Mg-protoporphyrin IX-monomethylester; NAB1 – forms a complex with mRNA encoding light harvesting proteins; NDH1 – NADH:ubiquinone oxidoreductase; OEE – oxygen evolving complex; PGR1 and PGR5 – protein thought to be part of the ferredoxin-quinone reductase complex; NPQ – non-photochemical quenching; PetA – cytochrome *f*; PSI – photosystem I; PSII – photosystem II; PSAH or PSIH – specific polypeptide associated with photosystem II; PSA – proteins associated with photosystem I (an additional letter indicates the subunit of the complex); PSB – protein associated with photosystem II (an additional letter indicates the subunit of the complex); PSBS – protein in the light harvesting family involved in qE-based quenching; PQ – plastoquinone; qE – quenching through the formation of an electrochemical gradient; qI – quenching through inhibition of photosystem II; qT – quenching through the formation of a state transition; *rbcL* – gene encoding the large subunit of ribulose-1,5-bisphosphate carboxylase; RNAi – RNA interference; RNA-seq – new generation of RNA sequencing; ROS – reactive oxygen species; RuBP – carboxylase ribulose-1,5-bisphosphate carboxylase; STN7 – Arabidopsis serine threonine protein kinase associated with state transition; STN8 – Arabidopsis protein with homology to STN7; STT7 – Chlamydomonas serine threonine protein kinase associated with state transition; TCA1 – protein involved in the translation of *petA* mRNA; TIC – proteins on the inner membrane of the chloroplast envelop involved in transporting proteins into the chloroplast; TOC – proteins on outer inner membrane of the chloroplast envelop involved in transporting proteins into the chloroplast; V – violaxanthin; Z – zeaxanthin; Z^+ – zeaxanthin cation

elucidate numerous biological processes. Perhaps the most in-depth analyses of Chlamydomonas have focused on defining proteins and complexes involved in the function and biogenesis of chloroplasts as well as the structure, assembly, and function of eukaryotic flagella (cilia); the latter are inherited from the common ancestor of animals and plants, but were lost during the evolution of land plants. This review emphasizes how Chlamydomonas has been used to elucidate a number of different activities associated with photosynthetic function. Many of these analyses were performed using classical genetic, biochemical and physiological approaches. However, recently, the DOE – Joint Genome Institute has sequenced the nuclear genome of Chlamydomonas (~120 Mb) and has helped the community of researchers perform comparative genomic analyses. Comparisons of deduced Chlamydomonas proteins have identified a set of proteins specifically present in the green lineage and other photosynthetic organisms, but not present in nonphotosynthetic organisms; this protein assemblage has been designated the GreenCut. Many proteins in the GreenCut are likely resident in the chloroplast and potentially associated with photosynthetic processes. Toward the end of this text we discuss the ways in which genomics has added a new dimension to our analyses of photosynthetic processes.

I. Introduction: Chlamydomonas as a Model Photosynthetic Eukaryote

A. Morphology and Phylogenetics of Chlamydomonas

Chlamydomonas reinhardtii (Chlamydomonas throughout) is a unicellular green alga present in soils and freshwater environments. It is a member of the Class Chlorophyceae, which diverged approximately one billion years ago from a progenitor that also evolved into the streptophyte (land plant) lineage. Chlamydomonas cells range in diameter from 5 to 10 μm, have a single, cup-shaped chloroplast that occupies approximately 50% of the cell volume, two flagella and associated basal bodies localized at the anterior region of the cell, and an eyespot associated with both chloroplast and cytoplasmic membranes (encompassing part of the thylakoid membrane, the chloroplast envelope and the plasma membrane). The eyespot helps orient the motile algal cells in a light gradient such that they move into light conditions favorable for increased photosynthesis. Interestingly, Chlamydomonas has retained many of the genes associated with protists and animal cells that were lost in the plant lineage. For example, the Chlamydomonas flagella and basal bodies are equivalent to cilia and centrioles, respectively, in animal cells; these structures do not occur in plant cells and the genes encoding components of the flagella/cilia and basal bodies/centrioles have been lost in vascular plants.

B. Why Use Chlamydomonas?

Organisms with relatively tractable genetic and molecular systems such as *Synechocystis*, Chlamydomonas and Arabidopsis are strong reference organisms for dissecting basic biological processes in plants, with Chlamydomonas also being an excellent reference for elucidating specific animal processes. Chlamydomonas has been fruitfully exploited for analyses of the basal body and flagellar structure and assembly (Dutcher, 1995; Goodenough et al., 1995; Lefebvre, 1995, 2009; Koutoulis et al., 1997; Fowkes and Mitchell, 1998; Tuxhorn et al., 1998; Bloodgood, 2009; King and Kamiya, 2009; Yang and Smith, 2009), intraflagellar transport (Scholey and Anderson, 2006; Scholey, 2008; Cole, 2009), establishing relationships between flagellar dysfunction and human diseases (Pazour and Witman, 2009), mitochondrial function and the properties of the mitochondrial genome (Funes et al., 2007; Cardol and Remacle, 2009), cell wall architecture and biogenesis (Goodenough, 1986; Woessner and Goodenough, 1989, 1992; Kurvari, 1997; Harris, 2009a), gametogenesis and mating (Goodenough et al., 1995, 2007; Ferris et al., 1996, 1997; Ferris and Goodenough, 1997; Harris, 2009b), phototaxis and photoperception (Pazour et al., 1995; Hegemann, 1997; Lamb et al., 1999; Dieckmann, 2003; Kottke et al., 2006; Schmidt et al., 2006; Wagner et al., 2008; Hegemann and Berthold, 2009; Sineshchekov et al., 2009), the characteristics of the chloroplast genome (Maul et al., 2002; Jiao et al., 2004; Higgs, 2009) and chloroplast

structure and function (de Vitry and Kuras, 2009; Finazzi et al., 2009; Gokhale and Sayre, 2009; Minagawa, 2009; Niyogi, 2009; Redding, 2009; Rochaix, 2009), regulation of gene expression and post-transcriptional processes in chloroplasts (Rochaix, 2001; Bollenbach et al., 2004; Drapier et al., 2007; Raynaud et al., 2007; Eberhard et al., 2008; Choquet and Wollman, 2009; Goldschmidt-Clermont, 2009; Herrin, 2009; Klein, 2009; Zerges and Hauser, 2009; Zimmer et al., 2009), and the activities and regulatory circuits associated with the assimilation of macronutrients (Camargo et al., 2007; Fernández and Galván, 2007, 2008; González-Ballester et al., 2008; Fernández et al., 2009; González-Ballester and Grossman, 2009; Moseley et al., 2009; Moseley and Grossman, 2009) and micronutrients (Tejada-Jimenez et al., 2007; Merchant et al, 2006; Kohinata et al., 2008; Long et al., 2008). Chlamydomonas is also an important model for studying light-driven H_2 production in photosynthetic organisms (Posewitz et al., 2009). A number of the references cited above are part of a new edition of the *Chlamydomonas Sourcebook*, edited by Harris, Stern and Witman (Harris, 2009c), which is an invaluable, up-to-date resource on most aspects of Chlamydomonas biology.

The physiological characteristics of Chlamydomonas make it an ideal organism for dissecting the structure and function of the photosynthetic apparatus. Chlamydomonas can be grown heterotrophically (in the dark or in the light) with acetate as a sole source of fixed carbon; mutants in which photosynthetic CO_2 fixation is completely blocked can be sustained on acetate-containing medium. This characteristic makes Chlamydomonas the only genetically tractable eukaryote in which lesions that eliminate photosynthesis are naturally conditional rather than lethal. Dark-grown, wild-type Chlamydomonas cells retain normal chloroplast structure and resume photosynthesis immediately following their transfer to the light (Harris, 1989). Hence, even mutants that are extremely sensitive to light (e.g., due to photooxidative reactions triggered even at low light intensity) can readily be maintained. By contrast, it is difficult to sustain light-sensitive mutants of vascular plants; homozygotes of such mutants of Arabidopsis experience extreme stress, even when supplied with a source of reduced carbon for growth. Since vegetative Chlamydomonas cells are haploid, mutant phenotypes are expressed almost immediately. Photosynthetic mutants were initially isolated as being unable to fix $^{14}CO_2$ in the light (Levine, 1960), and have subsequently been identified as acetate-requiring strains or as individual colonies in a population of thousands of colonies that exhibit aberrant chlorophyll (Chl) fluorescence. Furthermore, the duration of the Chlamydomonas life cycle (time from mating of gametes to analysis of tetrads) is only ~2 weeks, allowing for rapid genetic characterization of mutant strains. Several molecular techniques can be applied to Chlamydomonas, and cDNA libraries along with BAC, cosmid and plasmid genomic libraries are available. It is relatively simple to introduce genes into the Chlamydomonas genomes, control expression of those genes through a diversity of characterized promoters and tailor the transformed organisms for specific functionalities. Indeed, exogenous DNA can be introduced into all three Chlamydomonas genomes: nuclear, chloroplast and mitochondrial. Methods have also been developed to generate and identify tagged mutant alleles (Stevens et al., 1996; Sizova et al., 2001; Dent et al., 2005), and alleles that are not tagged can be identified by map-based cloning (Kathir et al., 2003; Rymarquis et al., 2005). Gene function can be evaluated by suppression of specific gene activities using antisense or RNAi constructs (Rohr et al., 2004; Schroda, 2006), and reporter genes have been developed to identify regulatory factors and *cis*-acting sequences involved in controlling gene expression (Davies et al., 1992; Leon-Banares et al., 2004). Recently a microRNA strategy has been developed to 'knock down' expression of specific genes (Molnar et al., 2007, 2009; Zhao et al., 2007, 2009).

Chlamydomonas is also one of the few green algae for which there is a full genome sequence (Merchant et al., 2007). This sequence information has been used to identify 259 tRNA genes and examine their clustering on the genome and their expression characteristics (Cognat et al., 2008). The sequence has also been used to identify and characterize gene families. Chlamydomonas displays gene duplications, which have led to the expansion of various gene families, including those encoding transporters (Moseley et al., 2006; Grossman et al., 2007; Merchant et al., 2007), thioredoxins (Moseley et al., 2006; Michelet et al., 2009), transcription factors

(Riano-Pachon et al., 2008), RNA modifying enzymes (Zimmer et al., 2008) and protein kinases (Wheeler et al., 2008). The genomic sequence has also helped in the generation of physical markers for mapping and isolating mutant alleles (Rymarquis et al., 2005), identifying genes encoding proteins critical for specific biosynthetic pathways (Lohr et al., 2005; Ball and Deschamps, 2009; Beale, 2009; Lohr, 2009; Reikhof and Benning, 2009; Vallon and Spalding, 2009), and constructing microarrays for genome-wide expression analyses (Zhang et al., 2004; Eberhard et al., 2006). The full genome sequence also allows for effective application of RNA-seq technology (Nagalakshmi et al., 2008; Wang et al., 2009; Wilhelm and Landry, 2009), in which the transcriptome is characterized through relatively short sequence reads that are mapped back to the genome and associated with specific gene models. Expression analyses based on microarray and in some cases RNA-seq data have been used to characterize the responses of Chlamydomonas to various stress conditions, including deprivation for iron (Fe), sulfur, phosphorus, copper and zinc (Castruita et al., 2010; González-Ballester et al., 2010) (Moseley, Casero, Pellegrini, Grossman {phosphorus}, unpublished; Urzica, Casero, Pellegrini, Merchant {Fe}, unpublished; Malasarn, Casero, Pellegrini, Merchant {zinc}, unpublished), and the responses of the organism to anaerobic conditions (Mus et al., 2007; Dubini et al., 2009) and excess excitation (Ledford et al., 2007). Overall, advances in our understanding of mechanistic aspects of a range of different biological processes in Chlamydomonas, our knowledge of proteins encoded on the Chlamydomonas genomes (nuclear, chloroplast, mitochondria) and their expression characteristics under various environmental conditions, and our ability to apply potent molecular and high throughput technologies to the study of Chlamydomonas, make its title "green yeast" (Goodenough, 1992; Rochaix, 1995) more appropriate than ever.

C. Historical Importance in Photosynthetic Processes

Genetic attributes of Chlamydomonas were clearly realized by Sager (Sager, 1960) and its ability to grow heterotrophically was exploited in early studies of Levine and colleagues for elucidating photosynthetic processes (Gorman and Levine, 1966; Givan and Levine, 1967; Lavorel and Levine, 1968; Levine, 1969; Levine and Goodenough, 1970; Moll and Levine, 1970; Sato et al., 1971). Indeed, many fundamental discoveries leading to today's knowledge of photosynthesis, including the sequence of carriers in the electron transfer pathway and the composition and mechanistic details of reaction centers, were derived from the application of physiological, biochemical and classical genetic approaches to dissect photosynthesis in Chlamydomonas. This initial work stimulated the development of highly sensitive, noninvasive screens for identifying photosynthetic mutants, mostly based on fluorescence (Bennoun and Levine, 1967; Shepherd et al., 1979; Bennoun and Delepelaire, 1982; Woessner et al., 1984; Bennoun and Beal, 1997), and systems to resolve thylakoid membrane proteins (Chua and Gillham, 1977; Delepelaire and Chua, 1979; Rolland et al., 2009; Stauber et al., 2009). Chlamydomonas was also central to the early studies of chloroplast ribosomes (Mets and Bogorad, 1971; Surzycki and Rochaix, 1971; Rochaix and Malnoe, 1978; Rochaix and Darlix, 1982).

Over the last two decades, Chlamydomonas has continued to dominate many aspects of photosynthesis research. The first physically defined lesion in the genome of the chloroplast was shown to be in the Chlamydomonas *rbcL* gene, which encodes the large subunit of ribulose-1,5-bisphosphate carboxylase (RuBP carboxylase) (Spreitzer and Mets, 1980; Dron et al., 1983). A change in the specificity factor for carboxylation was associated with the phenotype of an *rbcL* mutant (Chen et al., 1988). The labile 32 kDa D1 polypeptide was shown to be a photosystem II (PSII) component (Chua and Bennoun, 1975) and degradation of D1 was linked to photoinhibition (light-dependent inhibition of PSII activity) (Ohad et al., 1984). Studies of PSII and D1 and D2 in Chlamydomonas (Erickson et al., 1984, 1986, 1989; Ohad et al., 1984; Adir and Ohad, 1988) helped researchers recognize that these proteins were integral to the PSII reaction centers and functionally analogous to the L and M reaction center polypeptides of nonpurple sulfur bacteria (Deisenhofer et al., 1985; Michel and Deisenhofer, 1988), which led to the development of models describing the PSII repair cycle (Aro

et al., 1993a; Dannehl et al., 1995; Uniacke and Zerges, 2007). Studies of the multimeric PSI complex in Chlamydomonas have exploited biochemical approaches, targeted disruptions and site-directed mutagenesis to define interactions among subunits in the complex and the path of electron flow within the complex (Takahashi et al., 1992; Hallahan et al., 1995; Rodday et al., 1995, 1996; Webber et al., 1995, 1996; Fischer et al., 1997, 1998; Melkozernov et al., 1997; Hippler et al., 1998; Guergova-Kuras et al., 2001; Li et al., 2006; Nelson and Yocum, 2006). Other Chlamydomonas mutants are illuminating processes that control PSI stability (Boudreau et al., 1997; Moseley et al., 2000; Naumann et al., 2005) and assembly of photosynthetic "supercomplexes" (Nield et al., 2000; Swiatek et al., 2001; Melkozernov et al., 2005; Subramanyam et al., 2006).

Chlamydomonas has also served as a model for studies of the cell biology of photosynthesis and chloroplast biogenesis. Early work (Dobberstein et al., 1977; Schmidt et al., 1979) demonstrated that the small subunit of RuBP carboxylase was synthesized as a cytoplasmic precursor and suggested a role for the presequence in transport across the chloroplast envelope. These findings led to the development of *in vitro* systems in plants for analyzing the transport of polypeptides into chloroplasts and the energetic requirements of that process (Chua and Schmidt, 1978; Grossman et al., 1980; Schmidt and Mishkind, 1986; Cline and Henry, 1996). An elegant genetic selection was also developed that identified mutants with an altered ability to transport polypeptides into and across the thylakoid membranes (Smith and Kohorn, 1994; Bernd and Kohorn, 1998; Perret et al., 1998). Freeze-etch electron microscopy from mutant and wild-type Chlamydomonas thylakoid membranes led to the correlation of defined membrane particles with specific photosynthetic complexes (Goodenough and Staehelin, 1971). This work helped establish the fluidity of the photosynthetic apparatus within the membrane, exemplified by lateral movement of phosphorylated light-harvesting complexes in a process critical for light management termed state transition (see below) (Lemaire et al., 1986a; Wollman and Lemaire, 1988; Vernotte et al., 1990, 1992; Finazzi et al., 1999, 2001, 2002). Additional processes highlighting the dynamics of the photosynthetic apparatus were characterized, including non-photochemical quenching or NPQ (which is composed of at least three distinct processes; see below), control of cyclic electron flow, a regulatory phenomenon called control of epistasis of synthesis or CES (Choquet et al., 1998; Choquet and Vallon, 2000; Choquet and Wollman, 2009) and photosynthetic control (Eberhard et al., 2008). An abbreviated discussion of these processes in Chlamydomonas, and to some extent in vascular plants and cyanobacteria, is presented below. It is not our aim to be exhaustive in this discussion, but rather to highlight the dynamics and flexibility of photosynthetic electron transport and to point out numerous gaps in our understanding of these processes, many of which might be closed as we analyze and exploit information derived from the Chlamydomonas genome sequence and genome-wide analyses.

II. Dynamics of Photosynthetic Electron Transport

Photosynthetic functionality is exquisitely sensitive to environmental conditions, which markedly influence cellular metabolism and the growth potential of organisms. Furthermore, photosynthetic production of ATP and NADPH and the rate of utilization of these metabolites for the generation of carbon backbones and anabolic processes are strongly modulated by environmental conditions and coordinated through modification of photosynthetic activities in a process that has been designated photosynthetic control. The mechanistic aspects of this acclimatory control are numerous and may involve the extent of the pH gradient across the thylakoid membranes, the redox state of specific photosynthetic electron carriers, the levels of specific cellular metabolites and the generation of specific signaling molecules. In addition, some processes associated with acclimation of photosynthetic electron transport are short-term, initiating within seconds to minutes of the occurrence of a change in conditions. These acclimatory processes can rapidly alter the efficiency or functionality of the various complexes of the photosynthetic apparatus, but do not require *de novo* protein synthesis. Other changes may occur over minutes or hours and require protein modification, protein synthesis and turnover, resulting in remodeling of photosynthetic structures.

A. Components of NPQ: From Seconds to Hours

Non-photochemical quenching (NPQ) represents a group of processes that modulates the rate of excitation of the photosystems. It is mostly visualized by monitoring the fluorescence yields of PSII. In general, the level of NPQ correlates with the reduction state of the plastoquinone pool (PQ); increased reduction of this pool results in elevated NPQ and lowering of the maximal Chl fluorescence yield. The phenomenon of NPQ is comprised of three kinetically distinct components, a very fast component (short term) that occurs on the order of seconds to minutes, a slower component (moderate term) that occurs on the order of minutes to tens of minutes and a relatively slow component (long term) that can take many minutes to hours to become apparent. The short-term component of NPQ, designated qE, involves dissipation of energy absorbed by the PSII antenna bed as heat. The moderate-term component, designated qT, involves a state transition or a change in the relative cross section of the antenna pigments associated with PSI and PSII. The long-term component of NPQ, designated qI, results from photodamage of the D1 reaction center protein of PSII. Many aspects of NPQ have been extensively studied in Chlamydomonas and Arabidopsis, with biochemical and genetic approaches helping to identify specific components critical for this process (Niyogi 2009). The different mechanisms that cause NPQ and decreased PSII excitation are integrated with the rate of photosynthetic electron flow, the extent of the pH gradient across the thylakoid membranes and the redox state of photosynthetic electron carriers.

1. The qE Component

High energy state quenching or qE involves the dissipation of excess absorbed light energy within the antenna pigment-protein complexes of PSII. It generally occurs when organisms absorb excess excitation and the PQ pool becomes highly reduced. Most photosynthetic organisms have qE to a greater or lesser extent, and impairment of qE reduces the fitness of the organism to cope with dramatically fluctuating physical conditions in the natural environment (Kulheim et al., 2002). The biophysical consequence of qE is a shortening of the fluorescence lifetime of specific Chl molecules in the PSII antenna bed (Gilmore et al., 1995), which reflects a de-activation of singlet excited Chl molecules through a harmless, non-radiative process. Elicitation of qE requires a ΔpH across the thylakoid membranes and conversion of violaxanthin (V) to antheraxanthin (A) and zeaxanthin (Z), in a pathway designated the xanthophyll cycle (Jahns et al., 2009). V to A and Z conversion is catalyzed by a low pH-activated (often elicited by high light exposure), thylakoid membrane-bound, lumenal violaxanthin de-epoxidase (Pfundel and Dilley, 1993). The qE-based quenching relaxes when algae/plants are placed in the dark or in low light; the pH gradient across the thylakoid membrane dissipates and Z is converted back to V by Z epoxidase.

Detection of qE is non-invasive and easily visualized (Niyogi, 1999; Pogson and Rissler, 2000; Holt et al., 2004; Horton and Ruban, 2005; Szabo et al., 2005); it is readily monitored using a video imaging system. Such a system has been used to identify several Arabidopsis and Chlamydomonas mutants with aberrant qE (Niyogi et al., 1997, 1998; Niyogi, 1999), many of which have a defective xanthophyll cycle. While the precise mechanism for qE is still controversial (Polivka et al., 1999), the S1 energy state of Z (and A) has been proposed to be below the Q_y (energy state) transition of Chl, allowing excitation energy to be directly transferred from Chl to carotenoids (Frank and Cogdell, 1996). Quenching may also involve a charge transfer within a complex containing interacting Chl-carotenoid molecules to generate Chl^- and Z^+ followed by a rapid relaxation (back reaction) to the ground state (Holt et al., 2005); this process may be impacted by low abundance light harvesting complex (LHC) polypeptides, including PSII-S (or PSBS) in plants (Li et al., 2000, 2002a, b, 2004; Niyogi et al., 2005) and LHCSR2/LHCSR3 polypeptides in Chlamydomonas (Peers et al., 2009). The LHCSR proteins are encoded on the genomes of other algae but have not been identified on the genomes of land plants. It should be kept in mind that several molecules and factors likely influence qE, including the levels of specific thylakoid membrane polypeptides, the conformational and aggregation state of the pigments and

proteins of antenna complexes and the population of carotenoids bound to the antenna complexes (Horton and Ruban, 2005; Ruban et al., 2007; Mozzo et al., 2008; Jahns et al., 2009).

2. The qT Component

State transition, which is elicited by changes in the redox state of photosynthetic electron carriers, involves reversible redistribution of mobile LHCII antennae complexes between the two photosystems. This phenomenon, most prominent in unicellular photosynthetic organisms, was first recognized in the green alga *Chlorella pyrenoidosa* (Bonaventura and Meyers, 1969) and the red alga *Porphyridium cruentum* (Murata, 1969). Initially, it was attributed a major function in balancing the absorption of excitation energy by the two photosystems as the wavelengths of light in the environment fluctuate, but it also may help photosynthetic organisms adjust to low but fluctuating light intensities (Tikkanen et al., 2006). While state transitions do occur in vascular plants, studies of Chlamydomonas have been essential for elucidating mechanisms associated with this phenomenon (Rochaix, 2002, 2009).

When PSII and the PQ pool are oxidized, most mobile antenna are associated with PSII; the cells are described as being in state 1. As PSII absorbs more excitation energy, electron carriers on the reducing side of PSII become reduced and the antenna proteins move from PSII to PSI; the cells are described as being in state 2 (Vener et al., 1997; Wollman, 2001). This movement of antenna from PSII (mostly in the appressed or "grana" membranes) to PSI (mostly in the non-appressed membranes, which include the stroma lamellae and the grana margins) is dependent upon the activity of a membrane-bound protein kinase (Depege et al., 2003; Rochaix, 2007) that phosphorylates LHCII polypeptides (Bennett et al., 1980; Allen, 1992; Nilsson et al., 1997), and also the interaction of antenna proteins with PSAH (or PSIH) (Lunde et al., 2000). Regulation of a state transition-associated protein kinase activity by electron flow through the cytochrome b_6f complex was first recognized in Chlamydomonas (Lemaire et al., 1986a; Wollman and Lemaire, 1988), and the kinase itself was first identified in Chlamydomonas using a fluorescence-based screen to identify strains aberrant for state transitions (Depege et al., 2003). The first state transition mutant identified, *stt7*, was locked in state 1 and was unable to phosphorylate LHCII proteins under conditions that would elicit a transition to state 2. The *STT7* gene encodes a thylakoid membrane-associated serine–threonine protein kinase that phosphorylates LHCII. Arabidopsis has two STT7 homologs, STN7, and STN8 (Bellafiore et al., 2005; Bonardi et al., 2005; Vainonen et al., 2005). In addition to phosphorylation of some of the major LHC polypeptides that form trimeric complexes, CP29, a low-abundance component of the PSII antenna, also undergoes a reversible phosphorylation (Turkina et al., 2006) and appears to be essential for state transitions (Iwai et al., 2008; Tokutsu et al., 2009); it plays a role in the docking of LHCII proteins to PSI during the transition to state 2.

3. Cyclic Electron Flow

Cyclic electron flow was first discovered by Arnon (1959), although the exact mechanisms and components required for this process are still under debate. Furthermore, there appears to be a connection between cyclic electron flow and state transitions. While potential functions of a state transition are to balance electron flow between the photosystems and prevent hyper-reduction of PSI (Johnson, 2005), the state transition may also drive interconversion of photosynthetic electron transport, especially in algal cells, between predominantly linear vs. predominantly cyclic electron flow (Wollman, 2001; Finazzi and Forti, 2004). While linear electron flow produces both reductant and ATP, cyclic electron flow generates a ΔpH and hence only ATP. This dynamic in the function of photosynthetic electron transport provides an 'energetic' flexibility that helps tune cells to a continually changing environment. For example, cells stressed and unable to grow (or grow only very slowly) would not require photosynthetically-derived reductant to generate carbon skeletons, but would need energy for cell maintenance and repair. Hence, the contribution of cyclic relative to linear electron flow may be greater in cells experiencing suboptimal growth conditions (Rumeau et al., 2007; Shikanai, 2007). One mechanism for photoinhibition is initial damage to the oxygen-evolving complex through the absorption of light by manganese (Hakala

et al., 2005; Ohnishi et al., 2005; Nishiyama et al., 2006; Tyystjarvi, 2008) followed by light-dependent damage to PSII reaction centers (Hakala et al., 2005; Ohnishi et al., 2005). Cyclic electron flow helps ameliorate photoinhibitory processes by at least two different mechanisms. First, it is important for establishing a ΔpH across the thylakoid membranes, which would stimulate qE, and recent work suggests that qE is needed for efficient de novo protein synthesis in chloroplasts and D1 repair; when qE is low there may be damage to the translational machinery through ROS production. Cyclic electron flow also appears to impact the repair process in a qE-independent manner (Takahashi et al., 2009).

Presently, at least two major routes for cyclic electron flow have been identified. The first involves the chloroplast NDH1 complex, which is analogous to mitochondrial complex I; the complex catalyzes electron transfer from NADPH (or NADH) to the PQ pool (Peltier and Cournac, 2002; Battchikova and Aro, 2007; Rumeau et al., 2007; Endo et al., 2008). While several mutants have been identified that exhibit aberrant assembly of the NDH1 complex (Shimizu and Shikanai, 2007; Shimizu et al., 2008; Ishida et al., 2009; Martin et al., 2009), the quantitative contribution of this pathway to cyclic electron flow is not clear since maximal rates of cyclic flow in *ndh* mutants are not significantly altered (Nandha et al., 2007). Another potential route for cycling electrons around PSI involves a complex designated the Ferredoxin-Quinone Reductase (FQR), which reduces PQ with electrons derived directly from ferredoxin (Fd). The activity of this complex appears to depend on the PGR5 and PGRL1 polypeptides, which may be part of an FQR complex (Shikanai, 2007; DalCorso et al., 2008), and on environmental conditions, with activity increasing at high light intensity (Ooyabu et al., 2008). The conclusion concerning the role of PGR proteins in FQR is mainly based on the diminished ability of these mutants to reduce PQ (Munekage et al., 2004). Furthermore, the plant PGR5 homolog in cyanobacteria is required for FQR activity, which operates in parallel to NDH1-driven cyclic electron flow (Yeremenko et al., 2005). Distinct roles of the NDH1 and FQR complexes, their functional interactions and components of each of the complexes (and potentially incorporation into supercomplexes) need to be further elucidated.

4. The qI Component

PSII and water oxidation are inactivated by a process designated photoinhibition (Aro et al., 2005; Osmond and Forster, 2006; Edelman and Mattoo, 2008; Takahashi and Murata, 2008), which occurs when the rate of light energy conversion or the utilization of excitation energy for photochemistry is lower than the rate of light energy absorption. This imbalance in energy absorption and utilization can result in ROS accumulation and damage to PSII reaction centers, and especially to the D1 reaction center protein. An efficient system to repair this damage has evolved that requires partial disassembly of inactive PSII, proteolytic degradation of the photodamaged D1 protein and co-translational insertion of newly synthesized D1 into PSII (Barber and Andersson, 1992; Aro et al., 2005). Therefore, it is only when the rate of PSII repair lags behind the rate of photodamage that loss of PSII activity is observed. Furthermore, the distribution of photosynthetic complexes in the thylakoid membranes is not uniform. Dimeric, functional PSII is located mostly in appressed membranes, while PSI resides mostly in non-appressed membranes (Andersson and Anderson, 1980; Albertsson, 2001; Danielsson et al., 2006), although there is a finer structural segregation of the complexes within each of these membrane types that may correlate with the aggregation state/subunit composition and potentially other features of the photosynthetic complexes (Danielsson and Albertsson, 2009). In contrast, the cytochrome b_6f complex exhibits a relatively homogeneous distribution in the thylakoid membrane, although in Chlamydomonas the distribution does change dependent on whether the cells are in state 1 or state 2 (Vallon et al., 1991). In plants, the repair cycle requires lateral migration of PSII complexes between the appressed and non-appressed membranes (Aro et al., 2005), which is facilitated by a phosphorylation-elicited dissociation of reaction centers into monomers and migration of the monomers out of the appressed membranes (Aro et al., 1993b). The STN8 protein kinase of Arabidopsis appears to play a role in this feature of the repair cycle (Tikkanen et al., 2008a, b). Furthermore, degradation of damaged D1 may involve dephosphorylation of specific amino acid residues at the amino terminus; this dephosphorylation occurs rapidly at elevated temperatures

(Komayama et al., 2007; Yamamoto et al., 2008). The activities of the proteases FTSH and potentially DEG-P2 (Bailey et al., 2002; Komenda et al., 2006, 2007; Yoshioka et al., 2006; Cheregi et al., 2007; Kapri-Pardes et al., 2007) are involved in the degradation pathway. Importantly, chloroplasts house a number of other proteases likely to contribute to chloroplast function and the dynamics of photosynthetic activities (Adam and Clarke, 2002; Adam et al., 2006; Sakamoto, 2006), although specific functionalities of these proteases are largely unexplored.

B. Environment and the Photosynthetic Apparatus

Changes in environmental conditions trigger altered synthesis and accumulation of photosynthetic complexes and changes in the stoichiometry of the photosystems (Anderson et al., 1995; Walters, 2005; Foyer et al., 2009). These phenomena have been most thoroughly examined with respect to changing light conditions and also under some conditions of nutrient limitation. However, detailed molecular interactions critical for acclimation have not been completely defined.

1. Light and Antenna Changes

Dramatic alterations in gene expression have been observed when plants and algae are exposed to high light, different wavelengths of light, or oxidative stress conditions (Ma et al., 2001; Jiao et al., 2005; Walters, 2005; Fischer et al., 2006; Ledford et al., 2007). Some of the most apparent changes associated with light involve altered pigment composition, and the size (Walters, 2005) and composition of the LHCs (Jackowski et al., 2003; Caffarri et al., 2005; Ballottari et al., 2007; Frigerio et al., 2007). The remodeling of the light-harvesting antenna is mostly a consequence of changes in the abundance of proteins in the PSII LHCB family, with little change in PSI LHCA protein content (Ballottari et al., 2007). Protein degradation (Bailey et al., 2001; Jackowski et al., 2003; Zelisko et al., 2005; Frigerio et al., 2007) and/or inhibition of translation (Mussgnug et al., 2005; McKim and Durnford, 2006), followed by dilution as cell division continues, impact the absolute quantities and ratios of light-harvesting components. Although the protease(s) involved in degradation of LHC proteins is not known, a metalloprotease in the FTSH family is potentially involved (Garcia-Lorenzo et al., 2005). Furthermore, while numerous environmental factors impact restructuring of the photosynthetic apparatus, some of the changes are dependent on the quantity/quality of incident light and have been linked to photoreceptor control (Ghassemian et al., 2006; Im et al., 2006; Li et al., 2009). Interestingly, in Arabidopsis, light-dependent changes in antenna size were not observed in the *STN7* mutant (Bonardi et al., 2005) (see below).

2. Nutrient Deprivation

Marked changes in the photosynthetic apparatus are also observed in response to changes in the nutrient environment. Alterations in photosynthetic activity during phosphorus and sulfur deprivation include reduced electron flow, accumulation of photoinhibitied PSII reaction centers with a reduced rate of Q_B reduction and increased elimination of excess absorbed excitation energy as heat (Davies et al., 1996; Wykoff et al., 1998). Specific light harvesting proteins may also accumulate; sulfur deprivation of Chlamydomonas leads to the decline of all LHCBM and LHCA polypeptides except for LHCBM9 (Nguyen et al., 2008; González-Ballester et al., 2009). The increase in the level of the *LHCBM9* transcript seems to be specific for sulfur deprivation and the encoded polypeptide has fewer sulfur-containing amino acids than the other LHCBMs, which suggests the occurrence of sulfur sparing in sulfur-starved cells. There is also an increase in the transcripts encoding the LHCSR proteins; these proteins are related to light-harvesting proteins, are not found in vascular plants, and have been associated with stress responses in Chlamydomonas (Savard et al., 1996; Richard et al., 2000; Nguyen et al., 2008; González-Ballester et al., 2009). Recent work suggests that these proteins are critical for NPQ and survival at high light intensity (Peers et al., 2009). A more detailed molecular analysis of these processes is required to better understand their physical and functional consequences.

Fe is also a critical element in photosynthesis, with 12 Fe atoms in each PSI reaction center and 5 Fe atoms in the cytochrome b_6f complex. Several photosynthetic organisms, including the diatoms (Strzepek and Harrison, 2004), *Ostreococcus*

(Cardol et al., 2008), Chlamydomonas (Moseley et al., 2002), and vascular plants (Sharma, 2007) exhibit a specific reduction in PSI content when Fe becomes scarce (Pilon et al., 2006); other effects on photosynthetic activities have also been noted (Larbi et al., 2006; Varsano et al., 2006; Naumann et al., 2007). In cyanobacteria, upon Fe deprivation the reduction in the amount of PSI reaction centers relative to other photosynthetic complexes can be marked (Guikema and Sherman, 1984; Sandstrom et al., 2002; Chen et al., 2005); in some cases the reduction in PSI is constitutive and linked to the evolution of cells in a low Fe environment (Bailey et al., 2008). Upon Fe deprivation, cyanobacteria exhibit a dramatic loss in Chl and phycobilisomes (mobile antenna mostly associated with PSII) (Guikema and Sherman, 1983; Pakrasi et al., 1985), remodel PSI-associated antenna (Bibby et al., 2001; Nield et al., 2003; Murray et al., 2006), and make IdiA, which is associated with PSII (Michel and Pistorius, 2004; Lax et al., 2007), and IsiB, which is a flavodoxin that substitutes for the Fe-containing Fd (Leonhardt and Straus, 1994). The restructuring of PSI during Fe deprivation has been extensively studied and involves the synthesis of IsiA, a Chl-binding protein similar to the tightly bound, PSII antenna polypeptide CP43. The newly synthesized IsiA forms a ring around the PSI cores (Nield et al., 2003; Murray et al., 2006; Wilson et al., 2007). Having fewer PSI cores associated with large antenna complexes allows the PSI centers to more efficiently capture incident excitation. Furthermore, organisms with low PSI that thrive in environments in which Fe is never abundant (oligotrophic oceans) might also have significant plastoquinol terminal oxidase (PTOX) activity; this enzyme appears to use electrons from the PQ pool and reduces O_2 and H^+ to H_2O (Rumeau et al., 2007; Bailey and Grossman, 2008; Bailey et al., 2008), thereby ameliorating the effect of absorption of excess excitation by PSII and maintaining the proper redox state of PSI (an overabundance of electrons in the iron sulfur clusters of PSI could lead to the formation of ROS).

The remodeling of the PSI antenna upon Fe deprivation has also been characterized in the green algae *Dunaliella* and Chlamydomonas. When *Dunalliela salina* is deprived of Fe, the level of PSI reaction centers declines, the size of isolated PSI particles increases and the cells accumulate a 45 kDa polypeptide designated Tidi, which contains peptide sequences similar to those of LHC proteins. Furthermore, Tidi polypeptides appear to be associated with PSI, increasing the absorption efficiency of the reduced number of PSI core complexes in Fe-starved cells (Varsano et al., 2006).

In Chlamydomonas, Fe deprivation elicits a reduction in the level of the photosystems and the cytochrome b_6f complex. Deprived cells also remodel the PSI antenna even before there are marked visual signs of starvation (e.g., chlorosis). The sequential changes in PSI characteristics that are observed as cells acclimate to low Fe conditions are: (1) uncoupling of antenna complexes from the PSI core, (2) degradation of the original complement of LHC polypeptides, (3) expression of genes encoding new LHC polypeptides, and (4) assembly of new LHC-PSI complexes (Moseley et al., 2002). These changes may protect PSI from oxidative damage as Fe levels decline and cell growth diminishes. The uncoupling of the antenna from PSI likely results from a loss of the Chl-binding PSAK polypeptide, which appears to connect specific PSI antenna polypeptides to the core complex (Jansson et al., 1996; Jensen et al., 2000; Jordan et al., 2001). Elimination of PSAK from PSI may occur because the nascent polypeptide may not effectively compete for Chl molecules under conditions in which Chl levels decline. A decline in Chl may be a consequence of reduced levels of CRD1, an Fe-requiring cyclase that synthesizes the fifth ring of Chl molecules (Tottey et al., 2003; Allen et al., 2008). The importance of remodeling PSI antenna was demonstrated using the light sensitive mutant *psaF*. This strain grew better when it was starved for Fe (on 0.1 μM Fe relative to 18 μM Fe), which would trigger remodeling of PSI and reduced delivery of light to the PSI reaction center Chl. A *CRD1* mutant also grew better under Fe-starvation conditions since it already has reduced levels of PSAK and uncoupled PSI antenna (similar to the situation that develops during Fe deprivation). This work clearly shows the evolution of specific remodeling processes that enhance cell survivability as nutrient conditions in the environment change. Note that the changes described for Chlamydomonas are documented in photoheterotrophically grown cells. The changes in phototrophically grown cells are distinct (Merchant, unpublished). Remodeling of

the photosynthetic apparatus when the cells are deprived of nutrients is likely to be critical since the photosynthetic apparatus would be experiencing high excitation pressure even at low/moderate light levels since starved cells generally cannot fix CO_2 at high rates. Changes in the composition/organization of the photosynthetic apparatus appear to be essential to the survival of Chlamydomonas cells as they acclimate to sulfur or phosphorus deprivation (Davies et al., 1996; Moseley et al., 2006; González-Ballester et al., 2008), although a more detailed molecular analysis of these processes is required to understand its physical and functional consequences.

C. The Signals

The functional flexibility of the photosynthetic apparatus allows photoautotrophs to cope with dramatic environmental fluctuations. Coordinating cellular responses to environmental conditions requires communication among subcellular compartments since many of the complexes are comprised of components encoded by both nuclear and organellar genomes. There is substantial evidence that retrograde signaling by compounds generated in the chloroplast impacts nuclear gene expression (Beck, 2005; Leister, 2005; Laloi et al., 2006; Nott et al., 2006; Fischer et al., 2007; Koussevitzky et al., 2007; Pesaresi et al., 2007; Oelze et al., 2008; Pogson et al., 2008). Furthermore, many molecules synthesized in the cytoplasm of the cell and imported into chloroplasts exert post-transcriptional control on the accumulation of chloroplast-encoded proteins (Bollenbach et al., 2009).

At this point, it is not clear which signals in photosynthetic control are dominant in controlling accumulation of the various proteins (e.g. LHC polypeptides) involved in photosynthetic function. However, relevant factors include ROS (Apel and Hirt, 2004; Mittler et al., 2004; Hideg et al., 2006; Laloi et al., 2006; Moller et al., 2007), pigments and potentially intermediates in tetrapyrrole biosynthesis (Beck, 2005; Koussevitzky et al., 2007) (but see below), the redox state of carriers in the photosynthetic electron transport system (Fey et al., 2005; Foyer et al., 2009), photoperception (Li et al., 2009), and the potentially complex integration of light, redox, metabolic and developmental elements (Ruckle et al., 2007; Larkin and Ruckle, 2008; Lepisto et al., 2009). A number of specific regulatory molecules have been shown to be involved in transcriptional and post-transcriptional events that govern levels of proteins involved in photosynthesis, and especially light harvesting proteins. The GOLDEN-like kinase (GLK) transcriptional regulators play an important role in the biogenesis of chloroplasts, especially with respect to controlling activities of genes associated with light harvesting and Chl biosynthesis (Waters et al., 2009). The nuclear factor NAB1 forms a complex with *LHCBM* mRNA in the cytoplasm of the cell, which potentially results in sequestration of the mRNA (Mussgnug et al., 2005). Chl *b* is required for the import of light harvesting polypeptides (Kuttkat et al., 1997; Espineda et al., 1999; Reinbothe et al., 2006, 2008). Thioredoxins may also modulate the import of proteins into plastids (Bartsch et al., 2008). Furthermore, two-component regulators may control expression of some chloroplast genes (Puthiyaveetil and Allen, 2009), and it is possible that they couple photosynthetic activity and the redox state of the plastid with the regulation of chloroplast gene activity (Puthiyaveetil et al., 2008). It is likely that an integration of intracellular and extracellular cues and specific biochemical interactions will be critical for establishing the final levels of photosynthetic proteins.

Characterizations of nuclear mutants that impact photosynthetic function have been invaluable in elucidating the biogenesis of several chloroplast complexes, and the ways in which the synthesis of chloroplast-encoded proteins are controlled. A number of studies on the biogenesis of the cytochrome b_6f complex have focused on how unassembled subunits of the complex control levels of the other component polypeptides (Lemaire et al., 1986b; Howe and Merchant, 1992, 1994; Girard-Bascou et al., 1995; Gumpel et al., 1995; Inoue et al., 1997; Kuras et al., 1997; Drager et al., 1998; Wollman, 1998; Xie and Merchant, 1998; Kuras et al., 2007; Lezhneva et al., 2008). Chlamydomonas mutants unable to synthesize subunit IV or cytochrome b_6 exhibited a marked reduction in cytochrome *f* (Kuras and Wollman, 1994); this reduction in cytochrome *f* polypeptide abundance was a consequence of autoregulation of the initiation of translation from the cytochrome *f*-encoding *petA* mRNA (Choquet

et al., 1998) in a process mediated by the binding of the C-terminal domain of unassembled cytochrome *f* to *petA* messenger RNA (Choquet et al., 2003). This regulatory process in which a polypeptide subunit of a multisubunit complex is required for sustained synthesis of other subunits in the complex is termed CES (control by epistasy of synthesis) regulation (Wollman et al., 1999; Choquet and Vallon, 2000; Choquet and Wollman, 2002). This regulatory phenomenon is not unique to the cytochrome b_6f complex, but has also been noted for chloroplast DNA-encoded polypeptides of PSII as well as for PSI reaction center subunits. The synthesis of PSII core polypeptides is dependent on partner proteins in the PSII core complex. Chlamydomonas mutants lacking D1 also have diminished levels of CP47, but not of the D2 polypeptide (Bennoun et al., 1986; Jensen et al., 1986; de Vitry et al., 1989), while mutants lacking D2 exhibit diminished D1 and apoCP47 synthesis (Bennoun et al., 1986; Erickson et al., 1986). Similarly, the PsaA and PsaC proteins are under CES control (Wostrikoff et al., 2004). This work, performed using Chlamydomonas, has helped elucidate biosynthetic processes that are very difficult to examine in vascular plants and highlights the fact that chloroplast biogenesis is often controlled at the post-transcriptional level (Mayfield, 1996; Rochaix, 1996; Yohn et al., 1996, 1998a, b; Kim and Mayfield, 1997; Stern et al., 1997; Stern and Drager, 1998). CES has been thoroughly reviewed recently (Choquet and Wollman, 2009).

1. Nuclear Proteins and Chloroplast Gene Expression

Detailed studies with Chlamydomonas and to some extent with Arabidopsis have demonstrated the occurrence of nucleus-encoded, chloroplast proteins that impact chloroplast mRNAs abundance. Many of these proteins act post-transcriptionally on either a single or a small subset of chloroplast transcripts to influence transcript maturation, half-life or translation (Herrin and Nickelsen, 2004; Rochaix, 2004; Rochaix et al., 2004; Merendino et al., 2006; Rymarquis et al., 2006, 2007; Goldschmidt-Clermont et al., 2008; Goldschmidt-Clermont, 2009; Herrin, 2009). There is still considerable work required to determine the specificity and precise activity of these nucleus-encoded regulatory factors, whether they are integrated into larger complexes, and the relationships among polypeptides of the various multimeric complexes that function in posttranscriptional control of gene expression. Some of these regulatory elements are also linked to the acclimation of organisms to fluctuating environmental conditions, such as light and/or the redox status of the photosynthetic electron transport chain. For example, Kim and Mayfield (2002) proposed that a specific nuclear polypeptide in Chlamydomonas is part of a complex that controls D1 translation in response to redox and light signals, although there are still issues concerning the activity and specificity of these factors (Schwarz et al., 2007). Other nucleus-encoded regulatory elements include ATAB2, an RNA binding protein that controls the translation of photosystem proteins in Arabidopsis (Barneche et al., 2006). Furthermore, the Chlamydomonas MCA1 protein is required for stable accumulation of *petA* mRNA (encodes cytochrome *f*) while the TCA1 protein controls translation of the *petA* transcript (Raynaud et al., 2007). These nucleus-encoded regulatory elements may be important for modulating *petA* expression in response to nutrient levels (Wostrikoff et al., 2001; Raynaud et al., 2007; Loiselay et al., 2008). When environmental nitrogen levels decline, there is a congruent decline in TCA1, MCA1 and the polypeptides of the cytochrome b_6f complex (Raynaud et al., 2007). Several nucleus-encoded proteins are also involved in the trans-splicing of the chloroplast encoded *psaA* transcript in Chlamydomonas.

2. Retrograde Signaling

Retrograde control involves metabolic signals generated in chloroplasts that regulate expression of nuclear genes. Although there is still controversy about the nature of some of the molecules involved in this process, there are a number of pathways and metabolites potentially involved. Work from the laboratory of Beck (Kropat et al., 1997, 2000; Chekounova et al., 2001) has implicated Mg-protoporphyrin IX (Mg-ProtoIX) and Mg-protoporphyrin IX-monomethylester (Mg-ProtoIXMe), which are intermediates in Chl biosynthesis, and possibly also the Mg-chelatase subunit H, in retrograde signaling. In one study it was shown that Mg-ProtoIX but not ProtoIX

could elicit expression of the Chlamydomonas HSP70 gene (normally light induced) in the dark (Kropat et al., 2000). Mg-ProtoIX and Mg-ProtoIXMe and Mg-chelatase have also been associated with retrograde signaling in Arabidopsis. Based on a mutant screen in which regulation of the *LHCB* genes was uncoupled from light regulation, it was proposed that Mg-ProtoIX and Mg-ProtoIXMe served to inhibit expression of the *LHCB* genes. However, the interpretation of some of these results is complicated and recent studies indicate that the accumulation of intermediates in Chl biosynthesis (e.g., Mg-Protoporphyrin IX) do not correlate with expression of LHCB1, suggesting that they do not play a direct role in retrograde signaling (Mochizuki et al., 2008; Moulin et al., 2008). However, perturbations of tetrapyrrole synthesis may alter the redox state of plastids or provoke local accumulation of ROS, either of which may participate in the generation of retrograde signals.

ROS, generated as a consequence of photosynthetic electron transport, may also be critical for retrograde signaling. ROS production, light acclimation, programmed cell death, and plant defense responses all appear to be sensitive to changes in the redox state of the PQ pool (Muhlenbock et al., 2008); the relationships between ROS production and many of the downstream phenotypic responses still need to be established (although see below). The three ROS types produced in chloroplasts are $^1O_2$, mainly a byproduct of PSII activity, and O_2^-, which leads to H_2O_2 production, formed mainly on the acceptor side of PSI. Elevated levels of H_2O_2 and O_2^- in cells stimulate the accumulation of some transcripts encoding photosynthetic proteins, while others decline (Desikan et al., 2001; Fischer et al., 2007; Krieger-Liszkay et al., 2008; Shao et al., 2008; Li et al., 2009). Furthermore, Arabidopsis plants in which chloroplast superoxide dismutase was 'knocked down' bleached to some extent, with an accompanying decline in abundances of transcripts encoding LHC proteins, the oxygen-evolving complex proteins (OEEs), and ferredoxin-NADPH-oxidoreductase while other proteins required for chloroplast function increased (Rizhsky et al., 2003). A number of studies have now shown that photosynthetically-generated ROS can markedly impact expression of numerous nuclear genes (Apel and Hirt, 2004; Mittler et al., 2004; Gadjev et al., 2006; Moller et al., 2007). However, oxidative conditions were also shown to control the translation of *rbcL* transcripts: elevated levels of ROS lead to translation arrest (Irihimovitch and Shapira, 2000; Cohen et al., 2006). Since ROS can interact with and modify molecules both in the membranes and cytoplasm of the cells, the distinct effects of each ROS type, as well as the molecules directly involved in the signaling, are difficult to define.

A recent and important contribution to the field of oxidative signaling was the identification of a pathway involved in $^1O_2$-specific retrograde signaling in Arabidopsis based on analysis of the *flu* mutant. Unlike in wild-type plants, the *flu* mutant of Arabidopsis accumulates protochlorophyllide in the dark. This is a consequence of a regulatory loop involving the FLU protein, in which tetrapyrrole intermediates inhibit precursor conversion (Wettstein et al., 1974; Meskauskiene et al., 2001). When *flu* mutants that had been maintained in the dark are exposed to light, the protochlorophyllide that accumulates in these strains becomes photosensitized and interacts with O_2 to form $^1O_2$, which in turn can cause bleaching and inhibit growth. These responses are similar to oxidative stress responses that are observed during pathogen attack, wounding and exposure to high light intensity. Suppressors of the *flu* mutant phenotypes were found in genes encoding two thylakoid polypeptides, EXECUTER 1 and 2 (Wagner et al., 2004a; Lee et al., 2007). Furthermore, the EXECUTER proteins are required for the induction of approximately 30 nucleus-encoded chloroplast genes; the functions of the proteins encoded by these genes and their role in oxidative stress responses require further examination.

3. Redox Signals

Redox signals are generated on the acceptor side of PSI through the action of thioredoxins, Fd, and NADPH, and on the acceptor side of PSII through the action of PQ and the cytochrome b_6f complex. In an early study, the redox state of the intersystem electron transport chain in *Dunaliella* (Escoubas et al., 1995) was shown to impact *LCHB* mRNA levels, and evidence has accumulated that the cellular redox status is important for expression of many genes encoding photosynthesis-related proteins (Eberhard et al., 2008; Foyer

et al., 2009; Lawlor, 2009). The cytochrome b_6f complex appears to be critical for redox control of genes involved in Chl biosynthesis. In Chlamydomonas it has been linked to light-induced accumulation of transcripts from *HEMA*, *GSA*, *ALAD*, *CPXI*, *HLD*, *CHLH1*, *CHLI1*, *CTH1*; these transcripts no longer accumulate upon illumination of cytochrome b_6f mutants (Shao et al., 2006). At this point, we know little about the specific elements that may link the redox state of chloroplasts to changes in nuclear gene activities. The activity of the cytochrome b_6f complex through quinol binding to the complex appears to impact the activity of the STT7 and STN7 protein kinases (required for state transition, as discussed above) of Chlamydomonas and Arabidopsis, respectively. The activities of these kinases may be modulated by thiol/disulfide redox couples that depend on stromal redox conditions and the binding of quinol to the Q_o site of cytochrome b_6f (Zito et al., 1999). In addition to state transitions, the STT7/STN7 kinases may participate in signaling, either directly or indirectly, that impacts expression of various genes encoding proteins critical for photosynthesis. This possibility is suggested by the finding that the *stn7* mutant of Arabidopsis is aberrant for acclimation to changes in light conditions (Bonardi et al., 2005), suggesting that the inability of the mutant to perform a state transition, which would alter the redox state of the PQ pool, may critically affect regulation of nuclear genes, and/or the STN7 kinase may have a role in signaling independent of state transition effects. A related light/redox-regulated kinase in Arabidopsis, STN8, appears to specifically phosphorylate CP43 and PSBH proteins of PSII (Vainonen et al., 2005), although the exact regulatory function association with the protein modification is not clear.

While the above discussion illustrates tremendous progress, key biological and technical frontiers remain. We know little about photosynthetic protein maturation and degradation, and the various ways in which transcription and translation are controlled. We also know very little about the mechanisms that control the stoichiometries of photosynthetic complexes and how photosynthetic structure and function are altered in response to environmental change (e.g., macronutrient, micronutrient deprivation, drought, elevated temperature). With the development of technologies that allow for rapid sequencing of genomes, in-depth analyses of transcriptomes and informatic comparisons among genomes, it will be extremely beneficial to exploit genomic tools to generate information for elucidating aspects of chloroplast function and regulation that have not yet been explored. Chlamydomonas is one of the most versatile organisms for this purpose. It is our feeling that genome-wide approaches will also catapult Chlamydomonas into a new era as a model organism for studying processes such as cell motility and flagellar structure, function and evolution, photoperception, vesicle trafficking, mitochondrial biochemistry, numerous acclimation responses, transport processes, cell wall structure, and other aspects of cell physiology.

III. In the Direction of Genomics

A. Introduction

The explosion of genomic information over the last decade is being used to help researchers understand the gene content of organisms, how gene content and expression patterns explain ecological differences among related organisms, the ways in which gene content has been arranged and modified by evolutionary processes, the transfer of genes between organisms, and mechanisms critical for modulating gene expression in response to developmental processes and environmental fluctuation. With the massive influx of genomic information and comparative genomic tools, it is becoming clear just how much is not understood about many biological processes, including those that are integral to global productivity, biogeochemical cycling, the structure and composition of ecological habitats and the ways in which biological processes impact the geochemistry and geophysics of the Earth. Many researchers have begun to mine algal and cyanobacterial genomic information (Rocap et al., 2003; Armbrust et al., 2004; Matsuzaki et al., 2004; Barbier et al., 2005; Misumi et al., 2005; Mulkidjanian et al., 2006; Palenik et al., 2007; Bowler et al., 2008; Vardi et al., 2008; Maheswari et al., 2009; Worden et al., 2009), with many algal whole genome sequences, such as those of *Chlorella* NC64A, *Aureococcus anophagefferens*, *Emiliania huxleyi* CCMP1516, *Bathycoccus* sp. (BAN7), *Chondrus crispus*,

Ectocarpus siliculosus, *Micromonas pusilla* CCMP1545, *Micromonas strain* RCC299, *Selaginella moellendorffii*, *Volvox carteri*, either just completed or in the process of being completed (http://genome.jgi-psf.org/euk_cur1.html;http://www.genoscope.cns.fr/spip/spip.php?lang=en). This list is rapidly expanding.

We and other researchers have been exploring the genomics of Chlamydomonas (Grossman et al., 2003, 2007; Gutman and Niyogi, 2004; Ledford et al., 2004, 2007; Merchant et al., 2007) in the context of a number of other algae, photosynthetic microbes and plants. This is complemented by strong proteomic programs (Stauber and Hippler, 2004; Wagner et al., 2004b; Keller et al., 2005; Schmidt et al., 2006; Naumann et al., 2007; Rolland et al., 2009; Wagner and Mittag, 2009) and integrative systems databases (May et al., 2008, 2009). Much of our attention has been focused on mechanisms of photosynthetic electron transport and its regulation, identification of specific genes/proteins associated with functional and regulatory aspects of photosynthesis, with special emphasis on the acclimation of the photosynthetic apparatus to environmental change. We are now in a position to exploit the genomic information generated from both photosynthetic and nonphotosynthetic organisms to unveil aspects of photosynthesis that have not yet been identified, and the functions of numerous proteins that have been linked to photosynthesis (e.g., through comparative genomic tools, subcellular localization analyses, expression studies) with no assigned, specific function.

B. The Initial Sequence

Having decided that it was critical to promote the use of Chlamydomonas genomics, it was important to determine which strains would be used for the various genomic analyses. The Chlamydomonas genome sequence was generated by the Joint Genome Institute (JGI) from CC-503 *cw92 mt+*, a cell-wall deficient strain for which a BAC library is available from the Clemson University Genomics Institute (https://www.genome.clemson.edu/cgi-bin/orders?page=serviceSearch&service=bacrc&libtype=BAC&action=search). Most of the EST libraries sequenced at Stanford were prepared from RNA isolated from strain CC-1690 21 gr *mt+*, while the cDNA libraries analyzed in Japan were constructed from C-9 *mt*$^{-}$. Both of these strains are related to CC-503 since they were derived from the same field isolate collected in Massachusetts in 1945; the strains have been separated since the mid-1950s. The mating partner used for mapping of mutant alleles in Chlamydomonas is designated S1D2, which is a field isolate (collected in Minnesota in the 1980s) for which significant EST information has been generated. The EST sequences from S1D2 have been used to generate physical markers for fine scale map-based cloning of mutant alleles.

There have been a number of versions of the Chlamydomonas genome assembled by JGI, with the most recent, version 4.0, released in March 2009 (http://genome.jgi-psf.org/Chlre4/Chlre4.home.html). For the initial genome analysis (Merchant et al., 2007), the version 3.0 assembly was used. This assembly represents ~13X coverage of the genome, which is ~125 Mb. The number of predicted gene models, determined using *ab initio* and homology-based predictions, is 15,143. Furthermore, certain genes are represented by extensive gene families, including the adenylyl/guanylyl cyclases, and a number of regions of the genome exhibited small duplications and tandem gene repeats.

The current version 4.0 assembly is comprised of 88 scaffolds with 112 Mb of genomic sequence information. The genome scaffolds have numerous gaps that comprise ~7.5% of the total scaffold lengths. After filtering the dataset for cDNA/EST and homology-based support, manual annotation of many genes was completed, and the genome now has a total of 16,709 gene models. Based on these models, the average gene has 3,895 nucleotides, the average transcript has 1,768 nucleotides and the average exon and intron are 240 and 336 nucleotides, respectively. However, there are still a number of models that are not correct, and it will take more extensive transcriptome analyses to establish an accurate reference model set. Therefore, the numbers reported above will not remain static.

Based on targeted studies to generate and characterize Chlamydomonas EST (expressed sequence tag) libraries, over 300,000 ESTs were generated from cDNAs constructed from RNA isolated from cultures of Chlamydomonas exposed to diverse physiological conditions (Asamizu et al., 1999, 2000; Shrager et al., 2003; Jain et al., 2007). Although the use of normalized libraries increased the representation of transcripts in the EST data-

base, the existing dataset covers a little over half of the predicted protein-coding gene models, and only about half of those represent full-length (or very near to full length) mRNAs. Accordingly, despite the importance of Chlamydomonas as a reference for the study of photosynthesis and energy metabolism, only ~25% of the protein-coding gene models are accurately computed and verified by transcript maps. Comparisons to the genome of the close-relative Volvox on the Vista track of the JGI browser and to available cDNA information suggest that many Chlamydomonas models have under-predicted exons; this is true for both computed *ab initio* and solely homology-based models. For models in the latter category, domains that are not as highly conserved and the 5′ and 3′ UTRs may be completely or partially missing. Since in-depth sequencing of cDNA libraries may still not identify genes encoding low abundance transcripts, and maximizing sequence information from the libraries is neither time-efficient nor cost-effective, it is best to use next-generation sequence technologies to generate and correct gene models, achieving a near-complete transcriptome (see below). Projects based on next generation technologies have already begun, with the goal of examining differential gene regulation when Chlamydomonas is exposed to different environmental conditions: anoxia, deficiency in Cu, Fe, Zn, sulfur and phosphorus, treatment with Rose Bengal, and culture to various densities (0.5, 1, 2, 4 and 8 × 10^6 cells/ml) (Merchant, Pellegrini, Grossman, Niyogi, Ghirardi, unpublished). Over 80% of the reads (35 nt each) generated through RNA-seq can be uniquely positioned on the genome when two mismatches are allowed, and ~90% can be placed when two mismatches and two or more genome hits are permitted (e.g., gene families, repeated regions of the genome). The Illumina platform presently yields individual read lengths of ~125 nucleotides. The reads can be mapped to the version 4.0 genome and displayed on the genome browser; the set of reads from each experiment could be given its own track. The data are critical for correcting gene model predictions made using the *ab initio* procedures developed by JGI. Furthermore, alignments of unplaced reads across exon-exon junctions and procedures for generating paired-end sequence are facilitating the proper placement of intron-exon boundaries and also identifying new exons and alternative transcript spliced forms (Pellegrini and Merchant, unpublished). The corrected models generated from the RNA-seq information will be used to repopulate the models currently displayed on the JGI genome browser. In addition, the JGI is generating additional cDNA reads on the 454 and Illumina platforms. RNAs are pooled from diverse conditions (along the vegetative cell cycle, the mating process, re-synthesis of flagella, heterotrophic, phototrophic growth, macro- and micro-nutrient deprivation). The goal is to increase coverage of regulated genes whose transcripts might be low abundance in most growth conditions.

C. Mining the Genomic Sequence Using Informatic Approaches

In initial comparative analyses, all of the predicted Chlamydomonas proteins were analyzed by BLAST against both the Arabidopsis and human protein sequences and the best-hit scores for each Chlamydomonas protein relative to the two genomes were plotted. As shown in Fig. 6.1, there were subsets of Chlamydomonas proteins that were more similar to those of Arabidopsis, while others were more similar to those of humans. For example, thylakoid and stromal proteins were significantly more similar to proteins in Arabidopsis than to proteins in humans, while the flagellar proteins were more similar to polypeptides in humans than Arabidopsis. These results suggest that there are processes such as photosynthesis that have been preserved in Chlamydomonas and Arabidopsis (Streptophyte lineage, land plants) but not in humans (animal lineage), while other processes, such as those associated with the structure and function of flagella, have been preserved in Chlamydomonas and humans but not in land plants. Chlamydomonas has numerous genes that were lost in the Streptophyte lineage (evolved into land plants), but maintained in animal lineages (Merchant et al., 2007).

A more systematic analysis was performed to capture genes encoding proteins associated with the flagella and basal bodies, features lost during the evolution of the plant lineage, and genes for proteins specific for the green lineage and potentially targeted to chloroplasts and involved in photosynthetic function. First, families of proteins that included at least one Chlamydomonas protein and homologous proteins from a group

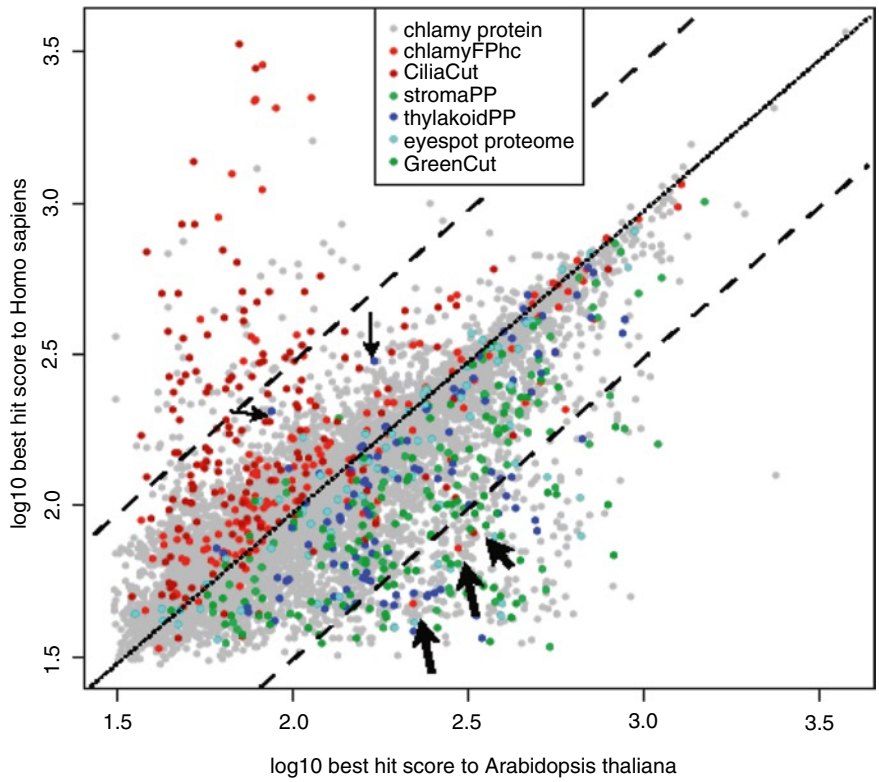


Fig. 6.1. Scatter plot of best BLASTP hit score of *Chlamydomonas* proteins to *Arabidopsis* proteins versus best BLASTP hit score of *Chlamydomonas* proteins to human proteins. Functional or genomic groupings are colored (see inset key): *Chlamydomonas* flagellar proteome high confidence set (chlamyFPhc); CiliaCut; *Arabidopsis* stroma plastid proteome (stromaPP); *Arabidopsis* thylakoid plastid proteome (thylakoidPP); eyespot proteome; GreenCut; remaining proteins are gray. The thin arrows mark some thylakoid-associated Chlamydomonas proteins (blue dots) that match proteins in Homo sapiens better than proteins in Arabidopsis thaliana while the thick arrows show some flagella-associated Chlamydomonas proteins (red dots) that match proteins in Arabidopsis thaliana better than proteins in Homo sapiens (This figure and the legend are taken from Merchant et al. (2007)).

of both photosynthetic and nonphotosynthetic organisms were constructed, as described in the supplemental material of Merchant et al. (2007). Briefly, all protein sequences in Chlamydomonas were compared by BLAST to all protein sequences in the red alga *Cyanidioschyzon,* the green algae *Ostreococcus tauri* and *O. lucimarinus*, the land plants *Arabidopsis thaliana* and *Physcomitrella patens,* the cyanobacteria *Synechocystis* sp. strain PCC6803 and *Prochlorococcus marinus* strain MIT9313*,* the bacteria *Pseudomonas aeruginosa*, the Archaea *Methanosarcina acetivorans* strain C2A and *Sulfolobus solfataricus* strain P2, the oomycetes *Phytophthora ramorum* and *P. sojae*, the diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*, the amoeba *Dictyostelium discoideum*, the fungus *Neurospora crassa*, and the metazoans human and *Caenorhabditis elegans*. Initially all possible orthologous pairs of proteins, in which one was a Chlamydomonas protein, were generated; orthologous proteins were defined as those in which there was a mutual best BLAST hit. Paralogs were then added to each of the families, based on specific evolutionary distance considerations (Merchant et al., 2007). Finally, all families of orthologs with the accompanying paralogs were merged if they contained the same Chlamydomonas protein, which resulted in the generation of 6,968 families, each containing one or more Chlamydomonas paralog(s), all of the mutual best BLAST hits to proteins of other species (orthologs), and their associated paralogs. These families were used for the comparative analyses that allowed for identification of polypeptides potentially associated with cilia ('CiliaCut') and chloroplast ('GreenCut') functions based on identification of specific sets of organisms that

contained (or did not contain) members from a specific protein family. Analysis of the CiliaCut proteins led to the identification of proteins associated with motile as well as nonmotile sensory cilia; many of the identified proteins had not been previously assigned a function, and had not been previously associated with cilia/flagella.

D. The GreenCut

The establishment of the GreenCut is most relevant to this article since it identifies genes/proteins likely to be associated with chloroplasts and potentially with photosynthetic function. Having constructed families of putative homologous proteins, which were centered on Chlamydomonas proteins, comparative genomic analyses were used to identify protein families for which all members were in the green lineage of the Plantae, which included Chlamydomonas, the prasinophyte algae *Ostreococcus* spp., the angiosperm *Arabidopsis*, and the bryophyte *Physcomitrella* (http://www.jgi.doe.gov/sequencing/why/CSP2005/physcomitrella.html), but were not present in the genomes of nonphotosynthetic eukaryotes and prokaryotes.

1. Categories Within the GreenCut

The set of 349 polypeptides of Chlamydomonas that met these criteria were considered members of the GreenCut. The proteins of the GreenCut could be further divided based on whether or not they are present in diatoms and cyanobacteria; about 90 of the 349 GreenCut proteins are present in cyanobacteria. Furthermore, at the time that the GreenCut was generated, 135 of the proteins had a known or inferred (i.e., physiological function based on comparisons with proteins from other organisms) function while 214 were unknowns. Of course, as more functional information on proteins is generated, more of the unknowns will be placed into the known or inferred categories. In the 2 years since the initial analysis, functions have been demonstrated or suggested for over 20 unknown proteins. Proteins encoded by large gene families were underrepresented because gene duplication and divergence of individuals within such families make it difficult to make precise orthology/paralogy assignments (there may be no mutual best BLAST hit) and the situation may also cause the introduction of non-homologous proteins into some of the protein families. This problem was particularly apparent for the LHC proteins: only 2 of the nearly 20 Chlamydomonas LHC proteins were retrieved in the analysis.

2. Estimate of False Negative and Positive Frequencies

Because of the conservative nature of the algorithm, some proteins that should be included in the green lineage might be lost. To examine this possibility, we generated a list of specific proteins with known functions in photosynthesis; these presumably green-lineage-specific proteins should be captured in the GreenCut. This analysis did not include those chloroplast-encoded proteins or polypeptides of the photosynthetic apparatus that exhibit sequence similarities to proteins that function in bacterial respiration; polypeptides in this category are the subunits of the cytochrome b_6f and ATP synthase complexes. The expected green-lineage-specific proteins used for the analysis included PSBO, P, Q, R, S, W, X, Y, PSAD, E, F, G, H, K, L, O, plastocyanin, Fd, FNR, RbcS and phosphoribulokinase. Of these 21 proteins, 18 were present in the GreenCut, yielding a potential false negative frequency of ~14%.

To generate a statistic for potential false positives, we used all of the 135 known proteins and determined which ones function in chloroplasts and which may reside in other cellular compartments. Of the 135 polypeptides in the GreenCut with known functions, at least 85% (115/135) are predicted to be in the chloroplast based on both experimental and bioinformatics (subcellular localization prediction tools) evidence. Of the remaining 20 polypeptides, 12 were predicted to be in other subcellular compartments and for 8 there was no strong localization information. Furthermore, a number of the proteins that do not localize to the chloroplasts do function in processes that are unique to plants or that have strongly diverged in plants relative to animals and bacteria. Overall, we estimated a false positive rate of approximately 8%. The high percentage of proteins that are chloroplast localized or that have plant specific functions that are present in the GreenCut suggest that the methods used were valid, giving us confidence that 90% or more of the unknowns generated in the analyses would

likely be localized to chloroplasts and potentially involved in photosynthetic processes, while some might also have functions unique to plants. A revised version of the GreenCut which captures nearly 600 proteins has recently been published (Karpowicz et al., 2011).

3. Functionality of the Unknowns Using Informatic Approaches

Chloroplast polypeptides of known and unknown function can be categorized based on the presence of functional groups or motifs, which include pfam domains (e.g., DNA binding, RNA binding, kinase domains), co-regulation expression profiles (e.g., tissue-specific expression) and potential subcellular location based on the presence/absence of a recognizable transit peptide (this analysis is most effective using Arabidopsis orthologs), which targets the polypeptide to the chloroplast. Figure 6.2a shows the expression of three GreenCut genes in Arabidopsis. The gene for protein of unknown function CGLD14 shares an expression profile with *CYN38* and *PSBY*. *CYN38* encodes a cyclophilin that assembles and maintains the proteins in the supercomplex of

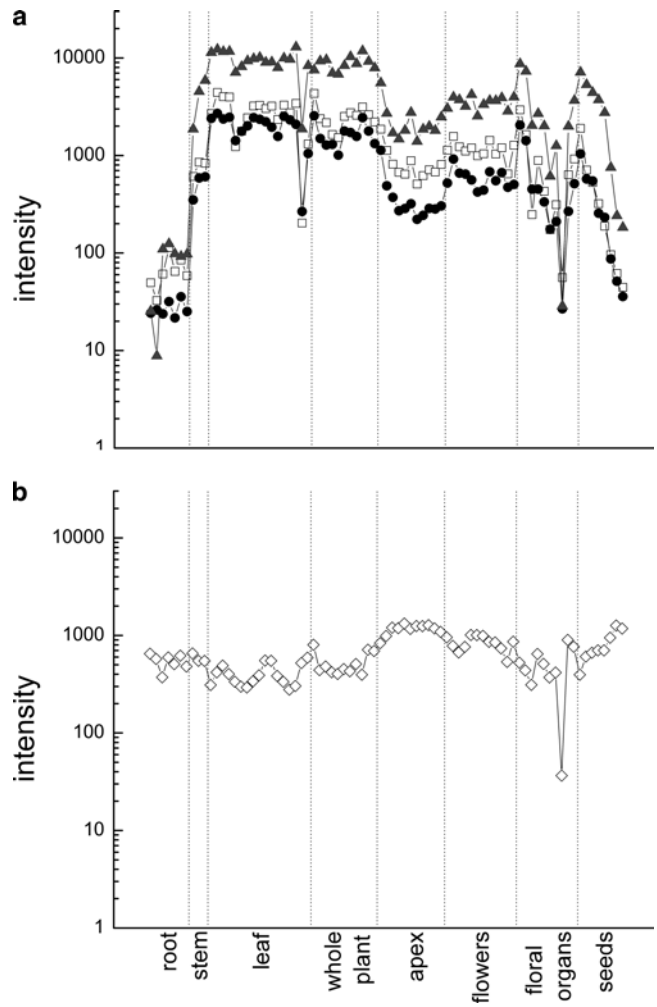


Fig. 6.2. Patterns of transcript abundance of GreenCut genes in Arabidopsis tissue. (**a**) *CGLD14* (AT1G76450; *closed circles*), a conserved expressed protein of unknown function, shares a similar expression profile to *CYN38* (AT3G01480; Pearson correlation coefficient 0.93; *open squares*) and *PSBY* (AT1G67740; Pearson 0.79; *gray triangles*), which both encode PSII-associated proteins. (**b**) *CGLD14* does not share a profile with *TOC159* (AT2G16640; Pearson −0.38; *open diamonds*), which encodes a translocon of the outer envelope membrane of chloroplasts (Microarray expression data are from Schmid et al. (2005)).

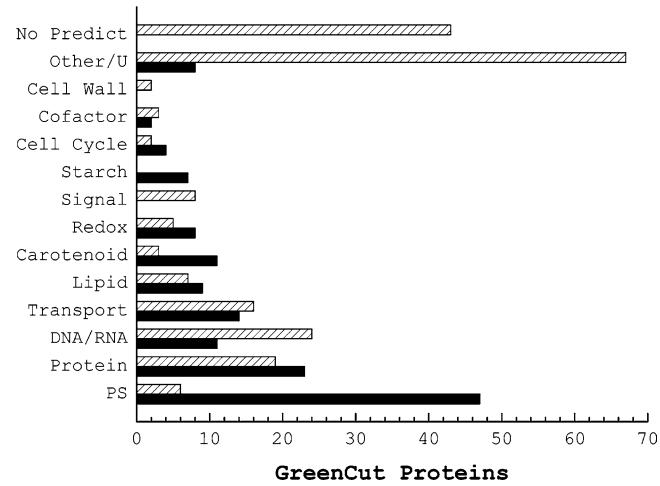


Fig. 6.3. Classification of GreenCut proteins into functional groups. MAPMAN (Thimm et al., 2004) functional descriptors and pfam domains were retrieved for the Arabidopsis GreenCut homologs and used to group proteins into major functional categories. Unknown proteins containing a predicted domain but lacking an informative MAPMAN descriptor were manually added to an appropriate category. *No predict* no functional prediction, *Other/U* other or undefined domain function, *Cell Wall* cell wall biosynthesis and maintenance, *Cofactor* cofactor biosynthesis, *Cell Cycle* cell division and regulation, *Starch* starch biosynthesis and degradation, *Signal* signal transduction, *Redox* electron transport, *Carotenoid* carotenoid and tetrapyrrole biosynthesis, *Lipid* lipid biosynthesis; *Transport* protein or small molecule trafficking; *DNA/RNA* nucleic acids and transcription, *Protein* protein maturation and degradation, *PS* photosynthesis. Proteins of known function are represented by *solid bars*; those of unknown function by *hatched bars*.

PSII (Fu et al., 2007). The product of *PSBY* is a poorly understood thylakoid membrane protein that associates with PSII (Gau et al., 1998). This shared profile is typical of proteins associated with photosynthetic function; there is high expression in green organs like stems and leaves with little expression in the root and floral organs. In contrast, *TOC159*, whose protein is a translocon of the outer envelope membrane of the chloroplast, is expressed equally in all Arabidopsis tissues, as shown in Fig. 6.2b. We would suggest that the protein of unknown function likely has a role in the photosynthetic apparatus. Based on the above analyses, a number of the unknowns can be placed in functional categories. As shown in the functional categorizations presented in Fig. 6.3, the specific activities of most proteins associated with photosynthetic function are known, with only six having a function that has not been clearly defined. In contrast, all eight of the proteins associated with signaling have an unknown function while the majority of proteins associated with DNA- and RNA-dependent processes, including transcription and transcript stability/maturation, have no known physiological function. Furthermore, there are nearly 50 proteins that have no potential functionality assigned; these include proteins that contain no characterized motif, or that have a motif that still does not provide enough information to decide on the potential protein function.

4. Analysis of Photosynthetic Phenotypes in Mutants Associated with the Unknowns of the GreenCut

Thus far, approximately 40 strains of Chlamydomonas (Dent et al., 2005) and well over 100 strains of Arabidopsis (Bailey, Hamel and Grossman, unpublished) have been identified with insertions in genes encoding GreenCut proteins of unknown function. Analyses of both sets of mutants have been initiated and a number have a distinct mutant phenotype. To visualize this, we have included a fluorescence analysis (F_v/F_m) of Chlamydomonas mutants in GreenCut genes growing as colonies on plates, which is shown in Fig. 6.4. The most obvious mutants are blue colonies, which represents a false color image based on the F_v/F_m value, which essentially lack F_v and thereby presumably PSII activity. The reasons for the apparent lack of PSII activity in these two photosynthetic mutants have not yet

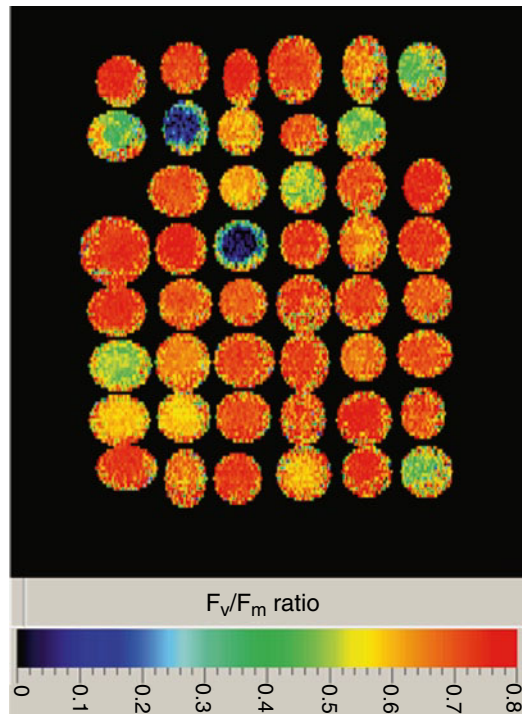


Fig. 6.4. Fluorescence imaging of mutants in genes in the GreenCut of Chlamydomonas. The fluorescence images represent the magnitude of the F_v/F_m (the degree to which the PSII traps are open after a 5–10 min dark incubation); blue represents strains which have little PSII activity while red represents functional traps that are fully open in the dark. The two most apparent mutants appear as blue colonies, but a number of the other strains (*light green*) also show aberrant photosynthesis. The mutants were kindly provided by Rachel Dent (Dent et al., 2005).

been clarified. One of the mutants has an insertion in the GreenCut gene CGL39 while the other has an insertion in CGL28 (see Merchant et al., 2007). Interestingly, while CGL39 has no motif that provides clues concerning its function, CGL28 has an RNA binding motif. This protein may be involved in the maturation or stabilization of transcripts associated with the function and/or the biogenesis of PSII. A number of the other colonies shown on this plate have lower F_v/F_m values, as indicated by their light green color on the false color map in Fig. 6.4. Most of these appear to have reduced electron flow through PSII, PSI or between the two photosystems. Some of the mutants appear to be defective for all aspects of photosynthetic electron transport and may represent strains unable to acclimate to changes in the environment (so the photosynthetic apparatus remains in a single, default acclimation state); this is currently being tested. Based on motifs, the proteins eliminated in some of these mutants may function in lipid metabolism, the transport of proteins into the nucleus and the maturation of transcripts. The functions of two of the mutants presented in this analysis (CCS1 and CCB4) have recently been identified and were shown to be involved in assembly of the covalent heme on the cytochrome b_6 polypeptide of the electron transport chain (Kuras et al., 2007). More genetic/molecular information is required that clearly links the insertion of the marker gene (which is inserted into the GreenCut gene) to the mutant phenotype.

IV. Concluding Remarks

While studies over the last half century have identified many of the proteins with specific catalytic activities critical for photosynthetic electron transport and CO_2 fixation, the photosynthetic apparatus – also as a function of the developmental stage of the organism – is exquisitely sensitive to environmental change, and this responsivity may

encompass multiple tiers of regulation. Fluctuating environmental conditions can rapidly alter the activities of various photosynthetic complexes by directly modulating the activity of the complex, and by controlling both the levels and types of components within the complex. Short-term responses may reflect changes in protonation, phosphorylation and the association of various pigment and protein components of the complexes. Longer-term responses may reflect changes in subunit stoichiometries and insertion of novel proteins into complexes. With the development of a broad genomic view coupled with results from high-throughput and strong molecular and genetic technologies, we are now in a position to unravel new aspects of photosynthetic control and the impact of such control on the overall physiological and metabolic processes in the cell.

Acknowledgments

Much of the genome analysis discussed in this manuscript could not have been performed without the help of Dan Rokhsar and Simon Prochnik, who are responsible for the Chlamydomonas genome project at DOE's Joint Genome Institute. Aspects of the work presented were also supported by NSF grants MCB0235878, MCB0951094 and Department of energy grant DE-FG0207ER64427 awarded to ARG and by the Marie Curie OIF-6 APOSD grant (EU). SM acknowledges support from the Department of Energy (DE-FC03-02ER63421 and DE-FG02-07ER15229). SJK was supported in part by a Ruth L. Kirschstein National Research Service Award GM07185.

References

Adam Z and Clarke AK (2002) Cutting edge of chloroplast proteolysis. Trends Plant Sci 7: 451–456

Adam Z, Rudella A and van Wijk KJ (2006) Recent advances in the study of Clp, FtsH and other proteases located in chloroplasts. Curr Opin Plant Biol 9: 234–240

Adir N and Ohad I (1988) Structural properties of the D1 and surrounding photosystem II polypeptides as revealed by their interaction with cross-linking reagents. J Biol Chem 263: 283–289

Albertsson P (2001) A quantitative model of the domain structure of the photosynthetic membrane. Trends Plant Sci 6: 349–358

Allen JF (1992) Protein phosphorylation in regulation of photosynthesis. Biochim Biophys Acta 1098: 275–335

Allen MD, Kropat J and Merchant SS (2008) Regulation and localization of isoforms of the aerobic oxidative cyclase in *Chlamydomonas reinhardtii*. Photochem Photobiol 84: 1336–1342

Anderson JM, Chow WS and Park YI (1995) The grand design of photosynthesis: Acclimation of the photosynthetic apparatus to environmental cues. Photosyn Res 46: 129–139

Andersson B and Anderson JM (1980) Lateral heterogeneity in the distribution of chlorophyll-protein complexes of the thylakoid membranes of spinach chloroplasts. Biochim Biophys Acta 593: 427–440

Apel K and Hirt H (2004) Reactive oxygen species: metabolism, oxidative stress, and signal transduction. Annu Rev Plant Biol 55: 373–399

Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH et al. (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science 306: 79–86

Arnon DI (1959) Conversion of light into chemical energy in photosynthesis. Nature 184: 10–21

Aro EM, McCaffrey S and Anderson JM (1993a) Photoinhibition and D1 protein degradation in peas acclimated to different growth irradiances. Plant Physiol 103: 835–843

Aro E-M, Virgin I and Andersson B (1993b) Photoinhibition of photosystem II. Inactivation, protein damage and turnover. Biochim Biophys Acta 1143: 113–134

Aro EM, Suorsa M, Rokka A, Allahverdiyeva Y, Paakkarinen V, Saleem A et al. (2005) Dynamics of photosystem II: a proteomic approach to thylakoid protein complexes. J Exp Bot 56: 347–356

Asamizu E, Nakamura Y, Sato S, Fukuzawa H and Tabata S (1999) A large scale structural analysis of cDNAs in a unicellular green alga *Chlamydomonas reinhardtii*. Generation of 3,433 non-redundant expressed sequence tags. DNA Res 6: 369–373

Asamizu E, Miura K, Kucho K, Inoue Y, Fukuzawa H, Ohyama K et al. (2000) Generation of expressed sequence tags from low-CO_2 and high-CO_2 adapted cells of *Chlamydomonas reinhardtii*. DNA Res 7: 305–307

Bailey S and Grossman A (2008) Photoprotection in cyanobacteria: regulation of light harvesting. Photochem Photobiol 84: 1410–1420

Bailey S, Walters RG, Jansson S and Horton P (2001) Acclimation of Arabidopsis thaliana to the light environment: the existence of separate low light and high light responses. Planta 213: 794–801

Bailey S, Thompson E, Nixon PJ, Horton P, Mullineaux CW, Robinson C and Mann NH (2002) A critical role for the Var2 FtsH homologue of *Arabidopsis thaliana* in the photosystem II repair cycle in vivo. J Biol Chem 277: 2006–2011

Bailey S, Melis A, Mackey KR, Cardol P, Finazzi G, van Dijken G et al. (2008) Alternative photosynthetic electron flow to oxygen in marine *Synechococcus*. Biochim Biophys Acta 1777: 269–276

Ball S and Deschamps P (2009) Starch metabolism. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 1–40. Elsevier, Amsterdam

Ballottari M, Dall'Osto L, Morosinotto T and Bassi R (2007) Contrasting behavior of higher plant photosystem I and II antenna systems during acclimation. J Biol Chem 282: 8947–8958

Barber J and Andersson B (1992) Too much of a good thing: light can be bad for photosynthesis. Trends Biochem Sci 17: 61–66

Barbier G, Oesterhelt C, Larson MD, Halgren RG, Wilkerson C, Garavito RM et al. (2005) Comparative genomics of two closely related unicellular thermoacidophilic redalgae, *Galdieria sulphuraria* and *Cyanidioschyzon merolae*, reveals the molecular basis of the metabolic flexibility of *Galdieria sulphuraria* and significant differences in carbohydrate metabolism of both algae. Plant Physiol 137: 460–474

Barneche F, Winter V, Crevecoeur M and Rochaix JD (2006) ATAB2 is a novel factor in the signaling pathway of light-controlled synthesis of photosystem proteins. Embo J 25: 5907–5918

Bartsch S, Monnet J, Selbach K, Quigley F, Gray J, von Wettstein D et al. (2008) Three thioredoxin targets in the inner envelope membrane of chloroplasts function in protein import and chlorophyll metabolism. Proc Natl Acad Sci USA 105: 4933–4938

Battchikova N and Aro EM (2007) Cyanobacterial NDH-1 complexes: multiplicity in function and subunit composition. Physiol Plant 131: 22–32

Beale SI (2009) Biosynthesis of chlorophylls and hemes. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 731–798. Elsevier, Amsterdam

Beck CF (2005) Signaling pathways from the chloroplast to the nucleus. Planta 222: 743–756

Bellafiore S, Barneche F, Peltier G and Rochaix JD (2005) State transitions and light adaptation require chloroplast thylakoid protein kinase STN7. Nature 433: 892–895

Bennett J, Steinback KE and Arntzen CJ (1980) Chloroplast phosphoproteins: regulation of excitation energy transfer by phosphorylation of thylakoid membrane polypeptides. Proc Natl Acad Sci USA 77: 5253–5257

Bennoun P and Levine RP (1967) Detecting mutants that have impaired photosynthesis by their increased level of fluorescence. Plant Physiol 42: 1284–1287

Bennoun P and Delepelaire P (1982) Isolation of photosynthesis mutants in *Chlamydomonas*. In: Edelman M, Chua N-H, Hallick RB (eds) Methods in Chloroplast Molecular Biology, pp 25–38. Elsevier Biomedical Press, Amsterdam

Bennoun P and Beal D (1997) Screening algal mutant colonies with altered thylakoid electrochemical gradient through fluorescence and delayed luminescence digital imaging. Photosyn Res 51: 161–165

Bennoun P, Spierer-Herz M, Erickson J, Girard-Bascou J, Pierre Y, Delosme M and Rochaix JD (1986) Characterization of photosystem II mutants of *Chlamydomonas reinhardtii* lacking *psbA* gene. Plant Mol Biol 6: 151–160

Bernd KK and Kohorn BD (1998) Tip loci: Six Chlamydomonas nuclear suppressors that permit the translocation of proteins with mutant thylakoid signal sequences. Genetics 149: 1293–1301

Bibby TS, Nield J, Partensky F and Barber J (2001) Oxyphotobacteria. Antenna ring around photosystem I. Nature 413: 590

Bloodgood RA (2009) The Chlamydomonas flagellar membrane and its dynamic properties. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 309–368. Elsevier, Amsterdam

Bollenbach TJ, Schuster G and Stern DB (2004) Cooperation of endo- and exoribonucleases in chloroplast mRNA turnover. Prog Nucleic Acid Res Mol Biol 78: 305–337

Bollenbach TJ, Sharwood RE, Gutierrez R, Lerbs-Mache S and Stern DB (2009) The RNA-binding proteins CSP41a and CSP41b may regulate transcription and translation of chloroplast-encoded RNAs in Arabidopsis. Plant Mol Biol 69: 541–552

Bonardi V, Pesaresi P, Becker T, Schleiff E, Wagner R, Pfannschmidt T et al. (2005) Photosystem II core phosphorylation and photosynthetic acclimation require two different protein kinases. Nature 437: 1179–1182

Bonaventura C and Meyers J (1969) Fluorescence and oxygen evolution from *Chlorella pyrenoidosa*. Biochim Biophys Acta 189: 366–383

Boudreau E, Takahashi Y, Lemieux C, Turmel M and Rochaix J-D (1997) The chloroplast *ycf3* and *ycf4* open reading frames of *Chlamydomonas reinhardtii* are required for the accumulation of the photosystem I complex. EMBO J 16: 6095–6104

Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A et al. (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. Nature 456: 239–244

Caffarri S, Frigerio S, Olivieri E, Righetti PG and Bassi R (2005) Differential accumulation of Lhcb gene products in thylakoid membranes of *Zea mays* plants grown under contrasting light and temperature conditions. Proteomics 5: 758–768

Camargo A, Llamas A, Schnell RA, Higuera JJ, Gonzalez-Ballester D, Lefebvre PA et al. (2007) Nitrate signaling by the regulatory gene NIT2 in Chlamydomonas. Plant Cell 19: 3491–3503

Cardol P and Remacle C (2009) The mitochondrial genome. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 445–468. Elsevier, Amsterdam

Cardol P, Bailleul B, Rappaport F, Derelle E, Béal D, Breyton C, Bailey S, Wollman FA, Grossman A, Moreau H and Finazzi G. (2008) An original adaptation of

photosynthesis in the marine green alga Ostreococcus. Proc Natl Acad Sci USA 105: 7881–7886

Castruita M, Casero D, Karpowicz SJ, Kropat J, Vieler A, Hsieh SI, Yan W, Cokus S, Loo JA, Benning C, Pellegrini M and Merchant SS (2010) Systems biology approach in *Chlamydomonas* reveals connections between copper nutrition and multiple metabolic steps. Plant Cell. 2011 23: 1273–1292

Chekounova E, Voronetskaya V, Papenbrock J, Grimm B and Beck CF (2001) Characterization of *Chlamydomonas* mutants defective in the H subunit of Mg-chelatase. Mol Gen Genom 266: 363–373

Chen M, Bibby TS, Nield J, Larkum A and Barber J (2005) Iron deficiency induces a chlorophyll d-binding Pcb antenna system around Photosystem I in *Acaryochloris marina*. Biochim Biophys Acta 1708: 367–374

Chen Z, Chastain CJ, Al-Abed SR, Chollet R and Spreitzer RJ (1988) Reduced CO_2/O_2 specificity of ribulose-1, 5-biphosphate carboxylase/oxygenase in a temperature-sensitive chloroplast mutant of *Chlamydomonas reinhardtii*. Proc Natl Acad Sci USA 85: 4696–4699

Cheregi O, Sicora C, Kos PB, Barker M, Nixon PJ and Vass I (2007) The role of the FtsH and Deg proteases in the repair of UV-B radiation-damaged photosystem II in the cyanobacterium *Synechocystis* PCC 6803. Biochim Biophys Acta 1767: 820–828

Choquet Y and Vallon O (2000) Synthesis, assembly and degradation of thylakoid membrane proteins. Biochimie 82: 615–634

Choquet Y and Wollman FA (2002) Translational regulations as specific traits of chloroplast gene expression. FEBS Lett 529: 39–42

Choquet Y and Wollman FA (2009) The CES process. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 1027-1063. Elsevier, Amsterdam

Choquet Y, Zito F, Wostrikoff K and Wollman FA (2003) Cytochrome f translation in Chlamydomonas chloroplast is autoregulated by its carboxyl-terminal domain. Plant Cell 15: 1443–1454

Choquet Y, Stem DB, Wostrikoff K, Kuras R, Girard-Bascou J and Wollman F-A (1998) Translation of cytochrome is autoregulated through the 5'-untranslated region of *petA* mRNA in Chlamydomonas chloroplasts. Proc Natl Acad Sci USA 95: 4380–4385

Chua N-H and Bennoun P (1975) Thylakoid membrane polypeptides of *Chlamydomonas reinhardtii*: Wild-type and mutant strains deficient in photosystem II reaction center. Proc Natl Acad Sci USA 72: 2175–2179

Chua N-H and Gillham NW (1977) The sites of synthesis of the principal thylakoid membrane polypeptides in *Chlamydomonas reinhardtii.* J Cell Biol 74: 441–452

Chua NH and Schmidt GW (1978) Post-translational transport into intact chloroplasts of a precursor to the small subunit of ribulose-1,5-bisphosphate carboxylase. Proc Natl Acad Sci USA 75: 6110–6114

Cline K and Henry R (1996) Import and routing of nucleus-encoded chloroplast proteins. Annu Rev Cell Devel Biol 12: 1–26

Cognat V, Deragon JM, Vinogradova E, Salinas T, Remacle C and Marechal-Drouard L (2008) On the evolution and expression of *Chlamydomonas reinhardtii* nucleus-encoded transfer RNA genes. Genetics 179: 113–123

Cohen I, Sapir Y and Shapira M (2006) A conserved mechanism controls translation of Rubisco large subunit in different photosynthetic organisms. Plant Physiol 141: 1089–1097

Cole DG (2009) Intraflagellar transport. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 71–114. Elsevier, Amsterdam

DalCorso G, Pesaresi P, Masiero S, Aseeva E, Schunemann D, Finazzi G et al (2008) A complex containing PGRL1 and PGR5 is involved in the switch between linear and cyclic electron flow in Arabidopsis. Cell 132: 273–285

Danielsson R and Albertsson PA (2009) Fragmentation and separation analysis of the photosynthetic membrane from spinach. Biochim Biophys Acta 1787: 25–36

Danielsson R, Suorsa M, Paakkarinen V, Albertsson PA, Styring S, Aro EM and Mamedov F (2006) Dimeric and monomeric organization of photosystem II. Distribution of five distinct complexes in the different domains of the thylakoid membrane. J Biol Chem 281: 14241–14249

Dannehl H, Herbik A and Godde D (1995) Stress induced degradation of the photosynthetic apparatus is accompanied by changes in thylakoid protein turnover and phosphorylation. Physiol Plant 93: 179–186

Davies J, Yildiz F and Grossman, AR (1996) *Sac1*, a putative regulator that is critical for survival of *Chlamydomonas reinhardtii* during sulfur deprivation. Embo J 15: 2150–2159

Davies JP, Weeks DP and Grossman AR (1992) Expression of the arylsulfatase gene from the $ß_2$-tubulin promoter in *Chlamydomonas reinhardtii.* Nucleic Acids Res 20: 2959–2965

de Vitry C and Kuras R (2009) The cytochrome b_6f complex. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 603–638. Elsevier, Amsterdam

de Vitry C, Olive J, Drapier D, Recouvreur M and Wollman FA (1989) Posttranslational events leading to the assembly of photosystem II protein complex: a study using photosynthesis mutants from *Chlamydomonas reinhardtii*. J Cell Biol 109: 991–1006

Deisenhofer J, Epp O, Miki K, Huber R and Michel H (1985) Structure of the protein subunits in the photosynthetic reaction centre of *Rhodopseudomonas viridis* at 3 Å resolution. Nature 318: 618–624

Delepelaire P and Chua N-H (1979) Lithium dodecyl sulfate/polyacrylamide gel electrophoresis of thylakoid membranes at 4°C: Characterization of two additional chlorophyll *a*-protein complexes. Proc Natl Acad Sci USA 76: 111–115

Dent RM, Haglund CM, Chin BL, Kobayashi MC and Niyogi KK (2005) Functional genomics of eukaryotic photosynthesis using insertional mutagenesis of *Chlamydomonas reinhardtii*. Plant Physiol 137: 545–556

Depege N, Bellafiore S and Rochaix JD (2003) Role of chloroplast protein kinase Stt7 in LHCII phosphorylation and state transition in Chlamydomonas. Science 299: 1572–1575

Desikan R, A-H-Mackerness S, Hancock JT and Neill SJ (2001) Regulation of the Arabidopsis transcriptome by oxidative stress. Plant Physiol 127: 159–172

Dieckmann CL (2003) Eyespot placement and assembly in the green alga Chlamydomonas. Bioessays 25: 410–416

Dobberstein B, Blobel G and Chua N-H (1977) *In vivo* synthesis and processing of a putative precursor for the small subunit of ribulose- 1,5-bisphosphate carboxylase of *Chlamydomonas reinhardtii*. J Plant Physiol 126: 467–473

Drager RG, Girard-Bascou J, Choquet Y, Kindle KL and Stern DB (1998) *In vivo* evidence for 5'–>3' exoribonuclease degradation of an unstable chloroplast mRNA. Plant J 13: 85–96

Drapier D, Rimbault B, Vallon O, Wollman FA, and Choquet Y (2007) Intertwined translational regulations set uneven stoichiometry of chloroplast ATP synthase subunits. EMBO J 26: 3581–3591

Dron M, Rahire M, Rochaix J-D, and Mets L (1983) First DNA sequence of a chloroplast mutation: A missense alteration in the ribulosebisphosphate carboxylase large subunit gene. Plasmid 9: 321–324

Dubini A, Mus F, Seibert M, Grossman AR and Posewitz MC (2009) Flexibility in anaerobic metabolism as revealed in a mutant of *Chlamydomonas reinhardtii* lacking hydrogenase activity. J Biol Chem 284: 7201–7213

Dutcher SK (1995) Flagellar assembly in two hundred and fifty easy-to-follow steps. Trends Genet 11: 398–404

Eberhard S, Jain M, Im CS, Pollock S, Shrager J, Lin Y et al (2006) Generation of an oligonucleotide array for analysis of gene expression in *Chlamydomonas reinhardtii*. Curr Genet 49: 1–19

Eberhard S, Finazzi G and Wollman FA (2008) The dynamics of photosynthesis. Annu Rev Genet 42: 463–515

Edelman M and Mattoo AK (2008) D1-protein dynamics in photosystem II: the lingering enigma. Photosynth Res 98: 609–620

Endo T, Ishida S, Ishikawa N and Sato F (2008) Chloroplastic NAD(P)H dehydrogenase complex and cyclic electron transport around photosystem I. Mol Cell 25: 158–162

Erickson J, Rahire M, Malnoe P, Girard-Bascou J, Pierre Y, Bennoun P and Rochaix JD (1986) Lack of D2 protein in a *Chlamydomonas reinhardtii psbD* mutant affects photosystem II stability and D1 expression. Embo J 5: 1745–1754

Erickson JM, Rahire M, Rennoun P, Delepelaire P, Diner B and Rochaix, J-D (1984) Herbicide resistance in *Chlamydomonas reinhardtii* results from a mutation in the chloroplast gene for the 32-kilodalton protein of photosystem II. Proc Natl Acad Sci USA 81: 3617–3621

Erickson JM, Pfister K, Rahire M, Togasaki RK, Mets L and Rochaix JD (1989) Molecular and biophysical analysis of herbicide-resistant mutants of *Chlamydomonas reinhardtii*: structure-function relationship of the photosystem II D1 polypeptide. Plant Cell 1: 361–371

Escoubas J-M, Lomas MW, LaRoche J and Falkowski PG (1995) Light intensity regulation of *cab* gene transcription is signaled by the redox state of the plastoquinone pool. Proc Natl Acad Sci USA 92: 10237–10241

Espineda CE, Linford AS, Devine D and Brusslan JA (1999) The AtCAO gene, encoding chlorophyll a oxygenase, is required for chlorophyll b synthesis in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 96: 10507–10511

Fernández E and Galván, Á (2007) Inorganic nitrogen assimilation in Chlamydomonas. J Exp Bot 58: 2279–2287

Fernández E and Galván Á (2008) Nitrate assimilation in Chlamydomonas. Eukaryot Cell 7: 555–559

Fernández E, Llamas A and Galván Á (2009) Nitrogen assimilation and its regulation. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 69–114. Elsevier, Amsterdam

Ferris PJ and Goodenough UW (1997) Mating type in *Chlamydomonas* is specified by *mid*, the *minus*-dominance gene. Genetics 146: 859–869

Ferris PJ, Woessner JP and Goodenough UW (1996) A sex recognition glycoprotein is encoded by the *plus* mating-type gene *fusl* of *Chlamydomonas reinhardtii*. Mol Biol Cell 7: 1235–1248

Ferris PJ, Pavolvic C, Fabrey S and Goodenough UW (1997) Rapid evolution of sex-related genes in *Chlamydomonas.* Proc Natl Acad Sci USA 94: 8634–8639

Fey V, Wagner R, Brautigam K, Wirtz M, Hell R, Dietzmann A et al. (2005) Retrograde plastid redox signals in the expression of nuclear genes for chloroplast proteins of *Arabidopsis thaliana*. J Biol Chem 280: 5318–5328

Finazzi G and Forti G (2004) Metabolic flexibility of the green Alga *Chlamydomonas reinhardtii* as revealed by the link between state transitions and cyclic electron flow. Photosynth Res 82: 327–338

Finazzi G, Drapier D and Rappaport F (2009) The CF_oF_1 ATP-synthase complex of photosynthesis. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 639–670. Elsevier, Amsterdam

Finazzi G, Furia A, Barbagallo RP and Forti G (1999) State transitions, cyclic and linear electron transport and photophosphorylation in *Chlamydomonas reinhardtii*. Biochim Biophys Acta 1413: 117–129

Finazzi G, Zito, F, Barbagallo, RP and Wollman FA (2001) Contrasted effects of inhibitors of cytochrome b_6f complex on state transitions in *Chlamydomonas reinhardtii*: the role of Qo site occupancy in LHCII kinase activation. J Biol Chem 276: 9770–9774

Finazzi G, Rappaport F, Furia A, Fleischmann M, Rochaix JD, Zito F and Forti G (2002) Involvement of state transitions in the switch between linear and cyclic electron flow in *Chlamydomonas reinhardtii*. EMBO Reports 3: 280–285

Fischer BB, Wiesendanger M and Eggen RI (2006) Growth condition-dependent sensitivity, photodamage and stress response of *Chlamydomonas reinhardtii* exposed to high light conditions. Plant Cell Physiol 47: 1135–1145

Fischer BB, Krieger-Liszkay A, Hideg E, Snyrychova I, Wiesendanger M and Eggen RI (2007) Role of singlet oxygen in chloroplast to nucleus retrograde signaling in *Chlamydomonas reinhardtii*. FEBS Lett 581: 5555–5560

Fischer N, Setif P and Rochaix J-D (1997) Targeted mutations in the *psaC* gene of *Chlamydomonas reinhardtii*: Preferential reduction of FB at low temperature is not accompanied by altered electron flow from photosystem I to ferredoxin. Biochem 36: 93–102

Fischer N, Hippler M, Setif P, Jacquot JP and Rochaix JD (1998) The PsaC subunit of photosystem I provides an essential lysine residue for fast electron transfer to ferredoxin. EMBO J 17: 849–858

Fowkes ME and Mitchell DR (1998) The role of preassembled cytoplasmic complexes in assembly of flagellar dynein subunits. Mol Biol Cell 9: 2337–2347

Foyer CH, Noctor G, Buchanan B, Dietz KJ and Pfannschmidt T (2009) Redox regulation in photosynthetic organisms: Signaling, acclimation, and practical implications. Antioxid Redox Signal 11: 861–905

Frank HA and Cogdell RJ (1996) Carotenoids in photosynthesis. Photochem Photobiol 63: 257–264

Frigerio S, Campoli C, Zorzan S, Fantoni LI, Crosatti C, Drepper F et al (2007) Photosynthetic antenna size in higher plants is controlled by the plastoquinone redox state at the post-transcriptional rather than transcriptional level. J Biol Chem 282: 29457–29469

Fu A, He Z, Cho HS., Lima A, Buchanan BB and Luan S (2007) A chloroplast cyclophilin functions in the assembly and maintenance of photosystem II in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 104: 15947–15952

Funes S, Franzen LG and Gonzalez-Halphen D (2007) *Chlamydomonas reinhardtii*: the model of choice to study mitochondria from unicellular photosynthetic organisms. Methods Mol Biol 372: 137–149

Gadjev I, Vanderauwera, S, Gechev, T.S, Laloi C, Minkov IN, Shulaev V et al. (2006) Transcriptomic footprints disclose specificity of reactive oxygen species signaling in Arabidopsis. Plant Physiol 141: 436–445

Garcia-Lorenzo M, Zelisko A, Jackowski G and Funk C (2005) Degradation of the main photosystem II light-harvesting complex. Photochem Photobiol Sci 4: 1065–1071

Gau AE, Thole HH, Sokolenko A, Altschmied L, Hermann RG and Pistorius EK (1998) PsbY, a novel manganese-binding, low-molecular-mass protein associated with photosystem II. Mol Gen Genet 260: 56–68

Ghassemian M, Lutes J, Tepperman JM, Chang HS, Zhu T, Wang X et al (2006) Integrative analysis of transcript and metabolite profiling data sets to evaluate the regulation of biochemical pathways during photomorphogenesis. Arch Biochem Biophys 448: 45–59

Gilmore AM, Hazlett TL and Govindjee (1995) Xanthophyll cycle-dependent quenching of photosystem II chlorophyll a fluorescence: Formation of a quenching complex with a short fluorescence lifetime. Proc Natl Acad Sci USA 92: 2273–2277

Girard-Bascou J, Choquet U, Gimpel NJ, Culler D, Purton S, Merchant S et al (1995) Nuclear control of the expression of the chloroplast *pet* genes in *Chlamydomonas reinhardtii*. In: Mathis P (ed) Photosynthesis: From Light to Biosphere, pp 683–686. Kluwer Academic Publishers, Dordrecht

Givan AL and Levine RP (1967) The photosynthetic electron transport chain of *Chlamydomonas reinhardtii*. VII. Photosynthetic phosphorylation by a mutant strain of *Chlamydomonas reinhardtii* deficient in active P700. Plant Physiol 42: 1264–1268

Gokhale X and Sayre RT (2009) Photosystem II, a structural perspective. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 573–602. Elsevier, Amsterdam

Goldschmidt-Clermont M (2009) Chloroplast RNA splicing. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 915–936. Elsevier, Amsterdam

Goldschmidt-Clermont M, Rahire M and Rochaix JD (2008) Redundant cis-acting determinants of 3' processing and RNA stability in the chloroplast rbcL mRNA of Chlamydomonas. Plant J 53: 566–577

González-Ballester D, Casero D, Cokus S, Pellegrini M, Merchant SS, Grossman AR (2010) *RNA-seq analysis of sulfur-deprived Chlamydomonas cells reveals aspects of acclimation critical for cell survival*. Plant Cell 22: 2058–2084

González-Ballester D, Pollock SV, Pootakham W and Grossman AR (2008) The central role of a SNRK2 kinase in sulfur deprivation responses. Plant Physiol 147: 216–227

Goodenough U, Lin H and Lee JH (2007) Sex determination in Chlamydomonas. Semin Cell Dev Biol 18: 350–361

Goodenough UW (1986) Experimental analysis of the adhesion reaction between isolated *Chlamydomonas* flagella. Exp Cell Res 166: 237–246

Goodenough UW (1992) Green Yeast. Cell 70: 533–538

Goodenough UW and Staehelin LA (1971) Structural differentiation of stacked and unstacked chloroplast membranes. Freeze-etch electron microscopy of wild-type and mutant strains of *Chlamydomonas*. J Cell Biol 48: 594–619

Goodenough UW, Armbrust EV, Campbell AM and Ferris PJ (1995) Molecular genetics of sexuality in *Chlamydomonas*. Annu Rev Plant Physiol Plant Mol Biol 46: 21–44

Gorman DS and Levine RP (1966) Photosynthetic electron transport chain of *Chlamydomonas reinharditi*. VI. Electron

transport in mutant strains lacking either cytochrome 553 or plastocyanin. Plant Physiol 41: 1648–1656

Grossman AR, Bartlett SG and Chua N-H (1980) Energy-dependent uptake of cytoplasmically-synthesized polypeptides by chloroplasts. Nature 285: 625–628

Grossman AR, Croft M, Gladyshev VN, Merchant SS, Posewitz MC, Prochnik S and Spalding MH (2007) Novel metabolism in Chlamydomonas through the lens of genomics. Curr Opin Plant Biol 10: 190–198

Grossman AR, Harris EE, Hauser C, Lefebvre PA, Martinez D, Rokhsar D et al (2003) *Chlamydomonas reinhardtii* at the crossroads of genomics. Eukaryot Cell 2: 1137–1150

Guergova-Kuras M, Boudreaux B, Joliot A, Joliot P and Redding K (2001) Evidence for two active branches for electron transfer in photosystem I. Proc Natl Acad Sci USA 98: 4437–4442

Guikema JA and Sherman LA (1983) Organization and function of chlorophyll in membranes of cyanobacteria during iron starvation. Plant Physiol 73: 250–256

Guikema JA and Sherman LA (1984) Influence of iron deprivation on the membrane composition of *Anacystis nidulans*. Plant Physiol 74: 90–95

Gumpel NJ, Ralley L, Girard-Bascou J, Wollman FA, Nugent JH and Purton S (1995) Nuclear mutants of *Chlamydomonas reinhardtii* defective in the biogenesis of the cytochrome b_6f complex. Plant Mol Biol 29: 921–932

Gutman BL and Niyogi KK (2004) Chlamydomonas and Arabidopsis. A dynamic duo. Plant Physiol 135: 607–610

Hakala M, Tuominen I, Keranen M, Tyystjarvi T, and Tyystjarvi E (2005) Evidence for the role of the oxygen-evolving manganese complex in photoinhibition of Photosystem II. Biochim Biophys Acta 1706: 68–80

Hallahan BJ, Purton S, Ivison A, Wright D and Evans MCW (1995) Analysis of the proposed Fe-Sx binding region in *Chlamydomonas reinhardtii.* Photosyn Res 46: 257–264

Harris EH (1989) The Chlamydomonas Sourcebook. A Comprehensive Guide to Biology and Laboratory Use. Academic Press, San Diego

Harris EH (2009a) Cell architecture. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 25–64. Elsevier, Amsterdam

Harris EH (2009b) The sexual cycle. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 119–158. Elsevier, Amsterdam

Hegemann P (1997) Vision in microalgae. Planta 203: 265–274

Hegemann P and Berthold P (2009) Sensory photoreceptors and light control of flagellar activity. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 395–430. Elsevier, Amsterdam

Herrin DL (2009) Chloroplast RNA processing and stability. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 937–966. Elsevier, Amsterdam

Herrin DL and Nickelsen J (2004) Chloroplast RNA processing and stability. Photosynth Res 82: 301–314

Hideg E, Kalai T, Kos PB, Asada K and Hideg K (2006) Singlet oxygen in plants – its significance and possible detection with double (fluorescent and spin) indicator reagents. Photochem Photobiol 82: 1211–1218

Higgs D (2009) The chloroplast genome. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 871–892. Elsevier, Amsterdam

Hippler, M, Drepper F, Haehnel W and Rochaix JD (1998) The N-terminal domain of PsaF: Precise recognition site for binding and fast electron transfer from cytochrome c-6 and plastocyanin to photosystem I of *Chlamydomonas reinhardtii*. Proc Natl Acad Sci USA 95: 7339–7344

Holt NE, Fleming GR and Niyogi KK (2004) Toward an understanding of the mechanism of nonphotochemical quenching in green plants. Biochem 43: 8281–8289

Holt NE, Zigmantas D, Valkunas L, Li XP, Niyogi KK and Fleming GR (2005) Carotenoid cation formation and the regulation of photosynthetic light harvesting. Science 307: 433–436

Horton P and Ruban A (2005) Molecular design of the photosystem II light-harvesting antenna: photosynthesis and photoprotection. J Exp Bot 56: 365–373

Howe B and Merchant SS (1992) The biosynthesis of membrane and soluble plastidic *c*-type cytochromes of *Chlamydomonas reinhardtii* is dependent on multiple common gene products. EMBO J 11: 2789–2801

Howe G and Merchant S (1994) Role of heme in the biosynthesis of cytochrome c_6. J Biol Chem 269: 5824–5832

Im C-S, Eberhard S, Huang K, Beck C and Grossman AR (2006) Phototropin involvement in expression of genes encoding chlorophyll and carotenoid biosynthesis enzymes and LHC apoproteins in *Chlamydomonas reinhardtii*. Plant J 48: 1–16

Inoue K, Dreyfuss BW, Kindle KL Stern DB, Merchant S and Sodeinde OA (1997) *Ccs1*, a nuclear gene required for the post-translational assembly of chloroplast c-type cytochromes. J Biol Chem 272: 31747–31754

Irihimovitch V and Shapira M (2000) Glutathione redox potential modulated by reactive oxygen species regulates translation of Rubisco large subunit in the chloroplast. J Biol Chem 275: 16289–16295

Ishida S, Takabayashi A, Ishikawa N, Hano Y, Endo T and Sato F (2009) A novel nuclear-encoded protein, NDH-dependent cyclic electron flow 5, is essential for the accumulation of chloroplast NAD(P)H dehydrogenase complexes. Plant Cell Physiol 50: 383–393

Iwai M, Takahashi Y and Minagawa J (2008) Molecular remodeling of photosystem II during state transitions in *Chlamydomonas reinhardtii*. Plant Cell 20: 2177–2189

Jackowski G, Olkiewicz P and Zelisko A (2003) The acclimative response of the main light harvesting chlorophyll a/b-protein complex of photosystem II (LHCII) to elevated irradiances at the level of trimeric subunits. J Photochem. Photobiol B Biol 70: 163–170

Jahns P, Latowski, D and Strzalka K (2009) Mechanism and regulation of the violaxanthin cycle: the role of antenna proteins and membrane lipids. Biochim Biophys Acta 1787: 3–14

Jain M, Shrager J, Harris EH, Halbrook R, Grossman AR, Hauser C and Vallon O (2007) EST assembly supported by a draft genome sequence: an analysis of the *Chlamydomonas reinhardtii* transcriptome. Nucleic Acids Res 35: 2074–2083

Jansson S, Andersen B and Scheller HV (1996) Nearest-neighbor analysis of higher-plant photosystem I holocomplex. Plant Physiol 112: 409–420

Jensen KH, Herrin DL, Plumley FG and Schmidt GW (1986) Biogenesis of photosystem II complexes: transcriptional, translational, and posttranslational regulation. J Cell Biol 103: 1315–1325

Jensen PE, Gilpin M, Knoetzel J and Scheller HV (2000) The PSI-K subunit of photosystem I is involved in the interaction between light-harvesting complex I and the photosystem I reaction center core. J Biol Chem 275: 24701–24708

Jiao HS, Hicks A, Simpson C and Stern DB (2004) Short dispersed repeats in the Chlamydomonas chloroplast genome are collocated with sites for mRNA 3' end formation. Curr Genet 45: 311–322

Jiao Y, Jia P, Wang X, Su N, Yu S, Zhang D et al. (2005) A tiling microarray expression analysis of rice chromosome 4 suggests a chromosome-level regulation of transcription. Plant Cell 17: 1641–1657

Johnson GN (2005) Cyclic electron transport in C3 plants: fact or artefact? J Exp Bot 56: 407–416

Jordan P, Fromme P, Witt HT, Klukas O, Saenger W and Krauss N (2001) Three-dimensional structure of cyanobacterial photosystem I at 2.5 A resolution. Nature 411: 909–917

Karpowicz SJ, Prochnik SE, Grossman AR and Merchant SS (2011) The GreenCut2 resource: a phyologenomically derived inventory of proteins specific to the plant lineage. J Biol Chem. In Press (Epub ahead of print)

Kapri-Pardes E, Naveh L and Adam Z (2007) The thylakoid lumen protease Deg1 is involved in the repair of photosystem II from photoinhibition in Arabidopsis. Plant Cell 19: 1039–1047

Kathir P, LaVoie M, Brazelton WJ, Haas NA, Lefebvre PA and Silflow CD (2003) Molecular Map of the *Chlamydomonas reinhardtii* Nuclear Genome. Eukaryot Cell 2: 362–379

Keller LC, Romijn EP, Zamora I, Yates JR 3 rd and Marshall WF (2005) Proteomic analysis of isolated Chlamydomonas centrioles reveals orthologs of ciliary-disease genes. Curr Biol 15: 1090–1098

Kim J and Mayfield SP (1997) Protein disulfide isomerase as a regulator of chloroplast translational activation. Science 278: 1954–1957

Kim J and Mayfield SP (2002) The active site of the thioredoxin-like domain of chloroplast protein disulfide isomerase, RB60, catalyzes the redox-regulated binding of chloroplast poly(A)-binding protein, RB47, to the 5' untranslated region of psbA mRNA. Plant Cell Physiol 43: 1238–1243

King SM and Kamiya R (2009) Axonemal dyneins: Assembly, structure and force generation. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 131–208. Elsevier, Amsterdam

Klein U (2009) Chloroplast transcription. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 893–914. Elsevier, Amsterdam

Kohinata T, Nishino H and Fukuzawa H (2008) Significance of zinc in a regulatory protein, CCM1, which regulates the carbon-concentrating mechanism in *Chlamydomonas reinhardtii*. Plant Cell Physiol 49: 273–283

Komayama K, Khatoon M, Takenaka D, Horie J, Yamashita A, Yoshioka M et al (2007) Quality control of Photosystem II: cleavage and aggregation of heat-damaged D1 protein in spinach thylakoids. Biochim Biophys Acta 1767: 838–846

Komenda J, Barker M, Kuvikova S, de Vries R, Mullineaux C and Nixon PJ (2006) The FtsH protease Slr0228 is important for quality control of photosystem II in the thylakoid membrane of *Synechocystis* sp. PCC 6803. J Biol Chem 281: 1145–1151

Komenda J, Tichy M, Prasil O, Knoppova J, Kuvikova S, de Vries R and Nixon PJ (2007) The exposed N-terminal tail of the D1 subunit is required for rapid D1 degradation during photosystem II repair in *Synechocystis* sp PCC 6803. Plant Cell 19: 2839–2854

Kottke T, Hegemann P, Dick B and Heberle J (2006) The photochemistry of the light-, oxygen-, and voltage-sensitive domains in the algal blue light receptor phot. Biopolymers 82: 373–378

Koussevitzky S, Nott A, Mockler TC, Hong F, Sachetto-Martins G, Surpin M et al (2007) Signals from chloroplasts converge to regulate nuclear gene expression. Science 316: 715–719

Koutoulis A, Pazour GJ, Wilkerson CG, Inaba K, Sheng H, Takada S and Witman GB (1997) The *Chlamydomonas reinhardtii ODA3* gene encodes a protein of the outer dynein arm docking complex. J Cell Biol 137: 1069–1080

Krieger-Liszkay A, Fufezan C and Trebst A (2008) Singlet oxygen production in photosystem II and related protection mechanism. Photosynth Res 98: 551–564

Kropat J, Oster U, Rudiger W and Beck CF (1997) Chlorophyll precursors are signals of chloroplast origin involved in light induction of nuclear heat-shock genes. Proc Natl Acad Sci USA 94: 14168–14172

Kropat J, Oster U, Rudiger W and Beck CF (2000) Chloroplast signalling in the light induction of nuclear HSP70 genes requires the accumulation of chlorophyll precursors and their accessibility to cytoplasm/nucleus. Plant J 24: 523–531

Kulheim C, Agren J and Jansson S (2002) Rapid regulation of light harvesting and plant fitness in the field. Science 297: 91–93

Kuras R and Wollman F-A (1994) The assembly of cytochrome b_6f complexes: an approach using genetic transformation of the green alga *Chlamydomonas reinhardtii*. Embo J 13: 1019–1027

Kuras R, Saint-Marcoux D, Wollman FA and de Vitry C (2007) A specific c-type cytochrome maturation system is required for oxygenic photosynthesis. Proc Natl Acad Sci USA 104: 9906–9910

Kuras R, DeVitry C, Choquet Y, Girard-Bascou J, Culler D, Buschlen S et al (1997) Molecular genetic identification of a pathway for heme binding to cytochrome b_6. J Biol Chem 272: 32427–32435

Kurvari V (1997) Cell wall biogenesis in Chlamydomonas: molecular characterization of a novel protein whose expression is up-regulated during matrix formation. Mol Gen Genet 256: 572–580

Kuttkat A, Edhofer I, Eichacker LA and Paulsen H (1997) Light-harvesting chlorophyll *a/b*-binding protein stably inserts into etioplast membranes supplemented with Zn-pheophytin *a/b*. J Biol Chem 272: 20451–20455

Laloi C, Przybyla D and Apel K (2006) A genetic approach towards elucidating the biological activity of different reactive oxygen species in *Arabidopsis thaliana*. J Exp Bot 57: 1719–1724

Lamb MR Dutcher SK, Worley CK and Dieckmann CL (1999) Eyespot-assembly mutants in *Chlamydomonas reinhardtii*. Genetics 153: 721–729

Larbi A, Abadia A, Abadia J and Morales F (2006) Down co-regulation of light absorption, photochemistry, and carboxylation in Fe-deficient plants growing in different environments. Photosynth Res 89: 113–126

Larkin RM and Ruckle ME (2008) Integration of light and plastid signals. Curr Opin Plant Biol 11: 593–599

Lavorel J and Levine RP (1968) Fluorescence properties of wild-type *Chlamydomonas reinhardtii* and three mutant strains having impaired photosynthesis. Plant Physiol 43: 1049–1055

Lawlor DW (2009) Musings about the effects of environment on photosynthesis. Ann Bot (Lond) 103: 543–549

Lax JE, Arteni AA, Boekema EJ, Pistorius EK, Michel KP and Rögner M (2007) Structural response of Photosystem 2 to iron deficiency: characterization of a new photosystem 2-IdiA complex from the cyanobacterium *Thermosynechococcus elongatus* BP-1. Biochim Biophys Acta 1767: 528–534

Ledford HK, Chin BL and Niyogi KK (2007) Acclimation to singlet oxygen stress in *Chlamydomonas reinhardtii*. Eukaryot Cell 6: 919–930

Ledford HK, Baroli I, Shin JW, Fischer BB, Eggen RI and Niyogi KK (2004) Comparative profiling of lipid-soluble antioxidants and transcripts reveals two phases of photooxidative stress in a xanthophyll-deficient mutant of *Chlamydomonas reinhardtii*. Mol Genet Genomics 272: 470–479

Lee KP, Kim C, Landgraf F and Apel K (2007) EXECUTER1- and EXECUTER2-dependent transfer of stress-related signals from the plastid to the nucleus of *Arabidopsis thaliana*. Proc Natl Acad Sci USA 104: 10270–10275

Lefebvre PA (1995) Flagellar amputation and regeneration in Chlamydomonas. Methods Cell Biol 47: 3–7

Lefebvre PA (2009) Flagellar length control. In *The Chlamydomonas Sourcebook*. Harris EH, Stern D and Witman GB (eds). Amsterdam: Elsevier, pp. 115–130

Leister D (2005) Genomics-based dissection of the cross-talk of chloroplasts with the nucleus and mitochondria in Arabidopsis. Gene 354: 110–116

Lemaire C, Girard-Bascou J and Wollman F-A (1986a) Characterization of the b_6f complex subunits and studies on the LHC-Kinase in *Chlamydomonas reinhardtii* using mutant strains altered in the b_6/f complex. In: Biggins J (ed) Progress in Photosynthesis Research, pp 655–658. Nijhoff, Dordrecht

Lemaire C, Girard-Bascou J, Wollman F-A and Bennoun P (1986b) Studies on the cytochrome b_6f complex. I. Characterization of the complex subunits in *Chlamydomonas reinhardtii*. Biochim Biophys Acta 851: 229–238

Leon-Banares R, Gonzalez-Ballester D, Galvan A and Fernandez, E. (2004) Transgenic microalgae as green cell-factories. Trends Biotechnol 22: 45–52

Leonhardt K and Straus NA (1994) Photosystem II genes *isiA*, *psbDI* and *psbC* in *Anabaena* sp. PCC 7120: cloning, sequencing and the transcriptional regulation in iron-stressed and iron-repleted cells. Plant Mol Biol 24: 63–73

Lepisto A, Kangasjarvi S, Luomala EM, Brader G, Sipari N, Keranen M et al (2009) Chloroplast NADPH-thioredoxin reductase interacts with photoperiodic development in Arabidopsis. Plant Physiol 149: 1261–1276

Levine RP (1960) A screening technique for photosynthetic mutants in unicellular algae. Nature 188: 339–340

Levine RP (1969) The analysis of photosynthesis using mutant strains of algae and higher plants. Annu Rev Plant Physiol 20: 523–540

Levine RP and Goodenough UW (1970) The genetics of photosynthesis and of the chloroplast in *Chlamydomonas reinhardii*. Annu Rev Genetics 4: 397–408

Lezhneva L, Kuras R, Ephritikhine G and de Vitry C (2008) A novel pathway of cytochrome *c* biogenesis is involved in the assembly of the cytochrome b_6f complex in Arabidopsis chloroplasts. J Biol Chem 283: 24608–24616

Li XP, Bjorkman O, Shih C, Grossman A, Rosenquist M, Jansson C and Niyogi KK (2000) A pigment-binding protein essential for regulation of photosynthetic light harvesting. Nature 403: 391–395

Li XP, Gilmore AM and Niyogi KK (2002a) Molecular and global time-resolved analysis of a *psbS*

gene dosage effect on pH- and xanthophyll cycle-dependent nonphotochemical quenching in photosystem II. J Biol Chem 277: 33590–33597

Li XP, Muller-Moule P, Gilmore AM and Niyogi KK (2002b) PsbS-dependent enhancement of feedback de-excitation protects photosystem II from photoinhibition. Proc Natl Acad Sci USA 99: 15222–15227

Li XP, Gilmore AM, Caffarri S, Bassi R, Golan T, Kramer D and Niyogi KK (2004) Regulation of photosynthetic light harvesting involves intrathylakoid lumen pH sensing by the PsbS protein. J Biol Chem 279: 22866–22874

Li Y, van der Est A, Lucas MG, Ramesh VM, Gu F, Petrenko A et al (2006) Directing electron transfer within Photosystem I by breaking H-bonds in the cofactor branches. Proc Natl Acad Sci USA 103: 2144–2149

Li Z, Wakao S, Fischer BB and Niyogi KK (2009) Sensing and responding to excess light. Annu Rev Plant Biol. 60: 239–260

Lohr M (2009) Carotenoids. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 799–818. Elsevier, Amsterdam

Lohr M, Im CS and Grossman AR (2005) Genome-based examination of chlorophyll and carotenoid biosynthesis in *Chlamydomonas reinhardtii*. Plant Physiol 138: 490–515

Loiselay C, Gumpel NJ, Girard-Bascou J, Watson AT, Purton S, Wollman FA and Choquet Y (2008) Molecular identification and function of cis- and trans-acting determinants for *petA* transcript stability in *Chlamydomonas reinhardtii* chloroplasts. Mol Cell Biol 28: 5529–5542

Long JC, Sommer F, Allen MD, Lu SF and Merchant SS (2008) FER1 and FER2 encoding two ferritin complexes in *Chlamydomonas reinhardtii* chloroplasts are regulated by iron. Genetics 179: 137–147

Lunde C, Jensen PE, Haldrup A, Knoetzel J and Scheller HV (2000) The PSI-H subunit of photosystem I is essential for state transitions in plant photosynthesis. Nature 408: 613–615

Ma L, Li J, Qu L, Hager J, Chen Z, Zhao H and Deng XW (2001) Light control of Arabidopsis development entails coordinated regulation of genome expression and cellular pathways. Plant Cell 13: 2589–2607

Maheswari U, Mock T, Armbrust EV and Bowler C (2009) Update of the Diatom EST Database: a new tool for digital transcriptomics. Nucleic Acids Res 37: D1001–1005

Martin M, Funk HT, Serrot PH, Poltnigg P and Sabater B (2009) Functional characterization of the thylakoid Ndh complex phosphorylation by site-directed mutations in the *ndhF* gene. Biochim Biophys Acta 1787: 920–928

Matsuzaki M, Misumi O, Shin IT, Maruyama S, Takahara M, Miyagishima SY et al (2004) Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. Nature 428: 653–657

Maul JE, Lilly JW, Cui L, dePamphilis CW, Miller W, Harris EH and Stern DB (2002) The *Chlamydomonas reinhardtii* plastid chromosome: islands of genes in a sea of repeats. Plant Cell 14: 2659–2679

May P, Wienkoop S, Kempa S, Usadel B, Christian N, Rupprecht J et al (2008) Metabolomics- and proteomics-assisted genome annotation and analysis of the draft metabolic network of *Chlamydomonas reinhardtii*. Genetics 179: 157–166

May P, Christian JO, Kempa S and Walther D (2009) ChlamyCyc: an integrative systems biology database and web-portal for *Chlamydomonas reinhardtii*. BMC Genomics 10: 209

Mayfield S (1996) Double agent: Translational regulation by a transcription factor. Chem Biol (London) 3: 415–418

McKim SM and Durnford DG (2006) Translational regulation of light-harvesting complex expression during photoacclimation to high-light in *Chlamydomonas reinhardtii*. Plant Physiol Biochem 44: 857–865

Melkozernov AN, Kargul J, Lin S, Barber J and Blankenship RE (2005) Spectral and kinetic analysis of the energy coupling in the PS I-LHC I supercomplex from the green alga *Chlamydomonas reinhardtii* at 77 K. Photosynth Res 86: 203–215

Melkozernov AN, Su H, Lin S, Bingham S, Webber AN and Blankenship RE (1997) Specific mutations near the primary donor in Photosystem I from *Chlamydomonas reinhardtii* alters the trapping time and spectroscopic properties of P700. Biochemistry 36: 2898–2907

Merchant SS, Allen MD, Kropat J, Moseley JL, Long JC, Tottey S and Terauchi AM (2006) Between a rock and a hard place: trace element nutrition in Chlamydomonas. Biochim Biophys Acta 1767: 578–594

Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB et al (2007) The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science 318: 245–250

Merendino L, Perron K, Rahire M, Howald I, Rochaix JD and Goldschmidt-Clermont M. (2006) A novel multifunctional factor involved in trans-splicing of chloroplast introns in Chlamydomonas. Nucleic Acids Res 34: 262–274

Meskauskiene R, Nater M, Goslings D, Kessler F op den Camp R and Apel K (2001) FLU: a negative regulator of chlorophyll biosynthesis in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 98: 12826–12831

Mets L and Bogorad L (1971) Mendelian and uniparental alterations in erythromycin binding by plastid ribosomes. Science 174: 707–709

Michel H and Deisenhofer J (1988) Relevance of the photosynthetic reaction center from purple bacteria to the structure of photosystem II. Biochem 27: 1–7

Michel KP and Pistorius EK (2004) Adaptation of the photosynthetic electron transport chain in cyanobacteria to iron deficiency: The function of IdiA and IsiA. Physiol Plant 120: 36–50

Michelet L, Zaffagnini M and Lemaire SD (2009) Thioredoxins and related proteins. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 401–444. Elsevier, Amsterdam

Minagawa J (2009) Light-harvesting proteins. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 503–540. Elsevier, Amsterdam

Misumi O, Matsuzaki M, Nozaki H, Miyagishima SY, Mori T, Nishida K et al (2005) *Cyanidioschyzon merolae* genome. A tool for facilitating comparable studies on organelle biogenesis in photosynthetic eukaryotes. Plant Physiol 137: 567–585

Mittler R, Vanderauwera S, Gollery M and Van Breusegem F (2004) Reactive oxygen gene network of plants. Trends Plant Sci 9: 490–498

Mochizuki N, Tanaka R, Tanaka A, Masuda T and Nagatani A (2008) The steady-state level of Mg-protoporphyrin IX is not a determinant of plastid-to-nucleus signaling in Arabidopsis. Proc Natl Acad Sci USA 105: 15184–15189

Moll B and Levine RP (1970) Characterization of a photosynthetic mutant strain of *Chlamydomonas reinhardi* deficient in phosphoribulokinase activity. Plant Physiol 46: 576–580

Moller IM, Jensen PE and Hansson A (2007) Oxidative modifications to cellular components in plants. Annu Rev Plant Biol 58: 459–481

Molnar A, Schwach F, Studholme DJ, Thuenemann EC and Baulcombe DC (2007) miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*. Nature 447: 1126–1129

Molnar A, Bassett A, Thuenemann E, Schwach F, Karkare S, Ossowski S et al (2009) Highly specific gene silencing by artificial microRNAs in the unicellular alga *Chlamydomonas reinhardtii*. Plant J 58: 165–174

Moseley J and Grossman AR (2009) Phosphorus limitation from the physiological to the genomic. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 189–216. Elsevier, Amsterdam

Moseley J, Quinn J, Eriksson M and Merchant SS (2000) The *Crd1* gene encodes a putative di-iron enzyme required for photosystem I accumulation in copper deficiency and hypoxia in *Chlamydomonas reinhardtii*. EMBO J 19: 2139–2151

Moseley J, Gonzalez-Ballester D, Pootakham W, Bailey S and Grossman AR (2009) Genetic interactions between regulators of Chlamydomonas phosphorus and sulfur deprivation responses. Genetics 181: 889–905

Moseley JL, Chang CW and Grossman AR (2006) Genome-based approaches to understanding phosphorus deprivation responses and PSR1 control in *Chlamydomonas reinhardtii*. Eukaryot Cell 5: 26–44

Moseley JL, Allinger T, Herzog S, Hoerth P, Wehinger E, Merchant S and Hippler M (2002) Adaptation to Fe-deficiency requires remodeling of the photosynthetic apparatus. EMBO J 21: 6709–6720

Moulin M, McCormac AC, Terry MJ and Smith AG (2008) Tetrapyrrole profiling in Arabidopsis seedlings reveals that retrograde plastid nuclear signaling is not due to Mg-protoporphyrin IX accumulation. Proc Natl Acad Sci USA 105: 15178–15183

Mozzo M, Passarini F, Bassi R, van Amerongen H and Croce R (2008) Photoprotection in higher plants: the putative quenching site is conserved in all outer light-harvesting complexes of Photosystem II. Biochim Biophys Acta 1777: 1263–1267

Muhlenbock P, Szechynska-Hebda M, Plaszczyca M, Baudo M, Mateo A, Mullineaux PM et al (2008) Chloroplast signaling and LESION SIMULATING DISEASE1 regulate crosstalk between light acclimation and immunity in Arabidopsis. Plant Cell 20: 2339–2356

Mulkidjanian AY, Koonin EV, Makarova KS, Mekhedov SL, Sorokin A, Wolf YI et al (2006) The cyanobacterial genome core and the origin of photosynthesis. Proc Natl Acad Sci USA 103: 13126–13131

Munekage Y, Hashimoto M, Miyake C, Tomizawa K, Endo T, Tasaka M and Shikanai, T (2004) Cyclic electron flow around photosystem I is essential for photosynthesis. Nature 429: 579–582

Murata N (1969) Control of excitation transfer in photosynthesis. I. Light-induced change of chlorophyll *a* fluorescence in *Porphyridium cruentum*. Biochim Biophys Acta 172: 242–251

Murray JW, Duncan, J and Barber J (2006) CP43-like chlorophyll binding proteins: structural and evolutionary implications. Trends Plant Sci 11: 152–158

Mus F, Dubini A, Seibert M, Posewitz MC and Grossman AR (2007) Anaerobic acclimation in *Chlamydomonas reinhardtii*: anoxic gene expression, hydrogenase induction, and metabolic pathways. J Biol Chem 282: 25475–25486

Mussgnug JH, Wobbe L, Elles I, Claus C, Hamilton M, Fink A et al (2005) NAB1 is an RNA binding protein involved in the light-regulated differential expression of the light-harvesting antenna of *Chlamydomonas reinhardtii*. Plant Cell 17: 3409–3421

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M and Snyder M (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320: 1344–1349

Nandha B, Finazzi G, Joliot P, Hald S and Johnson GN (2007) The role of PGR5 in the redox poising of photosynthetic electron transport. Biochim Biophys Acta 1767: 1252–1259

Naumann B, Stauber EJ, Busch A, Sommer F and Hippler M (2005) N-terminal processing of Lhca3 Is a key step in remodeling of the photosystem I-light-harvesting complex under iron deficiency in *Chlamydomonas reinhardtii*. J Biol Chem 280: 20431–20441

Naumann B, Busch A, Allmer J, Ostendorf E, Zeller M, Kirchhoff H and Hippler M (2007) Comparative quantitative proteomics to investigate the remodeling

of bioenergetic pathways under iron deficiency in *Chlamydomonas reinhardtii*. Proteomics 7: 3964–3979

Nelson N and Yocum CF (2006) Structure and function of photosystems I and II. Annu Rev Plant Biol 57: 521–565

Nguyen AV, Thomas-Hall SR, Malnoe A, Timmins M, Mussgnug JH, Rupprecht J et al (2008) The transcriptome of photo-biological hydrogen production induced by sulphur deprivation in the green alga *Chlamydomonas reinhardtii*. Eukaryot Cell 7: 1965–1979

Nield J, Morris EP, Bibby TS and Barber J (2003) Structural analysis of the photosystem I supercomplex of cyanobacteria induced by iron deficiency. Biochem 42: 3180–3188

Nield J, Kruse O, Ruprecht J, da Fonseca P, Buchel C and Barber J (2000) Three-dimensional structure of *Chlamydomonas reinhardtii* and *Synechococcus elongatus* photosystem II complexes allows for comparison of their oxygen-evolving complex organization. J Biol Chem 275: 27940–27946

Nilsson A, Stys D, Drakenberg T, Spangfort MD, Forsen S and Allen JF (1997) Phosphorylation controls the three-dimensional structure of plant light harvesting complex II. J Biol Chem 272: 18350–18357

Nishiyama Y, Allakhverdiev SI and Murata N (2006) A new paradigm for the action of reactive oxygen species in the photoinhibition of photosystem II. Biochim Biophys Acta 1757: 742–749

Niyogi K (2009) Photoprotection and high light responses. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 847–870. Elsevier, Amsterdam

Niyogi KK, Björkrman O and Grossman AR (1997) Chlamydomonas xanthophyll cycle mutants identified by video imaging of chlorophyll fluorescence quenching. Plant Cell 9: 1369–1380

Niyogi KK (1999) Photoprotection revisited: Genetic and molecular approaches. Annu Rev Plant Physiol Plant Mol Biol 50: 333–359

Niyogi KK, Grossman AR and Björkman O (1998) Arabidopsis mutants define a central role for the xanthophyll cycle in the regulation of photosynthetic energy conversion. Plant Cell 10: 1121–1134

Niyogi KK, Li XP, Rosenberg V and Jung HS (2005) Is PsbS the site of non-photochemical quenching in photosynthesis? J Exp Bot 56: 375–382

Nott A, Jung HS, Koussevitzky S and Chory J (2006) Plastid-to-nucleus retrograde signaling. Annu Rev Plant Biol 57: 739–759

Oelze ML, Kandlbinder A and Dietz KJ (2008) Redox regulation and overreduction control in the photosynthesizing cell: complexity in redox regulatory networks. Biochim Biophys Acta 1780: 1261–1272

Ohad I, Kyle DJ and Arntzen CJ (1984) Membrane protein damage and repair: removal and replacement of inactivated 32-kilodalton polypeptides in chloroplast membranes. J Cell Biol 99: 481–485

Ohnishi N, Allakhverdiev SI, Takahashi S, Higashi S, Watanabe M, Nishiyama Y, Murata N (2005) Two-step mechanism of photodamage to photosystem II: Step 1 occurs at the oxygen-evolving complex and step 2 occurs at the photochemical reaction center. Biochem 44: 8494–8499

Ooyabu J, Ohtsuka M, Kashino Y, Koike H and Satoh K (2008) The expression pattern of NAD(P)H oxidases and the cyclic electron transport pathway around photosystem I of *Synechocystis* sp. PCC6803 depend on growth conditions. Biosci Biotechnol Biochem 72: 3180–3188

Osmond B and Forster B (2006) Photoprotection, photoinhibition, gene regulation and environment. In: Demmig-Adams B, Adams III WW and Mattoo AK (eds) Photoinhibition: Then and Now, pp 11–22. Springer, Dordrecht

Pakrasi HB, Goldenberg A and Sherman LA (1985) Membrane development in the cyanobacterium, *Anacystis nidulans*, during recovery from iron starvation. Plant Physiol 79: 290–295

Palenik B, Grimwood J, Aerts A, Rouze P, Salamov A, Putnam N et al (2007) The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. Proc Natl Acad Sci USA 104: 7705–7710

Pazour G and Witman GB (2009) The Chlamydomonas flagellum as a model for human ciliary disease. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 445–478. Elsevier, Amsterdam

Pazour GJ, Sineshchekov OA and Witman GB (1995) Mutational analysis of the phototransduction pathway of *Chlamydomonas reinhardtii*. J Cell Biol 131: 427–440

Peers G, Truong TB, Ostendorf E, Elrad D, Grossman AR, Hippler M and Niyogi KK (2009) An ancient light-harvesting protein is critical for the regulation of algal photosynthesis. Nature 462: 518–21

Peltier G and Cournac L (2002) Chlororespiration. Annu Rev Plant Biol 53: 523–550

Perret MC, Bernd K.K and Kohorn BD (1998) Chloroplast protein translocation. In: Rochaix J-D, Goldschmidt-Clermont M and Merchant S (eds) Molecular Biology of Chlamydomonas: Chloroplasts and Mitochondria, pp 219–231. Kluwer Academic Publishers, Dordrecht

Pesaresi P, Schneider A, Kleine T and Leister D (2007) Interorganellar communication. Curr Opin Plant Biol 10: 600–606

Pfundel EE and Dilley RA (1993) The pH dependence of violaxanthin deepoxidation in isolated pea chloroplasts. Plant Physiol 101: 65–71

Pilon M, Abdel-Ghany SE, Van Hoewyk D, Ye, H and Pilon-Smits EA (2006) Biogenesis of iron-sulfur cluster proteins in plastids. Genet Eng (N Y) 27: 101–117

Pogson BJ and Rissler, HM (2000) Genetic manipulation of carotenoid biosynthesis and photoprotection. Philos Trans R Soc Lond B Biol Sci 355: 1395–1403

Pogson BJ, Woo NS, Forster B and Small ID (2008) Plastid signalling to the nucleus and beyond. Trends Plant Sci 13: 602–609

Polivka T, HerekJL, Zigmantas D, Akerland HE and Sundstrom V (1999) Direct observation of the (forbidden)

S1 state in carotenoids. Proc Natl Acad Sci USA 96: 4914–4917

Posewitz M, Dubini A, Meuser JE, Seibert M and Ghirardi ML (2009) Hydrogenase, hydrogen production and anoxia. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 218–256. Elsevier, Amsterdam

Puthiyaveetil S and Allen JF (2009) Chloroplast two-component systems: evolution of the link between photosynthesis and gene expression. Proc Biol Sci 276: 2133–2145

Puthiyaveetil S, Kavanagh TA, Cain P, Sullivan JA, Newell CA, Gray JC et al (2008) The ancestral symbiont sensor kinase CSK links photosynthesis with gene expression in chloroplasts. Proc Natl Acad Sci USA 105: 10061–10066

Raynaud C, Loiselay C, Wostrikoff K, Kuras R, Girard-Bascou J, Wollman FA and Choquet Y (2007) Evidence for regulatory function of nucleus-encoded factors on mRNA stabilization and translation in the chloroplast. Proc Natl Acad Sci USA 104: 9093–9098

Redding K (2009) Photosystem I. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 541–572. Elsevier, Amsterdam

Reikhof WR and Benning C (2009) Glycerolipid biosynthesis. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 41–68. Elsevier, Amsterdam

Reinbothe C, Pollmann S, Phetsarath-Faure P, Quigley F, Weisbeek P and Reinbothe S (2008) A pentapeptide motif related to a pigment binding site in the major light-harvesting protein of photosystem II, LHCII, governs substrate-dependent plastid import of NADPH:protochlorophyllide oxidoreductase A. Plant Physiol 148: 694–703

Reinbothe C, Bartsch S, Eggink LL Hoober JK, Brusslan J, Andrade-Paz R et al (2006) A role for chlorophyllide *a* oxygenase in the regulated import and stabilization of light-harvesting chlorophyll *a/b* proteins. Proc Natl Acad Sci USA 103: 4777–4782

Riano-Pachon DM, Correa LG, Trejos-Espinosa R and Mueller-Roeber B (2008) Green transcription factors: a Chlamydomonas overview. Genetics 179: 31–39

Richard C, Ouellet H and Guertin M (2000) Characterization of the LI818 polypeptide from the green unicellular alga *Chlamydomonas reinhardtii*. Plant Mol Biol 42: 303–316

Rizhsky L, Liang H and Mittler R (2003) The water-water cycle is essential for chloroplast protection in the absence of stress. J Biol Chem 278: 38921–38925

Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA et al (2003) Genome divergence in two Prochlorococcus ecotypes reflects oceanic niche differentiation. Nature 424: 1042–1047

Rochaix J-D (1995) *Chlamydomonas reinhardtii* as the photosynthetic yeast. Annu Rev Genetics 29: 209–230

Rochaix J-D (1996) Post-transcriptional regulation of chloroplast gene expression in *Chlamydomonas*. Plant Mol Biol 32: 327–341

Rochaix J-D (2009) State transitions. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 819–846. Elsevier, Amsterdam

Rochaix J-D and Malnoe P (1978) Anatomy of the chloroplast ribosomal DNA of *Chlamydomonas reinhardtii*. Cell 15: 661–670

Rochaix J-D and Darlix J-L (1982) Composite structure of the chloroplast 23 S ribosomal RNA genes of *Chlamydomonas reinhardtii*. Evolutionary and functional implications. J Mol Biol 159: 383–395

Rochaix JD (2001) Posttranscriptional control of chloroplast gene expression. From RNA to photosynthetic complex. Plant Physiol 125: 142–144

Rochaix JD (2002) Chlamydomonas, a model system for studying the assembly and dynamics of photosynthetic complexes. FEBS Lett 529: 34–38

Rochaix JD (2004) Genetics of the biogenesis and dynamics of the photosynthetic machinery in eukaryotes. Plant Cell 16: 1650–1660

Rochaix JD (2007) Role of thylakoid protein kinases in photosynthetic acclimation. FEBS Lett 581: 2768–2775

Rochaix JD, Perron K, Dauvillee D, Laroche F, Takahashi Y and Goldschmidt-Clermont M (2004) Post-transcriptional steps involved in the assembly of photosystem I in Chlamydomonas. Biochem Soc Trans 32: 567–570

Rodday SM, Webber AN, Bingham SE and Biggins J (1995) Evidence that the F_X domain in Photosystem I interacts with the subunit PsaC: Site-directed changes in PsaB destabilize the subunit interactions in *Chlamydomonas reinhardtii*. Biochem 34: 6328–6334

Rodday SM, Do LT, Chynwat V, Frank H and Biggins J (1996) Site-directed mutagenesis of the subunit PsaC establishes a surface exposed domain interacting with the Photosystem I core binding site. Biochem 35: 11832–11838

Rohr J, Sarkar N, Balenger S, Jeong BR and Cerutti H (2004) Tandem inverted repeat system for selection of effective transgenic RNAi strains in Chlamydomonas. Plant J 40: 611–621

Rolland N, Atteia A, Decottignies P, Garin J, Hippler M, Kreimer G et al (2009) Chlamydomonas proteomics. Curr Opin Microbiol. 12: 285–291

Ruban AV, Berera R, Ilioaia C, van Stokkum IH, Kennis JT, Pascal AA et al (2007) Identification of a mechanism of photoprotective energy dissipation in higher plants. Nature 450: 575–578

Ruckle ME, DeMarco SM and Larkin RM (2007) Plastid signals remodel light signaling networks and are essential for efficient chloroplast biogenesis in Arabidopsis. Plant Cell 19: 3944–3960

Rumeau D, Peltier G and Cournac L (2007) Chlororespiration and cyclic electron flow around PSI during photosynthesis and plant stress response. Plant Cell Environ 30: 1041–1051

Rymarquis LA, Higgs DC and Stern DB (2006) Nuclear suppressors define three factors that participate in both 5' and 3' end processing of mRNAs in Chlamydomonas chloroplasts. Plant J 46: 448–461

Rymarquis LA, Webster BR and Stern DB (2007) The nucleus-encoded factor MCD4 participates in degradation of nonfunctional 3' UTR sequences generated by cleavage of pre-mRNA in Chlamydomonas chloroplasts. Mol Genet Genomics 277: 329–340

Rymarquis LA, Handley JM, Thomas M and Stern DB (2005) Beyond complementation. Map-based cloning in *Chlamydomonas reinhardtii*. Plant Physiol 137: 557–566

Sager R (1960) Genetic systems in *Chlamydomonas*. Science 132: 1459–1465

Sakamoto W (2006) Protein degradation machineries in plastids. Annu Rev Plant Biol 57: 599–621

Sandstrom S, Ivanov AG, Park YI, Oquist G and Gustafsson P (2002) Iron stress responses in the cyanobacterium *Synechococcus* sp. PCC7942. Physiol Plant 116: 255–263

Sato V, Levine RP and Neumann J (1971) Photosynthetic phosphorylation in *Chlamydomonas reinhardti.* Effects of a mutation altering an ATP-synthesizing enzyme. Biochim Biophys Acta 253: 437–448

Savard F, Richard C and Guertin M (1996) The *Chlamydomonas reinhardtii* LI818 gene represents a distant relative of the cabI\II genes that is regulated during the cell cycle and in response to illumination. Plant Mol Biol 32: 461–473

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, et al (2005) A gene expression map of *Arabidopsis thaliana* development. Nat Genet 37: 501–506

Schmidt GW and Mishkind ML (1986) The transport of proteins into chloroplasts. Annu Rev Biochem 55: 879–912

Schmidt GW, Devillers-Thiery A, Desruisseaux H, Blobel G and Chua N-H (1979) NH_2 -terminal amino acid sequence of precursor and mature forms of the ribulose-1,5-bisphosphate carboxylase small subunit from *Chlamydomonas reinhardtii*. J Cell Biol 83: 615–622

Schmidt M, Gessner G, Luff M, Heiland I, Wagner V, Kaminski M et al (2006) Proteomic analysis of the eyespot of *Chlamydomonas reinhardtii* provides novel insights into its components and tactic movements. Plant Cell 18: 1908–1930

Scholey JM (2008) Intraflagellar transport motors in cilia: moving along the cell's antenna. J Cell Biol 180: 23–29

Scholey JM and Anderson KV (2006) Intraflagellar transport and cilium-based signaling. Cell 125: 439–442

Schroda M (2006) RNA silencing in Chlamydomonas: mechanisms and tools. Curr Genet 49: 69–84

Schwarz C, Elles I, Kortmann J, Piotrowski M and Nickelsen J (2007) Synthesis of the D2 protein of photosystem II in Chlamydomonas is controlled by a high molecular mass complex containing the RNA stabilization factor Nac2 and the translational activator RBP40. Plant Cell 19: 3627–3639

Shao N, Beck CF, Lemaire SD and Krieger-Liszkay A (2008) Photosynthetic electron flow affects H_2O_2 signaling by inactivation of catalase in *Chlamydomonas reinhardtii*. Planta 228: 1055–1066

Shao N, Vallon O, Dent R, Niyogi KK and Beck CF (2006) Defects in the cytochrome b6/f complex prevent light-induced expression of nuclear genes involved in chlorophyll biosynthesis. Plant Physiol 141: 1128–1137

Sharma S (2007) Adaptation of photosynthesis under iron deficiency in maize. J Plant Physiol 164: 1261–1267

Shepherd HS, Boynton JE and Gillham NW (1979) Mutations in nine chloroplast loci of *Chlamydomonas* affecting different photosynthetic functions. Proc Nat Acad Sci USA 76: 1353–1357

Shikanai T (2007) Cyclic electron transport around photosystem I: Genetic approaches. Annu Rev Plant Biol 58: 199–217

Shimizu H and Shikanai T (2007) Dihydrodipicolinate reductase-like protein, CRR1, is essential for chloroplast NAD(P)H dehydrogenase in Arabidopsis. Plant J 52: 539–547

Shimizu H, Peng L, Myouga F, Motohashi R, Shinozaki K and Shikanai T (2008) CRR23/NdhL is a subunit of the chloroplast NAD(P)H dehydrogenase complex in Arabidopsis. Plant Cell Physiol 49: 835–842

Shrager J, Hauser C, Chang CW, Harris EH Davies J, McDermott J et al (2003) *Chlamydomonas reinhardtii* genome project. A guide to the generation and use of the cDNA information. Plant Physiol 131: 401–408

Sineshchekov OA, Govorunova EG and Spudich JL (2009) Photosensory functions of channelrhodopsins in native algal cells. Photochem Photobiol 85: 556–563

Sizova I, Fuhrmann M and Hegemann P (2001) A Streptomyces rimosus *aphVIII* gene coding for a new type phosphotransferase provides stable antibiotic resistance to *Chlamydomonas reinhardtii*. Gene 277: 221–229

Smith T and Kohorn BD (1994) Mutations in a signal sequence for the thylakoid membrane identify multiple protein transport pathways and nuclear suppressors. J Cell Biol 126: 365–374

Spreitzer RJ and Mets LJ (1980) Non-mendelian mutation affecting ribulose-1,5- biosphosphate carboxylase/oxygenase. Nature 285: 114–115

Stauber EJ and Hippler M (2004) *Chlamydomonas reinhardtii* proteomics. Plant Physiol Biochem 42: 989–1001

Stauber EJ, Busch A, Naumann B, Svatos A and Hippler M (2009) Proteotypic profiling of LHCI from *Chlamydomonas reinhardtii* provides new insights into structure and function of the complex. Proteomics 9: 398–408

Stern DB and Drager RG (1998) Chloroplast RNA synthesis and processing. In: Rochaix J-D, Goldschmidt-Clermont M and Merchant S (eds) Molecular Biology of Chlamydomonas: Chloroplasts and Mitochondria, pp 165–183. Kluwer Academic Publishers, Dordrecht

Stern DB, Higgs DC and Yang J (1997) Transcriptional and translational activities of chloroplasts. Trends Plant Sci 2: 308–315

Stevens DR, Rochaix, J-D and Purton S (1996) The bacterial phleomycin resistance gene *ble* as a dominant selectable marker in Chlamydomonas. Mol Gen Genetics 251: 23–30

Strzepek RF and Harrison PJ (2004) Photosynthetic architecture differs in coastal and oceanic diatoms. Nature 431: 689–692

Subramanyam R, Jolley C, Brune DC, Fromme P and Webber AN (2006) Characterization of a novel Photosystem I-LHCI supercomplex isolated from *Chlamydomonas reinhardtii* under anaerobic (State II) conditions. FEBS Lett 580: 233–238

Surzycki SJ and Rochaix J-D (1971) Transcriptional mapping of ribosomal RNA genes of the chloroplast and nucleus of *Chlamydomonas reinhardtii*. J Mol Biol 62: 89–109

Swiatek M, Kuras R, Sokolenko A, Higgs D, Olive J, Cinque G et al (2001) The chloroplast gene *ycf9* encodes a photosystem II (PSII) core subunit, PsbZ, that participates in PSII supramolecular structure. Plant Cell 13: 1347–1367

Szabo I, Bergantino E and Giacometti GM (2005) Light and oxygenic photosynthesis: energy dissipation as a protection mechanism against photo-oxidation. Embo Rep 6: 629–634

Takahashi S and Murata N (2008) How do environmental stresses accelerate photoinhibition? Trends Plant Sci 13: 178–182

Takahashi S, Milward SE, Fan DY, Chow WS and Badger, MR (2009) How does cyclic electron flow alleviate photoinhibition in Arabidopsis? Plant Physiol 149: 1560–1567

Takahashi YM, Goldschmidt-Clermont M and Rochaix J-D (1992) Directed mutagenesis of the chloroplast gene *psaC* in *Chlamydomonas reinhardtii*. In: Murata N (ed) Research in Photosynthesis, Vol III, pp 393–396. Kluwer Academic Publishers, Dordrecht

Tejada-Jimenez M, Llamas A, Sanz-Luque E, Galván, A and Fernández E (2007) A high-affinity molybdate transporter in eukaryotes. Proc Natl Acad Sci USA 104: 20126–20130

Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Kruger P et al (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J 37: 914–939

Tikkanen M, Nurmi M, Kangasjarvi S and Aro EM (2008a) Core protein phosphorylation facilitates the repair of photodamaged photosystem II at high light. Biochim Biophys Acta 1777: 1432–1437

Tikkanen M, Nurmi M, Suorsa M, Danielsson R, Mamedov F, Styring S and Aro EM (2008b) Phosphorylation-dependent regulation of excitation energy distribution between the two photosystems in higher plants. Biochim Biophys Acta 1777: 425–432

Tikkanen M, Piippo M, Suorsa M, Sirpio S, Mulo P, Vainonen J et al (2006) State transitions revisited-a buffering system for dynamic low light acclimation of Arabidopsis. Plant Mol Biol 62: 779–793

Tokutsu R, Iwai M and Minagawa J (2009) CP29, a monomeric light-harvesting complex II protein, is essential for state transitions in *Chlamydomonas reinhardtii*. J Biol Chem 284: 7777–7782

Tottey S, Block MA, Allen M, Westergren T, Albrieux C, Scheller HV et al (2003) Arabidopsis CHL27, located in both envelope and thylakoid membranes, is required for the synthesis of protochlorophyllide. Proc Natl Acad Sci USA 100: 16119–16124

Turkina MV, Blanco-Rivero A, Vainonen JP, Vener AV and Villarejo A (2006) CO_2 limitation induces specific redox-dependent protein phosphorylation in *Chlamydomonas reinhardtii*. Proteomics 6: 2693–2704

Tuxhorn J, Daise T and Dentler WL (1998) Regulation of flagellar length in Chlamydomonas. Cell Motility and the Cytoskeleton 40: 133–146

Tyystjarvi E (2008) Photoinhibition of Photosystem II and photodamage of the oxygen evolving manganese cluster. Coordination Chem Rev 252: 361–376

Uniacke J and Zerges W (2007) Photosystem II assembly and repair are differentially localized in Chlamydomonas. Plant Cell 19: 3640–3654

Vainonen JP, Hansson M and Vener AV (2005) STN8 protein kinase in *Arabidopsis thaliana* is specific in phosphorylation of photosystem II core proteins. J Biol Chem 280: 33679–33686

Vallon O and Spalding MH (2009) Amino acid metabolism. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 115–158. Elsevier, Amsterdam

Vallon O, Bulte L, Dainese P, Olive J, Bassi R and Wollman FA (1991) Lateral redistribution of cytochrome b6f complexes along thylakoid membranes upon state transitions. Proc Natl Acad Sci USA 88: 8262–8266

Vardi A, Thamatrakoln K, Bidle K and Falkowski PG (2008) Diatom genomes come of age. Genome Biology 9: 245.1–245.5

Varsano T, Wolf SG and Pick U (2006) A chlorophyll a/b-binding protein homolog that is induced by iron deficiency is associated with enlarged photosystem I units in the eucaryotic alga *Dunaliella salina*. J Biol Chem 281: 10305–10315

Vener A, van Kan PJ, Rich PR, Ohad I and Andersson B (1997) Plastoquinol at the Qo-site of reduced cytochrome *b/f* mediates signal transduction between light and thylakoid phosphorylation. Thylakoid protein kinase deactivation by a single turnover flash. Proc Natl Acad Sci USA 94: 1585–1590.

Vernotte C, Astier C and Olive J (1990) State 1-state 2 adaptation in the cyanobacteria *Synechocystis* PCC 6714 wild type and *Synechocystis* PCC 6803 wild type and phycocyanin-less mutant. Photosyn Res 26: 203–212

Vernotte C, Picaud M, Kirilovsky D, Olive J, Ajlani G and Astier C (1992) Changes in the photosynthetic apparatus in the cyanobacterium *Synechocystis* sp. PCC 6714 following light-to-dark and dark-to-light transitions. Photosyn Res 32: 45–57

Wagner D, Przybyla D, Op den Camp R, Kim C, Landgraf F, Lee KP et al (2004a) The genetic basis of singlet oxygen-induced stress responses of *Arabidopsis thaliana*. Science 306: 1183–1185

Wagner V and Mittag M (2009) Probing circadian rhythms in *Chlamydomonas rheinhardtii* by functional proteomics. Methods Mol Biol 479: 173–188

Wagner V, Fiedler M., Markert C, Hippler M and Mittag M (2004b) Functional proteomics of circadian expressed proteins from *Chlamydomonas reinhardtii*. FEBS Lett 559: 129–135

Wagner V, Ullmann K, Mollwo A, Kaminski M, Mittag M and Kreimer G (2008) The phosphoproteome of a *Chlamydomonas reinhardtii* eyespot fraction includes key proteins of the light signaling pathway. Plant Physiol 146: 772–788

Walters RG (2005) Towards an understanding of photosynthetic acclimation. J Exp Bot 56: 435–447

Wang Z, Gerstein M and Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10: 57–63

Waters MT, Wang P, Korkaric M, Capper RG, Saunders NJ and Langdale JA (2009) GLK transcription factors coordinate expression of the photosynthetic apparatus in Arabidopsis. Plant Cell 21: 1109–1128

Webber AN, Bingham SE and Lee H (1995) Genetic engineering of thylakoid protein complexes by chloroplast transformation in *Chlamydomonas reinhardtii*. Photosyn Res 44: 191–205

Webber AN, Su H, Bingham SE, Käss H, Krabben L, Kuhn M et al (1996) Site-directed mutations affecting the spectroscopic characteristics and mid-point potential of the primary donor in photosystem I. Biochem 39: 12857–12863

Wettstein DV, Kahn A, Nielsen OF and Gough S (1974) Genetic regulation of chlorophyll synthesis analyzed with mutants in barley. Science 184: 800–802

Wheeler GL, Miranda-Saavedra D and Barton GJ (2008) Genome analysis of the unicellular green alga *Chlamydomonas reinhardtii* indicates an ancient evolutionary origin for key pattern recognition and cell-signaling protein families. Genetics 179: 193–197

Wilhelm BT and Landry JR (2009) RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. Methods 48: 249–257

Wilson A, Boulay C, Wilde A, Kerfeld CA and Kirilovsky D (2007) Light-induced energy dissipation in iron-starved cyanobacteria: roles of OCP and IsiA proteins. Plant Cell 19: 656–672

Woessner JP and Goodenough UW (1989) Molecular characterization of a zygote wall protein: an extensin-like molecule in *Chlamydomonas reinhardtii*. Plant Cell 1: 901–911

Woessner JP and Goodenough UW (1992) Zygote and vegetative cell wall proteins in *Chlamydomonas reinhardtii* share a common epitope, $(Ser\ Pro)_x$. Plant Science 83: 65–76

Woessner JP, Masson A, Harris EH, Bennoun P, Gillham NW and Boynton JE (1984) Molecular and genetic analysis of the chloroplast ATPase of *Chlamydomonas*. Plant Mol Biol 3: 177–190

Wollman F-A (1998) The Structure, function and biogenesis of cytochrome b_6f complexes. In: Rochaix J-D, Goldschmidt-Clermont M and Merchant S (eds) Molecular Biology of Chlamydomonas: Chloroplasts and Mitochondria, pp 459–476. Kluwer Academic Publishers, Dordrecht

Wollman FA (2001) State transitions reveal the dynamics and flexibility of the photosynthetic apparatus. Embo J 20: 3623–3630

Wollman FA and Lemaire C (1988) Studies on kinase-controlled *Chlamydomonas reinhardtii* which lack quinone-binding proteins. Biochim Biophys Acta 933: 85–94

Wollman FA, Minai L and Nechushtai R (1999) The biogenesis and assembly of photosynthetic proteins in thylakoid membranes 1. Biochim Biophys Acta 1411: 21–85

Worden AZ, Lee JH, Mock T, Rouze P, Simmons MP, Aerts AL et al (2009) Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes Micromonas. Science 324: 268–272

Wostrikoff K, Choquet Y, Wollman FA and Girard-Bascou J (2001) TCA1, A single nuclear-encoded translational activator specific for *petA* mRNA in *Chlamydomonas reinhardtii* chloroplast. Genetics 159: 119–132

Wostrikoff K, Girard-Bascou J, Wollman FA and Choquet Y (2004) Biogenesis of PSI involves a cascade of translational autoregulation in the chloroplast of Chlamydomonas. Embo J 23: 2696–2705

Wykoff D, Davies J and Grossman A (1998) The regulation of photosynthetic electron transport during nutrient deprivation in *Chlamydomonas reinhardtii*. Plant Physiol 117: 129–139

Xie Z and Merchant S (1998) A novel pathway for cytochromes c biogenesis in chloroplasts. Biochim Biophys Acta 1365: 309–318

Yamamoto Y, Aminaka R, Yoshioka M, Khatoon M, Komayama K, Takenaka D et al (2008) Quality control of photosystem II: impact of light and heat stresses. Photosynth Res 98: 589–608

Yang P and Smith EF (2009) The flagellar radial spokes. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 209–238. Elsevier, Amsterdam

Yeremenko N, Jeanjean R, Prommeenate P, Krasikov V, Nixon PJ, Vermaas WF et al (2005) Open reading frame *ssr2016* is required for antimycin A-sensitive photosystem I-driven cyclic electron flow in the cyanobacterium *Synechocystis* sp. PCC 6803. Plant Cell Physiol 46: 1433–1436

Yohn CB, Cohen A, Danon A and Mayfield SP (1996) Altered mRNA binding activity and decreased translational initiation in a nuclear mutant lacking translation of the chloroplast psbA mRNA. Mol Cell Biol 16: 3560–3566

Yohn CB, Cohen A, Danon A and Mayfield SP (1998a) A poly(A) binding protein functions in the chloroplast as a message-specific translation factor. Proc Natl Acad Sci USA 95: 2238–2243

Yohn CB, Cohen A, Rosch C, Kuchka MR and Mayfield SP (1998b) Translation of the chloroplast *psbA* mRNA requires the nuclear-encoded poly(A)-binding protein, RB47. J Cell Biol 142: 435–442

Yoshioka M, Uchida S, Mori H, Komayama K, Ohira S, Morita N et al (2006) Quality control of photosystem II. Cleavage of reaction center D1 protein in spinach thylakoids by FtsH protease under moderate heat stress. J Biol Chem 281: 21660–21669

Zelisko A, Garcia-Lorenzo M, Jackowski G, Jansson S and Funk C (2005) AtFtsH6 is involved in the degradation of the light-harvesting complex II during high-light acclimation and senescence. Proc Natl Acad Sci USA 102: 13699–13704

Zerges W and Hauser C (2009) Protein synthesis in the chloroplast. In: Harris EH, Stern D and Witman GB (eds) The Chlamydomonas Sourcebook, pp 967–1026. Elsevier, Amsterdam

Zhang Z, Shrager J, Jain M, Chang CW, Vallon O and Grossman AR (2004) Insights into the survival of *Chlamydomonas reinhardtii* during sulfur starvation based on microarray analysis of gene expression. Eukaryot Cell 3: 1331–1348

Zhao T, Li G, Mi S, Li S, Hannon GJ, Wang XJ and Qi Y (2007) A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. Genes Dev 21: 1190–1203

Zhao T, Wang W, Bai X and Qi Y (2009) Gene silencing by artificial microRNAs in Chlamydomonas. Plant J 58: 157–164

Zimmer SL, Fei Z and Stern DB (2008) Genome-based analysis of *Chlamydomonas reinhardtii* exoribonucleases and poly(A) polymerases predicts unexpected organellar and exosomal features. Genetics 179: 125–136

Zimmer SL, Schein A, Zipor G, Stern DB, and Schuster G (2009) Polyadenylation in Arabidopsis and Chlamydomonas organelles: the input of nucleotidyltransferases, poly(A) polymerases and polynucleotide phosphorylase. Plant J 59: 88–99

Zito F, Finazzi G, Delosme R, Nitschke W, Picot D and Wollman FA (1999) The Qo site of cytochrome b_6f complexes controls the activation of the LHCII kinase. EMBO J 18: 2961–2969

Chapter 7

Functional Genomics Dissection of Photosynthetic Mechanisms in *Arabidopsis thaliana*

Wei Chi and Lixin Zhang*
Photosynthesis Research Center, Key Laboratory of Photobiology, Institute of Botany, Chinese Academy of Sciences, Nanxincun 20, Xiangshan, Beijing 100093, China

Summary

The sequencing of the entire genome of *Arabidopsis thaliana* and the development of genomic technologies have radically changed the approaches to resolve the photosynthetic mechanisms of higher plants. Combining spectroscopic, biochemical, crystallographic, and physiological approaches, functional genomics has proven to be a highly successful approach for elucidating the functions of chloroplast genes in photosynthetic processes. The high-throughput techniques used in plant photosynthesis research include forward genetics, reverse genetics, microarrays, 2D electrophoresis, mass spectrometry and bio-informatics. We will give a brief review on the functional genomic approaches to dissect photosynthetic mechanisms in *Arabidopsis*.

* Author for correspondence, e-mail: zhanglixin@ibcas.ac.cn

R.L. Burnap and W.F.J. Vermaas (eds.), Functional Genomics and Evolution of Photosynthetic Systems, Advances in Photosynthesis and Respiration 33, pp. 177–198, DOI 10.1007/978-94-007-1533-2_7,

I. Introduction

Oxygenic photosynthesis is a crucial bioenergetic process that maintains the Earth's atmosphere and provides energy for nearly all living organisms on Earth. In plants, algae and cyanobacteria, the photosynthetic light reactions are catalyzed by a system that consists of several protein–pigment complexes embedded in the thylakoid membranes. These membrane protein complexes include photosystem II (PSII), the cytochrome b_6f complex, photosystem I (PSI) and the ATP synthase complex (Wollman et al., 1999; Nelson and Yocum, 2006). Light excitation energy is collected by the antennae of PSII and PSI and channeled to their reaction centers. At the PSII level, this leads to photooxidation of P680 chlorophyll and subsequent splitting of water with concomitant release of molecular oxygen, protons and electrons (Ferreira et al., 2004; Adachi et al., 2009). Subsequently the electrons are transferred from the plastoquinone pool to plastocyanin via the cytochrome b_6f complex and, ultimately, used to re-reduce the photooxidized PSI reaction center, which acts as a light-driven plastocyanin-ferredoxin oxidoreductase (Melkozernov et al., 2006). ATP is then synthesized from ADP and free phosphate in a process driven by the electrochemical potential across the thylakoid membrane and catalyzed by the ATP synthase complex.

In higher plants and algae, the thylakoid membranes and the soluble photosynthetic enzymes are contained in chloroplasts and are composed of subunits that are either nuclear- or chloroplast-encoded. For example, the extrinsic proteins of the PSII complex are encoded in the nucleus and must be post-translationally imported into the chloroplast lumen, where they are assembled with the intrinsic proteins of the complex. Our knowledge of the structure and functions of these protein complexes has greatly advanced over the years, but our understanding of many regulatory aspects of photosynthesis (e.g., the biogenesis and assembly of PSII) is still limited. The photosynthetic protein complexes are dynamic and able to adapt rapidly to environmental variables, such as light intensity (Allen, 1992; Wollman, 2001). The biogenesis and maintenance of the photosynthetic protein complexes depend on the coordinated action of the nuclear and chloroplast genetic systems. It is estimated that there are about 3,000 proteins in the chloroplast of *Arabidopsis*, of which only about 100 are encoded by the chloroplast genome (Abdallah et al., 2000). Therefore, the biogenesis and maintenance of the photosynthetic protein complexes are regulated mainly by nuclear-encoded auxiliary and regulatory proteins.

Genetic dissection of the function of chloroplast proteins in photosynthesis has become an exciting challenge. *Arabidopsis thaliana* has a small genome and a short life cycle compared with other flowering plants. In addition, its genome was the first to be completely sequenced of any flowering plant (The *Arabidopsis* Genome Initiative, AGI, 2000), and diverse tools have been developed that facilitate large-scale systematic investigation of gene functions in *Arabidopsis*, making it an excellent model for studying many aspects of photosynthetic and other mechanisms in plants. The advantages of using *Arabidopsis* as a model plant to analyze biological process are detailed in a review by Meinke et al. (1998).

The ability to analyze *Arabidopsis* using functional genomics tools has radically changed the biological questions posed by plant scientists and the approaches used to address them, permitting (*inter alia*) highly detailed characterization of the functions of related proteins encoded in the

Abbreviations: APX – ascorbate peroxidase; CRM – chloroplast RNA splicing and ribosome maturation; DE – dimensional electrophoresis; DHAR – dehydroascorbate reductase; EMS – ethylmethanesulfonate; ESI – electrospray ionization; FTICR – Fourier transform ion cyclotron resonance; HCF – high-chlorophyll-fluorescence; HPLC – high-performance liquid chromatography; HSP – heat-shock protein; LTQ – linear ion trap triple quadrupole; MALDI-TOF – matrix-assisted laser desorption/ionization time-of-flight; MS – mass spectrometry; MUDPIT – multidimensional protein identification technology; NDH – NAD(P)H-plastoquinone oxidoreductase; NEP – nuclear-encoded polymerase; NMR – nuclear magnetic resonance; NPQ – non-photochemical quenching; PAM – pulse amplitude modulation; PCR – polymerase chain reaction; PEP – plastid-encoded polymerase; PPR – Pentatricopeptide repeat; PSII – photosystem II; PSI – photosystem I; ROS – reactive oxygen species; RP-HPLC – reversed-phase HPLC; SDS-PAGE – sodium dodecyl sulfate polyacrylamide gel electrophoresis; SLF – sigma-like transcription factors; SOD – superoxide dismutase; SRP – signal recognition particle; TILLING – targeting-induced local lesions in genomes; TMDs – transmembrane domains; T-DNA – transferred DNA

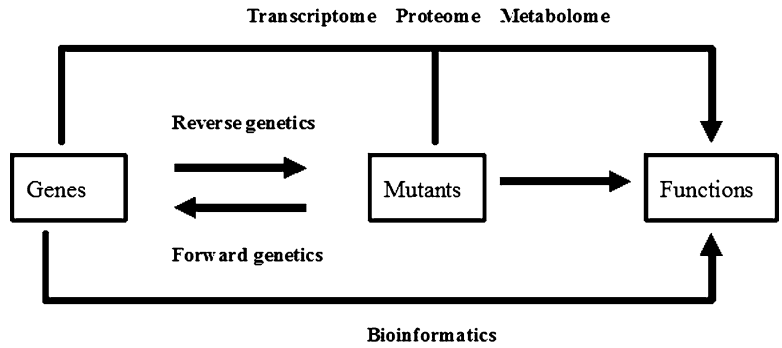


Fig. 7.1. Overview of functional genomics approaches.

genome. The term "functional genomics" is usually applied solely to large-scale reverse genetic and tagging approaches. However, in this review we apply it to an inter-related complex of high-throughput procedures that are used in plant photosynthesis research, including forward genetics, reverse genetics, microarrays, two-dimensional gel electrophoresis, mass spectrometry and bioinformatics (Neubauer et al., 1998; Somerville and Somerville, 1999; Richmond and Somerville, 2000; Yates, 2000), as illustrated in Fig. 7.1.

II. Reverse and Forward Genetics

Photosynthesis can be functionally dissected by isolating *Arabidopsis* mutants with defects in photosynthetic performance. To identify such mutants, several approaches have been applied. One forward genetics approach is to select pigmentation mutants with *albino*, bleached or pale-green phenotypes, which can be easily identified by visually inspecting plants grown in the greenhouse. Most of the affected genes in such mutants have functions in either protein translocation, e.g., Toc33, Toc159 and CAO (Jarvis et al., 1998; Klimyuk et al., 1999; Bauer et al., 2000); proteolysis, e.g., VAR1 and VAR2 (Chen et al., 2000; Sakamoto et al., 2002); enzyme activity, e.g., IM (Wu et al., 1999); protein assembly, e.g., Alb3 (Sundberg et al., 1997); or metal ion transport, e.g., IRT1 (Vert et al., 2002).

Mutants with altered chlorophyll fluorescence characteristics have also been isolated in several laboratories. For instance, Meurer et al. (1996) screened 7700 M2 progenies of ethylmethanesulfonate (EMS)-treated seeds of *Arabidopsis thaliana* and obtained 34 recessive photosynthetic mutants with high-chlorophyll-fluorescence (*hcf*) phenotypes, which were classified, on the basis of immunoblot and northern blot analyses, as having defects in PSI, PSII or the intersystem electron transport chain. Mutants of a pleiotropic nature, including a Chl *b*–deficient mutant, were also obtained in the cited study. Zhang's group screened 60,000 *Arabidopsis* activation tagging mutant lines and identified 90 *hcf* mutants (Peng et al., 2006). Other notable large-scale screenings have been performed by Niyogi et al. (1998), Shikanai et al. (1999) and Varotto et al. (2000). Niyogi's group screened 30,000 *Arabidopsis* M2 seeds sown on agar medium and identified 13 independent non-photochemical quenching (*npq*) mutants with defects in the dissipation of excess absorbed light energy (Niyogi et al., 1998). Shikanai's group screened approximately 21,000 M2 seeds (13,000 *Landsberg erecta* and 8,000 *Columbia*) by pulse-amplitude-modulation (PAM) chlorophyll fluorometry using two different actinic light intensities and identified 37 mutants, which were classified into three groups. Mutants of the first group had high F_0 fluorescence independent of photosynthetic activity, while those of the second group had significant reductions in the non-photochemical quenching formation but not in photosynthetic quantum yield. Members of the third group showed reductions in photosynthetic quantum yields, possibly due to a defect in their electron transport activity (Shikanai et al., 1999). In another forward-genetics approach, Varotto and his colleagues selected photosynthetic mutants with reduced effective PSII quantum yields and obtained 48 recessive mutants from 90,000 individual plants derived from En-Transposon and transferred DNA (T-DNA) mutants (Varotto et al., 2000).

Reverse genetics is a strategy to determine a particular gene's function by studying the phenotypes of individuals with alterations in its expression. Efficient reverse genetics is an essential component of functional genomics programs aimed at the functional characterization of large numbers of genes. The availability of large collections of *Arabidopsis* mutant lines has greatly stimulated efforts to systematically identify mutants lacking genes encoding specific chloroplast proteins and characterize their functions. Alonso et al. (2003) have determined the precise locations of more than 88,000 T-DNA insertions from over 225,000 independent *Agrobacterium* T-DNA insertion events in the genome of *Arabidopsis*, which resulted in the identification of mutations in more than 21,700 of the ~29,000 predicted *Arabidopsis* genes. In *Arabidopsis*, the identification of mutations in genes of interest can be performed by two approaches. The first is PCR screening of pooled mutant populations, at least one of which has been shown, by hybridization, to have a mutation in the gene of interest. Positive pools are deconvoluted sequentially until an individual mutant line is identified. This screening process is time-consuming, laborious and must be repeated for each gene of interest (Winkler et al., 1998; Sussman et al., 2000). An alternative approach is to sequence regions flanking insertion sites in individual plants from large insertion-mutant populations, thereby determining large numbers of insertion sites in advance. This approach eliminates the laborious and time-consuming process of PCR-based screening and deconvolution of pools (Parinov et al., 1999; Tissier et al., 1999). Other popular methods for identifying mutant lines with defects in the functions of genes of interest can also be used in reverse genetics, such as antisense, co-suppression and RNAi techniques. Further, for plastid-encoded genes, mutant alleles can be generated by homologous recombination, which leads to the replacement of wild-type genes by knockout alleles. This approach has been applied in tobacco (Svab et al., 1990) and the green alga *Chlamydomonas* (Boynton et al., 1988), but it has still not been successfully applied to *Arabidopsis*.

Forward and reverse genetics have proved to be powerful tools for the identification and functional characterization of novel genes involved in *Arabidopsis* photosynthesis. Genetic analysis of *Arabidopsis* mutants has revealed the existence of a surprisingly large number of nucleus-encoded factors that are involved in the regulation of chloroplast gene expression. Mutants with alterations in plastid gene transcription, RNA stability, RNA processing, RNA editing, splicing, translation initiation and elongation, and protein assembly have been isolated and characterized. Here, we review progress over the years on the identification of some key regulatory components involved in plastid gene expression and biogenesis of the photosynthetic machinery by forward and reverse genetic approaches in *Arabidopsis*. Some important regulatory components isolated by other approaches are not included here, but they are described in related references.

A. Transcription

Transcription has been shown to be a prime level of gene regulation in plastids of land plants and to be much more complex in higher plants than in photoautotrophic prokaryotes. Two types of polymerases responsible for RNA transcription have been identified in higher plant chloroplasts: plastid-encoded polymerases (PEPs) and nuclear-encoded polymerase (NEPs) (Hajdukiewicz et al., 1997; Hedtke et al., 1997; Maliga, 1988). PEPs contain core subunits encoded by *rpoA*, *rpoB*, *rpoC1* and *rpoC2* genes, while NEPs are each composed of a single subunit (Hedtke et al., 1997; Liere and Maliga, 1999). These two types of polymerases are responsible for the transcription of distinct sets of chloroplast genes (Allison et al., 1996; Hajdukiewicz et al., 1997). The chloroplast-encoded photosynthetic genes (such as *psbA*, *psbD* and *rbcL*) are exclusively transcribed by PEPs, a few genes (mostly encoding components of the transcription/translation apparatus, such as *rpoB* and *accD*) are exclusively transcribed by NEPs, and non-photosynthetic housekeeping genes are transcribed by both PEPs and NEPs. It is generally accepted that NEP activities are largely responsible for the transcription of housekeeping genes during early stages of chloroplast development and that PEP activities increase while NEP activities decrease during subsequent chloroplast development (Hanaoka et al., 2005). Therefore, NEP activities play important roles in the early stages of chloroplast development. Accordingly, mutations in RpoT2, a NEP polymerase that is dually

targeted to chloroplasts and mitochondria, result in only moderate defects in chloroplast gene expression and delayed greening of the leaves (Baba et al., 2004; Courtois et al., 2007), while insertion mutants and transgenic lines with low levels of the solely chloroplast-targeted NEP polymerase reportedly exhibit more severe defects in chloroplast development and gene expression (Hricová et al., 2006; Courtois et al., 2007; Swiatecka-Hagenbruch et al., 2008). PEP is a multisubunit (bacterial-type) polymerase. In addition to the chloroplast-encoded catalytic core subunits, a number of polypeptides have been shown to be associated with transcriptionally active chloroplast chromosomes (Pfalz et al., 2006). Inactivation of *pTAC-2*, *-6*, and *-12* resulted in a seedling-lethal phenotype, resembling those of *rpo* mutants, clearly indicating that they are involved in plastid gene expression (Pfalz et al., 2006).

The activity of the PEP core enzyme is regulated by sigma-like transcription factors (SLFs), six of which (SIG1–SIG6) have been described in *Arabidopsis*. The functions of these sigma factors have been examined by characterizing specific sigma knockout plants, which has indicated that these SLFs have overlapping as well as specific functions (Homann and Link, 2003; Privat et al., 2003). A SIG2-PEP holoenzyme specifically transcribes some of the tRNA genes and the *psaJ* gene (Kanamaru et al., 2001; Nagashima et al., 2004). SIG5 plays an important role in recognizing the blue light–dependent promoter of the *psbD* gene (Tsunoyama et al., 2002, 2004; Nagashima et al., 2004) and may also have specific functions during embryogenesis (Yao et al., 2003). SIG6 plays a more general role during early plastid differentiation and plant development (Ishizaki et al., 2005), and SIG4 is specifically required for *ndhF* gene transcription (Favory et al., 2005). SIG3 specifically transcribes the *psbN* gene in plastids (Zghidi et al., 2007). The activities of sigma factors are controlled by their phosphorylation modification state. Studies on transgenic *Arabidopsis* plants with or without the putative phosphorylation sites for SIG1 revealed that the phosphorylation of SIG1, which is regulated by redox signals, selectively inhibits the transcription of the *psaA* gene encoding a PS-I protein (Shimizu et al., 2010). Analysis of the *Arabidopsis sig6* mutants complemented with intact and mutant sigma cDNAs uncovered several potential phosphorylation sites in the unconserved region of SIG6, which could have widely different effects on the visual and/or molecular phenotypes (Schweer et al., 2010). Chi et al. identified a PPR protein, which is involved in the regulation of transcription of PEP-dependent chloroplast genes in *Arabidopsis* cotyledons by analysis of a *virescent* mutant *dg1* (Chi et al., 2008). The function of DG1 is accomplished *via* its interaction with SIG6 (Chi et al., 2010), which might indicate an additional regulatory pathway for SIG6 activity besides phosphorylation.

B. RNA Processing and Editing

Many chloroplast genes are organized in multiple transcription units that are transcribed into polycistronic transcripts. The translatable RNAs of these genes are generated by extensive RNA processing, which involves the splicing of introns, endonucleolytic cleavage in intercistronic regions, exonucleolytic and/or endonucleolytic trimming of RNA 5′ and 3′ ends, polyadenylation and RNA editing (Barkan and Goldschmidt-Clermont, 2000). Little is known about the auxiliary and regulatory factors that are needed to catalyze these reactions. However, *hcf* mutants with defective plastid RNA processing have been isolated, and the genes involved have been investigated. Westhoff's group isolated two *hcf* mutants, *hcf107* and *hcf152*, in which processing of chloroplast *psbB-psbT-psbH-petB-petD* transcripts was affected (Felder et al., 2001; Meierhoff et al., 2003). In the *hcf107* mutant, accumulation of processed *psbH*-containing RNAs is impaired and *psbH* transcripts with a −45 5′ end are absent, whereas processing at the *psbH* 3′ site proceeds normally. In the *hcf152* mutant, levels of spliced *petB* RNAs are reduced, and the accumulation of transcripts cleaved between the genes *psbH* and *petB* is affected. HCF152 encodes a pentatricopeptide repeat (PPR) protein with 12 putative PPR motifs and is located inside the chloroplast, consistent with its role in plastid RNA processing. Another conserved pentatricopeptide repeat protein, MRL1, was found to be necessary for the production/stabilization of the processed transcript of *rbcL*, presumably by acting as a barrier to 5′ > 3′ degradation, although the *Arabidopsis mrl1* mutant retains normal levels of the primary transcript (Johnson et al., 2010).

Analyses of chloroplast splicing mutants of *Arabidopsis* and other eukaryotes have led to the identification of several nuclear-encoded factors involved in regulating this process, including a family of proteins with a novel RNA-binding domain, denoted the chloroplast RNA splicing and ribosome maturation (CRM) domain (Barkan, 1998; Till et al., 2001; Ostheimer et al., 2003). Four CRM proteins have been shown to interact with chloroplast introns in chloroplasts and to promote their splicing: CAF1, CAF2, CRS1, and CFM2. *Arabidopsis* and maize mutants associated with these four proteins have been studied (Asakura and Barkan, 2006, 2007; Asakura et al., 2008). In general, CAF1, CAF2 and CRS1 appear to promote the splicing of plastid group II introns in both maize and *Arabidopsis*, although there are some differences in their activities between the two species. For example, CAF1 promotes the splicing of two group II introns, *rpoC1* and *clpP-intron 1*, that are found in *Arabidopsis* but not in maize (Asakura and Barkan, 2006). The *Arabidopsis* T-DNA insertion mutant of CFM2 was found to be impaired in splicing the *trnL* group I intron and the *ndhA*, *ycf3-int1*, and *clpP-int2* group II introns (Asakura and Barkan, 2007). These findings indicate that CFM2 functions dually in group I and group II intron splicing in plant chloroplasts (Asakura and Barkan, 2007). de Longevialle et al. recently found OPT51, a PPR protein with two LAGLIDADG motifs, that specifically promotes the splicing of the only group II intron. In the *Arabidopsis* mutant without OPT51, the splicing of *ycf3* intron 2 was defective (de Longevialle et al., 2008).

Another rather surprising finding was RNA editing in land plant chloroplasts. RNA editing is a process in which some plastid transcripts are posttranscriptionally modified by C-to-U conversion (Shikanai, 2006). Thirty-four sites are known to be edited in transcripts of eighteen genes in *Arabidopsis* plastids (Chateigner-Boutin and Small, 2007), and several nuclear-encoded factors responsible for specific RNA editing events have been discovered in genetic studies of *Arabidopsis* photosynthetic mutants. The *Arabidopsis crr4* and *crr21* mutants are specifically impaired in the RNA editing of sites 1 (*ndhD-1*) and 2 (*ndhD-2*) *ndhD* transcripts (Kotera et al., 2005; Okuda et al., 2007), which encodes a subunit of the NAD(P)H-plastoquinone oxidoreductase (NDH) complex. Studies on two yellow-phenotype mutants with impaired chloroplast development revealed that CLB19 is required for RNA editing of *rpoA* and *clpP* transcripts (Chateigner-Boutin et al., 2008) and YS1 is required for editing *rpoB* transcripts and the rapid development of chloroplasts during early growth (Zhou et al., 2008). Okuda et al. (2009) investigated two RNA editing mutants and reported that CRR22 and CRR28 are involved in multiple RNA editing events. All of these plastid proteins involved in RNA editing are members of the PPR family. The PPR family is one of the largest protein families in plants, which contains 450 PPR proteins in *Arabidopsis* (Lurin et al., 2004). The PPR family is divided into P and PLS subfamilies and the PLS subfamily is further divided into PLS, E and DYW subclasses based on the presence of E or DYW domains in the C-terminal sequences (Lurin et al., 2004). It is generally accepted that that PPR proteins with E and DYW domains are essential *trans*-factors for recognizing RNA editing sites and that different DYW family members may have acquired distinct functions in the evolution of RNA editing (Okuda et al., 2009). Many such PPR proteins, including LPA66, AtECB2, RARE1 and OTP80-86, were shown to be RNA-editing specificity factors by investigation of their corresponding *Arabidopsis* mutants. Loss of one of these factors leads to a specific defective editing of one or, at most, a few sites of plastid transcripts (Cai et al., 2009; Robbins et al., 2009; Yu et al., 2009; Hammani et al., 2009; Tseng et al., 2010; Okuda et al., 2010).

C. Translation

Translational regulation plays an important role in determining the levels of chloroplast proteins. Because the translational machinery is most closely related to that of eubacteria, chloroplast homologues of many components of the eubacterial translational system have been found in nuclear and chloroplast genomes of plants, e.g., 5S, 16S, 23S rRNAs, tRNAs, initiation factors, elongation factor EF-1A (EF-Tu), and ribosomal proteins (Harris et al., 1994; Sugiura et al., 1998). The application of various *in vitro* and *in vivo* approaches has permitted the identification of several *cis*-acting determinants in chloroplast RNAs for translation initiation, which probably represent the target sites for translational regulatory factors

(Danon, 1997; Zerges, 2000). However, to date, only a few nuclear genes whose products play a role in protein synthesis in the chloroplast have been identified by genetic approaches. Schult et al. (2007) characterized the high-chlorophyll-fluorescence mutant *hcf173* of *Arabidopsis thaliana*, which is impaired in the accumulation of PSII subunits. *In vivo* protein-labeling experiments showed that synthesis of the PSII reaction center protein D1 is dramatically reduced in this mutant. Furthermore, northern blot and polysome association experiments showed reductions in translation initiation and accumulation of the corresponding *psbA* mRNA. HCF173 is distantly related to the superfamily of the short-chain dehydrogenases/reductases and is mainly associated with the membranes in a high-molecular-mass complex. These results suggest that HCF173 is essential for translation of the *psbA* mRNA (Schult et al., 2007).

D. Translocation

Only about 100 chloroplast proteins are encoded by plastid genes. The others are encoded by nuclear genes and translated on cytosolic ribosomes as precursors containing an N-terminal transit peptide that is required for their entry into the chloroplasts, after which the transit peptide is cleaved. Some nuclear-encoded thylakoid polypeptides contain a bipartite transit peptide with an N-terminal part required for import into the chloroplast and a C-terminal part required for integration in, or transport across, the thylakoid membrane. There are at least four distinct translocation pathways: spontaneous protein insertion, a pathway related to the bacterial Sec pathway, the signal recognition particle pathway, and the pH gradient–dependent pathway (Robinson et al., 2001). To date, three membrane proteins (TatA, TatB, and TatC) have been identified that are required for Tat-dependent protein translocation in the plant system. Motohashi et al. (2001) isolated four transposon-inserted *albino Arabidopsis* mutants with mutations in the same gene, named *albino* and pale-green 2 (*apg2*), and showed that *APG2* encodes a chloroplast protein with similarity to bacterial TatC containing six putative transmembrane domains. Examination of *apg2* plastids revealed that they were highly vacuolated, lacked internal membrane structures and lamellae of the thylakoid membrane, and contained many densely staining globular structures, like undifferentiated proplastids, suggesting that *APG2* plays an essential role in chloroplast development (Motohashi et al., 2001). However, the precise function of APG2 in Tat-dependent protein transport remains to be determined.

E. Assembly

After synthesis and proper targeting to the thylakoid membrane, the subunits of the photosynthetic complexes must be coordinately inserted into the membrane and assembled into functional units, together with numerous pigments and redox cofactors. This raises important questions about how photosynthetic complexes are assembled. Due to the intricacy of photosynthetic complexes, the assembly of their subunits into functional units in the thylakoid membrane is likely to require the participation of many regulatory factors (Barkan and Goldschmidt-Clermont, 2000; Zhang and Aro, 2002; Rochaix, 2004). Genetic approaches have provided some insights regarding the processes involved and have shown a number of nucleus-encoded factors to be involved in the assembly of PSII. The first factor involved in *Arabidopsis* PSII assembly to be identified was discovered by studying the *hcf* mutant *HCF136*, which is essential for the stable assembly of PSII (Meurer et al., 1998). Protein-labeling analyses have shown that plastid-encoded proteins are synthesized in the *hcf136 Arabidopsis* mutant, but the assembly of PSII reaction centers is blocked, and no stable PSII complexes appear to accumulate in it. Characterizations of three *low PSII accumulation* (*lpa*) mutants of *Arabidopsis* have revealed two novel factors involved in PSII assembly (Peng et al., 2006; Ma et al., 2007; Cai et al., 2010). In these mutants, the assembly of PSII was found to be less efficient than in wild-type plants. LPA1 appears to be involved in efficient PSII assembly through direct interaction with the PSII reaction center protein D1, and LPA2 is likely to be involved in assisting CP43 assembly within PSII. LPA3, together with the previously identified LPA2 (Ma et al., 2007), may form a complex that regulates the incorporation of CP43 into PSII. These two proteins form a complex with Alb3, which belongs to the Oxa1p/YidC protein family and is involved in thylakoid membrane

biogenesis and cell division. Alb3 may function in assisting the PSII assembly processes through interactions with LPA2 and LPA3 (Cai et al., 2010). FKBP-2, an immunophilin of the chloroplast lumen, has been shown to function in the accumulation of PSII supercomplexes (Lima et al., 2006). Inactivation of *FKBP-2* resulted in elevated levels of PSII monomers and dimers and reduced accumulation of PSII supercomplexes. Another immunophilin of the cyclophilin type, CYP38, was also shown to be required for the assembly and stabilization of PSII in *Arabidopsis* (Fu et al., 2007; Sirpiö et al., 2008). HCF101 has also been found in the biogenesis of PSI by Stöckel and Oelmüller (2004). In *hcf101* mutants, photosystem I subunits are synthesized but do not assemble into a stable complex.

F. Vesicular Transport

The photosynthetic protein complexes are embedded in a lipid matrix with 70–80% galactolipids. The galactolipids are synthesized at the inner envelope membrane, and there is likely an intraorganellar lipid transport system that ensures the transfer of lipids from their synthesis site to the thylakoids. Further, there is growing evidence that vesiculation of the envelope's inner membrane provides the pathway through which lipids are transported from the inner envelope to the thylakoids (Vothknecht et al., 2001; Westphal and Soll 2001). Vesicle budding from the inner envelope and accumulation of vesicles in the chloroplast stroma has been observed under certain conditions, such as cold treatment of leaves and mild heat-shock of *yellow* mutants of *Chlamydomonas* that are unable to synthesize chlorophyll in the dark. However, little was known about the molecular components of this vesicular transport system in *Arabidopsis* until the discovery of a mutant of the vesicle-inducing protein in plastids (VIPP): a pleiotropic *hcf* mutant of *Arabidopsis* (*hcf155*) identified by Kroll et al. (2001), which provides a genetic entry point for the analysis of plastid vesicular transport. Electron micrograph analysis showed that vesicles bud from the inner envelope in the plastids of wild-type *Arabidopsis* plants. However, no vesicles accumulate in the *hcf155* mutant upon exposure to low temperatures. The lack of vesicle formation in *hcf155* is paralleled by the block of thylakoid formation. The VIPP1 protein could be involved in transport from the envelope to the thylakoids, but it does not show any similarity to known proteins of the cytoplasmic vesicular transport system (Kroll et al., 2001). Instead, it appears to be unique to thylakoid formation in chloroplasts and cyanobacteria and may have evolved early in cyanobacterial evolution.

III. Transcriptomics

The aim of transcriptomics is to quantify the expression levels of genes of an organism during different developmental stages or in response to certain stimuli (Brown and Botstein, 1999; Duggan et al., 1999). The availability of the complete genome sequence and the collections of large numbers of expressed sequence tags enable large-scale gene expression analyses in *Arabidopsis*. One of the key tools for this is DNA microarray technology. Briefly, DNA microarrays are constructed by arraying thousands of target DNA fragments (cDNAs, genomic clones or oligonucleotides) in a grid format onto miniature solid supports such as polylysine-coated glass slides. They are then probed with a complex mixture of fluorescently labeled fragments made from polyA mRNA or total RNA (Schaffer et al., 2000). The development of successful microarray technologies has involved synergistic advances in methods for spotting probes, hybridization, the detection of hybridized targets and analysis of acquired data. Diverse DNA analysis systems have now been developed and commercialized. Notably, in the context of *Arabidopsis* research, the Affymetrix 22K ATH1 oligonucleotides array contains 22,500 probe sets and is currently widely used for the global evaluation of gene expression in *Arabidopsis* (Michael, 2002). Customized DNA microarrays for studying photosynthesis-related gene expression profiles have also appeared (Kurth et al., 2002; Biehl et al., 2005).

Photosynthetic organisms must acclimate to changes in light intensity in their environments (Boardman, 1977). Under high-light conditions, if energy supplied by light harvesting and electron transport exceeds its dissipation by CO_2 fixation and other energy-demanding processes, the photosynthetic electron transport components may become excessively reduced. This may result in excess production of reactive oxygen species

(ROS), leading to severe damage to many cellular components and processes (Karpinski et al., 1997; Huner et al., 1998; Foyer and Allen, 2003). However, plants have evolved the ability to sense such changes and communicate information about them that stimulate changes in the expression of both nuclear and chloroplast genes and thus avoid serious damage (provided that the stress is not too severe). DNA microarray analysis may facilitate the simultaneous identification of genes that play currently unknown roles in these responses and the signal transduction pathway(s) and greatly aid in elucidating the molecular mechanisms involved (Rossel et al., 2002; Kimura et al., 2003; Richly et al., 2003; Vanderauwera et al., 2005).

Rossel et al. (2002) monitored changes in transcript levels of *Arabidopsis* genes after exposure to high-light conditions. They detected increases in the expression of genes encoding proteins with antioxidant functions, such as APX1 and dehydroascorbate reductase (DHAR), as well as genes homologous to known regulatory genes and metabolic enzymes. In addition to many genes involved in detoxification, photoacclimation, protein folding, and de-aggregation, a substantial subset of genes that were found to be differentially expressed are also involved in other stress-sensing responses, indicating that there may be some common regulators or pathways for abiotic stress (Rossel et al., 2002). Besides the identification of novel genes regulated by high light, microarrays can also be useful for determining the novel functions of known genes. For instance, using the Affymetrix *Arabidopsis* 24 K GeneChip, Kleine et al. (2007) showed that 77 high light–responsive genes are regulated via CRY1, indicating that blue light photoreceptor CRY1 has a previously unsuspected role in mediating plant responses to high irradiances (Kleine et al., 2007). Chen et al. (2010) found that inactivation of any one of small chloroplast-targeted DnaJ proteins (AtJ8, AtJ11 and AtJ20) triggered a global stress response but both specific- and cross-talk functions exist between these small chloroplast-targeted DnaJ proteins under high light conditions (Chen et al., 2010).

DNA microarray approaches have also been widely used to analyze the regulation of gene expression by redox signals in chloroplasts, notably using GST (gene sequence tag) DNA microarrays with probes for typical nuclear genes encoding photosynthesis-related chloroplast proteins (Fey et al., 2005). Out of 2,133 identified light-regulated genes, 286 appear to be directly regulated by redox signals from the photosynthetic electron transport chain; 86 were up-regulated by a reduction signal, while 200 were down-regulated by it. Furthermore, 232 genes out of these 286 redox-regulated genes appear to be regulated by more than one redox variable (Fey et al., 2005). In addition, genome-wide microarrays have provided ways to assess the specificity of ROS signaling by analyzing expression data generated from ROS-related microarray experiments with samples of plants that were treated by oxidative stress-causing agents (methyl viologen, *Alternaria alternata* toxin, 3-aminotriazole, and ozone) (op den Camp et al., 2003; Gechev et al., 2004; Gechev and Hille, 2005; Vanderauwera et al., 2005) and from genetic mutants and transgenic plants, in which the activity of an individual antioxidant enzyme was perturbed (catalase, cytosolic ascorbate peroxidase or copper/zinc superoxide dismutase) (Rizhsky et al., 2003; Davletova et al., 2005; Umbach et al., 2005; Vanderauwera et al., 2005). In addition, Gadjev et al. (2006) have presented a broad integrative comparison of ROS-regulated gene expression profiles in *Arabidopsis*, based on analyses of changes in the abundance of nearly 26,000 transcripts in response to different ROS, including three-, four-, or fivefold increases in levels of 8,056, 5,312, and 3,925 transcripts, respectively.

It should be noted that several difficulties and possible artifacts might be associated with the application of this DNA microarray technique. First, some genes, such as those encoding regulatory components (e.g., signal transduction and transcription factors), are of great interest, but they are often not detected due to their low abundance. Second, highly similar genes cannot be distinguished, so the presence of highly similar genes may result in cross-hybridization and hence misleading results. Third, shifts in abundance due to changes in synthesis rates and changes in degradation rates cannot be distinguished with DNA microarray analysis (Hihara et al., 2001), although they could be resolved by performing microarray analyses in the presence or absence of inhibitors of RNA synthesis.

IV. Proteomics

The aims of *Arabidopsis* plastid proteomics analysis are to identify the proteins and isoforms that are expressed in the plastid of *Arabidopsis*; to determine the post-translational modifications that occur in these proteins; to elucidate differences in these patterns under different developmental, physiological, and stressful conditions; and to synthesize the information acquired to develop a better understanding of the molecular basis of photosynthesis and other processes in plastids. Briefly, plastid proteomics includes determination of protein expression levels, protein–protein interactions and sub-chloroplast localizations (Tyers and Mann, 2003).Advances in proteomics and mass spectrometry (MS) have enabled the systematic identification of proteins in plants, as demonstrated in several studies. Nevertheless, although modern MS has very high sensitivity, identification of low-abundance proteins is still relatively difficult and may require multistep fractionation and concentration procedures. Membrane proteomes are especially challenging in this respect, and very few membrane proteomes have been successfully analyzed (van Wijk, 2004). However, techniques and media for reversed-phase high-performance liquid chromatography (HPLC) and gel filtration of intact proteins, using various solvent systems, are rapidly emerging that are capable of purifying different integral membrane proteins and separating abundant members of the thylakoid proteome. Fractionation with organic solvents, followed by one-dimensional discontinuous sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) has also been used to reduce membrane proteome complexity and to remove soluble proteins from extracts of integral membrane proteins. Such organic solvent fractionation has proven to be useful because it enriches the samples in hydrophobic proteins. Developments in chloroplast proteomics approaches have been reviewed by van Wijk (2004) and Baginsky and Gruissem (2006).

An intact plastid of *Arabidopsis* is composed of four compartments: (a) the outer and inner envelope membranes surrounding it, (b) the soluble stroma within it, (c) the thylakoid membrane, and (d) the thylakoid lumen. Each of these compartments has its own subset of proteins, or subproteome. In addition, proteins can be present in different sub-chloroplast locations, e.g., associated with thylakoid and envelope membranes. These chloroplast sub-proteomes have been investigated using a variety of approaches, with diverse purification and fractionation techniques. We describe attempts to characterize the proteomes of these plastid compartments in *Arabidopsis* below.

A. The Envelope Proteome

Located at the interface between the stroma and the cytosol, the envelope forms the barrier between the cytosol and the chloroplast stoma. It is also the site of transport and exchange of various ions and metabolites required for integration of the plastid metabolism within the plant cell (Neuhaus and Wagner, 2000; Weber et al., 2000). Several proteomic approaches have been described that allow the identification of plastid envelope proteins in *Arabidopsis* (Seigneurin-Berny et al., 1999; Ferro et al., 2002, 2003; Froehlich et al., 2003).

Ferro et al. (2003) reported a procedure to prepare highly purified envelope membranes from *Arabidopsis* chloroplasts, in which envelope proteins are extracted using several methods, e.g., chloroform/methanol extraction and alkaline or saline treatments, and each envelope sub-fraction was analyzed by liquid chromatography–tandem mass spectrometry (Ferro et al., 2003). More than 100 proteins were identified, of which more than 50% have one or more predicted transmembrane domains (TMDs). These include ion and metabolite transporters; proteins involved in vitamin, pigment, fatty acid and glycerol lipid metabolism; components of the Tic and Toc complex involved in protein import from the cytosol; and proteins of unknown functions (Ferro et al., 2003). Froehlich et al. (2003) used a high-throughput mass spectrometry–based proteomics approach to identify and characterize membrane proteins localized to the *Arabidopsis thaliana* chloroplast envelope membrane. *Arabidopsis* chloroplasts were isolated using a novel isolation procedure, and "mixed" envelopes were subsequently isolated using sucrose step gradients. Then, total envelope proteins were analyzed by two alternative methods: off-line multidimensional protein identification technology (MUDPIT) and one-dimensional gel electrophoresis followed by proteolytic digestion and liquid chromatography coupled with tandem mass spectrometry (Gel-C-MS/MS). This

proteomic procedure enabled the identification of 392 non-redundant proteins, of which 113 had more than one predicted TMD (Froehlich et al., 2003). Altogether, 429 proteins have been identified in the plastid envelope, including some that are preferentially located in the stroma and thylakoid.

B. The Lumen Proteome

Key functions of the lumen are to accumulate protons and balance the ion current across the thylakoid membrane during photosynthetic electron transfer processes in thylakoids. Until the 1990s, few proteins were known to be localized in the lumen, although this small set included three important extrinsic photosystem II proteins (PsbO, PsbP, and PsbQ) and plastocyanin. Several other lumenal proteins were identified in the 1990s, including violaxanthin de-epoxidase (Hager and Holocher, 1994), polyphenol oxidase (Sommer et al., 1994), the extrinsic photosystem I protein PsaN (He and Malkin, 1992), and the carboxyl-terminal processing protease for the D1 protein (Oelmüller et al., 1996). Analysis of the thylakoid lumen proteome has greatly benefited from the development of techniques to isolate the lumen proteins (Peltier et al., 2000, 2002; Schubert et al., 2002). Using such techniques, Schubert et al. (2002) carried out a systematic characterization of the thylakoid luminal proteins from *Arabidopsis*, showing that the thylakoid lumen has a specific proteome, of which 36 proteins were identified. Besides a large group of peptidyl-prolyl *cis-trans* isomerases and proteases, they detected a family of novel PsbP domain proteins. Of 36 luminal precursors, 19 had a twin-arginine motif for import via the Tat pathway (Schubert et al., 2002). In another notable analysis, Peltier et al. (2002) characterized 81 thylakoid luminal proteins using a unique combination of experimentation and localization predictions.

C. The Stroma Proteome

The stroma is both the site of carbon fixation, where the Calvin–Benson cycle takes place, and the site of synthesis of many essential compounds, such as plant hormones, fatty acids and lipids, amino acids, vitamins, purine and pyrimidine nucleotides, tetrapyrroles, and nitrogen and sulfur assimilation. The stroma also contains the chloroplast gene expression system, including the plastid transcriptional and translation machinery.

An analysis of the stromal proteome from highly purified chloroplasts of *Arabidopsis* has been presented by Peltier et al. (2000, 2006), in which purified intact chloroplasts were lysed under non-denaturing conditions, and chloroplast stromal proteins and protein complexes were separated based on native mass using CN (Colorless-Native)-PAGE and subsequently SDS-PAGE. In total, 241 non-redundant proteins were identified and classified according to their biological functions (Peltier et al., 2006). Proteins related to chloroplast protein synthesis, biogenesis, and degradation represented nearly 10% of the total identified stroma proteins. Of the total protein mass, proteins involved in the oxidative pentose phosphate pathway, glycolysis and the Calvin cycle accounted for about 75%, nitrogen assimilation 5–7%, and all other pathways, such as biosynthesis of fatty acids, amino acids, nucleotides, tetrapyrroles and vitamins B1 and B2, each <1%. Several proteins with diverse functions other than primary organic carbon metabolism, such as the isomerase ROC4, lipoxygenase 2 (involved in jasmonic acid biosynthesis), and a carbonic anhydrase (CA1), were surprisingly abundant, in the range of 0.75–1.5% of the total stromal proteome (Peltier et al., 2006). Interestingly, about 60% of the proteins were found to be present as homo-oligomeric complexes or monomers rather than heteromeric complexes (Peltier et al., 2006).

D. The Thylakoid Proteome

The chloroplast of *Arabidopsis* contains photosynthetic thylakoid membranes with four multisubunit protein complexes (PSI, PSII, ATP synthase, and cytochrome b_6f complexes), each with multiple cofactors. These four complexes are composed of at least 70 different proteins that are involved in the photosynthetic reactions (Aro et al., 2005). In addition to the abundant proteins of the photosynthetic apparatus, thylakoid membranes also contain proteins with unknown functions, which might be involved in the regulation of biogenesis and maintenance of photosynthetic proteins (Rochaix, 2004). Such proteins are expected to be present at much lower levels than the structural proteins of the photosynthetic

complexes and hence challenging to be identified by proteomic approaches.

The *Arabidopsis* thylakoid membrane proteome, including lumenal and/or peripheral proteins associated with the membranes, has been analyzed by traditional 2-DE/matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) MS, and off- or online electrospray ionization (ESI)-MS/MS and Edman degradation sequencing (Kieselbach et al., 1998; Peltier et al., 2000, 2002; Friso et al., 2004). Friso et al. (2004) identified 17 novel integral membrane proteins that are likely to play important roles in thylakoid biogenesis. In addition, they identified 83 known proteins of the photosynthetic apparatus, including several new paralogues; 20 proteins involved in protein insertion, assembly, folding, or proteolysis; and 16 involved in translation.

Plastid proteomics studies have generated large amounts of detailed, interrelated data that have been compiled in several proteome databases, including two dedicated to plant plastids (PPDB at http://ppdb.tc.cornell.edu/ and 'plprot' at htpp://www.plprot.ethz.ch), through which plastid proteomics data can be easily accessed (Kleffmann et al., 2006; Sun et al., 2009).

E. Plastid Protein Expression Profiling by Proteomics

Proteomics is used routinely to study global protein changes associated with cellular responses to both abiotic and biotic factors, as well as developmental cues. Such comparative proteomics are challenging because of the experimental difficulties in obtaining sufficient dynamic resolution of the stromal proteome and the challenging nature of the hydrophobic membrane proteomes. However, interesting findings have been obtained in these analyses, as outlined below.

Comparative proteomics approaches have been used to analyze responses to high-light stress in *Arabidopsis* by Phee et al. (2004), who identified 64 proteins as candidate factors that responded to high-light stress, and the expression of 35 of them was found to decrease during high-light stress. Most of the proteins that were down-regulated during high-light stress are involved in photosynthesis pathways. In contrast, 14 up-regulated proteins were identified as high-light stress–related proteins that had already been well characterized, such as heat-shock proteins (HSPs), DHAR, and SOD. In another proteomic analysis, Giacomelli et al. (2006) examined the quantitative response of the thylakoid-associated proteome of wild type and ascorbate-deficient *vtc2-2* mutant *Arabidopsis* plants after exposure to high light, using two-dimensional gels followed by tandem mass spectrometry. Overall, 45 proteins were significantly differentially expressed in relation to genotype, light treatment, or both. The most significant responses were the up-regulation of thylakoid YCF37, likely involved in photosystem I assembly, and specific fibrillins, a flavin reductase-like protein, and an aldolase, all located in thylakoid-associated plastoglobules. The expression of Fe-SOD was down-regulated in *vtc2-2*, while Cu,Zn-SOD was up-regulated. The levels of a number of other stress-related proteins, such as thylakoid proteases and lumenal isomerases, did not change, while those of PsbS increased in wild-type plants upon light stress. These comparative proteomic findings may facilitate understanding the multiple proteins and functional network involved in antioxidative defense.

Kubis et al. (2003) also used 2-DE gels to compare total chloroplast proteins from an *Arabidopsis* mutant disrupted in *Toc33*, an important component of the protein import apparatus in the outer envelope. These authors showed that two abundant soluble subunits of the water-splitting complex (PsbP and PsbO) were less abundant, whereas several chaperones (a Hsp70, a Cpn60 and a HSP90) were apparently up-regulated in the mutant. Finally, Rutschow et al. (2008) compared total leaf proteomes of young seedlings and chloroplast proteomes of fully developed leaves from *Arabidopsis* cpSRP54 (chloroplast signal recognition particle 54) mutant *ffc1-2*, finding that cpSRP54 deletion led to changes in light-harvesting complex composition, increases in PsbS levels, and reductions in the PSI/PSII ratio. Zybailov et al. (2009) used large-scale comparative proteomics to study *clpr2-1* and wild-type leaf proteomes of young seedlings and suggested the important roles of Clp proteolysis in the regulation of chloroplast homeostasis and metabolism.

F. Protein–Protein Interactions

There must be a very large number of stable and transient interactions between the proteins in the

chloroplast, and the proteomics approach provides the powerful tools to analyze such protein–protein interactions. In chloroplasts, three types of thioredoxins (*f*, *m* and *x*) are present, which play important roles in regulating enzymatic activities, but few of their targets were known (Collin et al., 2003). However, Balmer et al. (2003) identified many putative targets of the chloroplast thioredoxins *f* and *m* using an innovative proteomics approach. They identified 15 potential targets that function in 10 chloroplast processes that were not known to be thioredoxin-linked. These targets include proteins that seem to function in plastid-to-nucleus signaling and in a previously unrecognized type of oxidative regulation. They also identified 11 previously unknown and nine confirmed target proteins that are members of pathways known to be regulated by thioredoxin (Balmer et al., 2003). Recent studies have shown that Trx targets are also present in thylakoid lumen. Using complementary proteomics approaches, Hall et al. (2010) identified 19 Trx target proteins, covering more than 40% of the currently known lumenal chloroplast proteome (Hall et al., 2010).

The identification of protein complexes in chloroplasts is important for understanding the interactions amongst them. Several chloroplast protein complexes have been identified to date, including a heteromeric Clp protease complex of 325–350 kDa with 11 different proteins (Peltier et al., 2004), the 200–240-kDa heterotetrameric ADP-glucose pyrophosphorylase (or glucose-1-phosphate adenylyltransferase) (Kavakli et al., 2002), the stromal signal recognition particle (Schunemann, 2004), and the 150-kDa heterotetrameric tryptophan synthase (Radwanski et al., 1996). In the Clp protease complex, ten different subunits have been identified, of which ClpR1, 2, and 3 form part of the heptameric rings, while ClpS1 is a regulatory subunit positioned at the axial opening of the ClpP/R core (Peltier et al., 2004).

G. Post-translational Modifications

Post-translational modifications play important roles in the regulation of protein functions. Modifications of hydrophobic moieties can help to anchor proteins to membranes, phosphorylation modifications can regulate activity or protein interactions, and glycosylation and N-terminal formylation can stabilize proteins. Therefore, systematic experimental analysis of post-translational modifications of *Arabidopsis* chloroplast proteins is important but often challenging (Vener et al., 2001; Zabrouskov et al., 2003; Zybailov et al., 2008).

Both "top-down" and "bottom-up" Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) strategies have been used to analyze plastid protein post-translational modifications because many peptide and other bonds can be broken in a single MS analysis and FTICR-MS instruments have very high mass accuracies. Using a top-down method, Zabrouskov et al. (2003) analyzed different chloroplast protein fractions and showed that even in a complex protein mixture, intact proteins can be identified and modifications determined. In addition, Vener et al. (2001) studied phosphorylated thylakoid proteins using metal affinity chromatography and identified several new phosphorylated proteins, as well as phosphorylation sites on well-known photosynthetic proteins, while Zybailov et al. (2008) used a linear ion trap triple quadrupole (LTQ) system to evaluate modifications of the N-termini of 54 identified chloroplast proteins and associated features, revealing a novel N-terminal acetylation motif.

The proteomic technologies have been applied relatively successfully in large-scale protein separation during the past 20 years. However, limitations of current proteomic tools still exist. The isolation of low-abundance proteins, membrane and hydrophobic proteins, and basic proteins are still relatively difficult. The existing 2-DE technologies are still time-consuming, laborious and messy. Such problems may limit the wider use of the current proteomic technologies in chloroplasts. Therefore, further improvements to overcome these problems are required.

V. Other Approaches

A. Metabolomics

The ultimate goal of metabolomics is to analyze the complete set of small molecules (metabolites) present in cells at particular developmental stages, under particular environmental conditions, and/or in the context of various genetic modifications

(Oliver et al., 1998; Fiehn, 2002). Generally, in metabolomic studies, information on identified and unidentified metabolites is obtained and then used for metabolite profiling and metabolic network analysis. Ultimately, dynamic information regarding the fluxes through the different pathways needs to be obtained because steady-state levels of intermediates may not reflect the quantitative significance of a given metabolite to overall metabolism. Using metabolomics approaches, researchers can compare levels of metabolites of specific classes or pathways and/or perform global analyses of metabolites to detect signature patterns such as metabolite–metabolite correlations (Weckwerth et al., 2004; Dettmer et al., 2007).

Mass spectrometry–based metabolic profiling has been widely used to identify metabolites in plants, and combinations of transcriptomic and metabolomic approaches enable very large-scale analyses of metabolites and corresponding mRNA species, permitting holistic systems analysis of model plants such as *Arabidopsis* (Cook et al., 2004; Hirai et al., 2004; Kaplan et al., 2004, 2007; Yonekura-Sakakibara et al., 2007). To date, however, there have been few metabolomic investigations of photosynthetic mechanisms, although comprehensive analysis of *albino* mutant lines of *Arabidopsis* by a combination of metabolic fingerprinting using one-dimensional ^{1}H nuclear magnetic resonance (NMR) spectroscopy and multidimensional NMR spectroscopy with stable isotope labeling has been attempted (Tian et al., 2007). One-dimensional ^{1}H NMR metabolic fingerprinting revealed global metabolic changes in the *albino* mutants, notably reductions in levels of aromatic metabolites and changes in aliphatic metabolites. In addition, NMR measurements of plants fed $^{13}C_6$-glucose showed that the *albino* lines had dramatically different ^{13}C-labeling patterns and increased levels of several amino acids, especially Asn and Gln, compared to wild-type counterparts. This study illustrated the extent to which the metabolite network is affected by chloroplast function in plants and demonstrated the effectiveness of NMR-based metabolic analysis for metabolite profiling (Tian et al., 2007). As the plastid is a metabolic center where many essential compounds are synthesized, such as plant hormones, fatty acids and lipids, amino acids, vitamins, purine and pyrimidine nucleotides, tetrapyrroles, and isoprenoids, metabolomics approaches are likely to be widely used in studies of chloroplast functions in the near future.

The metabolomics approaches sound very promising, but it is still extremely challenging to use them successfully in plants. For a given organism, the chemical properties of metabolites are complex and diverse. They include ionic inorganic species, carbohydrates, hydrophobic lipids, and complex natural products and etc. Therefore, it is very difficult to profile all of the metabolome simultaneously with current analytical techniques. In the future, multiple and parallel approaches for comprehensive analyses should be incorporated to circumvent this obstacle.

B. TILLING

Although reverse genetics has become an important approach for many biologists, reverse genetic methods are not equally applicable to all organisms. For example, T-DNA insertional mutagenesis has turned the problem of obtaining a gene knockout into an *in silico* procedure for >70% of *Arabidopsis* genes (Alonso et al., 2003), but no comparable resources are available for rice (*Oryza sativa*) or maize (*Zea mays*). Similarly, RNAi-based silencing is an exciting strategy for reverse genetics, but its throughput is limited by the difficulty of delivering siRNAs to target loci (Waterhouse et al., 1998). However, in the last decade, a novel reverse genetic approach called TILLING (targeting-induced local lesions in genomes) has been developed (McCallum et al., 2000; Henikoff et al., 2004). High-throughput TILLING permits the rapid, low-cost discovery of induced point mutations in populations of chemically mutagenized individuals. Because chemical mutagenesis is widely applicable and mutation detection by TILLING is solely dependent on obtaining sufficient yields of PCR products, TILLING can be applied to most organisms, including some important crop plants, such as rice and maize, and model systems such as *Arabidopsis* (Till et al., 2003; Henikoff et al., 2004).

For TILLING in *Arabidopsis*, seeds are mutagenized by treatment with EMS. The resulting M1 plants are self-fertilized, and M2 individuals are used to prepare DNA samples for mutational screening. The DNA samples are pooled, and the pools are amplified using gene-specific primers.

Then the amplified products are incubated with the CEL I endonuclease (Oleykowski et al., 1998), which cleaves the 3′ side of mismatches and loop-outs in heteroduplexes between wild-type and mutant DNA, while leaving duplexes intact. Cleaved products are electrophoretically separated, and then a standard commercial image-processing program is used to examine the gel readout. Differential double end-labeling of amplification products permits rapid confirmation of mutations, which can be easily distinguished from amplification artifacts. Till et al. (2003) have developed TILLING as a service to the *Arabidopsis* community, known as the *Arabidopsis* TILLING project. Its goal is to rapidly deliver allelic series of ethylmethanesulfonate-induced mutations in target 1-kb loci, as requested by the international research community. Some chloroplast proteins are also involved in embryo development, so their null T-DNA mutants could lead to seed-lethal phenotypes. *Arabidopsis* TILLING may provide a good strategy for analyzing the functions of these genes.

VI. Conclusions and Perspectives

The ultimate goals of photosynthesis research are to identify all the genes that are involved in photosynthesis and to understand how the proteins they encode function and interact with one another. Combining spectroscopic, biochemical, crystallographic, and physiological approaches, functional genomics has proven to be a highly successful approach for elucidating the functions of chloroplast genes in photosynthetic processes and thus achieving some of the ultimate goals. However, although these studies have provided a wealth of information on the biogenesis of the photosynthetic apparatus and regulation of photosynthetic mechanisms, we have only touched the tip of the iceberg in understanding the complex photosynthetic processes. For instance, we are just beginning to understand how a functional complex is assembled in the thylakoid membrane together with its numerous pigment and redox cofactors. *Arabidopsis* appears to the best model for plant functional genomics analysis in many respects. However, *Arabidopsis* is not an ideal organism for all fields of photosynthetic research. For example, chloroplast transformation is not currently practical in *Arabidopsis*. In addition, because *Arabidopsis* is an annual C3 plant, it is not suitable for C4 photosynthetic research or for investigating photosynthesis-related aspects of dormancy in temperate perennials. Therefore, functional genomic data regarding photosynthesis in *Arabidopsis* and other model plants, such as tobacco, maize, *Chlamydomonas* and *Populus*, can be valuably compared and contrasted to gain a global understanding of the regulation of photosynthetic mechanisms.

For scientists studying photosynthesis, the new challenge is in integrating the plethora of data obtained from genetic, transcriptomic, proteomic, metabolomic, and other analyses. Because they are derived from different analytical platforms, it is important to normalize the data rigorously, and procedures to do so, against thoroughly validated standards, are being introduced. With further developments to tackle remaining problems, the emerging plant systems biology approach, combining transcriptomics, proteomics and metabolomics, is likely to be an ideal strategy for addressing many challenging issues in future studies. It has been suggested that systems biology has great potential to better understand signal transduction between plastids and the nucleus (Jung and Chory, 2010).

Acknowledgments

This work of Dr. Lixin Zhang's group was supported in part by National Science Foundation of China (30725003, 30970217) and the Major State Basic Research Development Program (2009CB118500).

References

Abdallah F, Salamini F and Leister D (2000) A prediction of the size and evolutionary origin of the proteome of chloroplasts of *Arabidopsis*. Trends Plant Sci 5: 141–142

Adachi H, Umena Y, Enami I, Henmi T, Kamiya N and Shen JR (2009) Towards structural elucidation of eukaryotic photosystem II: Purification, crystallization and preliminary X-ray diffraction analysis of photosystem II from a red alga. Biochim Biophys Acta 1787: 121–128

Allen JF (1992) Protein phosphorylation in regulation of photosynthesis. Biochim Biophys Acta 1098: 275–335

Allison LA, Simon LD and Maliga P (1996) Deletion of *rpoB* reveals a second distinct transcription system in plastids of higher plants. EMBO J 15: 2802–2809

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, et al (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science 301: 653–657

Aro EM, Suorsa M, Rokka A, Allahverdiyeva Y, Paakkarinen V, Saleem A, Battchikova N and Rintamaki E (2005) Dynamics of photosystem II: a proteomic approach to thylakoid protein complexes. J Exp Bot 56: 347–335

Asakura Y and Barkan A (2006) *Arabidopsis* orthologs of maize chloroplast splicing factors promote splicing of orthologous and species-specific group II introns. Plant Physiol 142: 1656–1663

Asakura Y and Barkan A (2007) A CRM domain protein functions dually in group I and group II intron splicing in land plant chloroplasts. Plant Cell 19: 3864–3875

Asakura Y, Bayraktar O and Barkan A (2008) Two CRM protein subfamilies cooperate in the splicing of group IIB introns in chloroplasts. RNA 14: 2319–2332

Baba K, Schmidt J, Espinosa-Ruiz A, Villarejo A, Shiina T, Gardestrom P, Sane AP and Bhalerao RP (2004) Organellar gene transcription and early seedling development are affected in the rpoT; 2 mutant of *Arabidopsis*. Plant J 38: 38–48

Baginsky S and Gruissem W (2006) *Arabidopsis thaliana* proteomics: from proteome to genome. J Exp Bot 57: 1485–1491

Balmer Y, Koller A, Del Val G, Manieri W, Schurmann P and Buchanan BB (2003) Proteomics gives insight into the regulatory function of chloroplast thioredoxins. Proc Natl Acad Sci USA 100: 370–375

Barkan A (1998) Approaches to investigating nuclear genes that function in chloroplast biogenesis in land plants. Methods Enzymol 297: 38–57

Barkan A and Goldschmidt-Clermont M (2000) Participation of nuclear genes in chloroplast gene expression. Biochimie 82: 559–572

Bauer J, Chen K, Hiltbunner A, Wehrli E, Eugster M, Schnell D and Kessler F (2000) The major protein import receptor of plastids is essential for chloroplast biogenesis. Nature 403: 203–207

Biehl A, Richly E, Noutsos C, Salamini F and Leister D (2005) Analysis of 101 nuclear transcriptomes reveals 23 distinct regulons and their relationship to metabolism, chromosomal gene distribution and co-ordination of nuclear and plastid gene expression. Gene 344: 33–41

Boardman NK (1977) Comparative photosynthesis of sun and shade plants. Annu Rev Plant Physiol 28: 355–377

Boynton JE, Gillham NW, Harris EH, Hosler JP, Johnson AM, Jones AR, Randolph-Anderson BL, Robertson D, Klein TM, Shark KB et al (1988) Chloroplast transformation in *Chlamydomonas* with high velocity microprojectiles. Science 240: 1534–1538

Brown PO and Botstein D (1999) Exploring the new world of the genome with DNA microarrays. Nat Genet 21: 33–37

Cai WH, Ji DL, Peng LW, Guo JK, Ma JF, Zou MJ, Lu CM and Zhang LX (2009) LPA66 is required for editing *psbF* chloroplast transcripts in *Arabidopsis*. Plant Physiol 150: 1260–1271

Cai WH, Ma JF, Chi W, Zou M, Guo JK, Lu CM and Zhang LX (2010) Cooperation of LPA3 and LPA2 is essential for photosystem II assembly in *Arabidopsis*. Plant Physiol 154: 109–120

Chateigner-Boutin AL and Small I (2007) A rapid high-throughput method for the detection and quantification of RNA editing based on high-resolution melting of amplicons. Nucleic Acids Res 35: e114

Chateigner-Boutin AL, Ramos-Vega M, Guevara-Garcia A, Andres C, Gutierrez-Nava MD, Cantero A, Delannoy E, Jimenez LF, Lurin C, Small I and Leon P (2008) CLB19, a pentatricopeptide repeat protein required for editing of *rpoA* and *clpP* chloroplast transcripts. Plant J 56: 590–602

Chen KM, Holmström M, Raksajit W, Suorsa M, Piippo, M and Aro EM (2010) Small chloroplast-targeted DnaJ proteins are involved in optimization of photosynthetic reactions in *Arabidopsis thaliana*. BMC Plant Biol 10: 43

Chen M, Choi Y, Voytas DF and Rodermel S (2000) Mutations in the *Arabidopsis* VAR2 locus cause leaf variegation due to the loss of a chloroplast FtsH protease. Plant J 22: 303–313

Chi W, Ma J, Zhang D, Guo J, Chen F, Lu C and Zhang L (2008) The pentatricopeptide repeat protein DELAYED GREENING1 is involved in the regulation of early chloroplast development and chloroplast gene expression in *Arabidopsis*. Plant Physiol 147: 573–584

Chi W, Mao J, Li QN, Ji DL, Zou MJ, Lu CM and Zhang LX (2010) Interaction of the pentatricopeptide-repeat protein DELAYED GREENING1 with sigma factor SIG6 in the regulation of chloroplast gene expression in *Arabidopsi*s cotyledons. Plant J 64: 14–25

Collin V, Issakidis-Bourguet E, Marchand C, Hirasawa M, Lancelin J-M, Knaff DB and Miginiac-Maslow M (2003) The *Arabidopsis* plastidial thioredoxins: new functions and new insights into specificity. J Biol Chem 278: 23747–23752

Cook D, Fowler S, Fiehn O and Thomashow MF (2004) A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of *Arabidopsis*. Proc Natl Acad Sci USA 101: 15243–15248

Courtois F, Merendino L, Demarsy E, Mache R and Lerbs-Mache S (2007) Phage-type RNA polymerase RPOTmp transcribes the *rrn* operon from the PC promoter at early developmental stages in *Arabidopsis*. Plant Physiol 145: 712–721

Danon A (1997) Translational regulation in the chloroplast. Plant Physiol 115: 1293–1298

Davletova S, Rizhsky L, Liang H, Zhong S, Oliver DJ, Coutu J, Shulaev V, Schlauch K and Mittler R (2005) Cytosolic ascorbate peroxidase 1 is a central component of the reactive oxygen gene network of *Arabidopsis*. Plant Cell 17: 268–281

de Longevialle AF, Hendrickson L, Taylor NL, Delannoy E, Lurin C, Badger M, Millar AH and Small I (2008) The pentatricopeptide repeat gene OTP51 with two LAGLIDADG motifs is required for the cis-splicing of plastid ycf3 intron 2 in *Arabidopsis thaliana*. Plant J 56: 157–168

Dettmer K, Aronov PA and Hammock BD (2007) Mass spectrometry-based metabolomics. Mass Spectrom Rev 26: 51–78

Duggan DJ, Bittner M, Chen Y, Meltzer P and Trent JM (1999) Expression profiling using cDNA microarrays. Nat Genet 21: 10–14

Favory JJ, Kobaysh M, Tanaka K, Peltier G, Kreis M, Valay JG and Lerbs-Mache S (2005) Specific function of a plastid sigma factor for *ndhF* gene transcription. Nucleic Acids Res 33: 5991–5999

Felder S, Meierhoff K, Sane AP, Meurer J, Driemel C, Plucken H, Klaff P, Stein B, Bechtold N and Westhoff P (2001) The nucleus-encoded HCF107 gene of *Arabidopsis* provides a link between intercistronic RNA processing and the accumulation of translation-competent *psbH* transcripts in chloroplasts. Plant Cell 13: 2127–2141

Fey V, Wagner R, Brautigam K, Wirtz M, Hell R, Dietzmann A, Leister D, Oelmuller R and Pfannschmidt T (2005) Retrograde plastid redox signals in the expression of nuclear genes for chloroplast proteins of *Arabidopsis thaliana*. J Biol Chem 280: 5318–5328

Ferreira KN, Iverson TM, Maghlaoui K, Barber J and Iwata S (2004) Architecture of the photosynthetic oxygen-evolving center. Science 303: 1831–1838

Ferro M, Salvi D, Rivière-Rolland H, Vermat T, Seigneurin-Berny D, Grunwald D, Garin J, Joyard J and Rolland N (2002) Integral membrane proteins of the chloroplast envelope: Identification and subcellular localization of new transporters. Proc Natl Acad Sci USA 99: 11487–11492

Ferro M, Salvi D, Brugiere S, Miras S, Kowalski S, Louwagie M, Garin J, Joyard J and Rolland N (2003) Proteomics of the chloroplast envelope membranes from *Arabidopsis thaliana*. Mol Cell Proteomics 2: 325–345

Fiehn O (2002) Metabolomics - the link between genotypes and phenotypes. Plant Mol Biol 48: 155–171

Foyer CH and Allen JF (2003) Lessons from redox signaling in plants. Antioxid Redox Signal 5: 3–5

Friso G, Ytterberg AJ, Giacomelli L, Peltier JB, Rudella A, Sun Q and van Wijk KJ (2004) In-depth analysis of the thylakoid membrane proteome of *Arabidopsis thaliana* chloroplasts: new proteins, new functions, and a plastid proteome database. Plant Cell 16: 478–499

Froehlich JE, Wilkerson CG, Ray WK, McAndrew RS, Osteryoung KW and Gage DA et al (2003) Proteomic study of the *Arabidopsis thaliana* chloroplastic envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis. J Proteome Res 2: 413–425

Fu A, He Z, Cho HS, Lima A, Buchanan BB and Luan S (2007) A chloroplast cyclophilin functions in the assembly and maintenance of photosystem II in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 104: 15947–15952

Gadjev I, Vanderauwera S, Gechev TS, Laloi C, Minkov IN, Shulaev V, Apel K, Inzé D, Mittler R and van Breusegem F (2006) Transcriptomic footprints disclose specificity of reactive oxygen species signaling in *Arabidopsis*. Plant Physiol 141: 436–445

Gechev TS, Gadjev IZ and Hille J (2004) An extensive microarray analysis of AAL-toxin-induced cell death in *Arabidopsis thaliana* brings new insights into the complexity of programmed cell death in plants. Cell Mol Life Sci 61: 1185–1197

Gechev TS and Hille J (2005) Hydrogen peroxide as a signal controlling plant programmed cell death. J Cell Biol 168: 17–20

Giacomelli L, Rudella A and van Wijk KJ (2006) High light response of the thylakoid proteome in *Arabidopsis* wild type and the ascorbate-deficient mutant *vtc2-2*: a comparative proteomics study. Plant Physiol 141: 685–701

Hager A and Holocher K (1994) Localization of the xanthophyll-cycle enzyme violaxanthin de-epoxidase within the thylakoid lumen and abolition of its mobility by a (light-dependent) pH decrease. Planta 192: 581–589

Hajdukiewicz PTJ, Allison LA and Maliga P (1997) The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. EMBO J 16: 4041–4048

Hall M, Mata-Cabana A, Akerlund HE, Florencio FJ, Schröder WP, Lindahl M and Kieselbach T (2010) Thioredoxin targets of the plant chloroplast lumen and their implications for plastid function. Proteomics 10: 987–1001

Hammani K, Okuda K, Tanz SK, Chateigner-Boutin AL, Shikanai T and Small I (2009) A study of new *Arabidopsis* chloroplast RNA editing mutants reveals general features of editing factors and their target sites. Plant Cell 21: 3686–3699

Hanaoka M, Kanamaru K, Fujiwara M, Takahashi H and Tanaka K (2005) Glutamyl-tRNA mediates a switch in RNA polymerase use during chloroplast biogenesis. EMBO Rep 6: 545–550

Harris EH, Boynton JE and Gillham NW (1994) Chloroplast ribosomes and protein synthesis. Microbiol Rev 58: 700–754

He WZ and Malkin R (1992) Specific release of a 9-kDa extrinsic polypeptide of photosystem I from spinach chloroplasts by salt washing. FEBS Lett 308: 298–300

Hedtke B, Börner T and Weihe A (1997) Mitochondrial and chloroplast phage-type RNA polymerases in *Arabidopsis*. Science 277: 809–811

Henikoff S, Till BJ and Comai L (2004) TILLING. Traditional mutagenesis meets functional genomics. Plant Physiol 135: 630–636

Hihara Y, Kamei A, Kanehisa M, Kaplan A and Ikeuchi M (2001) DNA microarray analysis of cyanobacterial gene expression during acclimation to high light. Plant Cell 13: 793–806

Hirai MY, Yano M, Goodenowe DB, Kanaya S, Kimura T, Awazuhara M, Arita M, Fujiwara T and Saito K (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 101: 10205–10210

Homann A and Link G (2003) DNA-binding and transcription characterization of three cloned sigma factors from mustard (*Sinapis alba* L.) suggest overlapping and distinct roles in plastid gene expression. Eur J Biochem 270: 1288–1300

Hricová A, Quesada V and Micol JL (2006) The SCABRA3 nuclear gene encodes the plastid RpoTp RNA polymerase, which is required for chloroplast biogenesis and mesophyll cell proliferation in *Arabidopsis*. Plant Physiol 141: 942–958

Huner NPA, Öquist G and Sarhan F (1998) Energy balance and acclimation to light and cold. Trends Plant Sci 3: 224–230

Ishizaki Y, Tsunoyama Y, Hatano K, Ando K, Kato K, Shinmyo A, Kobori M, Takeba G, Nakahira Y and Shiina T (2005) A nuclear-encoded sigma factor, *Arabidopsis* SIG6, recognizes sigma-70 type chloroplast promoters and regulates early chloroplast development in cotyledons. Plant J 42: 133–144

Jarvis P, Chen LJ, Li HM, Pete CA, Fankhauser C and Chory J (1998) An *Arabidopsis* mutant defective in the plastid general protein import apparatus. Science 282: 100–103

Johnson X, Wostrikoff K, Finazzi G, Kuras R, Schwarz C, Bujaldon S, Nickelsen J, Stern DB, Wollman FA and Vallon O (2010) MRL1, a conserved pentatricopeptide repeat protein, is required for stabilization of *rbcL* mRNA in *Chlamydomonas* and *Arabidopsis*. Plant Cell 22: 234–248

Jung HS and Chory J (2010) Signaling between chloroplasts and the nucleus: can a systems biology approach bring clarity to a complex and highly regulated pathway? Plant Physiol 152:453–459

Kanamaru K, Nagashima A, Fujiwara M, Shimada H, Shirano Y, Nakabayashi K, Shibata D, Tanaka K and Takahashi H (2001) An *Arabidopsis* sigma factor (SIG2)-dependent expression of plastid-encoded tRNAs in chloroplasts. Plant Cell Physiol 42: 1034–1043

Kaplan F, Kopka J, Haskell DW, Zhao W, Schiller KC, Gatzke N, Sung DY and Guy CL (2004) Exploring the temperature-stress metabolome of *Arabidopsis*. Plant Physiol 136: 4159–4168

Kaplan F, Kopka J, Sung DY, Zhao W, Popp M, Porat R and Guy CL (2007) Transcript and metabolite profiling during cold acclimation of *Arabidopsis* reveals an intricate relationship of cold-regulated gene expression with modifications in metabolite content. Plant J 50: 967–981

Karpinski S, Escobar C, Karpinska B, Creissen G and Mullineaux PM (1997) Photosynthetic electron transport regulates the expression of cytosolic ascorbate peroxidase genes in *Arabidopsis* during excess light stress. Plant Cell 9: 627–640

Kavakli IH, Kato C, Choi SB, Kim KH, Salamone PR, Ito H and Okita TW (2002) Generation, characterization, and heterologous expression of wild-type and up-regulated forms of *Arabidopsis thaliana* leaf ADP-glucose pyrophosphorylase. Planta 215: 430–439

Kieselbach T, Hagman A, Andersson B and Schroder WP (1998) The thylakoid lumen of chloroplasts. Isolation and characterization. J Biol Chem 273: 6710–6716

Kimura M, Yamamoto YY, Seki M, Sakurai T, Sato M, Abe T, Yoshida S, Manabe K, Shinozaki K and Matsui M (2003) Identification of *Arabidopsis* genes regulated by high light-stress using cDNA microarray. Photochem Photobiol 77: 226–233

Kleffmann T, Hirsch-Hoffmann M, Gruissem W and Baginsky S (2006) plprot: a comprehensive proteome database for different plastid types. Plant Cell Physiol 47: 432–436

Kleine T, Kindgren P, Benedict C, Hendrickson L and Strand A (2007) Genome-wide gene expression analysis reveals a critical role for CRYPTOCHROME1 in the response of *Arabidopsis* to high irradiance. Plant Physiol 144: 1391–1406

Klimyuk VI, Persello-Cartieaux F, Havaux M, Contard-David P, Schuenemann D, Meiherhoff K, Gouet P, Jones JD, Hoffman NE and Nussaume L (1999) A chromodomain protein encoded by the *Arabidopsis* CAO gene is a plant-specific component of the chloroplast signal recognition particle pathway that is involved in LHCP targeting. Plant Cell 11: 87–100

Kotera E, Tasaka M and Shikanai T (2005) A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. Nature 433: 326–330

Kroll D, Meierhoff K, Bechtold N, Kinoshita M, Westphal S, Vothknecht UC, Soll J and Westhoff P (2001) VIPP1, a nuclear gene of *Arabidopsis thaliana* essential for thylakoid membrane formation. Proc Natl Acad Sci USA 98: 4238–4242

Kubis S, Baldwin A, Patel A, Razzaq A, Dupree P, Lilley K, Kurth J, Leister D and Jarvi P (2003) The *Arabidopsis ppi1* mutant is specifically defective in the expression, chloroplast import, and accumulation of photosynthetic proteins. Plant Cell 15: 1859–1871

Kurth J, Varotto C, Pesaresi P, Biehl A, Richly E, Salamini F and Leister D (2002) Gene-sequence-tag expression analyses of 1800 genes related to chloroplast functions. Planta 215: 101–109

Liere K and Maliga P (1999) *In vitro* characterization of the tobacco *rpoB* promoter reveals a core sequence motif conserved between phage-type plastid and plant mitochondrial promoters. EMBO J 18: 249–257

Lima A, Lima S, Wong JH, Philips RS, Buchanan BB and Luan S (2006) A redox-active FKBP-type immunophilin

functions in accumulation of the photosystem II supercomplex in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 103: 12631–12636

Lurin C, Andres C, Aubourg S et al (2004) Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell, 16: 2089–2103

Ma JF, Peng LW, Guo JK, Lu QT, Lu CM and Zhang LX (2007) LPA2 is required for efficient assembly of photosystem II in *Arabidopsis thaliana*. Plant Cell 19: 1980–1993

Maliga P (1988) Two plastid RNA polymerases of higher plants: an evolving story. Trends Plant Sci 3: 4–6

McCallum CM, Comai L, Greene EA and Henikoff S (2000) Targeted screening for induced mutations. Nat Biotechnol 18: 455–457

Meierhoff K, Felder S, Nakamura T, Bechtold N and Schuster G (2003) HCF152, an *Arabidopsis* RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast *psbB-psbT-psbH-petB-petD* RNAs. Plant Cell 15: 1480–1495

Meinke DW, Cherry JM, Dean CD, Rounsley S and Koornneef M (1998) *Arabidopsis thaliana*: a model plant for genome analysis. Science 282: 662–682

Melkozernov AN, Barber J and Blankenship RE (2006) Light harvesting in photosystem I supercomplexes. Biochemistry 45: 331–345

Meurer J, Meierhoff K, and Westhoff P (1996) Isolation of high-chlorophyll-fluorescence mutants of *Arabidopsis thaliana* and their characterization by spectroscopy, immunoblotting and northern hybridization. Planta 198: 385–396

Meurer J, Plücken H, Kowallik KV and Westhoff P (1998) A nuclear-encoded protein of prokaryotic origin is essential for the stability of photosystem II in *Arabidopsis thaliana*. EMBO J 17: 5286–5297

Michael JH (2002) DNA microarray technology: Devices, systems, and applications. Annu Rev Biomed Eng 4: 129–153

Motohashi R, Nagata N, Ito T, Takahashi S, Hobo T, Yoshida S and Shinozaki K (2001) An essential role of a TatC homologue of a Delta pH-dependent protein transporter in thylakoid membrane formation during chloroplast development in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 98: 10499–10504

Nagashima A, Hanaoka M, Shikanai T, Fujiwara, M, Kanamaru K, Takahashi H and Tanaka K (2004) The multiple-stress responsive plastid sigma factor, SIG5, directs activation of the *psbD* blue light-responsive promoter (BLRP) in *Arabidopsis thaliana*. Plant Cell Physiol 45: 357–368

Nelson N and Yocum CF (2006) Structure and function of photosystem I and II. Annu Rev Plant Biol 57: 521–565

Neubauer G, King A, Rappsilber J, Calvio C, Watson M, Ajuh P, Sleeman J, Lamond A and Mann M (1998) Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. Nat Genet 20: 46–50

Neuhaus HE and Wagner R (2000) Solute pores, ion channels, and metabolite transporters in the outer and inner envelope membranes of higher plant plastids. Biochim Biophys Acta 1465: 307–323

Niyogi KK, Grossman AR and Bjorkman O (1998) *Arabidopsis* mutants define a central role for the xanthophyll cycle in the regulation of photosynthetic energy conversion. Plant Cell 10: 1121–1134

Oelmüller R, Herrmann RG and Pakrasi HB (1996) Molecular studies of CtpA, the carboxyl-terminal processing protease for the D1 protein of the photosystem II reaction center in higher plants. J Biol Chem 271: 21848–21852

Okuda K, Chateigner-Boutin AL, Nakamura T, Delannoy E, Sugita M, Myouga F, Motohashi R, Shinozaki K, Small I and Shikanai T (2009) Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in *Arabidopsis* chloroplasts. Plant Cell 21: 146–156

Okuda K, Myouga F, Motohashi R, Shinozaki K and Shikanai T (2007) Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing. Proc Natl Acad Sci USA 104: 8178–8183

Okuda K, Hammani K, Tanz SK, Peng L, Fukao Y, Myouga F, Motohashi R, Shinozaki K, Small I and Shikanai T (2010) The pentatricopeptide repeat protein OTP82 is required for RNA editing of plastid *ndhB* and *ndhG* transcripts. Plant J 61: 339–349

Oleykowski CA, Bronson Mullins CR, Godwin AK and Yeung AT (1998) Mutation detection using a novel plant endonuclease. Nucleic Acids Res 26: 4597–4602

Oliver S, Winson MK, Kell DB and Baganz F (1998) Systematic functional analysis of the yeast genome. Trends Biotechnol 16: 373–378

op den Camp RGL, Przybyla D, Ochsenbein C, Laloi C, Kim C, Danon A, Wagner D, Hideg E, Göbel C, Feussner I, et al (2003) Rapid induction of distinct stress responses after the release of singlet oxygen in *Arabidopsis*. Plant Cell 15: 2320–2332

Ostheimer G, Williams-Carrier R, Belcher S, Osborne E, Gierke J and Barkan A (2003) Group II intron splicing factors derived by diversification of an ancient RNA binding module. EMBO J 22: 3919–3929

Parinov S, Sevugan M, Ye D, Yang W, Kumaran M and Sundaresan V (1999) Analysis of flanking sequences from dissociation insertion lines: A database for reverse genetics in *Arabidopsis*. Plant Cell 11: 2263–2270

Peltier JB, Cai Y, Sun Q, Zabrouskov V, Giacomelli L, Rudella A, Ytterberg AJ, Rutschow H and van Wijk KJ (2006) The oligomeric stromal proteome of *Arabidopsis thaliana* chloroplasts. Mol Cell Proteomics 5: 114–133

Peltier JB, Emanuelsson O, Kalume DE, Ytterberg J, Friso G, Rudella A, Liberles DA, Soderberg L, Roepstorff P, von Heijne G and van Wijk KJ (2002) Central functions of the

lumenal and peripheral thylakoid proteome of *Arabidopsis* determined by experimentation and genome-wide prediction. Plant Cell 14: 211–236

Peltier JB, Friso G, Kalume DE, Roepstorff P, Nilsson F, Adamska I and van Wijk KJ (2000) Proteomics of the chloroplast: systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins. Plant Cell 12: 319–341

Peltier JB, Ripoll DR, Friso G, Rudella A, Cai Y, Ytterberg J, Giacomelli L, Pillardy J and van Wijk KJ (2004) Clp protease complexes from photosynthetic and non-photosynthetic plastids and mitochondria of plants, their predicted three-dimensional structures, and functional implications. J Biol Chem 279: 4768–4781

Peng LW, Ma JF, Chi W, Guo JK, Zhu SY, Lu QT, Lu CM and Zhang LX (2006) Low PSII accumulation1 is involved in the efficient assembly of photosystem II in *Arabidopsis thaliana*. Plant Cell 18: 955–969

Pfalz J, Liere K, Kandlbinder A, Dietz KJ and Oelmüller R (2006) pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. Plant Cell 18: 176–197

Phee BK, Cho JH, Park S, Jung JH, Lee YH, Jeon JS, Bhoo SH and Hahn TR (2004) Proteomic analysis of the response of *Arabidopsis* chloroplast proteins to high light stress. Proteomics 4: 3560–3568

Privat I, Hakimi MA, Buhot L, Favory JJ and Lerbs-Mache S (2003) Characterization of *Arabidopsis* plastid sigma-like transcription factors SIG1, SIG2 and SIG3. Plant Mol Biol 55: 385–399

Radwanski ER, Barczak AJ and Last RL (1996) Characterization of tryptophan synthase alpha subunit mutants of *Arabidopsis thaliana*. Mol Gen Genet 253: 353–361

Richly E, Dietzmann A, Biehl A, Kurth J, Laloi C, Apel K, Salamini F and Leister D (2003) Covariations in the nuclear chloroplast transcriptome reveal a regulatory master-switch. EMBO Rep 4: 491–498

Richmond T and Somerville S (2000) Chasing the dream: plant EST microarrays. Curr Opin Plant Biol 3: 108–116

Rizhsky L, Liang H and Mittler R (2003) The water-water cycle is essential for chloroplast protection in the absence of stress. J Biol Chem 278: 38921–38925

Robinson C, Thompson SJ and Woolhead C (2001) Multiple pathways used for the targeting of thylakoid proteins in chloroplasts. Traffic 2: 245–251

Rochaix JD (2004) Genetics of the biogenesis and dynamics of the photosynthetic machinery in eukaryotes. Plant Cell 16: 1650–1660

Rossel JB, Wilson IW and Pogson BJ (2002) Global changes in gene expression in response to high light in *Arabidopsis*. Plant Physiol 130: 1109–1120

Robbins JC, Heller WP and Hanson MR (2009) A comparative genomics approach identifies a PPR-DYW protein that is essential for C-to-U editing of the *Arabidopsis* chloroplast *accD* transcript. RNA 15: 1142–1153.

Rutschow H, Ytterberg AJ, Friso G, Nilsson R and van Wijk KJ (2008) Quantitative proteomics of a chloroplast SRP54 sorting mutant and its genetic interactions with CLPC1 in *Arabidopsis thaliana*. Plant Physiol 148: 156–175

Sakamoto W, Tamura T, Hanba-Tomita Y and Murata M (2002) The VAR1 locus of *Arabidopsis* encodes a chloroplastic FtsH and is responsible for leaf variegation in the mutant alleles. Genes Cells 7: 769–780

Schaffer R, Landgraf J, Perez-Amador M and Wisman E (2000) Monitoring genome-wide expression in plants. Curr Opin Biotechnol 11: 162–167

Schweer J, Türkeri H, Link B and Link G (2010) AtSIG6, a plastid sigma factor from *Arabidopsis*, reveals functional impact of cpCK2 phosphorylation. Plant J 62: 192–202

Shimizu M, Kato H, Ogawa T, Kurachi T, Nakagawa Y and Kobayashi H (2010) Sigma factor phosphorylation in the photosynthetic control of photosystem stoichiometry. Proc Natl Acad Sci USA 107:10760–10764

Schubert M, Petersson UA, Haas BJ, Funk C, Schroder WP and Kieselbach T (2002) Proteome map of the chloroplast lumen of *Arabidopsis thaliana*. J Biol Chem 277: 8354–8365

Schult K, Meierhoff K, Paradies S, Töller T, Wolff P and Westhoff P (2007) The nuclear-encoded factor HCF173 is involved in the initiation of translation of the *psbA* mRNA in *Arabidopsis thaliana*. Plant Cell 19: 1329–1346

Schunemann D (2004) Structure and function of the chloroplast signal recognition particle. Curr Genet 44: 295–304

Seigneurin-Berny D, Rolland N, Garin J and Joyard J (1999) Technical advance: differential extraction of hydrophobic proteins from chloroplast envelope membranes: a subcellular-specific proteomic approach to identify rare intrinsic membrane proteins. Plant J 19: 217–228

Shikanai T (2006) RNA editing in plant organelles: Machinery, physiological function and evolution. Cell Mol Life Sci 63: 698–708

Shikanai T, Munekage Y, Shimizu K, Endo T and Hashimoto T (1999) Identification and characterization of *Arabidopsis* mutants with reduced quenching of chlorophyll fluorescence. Plant Cell Physiol 40: 1134–1142

Sirpiö S, Khrouchtchova A, Allahverdiyeva Y, Hansson M, Fristedt R, Vener AV and Aro EM (2008) AtCYP38 ensures early biogenesis, correct assembly and sustenance of photosystem II. Plant J 55: 639–651

Somerville C and Somerville S (1999) Plant functional genomics. Science 285: 380–383

Sommer A, Ne'eman E, Steffens JC, Mayer AM and Harel E (1994) Import, targeting, and processing of a plant polyphenol oxidase. Plant Physiol 105: 1301–1311

Stöckel J and Oelmüller R (2004) A novel protein for photosystem I biogenesis. J Biol Chem 279: 10243–10251

Sugiura M, Hirose T and Sugita M (1998) Evolution and mechanisms of translation in chloroplasts. Annu Rev Genet 32: 437–459

Sun Q, Zybailov B, Majeran W, Friso G, Olinares PD and van Wijk KJ (2009) PPDB, the plant proteomics database at Cornell. Nucleic Acids Res 37: D969–974

Sundberg E, Slagter JG, Fridborg I, Cleary SP, Robinson C and Coupland G (1997) ALBINO3, an *Arabidopsis* nuclear gene essential for chloroplast differentiation, encodes a chloroplast protein that shows homology to proteins present in bacterial membranes and yeast mitochondria. Plant Cell 9: 717–730

Sussman MR, Amasino RM, Young JC, Krysan PJ and Austin-Phillips S (2000) The *Arabidopsis* knockout facility at the University of Wisconsin-Madison. Plant Physiol 124: 1465–1467

Svab Z, Hajdukiewicz P and Maliga P (1990) Stable transformation of plastids in higher plants. Proc Natl Acad Sci USA 87: 8526–8530

Swiatecka-Hagenbruch M, Emanuel C, Hedtke B, Liere K and Börner T (2008) Impaired function of the phage-type RNA polymerase RpoTp in transcription of chloroplast genes is compensated by a second phage-type RNA polymerase. Nucleic Acids Res 36: 785–792

The Arabidopsis Genome Initiative (AGI) (2000) Analysis of the genome of the flowering plant *Arabidopsis thaliana*. Nature 408: 796–815

Tian C, Chikayama E, Tsuboi Y, Kuromori T, Shinozak K, Kikuchi J and Hirayama T (2007) Top-down phenomics of *Arabidopsis thaliana*. J Biol Chem 282: 18532–18541

Till B, Schmitz-Linneweber C, Williams-Carrier R and Barkan A (2001) CRS1 is a novel group II intron splicing factor that was derived from a domain of ancient origin. RNA 7: 1227–1238

Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR, Young K, Taylor NE, et al (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. Genome Res 13: 524–530

Tissier AF, Marillonnet S, Klimyuk V, Patel K, Torres MA, Murphy G and Jones JD (1999) Multiple independent defective suppressor-mutator transposon insertions in *Arabidopsis*: A tool for functional genomics. Plant Cell 11: 1841–1852

Tseng CC, Sung TY, Li YC, Hsu SJ, Lin CL and Hsieh MH (2010) Editing of *accD* and *ndhF* chloroplast transcripts is partially affected in the *Arabidopsis vanilla cream1* mutant. Plant Mol Biol 73: 309–323

Tsunoyama Y, Morikawa K, Shiina T and Toyashima Y (2002) Blue light specific and differential expression of a plastid σ factor Sig5 in *Arabidopsis thaliana*. FEBS Lett 516: 225–228

Tsunoyama Y, Ishizak Y, Morikawa K, Kobori M, Nakahira Y, Takeba G, Toyoshima Y and Shiina T (2004) Blue light-induced transcription of plastid-encoded *psbD* gene is mediated by a nuclear-encoded transcription initiation factor, AtSig5. Proc Natl Acad Sci USA 101: 3304–3309

Tyers M and Mann M (2003) From genomics to proteomics. Nat Rev Genet 422: 193–197

Umbach AL, Fiorani F and Siedow JN (2005) Characterization of transformed *Arabidopsis* with altered alternative oxidase levels and analysis of effects on reactive oxygen species in tissue. Plant Physiol 139: 1806–1820

Vanderauwera S, Zimmermann P, Rombauts S, Vandenabeele S, Langebartels C, Gruissem W, Inze D and Van Breusegem F (2005) Genome-wide analysis of hydrogen peroxide-regulated gene expression in *Arabidopsis* reveals a high light-induced transcriptional cluster involved in anthocyanin biosynthesis. Plant Physiol 139: 806–821

van Wijk KJ (2004) Plastid proteomics. Plant Physiol Biochem 42: 976--977

Varotto C, Pesaresi P, Maiwald D, Kurth J, Salamini F and Leister D (2000) Identification of photosynthetic mutants of *Arabidopsis* by automatic screening for altered effective quantum yield of photosystem 2. Photosynthetica 38: 497–504

Vener AV, Harms A, Sussman MR and Vierstra RD (2001) Mass spectrometric resolution of reversible protein phosphorylation in photosynthetic membranes of *Arabidopsis thaliana*. J Biol Chem 276: 6959–6966

Vert G, Grotz N, Dedaldechamp F, Gaymard F, Guerinot ML, Briat JF and Curie C (2002) IRT1, an *Arabidopsis* transporter essential for iron uptake from the soil and for plant growth. Plant Cell 14: 1223–1233

Vothknecht UC, Westhoff P, Kroll D, Meierhoff K, Bechtold N, Kinoshita M, Westphal S and Soll J (2001) Biogenesis and origin of thylakoid membranes. Biochim Biophys Acta 1541: 91–101

Waterhouse PM, Graham MW and Wang MB (1998) Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. Proc Natl Acad Sci USA 95: 13959–13964

Weber A, Servaites JC, Geiger DR, Kofler H, Hille D, Gröner F, Hebbeker U and Flügge UI (2000) Identification, purification, and molecular cloning of a putative plastidic glucose translocator. Plant Cell 12: 787–801

Weckwerth W, Loureiro ME, Wenzel K and Fiehn O (2004) Differential metabolic networks unravel the effects of silent plant phenotypes. Proc Natl Acad Sci USA 101: 7809–7814

Westphal S, Heins L, Soll J and Vothknecht UC (2001) Vipp1 deletion mutant of *Synechocystis*: a connection between bacterial phage shock and thylakoid biogenesis? Proc Natl Acad Sci USA 98: 4243–4248

Winkler R, Frank M, Galbraith D, Feyereisen R and Feldmann K (1998) Systematic reverse genetics of transfer-DNA-tagged lines of *Arabidopsis*. Plant Physiol 118: 743–750

Wollman FA, Minai L and Nechushtai R (1999) The biogenesis and assembly of photosynthetic proteins in thylakoid membranes. Biochim Biophys Acta 1141: 21–85

Wollman FA (2001) State transitions reveal the dynamics and flexibility of the photosynthetic apparatus. EMBO J 20: 3623–3630

Wu D, Wright DA, Wetzel C, Voytas DF and Rodermel S (1999) The IMMUTANS variegation locus of *Arabidopsis* decodes a mitochondrial alternative oxidase homolog that functions during early chloroplast biogenesis. Plant Cell 11: 43–55

Yao J, Roy-Chowdhury S and Allison LA (2003) AtSig5 is an essential nucleus-encoded Arabidopsis σ-like factor. Plant Physiol 132: 739–747

Yates JR III (2000) Mass spectrometry. From genomics to proteomics. Trends Genet 16: 5–8

Yonekura-Sakakibara K, Tohge T, Niida R and Saito K (2007) Identification of a flavonol 7-O-rhamnosyltransferase gene determining flavonoid pattern in *Arabidopsis* by transcriptome coexpression analysis and reverse genetics. J Biol Chem 282: 14932–1494

Yu QB, Jiang Y, Chong K and Yang ZN (2009) AtECB2, a pentatricopeptide repeat protein, is required for chloroplast transcript *accD* RNA editing and early chloroplast biogenesis in *Arabidopsis thaliana.* Plant J 59: 1011–1023

Zabrouskov V, Giacomelli L, Van Wijk KJ and McLafferty FW (2003) A new approach for plant proteomics: characterization of chloroplast proteins of *Arabidopsis thaliana* by top-down mass spectrometry. Mol Cell Proteomics 2: 1253–1260

Zerges W (2000) Translation in chloroplasts. Biochimie 82: 583–607

Zghidi W, Merendino L, Cottet A, Mache R and Lerbs-Mache S (2007) Nucleus-encoded plastid sigma factor SIG3 transcribes specifically the *psbN* gene in plastids. Nucleic Acids Res 35: 455–464

Zhang LX and Aro E-M (2002) Synthesis, membrane insertion and assembly of the chloroplast-encoded D1 protein into photosystem II. FEBS Lett 512: 13–18

Zhou W, Cheng Y, Yap A, Chateigner-Boutin AL, Delannoy E, Hammani K, Small I and Huang J (2008) The *Arabidopsis* gene YS1 encoding a DYW protein is required for editing of *rpoB* transcripts and the rapid development of chloroplasts during growth. Plant J 58: 82–96

Zybailov B, Rutschow H, Friso G, Rudella A, Emanuelsson O, Sun Q and van Wijk KJ (2008) Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. PLoS ONE 3: e1994

Zybailov B, Friso G, Kim J, Rudella A, Rodríguez VR, Asakura Y, Sun Q and van Wijk KJ (2009) Large scale comparative proteomics of a chloroplast Clp protease mutant reveals folding stress, altered protein homeostasis, and feedback regulation of metabolism. Mol Cell Proteomics 8: 1789–1810

Chapter 8

Evolutionary Integration of Chloroplast Metabolism with the Metabolic Networks of the Cells

Marc Linka and Andreas P.M. Weber*
Institut für Biochemie der Pflanzen, Heinrich-Heine-Universität, Universitätsstraße 1, D-40225 Düsseldorf, Germany

Summary

The merger of a prokaryotic cyanobacterium with a primitive eukaryotic mitochondriate cell was a crucial event in evolution, leading to the first photosynthetic eukaryote, the protoalga. A certainly essential step in this merger of two previously independent entities was the metabolic integration of cyanobacterial photosynthetic metabolism with that of the heterotrophic host. This required the insertion of specific metabolite transporters into the membrane system that separated the cyanobiont from the surrounding cytoplasm of the host cell. Based on phylogenetic and phylogenomic analyses of extant plastids, this metabolic connection between host and endosymbiont was predominantly established by routing of pre-existing host transporters to the cyanobiont plasma membrane. In this chapter, we will review the current knowledge on connecting host with cyanobiont metabolism via metabolite transporters. We further discuss possible syntrophic associations between cyanobacteria and primitive eukaryotic cells that might have paved the way into endosymbiosis.

*Author for correspondence, e-mail: andreas.weber@uni-duesseldorf.de

R.L. Burnap and W.F.J. Vermaas (eds.), Functional Genomics and Evolution of Photosynthetic Systems, Advances in Photosynthesis and Respiration 33, pp. 199–224, DOI 10.1007/978-94-007-1533-2_8,

I. Introduction

Approximately three billion years ago, an ancestor of extant cyanobacteria introduced oxygenic photosynthesis to a literally oxygen-free atmosphere (Blankenship, 1992; Falkowski and Godfrey, 2008). Descendants of these early oxygenic photosynthetic organisms are sustaining the oxygen concentration in the present Earth's atmosphere. Sunlight is absorbed and utilized by the photosynthetic apparatus and used for withdrawing electrons from water, thereby generating protons and oxygen (O_2) as a by-product. The electrons and protons are stored as usable chemical energy to assimilate inorganic nutrients (i.e., C, N, P, and S) into organic matter. Thus, autotrophic organisms are of central importance for the Earth, providing the essential O_2 for cellular respiration of aerobic organisms and are the primary producers at the base of almost every food chain.

More than one billion years ago a phagotrophic eukaryote gained the ability to conduct photosynthesis by engulfing a cyanobacterium-like prokaryote (Yoon et al., 2004), thus changing from a heterotrophic to a mixotrophic and subsequently autotrophic lifestyle. This permitted the invasion of and thriving in new habitats. Over millions of years the endosymbiotic cyanobacterium evolved into the organelle plastid. Today's photoautotrophic organisms exist ubiquitously on the Earth, with a remarkable diversity of species and appearance (Bhattacharya et al., 2004). The smallest and largest photosynthetic eukaryotes described to date are the green micro-alga *Ostreococcus tauri* (Lanier et al., 2008; Palenik et al., 2007) and the Giant Sequoia tree, respectively.

Plastids are the site of carbon dioxide, nitrite, and sulfur assimilation. They partially or fully conduct the biosynthesis of various metabolite classes, such as fatty acids, starch, amino acids, nucleic acids, isoprenoids, or phenylpropanoids (Weber and Flügge, 2002; Weber, 2006). To understand and fully utilize the potential of photosynthetic eukaryotes it is crucial reconstruct the events that led to plastid origin to understand the common features and specific adaptations in extant, distantly related photosynthetic organisms.

Plastid origin from a cyanobacterial prokaryote has been postulated already in the late nineteenth century by Andreas Schimper (Schimper, 1885) and led in 1971 to the "Theory of Endosymbiosis" by Lynn Margulis (1971a, b). Based on numerous single and multi-gene phylogenetic analyses from diverse plastid genomes and nuclear-encoded plastid-targeted genes it is now commonly accepted that the primary endosymbiosis had been a single event in eukaryotic evolution and all photosynthetic species containing primary plastids have been recently summarized in the supergroup Archaeplastida (Adl et al., 2005). Despite its monophyletic origin, the evolutionary biogenesis of plastids has remained particularly challenging. Within approx. 0.15 billion years after the unicellular primordial alga (i.e., the protoalga) had been established, it gave rise to the three major lineages of the Archaeplastida: Glaucophyta (glaucophyte algae), Rhodophyceae (red algae) and Chloroplastida (green algae and plants). Multicellularity evolved independently in the Rhodophyceae and in the Chloroplastida, whereas only the Chloroplastida mastered the colonization of the land. This was accompanied by an extraordinary radiation of species. Heterotrophic eukaryotes later captured green or red algae, respectively, transforming them to secondary plastids of green algal or red algal origin (Bhattacharya et al., 2004). The process of secondary endosymbiosis gave rise to a plethora of new species, mainly

Abbreviations: 2-OG – 2-oxoglutarate; 2-PG – 2-phosphoglyceric acid; 3-PGA – 3-phosphoglycerate; AAC – ATP ADP carrier; DHAP – dihydroxyacetone phosphate; DiT – dicarboxylate translocator; DMT – drug/metabolite transporter; DOXP – deoxyxylulose 5-phosphate; E4P – erythrose 4-phosphate; EGT – endosymbiotic gene transfer; ER – endoplasmatic reticulum; Fd – ferredoxin; GAP – glyceraldehyde 3-phosphate; GADPH – glyceraldehyde phosphate dehydrogenase; Glc6P – glucose-6-phosphate; GOGAT – glutamine:oxoglutarate aminotransferase; GPT – glucose 6-phosphate/phosphate translocator; GS – glutamine synthetase; GYA – giga years ago; IPP – isopentenyl pyrophosphate; MCF – mitochondrial carrier family; MEP – 2-methylerythritol-4-phosphate; MVA – mevalonic acid; NiR – nitrite reductase; NR – nitrate reductase; NST – nucleotide sugar transporter; NTT – nucleotide transporter; OAA – oxaloacetate; PEP – phosphoenolpyruvate; PEPC – phosphoenolpyruvate carboxylase; PPT – phosphoenolpyruvate/phosphate translocator; PT – phosphate translocator; pPT – plastidic phosphate translocator; Rubisco – ribulosebisphosphate carboxylase oxygenase; RubP – ribulose 1,5-*bis*phosphate; SAHC – S-adenosylhomocysteine; SAM – S-adenosylmethionine; THF – tetrahydrofolate; TIM – triose phosphate isomerase; TP – triose phosphate; TPT – triose-phosphate/phosphate translocator; XPT – xylulose 5-phosphate/phosphate translocator

micro- and macro-algae dominating aquatic environments as major primary producers. Serial secondary endosymbiosis, occurring by the replacement of a secondary plastid by another one, and even tertiary endosymbiosis via an engulfment of algae harboring a secondary plastid by another heterotrophic eukaryote, are evident among the diverse set of extant eukaryotic autotrophs. Even more obscure is the complete or partial loss of plastid functions in several species. Noteworthy is also continuing evolution of novel photosynthetic organisms of primary endosymbiotic origin, as has been recently reported for the thecamoeba *Paulinella chromatophora* (Marin et al., 2005; Nowack et al., 2008).

The antiquity of the last common ancestor, the recurrent acquisition of the plastid, and species radiation has scrambled the evolutionary signals of plastid origin. In the last decade, plastid-, EST- and complete genome sequence projects from a diverse group of photosynthetic organisms were initiated by the research community to clarify the complex host-endosymbiont partnership. Multiple cyanobacterial and plastid genomes are available. The green lineage has been followed from the unicellular and multi-cellular green algae to mosses, gymnosperms, and several mono- and dicotyledonous plants. A catalog of sequences from diverse micro- and macro-algae with primary, secondary, and even tertiary plastids was initiated to compile red algae evolution. The genome of the glaucophyte alga *Cyanophora paradoxa* is currently being sequenced to complement the available sequence resources with the third major lineage of the Archaeplastida. Besides all the diversity, common themes and specialization in plastid evolution can be now validated and addressed.

Endosymbiontic gene transfer (EGT) to the host nucleus (Weeden, 1981) massively reduced the plastid genome size. The model-plant *Arabidopsis thaliana* has a chloroplast genome of ~155 kb with 87 protein-coding genes, which is just a small fraction of extant cyanobacterial genomes. On the other hand, bioinformatic and proteomic predictions assume in excess of 3,000 proteins in this organelle (Sun et al., 2004; Zybailov et al., 2008). Expression and re-import of these nucleus-encoded plastid targeted proteins has been predicted for a long time to be a hallmark for a successful organelle establishment. An intriguing question is to what extent and for what purpose cyanobacterial genes had been transferred and maintained in the host nucleus. According to Sato et al. (2005) and Martin et al. (2002), the *A. thaliana* nuclear genome harbors between 1,200 and 4,500 genes of cyanobacterial origin, respectively. The majority of these genes have non-plastid function, hence, the present plastid proteome evolved from multiple sources including cyanobacterial, host proteins, and eubacterial genes acquired by horizontal gene transfer. Data from glaucophytes and red algae suggest a much smaller portion of transferred cyanobacterial genes and in contrary to *A. thaliana*, the majority of these candidates are plastid targeted (Reyes-Prieto et al., 2006). Although these findings are not in full agreement, they still demonstrate a significant and complex contribution of EGT to the evolution of the host cell. Host genes and endosymbiont genes have been replaced by one another, duplicated, diverged and several times redirected to a new compartment or even existing homologous genes recombined with each other.

In this chapter, we will address the question how the evolving plastid was connected and integrated into the metabolic network of its host cell. We will discuss that integration of host and chloroplast metabolism was predominantly a host-driven process that involved recruitment of many host-derived metabolite transporters to the plastid envelope membrane. We will also discuss alternative syntrophic scenarios of plastid origin.

II. Metabolic Links Between Extant Plastids and Cytosol

Plastids are metabolically extraordinary versatile and active cellular organelles. They use solar energy to convert carbon dioxide and water to carbohydrates in the process of photosynthesis. In addition, a glut of other anabolic and catabolic routes are located in the plastid stroma, including nitrogen and sulfur assimilation, and fatty acid, amino acid, and terpenoid biosyntheses (Weber and Flügge, 2002; Weber, 2006). The exchange of metabolic precursors and intermediates, and of the end products of plastid-localized metabolic pathways requires considerable routing of small molecules across the plastid envelope membrane (Tegeder and Weber, 2006). Most solutes pass the outer envelope membrane of primary plastids, which has the characteristics of a wide-meshed molecular sieve, through a range of substrate-specific

and general porins (Flügge, 2000; Flügge and Benz, 1984; Soll et al., 2000). The inner envelope membrane, however, represents the actual permeability barrier and selectivity filter between plastid stroma and surrounding cytosol (Flügge, 1998; Flügge and Heldt, 1991) and solute transport across the inner envelope membrane is catalyzed by highly specific metabolite transporters (Flügge, 1992, 1998). In the following paragraphs, we will give a brief overview of plastid-localized metabolic pathways and review what is currently known about plastid envelope membrane transporters. Additional details can be found in several recent reviews covering this topic (Block et al., 2007; Linka and Weber, 2010; Tegeder and Weber, 2006; Weber, 2004, 2006; Weber and Fischer, 2007; Weber et al., 2005).

A. Overview of Plastid-Localized Metabolic Pathways

Figure 8.1 presents an overview of the many functions that plastids have. In this section a variety of these functions will be highlighted.

1. Primary Carbon Metabolism

Chloroplasts are the exclusive sites of photosynthetic carbon assimilation in eukaryotic photosynthetic cells. Triose phosphates, which are the net products of carbon dioxide assimilation through the Calvin-Benson cycle, represent the principal precursors for all other biosynthetic routes in photosynthetic eukaryotes. Recently assimilated carbon is allocated to storage (e.g., starch), soluble (e.g., sucrose), and structural (e.g., cellulose) carbohydrate biosynthesis, nitrogen and sulfur metabolism, fatty acid biosynthesis, secondary metabolism, and a host of other metabolic pathways. Carbon allocation to these pathways is regulated by control of enzyme activity at the transcriptional, translational, and posttranslational levels and by allosteric regulation. In addition, sub-cellular compartmentation, and, in multicellular plants and algae, the distribution of specific pathways between different cell types and tissues provides additional options for metabolic regulation.

Inorganic carbon dioxide is assimilated into organic C-compounds in the chloroplast stroma by the Calvin-Benson Cycle (reductive pentose phosphate pathway). Ribulose 1,5 *bis*phosphate carboxylase/oxygenase (Rubisco) catalyzes the carboxylation of the carbon dioxide acceptor ribulose 1,5-*bis*phosphate (RubP), yielding two molecules of 3-phosphoglyceric acid (3-PGA). Reduction of 3-PGA by the consecutive actions of glyceraldehyde-3-phosphate (GAP) phosphoglycerate kinase and NADPH-dependent glyceraldehyde phosphate dehydrogenase (GAPDH) yields GAP, which is freely interconvertible with dihydroxyacetone 3-phosphate (DHAP) by the activity of triosephosphate isomerase (TIM). One

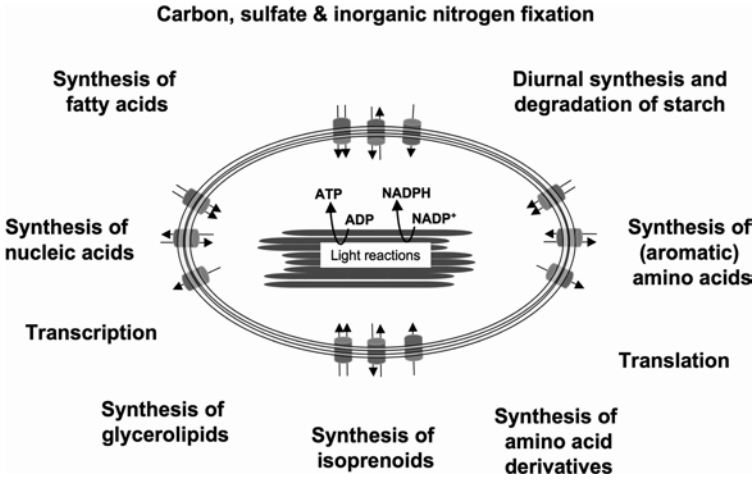


Fig. 8.1. Schematic representation of major plastid localized pathways and functions. Depending on the differentiation and developmental stage of the plastid, the organelle either depends on supply with amino acids, energy, and nucleotides to drive plastidial transcription and translation or it serves as a major production site for a large array of compounds in photosynthetic eukaryotes. Not all of the listed pathways, such as starch biosynthesis, are plastid-localized in all photosynthetic eukaryotes.

out of six synthesized triose-phosphates (TP) can be withdrawn from the Calvin-Benson Cycle for biosynthetic reactions whereas the remainder is required for regeneration of the CO_2 acceptor RubP. TPs can either be retained inside the chloroplast to enter plastid-localized pathways, or they can be exported to the cytosol in strict counter-exchange with *ortho*-phosphate by a triose-phosphate/phosphate translocator (TPT, see below) that resides in the inner chloroplast envelope membrane (Flügge, 1999). While a large share of recently assimilated carbon is directed into starch and sucrose biosynthesis, a sizable portion serves as precursor for various other primary and secondary metabolic routes. Reduced carbon can be either directly withdrawn from the Calvin-Benson Cycle, for example as erythrose 4-phosphate (E4P) to drive the shikimic acid pathway, or it is exported to the cytosol as TPs that serve as precursor for the biosynthesis of organic acids, amino acids, etc. (Tegeder and Weber, 2006).

2. Assimilation of Inorganic Ions (Sulfate, Nitrate)

Inorganic nitrogen assimilation frequently represents the second-most abundant metabolic flux in plastids. The incorporation of inorganic nitrogen into organic compounds such as glutamate is usually an essential biochemical pathway in plant cells, although it must be noted that plants and many algae are able to survive and thrive on amino acids as sole source of nitrogen (Forsum et al., 2008; Komarova et al., 2008; Lee et al., 2007; Rentsch et al., 2007; Svennerstam et al., 2008). The principal form of nitrogen that is converted into organic N-containing compounds is ammonia. If ammonia is not taken up from the external medium, it can be generated from nitrate by the consecutive actions of assimilatory NADH-dependent nitrate reductase (NR) that resides in the cytosol and of nitrite reductase (NiR), which is localized in the plastid stroma. The intermediate nitrite must thus be transported from the cytoplasm into the chloroplast stroma. Ammonia is then assimilated by the joint action of glutamine synthetase (GS) and ferredoxin or NADH-dependent glutamate synthase (Fd/NADH-GOGAT) (Ireland and Lea, 1999; Miflin and Lea, 1976, 1980).

The reaction is as follows:

$$\text{2-oxoglutarate} + \text{glutamate} + \text{ATP} + 2\,\text{Fd}_{\text{red}} + \text{NH}_4^+ \rightarrow 2\,\text{glutamate} + \text{ADP} + \text{P}_\text{i} + 2\,\text{Fd}_{\text{ox}}$$

In summary, the GS/GOGAT cycle produces one molecule of glutamate from one molecule of each, 2-oxoglutarate and ammonia. The GS/GOGAT system is not only required for primary ammonia assimilation but also for re-assimilation of ammonia that is produced by the mitochondrial glycine decarboxylase complex during photorespiration (Lea et al., 1978; Linka and Weber, 2005). While all land plants analyzed to date possess cytosolic and plastidial isoforms of GS (e.g., Brugière et al., 2000; Kichey et al., 2005; Paczek et al., 2002), the currently known red algal genomes encode only a cytosolic isozyme of glutamine synthetase (Terashita et al., 2006). In plants, the various GOGAT isozymes are nuclear encoded, whereas it is encoded by a single gene on the plastid genome of red algae (Glöckner et al., 2000; Ohta et al., 2003), suggesting that the gene product is exclusively localized in the chloroplast stroma.

The third inorganic compound in addition to carbon dioxide and ammonia that is assimilated in the plastid stroma is sulfate. Cysteine is the first organic product resulting from assimilation of inorganic sulfur. It thus represents the precursor for all other sulfur-containing compounds in plants (Leustek et al., 2000; Martin et al., 2005; Rotte and Leustek, 2000). Also the *de novo* biosynthesis of methionine is localized in plastids (Ravanel et al., 2004). While chloroplasts of the green alga *Chlamydomonas reinhardtii* possess a tri-partite ABC transporter system for the uptake of sulfate (Chen and Melis, 2004; Chen et al., 2003), it is currently unknown how sulfate is taken up into the chloroplasts of land plants. A sulfate transporter previously assigned to the chloroplast envelope of *Arabidopsis thaliana* (Takahashi et al., 1999) was recently shown to be localized at the tonoplast (Kataoka et al., 2004).

3. Amino Acids, Fatty Acids, Isoprenoids

In most photosynthetic organisms, over 50% of the total polar lipids consist of galactolipids, such as monogalactosyldiacylglycerol (MGDG) and digalactosyldiacylglycerol (DGDG) (Kelly and Dormann, 2004). In many photosynthetic eukaryotes, such as Arabidopsis, galactolipids are derived from two distinct pathways that localize to

plastids (prokaryotic) and the ER (eukaryotic), respectively. In Arabidopsis, both pathways contribute almost equally to galactolipid synthesis (Benning et al., 2006). In the prokaryotic pathway, fatty acids synthesized in plastids are directly incorporated into polar lipids. In the eukaryotic pathway, however, fatty acids are transported from the plastid to the ER where the eukaryotic lipids are synthesized (Kelly and Dormann, 2004). The molecular nature of the compound that is shuttled back to the plastids is not yet completely understood, but it is likely that phosphatidylcholine and/or phosphatidic acid (PA) are transported from the ER to the plastids where they are converted to diacylglycerol (DAG), the precursor for galactolipid synthesis. Hence, plastidic lipid biosynthesis requires intense traffic of fatty acids and lipids between plastids and ER. Recently, forward genetics using the Arabidopsis model provided the first insights into the mechanism of fatty acid shuttling between ER and plastids. Two non-allelic mutants, which accumulated the unusual lipid trigalactosyldiacylglycerol (*tgt1*, *tgt2*), were identified that accumulate oligogalactolipids and phosphatidic acid (Awai et al., 2006a; Xu et al., 2003, 2006). *TGT1* (*At1g19800*) encodes a membrane protein with similarity bacterial ABC transporters (Xu et al., 2003, 2006) and *TGT2* (*At3g20320*) encodes a putative substrate binding protein (Awai et al., 2006b). Both genes are in bacteria organized in a single operon suggesting that they are involved in the same process. It was further shown that TGT2 binds phosphatidic acid, indicating it is indeed a binding protein. The current model of phosphatidic acid transport between ER and plastid thus includes TGT1 and TGT2 as residents of the inner plastid envelope membrane. It was hypothesized that the binding protein TGT2 binds PA and delivers it to the TGT1 permease, which transports it into the plastid (Awai et al., 2006b).

In addition to fatty acid biosynthesis, chloroplasts also represent the major site of amino acid (AA) biosynthesis in plants. For example, the minor amino acids Val, Arg, Ile, His, Leu, Trp, Tyr, Met, Phe, Cys, and Lys, and the major amino acid Glu are predominantly or exclusively synthesized in the plastid stroma. Based on the fact that several of the enzymes required for Thr biosynthesis carry plastid targeting signals, it is believed that Thr biosynthesis is also localized in plastids (Curien et al., 1996; Lee and Leustek, 1999). Further, at least in land plants, chloroplasts are the sole site of *de novo* glutamate biosynthesis from 2-OG by the various GOGAT isozymes (Coruzzi, 2003). Other amino acids, such as Asp, Gln, Ser, and Gly, can be synthesized in multiple compartments. The sole or predominant localization of some amino acid biosynthetic pathways in the plastid stroma requires (i) import systems for carbon precursors and cofactors of amino acid biosynthetic pathways and (ii) export systems for amino acids that are required in other cellular compartments for protein biosynthesis or as amino group donors for other biosynthetic pathways. In addition, amino uptake systems are needed for those amino acids that are not synthesized inside the plastid (e.g., proline) but are necessary for protein biosynthesis in the plastid stroma. Similarly, mitochondria also need amino acid uptake systems to provide mitochondrial protein biosynthesis with precursors. With the exception of the plastidic glutamate/malate transporter DiT2, which serves as the major Glu export pathway from plastids (Renné et al., 2003), none of the plastidic amino acid transporters has been unequivocally identified at the molecular level.

Plants harbor two distinct pathways for isoprenoid biosynthesis, one that is localized in the cytoplasm, the other one residing in the chloroplast stroma. The latter pathway, known as the 2-methylerythritol-4-phosphate (MEP) pathway, was only recently discovered (Lichtenthaler et al., 1997; Schwender et al., 1996). The cytosolic pathway, also known as the mevalonic acid (MVA) pathway, uses acetyl-CoA as precursor for the biosynthesis of the isoprenoid precursor isopentenyl pyrophosphate (IPP), whereas the plastidial MEP pathway requires pyruvate and GAP as precursors (Lichtenthaler et al., 1997). Most, if not all plastid-synthesized isoprenoids, such as carotenoids, tocopherol, isoprene, and the phytol side chain of chlorophyll, are derived from the MEP pathway. GAP, at least in photosynthetically active plastids, can be directly withdrawn from the Calvin cycle, or, in non-photosynthetic plastids, from the reductive pentose phosphate pathway. However, at least in most land plants, pyruvate must be imported from the cytosol because photosynthetic plastids lack the glycolytic sequence from GAP to PEP (Bagge and Larsson, 1986; Borchert et al., 1993; Van der Straeten et al., 1991).

The corresponding pyruvate transporters have been biochemically characterized in a number of C_4 (Aoki et al., 1992; Huber and Edwards, 1977) and C_3 (Proudlove and Thurman, 1981) plants but the corresponding genes are unknown. Alternatively, pyruvate could be generated from PEP that is taken up by the PEP/phosphate translocator by pyruvate kinase (Andre and Benning, 2007; Andre et al., 2007; Voll et al., 2003). The plastidal MEP and the cytosolic MVA pathways for IPP biosynthesis are not fully separated, but communicate through the exchange of intermediates across the plastid envelope membrane (Hemmerlin et al., 2003). Transport of DOXP and IPP across the plastid envelope membrane was demonstrated in isolated plastids, thus providing a mechanism for the cross talk between the MEP and MVA pathways (Flügge and Gao, 2005). Based in kinetic and inhibitor studies, it was concluded that DOXP is preferentially transported by the xylulose 5-phosphate/phosphate translocator (XPT) (Eicks et al., 2002; Flügge and Gao, 2005). IPP is likely taken up by facilitated diffusion (Flügge and Gao, 2005).

B. Overview of Solute Transporters Residing in the Plastid Envelope Membrane

Figure 8.2 presents an overview of many of the transporters in the plastid membranes. Some of these will be discussed in the following sections.

1. Phosphate Translocators as Major Carbon Exporters and Importers

In all photosynthetic eukaryotes analyzed to date, the major pathway for the export of recently assimilated carbon from the chloroplast is via triose-phosphate/phosphate antiporters. These translocators catalyze the strict counter-exchange of one molecule of triose-phosphate with one molecule of inorganic phosphate, thereby providing a route for the export of phosphorylated and reduced organic carbon while maintaining the phosphate homeostasis of the plastid. Coupling of triose-phosphate to phosphate counter-exchange also provides a means for metabolic coordination between plastid and cytosol, such as between starch biosynthesis in the plastid and sucrose

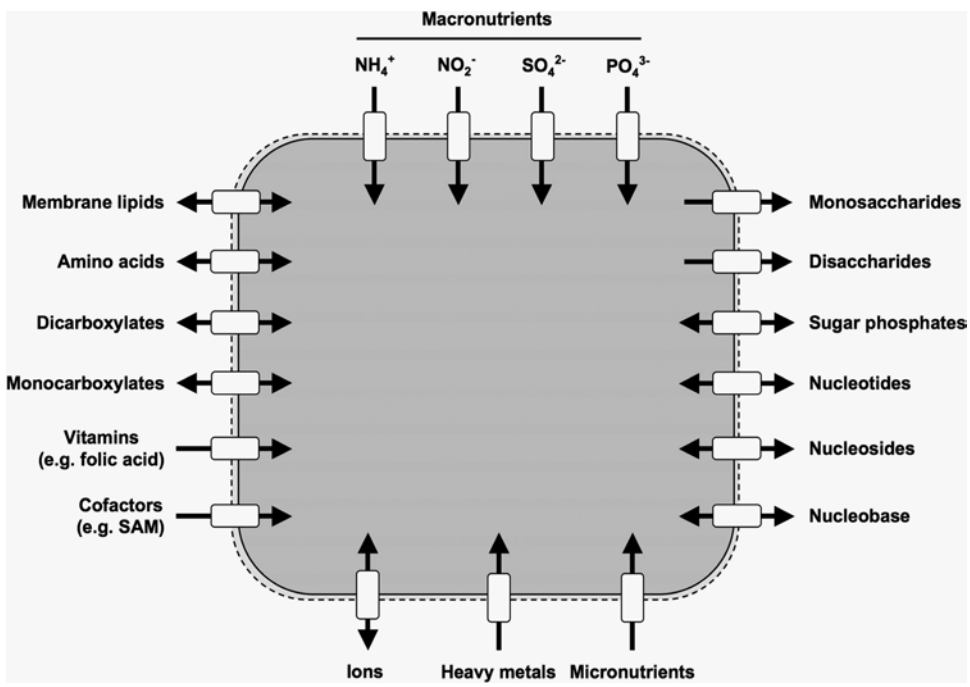


Fig. 8.2. Schematic representation of major plastid-localized transporters. Due to the complex differences between the plastid permeomes of the different eukaryotic photosynthetic lineages, not all transporters are present in plastids of all phylogenetic groups.

biosynthesis in the cytosol in the Chloroplastida. Multiple recent reviews have covered the biochemical and functional aspects of plastidic phosphate translocators (Fischer and Weber, 2002; Flügge, 1999; Flügge et al., 2003; Weber, 2004; Weber et al., 2004a); we therefore only briefly introduce the characteristics of this important carbon transporter family as required for understanding their role in the evolutionary context.

In the model land plant *A. thaliana*, four different subfamilies of plastidic phosphate antiporters can be detected, which have partially overlapping but distinct specificities for phosphorylated C3-, C5- and C6-compounds that are exchanged with inorganic phosphate. As discussed in detail in this chapter, the phosphate translocators (PTs) all belong to the Drug/Metabolite Transporter Family (DMT-family), which is ubiquitously present in all eukaryotic cells, but absent from prokaryotes. Plastidic phosphate translocators have evolved from endomembrane-localized members of the DMT-family.

The triose phosphate/phosphate translocator (TPT) provides the main pathway for carbon export from chloroplasts during the day. Triose-phosphates that can be withdrawn from the Calvin-Benson Cycle are exported to the cytosol in strict counter-exchange with inorganic P_i. Coupling of TP-export to P_i import is important because ATP biosynthesis by the light reactions of photosynthesis requires P_i as a substrate. While in the Rhodophyceae, TPT likely also serves to supply the plastid stroma with reduced carbon in the absence of photosynthesis, this role is taken over by specialized hexose-phosphate and pentose-phosphate transporters in the Chloroplastida.

In the Chloroplastida, plastids of non-green tissues depend on carbon import from the cytosol in the form of glucose 6-phosphate by the glucose 6-phosphate/phosphate translocator (GPT; Kammerer et al., 1998). While the genome of Arabidopsis encodes two paralogous GPT genes with identical substrate specificity (AtGPT1 and AtGPT2), the orthologous gene of the red alga *G. sulphuraria* does not accept hexose- or pentose-phosphates as substrates in a physiological range of concentrations (Linka et al., 2008). In the Chloroplastida, the GPT serves to provide glucose 6-phosphate as substrate for the oxidative pentose phosphate cycle to create reducing power and as a precursor for starch biosynthesis in amyloplasts (Niewiadomski et al., 2005).

A third subtype of plastidial phosphate translocators catalyzes the strict counter-exchange of phosphoenolpyruvate (PEP) with P_i and is thus called the phosphoenolpyruvate/phosphate translocator (PPT, Fischer et al., 1997). The main function of the PPT is the import of PEP into chloroplasts where, together with E4P that can be withdrawn from the pentose phosphate pathway, it is required as precursor for the plastid-localized shikimate pathway, from which aromatic amino acids and precursors for phenolic secondary plant products derive (Fischer et al., 1997; Streatfield et al., 1999; Voll et al., 2003). The housekeeping function of the PPT in providing PEP for the shikimate pathway is also conserved in the red algae, where the orthologous gene has identical substrate specificity and kinetic constants as its green plant relative (Linka et al., 2008).

The last functionally characterized member of the plastidal phosphate translocator family was discovered relatively recently. It exchanges pentose phosphates, such as xylulose 5P and ribose 5P with inorganic phosphate and is thus called the xylulose 5P/phosphate translocator (XPT, Eicks et al., 2002). In Arabidopsis, the XPT is thought to exchange pentose phosphates between the cytosol and the plastids in which pentose phosphates are intermediates of both the oxidative pentose phosphate pathway and the Calvin-Benson cycle, although the physiological role of the XPT has not yet been demonstrated by knock-down or knockout lines. Plastids of the red algae do not possess a pentose phosphate translocator (Linka et al., 2008).

2. Transporters of Mono- and Disaccharides, Amino Acids, Organic Acids, Nucleotides, Metabolic Intermediates, Inorganic Ions

Common to the plastids of all photosynthetic eukaryotes are also the non-mitochondrial ATP transporters or nucleotide transporters (NTTs, Kampfenkel et al., 1995; Linka et al., 2003; Winkler and Neuhaus, 1999). These transporters serve to provide the plastid stroma with ATP in the absence of photosynthetic ATP production or in specialized plastid subtypes that do not harbor a photosynthetic electron transport chain and ATP synthase complex (Reinhold et al., 2007; Reiser et al., 2004). While the plastidial phosphate translocators and the ATP-transporters are likely a conserved feature of most, if not all,

photosynthetic eukaryotes, several other transporter families are exclusive to the Chloroplastida or have been acquired through secondary or tertiary endosymbioses in the Chromalveolates (Linka et al., 2008; Tyra et al., 2007). For example, the uptake of 2-oxoglutarate into and the export of glutamate from chloroplasts (Weber and Flügge, 2002) by a two-translocator system consisting of an 2-oxoglutarate/malate antiporter (DiT1) and a glutamate/malate antiporter (DiT2) (Schneidereit et al., 2006; Weber and Flügge, 2002; Weber et al., 1995; Woo et al., 1987) is a specific feature of the green lineage and some protists with secondary plastids, with the latter ones probably originating from lateral gene transfer from a member of the Chloroplastida (Tyra et al., 2007). Plastidic dicarboxylate translocators are not encoded by the genomes of red algae or diatoms (Armbrust et al., 2004; Barbier et al., 2005a, b; Matsuzaki et al., 2004; Weber, 2006; Weber et al., 2004b).

Also the presence of monosaccharide (i.e., glucose) and disaccharide (i.e., maltose) transporters is a unique feature of the plastid envelope membrane of the Chloroplastida. The glucose (Weber et al., 2000) and maltose transporters (Niittyla et al., 2004) serve to export the breakdown products of starch (i.e., glucose and maltose) during the night (Fischer and Weber, 2002; Weise et al., 2004). Since the presence of starch in plastids is a specific feature of the Chloroplastida (Deschamps et al., 2008), it is not surprising that the presence of transporters for starch breakdown products is exclusive to the Chloroplastida. While the plastidial glucose transporter has clearly evolved from monosaccharide transporters of the plasma membrane and other internal membranes, the origin of the maltose transporter is unclear (Tyra et al., 2007).

As indicated in Fig. 8.2, a large variety of small molecules are transported across the chloroplast envelope, requiring the presence of a diverse set of inorganic ion and metabolite transporters. While some of these transporters will be discussed in more detail in the following paragraphs, a complete coverage of all plastid envelope membrane transporters known to date is beyond the scope of this chapter. The readers are referred to several recent reviews covering this topic (Block et al., 2007; Linka and Weber, 2010; Neuhaus and Wagner, 2000; Tegeder and Weber, 2006; Weber, 2006; Weber and Fischer, 2007; Weber et al., 2004b, 2005).

III. Establishment of Endosymbiosis and Evolution of Metabolic Links Between Plastid and Cytosol

A multitude of different scenarios for chloroplast origin can be envisaged. One commonly discussed scenario is the phagotrophic origin of plastids. Under this assumption, a primitive mitochondriate eukaryote made prey on free-living cyanobacteria, likely by phagocytosis. Eventually, a captured cyanobacterium remained intact in the eukaryotic food vacuole, continuing with photosynthetic CO_2 assimilation. Somehow the eukaryote managed to benefit from cyanobacterial photosynthesis, thus providing an evolutionary benefit to the association. However, converting an intracellular symbiont into an organelle requires more than just export of reduced carbon to feed the host. Among other prerequisites, it requires synchronization and coordination of cell cycle and cell division. That is, the cyanosymbiont can only be passed on to the next generation if it undergoes cell division before the host cell divides; otherwise one daughter cell would remain without a plastid. Conversely, also uncontrolled proliferation of the cyanobiont within the host cell must be prevented. Possibly, consecutive division of cyanobiont and host cell initially solved this problem. Under this hypothesis, cyanobiont division would trigger eukaryotic cell division, a scenario that does not require control of the host cell over the cyanobiont cell cycle. In fact, even in extant red algae and land plants, duplication of nuclear DNA before mitosis requires duplication of plastidal DNA. Completion of organellar DNA duplication is communicated to the host cell nucleus by a tetrapyrrole signal, which might be a remnant of the ancient pathway that coordinated host with cyanosymbiont division (Kobayashi et al., 2009).

In addition to the coordination of cell cycle and division, multiple other requirements had to be fulfilled before the cyanobiont evolved into a *bona fide* organelle. Endosymbiotic gene transfer (EGT) from the genetically isolated endosymbiont to the host cell nucleus was likely required to evade the accumulation of deleterious mutations in the non-recombining organellar genome (Muller's ratchet, Muller, 1964), although the remaining chloroplast-encoded genes are seemingly protected from the accumulation of deleterious mutations by a gene conversion mechanism (Khakhlova

and Bock, 2006). EGT appears to be frequent in extant plants and algae, and has recently been observed even between algae and sea slugs (Rumpho et al., 2008). Also, horizontal gene transfer is frequently observed in phagotrophic protists (Andersson et al., 2003; Richards et al., 2003). This indicates that the actual gene transfer did not represent a major hurdle to the establishment of endosymbiosis. However, while EGT likely protected cyanobacterial genes from deleterious mutations by genetic recombination, provided selective advantage to the protoalga through additional functions, and contributed to establishing host control over plastid gene expression, it also generated a new issue – somehow some of the now nuclear-encoded gene products that are required for plastid function must make their way back to the plastid. This required the establishment of a protein targeting and import system that enabled efficient import of nuclear-encoded and cytosolically translated proteins into the chloroplast. A detailed discussion of this process is beyond the scope of this chapter, but it is reasonable to hypothesize that protein targeting to the plastid initially made use of both the eukaryotic and the cyanobacterial protein secretion systems, the latter eventually evolving into what is now known as the plastidial protein translocation apparatus.

A critical point early during establishment of a functional endosymbiotic relationship between a eukaryotic heterotrophic mitochondriate host and prokaryotic photosynthetic symbiont was certainly the connection of the metabolism of both organisms. This leads to the questions of: (i) how did the eukaryotic host benefit from the cyanobacterium's ability to carry out oxygenic photosynthesis and *de novo* biosynthesis of carbohydrates from CO_2 and water, and (ii) in which way, if at all, did the cyanobacterial symbiont benefit from this interaction. The latter point is also important with respect to whether endosymbiosis can be considered a true symbiotic interaction that includes mutual benefits to both partners or whether the origin of the plastid was an unfriendly takeover, not a merger of equals.

Several different scenarios can be envisaged for the host to benefit from cyanobacterial photosynthetic carbon fixation: (i) excretion of fermentation products from the cyanobacterium; (ii) excretion of hydrogen from the cyanobacterium; (iii) excretion of glycolate from the cyanobacterium; and (iv) direct connection between the glycolytic pathways of cyanobacterium and host. Whereas (i) through (iii) do occur in free-living cyanobacteria, (iv) does not and it most likely requires the targeting of a membrane transporter from the host into the membrane surrounding the endosymbiont.

Also, several interactions beneficial to the cyanobacterium can be envisaged. Possibly, the close association of a cyanobacterium with a heterotrophic eukaryote might have permitted a more efficient cyanobacterial dinitrogen fixation because of (i) shielding of nitrogenase from molecular oxygen; (ii) efficient scavenging of molecular hydrogen by the host cell; and (iii) supply of the cyanobacterium with ATP. Whereas (i) and (ii) can be explained by diffusion of gases between host and symbiont, (iii) requires active transport of ATP from host cells into the cyanobacterium.

The above-mentioned syntrophic associations between eukaryotic heterotrophic cells and photosynthetic cyanobacteria are not easily compatible with the currently dominating phagotrophic model of plastid origin. However, the syntrophic model would provide time for the step-wise establishment of endosymbiosis, the final step being the permanent engulfment of a closely associated cyanobacterium by the eukaryotic cell. In the following paragraphs, we will discuss possible scenarios for syntrophic associations between photosynthetic cyanobacteria and eukaryotic cells that might have predated the endosymbiotic origin of plastids. While this discussion is not meant to discount the phagotrophic model of plastid origin, in particular since the phagotrophic model has gained additional support from the recent discovery of a second case of primary endosymbiosis in the filose thecamoeba *Paulinella chromatophora* (Marin et al., 2005), it is intended to highlight alternative scenarios that could have contributed to the initiation and success of the most important merger between two independent organisms in the history of our planet.

A. Anaerobic Cyanobacterial Carbon and Nitrogen Metabolism as a Possible Driver for Attracting a Mitochondriate Eukaryote into a Syntrophic Association

Some cyanobacteria that are able to carry out fermentation excrete the fermentation end products to the environment (Heyer and Krumbein, 1991;

Heyer et al., 1989; Vanderoost et al., 1989). Cyanobacterial fermentation occurs only under micro-aerobic or anaerobic conditions and is inhibited in the light (Heyer and Krumbein, 1991). Whereas the endosymbiont engulfed by a eukaryotic host cell was certainly able to produce soluble sugars, starch, and oxygen during the day, it was likely subject to micro-aerobic conditions in the dark and thus likely had to satisfy or complement its energy metabolism by fermentation. The fermentation products (e.g., lactate, acetate, formate, or ethanol) were excreted by the cyanobacterium and might have served as a carbon source to drive the host's energy metabolism and ATP biosynthesis by aerobic mitochondrial respiration. This scenario does not require any evolutionary inventions.

1. Cyanobacterial Nitrogen Fixation Provides Hydrogen to Eukaryotic Energy Metabolism

Another possible benefit of the host cell from the interaction with a cyanobacterium can be considered a casual extension of the hydrogen hypothesis that was originally put forward to explain the evolution of the first mitochondriate eukaryote (Martin and Muller, 1998). The engulfed unicellular cyanobacterium was likely able to fix atmospheric dinitrogen by using the enzyme nitrogenase (Deschamps et al., 2008; Deusch et al., 2008). Nitrogenase is highly sensitive to oxygen; therefore many present-day cyanobacterial species that are able to fix nitrogen have evolved specialized cells, called heterocysts, to spatially separate the process of nitrogen-fixation from the oxygen-generating process of photosynthesis and to thus protect the oxygen-sensitive nitrogenase from molecular oxygen. Unicellular non-colony forming cyanobacteria such as the diazotrophic cyanobacterium *Cyanothece* sp. do not have the option to spatially separate photosynthesis and nitrogen fixation. Instead, they temporally separate photosynthesis from dinitrogen fixation (Bergman et al., 1997; Reddy et al., 1993). That is, during the day, light energy is used to produce sugars and glycogen, whereas during the night, the chemical energy stored in these carbohydrates is used to drive dinitrogen fixation by nitrogenase. The dinitrogen reaction produces molecular hydrogen (Alberty, 2005; Tamagnini et al., 2002), which, if not re-oxidized by uptake hydrogenase, leaves the cyanobacterial cytoplasm by diffusion (Bothe et al., 2008; Roeselers et al., 2008) to enter the surrounding host cytoplasm. At this point, it is helpful to briefly recapitulate the hydrogen hypothesis (Martin and Muller, 1998). According to this hypothesis, the host cell of the hydrogen-producing α-proteobacterium that evolved into mitochondria and/or hydrogenosomes was an anaerobic, strictly hydrogen-dependent autotrophic archaeon. The hydrogen hypothesis posits that the dependence of the host on molecular hydrogen forged its interaction with a hydrogen-producing precursor of mitochondria. It further posits that the host's ability to use molecular hydrogen as energy source was lost in the course of establishing the symbiosis between the mitochondrion and its host cell and was replaced by a glycolytic pathway derived from the α-proteobacterium (Martin and Muller, 1998). The time scale of this process is unclear, but it is reasonable to posit that the ability to carry out glycolysis coexisted for some time with the ability to use molecular hydrogen as an energy source. Hence, hydrogen generated by cyanobacterial nitrogenase may have been an attractive energy source for a mitochondriate eukaryotic cell that was still able to use molecular hydrogen. Possibly, dihydrogen and CO_2 would have been converted to acetyl-CoA and further to pyruvate by the acetyl-CoA pathway in the host cell cytosol. Pyruvate would enter the mitochondria and would be respired by the mitochondrial electron transport chain, thus using molecular oxygen (thus protecting the nitrogenase reaction from oxygen) and producing CO_2 (required for the acetyl-CoA pathway) and generating ATP.

2. Possible Benefits of a Syntrophic Association with the Cyanobacterial Partner

This hypothetical hydrogen-based pathway is attractive because it also provides important benefits for the cyanobacterium: (i) enclosure in the host cell cytosol and aerobic mitochondrial respiration decrease oxygen partial pressure in the dark and thus contribute to efficient dinitrogen fixation; (ii) moreover, use of molecular hydrogen by the cytosolic acetyl-CoA pathway removes molecular hydrogen from the equilibrium of the nitrogenase reaction and thus drives the nitrogenase reaction forward. These two points are

important because they provide selective advantage to a cyanobacterium entering a syntrophic (and, eventually, endosymbiotic) relationship with a mitochondriate eukaryote: nitrogen was certainly a limiting factor for growth and dinitrogen fixation is an extremely energy consuming process, costing 16 ATP, 8 electrons and 8 protons per mole of dinitrogen fixed (Alberty, 2005). Even a slight increase in the efficiency of the reaction by efficient removal one of the reaction products (i.e., H_2) would have certainly benefited the cyanobacterium and would have caused a significant competitive advantage for the cyanobacterium cohabiting with an eukaryotic mitochondriate host. In addition, the reaction was protected from inhibitory oxygen, which became increasingly important due to the significant rise of atmospheric oxygen levels 2.2 GYA (Falkowski and Godfrey, 2008; Goldblatt et al., 2006).

Another significant possible benefit to the symbiotic cyanobacterium would have been nocturnal supply with ATP by the host cell. As stated above, nitrogen fixation by nitrogenase is an extremely energy-demanding process. In addition, much of the invested energy, if not recovered by uptake hydrogenase, is lost in the form of hydrogen, an inevitable by-product of the nitrogenase reaction. Conversion of this hydrogen to ATP by the host cell and export of some of this ATP to the cyanobacterium could have generated a massive benefit for the nitrogen-fixing symbiont. However, in contrast to molecular hydrogen, ATP is not membrane-permeable and a transporter is required to catalyze its efficient transport across a lipid bilayer membrane (Winkler and Neuhaus, 1999). Two types of adenylate transporters are known, the mitochondrial and the non-mitochondrial adenylate translocators (Winkler and Neuhaus, 1999). In extant plants, the plastidic adenylate transporter is a member of the non-mitochondrial adenylate transporter family and it primarily serves to supply plastidic dark metabolism with ATP from the cytosol (Reinhold et al., 2007). It was very likely already present in the ancestor of all eukaryotic photosynthetic cells, possibly before the cyanobacterial precursor of the chloroplast was engulfed (Linka et al., 2003). This hypothesis is supported by the fact that the ATP-transporter is encoded by the nuclear genomes of all three members of the Archaeplastida and was thus likely already present in the protoalga (Linka et al., 2003). This protein might have served initially to supply cyanobacterial dinitrogen fixation with ATP generated by mitochondrial respiration, thus providing an important incentive for the cyanobacterium to enter a symbiotic relationship. In this respect, the cyanobacterium might be considered an energy parasite. Importantly, the Plantae host very likely was inhabited by a chlamydial energy parasite before the entry of the cyanobacterium (Moustafa et al., 2008), thus the host cell was already accustomed to a drain on its ATP pool. Moreover, the plastidial ATP importer is clearly of chlamydial origin (Moustafa et al., 2008; Schmitz-Esser et al., 2004; Tyra et al., 2007). While highly speculative, it can be hypothesized that the highly ATP-demanding process of dinitrogen fixation led to very low ATP concentrations in the host cell cytosol, thereby depriving the chlamydial endoparasite of ATP and thus contributing to its elimination from the host cell.

To sum up this alternative scenario of plastid origin, we hypothesize that the initial establishment of a syntrophic and later on symbiotic relationship between a mitochondriate eukaryote and a cyanobacterium was driven by the exchange of hydrogen gas between both organisms, similar to what lead to the evolution of the ancestor of all eukaryotic cells according to the hydrogen hypothesis. The main benefit to the host would have been nocturnal supply with hydrogen as an energy source and a more efficient nitrogen fixation would have lured the cyanobacterium into the mutually beneficial relationship. This did not require any evolutionary inventions, is compatible with biochemical pathways found in existing organisms, and it lends (and draws) additional support to (from) the hydrogen hypothesis. In addition, strong evidence points to an evolutionary deep origin of the plastidic adenylate translocator (Moustafa et al., 2008; Schmitz-Esser et al., 2004; Tyra et al., 2007), indicating this protein was already present before the split of photosynthetic eukaryotes, possibly emerging simultaneously with the cyanobacterial endosymbiont. Targeting of the adenylate translocator to the cyanobacterial membrane would have enabled the host to supply the endosymbiont with ATP to drive dinitrogen fixation and other energy demanding processes, such as chlorophyll biosynthesis, at night. Moreover, host supply of ATP to the cyanobiont promotes the loss of ATP biosynthesis capability via oxidative

phosphorylation, which is now performed by the mitochondria. In this way, the cyanobiont becomes dependent on the host with respect to nocturnal energy supply. One major drawback of this scenario is that it does not provide significant benefits to the host or symbiont during the light.

3. Benefits of the Host from the Association

Two possible scenarios are proposed for beneficial interactions between host and symbiont metabolism during the light: (i) export of glycolate from the cyanobacterial symbiont and (ii) export of triose phosphates in counter-exchange with phosphate import.

a. Glycolate Export from Cyanobacteria

The key enzyme of CO_2 assimilation by the Calvin Cycle is ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco). Rubisco is a bifunctional enzyme, catalyzing both carboxylation and oxygenation of the acceptor molecule ribulose 1,5-bisphosphate (Tolbert, 1980). Whereas the carboxylation reaction is productive, yielding two molecules of 3-phosphoglyceric acid (3-PGA), the oxygenation reaction yields one 3-PGA and one 2-phosphoglycolate (2-PG). The ratio of oxygenation versus carboxylation reaction is increases with increasing O_2 partial pressures and with increasing temperature. Before the rise of atmospheric oxygen concentrations due to oxygenic photosynthesis, this was not an issue. However, the oxygenation reaction became increasingly problematic with the rising oxygen levels that occurred 2.2 GYA (Falkowski and Isozaki, 2008).

2-PG represents a metabolic dead end inside chloroplasts, but not in cyanobacteria (Eisenhut et al., 2008). In modern plants, 2-PG is dephosphorylated in the chloroplast, yielding glycolate, and glycolate is then exported from plastids to the cytosol. Two molecules of glycolate are converted by a complex pathway that involves peroxisomes and mitochondria into one molecule of glycerate that is imported into chloroplasts, phosphorylated to 3-PGA, and eventually fed into the Calvin cycle (Reumann and Weber, 2006). Many cyanobacteria have evolved carbon-concentrating mechanisms to increase the CO_2 concentration at the side of Rubisco (Raven et al., 2008), thus minimizing the production of glycolate. In addition, glycolate can be exported to the external medium (Heyer and Krumbein, 1991).

Likely, the endosymbiotic cyanobacterium, similar to its descendants, the chloroplasts, was also exporting some glycolate to the surrounding host cell cytosol. Many microorganisms such as *Pseudomonas* spp. (Lau and Armbrust, 2006) or *E. coli* (Furuya and Hayashi, 1963; Hansen and Hayashi, 1962) can grow on glycolate as the sole carbon source, which involves the conversion of glycolate into glycerate. Although the first step of this pathway, the oxidation of glycolate to glyoxylate is present in plants and algae, the eukaryotic glycolate oxidase enzyme shows some important biochemical differences in comparison to its bacterial glycolate dehydrogenase counterpart. Most importantly, the cyanobacterial enzyme consists of three subunits and it uses NAD as an electron acceptor, producing NADH and glyoxylate from NAD and glycolate. In contrast, the plant glycolate oxidase uses molecular oxygen as electron acceptor during glycolate oxidation, thus leading to the production of hydrogen peroxide. The location of eukaryotic glycolate oxidase in the peroxisomes provides a means for detoxification of hydrogen peroxide via catalase. The plastidic pathway via NAD-dependent glycolate dehydrogenase was lost during evolution of land plants and replaced with the complete photorespiratory pathway, as we know it today.

Although the export of glycolate was initially only beneficial to the host, supplying it with a source of carbon and energy, it eventually became beneficial to the chloroplast. Whereas the reduced carbon exported from the primitive plastid to the cytosol was a total loss, 75% of the carbon atoms are shuttled back to the plastid in photosynthetic cells that possess a complete photorespiratory pathway. In an increasingly oxygen-containing atmosphere, this must have represented a significant benefit to the symbiotic partnership.

Nevertheless, the export of glycolate from plastids represents a relatively inefficient means of carbon export because it depends on the rate of the oxygenation reaction. In addition, it does not provide means for feedback regulation since the export of glycolate from the chloroplast is not coupled with host cell metabolism. Also, the conversion of glycerate to triose phosphates, as required for gluconeogenesis, requires reducing equivalents and ATP, which are more abundant in the

photosynthetic plastid than in the heterotrophic cytosol. This may be the reason why the default pathway for glycerate metabolism in most modern photosynthetic eukaryotes is its import into chloroplasts where it is phosphorylated by glycerate kinase and reduced to triose phosphate in the Calvin Cycle. Moreover, this pathway is strictly dependent on photosynthetic activity and thus does not occur in the dark. Therefore, an alternative pathway for carbon export from plastids that is independent of the oxygenation reaction of Rubisco, that delivers ready-to-use reduced carbohydrates, and that is interfaced with the host cell metabolism by a feedback mechanism would provide more selective advantage than dependence on a salvage pathway.

b. Carbon Export by Sugar-Nucleotide Transporters Derived from the Host Endomembrane

The most prominent role of plastids in eukaryotic photosynthetic cells is the provision of reduced carbon to the surrounding cytosol, thus enabling photoautotrophy. As outlined above, a key question in understanding the formation of an endosymbiotic interaction between a cyanobacterium and a eukaryotic cell is how was the metabolic connection between these independent metabolic entities established. It is known for land plants that the major pathway for carbon export from chloroplasts is via the triose-phosphate/phosphate translocator (Flügge and Heldt, 1984). This polytopic membrane protein resides in the inner chloroplast envelope membrane and catalyzes the strict counter exchange of one molecule of triose phosphate with *ortho*-phosphate, thus exporting the products of the Calvin-Benson cycle to the cytosol while maintaining phosphate homeostasis of the plastid (Heldt et al., 1990). Recent progress in genomic sequencing permitted investigating the phylogenetic origin of the plastidal triose phosphate translocator and to analyze whether such plastidic phosphate translocators (pPTs) are a general feature of chloroplasts in all photosynthetic eukaryotic lineages (Weber et al., 2006).

Phylogenic analysis showed that the pPT family originated from nucleotide-sugar transporters that are present in the endomembrane system of all extant eukaryotes (Weber et al., 2006). These sugar nucleotide transporters serve to provide the ER and Golgi systems with activated monosaccharides for glycosylation reactions, such as protein glycosylation and cell-wall polysaccharide biosynthesis (Ashikov et al., 2005; Bakker et al., 2005; Luhn et al., 2004; Seifert, 2004; Suda et al., 2004). However, none of the extant and functionally studied plastidal phosphate translocator family members does transport nucleotide sugars (Flügge et al., 2003), thus leading to the question how a nucleotide sugar transporter might have functioned in the initial establishment of a metabolic connection between cyanobiont and host. This enigma was recently addressed by Deschamps et al. (2008), who put forward an elegant hypothesis explaining the importance of nucleotide-sugar transporters for merging the storage polysaccharide metabolism in host and cyanobiont. According to their hypothesis, both organisms initially harbored complete and independent sets of enzymes for synthesis, storage, and mobilization of carbon in their respective compartments. The insertion of a host-derived NDP-sugar transporter into the cyanobiont's plasma membrane would have allowed the host to withdraw nucleotide sugars, most likely ADP-glucose, from the cyanobiont. The key assumption of this hypothesis is that ADP-glucose represents an activated form of organic carbon that is no longer participating in primary metabolism but has been committed to storage carbohydrate (i.e., starch) biosynthesis. It can thus be withdrawn from the evolving plastid without disturbing its primary carbon metabolism. However, ADP-glucose is a poor substrate for the known cytosolic starch and/or glycogen synthases of extant heterotrophic eukaryotes. To counter this argument, it has been argued that the gene encoding ADP-glucose dependent soluble starch synthase of cyanobacterial origin has been transferred to the host nucleus at a very early stage during the formation of endosymbiosis and was subsequently expressed in the cytosol of the host. Importantly, this would only require the transfer of a single gene from the cyanobacterium to the host cell nucleus. The now nuclear-encoded gene product does not need to be retargeted to the cyanobacterium and the enzymatic machinery for the mobilization of the cytosolic α-glucans are already present. Thus, the proposed scenario is a very simple one in evolutionary terms – it does require only a single gene transfer and 'mistargeting' of a single endomembrane-derived transporter. The scenario would be even simpler if the ancestral host starch or glycogen

synthase would have accepted ADP-glucose as precursor, even with low affinity. In this case, the initial connection would not have required endosymbiotic gene transfer or evolutionary innovations besides re-targeting of an existing host nucleotide sugar transporter to the cyanobaterial plasma membrane.

The proposed scenario does nevertheless have important consequences for endosymbiont metabolism. Transfer of the key enzyme for starch biosynthesis to the host cell nucleus implies losing the cyanobiont's ability to store starch and thus massively impacts the cyanobiont's nocturnal carbon and energy metabolism. However, as outlined above, the evolving plastid acquired already at a very early stage an ATP/ADP exchanger of chlamydial origin. This transporter enabled nocturnal ATP-supply of the cyanobiont, thus attenuating the consequences of losing the starch store. Thus, both the plastidial ATP-importer and the nucleotide-sugar exporter played crucial and concerted roles in turning the cyanobiont into a fully functional organelle of endosymbiotic origin. Loss of the starch pool was likely accompanied or followed shortly after by loss of dinitrogen fixation, which, in the absence of a plastidal starch pool, would have massively impacted, due to its enormous ATP requirements, the host's energy metabolism in a non-sustainable manner. Interestingly, also in the second currently known case of a primary endosymbiotic plastid origin in the thecamoeba *Paulinella chromatophora*, the chromatophores' ability to synthesize starch was apparently lost early on. The chromatophores of *Paulinella* clearly derive from an endosymbiosis of a cyanobacterium in the *Synechococcus* clade. The chromatophore has a genome of approximately 1 megabase in size, which is intermediate between the 2 and 4 megabase genomes of free-living *Synechococcus* and the 0.15 megabase chloroplast genome. These features suggest that the only partial genome reduction has occurred following the initial establishment of the endosymbiont within thecameoba (Nowack et al., 2008). Unfortunately, though, to date nothing is known about the metabolic connection of the *Paulinella* chromatophore with its host cell.

Nucleotide-sugar transport proteins operate in a strict counter-exchange mode and the transport direction depends on the concentration of the substrates on either sides of the membrane. The export of NDP-sugar is compensated by an import of nucleotide-monophosphates, such as AMP or UMP. In the endosymbiosis context, this counter exchange mode of transport would have prevented the depletion of the cyanobacterial nucleotide pool. A relatively higher concentration of ADP-glucose at the site of its biosynthesis inside the endosymbiont would favor its export and immediate consumption in the cytosol. As outlined above, the transporter was likely inefficiently inserted and only carbon committed to storage was drained, thus carbon export likely did not affect the viability of the symbiont (Deschamps et al., 2008). However, a *bona fide* NDP-sugar export system has not yet been identified in the inner envelope membrane of any extant photosynthetic eukaryote. Identification of such a transporter would represent the missing link that would lend strong support to the "metabolic symbiosis and birth of the Plant kingdom" hypothesis, as put forward by Deschamps et al. (2008). According to current knowledge, the pPT gene family with at least three distinct members already evolved in the last common ancestor of the red algae and the green algae and plants, shortly after the initial endosymbiosis was established (Weber et al., 2006). They all are equipped with an N-terminal leader sequence for an efficient targeting to the chloroplast envelope membrane via the Toc and Tic system. Throughout plastid evolution, the major pathway for the export of carbohydrates from the chloroplast was re-directed from sugar-phosphates across towards specialized triose-phosphate (TP)/phosphate, phospho*enol*pyruvate (PEP)/phosphate, and glucose-6-phosphate (Glc6P)/phosphate shuttle systems. These transport systems perfectly suit the metabolic compartmentalization of higher plants: triose-phosphate is exported to the cytosol where it is mainly used for sucrose and/or cellulose biosynthesis. PEP is taken up from the cytosol to drive fatty acid biosynthesis and synthesis of phenolic compounds by the shikimate pathway (e.g., aromatic amino acids) in the plastid stroma. The glucose-6-phosphate transporter (GPT) provides Glc6P for starch synthesis and the oxidative pentose phosphate pathway (OPPP) in heterotrophic tissues.

In summary, the currently most likely scenario in evolutionary terms is that the NDP-sugar transporter represented the starting point for connecting the previously independent carbon metabolisms of host and cyanobiont. Gene duplications and

further radiation of the gene family delivered the genetic material for the evolution of functional diversification, eventually promoting the evolution of a more efficient targeting system to the plasma membrane of the endosymbiont and a fine-tuned mechanism for controlled exchange of metabolic intermediates across the envelope membrane.

The three pPT subfamilies that evolved from a host-derived sugar-nucleotide transporter do not simply encode differentially expressed or targeted gene products with identical functions. They evolved distinct substrate specificities and define the major flux of carbon across higher plant plastid envelope membranes (Flügge, 1999). Phylogenetic analysis indicates that three orthologous groups of sugar-phosphate transporters with the preferred substrates TP, PEP, and Glc6P, respectively, are also encoded by the genomes of the red algae *G. sulphuraria* and *C. merolae*. This is surprising, given the enormous differences between starch and soluble sugar metabolism in the green and the red lineages. In contrast to the ADP-glucose dependent enzyme of the Viridiplantae, the Rhodophyta harbor an UDP-glucose specific starch synthase of ancient eukaryotic origin, which is localized in the cytosol. The soluble starch synthase from green algae and plants is closely related to cyanobacterial genes and incorporates ADP-glucose into the polysaccharide chain in the stroma of the organelle (Deschamps et al., 2006). Probably, both UDP- and ADP-glucose dependent starch synthesis was still manifested in the genome of the proto-alga. By the time the *Rhodophyceae* and *Chloroplastida* diverged from the protoalga as distinct lineages, the carbon metabolism was likely far from being established as we find it in extant species. The starch biosynthesis pathway has been relocated to the plastid stroma in green plants and algae (Deschamps et al., 2008) or stayed and gradually evolved in the cytosol of Rhodophyceae and Glaucophytes. In addition, the predominant soluble carbohydrate in the red algae, floridoside, is composed of a glycerol and galactose moiety (Viola et al., 2001) whereas plants synthesize the disaccharide sucrose from UDP-glucose and fructose-6-phosphate. Based on phylogenomic analysis, it is tempting to hypothesize that the ability to transport glycolytic intermediates, i.e. PEP, triose-phosphate, and Glc6P, was a successful invention during plastid evolution that is conserved in all members of the Archaeplastida. If the hypothesis is correct, the homologues in *G. sulphuraria* should represent truly orthologous genes with identical substrate specificities as their counterparts in the green lineage. However, the major differences in carbon metabolism between photosynthetic eukaryotes challenge the hypothesis. Possibly, the transporters continue to evolve divergently in red algae and green plants.

These opposing hypotheses were recently tested by recombinant expression and biochemical characterization of the pPT genes from *G. sulphuraria* and comparison to their counterparts from land plants (Linka et al., 2008). It was found that the red alga *G. sulphuraria*, similar to the *Plantae*, possesses an export system for triose-phosphates, albeit it displays higher specificity and affinity. Also similar to green plants, the red alga harbors a PEP/P_i exchanger that has identical kinetic constants as the plant ortholog (Linka et al., 2008). In contrast to plants, though, no hexose or pentose phosphate transport activity could be detected. Hence, in red algae, two of the three main pPT functions (i.e., TP and PEP transport) are conserved, whereas the third (i.e., hexose and pentose phosphate transport) is missing. Thus, in the red algae, TPT controls the flux of reduced carbon across the plastid envelope during autotrophic and heterotrophic growth conditions. In contrast to plants, however, Glc6P transport activity does not exist in the red alga as a major import route for sugar-phosphates (Linka et al., 2008).

The different substrate affinities of green plant and red algal TPTs make sense in the context of their respective carbon metabolisms. In the green lineage, recent photosynthate must be partitioned between plastidial starch biosynthesis and cytosolic sucrose biosynthesis. In red algae, however, the vast majority of recent photosynthate must be exported to drive both starch and soluble carbohydrate biosynthesis in the cytosol. The low affinity TPT in green plants likely represents an adaptation that is required for maintaining a relatively high concentration of TP in the plastid stroma to permit plastidial starch biosynthesis. In contrast, the high affinity TPT of the red algae has evolved to permit the efficient export of TP to the cytosol (Linka et al., 2008).

c. The Role of the Endomembrane System for Metabolic Integration

The endomembrane system (ER, Golgi, nuclear envelope) likely evolved very early in the evolution of eukaryotic cells and almost certainly preceded the origin of mitochondria (Cavalier-Smith, 2003; Jekely, 2003, 2005; Reuzeau et al., 1997). The endomembrane system is an important component of endocytotic and/or phagocytosis processes and was thus critical for endosymbiosis (Cavalier-Smith, 2002). The endomembrane system can be considered as the 'mother of all intracellular compartments' (Jekely, 2005) and likely provided the initial mechanisms for protein and solute transport.

According to the currently dominant model of plastid origin through endosymbiosis, a complete cyanobacterium was engulfed by a phagotrophic eukaryote and escaped digestion in the food vacuole (Reyes-Prieto et al., 2007). Eventually, the membrane surrounding the food vacuole was lost or merged with the cyanobacterial outer membrane, resulting in two membranes marking the barrier to the new environment: the cytosol of the host (Cavalier-Smith, 2000). The composition of the outer membrane changed considerably over evolutionary time scales (Block et al., 2007). Eubacterial lipoproteins and lipopolysaccharides are absent from higher plant chloroplasts and the phospholipid phosphatidylcholine was specifically introduced into the outer envelope membrane from the endomembrane system of the host (Douce and Joyard, 1990). Extant Archaeplastida display an intimate connection between the endoplasmic reticulum (ER) and the envelopes of the plastid. Lipids are synthesized in the plastid stroma and a sophisticated trafficking system has been established between plastid and ER in both directions for the biogenesis of plastidial and extra-plastidial membranes (Benning, 2008). As a conserved feature of extant primary plastids, the Tic/Toc protein import apparatus recognizes most nuclear-encoded genes via a N-terminal leader sequence of 25–125 amino acids (Soll and Schleiff, 2004). Still, some plastid-targeted proteins do not exhibit an obvious leader sequence and an alternative non-canonical protein translocation route via the ER and Golgi vesicle sorting system has been recently discovered (Villarejo et al., 2005). The discovery of vesicle-mediated protein transport to plastids in extant plants lends support to the hypothesis that the Tic/Toc protein import apparatus was predated by a rudimentary and likely not very efficient vesicle-mediated targeting system. The vesicle-mediated system was eventually replaced by the Tic/Toc apparatus, with the exception of glycosylated proteins that required passage through the endomembrane system.

The crucial role of the endomembrane system for forging endosymbiosis is further emphasized and supported through plastids of secondary and tertiary endosymbiotic origin. Such plastids are surrounded by three to four envelope membranes (Bhattacharya et al., 2004) and all nuclear-encoded plastid-destined proteins are delivered to their final destination by a multi-partite leader sequence through the concerted action of the ER sorting system and the Tic/Toc related import machinery of the organelle (Kilian and Kroth, 2005; Lang et al., 1998).

These direct contacts between ER and plastid support the idea that vesicle transport to and from the plastid was initiated at an early stage of endosymbiosis. ER or Golgi-derived vesicles initially might have fused randomly with the outer membrane of the cyanobacterium, releasing their contents into the periplasmic space and the proteins were inserted into the cyanobacterial plasma membrane or taken up by the cyanobiont. Thus, even if insertion of host metabolite transporters into the cyanobiont's plasma membrane was initially not very inefficient, it would have given the host the opportunity to connect the cyanobiont with its cytosol and tap into photosynthates from the cyanobacterium. This could have become a selective advantage for the phagotrophic eukaryote and favored the establishment of a permanent partnership.

Given the origin of one the major plastidial carbon transporter families from nucleotide sugar transporters of the endomembrane system and the close interaction of between plastids and the endomembrane system, as outlined above, it is tempting to hypothesize that the endomembrane system served as a rich source of metabolite transporters for integrating the evolving plastid organelle into the metabolic network of the host cell. This hypothesis was recently tested by phylogenomic analysis of the permeome of extant plastids

(Tyra et al., 2007), making use of the well-annotated transporter proteins from *A. thaliana*. *In silico* searches for genes in the *A. thaliana* genome that encode a predicted plastid target sequence and hydrophobic domains that characteristic for membrane transporters provided an initial catalogue of candidate genes (Weber et al., 2005). Several plastid-specific proteomic studies confirmed the presence and extended the list of proteins to nearly ~130 putative metabolite transporter proteins, which are targeted to inner envelope membranes of plastids (Sun et al., 2009). This number does not include genes from the Toc, Tic and thylakoid-related protein import machinery and prominent thylakoid-localized electron-transporting proteins of the photosynthetic core complex, such as D1, D2, cytochrome b_6, cytochrome *f*, or H^+-coupled ATP-synthase subunits. Phylogenomic analysis of the extant plastid permeome revealed that the host has been the major source for metabolite transporter evolution (Tyra et al., 2007). Over 50% of the metabolite transporters residing in the membrane of *A. thaliana* plastids are of host origin whereas the cyanobacterial ancestor of the plastids contributed only a minor share. Many of these host-derived transporters have their authentic localization either in the plasma membrane or internal membranes, such as the tonoplast. In most documented cases, these transporters reach their final destination via vesicle-mediated transport that involves the endomembrane system. It thus seems likely that metabolite transporters were initially delivered to the plastid envelope membrane via mis-targeting of host internal and plasma membrane transporters. Eventually, after gene duplication and acquisition of specific targeting signals, a subset of host-derived metabolite transporters evolved into what is now established as the plastid permeome. Phylogenetic analysis also showed that most of the plastidial metabolite transport proteins were distributed in both green and red algae and to a much lesser extent present in one, but not the other phyletic group (Tyra et al., 2007). This suggests that a massive invasion of the envelope membrane with host-derived transporter proteins took place at the stage of the proto-alga, before the split into the Archaeplastida lineages. Some envelope membrane transporters, however, likely evolved in the green lineage after the split of the Archaeplastida. For example, the plastidic glucose and maltose transporters can only be detected in members of the Chloroplastida (Tyra et al., 2007). These transporters are required for the export of the products of starch breakdown (i.e., maltose and glucose) from chloroplasts. Since with the exception of the Viridiplantae all other photosynthetic eukaryotes store starch in the cytosol, these transporters likely represent a specific adaptation to starch metabolism in the green lineage.

d. Phosphate Translocators and Secondary/Tertiary Endosymbiosis

Apparently phosphate translocators are particularly suited for carbon export from chloroplasts because the corresponding genes have been transferred by lateral gene transfer to the nuclei of organisms harboring secondary or tertiary plastids (Weber et al., 2006), including the apicoplast-containing malaria parasite *Plasmodium falciparum* (Mullin et al., 2006; Lim et al., 2010). They can also be detected in photosynthetic protists of the Chromalveolata (Weber et al., 2006). However, of the three members of the PT-family in the red algae, only the gene encoding the TPT was transmitted to the nuclear genome of the host during secondary endosymbiosis. Once arrived in the host nucleus, the originally transferred gene underwent duplication, thus initiating the foundation of a novel gene family, likely accompanied by functional diversification and evolution of novel substrate specificities (Weber et al., 2006).

e. Mitochondria as Source of Plastidial Transporters

A range of plastididal transporters, such as the S-adenosylmethionine transporter, a folate transporter, and adenylate transporters have been recruited to the plastid envelope membrane from the mitochondrial carrier family (MCF).

S-adenosylmethionine (SAM) is required as methyl-group donor in plastids for a multitude of methylation reactions. However, the biosynthesis of SAM is restricted to the cytosol of plant cells (Ravanel et al., 2004), which means that SAM must be imported from the cytosol and reaction product S-adenosylhomocysteine (SAHC) needs to be exported for cytosolic regeneration of SAM. In Arabidopsis the plastidic SAM/SAHC antiporter (SAMT1) is encoded by gene *At4g39460* (Bouvier et al., 2006). SAMT1 belongs to the MCF but is

nevertheless targeted to chloroplasts by an N-terminal transit peptide (Bouvier et al., 2006). Phylogenomic analysis showed that SAMT1-related proteins can be detected in the genomes of both Chloroplastida and Rhodophyceae, indicating the recruitment of mitochondrial carrier family members to the chloroplast envelope membrane was an ancient event that likely took place before the split of the red and green lineages (Tyra et al., 2007).

Similar to SAM, tetrahydrofolate (THF) and its derivatives are essential cofactors for one-carbon transfer reactions in almost all organisms. Plants and most microorganisms synthesize folates *de novo* and in plants folate biosynthesis shows a complex compartmentation, involving the cytosol, plastids, vacuole, and mitochondria (Rebeille et al., 2006). Recently, the first plastidic folate carriers from plants have been described. In Arabidopsis, one class of plastidic folate transporters is encoded by the *At5g66380* (AtFOLT1) (Bedhomme et al., 2005) and the gene product belongs to the mitochondrial carrier family (MCF) (Picault et al., 2004). The most closely related protein in humans localizes to the mitochondria (Titus and Moran, 2000). Nevertheless, as shown by GFP-fusion analysis, AtFOLT1 is targeted to chloroplasts *in vivo*. AtFOLT1 thus represents a plastidal folate transporter that was recruited to the chloroplast envelope membrane from a pre-existing mitochondrial protein of the host cell. Both plants and cyanobacteria harbor a second class of folate transporters that are related to the *Leishmania* folate transporters (Klaus et al., 2005). While known cyanobacterial genomes encode only a single *Leishmania*-type folate transporter, the Arabidopsis genome encodes a small gene family with nine members. One of the gene family members is targeted to chloroplasts (*At2g32040*). Thus, this class of plant folate transporter represents one of a few examples of an inner chloroplast envelope membrane metabolite transporter that has evolved from a cyanobacterial ancestor.

In addition to a plastidal ATP/ADP transporter of chlamydial origin (see above), dicotyledonous plants possess a second class of nucleotide transporter that differs from the NTTs with respect to the mode of transport mechanism and its evolutionary origin (Leroch et al., 2005). This protein is closely related to the Brittle1 protein (BT1) from cereal endosperm plastids, which belong to the MCF. BT1 has a unique function in cereal endosperm, where it imports ADP-Glc synthesized in the cytosol into plastids where it serves as precursor for starch biosynthesis (Patron et al., 2004; Shannon et al., 1998). In dicotyledonous plants, where ADG-Glc biosynthesis is exclusively located in the chloroplast stroma (Beckles et al., 2001; Neuhaus et al., 2005), the BT1-related protein does not transport ADP-Glc but serves as a uniporter providing the cytosol and other compartments with adenine nucleotides (AMP, ADP, ATP) synthesized in plastids (Leroch et al., 2005).

In summary, at least three different transporters with distinct substrate specificity that belong to the mitochondrial carrier family actually reside in the chloroplast envelope membrane in plants. Likely, these transporters have been recruited from pre-existing MCF members of the host after the primary endosymbiosis.

f. Recruitment of Plastidial Transporters by Lateral Gene Transfer

While some plastidial transporters and transporter families have evolved from transporters that were already present in the primitive mitochondriate eukaryote that served as the host of the cyanobiont, others have been recruited to the plastid envelope by lateral gene transfer. For example, as outlined above, the dicarboxylate transporter family and the non-mitochondrial ATP transporter family have been recruited to photosynthetic eukaryotes by lateral transfer from chlamydiae. Given that a total of 55 genes in the Plantae has chlamydial origin (Moustafa et al., 2008) it is surprising that an unexpectedly high proportion of these chlamydiae-derived genes encode for plastid metabolite transporters. A similar result was demonstrated for the red alga *Cyanidioschyzon merolae* (Huang and Gogarten, 2007). Horizontal gene transfer from a parasitic *Chlamydia* bacterium thus provides a minor but apparently crucial third genome source for plastid evolution.

IV. Summary and Perspective

While many factors contributed to plastid evolution, integration of the plastid into the metabolic network of the host cell was certainly a crucial step in the establishment of a permanent endosymbiotic

association between cyanobiont and host cell. Multiple sources contributed to establishing the plastid permeome. However, some of the early and most likely crucial elements, the plastidic phosphate translocators and the ATP transporters, were likely already present in the host cell before the invasion by the cyanobiont. The phosphate translocators were essential for the host to tap into the photosynthetic carbon pool of the cyanobiont. The ATP transporters, in turn, were crucial for nocturnal energy supply of the cyanobiont in the absence of a plastidal starch pool. In this respect, the association between a eukaryotic mitochondriate and a cyanobacterium might be considered a mutually symbiotic association, with the cyanobiont providing reduced carbon to the host and the host providing energy to the cyanobiont. Together, host and cyanobiont have been immensely successful, having conquered almost all aquatic and land habitats. Nevertheless, many important questions remain unsolved. For example, the actual source of the cyanobiont is still unclear and controversially discussed (Deschamps et al., 2008; Deusch et al., 2008). The steady increase in completed genome sequences will contribute to addressing these and other questions related to plastid origin and evolution. Of particular interest within the next few years will be the upcoming nuclear genome sequences of the first member of the Glaucophyta, *C. paradoxa*, and of the host of the tecamoeba *P. chromatophora*, the second documented case of primary endosymbiotic plastid origin (Marin et al., 2005).

References

Adl SM, Simpson AG, Farmer MA, Andersen RA, Anderson OR, Barta JR, Bowser SS, Brugerolle G, Fensome RA, Fredericq S, James TY, Karpov S, Kugrens P, Krug J, Lane CE, Lewis LA, Lodge J, Lynn DH, Mann DG, McCourt RM, Mendoza L, Moestrup O, Mozley-Standridge SE, Nerad TA, Shearer CA, Smirnov AV, Spiegel FW and Taylor MF (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J Eukaryot Microbiol 52: 399–451

Alberty RA (2005) Thermodynamics of the mechanism of the nitrogenase reaction. Biophys Chem 114: 115–120

Andersson JO, Sjogren AM, Davis LA, Embley TM and Roger AJ (2003) Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes. Curr Biol 13: 94–104

Andre C and Benning C (2007) Arabidopsis seedlings deficient in a plastidic pyruvate kinase are unable to utilize seed storage compounds for germination and establishment. Plant Physiol 145:1670–1680

Andre C, Froehlich JE, Moll MR and Benning C (2007) A heteromeric plastidic pyruvate kinase complex involved in seed oil biosynthesis in Arabidopsis. Plant Cell 19: 2006–2222

Aoki N, Ohnishi J and Kanai R (1992) 2 different mechanisms for transport of pyruvate into mesophyll chloroplasts of C4 plants-a comparative study. Plant Cell Physiol 33: 805–809

Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M, Brzezinski MA, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Hellsten U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kroger N, Lau WW, Lane TW, Larimer FW, Lippmeier JC, Lucas S, Medina M, Montsant A, Obornik M, Parker MS, Palenik B, Pazour GJ, Richardson PM, Rynearson TA, Saito MA, Schwartz DC, Thamatrakoln K, Valentin K, Vardi A, Wilkerson FP and Rokhsar DS (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science 306: 79–86

Ashikov A, Routier F, Fuhlrott J, Helmus Y, Wild M, Gerardy-Schahn R and Bakker H (2005) The human solute carrier gene SLC35B4 encodes a bifunctional nucleotide sugar transporter with specificity for UDP-xylose and UDP-N-acetylglucosamine. J Biol Chem 280: 27230–27235

Awai K, Xu C, Lu B and Benning C (2006a) Lipid trafficking between the endoplasmic reticulum and the chloroplast. Biochem Soc Trans 34: 395–398

Awai K, Xu C, Tamot B and Benning C (2006b) A phosphatidic acid-binding protein of the chloroplast inner envelope membrane involved in lipid trafficking. Proc Natl Acad Sci USA 103: 10817–10822

Bagge P and Larsson C (1986) Biosynthesis of aromatic amino acids by highly purified spinach chloroplasts - Compartmentation and regulation of the reactions. Physiol Plant 68: 641–647

Bakker H, Routier F, Oelmann S, Jordi W, Lommen A, Gerardy-Schahn R and Bosch D (2005) Molecular cloning of two Arabidopsis UDP-galactose transporters by complementation of a deficient Chinese hamster ovary cell line. Glycobiology 15: 193–201

Barbier G, Oesterhelt C, Larson MD, Halgren RG, Wilkerson C, Garavito RM, Benning C and Weber APM (2005a) Genome Analysis. Comparative genomics of two closely related unicellular thermo-acidophilic red algae, *Galdieria sulphuraria* and *Cyanidioschyzon merolae*, reveals the molecular basis of the metabolic flexibility of *Galdieria* and significant differences in carbohydrate metabolism of both algae. Plant Physiol 137: 460–474

Barbier GG, Zimmermann M and Weber APM (2005b) Genomics of the thermo-acidophilic red alga *Galdieria sulphuraria*. In: Hoover RB, Levin GV, Rozanov AY and Gladstone GR (eds) Astrobiology and Planetary Missions, pp 590–609. SPIE, San Diego

Beckles DM, Smith AM and ap Rees T (2001) A cytosolic ADP-glucose pyrophosphorylase is a feature of graminaceous endosperms, but not of other starch-storing organs. Plant Physiol 125: 818–827

Bedhomme M, Hoffmann M, McCarthy EA, Gambonnet B, Moran RG, Rebeille F and Ravanel S (2005) Folate metabolism in plants: an Arabidopsis homolog of the mammalian mitochondrial folate transporter mediates folate import into chloroplasts. J Biol Chem 280: 34823–34831

Benning C (2008) A role for lipid trafficking in chloroplast biogenesis. Prog Lipid Res 47: 381–389

Benning C, Xu C and Awai K (2006) Non-vesicular and vesicular lipid trafficking involving plastids. Curr Opin Plant Biol 9: 241–247

Bergman B, Gallon JR, Rai AN and Stal LJ (1997) N_2 fixation by non-heterocystous cyanobacteria. FEMS Microbiol Rev 19: 139–185

Bhattacharya D, Yoon HS and Hackett JD (2004) Photosynthetic eukaryotes unite: endosymbiosis connects the dots. Bioessays 26: 50–60

Blankenship RE (1992) Origin and early evolution of photosynthesis. Photosynth Res 33: 91–111

Block MA, Douce R, Joyard J and Rolland N (2007) Chloroplast envelope membranes: a dynamic interface between plastids and the cytosol. Photosynth Res 92: 225–244

Borchert S, Harborth J, Schünemann D, Hoferichter P and Heldt HW (1993) Studies of the enzymic capacities and transport properties of pea root plastids. Plant Physiol 101: 303–312

Bothe H, Winkelmann S and Boison G (2008) Maximizing hydrogen production by cyanobacteria. Z Naturforsch [C] 63: 226–232

Bouvier F, Linka N, Isner JC, Mutterer J, Weber APM and Camara B (2006) Arabidopsis SAMT1 defines a plastid transporter regulating plastid biogenesis and plant development. Plant Cell 18: 3088–3105

Brugière N, Dubois F, Masclaux C, Sangwan RS and Hirel B (2000) Immunolocalization of glutamine synthetase in senescing tobacco (*Nicotiana tabacum* L.) leaves suggests that ammonia assimilation is progressively shifted to the mesophyll cytosol. Planta 211: 519–527

Cavalier-Smith T (2000) Membrane heredity and early chloroplast evolution. Trends Plant Sci 5: 174–182

Cavalier-Smith T (2002) The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. Int J Syst Evol Microbiol 52: 297–354

Cavalier-Smith T (2003) Genomic reduction and evolution of novel genetic membranes and protein-targeting machinery in eukaryote-eukaryote chimaeras (meta-algae). Philos Trans R Soc Lond B Biol Sci 358: 109–134

Chen HC and Melis A (2004) Localization and function of SulP, a nuclear-encoded chloroplast sulfate permease in *Chlamydomonas reinhardtii*. Planta 220: 198–210

Chen HC, Yokthongwattana K, Newton AJ and Melis A (2003) SulP, a nuclear gene encoding a putative chloroplast-targeted sulfate permease in *Chlamydomonas reinhardtii*. Planta 218: 98–106

Coruzzi GM (2003) Primary N-assimilation into amino acids in Arabidopsis. In: Somerville CR and Meyerowitz EM (eds) The Arabidopsis Book, pp 1–17. American Society of Plant Biologists, Rockville

Curien G, Dumas R, Ravanel S and Douce R (1996) Characterization of an *Arabidopsis thaliana* cDNA encoding an S-adenosylmethionine-sensitive threonine synthase. Threonine synthase from higher plants. FEBS Lett 390: 85–90

Deschamps P, Haferkamp I, Dauvillee D, Haebel S, Steup M, Buleon A, Putaux JL, Colleoni C, d'Hulst C, Plancke C, Gould S, Maier U, Neuhaus HE and Ball S (2006) Nature of the periplastidial pathway of starch synthesis in the cryptophyte *Guillardia theta*. Eukaryot Cell 5: 954–963

Deschamps P, Colleoni C, Nakamura Y, Suzuki E, Putaux JL, Buleon A, Haebel S, Ritte G, Steup M, Falcon LI, Moreira D, Loffelhardt W, Raj JN, Plancke C, d'Hulst C, Dauvillee D and Ball S (2008) Metabolic symbiosis and the birth of the plant kingdom. Mol Biol Evol 25: 536–548

Deusch O, Landan G, Roettger M, Gruenheit N, Kowallik KV, Allen JF, Martin W and Dagan T (2008) Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. Mol Biol Evol 25: 748–761

Douce R and Joyard J (1990) Biochemistry and function of the plastid envelope. Annu Rev Cell Biol 6: 173–216

Eicks M, Maurino V, Knappe S, Flügge UI and Fischer K (2002) The plastidic pentose phosphate translocator represents a link between the cytosolic and the plastidic pentose phosphate pathways in plants. Plant Physiol 128: 512–522

Eisenhut M, Ruth W, Haimovich M, Bauwe H, Kaplan A and Hagemann M (2008) The photorespiratory glycolate metabolism is essential for cyanobacteria and might have been conveyed endosymbiontically to plants. Proc Natl Acad Sci USA 105: 17199–17204

Falkowski PG and Godfrey LV (2008) Electrons, life and the evolution of Earth's oxygen cycle. Philos Trans R Soc Lond B Biol Sci 363: 2705–2716

Falkowski PG and Isozaki Y (2008) Geology. The story of O_2. Science 322: 540–542

Fischer K and Weber A (2002) Transport of carbon in non-green plastids. Trends Plant Sci 7: 345

Fischer K, Kammerer N, Gutensohn M, Arbinger B, Weber A, Häusler RE and Flügge UI (1997) A new class of plastidic phosphate translocators: A putative link between primary and secondary metabolism by the phosphoenolpyruvate/phosphate antiporter. Plant Cell 9: 453–462

Flügge UI (1992) Reaction mechanism and asymmetric orientation of the reconstituted chloroplast phosphate translocator. Biochim Biophys Acta 1110: 112–118

Flügge UI (1998) Metabolite transporters in plastids. Curr Opin Plant Biol 1: 201–206.

Flügge UI (1999) Phosphate translocators in plastids. Annu Rev Plant Physiol Plant Mol Biol 50: 27–45

Flügge UI (2000) Transport in and out of plastids: does the outer envelope membrane control the flow? Trends Plant Sci 5: 135–137

Flügge UI and Benz R (1984) Pore-forming activity in the outer membrane of the chloroplast envelope. FEBS Lett 169: 85–89

Flügge UI and Gao W (2005) Transport of isoprenoid intermediates across chloroplast envelope membranes. Plant Biol (Stuttg) 7: 91–97

Flügge UI and Heldt HW (1984) The phosphate-triose phosphate-phosphoglycerate translocator of the chloroplast. Trends Biochem Sci 9: 530–533

Flügge UI and Heldt HW (1991) Metabolite translocators of the chloroplast envelope. Annu Rev Plant Physiol Plant Mol Biol 42: 129–144

Flügge UI, Häusler RE, Ludewig F and Fischer K (2003) Functional genomics of phosphate antiport systems of plastids. Physiol Plantarum 118: 475–482

Forsum O, Svennerstam H, Ganeteg U and Nasholm T (2008) Capacities and constraints of amino acid utilization in Arabidopsis. New Phytol 179: 1058–1069

Furuya A and Hayashi JA (1963) Glycolic acid oxidation by *Escherichia coli* adapted to glycolate. J Bacteriol 85: 1124–1131

Glöckner G, Rosenthal A and Valentin K (2000) The structure and gene repertoire of an ancient red algal plastid genome. J Mol Evol 51: 382–390

Goldblatt C, Lenton TM and Watson AJ (2006) Bistability of atmospheric oxygen and the Great Oxidation. Nature 443: 683–686

Hansen RW and Hayashi JA (1962) Glycolate metabolism in Escherichia coli. J Bacteriol 83: 679–687

Heldt HW, Flügge UI, Borchert S, Brueckner G and Ohnishi J (1990) Phosphate translocators in plastids. Plant Biol 10: 39–54

Hemmerlin A, Hoeffler JF, Meyer O, Tritsch D, Kagan IA, Grosdemange-Billiard C, Rohmer M and Bach TJ (2003) Cross-talk between the cytosolic mevalonate and the plastidial methylerythritol phosphate pathways in tobacco bright yellow-2 cells. J Biol Chem 278: 26666–26676

Heyer H and Krumbein WE (1991) Excretion of fermentation products in dark and anaerobically incubated cyanobacteria. Arch Microbiol 155: 284–287

Heyer H, Stal L and Krumbein WE (1989) Simultaneous heterolactic and acetate fermentation in the marine cyanobacterium *Oscillatoria limosa* incubated anaerobically in the dark. Arch Microbiol 151: 558–564

Huang J and Gogarten JP (2007) Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? Genome Biol 8: R99

Huber SC and Edwards GE (1977) Transport in C_4 mesophyll chloroplasts: Characterization of the pyruvate carrier. Biochim Biophys Acta 462: 583–602

Ireland RJ and Lea PJ (1999) The enzymes of glutamine, glutamate, asparagine, and aspartate metabolism. In: Singh BK (ed) Plant amino acids: Biochemistry and biotechnology, pp 49–110. Dekker, New York

Jekely G (2003) Small GTPases and the evolution of the eukaryotic cell. Bioessays 25: 1129–1138

Jekely G (2005) Glimpsing over the event horizon: evolution of nuclear pores and envelope. Cell Cycle 4: 297–299

Kammerer B, Fischer K, Hilpert B, Schubert S, Gutensohn M, Weber A and Flügge UI (1998) Molecular characterization of a carbon transporter in plastids from heterotrophic tissues: the glucose-6-phosphate/phosphate antiporter. Plant Cell 10: 105–117

Kampfenkel KH, Möhlmann T, Batz O, van Montague M, Inzé D and Neuhaus HE (1995) Molecular characterization of an *Arabidopsis thaliana* cDNA encoding a novel putative adenylate translocator of higher plants. FEBS Lett 374: 351–355

Kataoka T, Watanabe-Takahashi A, Hayashi N, Ohnishi M, Mimura T, Buchner P, Hawkesford MJ, Yamaya T and Takahashi H (2004) Vacuolar sulfate transporters are essential determinants controlling internal distribution of sulfate in Arabidopsis. Plant Cell 16: 2693–2704

Kelly AA and Dormann P (2004) Green light for galactolipid trafficking. Curr Opin Plant Biol 7: 262–269

Khakhlova O and Bock R (2006) Elimination of deleterious mutations in plastid genomes by gene conversion. Plant J 46: 85–94

Kichey T, Le Gouis J, Sangwan B, Hirel B and Dubois F (2005) Changes in the cellular and subcellular localization of glutamine synthetase and glutamate dehydrogenase during flag leaf senescence in wheat (*Triticum aestivum* L.). Plant Cell Physiol 46: 964–974

Kilian O and Kroth PG (2005) Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. Plant J 41: 175–183

Klaus SM, Kunji ER, Bozzo GG, Noiriel A, de la Garza RD, Basset GJ, Ravanel S, Rebeille F, Gregory JF, 3rd and Hanson AD (2005) Higher plant plastids and cyanobacteria have folate carriers related to those of trypanosomatids. J Biol Chem 280: 38457–38463

Kobayashi Y, Kanesaki Y, Tanaka A, Kuroiwa H, Kuroiwa T and Tanaka K (2009) Tetrapyrrole signal as a cell-cycle coordinator from organelle to nuclear DNA replication in plant cells. Proc Natl Acad Sci USA 106: 803–807

Komarova NY, Thor K, Gubler A, Meier S, Dietrich D, Weichert A, Suter Grotemeyer M, Tegeder M and Rentsch

D (2008) AtPTR1 and AtPTR5 transport dipeptides in planta. Plant Physiol 148: 856–869

Lang M, Apt KE and Kroth PG (1998) Protein transport into "complex" diatom plastids utilizes two different targeting signals. J Biol Chem 273: 30973–30978

Lanier W, Moustafa A, Bhattacharya D and Comeron JM (2008) EST analysis of *Ostreococcus lucimarinus*, the most compact eukaryotic genome, shows an excess of introns in highly expressed genes. PLoS ONE 3: e2171

Lau WW and Armbrust EV (2006) Detection of glycolate oxidase gene glcD diversity among cultured and environmental marine bacteria. Environ Microbiol 8: 1688–6702

Lea PJ, Wallsgrove RM and Miflin BJ (1978) Photorespiratory nitrogen cycle. Nature 275: 741–743

Lee M and Leustek T (1999) Identification of the gene encoding homoserine kinase from Arabidopsis thaliana and characterization of the recombinant enzyme derived from the gene. Arch Biochem Biophys 372: 135–142

Lee YH, Foster J, Chen J, Voll LM, Weber APM and Tegeder M (2007) AAP1 transports uncharged amino acids into roots of Arabidopsis. Plant J 144: 32–42

Leroch M, Kirchberger S, Haferkamp I, Wahl M, Neuhaus HE and Tjaden J (2005) Identification and characterization of a novel plastidic adenine nucleotide uniporter from *Solanum tuberosum*. J Biol Chem 280: 17992–18000

Leustek T, Martin MN, Bick JA and Davies JP (2000) Pathways and regulation of sulfur metabolism revealed through molecular and genetic studies. Annu Rev Plant Physiol Plant Mol Biol 51: 141–165

Lichtenthaler HK, Schwender J, Disch A and Rohmer M (1997) Biosynthesis of isoprenoids in higher plant chloroplasts proceeds via a mevalonate-independent pathway. FEBS Lett 400: 271–274

Lim L, Linka M, Weber APM and McFadden GI (2010) The carbon and energy sources of the non-photosynthetic plastid in the malaria parasite. FEBS Lett 584: 549–554

Linka M and Weber APM (2005) Shuffling ammonia between mitochondria and plastids during photorespiration. Trends Plant Sci 10: 461–465

Linka N and Weber APM (2010) Intracellular metabolite transporters in plants. Mol Plant 3: 21–53

Linka N, Hurka H, Lang BF, Burger G, Winkler HH, Stamme C, Urbany C, Seil I, Kusch J and Neuhaus HE (2003) Phylogenetic relationships of non-mitochondrial nucleotide transport proteins in bacteria and eukaryotes. Gene 306: 27–35

Linka M, Jamai A and Weber AP (2008) Functional characterization of the plastidic phosphate translocator gene family from the thermo-acidophilic red alga *Galdieria sulphuraria* reveals specific adaptations of primary carbon partitioning in green plants and red algae. Plant Physiol 148: 1487–1496

Luhn K, Laskowska A, Pielage J, Klambt C, Ipe U, Vestweber D and Wild MK (2004) Identification and molecular cloning of a functional GDP-fucose transporter in Drosophila melanogaster. Exp Cell Res 301: 242–250

Margulis L (1971a) The origin of plant and animal cells. Am Sci 59: 230–235

Margulis L (1971b) Symbiosis and evolution. Sci Am 225: 48–57

Marin B, Nowack EC and Melkonian M (2005) A plastid in the making: evidence for a second primary endosymbiosis. Protist 156: 425–432

Martin W and Muller M (1998) The hydrogen hypothesis for the first eukaryote. Nature 392: 37–41

Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M and Penny D (2002) Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc Natl Acad Sci USA 99: 12246–12251

Martin MN, Tarczynski MC, Shen B and Leustek T (2005) The role of 5′-adenylylsulfate reductase in controlling sulfate reduction in plants. Photosynth Res 86: 309–323

Matsuzaki M, Misumi O, Shin IT, Maruyama S, Takahara M, Miyagishima SY, Mori T, Nishida K, Yagisawa F, Yoshida Y, Nishimura Y, Nakao S, Kobayashi T, Momoyama Y, Higashiyama T, Minoda A, Sano M, Nomoto H, Oishi K, Hayashi H, Ohta F, Nishizaka S, Haga S, Miura S, Morishita T, Kabeya Y, Terasawa K, Suzuki Y, Ishii Y, Asakawa S, Takano H, Ohta N, Kuroiwa H, Tanaka K, Shimizu N, Sugano S, Sato N, Nozaki H, Ogasawara N, Kohara Y and Kuroiwa T (2004) Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. Nature 428: 653–657

Miflin BJ and Lea PJ (1976) The pathway of nitrogen assimilation in plants. Phytochemistry 15: 873–885

Miflin BJ and Lea PJ (1980) Ammonia assimilation. In: Miflin BJ (ed) The Biochemistry of Plants, pp 169–202. Academic Press, New York

Moustafa A, Reyes-Prieto A and Bhattacharya D (2008) Chlamydiae has contributed at least 55 genes to Plantae with predominantly plastid functions. PLoS ONE 3: e2205

Muller HJ (1964) The relation of recombination to mutational advance. Mutat Res 106: 2–9

Mullin KA, Lim L, Ralph SA, Spurck TP, Handman E and McFadden GI (2006) Membrane transporters in the relict plastid of malaria parasites. Proc Natl Acad Sci USA 103: 9572–9577

Neuhaus HE and Wagner R (2000) Solute pores, ion channels, and metabolite transporters in the outer and inner envelope membranes of higher plant plastids. Biochim Biophys Acta 1465: 307–323

Neuhaus HE, Hausler RE and Sonnewald U (2005) No need to shift the paradigm on the metabolic pathway to transitory starch in leaves. Trends Plant Sci 10: 154–156

Niewiadomski P, Knappe S, Geimer S, Fischer K, Schulz B, Unte US, Rosso MG, Ache P, Flügge UI and Schneider A (2005) The Arabidopsis plastidic glucose 6-phosphate/

phosphate translocator GPT1 is essential for pollen maturation and embryo sac development. Plant Cell 17: 760–775

Niittyla T, Messerli G, Trevisan M, Chen J, Smith AM and Zeeman SC (2004) A previously unknown maltose transporter essential for starch degradation in leaves. Science 303: 87–89

Nowack EC, Melkonian M and Glockner G (2008) Chromatophore genome sequence of Paulinella sheds light on acquisition of photosynthesis by eukaryotes. Curr Biol 18: 410–418

Ohta N, Matsuzaki M, Misumi O, Miyagishima SY, Nozaki H, Tanaka K, Shin IT, Kohara Y and Kuroiwa T (2003) Complete sequence and analysis of the plastid genome of the unicellular red alga *Cyanidioschyzon merolae*. DNA Res 10: 67–77

Paczek V, Dubois F, Sangwan R, Morot-Gaudry JF, Roubelakis-Angelakis KA and Hirel B (2002) Cellular and subcellular localisation of glutamine synthetase and glutamate dehydrogenase in grapes gives new insights on the regulation of carbon and nitrogen metabolism. Planta 216: 245–254

Palenik B, Grimwood J, Aerts A, Rouze P, Salamov A, Putnam N, Dupont C, Jorgensen R, Derelle E, Rombauts S, Zhou K, Otillar R, Merchant SS, Podell S, Gaasterland T, Napoli C, Gendler K, Manuell A, Tai V, Vallon O, Piganeau G, Jancek S, Heijde M, Jabbari K, Bowler C, Lohr M, Robbens S, Werner G, Dubchak I, Pazour GJ, Ren Q, Paulsen I, Delwiche C, Schmutz J, Rokhsar D, Van de Peer Y, Moreau H and Grigoriev IV (2007) The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation. Proc Natl Acad Sci USA 104: 7705–7710

Patron NJ, Greber B, Fahy BF, Laurie DA, Parker ML and Denyer K (2004) The lys5 mutations of barley reveal the nature and importance of plastidial ADP-Glc transporters for starch synthesis in cereal endosperm. Plant Physiol 135: 2088–2097

Picault N, Hodges M, Palmieri L and Palmieri F (2004) The growing family of mitochondrial carriers in Arabidopsis. Trends Plant Sci 9: 138–146

Proudlove MO and Thurman DA (1981) The uptake of 2-oxoglutarate and pyruvate by isolated pea chloroplasts. New Phytol 88: 255–264

Ravanel S, Block MA, Rippert P, Jabrin S, Curien G, Rebeille F and Douce R (2004) Methionine metabolism in plants: chloroplasts are autonomous for de novo methionine synthesis and can import S-adenosylmethionine from the cytosol. J Biol Chem 279: 22548–22557

Raven JA, Cockell CS and De La Rocha CL (2008) The evolution of inorganic carbon concentrating mechanisms in photosynthesis. Philos Trans R Soc Lond B Biol Sci 363: 2641–2650

Rebeille F, Ravanel S, Jabrin S, Douce R, Storozhenko S and Van Der Straeten D (2006) Folates in plants: biosynthesis, distribution, and enhancement. Physiol Plantarum 126: 330–342

Reddy KJ, Haskell JB, Sherman DM and Sherman LA (1993) Unicellular, aerobic nitrogen-fixing cyanobacteria of the genus Cyanothece. J Bacteriol 175: 1284–1292

Reinhold T, Alawady A, Grimm B, Beran KC, Jahns P, Conrath U, Bauer J, Reiser J, Melzer M, Jeblick W and Neuhaus HE (2007) Limited nocturnal ATP import into Arabidopsis chloroplasts causes photooxidative damage. Plant J 50: 293–304

Reiser J, Linka N, Lemke L, Jeblick W and Neuhaus HE (2004) Molecular physiological analysis of the two plastidic ATP/ADP transporters from Arabidopsis. Plant Physiol 136: 3524–3536

Renné P, Dreßen U, Hebbeker U, Hille D, Flügge UI, Westhoff P and Weber APM (2003) The *Arabidopsis* mutant *dct* is deficient in the plastidic glutamate/malate translocator DiT2. Plant J 35: 316–331

Rentsch D, Schmidt S and Tegeder M (2007) Transporters for uptake and allocation of organic nitrogen compounds in plants. FEBS Lett 581: 2281–2289

Reumann S and Weber APM (2006) Plant peroxisomes respire in the light: Some gaps of the photorespiratory C_2 cycle have become filled - Others remain. Biochim Biophys Acta 1763: 1496–1510

Reuzeau C, McNally JG and Pickard BG (1997) The endomembrane sheath: a key structure for understanding the plant cell? Protoplasma 200: 1–9

Reyes-Prieto A, Hackett JD, Soares MB, Bonaldo MF and Bhattacharya D (2006) Cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions. Curr Biol 16: 2320–2325

Reyes-Prieto A, Weber APM and Bhattacharya D (2007) The origin and establishment of the plastid in algae and plants. Annu Rev Genet 41: 147–168

Richards TA, Hirt RP, Williams BA and Embley TM (2003) Horizontal gene transfer and the evolution of parasitic protozoa. Protist 154: 17–32

Roeselers G, Huisjes EH, van Loosdrecht MC and Muyzer G (2008) Diversity and expression of cyanobacterial hupS genes in pure cultures and in a nitrogen-limited phototrophic biofilm. FEMS Microbiol Ecol 63: 292–300

Rotte C and Leustek T (2000) Differential subcellular localization and expressin of ATP sulfurylase and 5′-adenylylsulfate reductase during ontogenesis of Arabidopsis leaves indicates that cytosolic and plastid forms of ATP sulfurylase may have specialized functions. Plant Physiol 124: 715–724

Rumpho ME, Worful JM, Lee J, Kannan K, Tyler MS, Bhattacharya D, Moustafa A and Manhart JR (2008) Horizontal gene transfer of the algal nuclear gene psbO to the photosynthetic sea slug *Elysia chlorotica*. Proc Natl Acad Sci USA 105: 17867–17871

Sato N, Ishikawa M, Fujiwara M and Sonoike K (2005) Mass identification of chloroplast proteins of endosymbiont

origin by phylogenetic profiling based on organism-optimized homologous protein groups. Genome Inform 16: 56–68

Schimper AFW (1885) Untersuchungen über die Chlorophyllkörner und die ihnen homologen Gebilde. Jahrb Wiss Bot 16: 1–247

Schmitz-Esser S, Linka N, Collingro A, Beier CL, Neuhaus HE, Wagner M and Horn M (2004) ATP/ADP translocases: a common feature of obligate intracellular amoebal symbionts related to Chlamydiae and Rickettsiae. J Bacteriol 186: 683–691

Schneidereit J, Häusler RE, Fiene G, Kaiser WM and Weber APM (2006) Antisense repression reveals a crucial role of the plastidic 2-oxoglutarate/malate translocator DiT1 at the interface between carbon and nitrogen metabolism. Plant J 45: 206–224

Schwender J, Seemann M, Lichtenthaler HK and Rohmer M (1996) Biosynthesis of isoprenoids (carotenoids, sterols, prenyl side-chain of chlorophylls and plastoquinone) via a novel pyruvate/glyceraldehyde 3-phosphate non-mevalonate pathway in the green alga *Scenedesmus obliquus*. Biochem J 316: 73–80

Seifert GJ (2004) Nucleotide sugar interconversions and cell wall biosynthesis: how to bring the inside to the outside. Curr Opin Plant Biol 7: 277–284

Shannon JC, Pien F-M, Cao H and Liu K-C (1998) Brittle-1, an adenylate translocator, facilitates transfer of extraplastidial synthesized ADP-glucose into amyloplasts of maize endosperms. Plant Physiol 117: 1235–1252

Soll J and Schleiff E (2004) Protein import into chloroplasts. Nat Rev Mol Cell Biol 5: 198–208

Soll J, Bolter B, Wagner R and Hinnah SC (2000) Response: the chloroplast outer envelope: a molecular sieve? Trends Plant Sci 5: 137–138

Streatfield SJ, Weber A, Kinsman EA, Häusler RE, Li JM, Post-Beittenmiller D, Kaiser WM, Pyke KA, Flügge UI and Chory J (1999) The phosphoenolpyruvate/phosphate translocator is required for phenolic metabolism, palisade cell development, and plastid-dependent nuclear gene expression. Plant Cell 11: 1609–1621

Suda T, Kamiyama S, Suzuki M, Kikuchi N, Nakayama K, Narimatsu H, Jigami Y, Aoki T and Nishihara S (2004) Molecular cloning and characterization of a human multisubstrate specific nucleotide-sugar transporter homologous to Drosophila fringe connection. J Biol Chem 279: 26469–26474

Sun Q, Emanuelsson O and van Wijk KJ (2004) Analysis of curated and predicted plastid subproteomes of Arabidopsis. Subcellular compartmentalization leads to distinctive proteome properties. Plant Physiol 135: 723–734

Sun Q, Zybailov B, Majeran W, Friso G, Olinares PD and van Wijk KJ (2009) PPDB, the Plant Proteomics Database at Cornell. Nucleic Acids Res 37: D969–974

Svennerstam H, Ganeteg U and Nasholm T (2008) Root uptake of cationic amino acids by Arabidopsis depends on functional expression of amino acid permease 5. New Phytol 180: 620–630

Takahashi H, Asanuma W and Saito K (1999) Cloning of an Arabidopsis cDNA encoding a chloroplast localizing sulphate transporter isoform. J Exp Bot 50: 1713–1714

Tamagnini P, Axelsson R, Lindberg P, Oxelfelt F, Wunschiers R and Lindblad P (2002) Hydrogenases and hydrogen metabolism of cyanobacteria. Microbiol Mol Biol Rev 66: 1–20

Tegeder M and Weber APM (2006) Metabolite transporters in the control of plant primary metabolism. In: Plaxton WC and McManus MT (eds) Control of Primary Metabolism in Plants, pp 85–120. Wiley-Blackwell, Oxford

Terashita M, Maruyama S and Tanaka K (2006) Cytoplasmic localization of the single glutamine synthetase in a unicellular red alga, *Cyanidioschyzon merolae* 10D. Biosci Biotechnol Biochem 70: 2313–2315

Titus SA and Moran RG (2000) Retrovirally mediated complementation of the glyB phenotype. Cloning of a human gene encoding the carrier for entry of folates into mitochondria. J Biol Chem 275: 36811–36817

Tolbert NE (1980) Photorespiration. In: Davies DD (ed) The Biochemistry of Plants, pp 487–523. Academic Press, New York

Tyra H, Linka M, Weber APM and Bhattacharya D (2007) Host origin of plastid solute transporters in the first photosynthetic eukaryotes. Genome Biol 8: R212

Van der Straeten D, Rodrigues-Pousada RA, Goodman HM and Van Montagu M (1991) Plant enolase: gene structure, expression, and evolution. Plant Cell 3: 719–735

Vanderoost J, Bulthuis Ba, Feitz S, Krab K and Kraayenhof R (1989) Fermentation metabolism of the unicellular cyanobacterium Cyanothece PCC 7822. Arch Microbiol 152: 415–419

Villarejo A, Buren S, Larsson S, Dejardin A, Monne M, Rudhe C, Karlsson J, Jansson S, Lerouge P, Rolland N, von Heijne G, Grebe M, Bako L and Samuelsson G (2005) Evidence for a protein transported through the secretory pathway en route to the higher plant chloroplast. Nat Cell Biol 7: 1224–1231

Viola R, Nyvall P and Pedersen M (2001) The unique features of starch metabolism in red algae. Proc R Soc Lond Ser B-Biol Sci 268: 1417–1422

Voll L, Häusler RE, Hecker R, Weber A, Weissenbock G, Fiene G, Waffenschmidt S and Flügge UI (2003) The phenotype of the Arabidopsis *cue1* mutant is not simply caused by a general restriction of the shikimate pathway. Plant J 36: 301–317

Weber APM (2004) Solute transporters as connecting elements between cytosol and plastid stroma. Curr Opin Plant Biol 7: 247–253

Weber APM (2006) Synthesis, export, and partitioning of the end products of photosynthesis. In: Wise RR and Hoober JK (eds) The Structure and Function of Plastids, pp 273–292. Springer, New York

Weber APM and Fischer K (2007) Making the connections – the crucial role of metabolite transporters at the interface

between chloroplast and cytosol. FEBS Lett 581: 2215–2222

Weber A and Flügge UI (2002) Interaction of cytosolic and plastidic nitrogen metabolism in plants. J Exp Bot 53: 865–874

Weber A, Menzlaff E, Arbinger B, Gutensohn M, Eckerskorn C and Flügge UI (1995) The 2-oxoglutarate/malate translocator of chloroplast envelope membranes: Molecular cloning of a transporter containing a 12-helix motif and expression of the functional protein in yeast cells. Biochemistry 34: 2621–2627

Weber A, Servaites JC, Geiger DR, Kofler H, Hille D, Gröner F, Hebbeker U and Flügge UI (2000) Identification, purification, and molecular cloning of a putative plastidic glucose translocator. Plant Cell 12: 787–801

Weber APM, Schneidereit J and Voll LM (2004a) Using mutants to probe the in vivo function of plastid envelope membrane metabolite transporters. J Exp Bot 55: 1231–1244

Weber APM, Oesterhelt C, Gross W, Bräutigam A, Imboden LA, Krassovskaya I, Linka N, Truchina J, Schneidereit J, Voll H, Voll LM, Zimmermann M, Jamai A, Riekhof WR, Yu B, Garavito RM and Benning C (2004b) EST-analysis of the thermo-acidophilic red microalga *Galdieria sulphuraria* reveals potential for lipid A biosynthesis and unveils the pathway of carbon export from rhodoplasts. Plant Mol Biol 55: 17–32

Weber APM, Schwacke R and Flügge UI (2005) Solute transporters of the plastid envelope membrane. Annu Rev Plant Biol 56: 133–164

Weber APM, Linka M and Bhattacharya D (2006) Single, ancient origin of a plastid metabolite translocator family in Plantae from an endomembrane-derived ancestor. Eukaryot Cell 5: 609–612

Weeden NF (1981) Genetic and biochemical implications of the endosymbiotic origin of the chloroplast. J Mol Evol 17: 133–139

Weise SE, Weber APM and Sharkey TD (2004) Maltose is the major form of carbon exported from the chloroplast at night. Planta 218: 474–482

Winkler HH and Neuhaus HE (1999) Non-mitochondrial ATP transport. Trends Biochem Sci 24: 64–68

Woo KC, Flügge UI and Heldt HW (1987) A two-translocator model for the transport of 2-oxoglutarate and glutamate in chloroplasts during ammonia assimilation in the light. Plant Physiol 84: 624–632

Xu C, Fan J, Riekhof W, Froehlich JE and Benning C (2003) A permease-like protein involved in ER to thylakoid lipid transfer in Arabidopsis. EMBO J 22: 2370–2379

Xu C, Yu B, Cornish AJ, Froehlich JE and Benning C (2006) Phosphatidylglycerol biosynthesis in chloroplasts of Arabidopsis mutants deficient in acyl-ACP glycerol-3-phosphate acyltransferase. Plant J 47: 296–309

Yoon HS, Hackett JD, Ciniglia C, Pinto G and Bhattacharya D (2004) A molecular timeline for the origin of photosynthetic eukaryotes. Mol Biol Evol 21: 809–818

Zybailov B, Rutschow H, Friso G, Rudella A, Emanuelsson O, Sun Q and van Wijk KJ (2008) Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. PLoS ONE 3: e1994

Chapter 9

Phototrophic CO_2 Fixation: Recent Insights into Ancient Metabolisms

Thomas E. Hanson*
College of Earth, Ocean and Environment, Department of Biological Sciences and Delaware Biotechnology Institute, University of Delaware, 127 DBI, 15 Innovation Way, Newark, DE 19711, USA

Birgit E. Alber
Department of Microbiology, The Ohio State University, 484 West 12th Avenue, Columbus, OH 43210-1292, USA

F. Robert Tabita
Department of Microbiology and The Plant Molecular Biology Biology/Biotechnology Program, The Ohio State University, 484 West 12th Avenue, Columbus, OH 43210-1292, USA

*Author for correspondence, e-mail: tehanson@udel.edu

R.L. Burnap and W.F.J. Vermaas (eds.), Functional Genomics and Evolution of Photosynthetic Systems, Advances in Photosynthesis and Respiration 33, pp. 225–251, DOI 10.1007/978-94-007-1533-2_9,

Summary

The production of all (autotrophy) or a substantial proportion (mixotrophy) of newly synthesized biomass from carbon dioxide is a hallmark of plants and a number of bacteria and archaea. Carbon dioxide fixation enables autotrophs to form the basis of entire ecosystem foodwebs as primary producers and mixotrophs to efficiently utilize simple forms of carbon present in various environments. During fixation, CO_2 must ultimately be reduced to the level of formaldehyde for assimilation into biomass and, for this reason, CO_2 fixation is also utilized by numerous organisms to dispose of excess reducing power. Currently, six distinct CO_2 fixation pathways are recognized amongst autotrophic prokaryotes and eukaryotes. The impact of genomics on our understanding of the distribution, function(s), and regulation of these pathways and their unique enzymes will be described as will two recently discovered pathways, one for autotrophic or mixotrophic CO_2 fixation in some photosynthetic bacteria and the other for acetate assimilation by a wide range of bacteria, including phototrophs, that involves a novel carboxylation step.

I. Overview of CO_2 Fixation Pathways

A. Physiological Functions of CO_2 Fixation

As this volume is dedicated to the comparative genomics of phototrophic organisms, the focus of most of this chapter will be on the subset of CO_2 fixation pathways that are relevant to these types of organisms. To provide a larger context for these discussions, a brief and basic overview of the functions of CO_2 fixation and the diversity of CO_2 fixing pathways will be presented first.

All heterotrophic organisms fix small amounts of CO_2 through the action of anaplerotic metabolism, also known as heterotrophic carbon fixation. For example, organisms rely on the tricarboxylic acid (TCA) cycle in part to provide carbon skeletons for biosynthesis of glutamate or glutamine. The pools of TCA cycle intermediates depleted by this flux are replenished via pyruvate carboxylase, phosphoenol pyruvate carboxylase or carboxykinase. In addition, precursors of odd numbered fatty acids are generated by acetyl-CoA carboxylase. Typically, 3–8% of cellular carbon is acquired by heterotrophs in this fashion (Roslev et al., 2004; Feisthauer et al., 2008) depending on the carbon source utilized.

With respect to carbon metabolism, autotrophs and mixotrophs differ from heterotrophs in two fundamental ways. First, auto-/mixotrophs derive far greater amounts of their cellular carbon from CO_2 fixation than do heterotrophs, such that autotrophs typically acquire all of their cellular carbon from CO_2 and mixotrophs obtain >30% of their total cellular carbon from CO_2. Second, autotrophs and mixotrophs utilize distinct multistep pathways for the reduction of CO_2 and its assimilation rather than the single enzyme shunts between central metabolic intermediates typical of heterotrophic carbon fixation. These CO_2 fixation pathways have unique and characteristic enzymes that can be biochemically assayed in cultures, identified in DNA samples by PCR amplification with gene specific primers, or by sequence homology in genomic sequence data. The six currently accepted autotrophic CO_2 fixation pathways are: the Calvin-Benson-Bassham (CBB) cycle, the reductive or reverse TCA (rTCA) cycle, the acetyl-CoA or Wood-Ljungdahl pathway, the 3-hydroxypropionate/malyl-CoA cycle (3-HPP), the dicarboxylate/4-hydroxybutyrate cycle and the 3-hydroxypropionate/4-hydroxybutyrate (3-HP/4-HB) pathway (Table 9.1).

While the acquisition of cellular carbon is the main function of autotrophic and mixotrophic pathways, an alternative function is the disposal of excess reducing power. Perhaps the most convincing demonstration of this function of CO_2 fixation pathways has been made in nonsulfur

Abbreviations: 3-HP – 3-hydroxypropionate; 3-HPP – 3-HP pathway; 4-HB – 4-hydroxybenzoate; ACL – ATP:citrate lyase; CBB cycle – Calvin-Benson-Bassham cycle; CCL – citryl CoA synthase; CCS – citryl CoA lyase; Fd – Ferredoxin; Fe-S – Iron sulfur cluster; NSP bacteria – Nonsulfur purple phototrophic bacteria; ORF – Open reading frame; RC1/2 – Reaction center type 1 (Fe-S reducing) or type 2 (quinone reducing); rTCA – reductive or reverse TCA cycle

Table 9.1. A comparison of CO_2 fixation pathways

Pathway	Key enzymes	Per triose-PO_4		Pathway known in			
		ATP	e^- donors	Phototrophs	Bacteria	Archaea	Eukarya
CBB	RubisCO Phosphoribulokinase	9	6 NAD(P)H	Yes	Yes	No[a]	Yes
rTCA	Pyruvate synthase (Pyr:Fd oxidoreductase) α-KG synthase (α-KG:Fd oxidoreductase) ATP:citrate lyase or Citryl-CoA synthetase+lyase	5	5 NAD(P)H 2 Fd	Yes	Yes	Yes	Yes
3-HPP	Malonyl-CoA reductase Propionyl-CoA synthase	8	5 NAD(P)H	Yes	Yes	No	No
Acetyl-CoA	CO dehydrogenase/acetyl-CoA synthase Pyruvate synthase (Pyr:Fd oxidoreductase)	5	4 NAD(P)H 3 Fd 1 H_2	No	Yes	Yes	No
3-HP/4-HB	Succinate semialdehyde reductase 4-Hydroxybutyryl-CoA synthetase	9	5 NAD(P)H 2 Fd	No	No	Yes	No
Dicarboxylate/4-HB	Succinate semialdehyde reductase 4-Hydroxybutyryl-CoA synthetase	8	2 NAD(P)H 8 Fd	No	No	Yes	No

[a]RubisCO activity has been demonstrated in numerous archaea. However, RuBP may not be synthesized via phosphoribulokinase in many archaea, but rather as an intermediate in nucleotide salvage pathways. Thus, a classical CBB cycle has yet to be demonstrated in archaea

3-HP 3-hydroxyporpionate, *4-HB* 4-hydroxybutyrate

purple phototrophic (NSP) bacteria. When grown photoheterotrophically, with light and a source of fixed carbon, these organisms still express CBB cycle genes and fix significant amounts of CO_2 (Wang et al., 1993; Joshi et al., 2009). Mutants lacking ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) are unable to grow under these conditions unless provided with an alternative external electron acceptor like dimethylsulfoxide (DMSO) (Wang et al., 1993). Spontaneously arising second site suppressor mutations that allow photoheterotrophic growth in CBB mutants of *Rhodobacter sphaeroides, Rhodobacter capsulatus, Rhodospirillum rubrum,* and *Rhodpseudomonas palustris* result in the utilization of nitrogenase (Joshi and Tabita, 1996; Tichi and Tabita, 2000; Laguna et al., 2010), carbon monoxide dehydrogenase (Joshi and Tabita, 2000), or as yet unidentified CO_2 fixation pathways (Wang et al., 1993) as alternative redox balancing pathways. As there is functional evidence for the presence of an rTCA cycle in the magnetotactic bacterium *Magnetococcus* strain MC1 (Williams et al., 2006), a member of the *Alphaproteobacteria* as are the *Rhodobacterales*, *Rhodospirillales*, and *Rhizobiales* (*Rps. palustris*), it may be that a variant of the rTCA cycle is operative as the unidentified carbon fixation pathway.

B. Comparison of CO_2 Fixation Pathways

CO_2 fixation pathways differ significantly from one another in the identity of key and/or unique enzymatic catalysts, the central metabolic intermediates produced from CO_2, the amount of ATP required, and types of reductant required (Table 9.1). Organisms that utilize CO_2 as a sole carbon source also tend to utilize strongly

thermodynamically favorable chemical reactions like the oxidation of hydrogen gas, reduced sulfur compounds, or ferrous iron (chemolithoautotrophs) with a suitable terminal electron acceptor or take advantage of the most readily available source of energy on the planet Earth, light (photolithoautotrophs). Chemolithoautotrophic organisms are known that use any of the six currently known autotrophic pathways. However, only the CBB cycle, rTCA cycle, and 3-HPP pathway are known among the photolithoautotrophs. These autotrophic pathways, along with the ethylmalonyl CoA pathway for acetate assimilation, will be the focus of this review. The latter pathway is included because of its importance to anoxygenic phototrophic bacteria and because it contains an unusual carboxylation step for heterotrophic CO_2 fixation.

1. Distribution of CO_2 Fixation Pathways in Phototrophs

For this article, we consider phototrophs to be those organisms that utilize a chlorophyll or bacteriochlorophyll based reaction center and associated electron transport complexes for the conservation of light energy as a proton gradient (Xiong and Bauer, 2002; Olson and Blankenship, 2004; Iverson, 2006). While halophilic archaea (Lozier et al., 1975) and marine bacteria and archaea (Frigaard et al., 2006) that contain rhodopsins are capable of conserving light energy, too little is known about their carbon metabolism to make definitive statements about their modes of growth. This is one area where targeted culturing and physiological studies will provide useful information in the future.

There is no obvious relationship between the mechanisms of light energy capture through the reaction center and the specific CO_2 fixation pathway utilized in phototrophs. Two types of reaction centers that convert light energy to chemical energy are known amongst the phototrophs, those that reduce a low potential Fe-S cluster (RC1) and those that reduce quinones (RC2) (Xiong and Bauer, 2002; Heinnickel and Golbeck, 2007). Bacterial photolithoautotrophs that possess only a single reaction center type (RC1 or RC2, green sulfur/nonsulfur bacteria, and purple sulfur/nonsulfur bacteria, the anoxygenic phototrophs) utilize many of the same electron donors as do chemolithoautotrophs, but do not have to direct electron flow towards an external electron acceptor thereby increasing energy efficiency and autotrophic growth yields. The proximal reason for this is that most electron donors do not have sufficient reducing power to directly reduce the electron donors used to fix CO_2 in autotrophic pathways: NADH or ferredoxin. For example, the oxidation of sulfide to polysulfide has a redox potential of ~ −270 mV relative to the standard hydrogen electrode while the redox potential for the $NADH/NAD^+$ couple is ~ −320 mV and that of ferredoxin is ~ −500 mV (Thauer, 2007). To overcome this thermodynamic barrier, chemolithoautotrophs must utilize some proton motive force via reverse electron transport to generate NADH and reduced ferredoxin (Stams and Plugge, 2009). This is proton motive force that cannot be used for substrate transport or ATP synthesis thereby decreasing growth efficiency. In phototrophs, however, light energy is captured in the reaction center to provide either additional proton motive force via electron transport from the quinone pool through the cytochrome bc_1 complex (RC2 type reaction centers) or electrons with sufficient potential to reduce NAD^+ and oxidized ferredoxin (RC1 type reaction centers). Currently, phototrophs are only known to utilize the CBB cycle, the rTCA cycle, or the 3-HPP pathway. Most of these organisms will also grow photomixotrophically or photoheterotrophically (with more complex carbon sources like pyruvate, malate, butyrate, glutamate, etc.) when given the option.

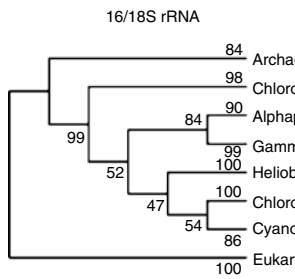
Photolithoautotrophs that possess two coupled reaction centers (RC1 + 2, cyanobacteria, prochlorophytes and all eukaryotic phototrophs) and an oxygen-evolving complex (the oxygenic phototrophs) have gained the ability to utilize water as an electron donor for photosynthesis (Iverson, 2006) and therefore have escaped the need for constant sources of H_2, H_2S or other reductants, allowing them to colonize a wider range of environments. These organisms appear to exclusively utilize the CBB cycle for autotrophic CO_2 fixation. Photoheterotrophy (Anderson and McIntosh, 1991) and dark heterotrophy (el-Refai et al., 1974; Graves et al., 1990; Summers et al., 1995) are known in some of the oxygenic phototrophs on a variety of carbon sources including sugars and acetate.

2. Evolution of Autotrophy/Photosynthesis: Summarizing Current Debates

As photoautotrophic primary productivity forms the basis of most ecosystems, the combination of autotrophic CO_2 fixation with chlorophyll and bacteriochlorophyll based phototrophic reaction centers in several organisms was a critical step in the evolution of the biosphere. Both phototrophic reaction centers and autotrophic CO_2 fixation pathways evolved early during the history of life on earth as there is abundant isotopic evidence (Zerkle et al., 2005) and fossil biomarker evidence (Olson, 2006) for the presence of cyanobacteria carrying out oxygenic photosynthesis by ~2.8 Ga ago followed by the "Great Oxidation Event" ca. 2.4 Ga ago. The earliest geological evidence for photosynthesis relies on carbon isotopic records preserved in rocks dating ~3.8 Ga ago where H_2, Fe^{2+}, or reduced sulfur compounds were the likely electron donors (Olson, 2006), though some evidence exists for the presence of oxygenic photosynthesis at these earliest dates (Buick, 2008). While it is often assumed that these signatures reflect photoautotrophy (Rothschild, 2008), it is difficult to state this with certainty as the evidence is correlative relying on the co-occurrence of chemical species, isotopic signatures, and biomarkers that may not have all been produced by the same organism or process. This makes the independent assessment of the early evolution of reaction centers and CO_2 fixation difficult. Moreover, it is still not clear that the first cells to arise were autotrophic. The current consensus is that a "heterotrophic first" phase of cellular evolution occurred in a prebiotic soup similar to that proposed by Oparin (1950). Supporting evidence is found in the classic Miller (1953) experiments on pre-biotic synthesis of organic compounds, including recent repetitions of these experiments in both reducing and neutral model archean atmospheres (Cleaves et al., 2008; Johnson et al., 2008).

If the "chemoautotrophic first" hypothesis as articulated by Wachtershauser (Wachtershauser, 1988, 1990) is correct, then it seems likely either the rTCA or the reductive acetyl-CoA pathway was the first autotrophic pathway and perhaps the first metabolic pathway to exist. The two key CO_2 fixing enzymes of the rTCA cycle, pyruvate synthase (Pyr:ferredoxin oxidoreductase) and α-ketoglutarate synthase (α-KG:Fd oxidoreductase), rely on Fe-S centers for internal electron transport to CO_2 that are similar to features present on charged Fe-S layers that occur in hydrothermal systems (Wachtershauser, 1990). Recently, a modified version of a prebiotic rTCA cycle has been proposed that explicitly involves hydrogen sulfide (Kalapos, 2007), but which is somewhat energetically less favorable than that originally proposed. However, an rTCA cycle first scenario is disputed by others on thermodynamic and mechanistic grounds (Orgel, 2008). The reductive acetyl-CoA pathway has been proposed as the first metabolic pathway because of a similar reliance on Fe-S chemistry and the argument that the tetrahydrofolate and tetrahydromethanopterin C_1 carriers in this pathway reflect an "RNA-world" origin with the advent of an incomplete rTCA cycle as a follow on event (Martin and Russell, 2007). The CBB cycle is usually considered to be a late evolving autotrophic pathway, most recently argued from the standpoint of a heterotrophic-first origin for core metabolism that puts the closure of the CBB cycle as a late event (Cunchillos and Lecointre, 2007). A far more detailed discussion of the evolution of photosynthesis and autotrophy can be found in a 2008 report from a Royal Society Discussion meeting on "Photosynthetic and atmospheric evolution" from which several references in this article (Buick, 2008; Rothschild, 2008; Tabita et al., 2008a) are taken.

Irrespective of how, or in what order, CO_2 fixation pathways and the harvesting of light energy evolved, in the modern world these two processes are tightly linked. Light harvesting reaction centers and CO_2 fixation pathways have likely merged within single genomes multiple times over the course of evolution to produce the combinations observed in currently living organisms (Fig. 9.1). The major correlation observed in this comparison is that the CBB cycle appears to be highly correlated with the presence of the quinone reducing RC2 reaction center in the *Proteobacteria*, *Cyanobacteria* and eukaryotic phototrophs. However this correlation is not universal as the *Chloroflexi* (filamentous anoxygenic phototrophic bacteria) possess an RC2 type of reaction center, but fix CO_2 via the 3-HPP cycle. All currently



	Reaction Center		Fixation/Assimilation Pathway CO_2			Acetate
	RC1	RC2	CBB	rTCA	3-HPP	EtMa-CoA
Archaea	–	–	–	+	+	–
Chloroflexi	–	+	–	–	+	–
Alphaproteobacteria	–	+	+	–	–	+
Gammaproteobacteria	–	+	+	–	–	–
Heliobacteria	+	–	–	Partial	–	–
Chlorobia	+	–	–	+	–	–
Cyanobacteria/Prochlorophytes	+	+	+	–	–	–
Eukarya	+	+	+	–	–	–

Fig. 9.1. Phototrophs, reaction centers and CO_2 fixation pathways. A Minimum Evolution phylogenetic tree constructed from 73 16S or 18S rRNA gene sequences of phototrophic bacteria and eukaryotes was constructed with the Archaea as a non-phototrophic outgroup. Numbers at nodes are bootstrap percentages from 1,000 replicates and branch lengths are not significant in this representation. The resulting tree was compared with the type of photosynthetic reaction center, type of CO_2 fixation pathway utilized by each lineage, and the presence or absence of the ethyl malonyl-CoA pathway for acetate assimilation (EtMa-CoA) that contains a novel heterotrophic CO_2 fixation reaction.

known oxygenic phototrophs that possess both RC1 and RC2 reaction centers fix CO_2 via the CBB cycle.

II. The Calvin-Benson-Bassham (CBB) Cycle

A. RubisCO/RLP: Genomics, RLP Function and Evolution

The advent of microbial genome sequencing allowed for gene discovery by homology search that is independent of classical methods of functional assessment for gene products. As archaeal and bacterial genome sequences accumulated, it was noted that a number of phylogenetically distantly related organisms possessed genes whose protein products were similar to the only two forms of RubisCO that had been biochemically characterized at the time, forms I and II. Careful phylogenetic analysis of these sequences indicated that these protein sequences could be sorted into a number of well defined and distantly related clades and led to the establishment of the RubisCO-like protein (RLP) family (Hanson and Tabita, 2001). Genetic and biochemical analysis of RubisCO and RLP genes and proteins has provided a significantly advanced understanding of the structure, function and evolution of this dynamic protein family. This has been recently and extensively reviewed (Tabita et al., 2007, 2008a, b). In addition, structures have been solved for RLPs from *C. tepidum* (Li et al., 2005), *Rps. palustris* (Tabita et al., 2007) and *Geobacillus kaustophilus* (Imker et al., 2007).

Our current understanding indicates that there exist two major sub-divisions in this protein superfamily: the bona fide RubisCO proteins, comprised of forms I, II, and III, and the RubisCO-like proteins (RLPs or form IV). Whereas all members of form I–III RubisCO contain all requisite active side residues and are catalytically active when assayed for both carboxylase and oxygenase activity, the RLPs all contain amino acid substitutions at key active site residues that render them incapable of catalyzing the classical RubisCO reactions. Distinct patterns of active site substitutions are found between the six distinct clades of RLPs that suggest they have evolved to carry out distinct cellular functions.

RubisCO is a highly promiscuous enzyme, capable of forming a variety of intermediates after enolization of its substrate ribulose 1, 5-bisphosphate. Thus, in hindsight it was not surprising that similar reactions using enolate intermediates are carried out by the RLPs. Those RLPs that have been functionally characterized catalyze the isomerization of thiolated and phosphorylated intermediates in methionine salvage pathways that differ between microorganisms. In *Bacillus subtilis* and *Geobacillus kaustophilus*, the RLP (MtnW/YkrW) catalyzes the tautomerization of 2,3-diketo-5-methylthiopentane 1-phosphate

(DKMTP) to 2-hydroxy-3-keto-5-thiomethylpent-1-en 1-phosphate in a pathway for methylthioadenosine recycling (Ashida et al., 2003; Imker et al., 2007). This reaction is reminiscent of the enolization reaction performed by bona fide RubisCO and, indeed, form II RubisCO can also catalyze the tautomerization of DKMTP as shown by the ability of the *R. rubrum cbbM* gene to complement a *B. subtilis ykrW* mutant (Ashida et al., 2003). The notion that distinct RLP clades act on distinct substrates was confirmed by the recent finding that the RLP of *R. rubrum* catalyzes the isomerization of 5-methylthio-D-ribulose-1-phosphate to a 3:1 mixture of 1-methylthioxylulose-5-phosphate and 1-methylthioribulose-5-phosphate (Fig. 9.2) in an unprecedented 1,3 proton transfer reaction via two successive 1,2 proton transfers (Imker et al., 2008). However, the *R. rubrum* RLP is inactive with DKMTP, the substrate for *B. subtilis* RLP (Imker et al., 2008). The RLPs of *R. rubrum* and *B. subtilis* belong to distinct sub-families within the RubisCO/RLP family as defined by phylogenetic relationships and distinct patterns of active site residue substitutions (Tabita et al., 2007). As *R. rubrum* can utilize methylthioadenosine as a sole sulfur source and the *R. rubrum* RLP is required for this process, it suggests that a modified pathway for methylthioadenosine recycling occurs in *R. rubrum* (Imker et al., 2008) and most likely in other organisms that possess RLPs.

The distribution of RubisCO and RLP subfamilies across phototrophic lineages is shown in Fig. 9.3. Based on a more detailed analysis of the distribution of RubisCO and RLP across phototrophic and non-phototrophic microorganisms, a model for the evolution of RubisCO and RLP superfamily proposes that a bona fide RubisCO first evolved in the archaea and later evolved into the current eight families of RubisCO and RLP (Tabita et al., 2007, 2008b). The complex pattern of distribution of RubisCO and RLP seen in extant phototrophs (Fig. 9.3) is a consequence of a long evolutionary history coupled with multiple events of lateral gene transfer and gene acquisition in eukaryotes by multiple endosymbiotic events. This model predicts that the evolution of phosphoribulokinase to efficiently produce RuBP, and not the evolution of RubisCO, was the critical step in the evolution and closure of the CBB cycle, which agrees with the previously noted argument that the CBB cycle is a relatively late evolving autotrophic pathway (Cunchillos and Lecointre, 2007). It should be noted that "late" in

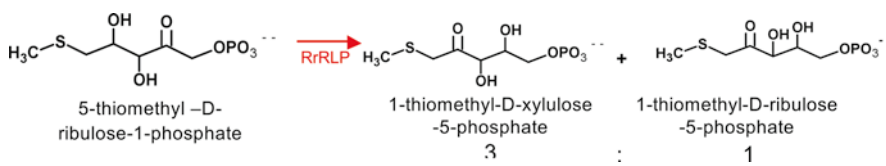
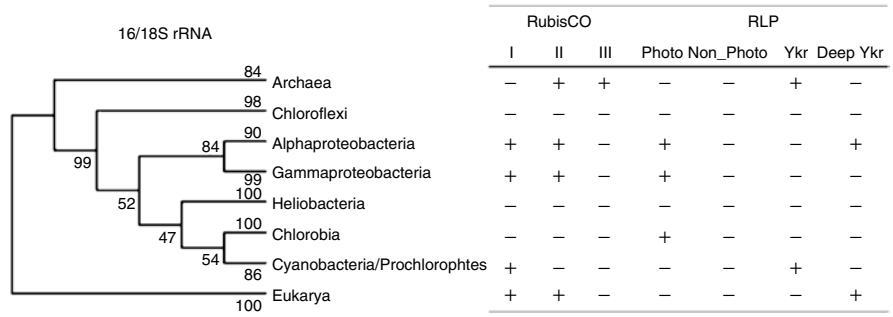


Fig. 9.2. Reaction catalyzed by R. rubrum RLP (From Imker et al., 2008).



	RubisCO			RLP			
	I	II	III	Photo	Non_Photo	Ykr	Deep Ykr
Archaea	–	+	+	–	–	+	–
Chloroflexi	–	–	–	–	–	–	–
Alphaproteobacteria	+	+	–	+	–	–	+
Gammaproteobacteria	+	+	–	+	–	–	–
Heliobacteria	–	–	–	–	–	–	–
Chlorobia	–	–	–	+	–	–	–
Cyanobacteria/Prochlorophtes	+	–	–	–	–	+	–
Eukarya	+	+	–	–	–	–	+

Fig. 9.3. Distribution of RubisCO and RLP forms across the phototrophs. The presence or absence of specific RubisCO (I, II, III) or RLP (Photo, Non-Photo, Ykr, DeepYkr) sub-families (Tabita et al., 2007) in phototrophic lineages (see Fig. 9.1 for tree construction) reflects the complex evolutionary history of this enzyme family.

this case still means that the CBB cycle had likely evolved by ca. 3.5 Ga ago as suggested by carbon isotopic records (Zerkle et al., 2005).

B. RubisCO Structure/Function and Regulation in NSP Bacteria

Because of the structural and functional diversity of RubisCO molecules obtained from phototrophic bacteria, these proteins have provided bountiful model systems to elucidate many aspects of the catalytic mechanism. Still, there remain many mysteries, including the molecular basis for CO_2/O_2 specificity, how residues distal from the active site influence catalysis and interactions with the gaseous substrates, and how the active site of this enzyme has become adapted or has evolved to accommodate vastly different intracellular milieus in diverse organisms. Several recent reviews treat various aspects of these issues and will not be repeated here (Tabita et al., 2007, 2008a, b). However, it might be instructive to underline the fact that these organisms, particularly the nonsulfur purple bacteria, provide useful systems to probe these remaining mysteries and other aspects of catalysis and regulation.

1. Bioselection of RubisCO Variants

In protein structure-function studies, it is desirable to provide an unbiased way to select for mutant forms of a protein that could shed light on key questions of function. That is, if a convenient and rapidly responding organism could be developed whose growth was dependent on the protein in question, randomly mutagenized genes of that protein could provide insights into structure-function relationships that would be both unpredictable and nonprejudicial. Among the many types of organisms that catalyze CO_2 fixation are the NSP phototrophic bacteria. A defining characteristic exhibited by these organisms is their metabolic versatility, as they are unique in their ability to employ all the known modes of metabolism; i.e., they are able to grow photolithoautotrophically (PA) and photoheterotrophically (PH) under anaerobic conditions, as well as chemolithoautotrophically (CA) and chemoheterotrophically (CH) under aerobic conditions (Romagnoli and Tabita, 2009). Most importantly, these organisms may be cultured under conditions where CO_2 is the sole source of carbon, both anaerobically, and for some strains, under aerobic growth conditions. These latter modes of growth of course require functional RubisCO. Initially, RubisCO knockout strains of *Rba. sphaeroides* were constructed (Falcone and Tabita, 1991), along with specific delivery systems that would allow complementation of this knockout strain with a variety of desired RubisCO genes. Although useful, *Rba. sphaeroides* does not grow well under aerobic CA conditions, even though gain of function mutant strains could be isolated that grew better under these conditions than the wild type (Paoli and Tabita, 1998). Because it would be desirable to select for mutant forms of RubisCO under conditions where oxygen is either present or absent when CO_2 is used as sole carbon source, a more suitable photosynthetic host organism that grew well under aerobic CA conditions was sought. *Rba. capsulatus* grows well under aerobic CA conditions (Madigan and Gest, 1979) and this organism, for the most part grows faster than *Rba. sphaeroides* under virtually all conditions, including under anaerobic PA conditions. After characterizing the *cbb* operons of *Rba. capsulatus*, it was recognized that it would be feasible to construct RubisCO knockout strains of this organism. A RubisCO knockout strain, strain SBI/II$^-$, was constructed and this strain could be complemented with any desired prokaryotic RubisCO gene(s) (Paoli et al., 1998). Improved broad host range expression vectors were constructed that were convenient for cloning randomly mutagenized RubisCO genes in *E. coli*. Such vectors also contained a highly active *R. rubrum cbbM* promoter and its cognate transcriptional activator gene, *cbbR*, to drive RubisCO gene expression for cloned *cbb* or *rbc* genes in *Rba. capsulatus* strain SBI/II$^-$ (Smith and Tabita, 2003). Most importantly, this construct insures that RubisCO levels are always high and nonvariable under selection conditions. A variety of eubacterial, cyanobacterial, and archaeal RubisCO genes have been expressed using this *Rba. capsulatus* system (Paoli et al., 1998; Finn and Tabita, 2003; Smith and Tabita, 2003) and strain SBI/II$^-$ has been extensively used for examining structure-function relationships with the highly studied *Synechococcus* spp. PCC6301 RubisCO (Smith and Tabita, 2004; Tabita et al., 2008a, b), a convenient close relative of the plant enzyme, for which recombinant systems are unavailable. This selection system has been especially useful to study a highly conserved

hydrophobic pocket and the role of residues that confer oxygen resistance to RubisCO (Tabita et al., 2008a, b; Satagopan et al., 2009). *Rps. palustris* (Yoshida et al., 2006) and an *E. coli*-based selection system (Greene et al., 2007) have also been recently employed, such that it is apparent that bioselection will be increasingly used to facilitate RubisCO structure-function studies.

2. Organization and Regulated Expression of CO_2 Fixation Genes

The metabolic diversity of the NSP phototrophic bacteria ensures that key processes such as CO_2 fixation are highly regulated, as CO_2 fixation and other important processes are dispensable under certain growth conditions. With facile genetic systems, such organisms are very amenable to detailed studies of molecular regulation. With respect to molecular mechanisms governing CO_2 assimilation, studies have primarily focused on model organisms such as *Rba. sphaeroides* and *Rba. capsulatus,* and to some extent *Rhodospirillum rubrum* and *Rps. palustris*. Each of these organisms provides unique advantages. Moreover, it should also be noted that NSP bacteria are also able to fix nitrogen, evolve hydrogen, and utilize oxygen or other diverse alternative electron acceptors for respiratory purposes. Interestingly, the control of these processes appears to be integrated with the regulation of CO_2 fixation (reviewed in Dubbs and Tabita, 2004; Romagnoli and Tabita, 2009). In addition to obvious functions of supplying needed carbon, CO_2 reduction plays a pivotal role in maintaining the redox balance of these organisms, as noted above. Finally, NSP bacteria are able to utilize a large variety of both organic and inorganic electron donors to support growth.

a. Regulation of CO_2 Fixation Gene Expression

In the NSP bacteria, the major metabolic route for carbon dioxide assimilation is via the Calvin-Benson-Bassham (CBB) reductive pentose phosphate cycle, with ribulose 1,5 bisphosphate carboxylase/oxygenase (RubisCO) catalyzing the actual CO_2 fixation reaction. Through the action of RubisCO and other enzymes of the pathway, necessary organic compounds may be synthesized to sustain all the major metabolic

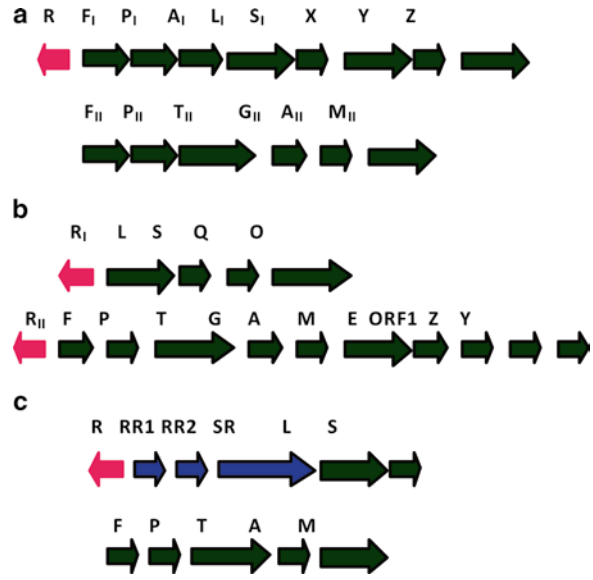


Fig. 9.4. Organization of *cbb* genes in NSP bacteria. Genes are grouped into cbb_I (upper cluster) and cbb_{II} (bottom cluster) operons in *Rba. spheroides* (**a**), *Rba. capsulatus* (**b**), and *Rps. palustris* (**c**). Note the unique RR1, RR2 and SR genes (in *blue*) in *Rps. palustris*.

requirements of the cell. In NSP bacteria, the genes that encode enzymes responsible for this process are for the most part organized in two separate loci, the cbb_I and cbb_{II} operons (Fig. 9.4). Each operon contains distinct structural genes encoding different forms of RubisCO; the *cbbLS* genes of the cbb_I operon encode, respectively, the large and small subunits of form I RubisCO, while the *cbbM* gene of the cbb_{II} operon encodes the single subunit of form II RubisCO. These loci also contain structural genes in single or double copies that encode other CBB cycle enzymes, in addition to putative short open reading frames of unknown function (reviewed in Gibson, 1995; Dubbs and Tabita, 2004). The arrangement of *cbb* genes into discrete operons is characteristic of several CO_2-fixing bacteria including chemoautotrophs and related NSP bacteria (Gibson and Tabita, 1996). Although the arrangement of genes differs, in all CO_2 fixers capable of also metabolizing organic carbon, *cbb* gene expression is regulated through the mediation of a transcriptional activator gene, *cbbR*, whose deduced amino acid sequence resembles LysR and other regulatory proteins (LTTR) of this group (Gibson, 1995). In all three organisms shown, *cbbR* is divergently transcribed and in both *Rba. sphaeroides* and *Rps. palustris* the product of a

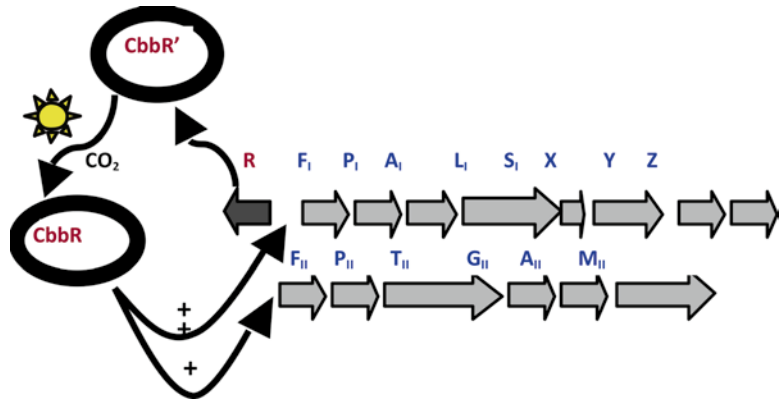


Fig. 9.5. Cartoon showing the importance of CbbR activation in Rb. sphaeroides under growth conditions that favor active CO_2 fixation. Under these conditions, RuBP (designated, R), an identified positive effector, binds to CbbR, altering its conformation (CbbR′ to CbbR) so that CbbR can now positively activate transcription of the *cbb* operons (Smith and Tabita, 2002; Tichi and Tabita, 2002; Dubbs et al., 2004; Dangel et al., 2005).

single *cbbR* gene (CbbR) regulates both operons, while each operon of *Rba. capsulatus* is regulated by a cognate and specific *cbbR* gene (Gibson and Tabita, 1993; Paoli et al., 1998; Vichivanives et al., 2000), In addition, *Rps. palustris* possesses an interesting three-protein two-component regulatory system that appears to modify the ability of CbbR to activate transcription (Romagnoli and Tabita, 2006) (the CbbRRS system comprising the *cbbRR1*, *cbbRR2*, and *cbbSR* genes in Fig. 9.4c).

CO_2 fixation gene expression is highly regulated and over the years work in the FRT laboratory has shown that gene expression is dependent on the activation of a specific regulator protein, CbbR. The cartoon (Fig. 9.5) best illustrates this as basically CbbR must first be activated by a positive effector (RuBP, abbreviated as R in the figure) made under certain growth conditions (i.e., when CO_2 is used as sole carbon source) before genes important for CO_2 fixation are transcribed (Smith and Tabita, 2002; Tichi and Tabita, 2002; Dubbs et al., 2004; Dangel et al., 2005). When CbbR is activated, it can then effectively bind to the cbb_I or cbb_{II} promoters, thus activating transcription (shown by the arrows in Fig. 9.5).

b. Structure and Function of CbbR

Although there have been hundreds of different LTTR proteins that have been described that regulate important processes in bacteria such as virulence, biodegradation, nitrogen metabolism, etc., little is known about their biochemistry and how they function. The structure of the first full-length LTTR regulator protein, CbnR, was initially reported (Muraoka et al., 2003), followed soon after by the structure of another LTTR, DntR (Smirnova et al., 2004). Moreover, structures of BenM and CatM with bound effectors were recently reported (Ezezika et al., 2007), providing the first close-up look at the actual binding sites within LTTR protein-effector complexes. The structures of CbnR, DntR, BenM, and CatM are thus very important as they clarify many aspects, such as the oligomerization state and parts of the protein important for function (Muraoka et al., 2003; Smirnova et al., 2004; Ezezika et al., 2007). While several very interesting studies have addressed issues of function for some important LTTRs, such as OccR (Akakura and Winans, 2002) and CysB (Lochowska et al., 2001), there is absolutely nothing known about the CbbR class and how these proteins function specifically to regulate CO_2 fixation. Recently, we developed a means to bioselect randomly mutagenized *cbbR* genes that both negatively and positively affect CbbR-mediated transcription. This has lead to a wealth of information on the structural basis for effector-mediated CbbR function (Dangel et al., 2005). In this study, a cbb_I promoter-*lacZ* fusion was inserted into the chromosome of a host strain of *Rba. sphaeroides* in which the endogenous *cbbR* gene was inactivated. This strain (strain 87) was subsequently used to detect mutations in CbbR that affect its function, assayed after separately mutagenizing *cbbR*

sequences, followed by noting the ability of plasmid copies of these sequences to complement strain 87 and yield a color change on X-gal plates; e.g., constitutively active CbbR proteins turn colonies bright blue on X-gal plates under normally repressive aerobic CH growth conditions. Thus, single or multiple residue mutant forms of CbbR that were more highly active in transcription than the wild-type protein, or activated transcription under normally repressive growth conditions, were easily isolated (Table 9.2). Most interestingly, some mutant *Rba. sphaeroides* CbbR proteins either responded or did not respond to effectors (RuBP) *in vivo* and *in vitro*, and some mutant proteins were poorly able to activate transcription. The behavior of these proteins in a variety of gel-mobility shift assays and DNaseI footprint experiments indicated that constitutive function and RuBP binding are not mutually exclusive for the mutant proteins (Dangel et al., 2005). These studies certainly suggest that more detailed structure-function studies for CbbR are necessary. In any case, for the first time it was possible to map several specific single residues on the protein that are important for CbbR specificity, activity, and interaction with effectors and residues that appeared to act synergistically were also noted (see Fig. 9.6).

Table 9.2. cbb_I promoter and Rubisco activities in *Rba. sphaeroides* strain 87 containing constitutively active CbbR proteins

	Chemoheterotrophic		Photoheterotophic	
CbbR	β-gal	RubisCO	β-gal	RubisCO
No CbbR	0	0	0.6	3
WT CbbR	0	0	6	22
G99R	51	40	34	95
R136C	103	78	354	96
R155H	110	60	314	128
P160L	41	27	79	82
P160S	75	46	306	109
V162M	6	7	42	42
E203K	8	8	33	41
V162M/E203K	45	26	80	102

All activities in nmol product/min/mg. β-galactosidase activities were derived from cbb_I – *lacZ* fusion. Examples were selected from Table 1 of Dangel et al. (2005)

c. CbbR May Form Complexes with Other Transcriptional Regulators

There are two separate two-component systems that regulate *cbb* gene transcription in NSP bacteria, including the Reg and CbbRRS systems of *Rhodobacter* and *Rhodopseudomonas*, respectively. The RegA/RegB two component system maintains global regulatory control over redox-affected operons in many NSP bacteria, especially during aerobic to anaerobic growth transitions (Dubbs and Tabita, 2004; Elsen et al., 2004). RegA from *Rba. sphaeroides* is thought to become activated to regulate certain promoters after phosphorylation of residue Asp-63, catalyzed by

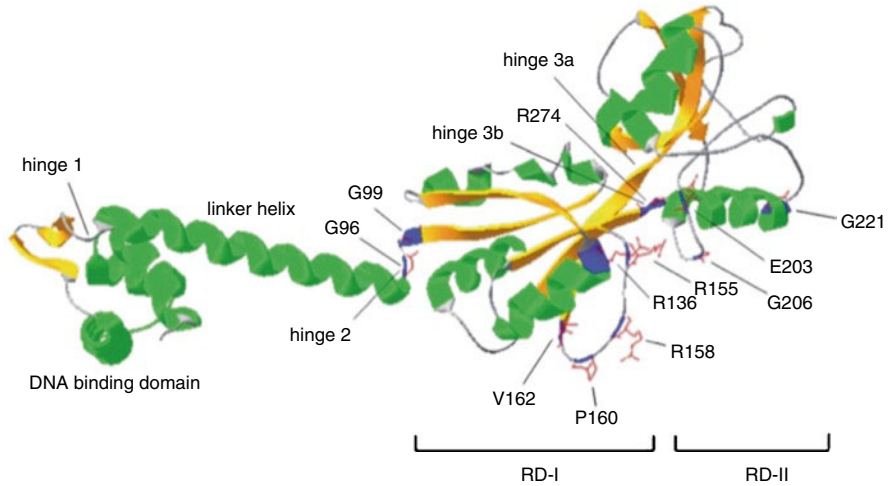


Fig. 9.6. Rba. sphaeroides CbbR structural model. The side chains of residues found to confer constitutive *cbb* transcription activity are indicated on the ribbon diagram of the peptide backbone (Dangel et al., 2005).

RegB (Inoue et al., 1995; Emmerich et al., 1999; Comolli et al., 2002). Among the operons under control of this two-component system are the cbb_I and cbb_{II} carbon dioxide fixation operons of *Rba. sphaeroides* (Qian and Tabita, 1996; Dubbs et al., 2000) and *Rba. capsulatus* (Vichivanives et al., 2000; Dubbs et al., 2004). Systems that are regulated by RegA and contain promoter loci with RegA binding sites include the *cbb*, nitrogen fixation (*nif*), photosystem biosynthesis, electron transport, and other energy-related operons (Joshi and Tabita, 1996; Dubbs and Tabita, 2004; Dangel and Tabita, 2009).

With respect to the cbb_I and cbb_{II} CO_2 fixation pathway operons of *Rba. sphaeroides*, it is intriguing that DNA binding sites for both CbbR and RegA are found within the regulatory regions of these operons and in some cases overlap (Dubbs and Tabita, 1998, 2003; Dubbs et al., 2000). For example, RegA binding site 1 overlaps the CbbR binding site just upstream of the transcription start site of the cbb_I operon in *Rba. sphaeroides*, as determined by prior footprint analyses. Protein-protein interactions between CbbR and RegA are thus plausible considering the proximity of the two proteins when bound to the cbb_I promoter. Indeed, in a recent study, it was demonstrated that CbbR and RegA interact and form a discrete complex in vitro, as illustrated by gel mobility shift experiments, direct isolation of the proteins from DNA complexes, and chemical cross-linking analyses (Dangel and Tabita, 2009). For CbbR/RegA interactions to occur, CbbR must be bound to the DNA, with the ability of CbbR to bind the cbb_I promoter enhanced by RegA. Conversely, RegA interactions with CbbR did not require RegA to bind the cbb_I promoter. From the available results, a model for CbbR and RegA complex formation on the cbb_I promoter was proposed, and the role of such interactions influencing transcription considered (Dangel and Tabita, 2009). This model includes a potential DNA loop intermediate, as well as a switch from RegA to RegA~P and an increase in RuBP concentration, signaling a change from chemoheterotrophic growth to autotrophic growth. RuBP is a necessary coinducer for CbbR to positively regulate the *cbb* operons (Smith and Tabita, 2002; Tichi and Tabita, 2002; Dubbs et al., 2004; Dangel et al., 2005), and as a global regulator, phophorylated RegA is necessary to activate transcription (Inoue et al., 1995; Emmerich et al., 1999; Comolli et al., 2002; Ranson-Olson et al., 2006). Since both the presence of RuBP and the phosphorylation of RegA are crucial for the transcription of cbb_I, it was postulated that the binding of RuBP to CbbR and the phosphorylation of RegA produces a conformational change to influence the CbbR/RegA interaction, subsequently allowing RNA polymerase to bind to the complex and initiate transcription. This model is eminently testable.

In *Rps. palustris*, the Reg system has no known role as mutations in RegR (equivalent to RegA) and RegS (equivalent to RegB) have no effect on photosystem biosynthesis, *nif* expression, or *cbb* transcription (Romagnoli and Tabita, 2009). However, the juxtaposition of the genes encoding the novel three protein CbbRRS two component system (*cbbRR1*, *cbbRR2*, and *cbbSR*) (Fig. 9.4c) between the *cbbR* and *cbbLS* genes warranted scrutiny. The *cbbRR1* and *cbbRR2* genes encode distinct response regulators, while *cbbRS* encodes a novel hybrid sensor kinase with potentially interesting PAS sensor domains (Romagnoli and Tabita, 2007). Thorough in vitro studies using recombinant wild-type and site-directed mutant constructs of the component proteins of the CbbRRS system were used to characterize the phosphotransfer pathway. These studies suggested a coherent model of signal transduction that took into account the phenotype of various mutant strains (Romagnoli and Tabita, 2007). Specifically, it was shown that the full-length sensor kinase was able to phosphorylate both response regulators, ruling out the necessity of a multi-step phosphorelay. Secondly, by utilizing two different-length truncated kinase constructs, it was demonstrated that the integrity of the N-terminal region of the sensor kinase, containing the first PAS motif, is important for discriminating which of the two response regulators might receive the phosphate from the sensor kinase. Finally, the receiver domain of the sensor kinase, along with the N-terminal region, were shown to contribute to the ATP affinity and stability of the sensor kinase and consequently the final signaling activity of the system.

In vivo, the CbbRRS system appears to be dispensable for photoautotrophic growth. However, deletions of single genes within the CbbRRS system resulted in the accumulation of less form I RubisCO and significantly lower total RubisCO activity; this effect was further amplified in the absence of *cbbM* (form II RubisCO). In fact,

strains deleted for components of the CbbRRS system in a *cbbM* background, showed between 10% and 25% of the wild-type RubisCO activity and displayed a significant growth phenotype (Romagnoli and Tabita, 2006). It was further found that cbb_{II} operon expression was constitutive in this organism (Romagnoli and Tabita, 2006; Joshi et al., 2009), and that benzoate-grown cells specifically induced the synthesis of form I RubisCO under conditions of CO_2 limitation (Joshi et al., 2009). These observations suggest that there might be some mechanism of redox regulation of form I RubisCO, requiring the CbbRRS system, that would allow *Rps. palustris* to overcome redox stress at low levels of carbon under photoautotrophic growth conditions or when the overall intracellular redox balance is changed by culturing cells with reduced organic carbon. It is intriguing that the components of the CbbRRS system, especially the response regulators, contain no apparent DNA binding domain, fueling speculation that this system somehow interacts with CbbR in order to influence *cbbLS* transcription. Studies using the bacterial two hybrid system (BacterioMatch® II Two-Hybrid System Vector Kit, Stratagene) have recently enabled the identification of a protein–protein interaction between the transcriptional regulator CbbR and CbbRR1, response regulator 1 of the CbbRRS system (Joshi et al., 2011). Further work identifying specific regions of the proteins critical for this interaction is in progress in order to understand the physiological role of the CbbRRS two-component system in vivo. Clearly, these latter studies lend credence to the concept that the response regulators might modify or somehow influence the interaction of the main transcriptional regulator protein, CbbR, with the *cbbLS* promoter.

III. The Reverse Tricarboxylic Acid (rTCA) Cycle

A. Importance and Distribution of the rTCA Cycle

The reductive TCA cycle is a reversal of the well-known oxidative TCA cycle that operates to serve both energy and biosynthetic demands in heterotrophic cells. Reversal of the oxidative cycle is accomplished by replacing large, membrane-bound, lipoic acid and thiamine pyrophosphate (TPP) dependent, pyruvate and α-ketoglutarate dehydrogenase complexes with relatively small, cytoplasmic, TPP-dependent, pyruvate and α-KG: ferredoxin (Fd) oxidoreductases (PFOR and KGOR, respectively). While pyruvate and α-KG dehydrogenases catalyze only the decarboxylation of their substrates, the Fd-linked oxidoreductases catalyze both oxidative decarboxylation and reductive carboxylation (CO_2 assimilatory or synthase) reactions. Thus, these enzymes are often referred to as pyruvate synthase and α-ketoglutarate synthase. The rTCA cycle is closed by the production of acetyl-CoA and oxaloacetate from citrate. Acetyl-CoA is the substrate for pyruvate synthase while oxaloacetate is converted to succinyl-CoA, the substrate of α-ketoglutarate synthase, by reactions common to the rTCA and TCA cycles to produce α-KG, which can be used for nitrogen assimilation or proceed to citrate acquiring an additional CO_2 by reversal of isocitrate dehydrogenase. Pyruvate can be recycled into the rTCA cycle via an additional CO_2 fixation step by PEP carboxylase to replenish α-KG lost to nitrogen assimilation via glutamate and glutamine synthesis.

As noted above, the rTCA cycle is considered to be one of the most ancient metabolic cycles and it has been proposed that the last universal common ancestor of all extant cellular life fixed CO_2 by this pathway (Zerkle et al., 2005). In addition, the rTCA cycle makes important contributions to primary productivity in modern hydrothermal systems like deep-sea hydrothermal vents and terrestrial hot springs. Primary productivity in these environments is driven mainly by chemolithoautotrophy and microbial communities of these ecosystems are commonly dominated by members of the *Epsilonproteobacteria* and often contain considerable numbers of *Aquificales*, a deeply branching bacterial lineage. Both of these groups are known to fix CO_2 by the rTCA cycle and subunits of ATP:citrate lyase have been increasingly utilized as molecular census targets to help measure the diversity of these microbial community assemblages in the environment (Campbell et al., 2006; Voordeckers et al., 2008) and in newly isolated cultures (Takai et al., 2005).

Catalysts of the rTCA cycle also have relevance to human disease as pathogenic members of the *Epsilonproteobacteria* (*Helicobacter pylori*) and certain amitochondriate eukaryotic pathogens (e.g. *Giardia*, *Entamoeba* and *Trichomonas*) all

contain genes with significant similarity to PFOR and KGOR, which may be of therapeutic use as humans lack these enzymes (Hughes et al., 1998; Srinivasan and Morowitz, 2006). In these cases, the enzymes are not proposed to function in CO_2 fixation, but rather as oxidative enzymes better suited to microaerophilic or anaerobic conditions where the reoxidation of NADH by typical NAD^+-linked dehydrogenases might be limited. Electron transfer from these enzymes via ferredoxin and/or flavodoxin is thought to be one explanation for the common susceptibility of these organisms to metronidazole (Kaihovaara et al., 1998).

B. Two Systems for Citrate Cleavage

A critical step in the rTCA cycle is the cleavage of citrate; this immediately provides acetyl-CoA for the synthesis of pyruvate by PFOR and oxaloacetate that is converted to succinyl-CoA, the substrate for α-ketoglutarate synthesis by KGOR. The enzyme catalyzing the cleavage of citrate was not elucidated at the time the rTCA cycle was first proposed in a green sulfur bacterium, but the need for a citrate cleavage system was evident (Evans et al., 1966). Subsequent studies confirmed that the activity was ATP:citrate lyase (Ivanovsky et al., 1980), but it was not until 31 years later that the enzyme was first purified to homogeneity from the green sulfur bacterium *Chlorobium tepidum* (Wahlund and Tabita, 1997) and in both this and a related organism (Kaneo et al., 2001), is found to consist of two subunits (Kim and Tabita, 2006) similar to the eukaryotic enzyme that is the first committed step in fatty acid biosynthesis. With the sequencing of the *C. tepidum* genome (Eisen et al., 2002), the two subunits of the ATP:citrate lyase (ACL) were found to be encoded by two linked genes, *aclBA*. Furthermore, both subunits are known to contribute to catalytic activity through their independent recombinant expression, and via reconstitution and site directed mutagenesis studies (Kim and Tabita, 2006).

Certain chemolithotrophic bacteria, like *Hydrogenobacter thermophilus (Aquificae)*, utilize a related system for citrate cleavage, a combination of citryl-CoA synthetase (CCS) and citryl-CoA lyase (CCL) whose activities can be separately purified and that are encoded by distinct genes (Aoshima et al., 2004a, b). The same protein domains are present in both ACL and CCS/CCL, and the overall reaction catalyzed by the two systems is identical. The CCS reaction requires two protein domains related to succinyl-CoA synthetase that reside on two polypeptides encoded by two genes while the CCL reaction requires one domain related to citrate synthase (Aoshima, 2007). Furthermore, the genes encoding the CCS and CCL enzymes in *Hydrogenobacter thermophilus (Aquificales)* are not co-localized on the genome and therefore occur on separate transcripts, which may therefore be differentially regulated depending on the needs of the cell. In the ACL of green sulfur bacteria, the CCL domain is fused with one of the CCS domains to form the large subunit of the enzyme, while the other CCS domain is carried by the small subunit of the enzyme. This explains in part why both large and small subunits of ACL are required to reconstitute any in vitro activities (Kim and Tabita, 2006).

Both ACL and the CCS/CCL system are now known to exist in relatively closely related members of the *Aquificae* (Hugler et al., 2007), an early diverging group within the bacteria as assessed by 16S rRNA phylogenies. Comparative sequence analyses of the genes in various strains was interpreted to suggest that ACL in the *Aquificae* was acquired by lateral gene transfer (Hugler et al., 2007). More experiments are required to determine what relative benefits the alternative citrate cleavage systems confer upon the organisms that utilize them. This will likely entail detailed catalytic comparisons of the enzymatic systems and gene swapping experiments where ATP:citrate lyase is replaced in an rTCA cycle organism with the genes encoding the CCS/CCL system to determine if the two systems are truly functionally interchangeable *in vivo*.

C. An rTCA Cycle in the Phototrophic Alphaproteobacteria?

As noted above, there is now functional evidence that at least one member of the *Alphaproteobacteria*, *Magnetococcus* sp. strain MC-1, utilizes the rTCA cycle for autotrophic CO_2 fixation. While evidence for PFOR and KGOR activity could be demonstrated in strain MC-1, ATP:citrate lyase activity was not found. In addition, PFOR was previously identified in *Rhodospirillum rubrum* as the NifJ protein even though a *nifJ* mutant was perfectly competent for nitrogen fixation under the

conditions initially tested (Lindblad et al., 1996). To support nitrogen fixation, PFOR would work primarily in the decarboxylating direction to produce reduced Fd as the electron donor to nitrogenase reductase. Subsequently, it was shown that the *R. rubrum* NifJ is likely important for electron transfer only under dark, anaerobic conditions where the other major electron donation pathway to nitrogenase reductase, the Fix system, cannot operate (Edgren and Nordlund, 2006). These results coupled with the observation of significant CO_2 fixation by a pathway other than the acetyl-CoA pathway in CBB cycle mutant strains (Wang et al., 1993) suggest that phototrophic *Alphaproteobacteria* may have a broader capacity for CO_2 fixation via the rTCA cycle than previously suspected.

To assess this possibility, BLASTP searches were conducted with relevant protein sequences for the key rTCA cycle enzymes against all phototrophic *Alphaproteobacteria* genomes available on the Integrated Microbial Genomes system (http://img.jgi.doe.gov). The results were mixed. Based on a comparison with *Magnetococcus* sp. strain MC-1, it appears that *R. rubrum* and several strains of *Rps. palustris* may possess the genetic potential for rTCA cycle activity. However, it is likely from the lack of clear homologs to CCL or ACL that an alternative citrate cleavage system may exist in these strains or at least the protein sequences in these organisms for the citrate cleavage system have diverged to the point that they are not easily recognized by BLAST searches. Additional experiments to test whether a functional rTCA cycle exists in these organisms should be performed. Studying the mechanisms of CO_2 fixation in CBB cycle deficient mutants by determining stable isotope labeling of lipids should be particularly valuable.

IV. The 3-Hydroxypropionate/ Malyl-CoA Pathway

A. Biochemistry of the Bicycle

Chloroflexus aurantiacus is a facultative autotrophic and moderate thermophilic filamentous green non-sulfur bacterium (Pierson and Castenholz, 1974). Although another member of the *Chloroflexi* class, *Oscillochloris trichoides*, uses the CBB cycle for autotrophic CO_2 fixation (Ivanovsky et al., 1999; Berg et al., 2005), both the CBB and rTCA cycles were ruled out early on for *C. aurantiacus* (Holo and Sirevag, 1986). Analysis of the recently available complete genome sequence of *C. aurantiacus* strain J-10-fl (NCBI accession number NC_010175) confirmed the absence of genes encoding the key enzymes of the CBB cycle, RubisCO, and of the rTCA cycle, ATP:citrate lyase. 3-Hydroxypropionate was identified as an intermediate during autotrophic growth (Holo, 1989; Strauss et al., 1992) and the subsequently discovered CO_2 fixation pathway was named accordingly (Strauss and Fuchs, 1993; Herter et al., 2001).

The 3-hydroxypropionate/malyl-CoA bicycle, as studied for *C. aurantiacus*, is shown in Fig. 9.6. The primary carboxylating enzymes are acetyl-CoA and propionyl-CoA carboxylase. These are biotin- and ATP-dependent enzymes that catalyze the addition of a bicarbonate molecule to acyl-CoA substrates. Carboxylation of acetyl-CoA forms malonyl-CoA. Malonyl-CoA is also a precursor for fatty acid synthesis. During CO_2 fixation, the activated carboxyl group of malonyl-CoA is reduced to the hydroxygroup of 3-hydroxypropionate. This two-step reduction, first to malonate semialdehyde and then to 3-hydroxypropionate, is catalyzed by a single enzyme, malonyl-CoA reductase, and requires NADPH as the source for reducing equivalents (Hugler et al., 2002). Malonyl-CoA reductase (NCBI accession number AAS20429) is an unusual protein with only limited sequence identity (and only over two 150-amino acid-stretches of the 1220-amino acid protein) to short chain alcohol dehydrogenases related to FabG (3-ketoacyl-(acyl-carrier-protein) reductase). 3-Hydroxypropionate is further reduced to propionyl-CoA, a conversion that formally requires three enzymatic steps, but is catalyzed by a single fusion protein (Alber and Fuchs, 2002). Propionyl-CoA synthase (NCBI accession number AAL47820) contains three functional domains: an acyl-CoA synthetase domain responsible for the activation of 3-hydroxypropionate to its CoA-ester, an enoyl-CoA reductase domain eliminating water from 3-hydroxypropionyl-CoA thereby forming acrylyl-CoA, and an enoyl-CoA reductase domain reducing acrylyl-CoA to propionyl-CoA using NADPH as the source for reducing equivalents. The three-step reduction of the activated carboxy-group of malonyl-CoA to the methyl-group of propionyl-

CoA, therefore, requires only two enzymes in *C. aurantiacus*: malonyl-CoA reductase and propionyl-CoA synthase. Carboxylation of propionyl-CoA forms (*S*)-methylmalonyl-CoA. Carbon skeleton rearrangement yields succinyl-CoA. It is not clear, if a bifunctional acetyl-CoA/propionyl-CoA carboxylase or two separate and substrate-specific enzymes are involved. Starting from succinyl-CoA, acetyl-CoA as the initial CO_2-acceptor is regenerated to close the first cycle (Fig. 9.7, left) and glyoxylate is formed as the primary CO_2-fixation product. Succinyl-CoA: L-malate CoA transferase as well as enzymes of the citric acid cycle are involved in this reaction sequence (Friedmann et al., 2006).

In the second cycle (Fig. 9.7, right), glyoxylate and CO_2 is converted to pyruvate that enters the central carbon metabolism. Propionyl-CoA is formed from acetyl-CoA and CO_2 as in the first cycle (by acetyl-CoA carboxylase, malonyl-CoA reductase, and propionyl-CoA synthase) and condenses with glyoxylate to form β-methylmalyl-CoA. This reaction is catalyzed by an enzyme that has three functions in the 3-hydroxypropionate/malyl-CoA cycle: L-malyl-CoA/β-methylmalyl-CoA/L-citramalyl-CoA lyase (NCBI accession number ABY33428) catalyzes (i) the reversible cleavage of L-malyl-CoA to glyoxylate and acetyl-CoA in the first cycle, (ii) the reversible condensation of glyoxylate and propionyl-CoA to β-methylmalyl-CoA, and (iii) the cleavage of L-citramalyl-CoA to pyruvate and acetyl-CoA in the second cycle (Herter et al., 2002). The enzyme is homologous to the β-subunit of citrate lyase an enzyme involved in citrate metabolism (Bott and Dimroth, 1994), but is not homologous to ATP citrate lyase or citryl-CoA lyase of the rTCA cycle (see Section III). The former belongs to the malate synthase class and the latter to the citrate synthase class of enzymes (Friedmann et al., 2007). The conversion of β-methylmalyl-CoA to L-citramalyl-CoA requires enzymes that are also unique (Zarzycki et al., 2009, Fig. 9.7), except for mesaconyl-CoA hydratase (GenBank accession number YP_001633816) which catalyzes the reversible dehydration of β-methylmalyl-CoA to mesaconyl-CoA (2-methylfumaryl-CoA) and is also involved in the ethylmalonyl-CoA pathway for acetyl-CoA assimilation (Zarzycki et al., 2008, Section V).

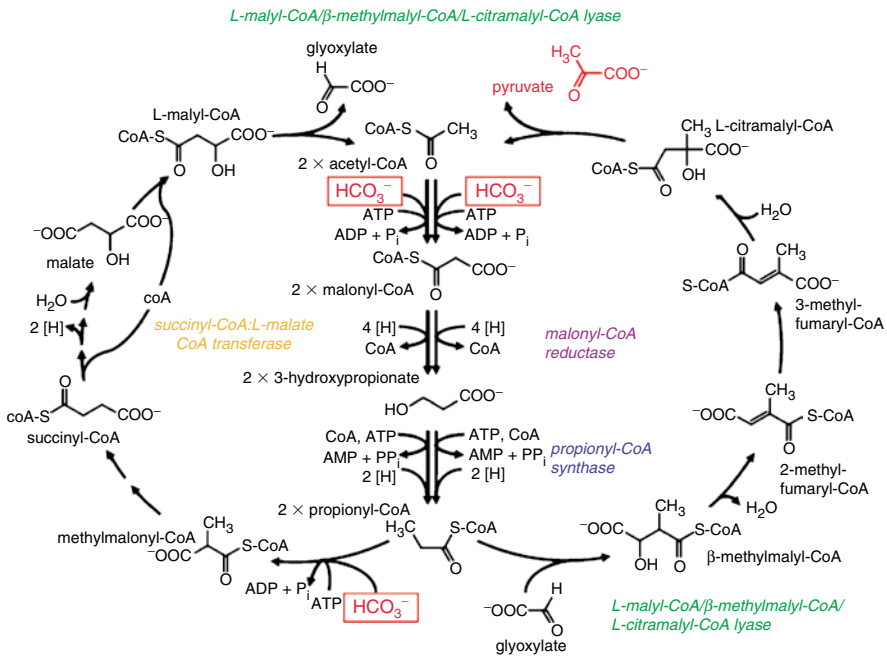


Fig. 9.7. The 3-hydroxypropionate/malyl-CoA bicycle in *Chloroflexus aurantiacus*. The key enzymes of the pathway are indicated: malonyl-CoA reductase (NCBI accession number: AAS20429), propionyl-CoA synthase (AAL47820), succinyl-CoA:L-malate CoA transferase (ABF14399, ABF14400), and L-malyl-CoA/β-methylmalyl-CoA/L-citramalyl-CoA lyase (ABY33428). The 3-hydroxypropionate/malyl-CoA bicycle for autotrophic growth converts three bicarbonate molecules into one molecule of pyruvate.

B. Comparative Genomics

Although, the reduction of malonyl-CoA to propionyl-CoA is a characteristic reaction sequence of the 3-hydroxypropionate/malyl-CoA cycle, it is shared by the 3-hydroxypropionate/4-hydroxybutyrate cycle (Menendez et al., 1999). This fifth pathway for autotrophic CO_2 fixation (Thauer, 2007), found in some archaea, differs from the 3-hydroxypropionate/malyl-CoA cycle in the way the CO_2 acceptor molecule acetyl-CoA is regenerated from succinyl-CoA (Berg et al., 2007). 4-Hydroxybutyrate is an intermediate and acetyl-CoA instead of glyoxylate is formed as the primary CO_2 fixation product (Berg et al., 2007). Interestingly, the proteins involved in the reductive conversion of malonyl-CoA to propionyl-CoA of the two pathways share no significant sequence similarity. This suggests convergent evolution of the two pathways. In the 3-hydroxypropionate/4-hydroxypropionate cycle as studied for *Metallosphaera sedula*, malonyl-CoA is reduced to 3-hydroxypropionate by two enzymes: malonyl-CoA reductase, a protein homologous to aspartate semialdehyde dehydrogenase, and succinate/malonate semialdehyde dehydrogenase (Alber et al., 2006a; Berg et al., 2007). The subsequent conversion of 3-hydroxypropionate to propionyl-CoA is catalyzed by three enzymes (Alber et al., 2008). Among them, acrylyl-CoA reductase of *M. sedula* is non-homologous to the enoyl-CoA reductase domain of propionyl-CoA synthase from *C. aurantiacus* (Teufel et al., 2009).

Genes encoding all key enzymes of the 3-hydroxypropionate/malyl-CoA cycle, propionyl-CoA synthase (*pcs*), malonyl-CoA reductase (*mcr*), and L-malyl-CoA/β-methylmalyl-CoA/L-citramalyl-CoA lyase (*mcl*) are present in the genomes of *C. aurantiacus* strain J-10-fl, *C. aurantiacus* strain Y-400-fl (NCBI accession number NZ_ABPG00000000), *C. aggregans* (NZ_AAUI00000000), *Roseiflexus* sp. RS-1 (NC_009523), and *R. castenholzii* (NC_009767). In case of the complete genome sequences it is apparent that *mcl* is always clustered with genes for mesaconyl-CoA hydratase (*mch*), succinyl-CoA:L-malate CoA transferase (*smtAB*), and a related enzyme that catalyzes the CoA transfer from mesaconyl-CoA (2-methylfumaryl-CoA) to 3-methylfumaryl-CoA (Zarzycki et al., 2009, Fig. 9.8). For the two *Roseiflexus* species the genes for propionyl-CoA synthase, malonyl-CoA reductase and an acyl-CoA carboxylase are also clustered; this is not the case for the *Chloroflexus* species (Klatt et al., 2007). The finding that *Roseiflexus* species apparently contain a functional 3-hydroxypropionate/malyl-CoA cycle was consistent with the prediction that autotrophic

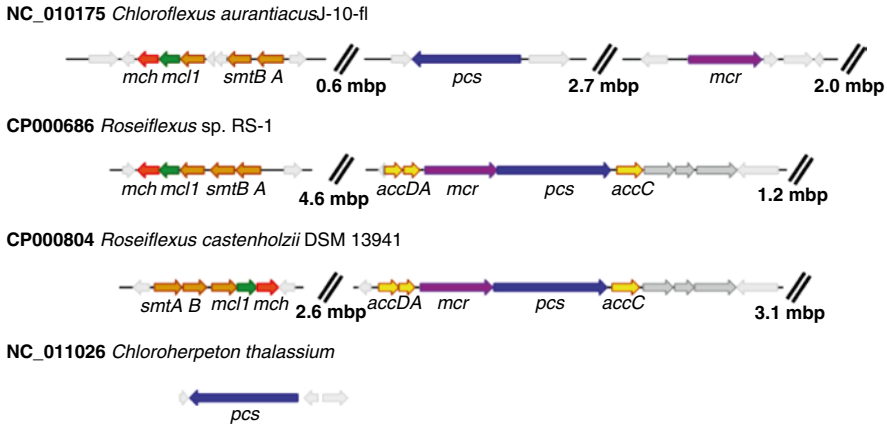


Fig. 9.8. Comparative genome analysis of key genes for autotrophic CO_2 fixation by the 3-hydroxypropionate/malyl-CoA cycle. Only complete genome sequences are considered. Propionyl-CoA is encoded by *pcs* and malonyl-CoA reductase by *mcr*. Genes encoding L-malyl-CoA/β-methylmalyl-CoA/L-citramalyl-CoA lyase (*mcl1*), mesaconyl-CoA hydratase (*mch*), and succinyl-CoA:L-malate CoA transferase (*smtAB*) are clustered. The gene upstream from *mcl* in ocher encodes a CoA transferase that converts 2-methylfumaryl-CoA to 3-methylfumaryl-CoA (Zarzycki et al., 2009). The genome of *C. thalassium* does not contain genes for the other key enzymes of the 3-hydroxypropionate/malyl-CoA cycle and autotrophic growth of the bacterium is assumed to occur by the rTCA cycle.

growth occurs in their natural habitat (van der Meer et al., 2000; Klatt et al., 2007), although these bacteria have not been grown with CO_2 as sole carbon source in pure culture.

Interestingly, the complete genome of *Chloroherpeton thalassium*, a green sulfur bacterium and member of the *Chlorobia* class, encodes a propionyl-CoA synthase-like protein (Fig. 9.8). This protein has 56% amino acid sequence identity to propionyl-CoA synthase from *C. aurantiacus* and it is, therefore, assumed that it catalyzes the same reaction. Genes for other key enzymes of the 3-hydroxypropionate/malyl-CoA cycle, however, are missing. *C. thalassium* is expected to fix CO_2 by the rTCA cycle, as are all other green sulfur bacteria (see Section III). The function of the propionyl-CoA synthase is, therefore, not clear.

C. Distribution of the 3-Hydroxypropionate/ Malyl-CoA Pathway

So far, the 3-hydroxypropionate/malyl-CoA cycle has only been described for bacteria that are not strict autotrophs, but have been shown to utilize a variety of organic carbon sources, including C_6-sugars, amino acids, and short-chain fatty acids anaerobically (photoheterotrophic growth) and aerobically (Madigan et al., 1974; Pierson and Castenholz, 1974; Krasilnikova et al., 1986; Hanada et al., 2002). In their natural habitat, they grow on metabolic products by other organisms present in the same habitat, yet stable carbon isotope signatures in lipid biomarkers also point to autotrophic growth (Bauld and Brock, 1973; Anderson et al., 1987; van der Meer et al., 2000). It is, therefore, feasible to suggest that reactions of the 3-hydroxypropionate/malyl-CoA cycle are also used for mixotrophic growth. Support comes from analysis of the almost complete genome of *Erythrobacter* sp. NAP-1 (NCBI accession number AAMW00000000), an aerobic anoxygenic phototroph that can co-assimilate CO_2 in the presence of an organic carbon source, but does not seem to grow autotrophically (Koblizek et al., 2003). Genes for malonyl-CoA reductase (ZP_01039179) and propionyl-CoA synthase (ZP_01039180) are present; however, enzymes involved in the regeneration of the acetyl-CoA acceptor molecule, only required for autotrophic growth, are not encoded by the genome.

Detection of genes encoding the key enzymes of the 3-hydroxypropionate/malyl-CoA cycle in metagenomic databases is somewhat hindered by the fact that both characteristic enzymes (malonyl-CoA reductase and propionyl-CoA synthase) are encoded by genes >3.5 kb, exceeding a typical DNA sequence read deposited in these databases.

V. The Ethylmalonyl CoA Pathway for Acetate Assimilation

A. Biochemistry of the Ethylmalonyl-CoA Pathway

Growth on organic compounds that enter the central carbon metabolism on the level of acetyl-CoA (such as acetate, fatty acids, C_1 compounds etc.) require a specialized pathway for its assimilation. In many organisms the glyoxylate cycle is used (Kornberg and Krebs, 1957). The ethylmalonyl-CoA pathway was discovered as an alternate route for the assimilation of acetyl-CoA in bacteria that lacked isocitrate lyase, a key enzyme of the glyoxylate cycle (Alber et al., 2006b; Erb et al., 2007). Overall, three molecules of acetyl-CoA, one molecule of CO_2, and one molecule of bicarbonate are converted to the two citric cycle intermediates, malate and succinyl-CoA (Fig. 9.9). The pathway has been studied in detail in *Rba. sphaeroides* (Alber et al., 2006b; Erb et al., 2007, 2008, 2009b, 2010; Zarzycki et al., 2008).

Initially two molecules of acetyl-CoA are condensed to form acetoacetyl-CoA that is reduced to (R)-3-hydroxybutyrate. This compound is the precursor for polyhydroxybutyrate, a carbon storage compound synthesized by *Rba. sphaeroides* (Kranz et al., 1997; Fales et al., 2001). During acetyl-CoA assimilation, (R)-3-hydroxybutyrate is dehydrated to crotonyl-CoA. It is not clear, if it is first epimerized to (S)-3-hydroxybutyrate or if a (R)-specific dehydratase is involved. In an unusual reaction, crotonyl-CoA is reductively carboxylated by crotonyl-CoA carboxylase/reductase (NCBI accession number ACJ71669; Erb et al., 2007, 2009a). The enantiomer of ethylmalonyl-CoA formed in this reaction is not the direct substrate for ethylmalonyl-CoA mutase. Instead (2S)-ethylmalonyl-CoA is converted to (2R)-ethylmalonyl-CoA by ethylmalonyl-CoA/methylmalonyl-CoA epimerase (NCBI accession

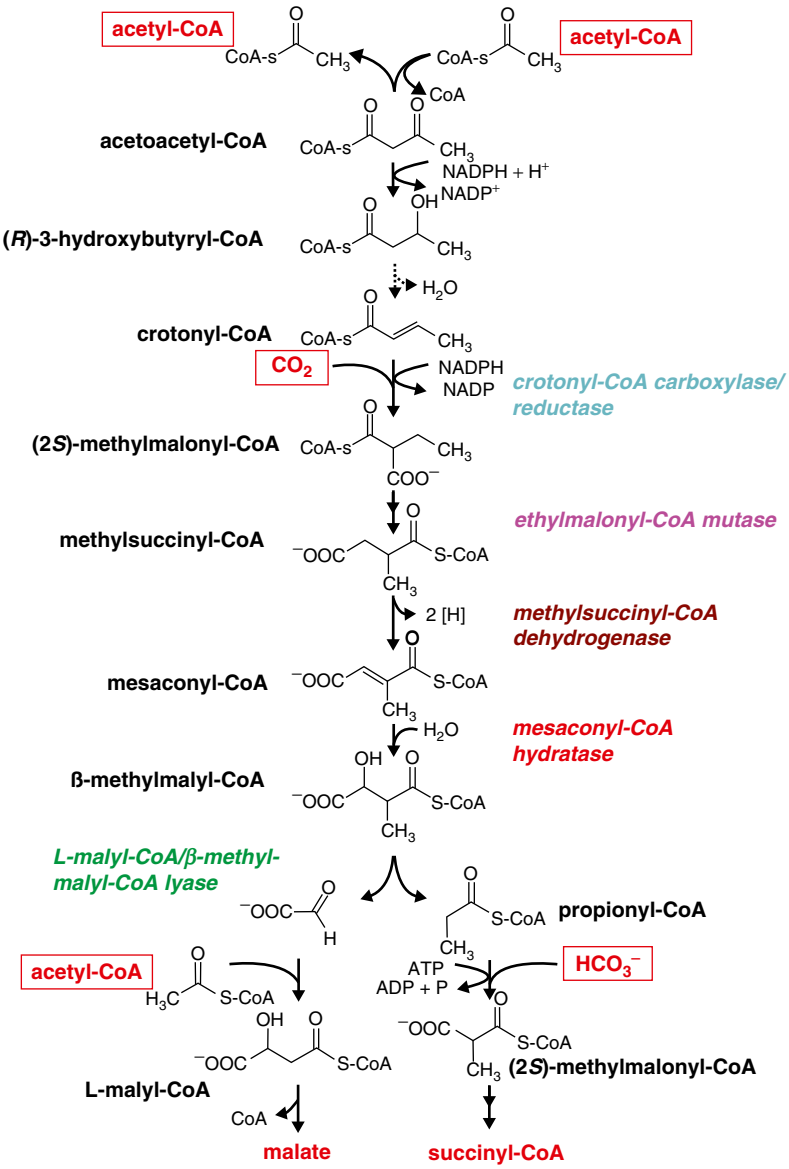


Fig. 9.9. The ethylmalonyl-CoA pathway in Rhodobacter sphaeroides. The key enzymes of the pathway are indicated: crotonyl-CoA carboxylase/reductase (NCBI accession number ACJ71669), ethylmalonyl-CoA mutase (ACJ71670), methylsuccinyl-CoA dehydrogenase, mesaconyl-CoA hydratase (ACJ71668), and L-malyl-CoA/β-methylmalyl-CoA lyase (ACJ71673). The ethylmalonyl-CoA pathway for acetyl-CoA assimilation converts three molecules of acetyl-CoA, one molecule of bicarbonate, and one molecule of CO_2 into one molecule of malate and one molecule of succinyl-CoA.

number ACJ71671). The enzyme has a second function in the ethylmalonyl-CoA pathway; it also catalyzes the interconversion of (2S)-methylmalonyl-CoA and (2R)-methylmalonyl-CoA (Erb et al., 2008). Carbon skeleton rearrangement of (2R)-ethylmalonyl-CoA to methylsuccinyl-CoA requires a specific ethylmalonyl-CoA mutase (NCBI accession number ACJ71670) that is distinct from methylmalonyl-CoA mutase needed later in the pathway (Erb et al., 2008). Methylsuccinyl-CoA is oxidized to mesaconyl-(C1)-CoA (2-methylfumaryl-CoA), the electron acceptor is unknown (Erb et al., 2009b). The following steps are common to the 3-hydroxypropionate/malyl-CoA cycle and homologous enzymes are involved: mesaconyl-CoA hydratase (NCBI

accession number ACJ71668) catalyzes the reversible dehydration of β-methylmalyl-CoA (Zarzycki et al., 2008) and L-malyl-CoA/β-methylmalyl-CoA lyase (NCBI accession number ACJ71673) catalyzes the reversible cleavage of β-methylmalonyl-CoA into glyoxylate and propionyl-CoA and the reversible condensation of glyoxylate and acetyl-CoA (Herter et al., 2002; Meister et al., 2005). L-Malyl-CoA is hydrolyzed to malate by malyl-CoA thioesterase (Erb et al., 2010). Propionyl-CoA that was formed from the cleavage of β-methylmalyl-CoA is assimilated by carboxylation to succinyl-CoA via methylmalonyl-CoA.

B. Comparative Genomics

All key enzymes of the ethylmalonyl-CoA pathway, crotonyl-CoA carboxylase/reductase, ethylmalonyl-CoA mutase, methylsuccinyl-CoA dehydrogenase, mesaconyl-CoA hydratase, and L-malyl-CoA/β-methylmalyl-CoA lyase are members of large enzyme families spread across all three domains of life. Various members of these families catalyze mechanistically similar reactions; however, they use different substrates. For all key enzymes of the ethylmalonyl-CoA pathway, the sequence distance relationship within a family reveals clustering of proteins based on function/substrate specificity rather than phylogenetic positioning (Meister et al., 2005; Erb et al., 2007, 2008). For example, ethylmalonyl-CoA mutase from *Rba. sphaeroides* clusters with the corresponding enzyme from *Streptomyces collinus* rather than with its own methylmalonyl-CoA mutase; both are members of the B_{12}-dependent acyl-CoA mutases (Erb et al., 2008). Careful sequence analysis, together with detection of gene-candidates for all enzymes of the ethylmalonyl-CoA pathway makes it straightforward to predict the presence of the pathway in organisms for which the genome sequence is known (Fig. 9.10). Among photosynthetic organisms, the ethylmalonyl-CoA pathway is found in some purple bacteria: genes for the ethylmalonyl-CoA pathway are present on chromosome 1 of all three sequenced strains of *Rba. sphaeroides* (NC_007493, NC_009049, NC_009428), *R. rubrum* (NC_007643). However, these genes are not found in strains of *Rps. palustris* even though all of these organisms belong to the order *Alphaproteobacteria*, though they reside in different classes (*Rhodobacterales*, *Rhodospirillales* and *Rhizobiales*). In addition, genes for the ethylmalonyl-CoA pathway are present in the

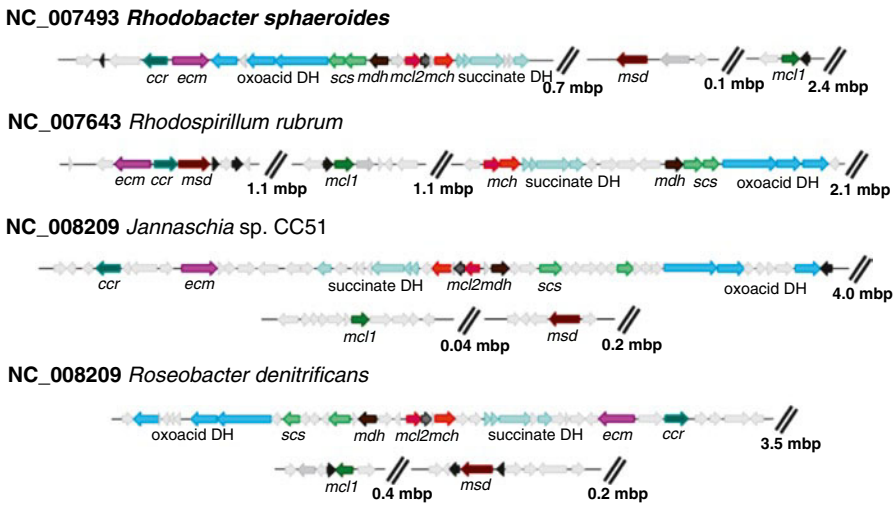


Fig. 9.10. Comparative genome analysis of key enzymes for acetyl-CoA assimilation by the ethylmalonyl-CoA pathway. Only complete genome sequences of phototrophic bacteria are considered. Crotonyl-CoA carboxylase/reductase is encoded by *ccr* and ethylmalonyl-CoA mutase by *ecm* and are usually found clustered. Methylsuccinyl-CoA dehydrogenase is encoded by *msd*, mesaconyl-CoA hydratase by *mch* and L-malyl-CoA/β-methylmalyl-CoA lyase by *mcl1*. Note, that *mch* and *mcl1* encode homologs of enzymes required for the 3-hydroxypropionate/malyl-CoA bicycle where they catalyze the same reactions (Figs. 9.7 and 9.8). Enzymes of the citric acid cycle are both subunits of succinyl-CoA synthetase (*scs*), α-ketoglutarate dehydrogenase (oxoacid DH), malate dehydrogenase (*mdh*), and succinate dehydrogenase (succinate DH).

complete genomes of anoxygenic aerobic phototrophs: *Roseobacter denitrificans* (NC_008209) and *Jannaschia* sp. CC51 (NC_007802).

C. Environmental Databases

The inferred presence of the ethylmalonyl-CoA pathway based on genomic analysis in members of the *Roseobacter* clade is reflected in the huge abundance of sequences corresponding to genes of the ethylmalonyl-CoA pathway in the "Global Ocean Sampling" database (Yooseph et al., 2007). Bacteria of this clade comprise up to 20% of marine coastal bacterioplankton communities (Gonzalez and Moran, 1997; Buchan et al., 2005). Among them are anoxygenic phototrophs that are strict aerobes and require organic molecules for growth, however light-stimulated uptake of CO_2 has been reported (Kolber et al., 2001). Because no autotrophic pathway could be inferred from the genomic sequence, an alternate route for co-assimilation of CO_2 during mixotrophic growth has been proposed (Swingley et al., 2007). Non-phototrophic members of this clade, e.g. *Silicibacter* sp. TM1040 (NC_008044), also have the genetic makeup for the ethylmalonyl-CoA pathway. There are several ways the ethylmalonyl-CoA pathway may participate in carbon flow in these organisms: for assimilation of compounds entering the central carbon metabolism on the level of acetyl-CoA (e.g. methylotrophic substrates) or for mixotrophic growth with organic compounds and CO_2 or CO.

VI. Conclusions and Perspective

The goal of this chapter was to present the latest insights into phototrophic CO_2 fixation, a demonstrably ancient set of metabolic pathways. While the data presented have relied on comparative genomics, we would like to emphasize our viewpoint that comparative genomics is best used for the generation of new hypotheses that must be subsequently subjected to metabolic, biochemical, and genetic scrutiny before being accepted or rejected. The case for the existence of RLPs and their functional divergence from RubisCO, while strongly suggested by comparative genomic data, have only recently become clearly established with demonstrated functions for a number of RLPs (Imker et al., 2007, 2008). Furthermore, hypothesis construction from comparative genomic data must be done cautiously. In the GOS data set, it was claimed that a wide diversity of novel RubisCO lineages had been discovered (Yooseph et al., 2007). However, a more careful and critical analysis of the same dataset suggested that only one new lineage of RLPs was present in that dataset (Tabita et al., 2008a). The main challenge to move our understanding of carbon fixation ahead in the post-genomic age will be the experimental assessment of metabolic functions suggested by rapidly growing metagenomic databases where very few representative organisms have been cultured from the sequenced community, This will require not only the continued development and application of stable isotope tools to analyze carbon flow in the environment and the expression of candidate genes in a suitable host, but a reinvestment of effort in the culturing of truly representative microorganisms from the environment for detailed studies in the laboratory to understand their physiology and metabolism as integrated organisms, not just collections of genes.

Acknowledgments

TEH was supported by a CAREER award from the National Science Foundation (MCB-0447649). BEA would like to thank G. Fuchs (Albert-Ludwigs Universität Freiburg) for continuous support. Research on the ethylmalonyl-CoA pathway was supported by DFG grant AL677/1-1 and NSF grant MCB-0842892. FRT was supported by DOE grants DE-FG02-07ER64489 and DE-FG02-08ER15976.

References

Akakura R and Winans SC (2002) Constitutive mutations of the OccR regulatory protein affect DNA bending in response to metabolites released from plant tumors. J Biol Chem 277: 5866–5874

Alber BE and Fuchs G (2002) Propionyl-coenzyme A synthase from *Chloroflexus aurantiacus*, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO_2 fixation. J Biol Chem 277: 12137–12143

Alber B, Olinger M, Rieder A, Kockelkorn D, Jobst B, Hügler M and Fuchs G. (2006a) Malonyl-coenzyme A reductase in the modified 3-hydroxypropionate cycle for

autotrophic carbon fixation in archaeal *Metallosphaera* and *Sulfolobus* spp. J. Bacteriol 188:8551–8559.

Alber BE, Spanheimer R, Ebenau-Jehle C and Fuchs G (2006b) Study of an alternate glyoxylate cycle for acetate assimilation by *Rhodobacter sphaeroide*s. Mol Microbiol 61: 297–309

Alber BE, Kung JW and Fuchs G (2008) 3-Hydroxypropionyl-coenzyme A synthetase from *Metallosphaera sedula*, an enzyme involved in autotrophic CO_2 fixation. J Bacteriol 190: 1383–1389

Anderson SL and McIntosh L (1991) Light-activated heterotrophic growth of the cyanobacterium *Synechocystis* sp. strain PCC 6803: a blue-light-requiring process. J Bacteriol 173: 2761–2767

Anderson KL, Tayne TA and Ward DM (1987) Formation and fate of fermentation products in hot spring cyanobacterial mats. Appl Environ Microbiol 53: 2343–2352

Aoshima M (2007) Novel enzyme reactions related to the tricarboxylic acid cycle: Phylogenetic/functional implications and biotechnological applications. Appl Microbiol Biotechnol 75: 249–255

Aoshima M, Ishii M and Igarashi Y (2004a) A novel enzyme, citryl-CoA lyase, catalysing the second step of the citrate cleavage reaction in *Hydrogenobacter thermophilus* TK-6. Mol Microbiol 52: 763–770

Aoshima M, Ishii M and Igarashi Y (2004b) A novel enzyme, citryl-CoA synthetase, catalysing the first step of the citrate cleavage reaction in *Hydrogenobacter thermophilus* TK-6. Mol Microbiol 52: 751–761

Ashida H, Saito Y, Kojima C, Kobayashi K, Ogasawara N and Yokota A (2003) A functional link between RubisCO-like protein of *Bacillus* and photosynthetic RubisCO. Science 302: 286–290

Bauld J and Brock TD (1973) Ecological studies of *Chloroflexus*, a gliding photosynthetic bacterium. Arch Mikrobiol 92: 267–284

Berg IA, Keppen OI, Krasil'nikova EN, Ugol'kova NV and Ivanovsky RN (2005) Carbon metabolism of filamentous anoxygenic phototrophic bacteria of the family *Oscillochloridaceae*. Microbiology 74: 258–264

Berg IA, Kockelkorn D, Buckel W and Fuchs G (2007) A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. Science 318: 1782–1786

Bott M and Dimroth P (1994) *Klebsiella pneumoniae* genes for citrate lyase and citrate lyase ligase: Localization, sequencing, and expression. Mol Microbiol 14: 347–356

Buchan A, Gonzalez JM and Moran MA (2005) Overview of the marine *Roseobacter* lineage. Appl Environ Microbiol 71: 5665–5677

Buick R (2008) When did oxygenic photosynthesis evolve? Philos Trans R Soc Lond B Biol Sci 363: 2731–2743

Campbell BJ, Engel AS, Porter ML and Takai K (2006) The versatile epsilon-proteobacteria: key players in sulphidic habitats. Nat Rev Microbiol 4: 458–468

Cleaves HJ, Chalmers JH, Lazcano A, Miller SL and Bada JL (2008) A reassessment of prebiotic organic synthesis in neutral planetary atmospheres. Orig Life Evol Biosph 38: 105–115

Comolli JC, Carl AJ, Hall C and Donohue T (2002) Transcriptional activation of the *Rhodobacter sphaeroides* cytochrome c_2 gene P2 promoter by the response regulator PrrA. J Bacteriol 184: 390–399

Cunchillos C and Lecointre G (2007) Ordering events of biochemical evolution. Biochimie 89: 555–573

Dangel AW and Tabita FR (2009) Protein-protein interactions between CbbR and RegA (PrrA), transcriptional regulators of the *cbb* operons of *Rhodobacter sphaeroides*. Mol Microbiol 71: 717–729

Dangel AW, Gibson JL, Janssen AP and Tabita FR (2005) Residues that influence in vivo and in vitro CbbR function in *Rhodobacter sphaeroides* and identification of a specific region critical for co-inducer recognition. Mol Microbiol 57: 1397–1414

Dubbs JM and Tabita FR (1998) Two functionally distinct regions upstream of the cbb_I operon of *Rhodobacter sphaeroides* regulate gene expression. J Bacteriol 180: 4903–4911

Dubbs JM and Tabita FR (2003) Interactions of the cbb_{II} promoter-operator region with CbbR and RegA (PrrA) regulators indicate distinct mechanisms to control expression of the two *cbb* operons of *Rhodobacter sphaeroides*. J Biol Chem 278: 16443–16450

Dubbs JM and Tabita FR (2004) Regulators of nonsulfur purple phototrophic bacteria and the interactive control of CO_2 assimilation, nitrogen fixation, hydrogen metabolism and energy generation. FEMS Microbiol Rev 28: 353–376

Dubbs JM, Bird TH, Bauer CE and Tabita FR (2000) Interaction of CbbR and RegA* transcription regulators with the *Rhodobacter sphaeroides* cbb_I promoter-operator region. J Biol Chem 275: 19224–19230

Dubbs P, Dubbs JM and Tabita FR (2004) Effector-mediated interaction of $CbbR_I$ and $CbbR_{II}$ regulators with target sequences in *Rhodobacter capsulatus*. J Bacteriol 186: 8026–8035

Edgren T and Nordlund S (2006) Two pathways of electron transport to nitrogenase in *Rhodospirillum rubrum*: The major pathway is dependent on the *fix* gene products. FEMS Microbiol Lett 260: 30–35

Eisen JA, Nelson KE, Paulsen IT, Heidelberg JF, Wu M, Dodson RJ, Deboy R, Gwinn ML, Nelson WC, Haft DH, Hickey EK, Peterson JD, Durkin AS, Kolonay JL, Yang F, Holt I, Umayam LA, Mason T, Brenner M, Shea TP, Parksey D, Nierman WC, Feldblyum TV, Hansen CL, Craven MB, Radune D, Vamathevan J, Khouri H, White O, Gruber TM, Ketchum KA, Venter JC, Tettelin H, Bryant DA and Fraser CM (2002) The complete genome sequence of *Chlorobium tepidum* TLS, a photosynthetic, anaerobic, green-sulfur bacterium. Proc Natl Acad Sci USA 99: 9509–9514

el-Refai AH, Nashat MA and Salem HM (1974) Heterotrophic growth of the nitrogen fixing blue-green alga *Nostoc muscorum*. Z Allg Mikrobiol 14: 297–302

Elsen S, Swem LR, Swem DL and Bauer CE (2004) RegB/RegA, a highly conserved redox-responding global two-component regulatory system. Microbiol Mol Biol Rev 68: 263–279

Emmerich R, Panglungtshang K, Strehler P, Hennecke H and Fischer HM (1999) Phosphorylation, dephosphorylation and DNA-binding of the *Bradyrhizobium japonicum* RegSR two-component regulatory proteins. Eur J Biochem 263: 455–463

Erb TJ, Berg IA, Brecht V, Müller M, Fuchs G and Alber BE (2007) Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: The ethylmalonyl-CoA pathway. Proc Natl Acad Sci USA 104: 10631–10636

Erb TJ, Retey J, Fuchs G and Alber BE (2008) Ethylmalonyl-CoA mutase from *Rhodobacter sphaeroides* defines a new subclade of coenzyme B_{12}-dependent acyl-CoA mutases. J Biol Chem 283: 32283–32293

Erb TJ, Brecht V, Fuchs G, Müller M and Alber BE (2009a) Carboxylation mechanism and stereochemistry of crotonyl-CoA carboxylase/reductase. Proc Natl Acad Sci USA 106: 8871–8876

Erb TJ, Fuchs G and Alber BE (2009b) (2S)-Methylsuccinyl-CoA dehydrogenase closes the ethylmalonyl-CoA pathway for acetyl-CoA assimilation. Mol Microbiol 73: 992–1008

Erb TJ, Frerichs-Revermann L, Fuchs G and Alber BE (2010) The apparent malate synthase activity of *Rhodobacter sphaeroides* is due to two paralogous enzymes, (3S)-malyl-coenzyme A (CoA)/β-methylmalyl-CoA lyase and (3S)-malyl-CoA thioesterase. J Bacteriol 192: 1249–1258

Evans MC, Buchanan BB and Arnon DI (1966) A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium. Proc Natl Acad Sci USA 55: 928–934

Ezezika OC, Haddad S, Clark TJ, Neidle EL and Momany C (2007) Distinct effector-binding sites enable synergistic transcriptional activation by BenM, a LysR-type regulator. J Mol Biol 367: 616–629

Falcone DL and Tabita FR (1991) Expression of endogenous and foreign ribulose 1,5-bisphosphate carboxylase-oxygenase (Rubisco) genes in a Rubisco deletion mutant of *Rhodobacter sphaeroides*. J Bacteriol 173: 2099–2108

Fales L, Kryszak L and Zeilstra-Ryalls J (2001) Control of *hemA* expression in *Rhodobacter sphaeroides* 2.4.1: Effect of a transposon insertion in the *hbdA* gene. J Bacteriol 183: 1568–1576

Feisthauer S, Wick LY, Kastner M, Kaschabek SR, Schlomann M and Richnow HH (2008) Differences of heterotrophic $13CO_2$ assimilation by *Pseudomonas knackmussii* strain B13 and *Rhodococcus opacus* 1CP and potential impact on biomarker stable isotope probing. Environ Microbiol 10: 1641–1651

Finn MW and Tabita FR (2003) Synthesis of catalytically active form III ribulose 1,5-bisphosphate carboxylase/oxygenase in archaea. J Bacteriol 185: 3049–3059

Friedmann S, Steindorf A, Alber BE and Fuchs G (2006) Properties of succinyl-coenzyme a:L-malate coenzyme a transferase and its role in the autotrophic 3-hydroxypropionate cycle of *Chloroflexus aurantiacus*. J Bacteriol 188: 2646–2655

Friedmann S, Alber BE and Fuchs G (2007) Properties of r-citramalyl-coenzyme a lyase and its role in the autotrophic 3-hydroxypropionate cycle of *Chloroflexus aurantiacus*. J Bacteriol 189: 2906–2914

Frigaard NU, Martinez A, Mincer TJ and DeLong EF (2006) Proteorhodopsin lateral gene transfer between marine planktonic bacteria and archaea. Nature 439: 847–850

Gibson JL (1995) Genetic analysis of CO_2 fixation genes. In: Blankenship RE, Madigan MT and Bauer CE (eds) Anoxygenic Photosynthetic Bacteria pp 1107–1124. Kluwer, Dordrecht

Gibson JL and Tabita FR (1993) Nucleotide sequence and functional analysis of *cbbR*, a positive regulator of the Calvin cycle operons of *Rhodobacter sphaeroides*. J Bacteriol 175: 5778–5784

Gibson JL and Tabita FR (1996) The molecular regulation of the reductive pentose phosphate pathway in proteobacteria and cyanobacteria. Arch Microbiol 166: 141–150

Gonzalez JM and Moran MA (1997) Numerical dominance of a group of marine bacteria in the alpha-subclass of the class *Proteobacteria* in coastal seawater. Appl Environ Microbiol 63: 4237–4242

Graves DA, Spradlin GM and Greenbaum E (1990) Effect of oxygen on photoautotrophic and heterotrophic growth of *Chlamydomonas reinhardtii* in an anoxic atmosphere. Photochem Photobiol 52: 585–590

Greene DN, Whitney SM and Matsumura I (2007) Artificially evolved *Synechococcus* PCC6301 Rubisco variants exhibit improvements in folding and catalytic efficiency. Biochem J 404: 517–524

Hanada S, Takaichi S, Matsuura K and Nakamura K (2002) *Roseiflexus castenholzii* gen. nov., sp. nov., a thermophilic, filamentous, photosynthetic bacterium that lacks chlorosomes. Int J Syst Evol Microbiol 52: 187–193

Hanson TE and Tabita FR (2001) A ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco)-like protein from *Chlorobium tepidum* that is involved with sulfur metabolism and the response to oxidative stress. Proc Natl Acad Sci USA 98: 4397–4402

Heinnickel M and Golbeck JH (2007) Heliobacterial photosynthesis. Photosynth Res 92: 35–53

Herter S, Farfsing J, Gad'On N, Rieder C, Eisenreich W, Bacher A and Fuchs G (2001) Autotrophic CO_2 fixation by *Chloroflexus aurantiacus*: Study of glyoxylate formation and assimilation via the 3-hydroxypropionate cycle. J Bacteriol 183: 4305–4316

Herter S, Busch A and Fuchs G (2002) L-malyl-coenzyme a lyase/beta-methylmalyl-coenzyme a lyase from *Chloroflexus aurantiacus*, a bifunctional enzyme involved in autotrophic CO_2 fixation. J Bacteriol 184: 5999–6006

Holo H (1989) *Chloroflexus aurantiacus* secretes 3-hydroxypropionate, a possible intermediate in the assimilation of CO_2 and acetate. Arch Microbiol 151: 252–256

Holo H and Sirevag R (1986) Autotrophic growth and CO_2 fixation of *Chloroflexus aurantiacus*. Arch Microbiol 145: 173–180

Hughes NJ, Clayton CL, Chalk PA and Kelly DJ (1998) *Helicobacter pylori porCDAB* and *oorDABC* genes encode distinct pyruvate:flavodoxin and 2-oxoglutarate: acceptor oxidoreductases which mediate electron transport to NADP. J Bacteriol 180: 1119–1128

Hugler M, Menendez C, Schagger H and Fuchs G (2002) Malonyl-coenzyme a reductase from *Chloroflexus aurantiacus*, a key enzyme of the 3-hydroxypropionate cycle for autotrophic fixation. J Bacteriol 184: 2404–2410

Hugler M, Huber H, Molyneaux SJ, Vetriani C and Sievert SM (2007) Autotrophic CO_2 fixation via the reductive tricarboxylic acid cycle in different lineages within the phylum *Aquificae*: Evidence for two ways of citrate cleavage. Environl Microbiol 9: 81–92

Imker HJ, Fedorov AA, Fedorov EV, Almo SC and Gerlt JA (2007) Mechanistic diversity in the RubisCO superfamily: The "enolase" In the methionine salvage pathway in *Geobacillus kaustophilus*. Biochemistry 46: 4077–4089

Imker HJ, Singh J, Warlick BP, Tabita FR and Gerlt JA (2008) Mechanistic diversity in the RubisCO superfamily: a novel isomerization reaction catalyzed by the RubisCO-like protein from *Rhodospirillum rubrum*. Biochemistry 47: 11171–11173

Inoue K, Kouadio JL, Mosley CS and Bauer CE (1995) Isolation and in vitro phosphorylation of sensory transduction components controlling anaerobic induction of light harvesting and reaction center gene expression in *Rhodobacter capsulatus*. Biochemistry 34: 391–396

Ivanovsky RN, Sintsov NV and Kondratieva EN (1980) ATP-linked citrate lyase activity in the green sulfur bacterium *Chlorobium limicola forma thiosulfatophilum*. Arch Microbiol 128: 239–241

Ivanovsky RN, Fal YI, Berg IA, Ugolkova NV, Krasilnikova EN, Keppen OI, Zakharchuc LM and Zyakun AM (1999) Evidence for the presence of the reductive pentose phosphate cycle in a filamentous anoxygenic photosynthetic bacterium, *Oscillochloris trichoides* strain DG-6. Microbiology 145: 1743–1748

Iverson TM (2006) Evolution and unique bioenergetic mechanisms in oxygenic photosynthesis. Curr Opin Chem Biol 10: 91–100

Johnson AP, Cleaves HJ, Dworkin JP, Glavin DP, Lazcano A and Bada JL (2008) The Miller volcanic spark discharge experiment. Science 322: 404

Joshi HM and Tabita FR (1996) A global two component signal transduction system that integrates the control of photosynthesis, carbon dioxide assimilation, and nitrogen fixation. Proc Natl Acad Sci USA 93: 14515–14520

Joshi HM and Tabita FR (2000) Induction of carbon monoxide dehydrogenase to facilitate redox balancing in a ribulose bisphosphate carboxylase/oxygenase-deficient mutant strain of *Rhodospirillum rubrum*. Arch Microbiol 173: 193–199

Joshi GS, Romagnoli S, Verberkmoes NC, Hettich RL, Pelletier D and Tabita FR (2009) Differential accumulation of form I RubisCO in *Rhodopseudomonas palustris* CGA010 under photoheterotrophic growth conditions with reduced carbon sources. J Bacteriol 191: 4243–4250

Joshi GS, Bobst CE and Tabita FR (2011) Unraveling the regulatory twist – regulation of CO_2 fixation in *Rhodopseudomonas palustris* CGA010 mediated by atypical response regulators. Mol Microbiol 80: 756–771

Kaihovaara P, Hook-Nikanne J, Uusi-Oukari M, Kosunen TU and Salaspuro M (1998) Flavodoxin-dependent pyruvate oxidation, acetate production and metronidazole reduction by *Helicobacter pylori*. J Antimicrob Chemother 41: 171–177

Kalapos MP (2007) The energetics of the reductive citric acid cycle in the pyrite-pulled surface metabolism in the early stage of evolution. J Theor Biol 248: 251–258

Kaneo T, Fukui T, Atomi H, Imanaka T (2001) ATP-citrate lyase from the green sulfur bacterium *Chlorobium limicola* is a heterodimeric enzyme composed of two distinct gene products. Eur J Biochem 268: 1670–1678

Kim W and Tabita FR (2006) Both subunits of ATP-citrate lyase from *Chlorobium tepidum* contribute to catalytic activity. J Bacteriol 188: 6544–6552

Klatt CG, Bryant DA and Ward DM (2007) Comparative genomics provides evidence for the 3-hydroxypropionate autotrophic pathway in filamentous anoxygenic phototrophic bacteria and in hot spring microbial mats. Environ Microbiol 9: 2067–2078

Koblizek M, Beja O, Bidigare RR, Christensen S, Benitez-Nelson B, Vetriani C, Kolber MK, Falkowski PG and Kolber ZS (2003) Isolation and characterization of *Erythrobacter* sp. strains from the upper ocean. Arch Microbiol 180: 327–338

Kolber ZS, Plumley FG, Lang AS, Beatty JT, Blankenship RE, VanDover CL, Vetriani C, Koblizek M, Rathgeber C and Falkowski PG (2001) Contribution of aerobic photoheterotrophic bacteria to the carbon cycle in the ocean. Science 292: 2492–2495

Kornberg HL and Krebs HA (1957) Synthesis of cell constituents from C2-units by a modified tricarboxylic acid cycle. Nature 179: 988–991

Kranz RG, Gabbert KK, Locke TA and Madigan MT (1997) Polyhydroxyalkanoate production in *Rhodobacter capsulatus*: Genes, mutants, expression, and physiology. Appl Environ Microbiol 63: 3003–3009

Krasilnikova EN, Keppen OI, Gorlenko VM and Kondrateva EN (1986) Growth of *Chloroflexus aurantiacus* on media with different organic-compounds and pathways of their metabolism. Microbiology 55: 325–329

Laguna R, Joshi GS, Dangel AW, Luther AK and Tabita FR (2010) Integrative control of carbon, nitrogen, hydrogen and sulfur metabolism: the central role of the Calvin-Benson-Bassham cycle. Adv Exp Med Biol 675: 265–271

Li H, Sawaya MR, Tabita FR and Eisenberg D (2005) Crystal structure of a Rubisco-like protein from the green sulfur bacterium *Chlorobium tepidum*. Structure 13: 779–789

Lindblad A, Jansson J, Brostedt E, Johansson M, Hellman U and Nordlund S (1996) Identification and sequence of a *nifJ*-like gene in *Rhodospirillum rubrum*: partial characterization of a mutant unaffected in nitrogen fixation. Mol Microbiol 20: 559–568

Lochowska A, Iwanicka-Nowicka R, Plochocka D and Hryniewicz MM (2001) Functional dissection of the LysR-type CysB transcriptional regulator. Regions important for DNA binding, inducer response, oligomerization, and positive control. J Biol Chem 276: 2098–2107

Lozier RH, Bogomolni RA and Stoeckenius W (1975) Bacteriorhodopsin: A Light-Driven Proton Pump in *Halobacterium halobium*. Biophys J 15: 955–962

Madigan MT and Gest H (1979) Growth of the photosynthetic bacterium *Rhodopseudomonas capsulata* chemoautotrophically in darkness with H_2 as the energy source. J Bacteriol 137: 524–530

Madigan MT, Petersen SR and Brock TD (1974) Nutritional studies on *Chloroflexus*, a filamentous photosynthetic, gliding bacterium. Arch Microbiol 100: 97–103

Martin W and Russell MJ (2007) On the origin of biochemistry at an alkaline hydrothermal vent. Philos Trans R Soc Lond B Biol Sci 362: 1887–1925

Meister M, Saum S, Alber BE and Fuchs G (2005) L-malyl-coenzyme a/beta-methylmalyl-coenzyme a lyase is involved in acetate assimilation of the isocitrate lyase-negative bacterium *Rhodobacter capsulatus*. J Bacteriol 187: 1415–1425

Menendez C, Bauer Z, Huber H, Gad'on N, Stetter KO and Fuchs G (1999) Presence of acetyl coenzyme a (CoA) carboxylase and propionyl-CoA carboxylase in autotrophic crenarchaeota and indication for operation of a 3-hydroxypropionate cycle in autotrophic carbon fixation. J Bacteriol 181: 1088–1098

Miller SL (1953) A production of amino acids under possible primitive earth conditions. Science 117: 528–529

Muraoka S, Okumura R, Ogawa N, Nonaka T, Miyashita K and Senda T (2003) Crystal structure of a full-length LysR-type transcriptional regulator, CbnR: unusual combination of two subunit forms and molecular bases for causing and changing DNA bend. J Mol Biol 328: 555–566

Olson JM (2006) Photosynthesis in the Archean era. Photosynth Res 88: 109–117

Olson JM and Blankenship RE (2004) Thinking about the evolution of photosynthesis. Photosynth Res 80: 373–386

Oparin AI (1950) [Noncellular form of life and the origin of cells.]. Vestn Akad Med Nauk SSSR 20: 118–122

Orgel LE (2008) The implausibility of metabolic cycles on the prebiotic earth. PLoS Biol 6: e18

Paoli GC and Tabita FR (1998) Aerobic chemolithoautotrophic growth and Rubisco function in *Rhodopseudomonas capsulata* and a spontaneous gain of function mutant of *Rhodobacter sphaeroides*. Arch Microbiol 170: 8–17

Paoli GC, Vichivanives P and Tabita FR (1998) Physiological control and regulation of the *Rhodobacter capsulatus cbb* operons. J Bacteriol 180: 4258–4269

Pierson BK and Castenholz RW (1974) A phototrophic gliding filamentous bacterium of hot springs, *Chloroflexus aurantiacus*, gen. and sp. nov. Arch Microbiol 100: 5–24

Qian Y and Tabita FR (1996) A global signal transduction system regulates aerobic and anaerobic CO_2 fixation in *Rhodobacter sphaeroides*. J Bacteriol 178: 12–18

Ranson-Olson B, Jones DF, Donohue TJ and Zeilstra-Ryalls JH (2006) In vitro and in vivo analysis of the role of PrrA in *Rhodobacter sphaeroides* 2.4.1 *hemA* gene expression. J Bacteriol 188: 3208–3218

Romagnoli S and Tabita FR (2006) A novel three-protein two-component system provides a regulatory twist on an established circuit to modulate expression of the cbb_1 region of *Rhodopseudomonas palustris* CGA010. J Bacteriol 188: 2780–2791

Romagnoli S and Tabita FR (2007) Phosphotransfer reactions of the CbbRRS three-protein two- component system from *Rhodopseudomonas palustris* CGA010 appear to be controlled by an internal molecular switch on the sensor kinase. J Bacteriol 189: 325–335

Romagnoli S and Tabita FR (2009) Carbon dioxide metabolism and its regulation in nonsulfur purple photosynthetic bacteria. In: Hunter CN, Daldal F, Thurnauer MC and Beatty JT (eds) The purple phototrophic bacteria, pp 563–576. Springer, Dordrecht, The Netherlands

Roslev P, Larsen MB, Jorgensen D and Hesselsoe M (2004) Use of heterotrophic CO_2 assimilation as a measure of metabolic activity in planktonic and sessile bacteria. J Microbiol Methods 59: 381–393

Rothschild LJ (2008) The evolution of photosynthesis…again? Philos Trans R Soc Lond B Biol Sci 363: 2787–2801

Satagopan S, Scott SS, Smith TG and Tabita FR (2009) A Rubisco mutant that confers growth under a normally "inhibitory" oxygen concentration. Biochemistry 48: 9076–9083

Smirnova IA, Dian C, Leonard GA, McSweeney S, Birse D and Brzezinski P (2004) Development of a bacterial biosensor for nitrotoluenes: The crystal structure of the transcriptional regulator DntR. J Mol Biol 340: 405–418

Smith SA and Tabita FR (2002) Up-regulated expression of the *cbb*(I) and *cbb*(II) operons during photoheterotrophic

growth of a ribulose 1,5-bisphosphate carboxylase-oxygenase deletion mutant of *Rhodobacter sphaeroides*. J Bacteriol 184: 6721–6724

Smith SA and Tabita FR (2003) Positive and negative selection of mutant forms of prokaryotic (cyanobacterial) ribulose-1,5-bisphosphate carboxylase/oxygenase. J Mol Biol 331: 557–569

Smith SA and Tabita FR (2004) Glycine 176 affects catalytic properties and stability of the *Synechococcus* sp. strain PCC6301 ribulose-1,5-bisphosphate carboxylase/oxygenase. J Biol Chem 279: 25632–25637

Srinivasan V and Morowitz HJ (2006) Ancient genes in contemporary persistent microbial pathogens. Biol Bull 210: 1–9

Stams AJ and Plugge CM (2009) Electron transfer in syntrophic communities of anaerobic bacteria and archaea. Nat Rev Microbiol 7: 568–577

Strauss G and Fuchs G (1993) Enzymes of a novel autotrophic CO_2 fixation pathway in the phototrophic bacterium *Chloroflexus aurantiacus*, the 3-hydroxypropionate cycle. Eur J Biochem 215: 633–643

Strauss G, Eisenreich W, Bacher A and Fuchs G (1992) 13C-NMR study of autotrophic CO_2 fixation pathways in the sulfur-reducing archaebacterium *Thermoproteus neutrophilus* and in the phototrophic eubacterium *Chloroflexus aurantiacus*. Eur J Biochem 205: 853–866

Summers ML, Wallis JG, Campbell EL and Meeks JC (1995) Genetic evidence of a major role for glucose-6-phosphate dehydrogenase in nitrogen fixation and dark growth of the cyanobacterium *Nostoc* sp. strain ATCC 29133. J Bacteriol 177: 6184–6194

Swingley WD, Sadekar S, Mastrian SD, Matthies HJ, Hao J, Ramos H, Acharya CR, Conrad AL, Taylor HL, Dejesa LC, Shah MK, O'Huallachain M E, Lince MT, Blankenship RE, Beatty JT and Touchman JW (2007) The complete genome sequence of *Roseobacter denitrificans* reveals a mixotrophic rather than photosynthetic metabolism. J Bacteriol 189: 683–690

Tabita FR, Hanson TE, Li H, Satagopan S, Singh J and Chan S (2007) Function, structure, and evolution of the RubisCO-like proteins and their RubisCO homologs. Microbiol Mol Biol Rev 71: 576–599

Tabita FR, Hanson TE, Satagopan S, Witte BH and Kreel NE (2008a) Phylogenetic and evolutionary relationships of RubisCO and the RubisCO-like proteins and the functional lessons provided by diverse molecular forms. Philos Trans R Soc Lond B Biol Sci 363: 2629–2640

Tabita FR, Satagopan S, Hanson TE, Kreel NE and Scott SS (2008b) Distinct form I, II, III, and IV RubisCO proteins from the three kingdoms of life provide clues about RubisCO evolution and structure/function relationships. J Exp Bot 59: 1515–1524

Takai K, Campbell BJ, Cary SC, Suzuki M, Oida H, Nunoura T, Hirayama H, Nakagawa S, Suzuki Y, Inagaki F and Horikoshi K (2005) Enzymatic and genetic characterization of carbon and energy metabolisms by deep-sea hydrothermal chemolithoautotrophic isolates of epsilonproteobacteria. Appl Environ Microbiol 71: 7310–7320

Teufel R, Kung JW, Kockelkorn D, Alber BE and Fuchs G (2009) 3-Hydroxypropionyl-coenzyme A dehydratase and acryloyl-coenzyme A reductase, enzymes of the autotrophic 3-hydroxypropionate/4-hydroxybutyrate cycle in Sulfolobales. J Bacteriol 191: 4572–4581

Thauer RK (2007) Microbiology. A fifth pathway of carbon fixation. Science 318: 1732–1733

Tichi MA and Tabita FR (2000) Maintenance and control of redox poise in *Rhodobacter capsulatus* strains deficient in the Calvin-Benson-Bassham pathway. Arch Microbiol 174: 322–333

Tichi MA and Tabita FR (2002) Metabolic signals that lead to control of *cbb* gene expression in *Rhodobacter capsulatus*. J Bacteriol 184: 1905–1915

van der Meer MT, Schouten S, de Leeuw JW and Ward DM (2000) Autotrophy of green non-sulphur bacteria in hot spring microbial mats: biological explanations for isotopically heavy organic carbon in the geological record. Environ Microbiol 2: 428–435

Vichivanives P, Bird TH, Bauer CE and Tabita FR (2000) Multiple regulators and their interactions in vivo and its in vitro with the *cbb* regulons of *Rhodobacter capsulatus*. J Mol Biol 300: 1079–1099

Voordeckers JW, Do MH, Hugler M, Ko V, Sievert SM and Vetriani C (2008) Culture dependent and independent analyses of 16S rRNA and ATP citrate lyase genes: a comparison of microbial communities from different black smoker chimneys on the mid-atlantic ridge. Extremophiles 12: 627–640

Wachtershauser G (1988) Before enzymes and templates: theory of surface metabolism. Microbiol Rev 52: 452–484

Wachtershauser G (1990) Evolution of the first metabolic cycles. Proc Natl Acad Sci USA 87: 200–204

Wahlund TM and Tabita FR (1997) The reductive tricarboxylic acid cycle of carbon dioxide assimilation: Initial studies and purification of ATP-citrate lyase from the green sulfur bacterium *Chlorobium tepidum*. J Bacteriol 179: 4859–4867.

Wang X, Falcone DL and Tabita FR (1993) Reductive pentose phosphate-independent CO_2 fixation in *Rhodobacter sphaeroides* and evidence that ribulose bisphosphate carboxylase/oxygenase activity serves to maintain the redox balance of the cell. J Bacteriol 175: 3372–3379

Williams TJ, Zhang CL, Scott JH and Bazylinski DA (2006) Evidence for autotrophy via the reverse tricarboxylic acid cycle in the marine magnetotactic coccus strain MC-1. Appl Environ Microbiol 72: 1322–1329

Xiong J and Bauer CE (2002) Complex evolution of photosynthesis. Annu Rev Plant Biol 53: 503–521

Yooseph S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, Remington K, Eisen JA, Heidelberg KB, Manning G, Li W, Jaroszewski L, Cieplak P, Miller CS, Li H, Mashiyama ST, Joachimiak MP, van Belle C, Chandonia

JM, Soergel DA, Zhai Y, Natarajan K, Lee S, Raphael BJ, Bafna V, Friedman R, Brenner SE, Godzik A, Eisenberg D, Dixon JE, Taylor SS, Strausberg RL, Frazier M and Venter JC (2007) The Sorcerer II global ocean sampling expedition: expanding the universe of protein families. PLoS Biol 5: e16

Yoshida S, Inui M, Yukawa H, Kanao T, Tomizawa K, Atomi H and Imanaka T (2006) Phototrophic growth of a Rubisco-deficient mesophilic purple nonsulfur bacterium harboring a type III Rubisco from a hyperthermophilic archaeon. J Biotechnol 124: 532–544

Zarzycki J, Schlichting A, Strychalsky N, Muller M, Alber BE and Fuchs G (2008) Mesaconyl-coenzyme a hydratase, a new enzyme of two central carbon metabolic pathways in bacteria. J Bacteriol 190: 1366–1374

Zarzycki J, Brecht V, Müller M and Fuchs G (2009) Identifying the missing steps of the autotrophic 3-hydroxypropionate CO_2 fixation cycle in *Chloroflexus aurantiacus.* Proc Natl Acad Sci USA 106: 21317–21322

Zerkle AL, House CH and Brantley SL (2005) Biogeochemical signatures through time as inferred from whole microbial genomes. Am J Sci 305: 467–502

Chapter 10

Evolutionary Relationships Among Antenna Proteins of Purple Phototrophic Bacteria

Andrew S. Lang
Department of Biology, Memorial University of Newfoundland, 232 Elizabeth Ave, St. John's, NL A1B 3X9, Canada

Caroline S. Harwood
Department of Microbiology, University of Washington, Box 357735, Seattle, WA 98195-7735, USA

J. Thomas Beatty*
Department of Microbiology and Immunology, University of British Columbia, 2350 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada

Summary

A core, or LH1, antenna complex is present in all purple phototrophic bacteria, whereas peripheral, or LH2, antenna complexes are found in the membranes of many, but not all purple phototrophic bacteria. The presence of an LH2 complex is thought to provide a phototrophic growth competitive advantage under low-intensity illumination. Some species have only one type of LH2 complex, whereas other species have two or more types of LH2 complex, giving rise to a variety of absorption spectrum profiles. The great diversity of LH2 genes, sometimes within a single bacterial strain, illustrates a gene family consisting of many individuals that have arisen over long evolutionary time scales. We use sequence alignments to infer evolutionary relationships of LH2 proteins, and compare phylogenetic trees to a 16S

*Author for correspondence, e-mail: jbeatty@interchange.ubc.ca

R.L. Burnap and W.F.J. Vermaas (eds.), Functional Genomics and Evolution of Photosynthetic Systems, Advances in Photosynthesis and Respiration 33, pp. 253–264, DOI 10.1007/978-94-007-1533-2_10,

rDNA-based tree to evaluate the relative extent of vertical and horizontal transmission of LH2 genes. A phylogenetic tree of LH1 and LH2 proteins indicates that they are derived from a common ancestor. The LH2 genes of *Phaeospirillum molischianum* appear to have evolved independently of the LH2 genes of other purple phototrophic bacteria.

I. Introduction

Much of what we know about light-harvesting (LH) complexes in purple phototrophic bacteria comes from studies of just a few species. Nevertheless, there are several themes that unify our understanding of the general properties of LH complexes. LH complexes in purple phototrophic bacteria may be subdivided into the core (LH1-containing) complex, and the more diverse peripheral (LH2) complexes (Gabrielsen et al., 2009). Some of the first gene cloning and sequencing work on *Rhodobacter capsulatus* revealed amino acid sequence identities of 27–34% between LH1 and LH2 proteins, indicating a common ancestor (Youvan and Ismail, 1985).

These LH complexes consist of repeating units comprised of small α and β proteins of 40–70 amino acids that are quite similar to each other in sequence. Each of these proteins has N- and C-terminal hydrophilic domains flanking a single membrane-spanning segment. The proteins bind bacteriochlorophyll (BChl) and carotenoid to absorb light energy. The LH1 complex closely encircles the photochemical reaction center (RC) (Qian et al., 2005), and in fact it is possible to isolate LH1/RC supercomplexes by solubilization of membrane proteins with detergents (Roszak et al., 2003; Abresch et al., 2005). The LH2 complexes are diverse in their absorption spectral profiles, and are thought to be located in membrane domains separate from but adjacent to the core complex.

Some species of purple bacteria, such as *Rhodospirillum rubrum*, lack an LH2 complex. At the other extreme, some *Rhodopseudomonas palustris* isolates have seven sets of LH2 genes – some of which clearly encode LH complexes that absorb light of wavelengths different from each other (Tharia et al., 1999; Braatsch et al., 2006; Oda et al., 2008). Thus there are paralogous LH2 genes in some genomes, but it is not known whether the multiple LH2 genes arose by lateral transfer between species or by gene duplication within a species. High-resolution structures based on X-ray diffraction analysis of crystals of LH2 complexes from three species have revealed many similarities (McDermott et al., 1995; Koepke et al., 1996), as have atomic force microscopy studies (Olsen et al., 2008; Scheuring, 2009). Although a limited number of gene expression experiments have been done so far, it is clear that LH2 gene sets within a strain can be differentially expressed depending on growth conditions (Evans et al., 2005; Braatsch et al., 2006; Rey et al., 2007; Jaubert et al., 2008). Additionally, absorption spectra of many cultivated strains indicate the presence of a variety of light absorption spectral profiles, thought to reflect differences in LH2 complexes (Gabrielsen et al., 2009). For example, *Rhodobacter* and *Rhodopseudomonas* species produce an LH2 that has absorption peaks at about 800 and 850 nm whereas, depending on growth conditions, *Rhodopseudomonas* species may produce an LH2 that absorbs maximally at 800 and 820 nm, or another LH2 that absorbs only at 800 nm (Hartigan et al., 2002; Gabrielsen et al., 2009).

This chapter is focused on the evolution of LH2 complexes, and the relationships of LH2 to LH1 in purple phototrophic bacteria. We begin by summarizing the structure and function of LH complexes in purple bacteria, and then move on to the distribution of LH2 complexes in evolutionary lines, based on a 16S rDNA phylogeny. We finish with several phylogenetic trees of LH proteins and an analysis to assess the relative likelihood of vertical or lateral transfer of the genes that encode LH2 proteins. It appears that instances of multiple LH2 genes within a given species arose predominantly by gene duplication before the divergence of different strains from each other. However there are exceptions. Likewise, although lateral transfer of LH2 genes appears to have been a relatively rare event, there is evidence that it has occurred. Lastly, sequences from a subset of LH1 and LH2 proteins are compared to reveal relationships between these two types of LH complex, and this

Abbreviations: BChl – bacteriochlorophyll; LH – light-harvesting; LH1 – light-harvesting complex 1; LH2 – light-harvesting complex 2; RC – reaction center

suggests an independent evolutionary origin for the *Phaeospirillum molischianum* LH2 proteins relative to other LH2 proteins.

II. Summary of Structural and Functional Properties of Light-Harvesting (LH) 2 Complexes

Light-harvesting complexes in purple phototrophic bacteria were first described on the basis of the wavelengths of light they absorb, in absorption spectra of cell membranes containing pigment-protein complexes. Therefore, a particular type of LH complex, such as the LH2 of *Rhodobacter capsulatus*, may be called the B800-850 complex to indicate absorption peaks at 800 and 850 nm. Alternatively, the practice has arisen to distinguish between two major types of LH complex called LH1 and LH2. LH1 encircles and is in direct contact with the photosynthetic reaction center (RC), and the LH1/RC together are often called the core complex of the photosystem. In contrast is the peripheral or LH2 type of complex (Gabrielsen et al., 2009).

LH2 (B800-850) complexes from *Rhodoblastus acidophilus* (formerly *Rhodopseudomonas acidophila*) and *Phaeospirillum molischianum* (formerly *Rhodospirillum molischianum*) were crystallized and the structures solved to fairly high resolution, revealing very similar structures (McDermott et al., 1995; Koepke et al., 1996). Both structures consist of multiple copies of an α and a β protein, each of which spans the membrane once, and with pairs of α/β proteins organized in a ring. The β proteins tend to be slightly longer than the α proteins, and the membrane-spanning region of both proteins has a conserved H (histidine) residue that binds the Mg^{2+} ion of BChl. The BChl and carotenoid pigments are more or less sandwiched between α and β protein transmembrane segments, with α proteins forming the inner wall and β proteins the outer wall of a cylinder. The *R. acidophilus* structure contains nine α/β subunits whereas the *P. molischianum* structure has eight, but the structures are otherwise fairly similar. Atomic force microscopy has yielded dramatic images of the core complex in cell membranes treated with detergent, and revealed that LH2 complexes are organized in multiple copies around core complex domains (Olsen et al., 2008; Scheuring, 2009).

It was discovered that some species of purple phototrophic bacteria produce one type of LH2 under certain cultural conditions, and a different type (on the basis of absorption spectrum) under other conditions. This has resulted in the coining of new names, such as LH3 for the *R. acidophilus* LH complex that absorbs at 800 and 820 nm, to distinguish it from the LH2 complex that absorbs at 800 and 850 nm. Similarly, a peripheral B800 antenna complex from *R. palustris* has been called LH4 (Evans et al., 2005). Crystal structures (McLuskey et al., 2001; Hartigan et al., 2002) show that these complexes share the general features (i.e., rings of α/β proteins sandwiching pigments) of B800-850 LH2 complexes, and the protein sequences are very similar (Tharia et al., 1999). We agree with Gabrielsen et al. (2009) that all of these complexes should generically be referred to as LH2, and the wavelength(s) of light absorbed given additionally if needed for clarity. All of these complexes bind exactly the same BChl molecule, and we show in detail below that the α and β proteins are evolutionarily related, and so these changes in protein sequence in duplicated LH2 genes presumably represent spectral tuning of LH complexes to provide an evolutionary adaptation that allows the cells to harvest light of diverse wavelengths.

LH2 complexes are thought to harvest light energy by delocalizing the energy over all of the BChl molecules in an individual complex as well as between adjacent complexes, with eventual transfer to LH1 and thence to the RC (Cogdell et al., 1999).

III. Species Distribution of LH2 Complexes

As of December 2008, we found that LH2 proteins are encoded in the genomes of twenty-one strains of alphaproteobacteria, two betaproteobacteria, and one gammaproteobacterium. Figure 10.1 shows a 16S rDNA tree of these organisms. There is a large number of purple bacteria that appear to produce an LH2 complex on the basis of absorption spectra, but unfortunately protein sequence information is limited to 23 of the 24 organisms listed in Fig. 10.1 (*R. rubrum* lacks an LH2 but is included in the 16S rDNA tree because an *R. rubrum* LH1 protein is in a tree comparing LH1 and LH2 β proteins).

IV. Evolutionary Trees of LH Proteins

As noted by Beverley Green and coworkers (Green et al., 2003) it is difficult to construct phylogenetic trees using short (single transmembrane) proteins because of "generic" hydropathic residues in the membrane-spanning region of the protein. However, the LH trees described below yield some interesting and significant information that we suggest improves the understanding of the evolutionary history of these short, single transmembrane LH proteins. Unfortunately, the trees of LH2 proteins generally have low statistical support values relative to the 16S rDNA tree (compare Fig. 10.1 to Figs. 10.2 and 10.3), making it difficult to be fully confident in some of the relationships suggested by the trees. Note that some proteins with the same amino acid sequence are encoded by more than one gene; for example, in *R. palustris* CGA009 the genes RPA1491 and RPA4291 encode LH2 β proteins with the same sequence, although the DNA sequences differ at four positions.

Our interpretations of LH protein trees follow. Because LH2 genes are typically found as an operon that contains a *pucB* (LH2 β) gene followed by a *pucA* (LH2 α) gene that evidently coevolved, similarities between the LH2 β and α trees gave us confidence that some surprising discoveries reflect evolutionary history as opposed to a treeing artifact. We found that the LH2 β proteins could be aligned with more ease and confidence, and that the LH2 β protein trees ended up with higher statistical support as reflected by higher bootstrap values. This is most likely because the β protein sequences have a slightly longer conserved sequence region than the α protein sequences. Therefore, we consider these sequences provide a better representation of evolutionary relationships than the α protein trees, and we defer to the relationships found in the β tree when there are differences between the β and α trees. Lastly, we conducted analyses to compare a subset of LH1 and LH2 β proteins, which indicate two independent evolutionary divergences of LH2 genes from LH1 genes.

A. LH2 β Proteins

With few exceptions, the LH2 β proteins yield a tree that is similar to the 16S rDNA tree (compare Figs. 10.1 and 10.2). Although there are low bootstrap and support values for some branches, we

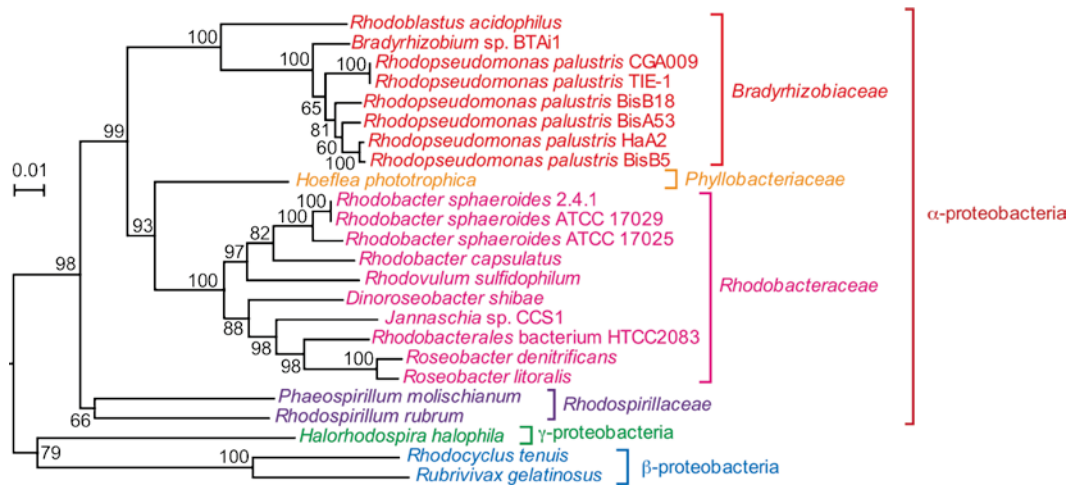


Fig. 10.1. Phylogenetic relationships of bacterial species and strains discussed in this chapter. Phylogenetic analyses based on 16S rDNA sequences were conducted using MEGA4 (Tamura et al., 2007). The evolutionary pattern was inferred using the Neighbor-Joining method (Saitou and Nei, 1987) and bootstrap values (Felsenstein, 1985) are shown as percentages based on 10,000 replicates. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al., 2004) and the scale bar indicates the number of base substitutions per site. The different taxonomic groups are indicated in different colors, with the taxonomic affiliations of the species shown beside the half-brackets. Sequences from species in the deltaproteobacteria were used as the outgroup (not shown).

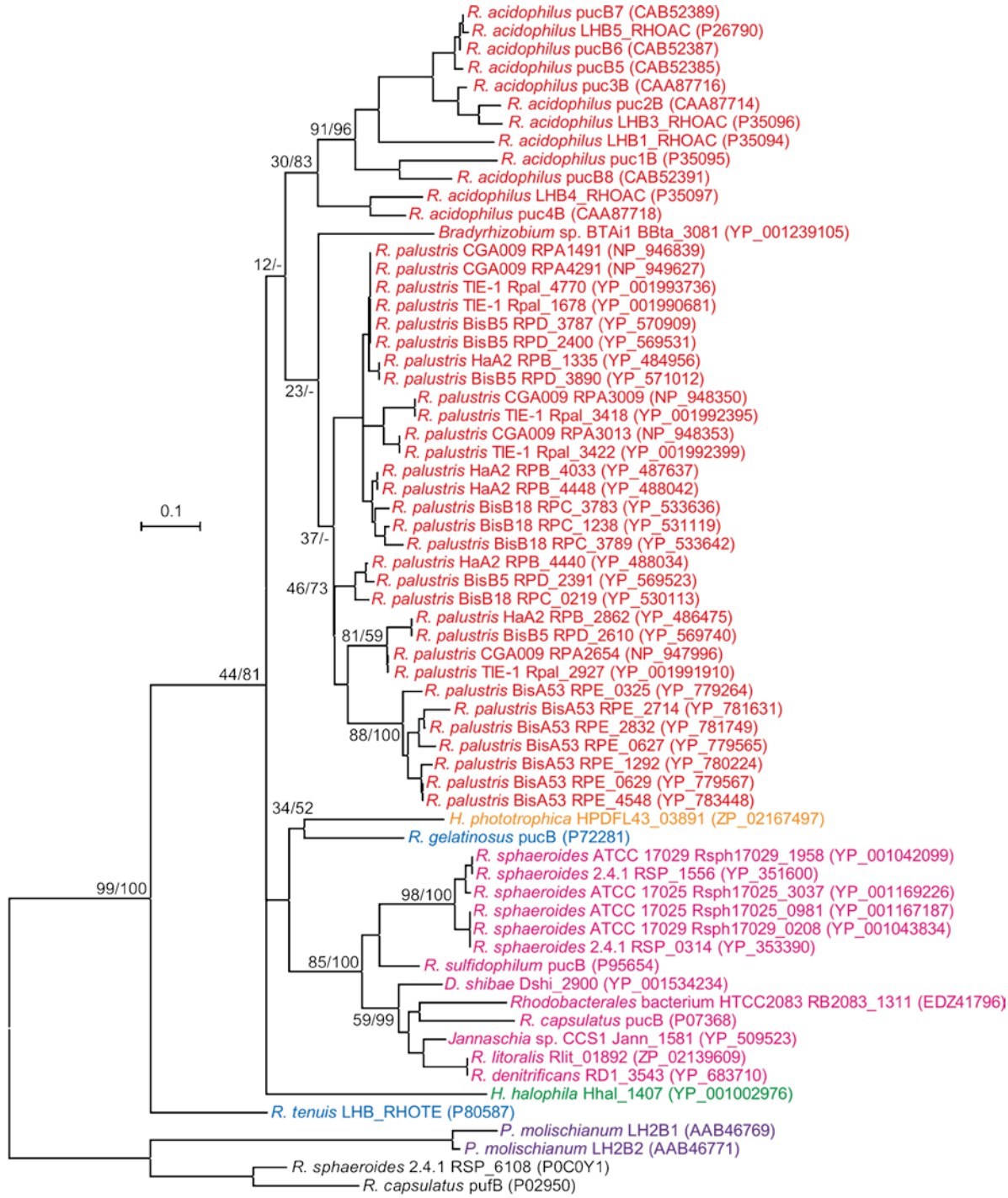


Fig. 10.2. Evolutionary relationships of LH2 β proteins. The protein identifications are colored corresponding to the taxonomic affiliations of the bacterial species as in the 16S rDNA-based phylogeny (Fig. 10.1). Proteins are identified by the species and strain (where appropriate) followed by the gene or locus designation from the GenBank submission, and the GenBank accession numbers are given in brackets. The evolutionary pattern was inferred using the Neighbor-Joining (Saitou and Nei, 1987) and Bayesian methods (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). The Neighbor-Joining bootstrap consensus tree inferred from 10,000 replicates (Felsenstein, 1985) using MEGA4 (Tamura et al., 2007) is shown. Bootstrap values (Felsenstein, 1985) are shown as percentages based on 10,000 replicates followed by Bayesian posterior probabilities (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) based on 300,000 generations for selected relevant nodes. A dash indicates the same cluster of sequences did not occur in the Bayesian consensus tree. The evolutionary distances were computed using the JTT matrix-based method (Jones et al., 1992) and the scale bar indicates the number of amino acid substitutions per site. LH1 β protein sequences were used as the outgroup.

suggest that the β protein tree reflects the genuine evolutionary relationship between LH2 genes because the LH2 β tree closely resembles the α tree (Fig. 10.3) in key instances.

In the alphaproteobacteria, all of the *R. acidophilus* β proteins form a distinct clade, and all of the *R. palustris* proteins appear to derive from a common ancestor after the divergence of the *Bradyrhizobium*/*R. palustris* lines from the *R. acidophilus* line because these sequences all fall within a single clade (Fig. 10.2). Many of the seven LH2 β sequences in *R. palustris* strain BisA53 are likely to represent gene duplications that occurred after this strain took a separate evolutionary path from the other *R. palustris* strains, because the strain BisA53 sequences group together whereas the other *R. palustris* strain sequences are intermingled. This differs from the LH2 α protein tree (see below), in which the strain BisA53 sequences do not form a single cluster, but as explained above we favor the grouping in the β protein tree.

The β proteins from the *Rhodobacteraceae* species (*Rhodobacter, Rhodovulum, Roseobacter*, etc.) form an independent clade, that matches the organismal phylogeny (compare Figs. 10.1 and 10.2). Although there are some differences and the support values are low, this general relationship also holds in the α protein tree (see below), and so we suggest that these LH genes descended in concert with the overall genome. Importantly, the *R. sphaeroides* β and α protein sequences both yield two clades that appear to be the result of a duplication event in this species before divergence of the three *R. sphaeroides* strains.

Hoeflea phototrophica is an alphaproteobacterium in the family *Phyllobacteriaceae* (Biebl et al., 2006). The sequences from this species do not show a consistent relationship with the other sequences making it difficult to speculate on their evolutionary history.

Despite some low support values, our analyses indicate that the betaproteobacterium *Rubrivivax gelatinosus* received LH2 genes by lateral transfer from an alphaproteobacterium, as has been suggested for the photosynthesis gene cluster in this species (Igarashi et al., 2001). Both the α and β proteins of *R. gelatinosus* are intermingled with alphaproteobacterial proteins in both trees and do not seem to be closely related to the other betaproteobacterial species *Rhodocyclus tenuis* (compare Figs. 10.2 and 10.3). This is different from the LH1 β proteins, where the betaproteobacterial sequences appear to be more closely related (Fig. 10.4).

The LH2 α and β sequences of the betaproteobacterium *R. tenuis* are external to the alphaproteobacterial protein sequences (Figs. 10.2 and 10.3), and this pattern matches the 16S rDNA-based tree. However, this branching order did not hold true when a subset of sequences was used for comparisons of LH1 and LH2 β proteins (Fig. 10.4), making it difficult to be confident of any interpretation. Similarly, it is difficult to be confident with interpretations of the exact relationships of the sequences from the gammaproteobacterium *Halorhodospira halophila* to the other LH2 sequences, because they do not show consistent and well-supported relationships to the other sequences.

Perhaps the most surprising discovery is that the alphaproteobacterium *P. molischianum* β protein sequences are divergent from all the other LH2 sequences, and the *P. molischianum* sequences appear to be more closely related to the LH1 β protein sequences that were used as the outgroup for this tree. Because this is consistent with the α protein tree, we suggest that the *P. molischianum* LH2 arose from LH1 independently of other LH2s, perhaps indicating an example of convergent evolution (see below).

B. LH2 α Proteins

The major groupings in the tree of LH2 α proteins (Fig. 10.3) are roughly similar to the 16S rDNA tree in Fig. 10.1, and the β protein tree in Fig. 10.2. The α proteins of the species in the *Bradyrhizobiaceae* family, *R. acidophilus, R. palustris* and *Bradyrhizobium* BTAi1, generally group together. Many of the proteins from species in the *Rhodobacteraceae* family, such as *R. sphaeroides*, *Rhodovulum sulfidophilum*, and *Roseobacter denitrificans*, cluster separately from the *Rhodopseudomonas* and *Rhodoblastus* proteins.

The *R. acidophilus* α sequences do not form a single clade, which differs from the β tree that groups *R. acidophilus* proteins together; for reasons summarized above, we favor the configuration of the β tree.

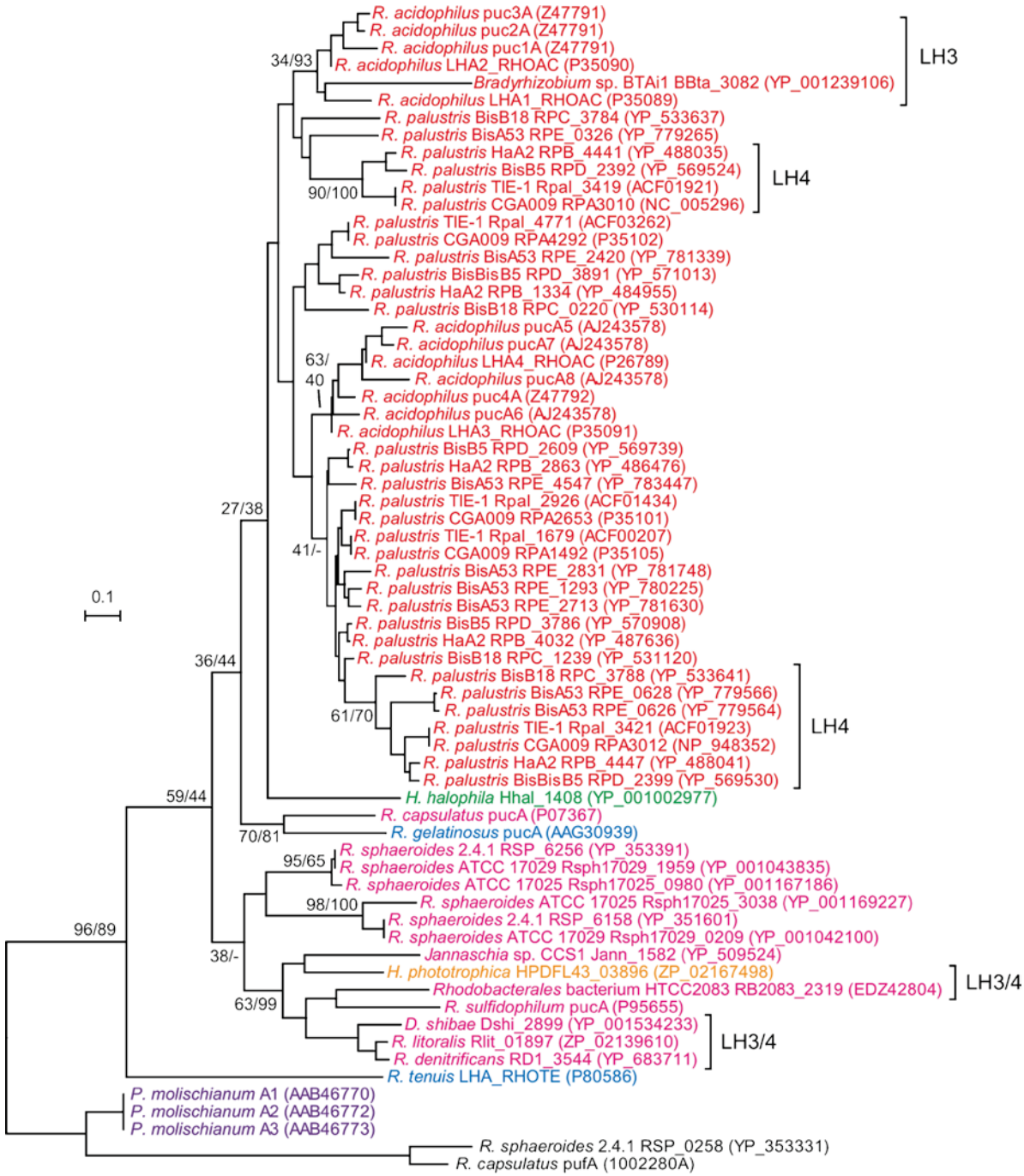


Fig. 10.3. Evolutionary relationships of LH2 α proteins. The protein identifications are colored corresponding to the taxonomic affiliations of the bacterial species as in the 16S rDNA-based phylogeny (Fig. 10.1). Large half-brackets indicate proteins that contain one or both of the LH3/4 motif residues (see text). Proteins are identified by the species and strain (where appropriate) followed by the gene or locus designation from the GenBank submission, and the GenBank accession numbers are given in brackets. The evolutionary pattern was inferred using the Neighbor-Joining (Saitou and Nei, 1987) and Bayesian methods (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). The Neighbor-Joining tree generated using MEGA4 (Tamura et al., 2007) is shown. Bootstrap values (Felsenstein, 1985) are shown as percentages based on 10,000 replicates followed by Bayesian posterior probabilities (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) based on 300,000 generations for selected relevant nodes. A dash indicates the same cluster of sequences did not occur in the Bayesian consensus tree. The evolutionary distances were computed using the JTT matrix-based method (Jones et al., 1992) and the scale bar indicates the number of amino acid substitutions per site. LH1 α protein sequences were used as the outgroup.

As in the β tree, most of the *R. palustris* α sequences are intermingled, and so we suggest that these LH2 genes evolved in a progenitor prior to the separation of five of these six strains whose genomes have been sequenced. The exception is *R. palustris* strain BisA53, in which the LH2 genes may have duplicated and diversified after speciation, based on the β protein tree (see above). However, we cannot rule out that BisA53 or any of the *R. palustris* strains acquired some of their LH2 genes by lateral gene transfer from closely related strains, including strains that exist in nature but have not yet been discovered.

The two *R. sphaeroides* α protein sequences appear to have diverged before divergence of these three strains, as in the β tree. One of the two *R. sphaeroides* α proteins (RSP_6158 and orthologs) is much larger (263 amino acids) than other α proteins. In *R. sphaeroides* strain 2.4.1 the two sets of LH2 genes are co-expressed, but the RSP_6158-encoded α protein does not participate in LH2 complex formation (Zeng et al., 2003).

The LH2 α protein of the betaproteobacterium *R. tenuis* branches separately from the alphaproteobacterial proteins, congruent with the 16S rDNA tree, whereas the betaproteobacterium *R. gelatinosus* protein is located within the alphaproteobacterial proteins. Because the bootstrap values of some of these branches are quite low, it is possible that the location of the *R. gelatinosus* α protein in this tree does not reflect the evolutionary relationship, but the similar treeing of the β protein (see above) indicates that the *R. gelatinosus* LH2 genes may have been acquired by lateral transfer from an alphaproteobacterium. The LH2 α protein of the gammaproteobacterium *H. halophila* trees within alphaproteobacterial proteins, but the bootstrap value is low and we favor the branching of the LH2 β protein tree (see above), which places the *H. halophila* LH2 separately from the alphaproteobacterial LH2s.

As in the LH2 β protein tree, the separation of the alphaproteobacterium *P. molischianum* α proteins from the other LH2 proteins has good bootstrap support, indicating a closer relationship of *P. molischianum* LH2 α proteins to LH1 α proteins (used as the outgroup for this tree) than to other LH2 α proteins. As discussed below, this is consistent with the idea that LH2 genes arose from LH1 genes, but indicates that there were two independent events: one giving rise to the LH2 genes in a progenitor of *P. molischianum*, and the other giving rise to LH2 genes in other purple bacterial species.

There are variants of the B800-850 LH2 complex that absorb light at 800 and 820 nm, or only at around 800 nm, and sometimes called LH3 or LH4 (Tharia et al., 1999; Evans et al., 2005; Gabrielsen et al., 2009). It was suggested that a key structural aspect modulating the spectral characteristics of these complexes is that B800-820 and B800 α proteins have F/L, F/T, or F/M at 13/14 residues C-terminal of the conserved H that ligates BChl, whereas the B800-850 α proteins have amino acids Y/W at these positions (Tharia et al., 1999). For simplicity, we call such proteins LH3/4-motif proteins. We inspected sequences of LH2 α proteins used in our trees to identify 20 protein sequences that have one or both of the LH3/4-motif residues, and these proteins are indicated in Fig. 10.3. The *Bradyrhizobium* sp. BTAi1, *R. palustris* BisB18 RPC_3784 and BisA53 RPE_0326, and *R. acidophilus* puc3A sequences have 1 of the 2 LH3/4 residues, and these sequences fall within a cluster of sequences that contains nine other LH3/4 proteins located at the top of the α protein tree (Fig. 10.3).

The LH3/4-motif-containing α proteins cluster in three branches: (1) the *R. acidophilus/R. palustris* proteins at the top of the tree; (2) the bottom of the *R. palustris* group; (3) the *Roseobacter* and related proteins. Gene context information described below and experimental data suggest that the sequences in the branch at the bottom of the *R. palustris* tree encode LH4-like complexes that absorb light at ~800 nm (Hartigan et al., 2002).

C. LH1 and LH2 β Proteins

Youvan and Ismail (1985) were among the first to suggest that LH1 (*pufBA*) and LH2 (*pucBA*) genes share a common ancestor. It is striking that the genes encoding proteins of both of these antenna complexes are transcriptionally organized with the β protein gene preceding the α protein gene, and that there is on the order of 30% amino acid sequence identity between LH1/2 α proteins and LH1/2 β proteins.

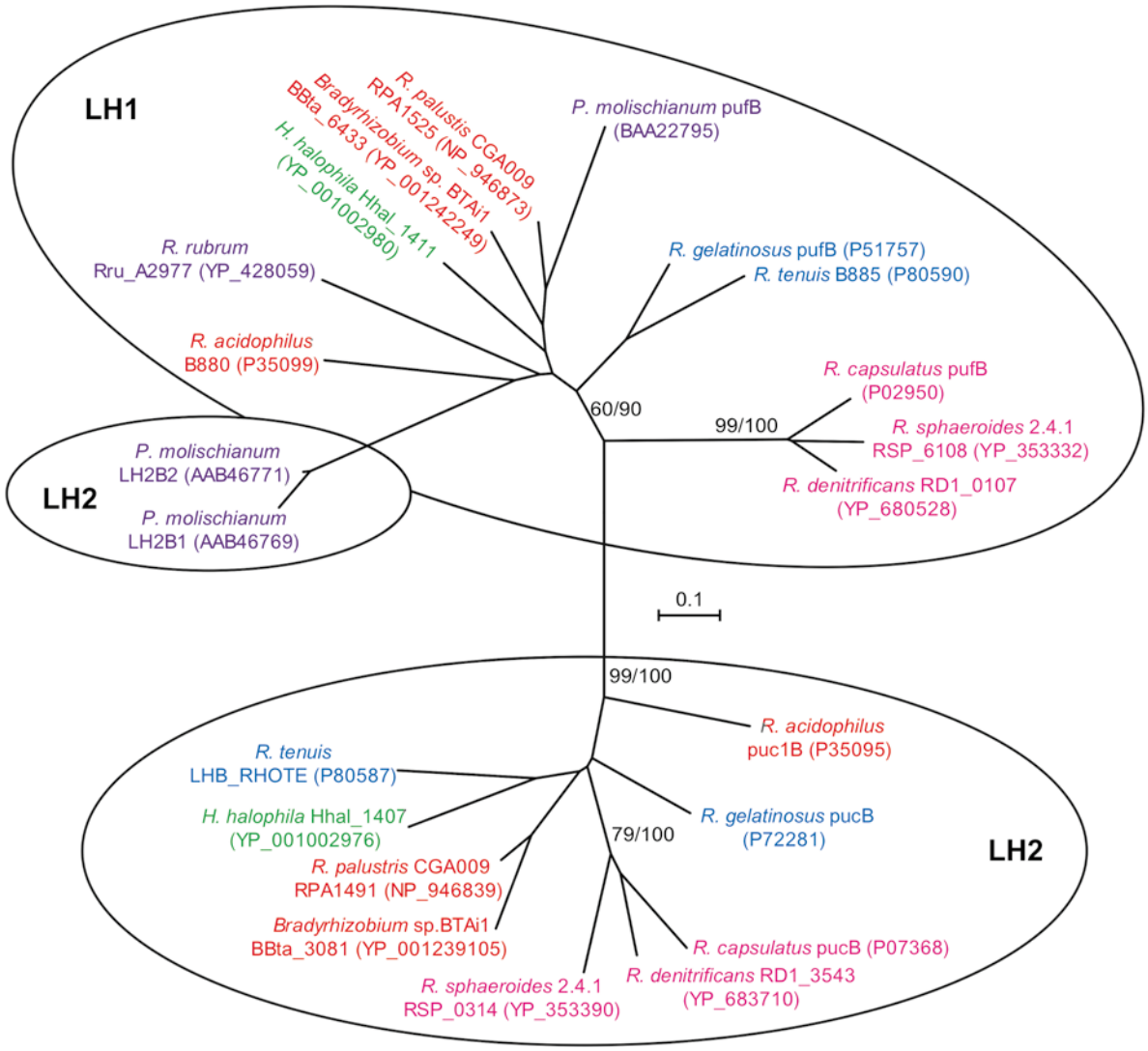


Fig. 10.4. Evolutionary relationships of selected LH1 and LH2 β proteins. The protein identifications are colored corresponding to the taxonomic affiliations of the bacterial species as in the 16S rDNA-based phylogeny (Fig. 10.1). Proteins are identified by the species and strain (where appropriate) followed by the gene or locus designation from the GenBank submission, and the GenBank accession numbers are given in brackets. The evolutionary pattern was inferred using the Neighbor-Joining (Saitou and Nei, 1987) and Bayesian methods (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). The unrooted Neighbor-Joining tree inferred from 10,000 replicates (Felsenstein, 1985) using MEGA4 (Tamura et al., 2007) is shown. Bootstrap values (Felsenstein, 1985) are shown as percentages based on 10,000 replicates followed by Bayesian posterior probabilities (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) based on 300,000 generations for selected relevant nodes. The evolutionary distances were computed using the JTT matrix-based method (Jones et al., 1992) and the scale bar indicates the number of amino acid substitutions per site.

A tree of LH1 *vs*. LH2 β proteins is given in Fig. 10.4. The encircled regions clearly indicate two major clusters of protein sequences, separated along LH1 *vs*. LH2 lines. Because LH1 is intimately associated with the heart of phototrophic energy transduction, the RC, we suggest that LH1 genes existed before and gave rise to LH2 genes.

The *P. molischianum* LH2 β sequences are clearly more closely related to LH1 β proteins than to other LH2 β proteins (Fig. 10.4). We suggest that LH2 arose from LH1 more than once, to give rise to *P. molischianum* LH2 and independently to give rise to the other LH2s, although additional sequence data from other

species may change this interpretation in the future. Presumably this indicates an evolutionary advantage is gained through the acquisition or evolution of LH2 systems because of the increased light-harvesting capability that results.

V. Implications of Genomic Context of LH2 Genes

Of the 23 phototrophs for which LH2 sequence data exist, complete genome sequence data are available for 18 strains. In general, LH2 genes are found as a set, with a *pucB* (LH2 β) gene followed by a *pucA* (LH2 α) gene in a genuine or predicted operon. One exception is the *R. palustris* BisA53 gene RPE_2420 (*pucA*), which encodes an LH2 α protein but is not preceded by a β (*pucB*) gene. Also, the *R. palustris* CGA009 gene RPA3010 is a frameshifted LH2 α (*pucA*) gene. The frameshift is likely a relatively recent event because the corresponding gene (Rpal_3419) from the very closely related strain TIE-1 is intact. In each of the *R. palustris* strains, in the *R. sphaeroides* strains, and in several other of the species considered here, *pucC* genes (required for proper assembly of LH2 complexes) are adjacent to one or more of the LH2 gene sets. Gene regions surrounding *pucBAC* gene sets in strains CGA009, TIE-1, HaA2 and BisB5 are conserved, and include a bacteriophytochrome gene (RPA1490 in strain CGA009) which controls LH2 expression in response to light or redox, depending on the strain (Vuillet et al., 2007). All six of the *R. palustris* genomes have two sets of LH2 genes very close to a set of regulatory genes that encodes two bacteriophytochromes, two response regulator receivers, and a transcriptional regulator (RPA3014-3018 in strain CGA009) that has been shown to regulate the synthesis of a B800 (LH4-type of) LH complex (Giraud et al., 2005). Correspondingly, this regulatory gene set is adjacent or very close to the seven α genes labeled LH4 in Fig. 10.3. We suggest that this regulatory region, which is slightly rearranged depending on the strain, evolved from a common ancestor and that the two sets of *pucBA* genes in this region arose by a gene duplication event. It is noteworthy that this regulatory gene set and the physically linked LH2 gene sets have different gene neighborhoods depending on the strain.

Most of the *R. palustris* BisA53 LH2 gene sets are present in locations and near to genes that do not match with those found in any of the other *R. palustris* genomes. This is consistent with the phylogeny in Fig. 10.2, suggesting that there was an expansion of the BisA53 LH2 gene set after its genome diverged from the other genomes. The one exception is the RPE_0325-0326 gene pair, which looks like it has landed in a *Rhodopseudomonas* gene neighborhood that is conserved in all the other *Rhodopseudomonas* genomes as well as the *Bradyrhizobium* BtAi1 genome. None of the other *R. palustris* genomes, except BisB18 has an LH2 gene pair in this neighborhood. One explanation for this is that strains BisA53 and BisB18 acquired RPE_0325-0326 and RPC_0219-220 by lateral gene transfer from a related *Rhodopseudomonas* strain, possibly before these two strains diverged from each other. Alternatively, the other strains may have lost this LH2 gene set. There is good evidence that the one LH2 gene set in *Bradyrhizobium* sp. BTAi1 was acquired by lateral gene transfer. This gene set is not present in *Bradyrhizobium* ORS278, nor in 107 photosynthetic *Bradyrhizobium* strains that were surveyed by PCR (Jaubert et al., 2008).

VI. Concluding Remarks

We have used the currently available LH2 sequences to infer the evolutionary histories of LH2 genes in the bacterial species that contain peripheral antenna complexes. Despite some problems caused by the nature of these LH2 proteins, we make several proposals about these histories. Most LH2 genes appear to have evolved through linear descent from a common ancestor that contained LH2 genes. The phylogenetic relationships that we have observed suggest there have also been a limited number of occasions of lateral transfer of LH2 genes. Our findings support the idea that LH2 complexes evolved from LH1 complexes, and we also believe there have been two independent LH2 evolution events, with one generating the LH2 complex present in the alphaproteobacterium *P. molischianum* and the other generating the complexes present in all other species discussed here.

Acknowledgments

For funding, ASL thanks the Canadian Natural Sciences and Engineering Research Council (NSERC), CSH thanks the US Department of Energy, and JTB thanks the NSERC.

References

Abresch EC, Axelrod HLA, Beatty JT, Johnson JA, Nechushtai R and Paddock ML (2005) Characterization of a highly purified, fully active, crystallizable RC/LH1/PufX core complex from *Rhodobacter sphaeroides*. Photosynth Res 86: 61–70

Biebl H, Tindall BJ, Pukall R, Lunsdorf H, Allgaier M and Wagner-Dobler I (2006) *Hoeflea phototrophica* sp. nov., a novel marine aerobic alphaproteobacterium that forms bacteriochlorophyll *a*. Int J Sys Evol Microbiol 56: 821–826

Braatsch S, Bernstein JR, Lessner F, Morgan J, Liao JC, Harwood CS and Beatty JT (2006) *Rhodopseudomonas palustris* CGA009 has two functional *ppsR* genes, each of which encodes a repressor of photosynthesis gene expression. Biochemistry 45: 14441–14451

Cogdell RJ, Isaacs NW, Howard TD, McLuskey K, Fraser NJ and Prince SM (1999) How photosynthetic bacteria harvest solar energy. J Bacteriol 181: 3869–3879

Evans K, Fordham-Skelton AP, Mistry H, Reynolds CD, Lawless AM and Papiz MZ (2005) A bacteriophytochrome regulates the synthesis of LH4 complexes in *Rhodopseudomonas palustris*. Photosynth Res 85: 169–180

Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39: 783–791

Gabrielsen M, Gardiner AT and Cogdell RJ (2009) Peripheral complexes of purple bacteria. In: Hunter CN, Daldal F, Thurnauer MC and Beatty JT (eds) The Purple Phototrophic Bacteria. Advances in Photosynthesis and Respiration, Vol 28, pp 135–153. Springer, Dordrecht.

Giraud E, Zappa S, Vuillet L, Adriano J-M, Hannibal L, Fardoux J, Berthomieu C, Bouyer P, Pignol D and Vermeglio A (2005) A new type of bacteriophytochrome acts in tandem with a classical bacteriophytochrome to control the antennae synthesis in *Rhodopseudomonas palustris*. J Biol Chem 280: 32389–32397

Green BR, Anderson JM and Parson WW (2003) Photosynthetic membranes and their light-harvesting antennas. In: Green BR and Parson WW (eds) Light-Harvesting Anetennas in Photosynthesis, pp 1–28. Kluwer, Dordrecht.

Hartigan N, Tharia HA, Sweeney F, Lawless AM and Papiz MZ (2002) The 7.5-Å electron density and spectroscopic properties of a novel low-light B800 LH2 from *Rhodopseudomonas palustris*. Biophys J 82: 963–977

Huelsenbeck JP and Ronquist F (2001) MRBAYES. Bayesian inference of phylogeny. Bioinformatics 17: 754–755

Igarashi N, Harada J, Nagashima S, Matsuura K, Shimada K and Nagashima KVP (2001) Horizontal transfer of the photosynthesis gene cluster and operon rearrangement in purple bacteria. J Molec Evol 52: 333–341

Jaubert M, Vuillet L, Hannibal L, Adriano J-M, Fardoux J, Bouyer P, Bonaldi K, Fleischman D, Giraud E and Vermeglio A (2008) Control of peripheral light-harvesting complex synthesis by a bacteriophytochrome in the aerobic photosynthetic bacterium *Bradyrhizobium* strain BTAi1. J Bacteriol 190: 5824–5831

Jones DT, Taylor WR and Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. CABIOS 8: 275–282

Koepke J, Hu X, Muenke C, Schulten K and Michel H (1996) The crystal structure of the light-harvesting complex II (B800-850) from *Rhodospirillum molischianum*. Structure 4: 581–597

McDermott G, Prince SM, Freer AA, Hawthornthwaite-Lawless AM, Papiz MZ, Cogdell RJ and Isaacs NW (1995) Crystal structure of an integral membrane light-harvesting complex from photosynthetic bacteria. Nature 374: 517–521

McLuskey K, Prince SM, Cogdell RJ and Isaacs NW (2001) The crystallographic structure of the B800-820 LH3 light-harvesting complex from the purple bacteria *Rhodopseudomonas acidophila* strain 7050. Biochemistry 40: 8783–8789

Oda Y, Larimer FW, Chain PSG, Malfatti S, Shin MV, Vergez LM, Hauser L, Land ML, Braatsch S, Beatty JT, Pelletier DA, Schaefer AL, and Harwood CS (2008) Multiple genome sequences reveal adaptations of a phototrophic bacterium to sediment microenvironments. Proc Natl Acad Sci USA 105: 18543–18548

Olsen JD, Tucker JD, Timney JA, Qian P, Vassilev C and Hunter CN (2008) The organization of LH2 complexes in membranes from *Rhodobacter sphaeroides*. J Biol Chem 283: 30772–30779

Qian P, Hunter CN and Bullough PA (2005) The 8.5Å projection structure of the core RC–LH1–PufX dimer of *Rhodobacter sphaeroides*. J Molec Biol 349: 948–960

Rey FE, Heiniger EK and Harwood CS (2007) Redirection of metabolism for biological hydrogen production. Appl Environ Microbiol 73: 1665–1671

Ronquist F and Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574

Roszak AW, Howard TD, Southall J, Gardiner AT, Law CJ, Isaacs NW and Cogdell RJ (2003) Crystal structure of the RC-LH1 core complex from *Rhodopseudomonas palustris*. Science 302: 1969–1972

Saitou N and Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molec Biol Evol 4: 406–425

Scheuring S (2009) The supramolecular assembly of the photosynthetic apparatus of purple bacteria investigated

by high-resolution atomic force microscopy. In: Hunter C, Daldal F, Thurnauer M and Beatty J (eds) The Purple Phototrophic Bacteria. Advances in Photosynthesis and Respiration, Vol 28, pp 941–952. Springer, Dordrecht.

Tamura K, Dudley J, Nei M and Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci USA 101: 11030–11035

Tamura K, Dudley J, Nei M and Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molec Biol Evol 24: 1596–1599

Tharia HA, Nightingale TD, Papiz MZ and Lawless AM (1999) Characterisation of hydrophobic peptides by RP-HPLC from different spectral forms of LH2 isolated from *Rps. palustris*. Photosynth Res 61: 157–167

Vuillet L, Kojadinovic M, Zappa S, Jaubert M, Adriano J, Fardoux J, Hannibal L, Pignol D, Verméglio A and Giraud E (2007) Evolution of a bacteriophytochrome from light to redox sensor. EMBO J 26: 3322–3321

Youvan DC and Ismail S (1985) Light harvesting II (B800-850 complex) structural genes from *Rhodopseudomonas capsulata*. Proc Natl Acad Sci USA 82: 58–62

Zeng X, Choudhary M and Kaplan S (2003) A second and unusual *pucBA* operon of *Rhodobacter sphaeroides* 2.4.1: genetics and function of the encoded polypeptides. J Bacteriol 185: 6171–6184

Chapter 11

The Extended Light-Harvesting Complex (LHC) Protein Superfamily: Classification and Evolutionary Dynamics

Johannes Engelken
Department of Biology, University of Konstanz, Universitätsstrasse 10, DE-78457 Constance, Germany; Institute of Evolutionary Biology (CSIC-UPF), Pompeu Fabra University, Barcelona, Spain

Christiane Funk
Department of Chemistry and Umeå Plant Science Centre (UPSC), Umeå University, S-90187 Umeå, Sweden

Iwona Adamska*
Department of Biology, University of Konstanz, Universitätsstrasse 10, DE-78457 Constance, Germany

*Author for correspondence, e-mail: Iwona.Adamska@uni-konstanz.dc

R.L. Burnap and W.F.J. Vermaas (eds.), Functional Genomics and Evolution of Photosynthetic Systems, Advances in Photosynthesis and Respiration 33, pp. 265–284, DOI 10.1007/978-94-007-1533-2_11,

Summary

The evolution of algae and land plants and their photosynthetic machineries is closely connected to the development of the extended light-harvesting complex (LHC) protein superfamily. Therefore, it is not surprising that the molecular organization, function and origin of the LHC proteins have been a central topic in plant biology and photosynthesis research during the last few years. The extended LHC protein superfamily in cyanobacteria and photosynthetic eukaryotes comprises different families, such as the LHC proteins and three groups of light stress-induced proteins, consisting of the LHC-like proteins, the red lineage CAB-like proteins and the photosystem II subunit S. This chapter provides a description of the different extended LHC superfamily members and shows their taxonomic distribution. Furthermore, an overview of scenarios suggested for the evolution of the extended LHC protein superfamily is provided and arised implications for light harvesting, stress responses and photoprotection are discussed.

I. Introduction

Sunlight is the origin of nearly all the metabolic energy that drives life processes, and plants, algae and photosynthetic bacteria are responsible for the conversion of solar energy into chemical forms that are used by all organisms. Light harvesting is the first step in this photosynthetic process, in which the absorbed light energy is transferred to the reaction center of a photosystem. The light-harvesting antennae of photosynthetic organisms have, therefore, to be regulated qualitatively and quantitatively in response to their physiological status and the environmental signals. In spite of the great variety of light-harvesting antennas among photosynthetic organisms, only three major types of light-absorbing chromophores and only few protein families are employed in this function in oxygenic photosynthesis. Chlorophylls (Chls) ligated to light-harvesting complex (LHC) proteins with three transmembrane (TM) alpha-helices (Kühlbrandt et al., 1994) serve as antenna in the majority of algae and land plants (Streptophyta). Since it was suggested that LHC proteins of all photosynthetic eukaryotes have a common origin, all these proteins are included into the extended LHC protein superfamily. In recent years more and more genomes of photosynthetic organisms were sequenced, allowing discoveries of new LHC-related proteins. Some of these proteins seem not to be involved in light harvesting but play a role in photoprotection. In the presence of oxygen the balance between light harvesting and light protection is important to prevent the formation of reactive oxygen species and to avoid photooxidative damage.

With regard to the evolution of the extended LHC protein superfamily, different models have been proposed (Green and Pichersky, 1994; Durnford et al., 1999; Montané and Kloppstech, 2000; Heddad and Adamska, 2002; Garczarek et al., 2003; Green, 2003; Six et al., 2005; Jansson, 2006; Koziol et al., 2007), mainly suggesting a four-helix intermediate as ancestor of various families, or alternatively, a protein resembling the cyanobacterial high light-induced protein (HLIP)/small CAB-like protein (SCP) (Green, 2003).

In this review we provide a coherent classification scheme for proteins of the extended LHC protein superfamily and discuss a new model for the evolution of these proteins. Furthermore, we summarize arising implications for photoprotection, stress response and light harvesting.

II. The Extended Light-Harvesting Complex (LHC) Protein Superfamily: Nomenclature and Taxonomic Distribution

The defining homologous structure of the extended LHC protein superfamily is a shared Chl-binding (CB) motif that is part of a TM alpha-helix located

Abbreviations: CAA – chlorophyll *a*-binding; CAB – chlorophyll *a/b*-binding; CAC – chlorophyll *a/c*-binding; CB – chlorophyll-binding; Chl – Chlorophyll; ELIP – early light-induced protein; FCP – fucoxanthin chlorophyll *a/c*-binding protein; HL – high light; HLIP – high light-induced protein; LHC – light-harvesting complex; LHL – high intensity light-inducible LHC-like; LIL – light-harvesting-like; LL – low light; NPQ – non-photochemical quenching; OHP – one-helix protein; PS – photosystem; PSBS – the S subunit of photosystem II; RedCAP – red lineage CAB-like protein; SCP – small chlorophyll *a/b*-binding-like proteins; SEP – stress-enhanced protein; TM – transmembrane

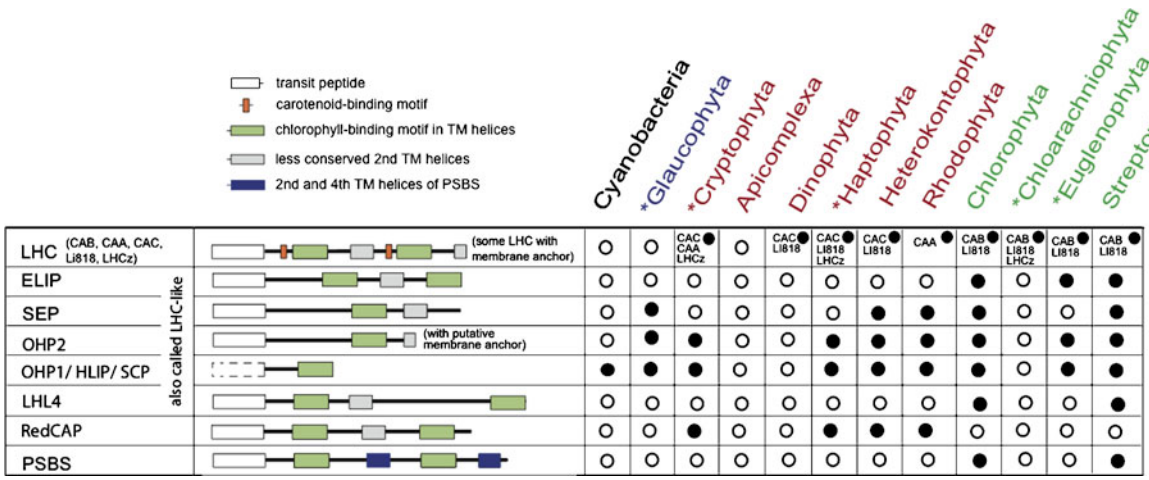


Fig. 11.1. Schematic overview of the extended LHC protein superfamily. *Left*: Predicted protein structures of representative members. A cleavable transit peptide (only present in nuclear-encoded members) is followed by a TM alpha-helix containing the homologous CB motif and other sequence motives. The drawing shows the approximate position and relative length of conserved regions and sequence motifs. *Right*: Distribution of CB proteins in photosynthetic organisms. Presence (*black circle*) and absence (*white circle*) of CB proteins in genomic and EST (marked by *stars*) databases is indicated.

in the thylakoid membrane. The extended LHC protein superfamily comprises several families (Fig. 11.1; reviewed in Green and Durnford, 1996; Jansson, 1999a; Montané and Kloppstech, 2000; Adamska, 2001; Funk, 2001; Green, 2003; Jansson, 2006), including the best characterized LHC proteins, the stress-induced LHC-like proteins and PSBS, an enigmatic protein involved in photoprotection. Despite diverse pigment content, primary sequence similarities indicate that all these proteins originated from a common ancestor with one TM helix (Green and Pichersky, 1994; Wolfe et al., 1994; Durnford et al., 1999; Heddad and Adamska, 2002; Garczarek et al., 2003; Jansson, 2006; Koziol et al., 2007; Engelken et al., 2010). These proteins are the main focus of this review. Non-homologous CB proteins building the inner LHC antennas, such as the Chl-binding protein (Pcb) of Prochlorophyta, the soluble peridinin-Chl *a*-protein (PCP) of Dinoflagellates, the Chl *a*-binding proteins CP43 and CP47 of higher plants and algae or their iron stress-induced (IsiA) homologs in cyanobacteria (Zhang et al., 2007) are not included in this review.

While the main function of the majority of LHC proteins is light harvesting, the members of the LHC-like protein family, PSBS, and probably of red lineage Chl *a/b*-binding (CAB)-like protein (RedCAP) families and the LI818 (LHCx) subfamily of LHC proteins are involved in mechanisms that act in response to excessive light and protect the reaction centers against photooxidative damage (Savard et al., 1996; Montané and Kloppstech, 2000; Li et al., 2000; Adamska, 2001; Funk, 2001; Teramoto et al., 2004). However, also typical LHC proteins are involved in photoprotection by participation in thermal energy dissipation (Pascal et al., 2005; Standfuss et al., 2005; Ruban et al., 2007) and state transitions (Haldrup et al., 2001; Wollman, 2001; Allen, 2003; Rochaix, 2007). The functional aspects of the extended LHC superfamily members are discussed in Section V.

A. The LHC Protein Family

The LHC protein family forms antenna complexes in algae and land plants (Fig. 11.1) and is the largest family of CB proteins in photosynthetic eukaryotes in terms of number of paralogs. All LHC family members have three TM helices with two CB motifs located in helices I and III (Fig. 11.1; Green and Durnford, 1996; Jansson, 1999a). This family is divided into subfamilies of the Chl *a*-binding (CAA) proteins, the Chl *a/b*-binding (CAB) proteins, the Chl *a/c*-binding (CAC) proteins and lesser known LHC clades including LI818 (LHCx, Richard et al., 2000) and LHCz (Koziol et al., 2007) (Fig. 11.1). A short summary of these subfamilies is provided below. The LHCz subfamily with members found in

some algae, like Cryptophyta, Haptophyta and Chlorachanophyta (Fig. 11.1; Koziol et al., 2007) are not described here in detail, since biochemical evidence hinting at their function or localization is lacking.

1. The Chlorophyll a-Binding (CAA) Protein Subfamily

The members of the CAA subfamily bind Chl *a* as the only Chl type and were reported only from red algae (Rhodophyta). LHCr genes encode these proteins, where the suffix 'r' indicates the algal group. Recently, sequences related to red algal CAA were found also in the genome of *Guillardia theta* (Cryptophyta) (Fig. 11.1). The first member of this family was isolated from the alga *Porphyridium cruentum* and shown to form the Chl *a*-containing antenna associated exclusively with photosystem I (PS I). Furthermore, CAA polypeptides were found to be immunogenically related to those of CAB and CAC (Wolfe et al., 1994). In the following years several other *LHCr* genes were discovered in the genomes of red alga *Galdieria sulphuraria*, *Cyanodioschyzon merolae*, and *Porphyra purpurea* (Fig. 11.2; Marquardt and Rhiel, 1997; Marquardt et al., 2001; Koziol et al., 2007). Interestingly, in the diatom Cyclotella (*Cyclotella cryptica)* sequences coding for 12 fucoxanthin Chl *a/c*-binding proteins (FCP) (LHCf1 through LHCf12) were identified, of which FCP4/LHCf4 is phylogenetically closely related to LHCr1 and LHCr2 proteins of *Porphyridium cruentum* (Eppard et al., 2000; Marquardt et al., 2001; Koziol et al., 2007). It has been suggested that FCP4/LHCf4 might have retained its original function as an LHCa protein in Cyclotella (Eppard et al., 2000). However, currently available data give no clear evidence that FCP4/LHCf4 is an antenna subunit of PS I.

2. The Chlorophyll a/b-Binding (CAB) Protein Subfamily

The CAB subfamily of green algae (Chlorophyta), Euglenophyta, Chlorarachniophyta and land plants (Fig. 11.1) is the most diverse from the LHC family concerning their presently known function and organization. This subfamily is composed of several nuclear-encoded CAB proteins associated with PS I (LHCa) or PS II (LHCb). Ten major types of CAB proteins are encoded in the genome of Arabidopsis (*Arabidopsis thaliana*), four of which are associated with PS I (LHCa1 through LHCa4) and six with PS II (LHCb1 through LHCb6) (Jansson, 1999a, 2006). In addition, the Arabidopsis genome carries three rarely expressed LHC genes (coding for LHCa5, LHCb7 and LHCb8) and one possible pseudogene (LHCa6) (Klimmek et al., 2006). The LHCb proteins are organized in major (LHCb1 through LHCb3) and minor (LHCb4 through LHCb6, called also CP29, CP26 and CP24, respectively) antenna systems. The major antenna is either tightly associated with the PS II core complex or forms an outer, loosely bound pool peripheral to the PS II core (Bassi and Dainese, 1992; Consoli et al., 2005). This pool can reversibly dissociate from and associate with PSII during a process called state transitions (Haldrup et al., 2001; Wollman, 2001; Allen, 2003; Rochaix, 2007).

The first resolved structures were reported for the major three-helix LHCb protein in spinach (*Spinacia oleracea*) (Kühlbrandt et al., 1994; Liu et al., 2004) and pea (*Pisum sativum*) (Standfuss et al., 2005). The monomeric unit of the complex is composed of three TM helices A-C and one amphipathic helix D exposed at the surface on the luminal side of the thylakoid membrane. The two central helices A and B are in close contact with each other, held together by two inter-helix ionic pairs, forming a left-handed supercoil with two-fold symmetry (Kühlbrandt et al., 1994; Liu et al., 2004). Furthermore, it was shown that the basic structural and functional unit of LHCb is the trimer (Liu et al., 2004; Nield and Barber, 2007). Interactions between adjacent trimers and PS II cores form a three-dimensional macrostructure in the stacked membranes of grana (Horton and Ruban, 2005).

The antenna system of PS I in Arabidopsis is composed of four major proteins LHCa1 through LHCa4 (Jansson, 1999a). Two other genes, *LHCa5* and *LHCa6*, are thought to code for additional PS I antenna subunits based on sequence similarity (Jansson, 1999a; Ganeteg et al., 2004). However, only LHCa5 has been detected in the antenna of PS I in very low amounts under high light (HL) conditions (Ganeteg et al., 2004). The crystal structure of the pea PS I supercomplex, containing a reaction center and its peripheral LHCa antenna, showed that LHCa consists of two

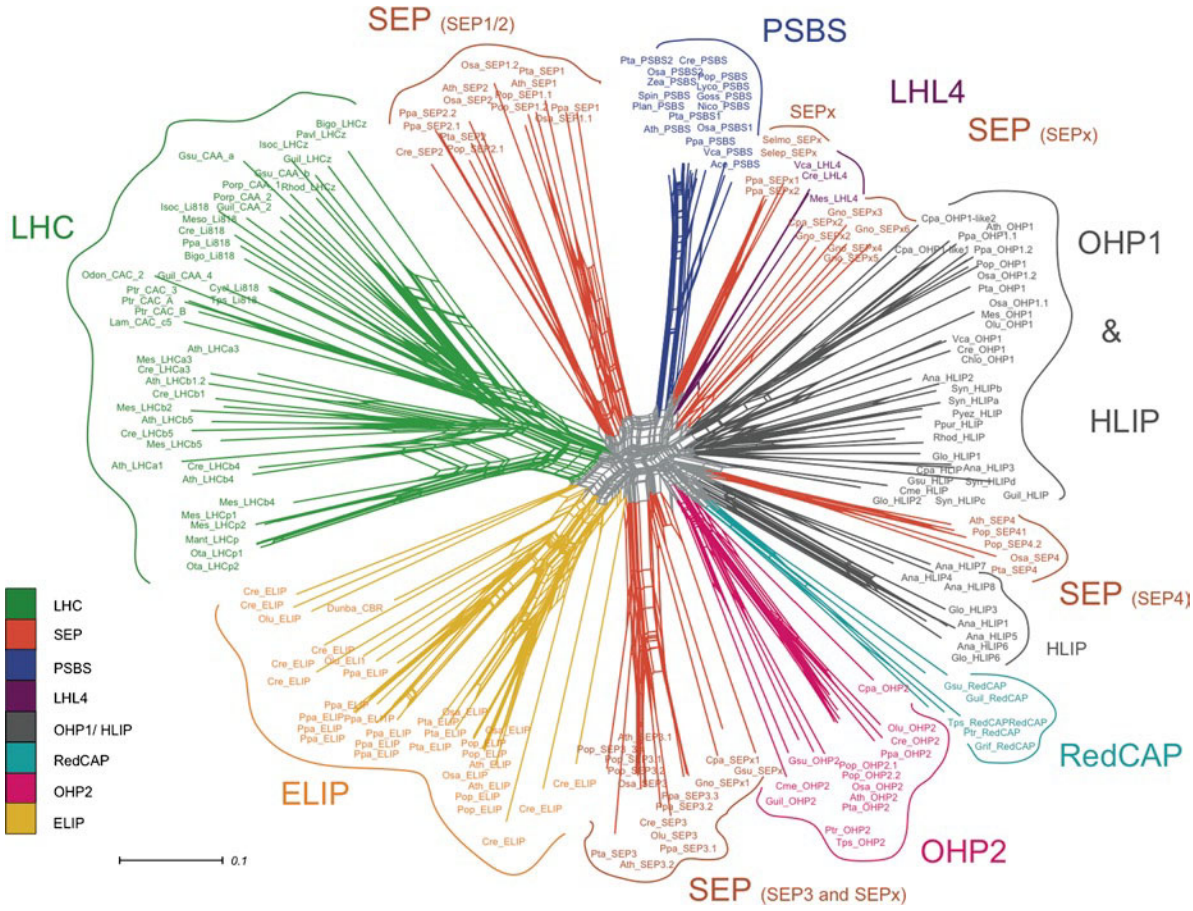


Fig. 11.2. Phylogenetic representation of the extended LHC protein superfamily showing its high diversity. Phylogenetic network based on 33 aligned amino acid positions (corresponding to the first TM helix) and 193 sequences (neighbor-net algorithm in SplitsTree 4.10; Huson and Bryant, 2006). Distantly related and slowly evolving representatives were chosen for each family. Due to the network analysis and the inherent short length of the alignment, this analysis has no power to resolve the deep relationships between families. Abbreviations for species names in alphabetical order: Ace, *Acetabularia acetabulum*; Ana, *Anabena* sp. PCC 7120; Ath, *Arabidopsis thaliana*; Bigo, *Bigelowiella natans*; Cme, *Cyanidioschyzon merolae*; Cpa, *Cyanophora paradoxa*; Cre, *Chlamydomonas reinhardtii*; Cyclo, *Cyclotella cryptica*; Dunba, *Dunaliella bardawil*; Eug, *Euglena variabilis*; Glo, *Gloeobacter violaceus* PCC 7421; Gno, *Glaucocystis nostochinearum*; Goss, *Gossypium mexicanum*; Grif, *Griffithsia japonica*; Gsu, *Galdieria sulphuraria*; Guil, *Guillardia theta*; Isoc, *Isochrysis galbana*; Lam, *Laminaria saccharina*; Lyco, *Lycopersium esculentum*; Mant, *Mantoniella squamata*; Meso, *Mesostigma viride*; Nico, *Nicotiana benthamiana*; Odon, *Odontella cryptica*; Olu, *Ostreococcus lucimarinus*; Osa, *Oryza sativa japonica*; Ota, *Ostreococcus tauri*; Pavl, *Pavlova lutheri*; Plan, *Plantago major*; Pop, *Populus trichocarpa*; Porp, *Porphyra cruentum*; Ppa, *Physcomitrella patens*; Ppur, *Porphyra purpurea*; Ptr, *Phaeodactylum tricornutum*; Pyez, *Porphyra yezoensis*; Rhod, *Rhodomonas* sp.; Selep, *Selaginella lepidophylla*; Selm, *Selaginella moellendorffi*; Spin, *Spinacea oleacea*; Syn, *Synechocystis* sp. PCC 6803; Tps, *Thalassiosira pseudonana*; Vca, *Volvox carteri*; Zea, *Zea mays*.

separate dimers (LHCa1/LHCa4 and LHCa2/LHCa3) that are arranged in a half-moon-shaped belt bound to one side of PS I (Ben-Shem et al., 2003; Amunts et al., 2007). Of the five expressed *LHCa* sequences (*LHCa1* through *LHCa5*) in Arabidopsis, the *LHCa3*-encoded protein is the only clear homolog identified in green algae examined so far (Koziol et al., 2007).

Our knowledge about the CAB subfamily in green algae is coming mainly from studies on Chlamydomonas (*Chlamydomonas reinhardtii*) and Ostreococcus (*Ostreococcus tauri*), both organisms with completely sequenced genomes (Merchant et al., 2007; Palenik et al., 2007). Similarly to Arabidopsis, LHCa and LHCb proteins in Chlamydomonas are encoded by multigene

families, with the exception of CP29 and CP26 (Elrad and Grossman, 2004). Nine *LHCb* genes encoding the major PS II antenna were identified in this alga (Elrad and Grossman, 2004).

Modeling of the PS II-LHCb supercomplex of Chlamydomonas revealed a very similar structure to the higher plant supercomplex, although three major antenna proteins (LHCb1 through LHCb3) and the minor antenna protein CP24 are missing in this alga (Nield et al., 2000; Elrad and Grossman, 2004). Similarly, CP24 is absent in the Ostreococcus genome (Six et al., 2005).

Nine LHCa proteins were reported in Chlamydomonas (Elrad and Grossman, 2004). Phylogenetic analysis of Chlamydomonas (Tokutsu et al., 2004) and Ostreococcus (Six et al., 2005) LHCa3 supported the presence of plant LHCa3 orthologs in green algae. Since in Chlamydomonas N-terminal processing of LHCa3 is involved in PS I antenna remodeling during iron deficiency, it was proposed that LHCa3 might act as a linker in the formation of the PS I-LHCa supercomplex (Naumann et al., 2005). Recently, two green algal-specific *LHCa* genes, called *LHCa2* (which is different from the plant *LHCa2* gene) and *LHCa9*, were identified (Koziol et al., 2007). It was proposed that the homolog of plant LHCa3 and green algal-specific proteins, LHCa2 and LHCa9 might form the LHCa PSI antenna belt in green algae.

In the euglenophyte *Euglena gracilis* polyproteins for LHCa and LHCb have been reported (Koziol and Durnford 2008). Such polyproteins encoded by a large mRNA are cleaved to individual LHC proteins after import into the chloroplast (Koziol and Durnford, 2008 and references within). However, no plant or green algal homologs of LHCa have been found in this alga (Koziol et al., 2007 and references within).

3. The Chlorophyll a/c-Binding (CAC) Protein Subfamily

CAC proteins are found in various algal taxa with complex red plastids, including Heterokontophyta, Haptophyta (also called Prymnesiophyta), Dinophyta and Cryptophyta (Fig. 11.1; Bathke et al., 1999; Durnford et al., 1999; Green, 2003; Koziol et al., 2007; Kereïche et al., 2008). It was shown that CAC are immunologically related to CAB proteins of higher plants and green algae (Green and Pichersky, 1994). Since some of the CAC proteins in Heterokontophyta, *e.g.*, in diatoms and brown algae, bind the xanthophyll fucoxanthin in addition of Chl *a* and *c,* they are also called FCPs (Green and Durnford, 1996; Durnford et al., 1999) or LHCf (Jansson 1999b).

Very little is known on the exact function of FCPs and their association with PS I and PS II. Büchel (2003) showed that FCPs in *Cyclotella meneghiniana* assemble into different oligomeric states, forming either trimers or higher oligomers. The higher oligomeric state in the diatom *Phaeodactylum tricornutum* consists either of six monomers or two tightly associated trimers (Lepetit et al., 2007). The excitation energy transfer between the FCP-bound pigments is more efficient in the oligomeric FCP complexes, indicating that these complexes may represent the native form of the diatom antenna system in the thylakoid membrane (Lepetit et al., 2007).

Up to 10 different LHC proteins could be identified in the cryptophyte *Rhodomonas* sp. to be associated with PS I or PS II (Bathke et al., 1999). Recent studies revealed that PS I has 6–8 monomers of Chl *a/c2* antenna complexes at both sides of the PS I core monomer showing no obvious similarities in size and shape with that of PS I-LHCI supercomplexes from green algae and plants (Kereïche et al., 2008). A PS II core dimer also contained three or four monomeric Chl *a/c2* proteins located on one side of the PS II core complex at positions corresponding to those of LHCb4, LHCb5 and LHCb6 of green plants (Kereïche et al., 2008).

4. The LI818 (LHCx) Protein Subfamily

The first LI818 protein was described from Chlamydomonas as a distant relative of the CAB protein subfamily although it exhibited a different expression pattern (Savard et al., 1996). In synchronized light/dark Chlamydomonas cultures *LI818* transcripts are expressed in a light-dependent and photosynthesis-independent manner before the accumulation of *CAB* transcripts (Savard et al., 1996; Richard et al., 2000). In microarray analyses *LI818* transcripts increased 20-fold following sulfur deprivation, where *LHC* transcripts generally declined (Richard et al., 2000; Elrad and Grossman, 2004). In Cyclotella transcripts for LI818 proteins, *FCP6/LHCf6, FCP7/LHCf7* and *FCP12/LHCf12,* accumulated in response to illumination with HL (Eppard

et al., 2000; Oeltien et al., 2002; Koziol et al., 2007). Similar results were reported for the diatom Thalassiosira (*Thalassiosira pseudonana*) (Zhu and Green, 2008). This suggests that similar to early light-induced protein (ELIP) and PSBS family members also LI818 proteins might have a photoprotective function, although LI818 proteins seem to be phylogenetically closer related to LHC proteins than to ELIPs (Elrad and Grossman, 2004).

Protein sequence comparisons suggest that LI818 proteins have three TM helices and several conserved residues involved in Chl binding (Richard et al., 2000). The exact localization of LI818 proteins is not yet known. It was shown for Chlamydomonas that LI818 is not tightly embedded in the stroma region of the thylakoid membranes and could be extracted with chaotropic agents and extreme alkaline pH treatment (Richard et al., 2000).

The LI818 subfamily has representatives in the major groups of algae, including green algae (Chlamydomonas, Ostreococcus, *Micromonas viride*), Chlorarachniophyta (*Bigelowiella natans*), Heterokontophyta (Cyclotella, Thalassiosira and *Phaeodactylum tricornutum)*, Haptophyta (*Isochrysis galbana*), as well as some of Euglenophyta and Dinophyta (Fig. 11.1; Savard et al., 1996; Richard et al., 2000; Six et al., 2005; Koziol et al., 2007; Zhu and Green, 2008). Recently, a LI818 ortholog was found in the moss Physcomitrella (*Physcomitrella patens*) (Alboresi et al., 2008) and the conifer *Picea glauca* (this review, GenBank accession CO250289). No LI818 proteins were found in angiosperms investigated so far.

B. The LHC-Like Protein Family

The LHC-like protein family consists of stress proteins located in thylakoid membranes of cyanobacteria and all photosynthetic eukaryotes investigated so far (Fig. 11.1; Montané and Kloppstech, 2000; Adamska, 2001; Heddad and Adamska, 2002; Engelken et al., 2010). The presence of genes coding for one-helix members of the LHC-like protein family has also been reported for cyanophyges (Bailey et al., 2004; Lindell et al., 2004).

The LHC-like protein family is divided into three-helix ELIPs (Montané and Kloppstech, 2000; Adamska, 2001), two-helix stress-enhanced proteins (SEPs) (Heddad and Adamska, 2000), also called light-harvesting-like (LIL) proteins (Jansson, 1999a), and one-helix proteins (OHPs) (Fig. 11.1; Jansson et al., 2000; Andersson et al., 2003). The latter are called also HLIPs or SCPs in cyanobacteria (Dolganov et al., 1995; Funk and Vermaas, 1999). The recent addition to the LHC-like protein family are high intensity light-inducible LHC-like (LHL4) proteins described from Chlorophyta (Teramoto et al., 2004, 2006). While ELIPs are not detected in thylakoid membranes under low light (LL) conditions and accumulate in response to illumination with HL (Adamska et al., 1992b), significant amounts of SEP and OHP proteins and LHL4 transcripts are present in the absence of light stress but their amounts increase during HL exposure (Heddad and Adamska, 2000; Andersson et al., 2003; Teramoto et al., 2004, 2006). A non-light-harvesting function was proposed for these proteins (Montané and Kloppstech, 2000; Adamska, 2001; Xu et al., 2002a; Teramoto et al., 2004). It is believed that LHC-like protein family members fulfill a protective role within thylakoids under stress conditions either by transient binding of free Chl molecules and preventing the formation of free radicals and/or by acting as sinks for excitation energy (Montané and Kloppstech, 2000; Adamska, 2001; Xu et al., 2002a, 2004; Teramoto et al., 2004). A role in Chl biosynthesis was also proposed for these proteins (Xu et al., 2002a, 2004; Tzvetkova-Chevolleau et al., 2007). These functions are discussed in details in Section V.

1. The Early Light-Induced Protein (ELIP) Subfamily

The occurrence of three-helix ELIPs is restricted to the green algal lineage and land plants (Fig. 11.1; Heddad and Adamska, 2002; Engelken et al., 2010). The expansion in the number of paralogous *ELIP* genes in Physcomitrella was proposed to represent an evolutionary strategy to avoid photooxidative damage in the new terrestrial environment (Rensing et al., 2008).

ELIPs contain two CB motifs located in the first and the third TM helices (Fig. 11.1; Adamska, 2001). The first ELIP was described by Klaus Kloppstech and co-workers to be induced very early during greening of etiolated pea (Meyer and

Kloppstech, 1984) and barley (*Hordeum vulgare*) (Grimm et al., 1989) seedlings. Since *ELIP* transcripts accumulated very rapidly after the transition from dark to light and preceded the accumulation of transcripts for other light-regulated genes, these proteins were named according to this feature. Some years later it was discovered that ELIPs in pea (Adamska et al., 1992a) and barley (Pötter and Kloppstech, 1993) are induced also in mature green plants exposed to light stress. The accumulation of ELIPs was controlled by blue and UVA light absorbed by the cryptochrome 1-type receptor (Adamska et al., 1992a, b; Kleine et al., 2007), occurred in a light intensity-dependent manner (Adamska et al., 1992a, 1993; Heddad et al., 2006; Kleine et al., 2007) and correlated with the degree of photoinactivation and photodamage of PS II reaction centers (Adamska et al., 1992b; Pötter and Kloppstech, 1993; Heddad et al., 2006). Induction of ELIPs during exposure of pea plants to low levels of UVB radiation was also reported (Sävenstrand et al., 2004). Other stress conditions were found to trigger a transient induction of these proteins in various plant species (Adamska, 2001).

Localization studies showed that under HL conditions ELIPs accumulated in the major LHCb antenna system of Arabidopsis (Heddad et al., 2006). Isolation of ELIP from light-stressed pea leaves and analysis of pigments revealed the presence of Chl *a* and lutein (Adamska et al., 1999). However, isolated ELIPs showed a weak excitonic coupling between Chl *a* molecules and a very high lutein content as compared with other CAB proteins (Adamska et al., 1999); the former suggests that ELIPs do not play a significant role in light harvesting.

2. The Stress-Enhanced Protein (SEP) Subfamily

Two-helix SEPs, also called LIL proteins (Jansson, 1999a), contain one CB motif located in the first TM helix (Fig. 11.1; Heddad and Adamska, 2000; Adamska, 2001). SEPs are ubiquitously distributed in photosynthetic eukaryotes, from Glaucophyta, red and green algal lineages to land plants (Fig. 11.1). One *SEP* sequence was found in the red alga *Galdieria sulphuraria*, six in Arabidopsis, and nine in the moss Physcomitrella (Engelken et al., 2010). In the diatoms Thalassiosira and *Phaeodactylum tricornutum* a single and rather divergent *SEP* sequence was found.

SEPs appear to be absent in cyanobacteria. A HLIP/SCP-type protein Hli5 with two predicted TM helices in *Synechococcus* strain OS-B' (Kilian et al., 2008) is not a SEP family member because the CB motif is located in the second TM helix (Engelken et al., 2010). Hli5 is a putative fusion protein of a "coh1" protein (cyanobacterial one-helix protein 1) and a HLIP/SCP (Kilian et al., 2008). Its mRNA level seems to be regulated in a way that differs from those reported for HLIP/SCPs in *Synechococcus* strain OS-B´ and the protein topology within the thylakoid membrane remains unknown (Kilian et al., 2008). For these reasons and due to its restricted taxonomic occurrence in two related *Synechococcus* strains the evolutionary significance of Hli5 seems to be low.

Up to now no detailed functional characterization of SEPs is available. Recently, LIL3 (SEP3) was demonstrated to assemble as a CB protein complex with Chl *a* and protochlorophyllide *a* during deetiolation of barley (Reisinger et al., 2008). Interestingly, in contrast to ELIPs the transcript level for two LIL3 orthologs in Arabidopsis, SEP3-1 (At4g17600) and SEP3-2 (At5g47110), was high in the dark and did not significantly increase during deetiolation as judged from the analysis of microarray data (Winter et al., 2007). Furthermore, the SEP3 transcript level remained unchanged upon transfer of Arabidopsis plants to HL (Jansson, 1999a). Therefore, it was proposed that LIL3 in barley might participate in the delivery of protochlorophyllide *a* to Chl synthase enzyme and transfer of esterified Chl from the enzyme to CB proteins of PS I or PS II (Reisinger et al., 2008).

3. The One-Helix Protein (OHP)/High Light-Induced Protein (HLIP)/Small CAB-Like Protein (SCP) Subfamilies

One-helix CB proteins contain a single TM helix carrying the CB motif (Fig. 11.1; Dolganov et al., 1995; Funk and Vermaas, 1999; Heddad and Adamska, 2000; Adamska, 2001; Andersson et al., 2003). They are widely distributed from cyanobacteria to higher plants (Fig. 11.1; Heddad and Adamska, 2002; Engelken et al., 2010). Two types of OHPs can be distinguished based on

sequence similarity: the OHP1/HLIP/SCP-type (Dolganov et al., 1995; Funk and Vermaas, 1999; Jansson et al., 2000) and the OHP2-type (Andersson et al., 2003). The OHP1/HLIP/SCP-type has a wide taxonomic distribution in cyanophages, cyanobacteria and photosynthetic eukaryota, whereas the OHP2-type is restricted to eukaryotic organisms and has a more pronounced stretch of around 75 evolutionary conserved amino acids.

Our knowledge about plant OHPs is very limited. In Arabidopsis accumulation of OHP1 (Jansson et al., 2000) and OHP2 (Andersson et al., 2003) is triggered by light stress and occurs in a light intensity-dependent manner. While OHP2 was reported to be associated with plant PS I (Andersson et al., 2003), more detailed localization and expression studies exist for cyanobacterial HLIP/SCPs. HLIP/SCPs from *Synechocystis* sp. PCC6803 can be reconstituted with Chl in vitro (Storm et al., 2008), thus confirming their CB nature. Originally, it was shown that *HLIP/SCP* transcripts accumulate in response to HL and therefore these proteins were designated HLIPs (Dolganov et al., 1995). Later it was shown that HLIP/SCPs are also induced under nutrition, cold and osmotic stresses (Funk and Vermaas, 1999; He et al., 2001; Mikami et al., 2002). An important function of these proteins was suggested by the finding that the strongly reduced genome of the HL-adapted ecotype of *Prochlorococcus marinus* strain Med4 encodes at least 24 *HLIP/SCP* genes (Bhaya et al., 2002). Relatives of HLIP/SCPs have been identified in cyanophages infecting marine cyanobacteria (Bailey et al., 2004; Lindell et al., 2004).

Synechocystis sp. PCC 6803 mutants with multiple deletions of *HLIP/SCP* genes are highly sensitive to illumination and show alteration in pigmentation and in the ability to perform thermal energy dissipation via a mechanism of non-photochemical quenching (NPQ) of Chl fluorescence (Havaux et al., 2003; Xu et al., 2004). Possible functions of HLIP/SCPs are discussed in more detail in Section V.

Biochemical as well as immunological studies localized HLIP/SCPs in PS II (Promnares et al., 2006; Yao et al., 2007; Kufryk et al., 2008), although contradictory reports are also existing. It was demonstrated for *Synechocystis* sp. PCC 6803 that HLIA/SCPC and HLIPB/SCPD proteins are located in trimeric PS I complexes, whereas two others, HLIC/SCPB and HLID/SCPE, are associated with the PSAL subunit of PS I or with a partially dissociated PS I complex, respectively (Wang et al., 2008).

4. The High Intensity Light-Inducible LHC-Like (LHL4) Protein Subfamily

The LHL4 protein subfamily is present in Chlorophyta (Fig. 11.1; Teramoto et al., 2004) and *Mesostigma viridae* (this review, GenBank accession EC730580), an early offshoot of the Streptophyta. Four *LHL* genes, *LHL1* through *LHL4*, were identified in Chlamydomonas (Teramoto et al., 2004). While three of these genes encoded members related to ELIPs, the *LHL4* gene encoded a novel protein that forms a separate clade (Fig. 11.2; Teramoto et al., 2004). It was demonstrated that *LHL4* transcripts accumulate in response to HL in a light intensity-dependent manner (Teramoto et al., 2004, 2006). A flavin-based photoreceptor sensing the UVA and blue light was involved in the regulation of the *LHL4* gene expression (Teramoto et al., 2006). A photoprotective function was proposed for LHL4 proteins.

C. The Red Lineage CAB-Like Protein (RedCAP) Family

Recently, the nuclear-encoded RedCAP sequences belonging to the extended LHC protein superfamily were found in the red algal lineage, including Rhodophyta, Heterokontophyta, Haptophyta and Cryptophyta (Fig. 11.1; S. Sturm, J. Engelken, A. Gruber, S. Vugrinec, I. Adamska, P.G. Kroth and J. Lavaud, unpublished). Apart from their three-helix structure with two CB motives, RedCAPs do not share any specific sequence similarity with ELIPs, PSBS or LHCs. Interestingly, second TM helices of RedCAPs are also conserved. No hints toward subcellular location and physiological function of RedCAPs are available.

D. The PSBS Protein Family

The PSBS proteins are unique within the extended LHC protein superfamily because of their four TM helices. Helices I and III with their CB motifs are similar as are helices II and IV (Fig. 11.1).

Similar to the ELIP subfamily the presence of PSBS is restricted to the green algal lineage and land plants (Fig. 11.1, Koziol et al., 2007; Bonente et al., 2008; Engelken et al., 2010). Based on various biochemical studies (Funk, 2001; Dominici et al., 2002; Thidholm et al., 2002) PSBS was considered to be a member of the PS II core complex. However, electron microscopic studies have so far not been successful in revealing its specific location. It was proposed that the PSBS location might be dependent on its oligomerization state (Bergantino et al., 2003). The PSBS monomer/dimer ratio was found to vary depending on the thylakoid lumen pH in a reversible manner, the monomer being the prevalent form at acidic and the dimer at alkaline pH. An association of the PSBS monomer and dimer with the LHC complex and the PS II core complexes, respectively, has been reported (Bergantino et al., 2003). Recently, Horton and co-workers suggested that PSBS might be localized outside the PS II supercomplex and be involved in the macro-organization of the PS II antenna (Horton et al., 2008).

It was demonstrated that PSBS isolated from spinach binds Chl *a* and Chl *b* (Funk et al., 1994, 1995b). However, the pigment-binding characteristics of PSBS are different from those of LHC proteins. Similar to ELIPs a low excitonic coupling was reported for PSBS (Funk et al., 1995b). In vivo, PSBS is stable without binding of pigments and accumulates in etiolated plants (Funk et al., 1995a). In higher plants PSBS has been found to be important to protect PS II against overexcitation by NPQ (Li et al., 2000).

III. Evolutionary Dynamics of the Extended LHC Protein Superfamily

Although a very early origin of cyanobacteria (2.8–3.6 billion years as reviewed by Olson, 2006) was recently challenged, cyanobacteria nevertheless date back to more than two billion years (Rasmussen et al., 2008). The HLIP/SCPs protein subfamily, which is widespread in cyanobacteria, including the early-branching *Gloeobacter* lineage has likely a similar early origin. The SEP subfamily and LHC protein family date back to the common ancestor of Plantae or the common ancestor of the red and green algal lineages, respectively, (Engelken et al., 2010) so they may have been in existence for around 1.5 billion years. Despite the old evolutionary age of many families within the extended LHC protein superfamily, recent findings favor the view of the superfamily as being a highly dynamic and quickly adapting, having evolved a large array of different functions in photosynthesis along the way. For example, certain marine cyanophages are known to exchange *HLIP/SCP* genes with their cyanobacterial hosts and these additional *HLIP/SCP* copies possibly enable them to maintain photosynthetic activity during infection (Lindell et al., 2004). In the long term this has triggered co-evolutionary process between host and viral genomes (Lindell et al., 2007). New protein families and subfamilies including three- and four-helix proteins have evolved repeatedly from a diverse pool of two-helix SEPs (Engelken et al., 2010). Furthermore, these families and superfamilies have undergone significant expansions in different lineages, for example the LHC and LHC-like families in Physcomitrella (Rensing et al., 2008). Sometimes, the genes are located in clusters of tandemly arrayed genes and they can be highly similar. However, the correct annotation of these clusters is challenging and depending on the quality of the genome assembly and it is not straightforward to infer the correct number of extended LHC superfamily members in some cases. In addition, it is well known (for example from the human genome) that the number of gene copies can vary not only between species but also between populations and individuals (Jakobsson et al., 2008), which further complicates the interpretation of such gene family expansions. Interestingly, such differences in the gene supplement can be adaptive, for example through gene dosage effects. These examples display the dynamic nature of the extended LHC protein superfamily, which ultimately helps plant species and populations to adapt to different and changing environments. Why the strongly reduced genome of the *Prochlorococcus marinus* HL-adapted ecotype Med4 encodes at least 24 *HLIP/SCPs* genes or whether the large number of SEPs was an evolutionary adaptation of the moss Physcomitrella to strong fluctuations in light and water supply in its semi-aquatic environment would be some of many interesting questions in this context. In extension, possible differences in *ELIP* gene copy number between different

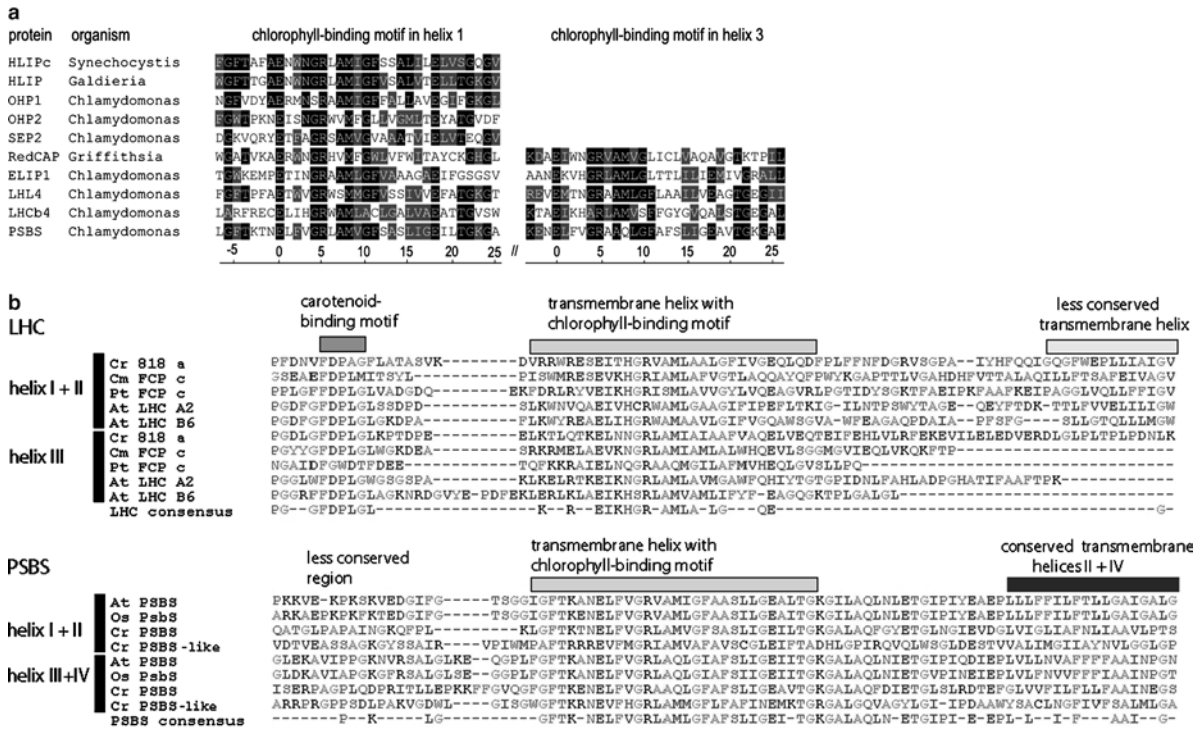


Fig. 11.3. Alignment of conserved sequence motifs within members from the extended LHC protein superfamily. (**a**) Alignment of conserved sequence motifs within representatives selected from each family and subfamily. (**b**), Alignment of conserved sequence motifs from the LHC and PSBS protein families. As representatives of the LHC family, one member of the LI818 subfamily, as well as two members from CAB and CAC protein subfamilies, were used. As representatives of the PSBS family two well-known PSBS, as well as a divergent PSBS-like sequence from Chlamydomonas, were chosen. For each sequence, the first two helices (I+II) and the third and fourth (III+IV, for PSBS) helices are shown and 50% consensus sequences are given. These intra- versus inter-sequence comparisons can be analyzed in a phylogenetic framework, showing that the tandem structure of LHC and PSBS sequences has arisen in independent internal duplication events (Engelken et al., 2010). Abbreviations for species names: At, *Arabidopsis thaliana*; Cm, *Cyanidioschyzon merolae*; Cr, *Chlamydomonas reinhardtii*; Os, *Oryza sativa* var. *japonica*; Pt, *Phaeodactylum tricornutum*.

Physcomitrella populations around the world could be investigated with regard to expression levels of *ELIP* genes/proteins and with regard to photoprotection in different environments.

A phylogenetic network of the extended LHC protein superfamily (Fig. 11.2) based on a 33 amino acid alignment (equivalent to first helix in Fig. 11.3a) of 193 CB sequences was constructed using SplitsTree 4.10 (Huson and Bryant, 2006). The phylogenetic tree reflects (but is not the base for) the proposed classification scheme of CB sequences to different families (Fig. 11.1), for which primary and secondary structure information was used (Engelken et al., 2010). SEPs and HLIP/SCPs are sometimes intermingled, whereas ELIPs, OHP1, OHP2, RedCAP, LHC and PSBS form weak clusters, a trend that is also found with maximum likelihood, posterior probability and even neighbor-joining analyses of the same sequences (data not shown).

IV. Suggested Model for Evolution of the Extended LHC Protein Superfamily

Many attempts have been made to solve the question in which order the different families and subfamilies of the extended LHC protein superfamily have evolved (Green and Pichersky, 1994; Green and Kühlbrandt, 1995; Montané and Kloppstech, 2000; Heddad and Adamska, 2002; Green, 2003;

Six et al., 2005; Jansson, 2006). Fifteen years ago, the existence of a protein with only two TM helices as the ancestor of three- and four-helix proteins was already predicted (Green and Pichersky, 1994; Green and Kühlbrandt, 1995). However, the first two-helix SEPs were described only several years later in Arabidopsis (Heddad and Adamska, 2000). Progressively, the number of CB protein sequences and their diversity increased. Due to the limited number of conserved sequence positions among the different families and despite the strong phylogenetic signal of parts of the CB sequences, molecular phylogenies of the extended LHC protein superfamily do not have enough resolving power to propose an evolutionary model for CB proteins with much certainty.

A "four-helix common ancestor" of LHC/ELIP and PSBS was postulated by Green and Pichersky (1994). According to this scenario the early LHC proteins would have evolved from this intermediate by the loss of the fourth helix and degeneration of their tandem structure except for the carotene-binding and CB motifs. However, such a scenario is rendered unlikely by the apparent absence of such a "four-helix common ancestor" in both glaucophytes and the red algal lineage. Strong and direct evidence for an independent origin of LHC and PSBS from two different internal gene duplication events comes from inter- and intra-sequence comparisons of their conserved CB motives using four-cluster likelihood mapping (Engelken et al., 2010). An informative selection of these CB motifs including flanking regions from different LHC and PSBS sequences is shown (Fig. 11.3b). An in-depth phylogenetic study of these helices provides for the first time direct evidence that LHC and PSBS have likely evolved independently (Engelken et al., 2010). In an extensive database search an unexpected diversity of two-helix SEP sequences in glaucophytes and the red algal lineage as well as in the green lineage was identified (many of these sequences are integrated in the overview network in Fig. 11.2), hereby dating the origin of SEPs to the common ancestor of Plantae and rendering members from this diverse subfamily the most plausible origin of LHC and PSBS (Engelken et al., 2010). Intriguingly, a number of SEPx sequences from Physcomitrella and Selaginella share similarity with the first and third CB motives in PSBS. The second helices are neither conserved within nor across SEPs or LHC and thus do not provide phylogenetic information. Previously, possibly due to the apparent absence of two-helix proteins in glaucophytes and the red lineage, Green (2003) mentioned also the possibility that LHCs may have evolved from duplicated *HLIP/SCP*-like genes with the second helix having arisen by the accumulation of hydrophobic residues. In the light of the currently available sequences and the conjectures given above, this possibility seems now less likely but cannot be fully discarded.

Some years ago, the model of a common origin of prokaryotic Pcb and eukaryotic LHC antenna proteins was proposed (Garczarek et al., 2003). However, this model is weakened considerably by the absence of primary sequence similarity between the two groups and by the lack of a two-helix protein ancestor in cyanobacteria and in plastids of photosynthetic eukaryota.

Based upon comparative genomics and phylogenetic tests of the duplication events leading to LHC and PSBS families a new scenario for the evolution of the extended LHC protein superfamily has been proposed (Fig. 11.4, Engelken et al., 2010). After the origin of the typical CB motif in cyanobacteria followed by the primary endosymbiosis, a stepwise evolution was suggested beginning with the gene transfer of a typical plastid HLIP/SCPs to the nuclear genome within the common ancestor of Plantae and the acquisition of a second TM helix yielding a two-helix SEP homolog. Repeated internal gene duplications of different SEPs and subsequent loss of the last helix yielded the early LHC and RedCAP proteins in the red lineage and the three-helix ELIPs in the green lineage. PSBS evolved independently from a different SEP ancestor in the green lineage and additional subfamilies of the extended LHC protein superfamily evolved in green algae or in plants (Fig. 11.4).

Gene duplication may result from unequal crossing over, retroposition, or segmental duplication (Zhang, 2003). For example, these duplications can result in tandemly arrayed genes, and an internal gene fusion may result in a changed secondary structure of a protein. Interestingly, the tertiary and quartery protein structure of CB proteins with its characteristic twofold symmetry would remain very similar to the

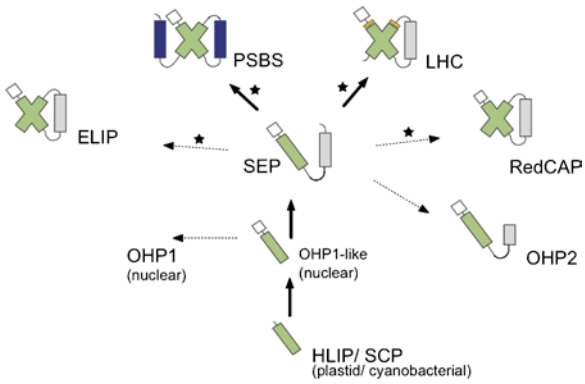


Fig. 11.4. Suggested model for the evolution of the extended LHC protein superfamily (modified from Engelken et al., 2010). Similar to other models also this model suggests a stepwise evolution from the cyanobacterial HLIP/SCPs to the central group of SEPs. The monophyletic group of OHP2 sequences could represent intermediates between nuclear-encoded OHP1/HLIP/SCPs and SEPs, or alternatively and more likely, they could be degenerated SEP sequences with their putative membrane anchors as leftover from the second SEP helix. Currently, there are not many LHL4 sequences available and their evolutionary position remains unclear. A novelty in the model is the postulation of independent origins of three-helix ELIPs and LHCs and four-helix PSBS proteins. Based on character evolution analysis, sequence motifs and the evidence for the independent origins of the PSBS and LHC families, it is proposed that: (**a**) early LHC proteins arose in the red/green ancestor from a SEP and subsequently diversified into different antenna proteins in the red (CAA and CAC) and green (CAB) lineages, (**b**) PSBS arose early in the green lineage from a different SEP, (**c**) ELIPs are neither ancestral to PSBS nor to LHC, but likely evolved independently from still other SEPs, and (**d**) RedCAP sequences evolved independently in the red algal lineage. *Asterisks* mark internal gene duplications.

three-dimensional structure of the original homo- or heterodimer.

V. Functional Aspects of the Extended LHC Protein Superfamily Members: Light Harvesting, Chlorophyll Biosynthesis and/or Photoprotection

Absorption of light and excitation energy transfer to the photosynthetic reaction center is the primary function of the LHC protein family. In addition, LHC family members participate also in photoprotective mechanisms, such as state transitions (Haldrup et al., 2001; Wollman, 2001; Allen, 2003; Rochaix, 2007), NPQ (Horton and Ruban, 2005; Ruban et al., 2007) and the xanthophyll cycle (Demmig-Adams and Adams, 1992).

State transitions balance excitation energy between the two photosystems by redistributing LHCb proteins between PS II and PS I. Such state transitions appear to be present in essentially all species with LHCb proteins and have been described, for example, in the green algae *Chlorella pyrenoidosa* (Bonaventura and Myers, 1969) and Chlamydomonas (Depege et al., 2003; Takahashi et al., 2006; Iwai et al., 2008; Tokutsu et al., 2009) and in Arabidopsis (Wollman, 2001; Bonardi et al., 2005; Tikkanen et al., 2006).

Under conditions of excess light the LHC antenna of plants is rapidly and reversibly switched into a photoprotected quenched state in which potentially harmful absorbed energy is dissipated as heat, a process called NPQ (Ruban et al., 2007). Two different mechanisms of NPQ were proposed. One of them is associated with a twist in the configuration of the LHCb-bound carotenoid neoxanthin as shown for Arabidopsis. Femtosecond transient absorption spectroscopy, performed on purified LHCb in the dissipative state shows that the energy is transferred from Chl *a* to a low-lying carotenoid excited state, identified as one of the two luteins in LHCb (Ruban et al., 2007). An alternative mechanism for NPQ through excess energy transfer from Chl singlet excited state to a Chl-zeaxanthin heterodimer was proposed. The Chl-zeaxanthin heterodimer undergoes than charge separation followed by change recombination and deexcitation (Holt et al., 2005).

In land plants and green algae LHC antenna proteins participate also in the xanthophyll cycle (Jahns et al., 2009). In plants, after the onset of excess light, violaxanthin bound to LHCa (Wehner et al., 2004) and LHCb (Bassi et al., 1993; Ruban et al., 1999; Wehner et al., 2006) is reversibly converted to zeaxanthin via antheraxanthin as an intermediate in a process known as the xanthophyll cycle. Zeaxanthin is a carotenoid that enhances NPQ (Niyogi et al., 1998), establishes a long-term quenching effect by binding to LHC proteins (Dall'Osto et al., 2006), and scavenges reactive oxygen species (Havaux and Niyogi, 1999). A two-component xanthophyll cycle consisting of diadinoxanthin and diatoxanthin is present in diatoms and Haptophyta (Lavaud et al., 2003; Beer et al., 2006; Goss et al., 2006).

Photoprotective functions are proposed for LHC-like family members. ELIPs might participate in transient binding of Chls released from photodamaged CB proteins and thus prevent the formation of singlet oxygen generated by reaction of free excited Chls in triplet state with molecular oxygen. Additionally, a function of ELIPs in NPQ is expected due the high abundance of bound carotenoids (Adamska et al., 1999; Montané and Kloppstech, 2000; Adamska, 2001; Havaux et al., 2003; Hutin et al., 2003). Experiments using the Arabidopsis *chaos* mutant affected in the posttranslational targeting of LHC proteins to the thylakoids and suppressed in the rapid accumulation of ELIPs during HL were supportive of a photoprotective function of these proteins (Hutin et al., 2003). Exposure of such a mutant to HL resulted in leaf bleaching and extensive photooxidative damage. Constitutive expression of *ELIP* genes in the *chaos* mutant before light stress restored the phototolerance of plants (Hutin et al., 2003). While the photoprotective function of ELIPs could be experimentally proven in the *chaos* mutant, the suppression of *ELIP1*, *ELIP2* or both in Arabidopsis did not affect tolerance to photoinhibition and photooxidative stress (Casazza et al., 2005; Rossini et al., 2006), probably due to a high functional redundancy of ELIPs, SEPs and OHPs. More recently, a role of ELIPs in the regulation of Chl biosynthesis was proposed for Arabidopsis (Tzvetkova-Chevolleau et al., 2007).

Similarly to higher plant ELIPs also HLIP/SCPs in cyanobacteria were proposed to prevent the formation of reactive oxygen species by serving as transient carriers of Chl and/or to participate in NPQ (Havaux et al., 2003). HLIP/SCPs were shown to retard the degradation of PS II-associated Chl in *Synechocystis* sp. PCC6803, which is consistent with the proposed involvement of these proteins in PS II re-assembly or/and repair processes by temporarily binding of released Chls (Vavilin et al., 2007). It was also demonstrated that the quadruple *hlip/scp* mutant showed a large decrease in Chl and carotenoid content without accumulation of early Chl intermediates (Xu et al., 2004). A regulation of the tetrapyrrole biosynthesis pathway by HLIP/SCPs as a function of Chl availability was proposed (Xu et al., 2002a, b, 2004). Additionally, HLIP/SCPs seem to trigger trimerization of PS I as recently reported (Wang et al., 2008). It was proposed that HLIP/SCP presence is associated with increased dissipation of absorbed excitation energy by NPQ either by the direct interaction of HLIP/SCPs with photosystems or indirectly by the modification of the organization and/or composition of photosynthetic complexes in mutant cells (Havaux et al., 2003). However, this effect was not very pronounced since only 7.5%-less energy was dissipated as heat in the quadruple *hlip/scp* mutant as compared to wild type.

The essential role of PSBS in the thermal dissipation of excess energy has been demonstrated for higher plants (Li et al., 2000). The main component of NPQ is a feedback regulatory process that depends on the trans-thylakoid δ-pH and is frequently referred to as qE (Briantais et al., 1979). It has been demonstrated that PSBS activates qE by sensing low pH in the thylakoid lumen through two exposed protonable glutamate residues (Li et al., 2004). However, it is still under debate whether PSBS is the actual site of NPQ or only the trigger (Niyogi et al., 2005; Horton et al., 2008).

The crucial role of PSBS in photoprotection and thus in plant fitness during natural conditions becomes evident under highly variable light intensity (Külheim et al., 2002). In constant HL conditions PSBS does not seem to be essential for photoprotection since other photoprotective mechanisms based on antioxidant molecules can overtake this function (Li et al., 2002; Dall'Osto et al., 2006; Golan et al., 2006). Therefore, evolution in limited light environments (e.g. deep water) would not have been dependend on the maintenance or expression of PSBS. This is in agreement with the lack of PSBS expression in Chlamydomonas and other unicellular green algae, and suggests a different mechanism for NPQ in these organisms (Bonente et al., 2008).

The PSBS expression is conserved in plants (*e.g.*, mosses and higher plants) indicating that the mechanism of photoprotection involving this protein is suitable for the highly variable conditions of the land environment. This is supported by the loss of the LI818 subfamily in higher plants, which members were proposed to be involved in NPQ in green algae (Peers et al., 2007) and diatoms (Zhu and Green, 2008).

VI. Closing Remarks and Future Perspectives

In the near future, many more genomes of photosynthetic organisms will become available and new sequencing technologies will also speed up transcriptome analysis of non-model plants, which will result in quickly increasing amounts of available CB sequences. Unfortunately, many evolutionary interesting organisms, *e.g.*, the red algae and the glaucophytes, are so far underrepresented in detailed analysis, although the genome projects of the glaucophyte *Cyanophora paradoxa* and the red algae *Porphyra purpurea* are important exceptions. This development and additional sequencing will further simplify valuable inferences from comparative genomics.

On the other hand, population-genetic studies based on genome-wide sets of markers are becoming increasingly feasible, *e.g.*, in Arabidopsis. Together with phenotypic data, both from the greenhouse as well as from field studies, this could provide interesting insights into recent evolution and adaptive events of plants and algae to different abiotic environments. Still another aspect would be the physiological acclimation of plants and algae to daily and seasonal differences in their environments. For example, gene expression studies of ELIP and PSBS proteins have highlighted mechanisms of acclimation to HL in an alpine evergreen during winter (Zarter et al., 2006). Apart from these important and very basic questions, research on CB proteins could also provide some insights into economically important issues, like stress responses and adaptation of commercial crop species and varieties to drought, extreme temperatures and light stress.

Clearly, at the moment there is a lack of in-depth functional studies of CB proteins apart from LHC or PSBS family members. This is especially apparent in the light of the broad taxonomic distribution of the OHP1/HLIP/SCPs, as well as the SEPs and OHP2. Biochemical, genetic and physiological approaches are needed to elucidate their functions in the photosynthetic machinery.

The extended LHC protein superfamily is an exciting example of protein evolution, showing how a simple and short transmembrane alpha-helix with a CB motif can diversify into numerous families and subfamilies of proteins with increasing complexity and how it hereby can acquire an impressive array of important functions.

Acknowledgments

We would like to thank BR Green (University of British Columbia, Vancouver, Canada) for the valuable comments on this manuscript. CF would like to acknowledge financial support from the Swedish Research Council (VR), the Swedish Energy Agency, the Royal Swedish Academy of Sciences and the Wenner-Gren Foundation and Umeå University. The IA work was supported by research grants from the Deutsche Forschungsgemeinschaft (grants AD92/7-1; AD92/7-2; AD92/7-3) and the Konstanz University grant. JE was kindly supported by a grant from the Volkswagenstiftung, "Förderungsinitiative Evolutionsbiologie".

References

Adamska I (2001) The Elip family of stress proteins in the thylakoid membranes of pro- and eukaryota. In: Aro E-M and Andersson B (eds) Advances in Photosynthesis and Respiration-Regulation of Photosynthesis, Vol 11, pp 487–505. Kluwer, Dordrecht

Adamska I, Kloppstech K and Ohad I (1992a) UV light stress induces the synthesis of the early light-inducible protein and prevents its degradation. J Biol Chem 267: 24732–24737

Adamska I, Ohad I and Kloppstech K (1992b) Synthesis of early light-inducible protein is controlled by blue light and related to light stress. Proc Natl Acad Sci USA 89: 2610–2613

Adamska I, Kloppstech K and Ohad I (1993) Early light-inducible protein in pea is stable during light stress but is degraded during recovery at low light intensity. J Biol Chem 268: 5438–5444

Adamska I, Roobol-Boza M, Lindahl M and Andersson B (1999) Isolation of pigment- binding early light-inducible proteins from pea. Eur J Biochem 260: 453–460

Alboresi AS, Caffarri S, Nogue F, Bassi R and Morosinotto T (2008) *In silico* and biochemical analysis of *Physcomitrella patens* photosynthetic antenna: identification of subunits which evolved upon land adaptation. PLoS ONE 3: e2033

Allen JF (2003) State transitions-a question of balance. Science 299: 1530–1532

Amunts A, Drory O and Nelson N (2007) The structure of a plant photosystem I supercomplex at 3.4Å resolution. Nature 447: 58–63

Andersson U, Heddad M and Adamska I (2003) Light stress-induced one helix protein of the chlorophyll *a/b*-binding family associated with photosystem I. Plant Physiol 132: 811–820

Bailey S, Clokie MR, Millard A and Mann NH (2004) Cyanophage infection and photoinhibition in marine cyanobacteria. Res Microbiol 155: 720–725

Bassi R and Dainese P (1992) A supramolecular light-harvesting complex from chloroplast photosystem-II membranes. Eur J Biochem 204: 317–326

Bassi R, Pineau B, Dainese P and Marquardt J (1993) Carotenoid-binding proteins of photosystem II. Eur J Biochem 212: 297–303

Bathke L, Rhiel E, Krumbein WE and Marquardt J (1999) Biochemical and immunological investigations on the light-harvesting system of the cryptophyte *Rhodomonas* sp.: Evidence for photosystem I specific antenna. Plant Biol 1: 516–523

Beer A, Gundermann K, Beckmann J and Büchel C (2006) Subunit composition and pigmentation of fucoxanthin-chlorophyll proteins in diatoms: evidence for a subunit involved in diadinoxanthin and diatoxanthin binding. Biochemistry 45: 13046–13053

Ben-Shem A, Frolow F and Nelson N (2003) Crystal structure of plant photosystem I. Nature 426: 630–635

Bergantino E, Segalla A, Brunetta A, Teardo E, Rigoni F, Giacometti GM and Szabò I (2003) Light- and pH-dependent structural changes in the PsbS subunit of photosystem II. Proc Natl Acad Sci USA 100: 15265–15270

Bhaya D, Dufresne A, Vaulot D and Grossman A (2002) Analysis of the *hli* gene family in marine and freshwater cyanobacteria. FEMS Microbiol Lett 215: 209–219

Bonardi V, Pesaresi P, Becker T, Schleiff E, Wagner R, Pfannschmidt T, Jahns P and Leister D (2005) Photosystem II core phosphorylation and photosynthetic acclimation require two different protein kinases. Nature 437: 1179–1182

Bonaventura C and Myers J (1969) Fluorescence and oxygen evolution from *Chlorella pyrenoidosa*. Biochim Biophys Acta 189: 366–383

Bonente G, Passarini F, Cazzaniga S, Mancone C, Buia MC, Tripodi M, Bassi R and Caffarri S (2008) The occurrence of the *psbS* gene product in *Chlamydomonas reinhardtii* and in other photosynthetic organisms and its correlation with energy quenching. Photochem Photobiol 84: 1359–1370

Briantais JM, Vernotte C, Picaud M and Krause GH (1979) A quantitative study of the slow decline of chlorophyll *a* fluorescence in isolated chloroplasts. Biochim Biophys Acta 548: 128–138

Büchel C (2003) Fucoxanthin-chlorophyll proteins in diatoms: 18 and 19kDa subunits assemble into different oligomeric states. Biochemistry 30: 333–337

Casazza AP, Rossini S, Rosso MG and Soave C (2005) Mutational and expression analysis of ELIP1 and ELIP2 in *Arabidopsis thaliana*. Plant Mol Biol 58: 41–51

Consoli E, Croce R, Dunlap DD and Finzi L (2005) Diffusion of light-harvesting complex II in the thylakoid membranes. EMBO Rep 6: 782–786

Dall'Osto L, Lico C, Alric J, Giuliano G, Havaux M and Bassi R (2006) Lutein is needed for efficient chlorophyll triplet quenching in the major LHCII antenna complex of higher plants and effective photoprotection in vivo under strong light. BMC Plant Biol 6: 32

Demmig-Adams B and Adams III WW (1992) Photoprotection and other responses of plants to high light stress. Annu Rev Plant Physiol Plant Mol Biol 43: 599–626

Depège N, Bellafiore S and Rochaix JD (2003) Role of chloroplast protein kinase Stt7 in LHCII phosphorylation and state transition in *Chlamydomonas*. Science 299: 1572–1575

Dolganov NA, Bhaya D and Grossman AR (1995) Cyanobacterial protein with similarity to the chlorophyll *a/b* binding proteins of higher plants: evolution and regulation. Proc Natl Acad Sci USA 92: 636–640

Dominici P, Caffarri S, Armenante F, Ceoldo S, Crimi M and Bassi R (2002) Biochemical properties of the PsbS subunit of photosystem II either purified from chloroplast or recombinant. J Biol Chem 277: 22750–22758

Durnford DG, Deane JA, Tan S, McFadden GI, Gantt E and Green BR (1999) A phylogenetic assessment of the eukaryotic light-harvesting antenna proteins, with implications for plastid evolution. J Mol Evol 48: 59–68

Elrad D and Grossman A (2004) A genome's-eye view of the light-harvesting polypeptides of *Chlamydomonas reinhardtii*. Curr Genet 45: 61–75

Engelken J, Brinkmann H and Adamska I (2010) Taxonomic distribution and origins of the extended LHC (light-harvesting complex) antenna protein superfamily. BMC Evol Biol 10: 233

Eppard M, Krumbein WE, von Haeseler A and Rhiel E (2000) Characterization of *fcp4* and *fcp12*, two additional genes encoding light-harvesting proteins of *Cyclotella cryptica* (Bacillariophyceae) and phylogenetic analysis of this complex gene family. Plant Biol 2: 283–289

Funk C (2001) The PsbS protein: A Cab-protein with a function of its own. In: Aro E-M and Andersson B (eds) Advances in Photosynthesis and Respiration-Regulation of Photosynthesis, Vol 11, pp 453–467. Kluwer, Dordrecht, Boston, London

Funk C and Vermaas W (1999) A cyanobacterial gene family coding for single-helix proteins resembling part of the light-harvesting proteins from higher plants. Biochemistry 38: 9397–9404

Funk C, Schröder WP, Green BR, Renger G and Andersson B (1994) The intrinsic 22kDa protein is a chlorophyll-binding subunit of photosystem II. FEBS Lett 342: 261–266

Funk C, Adamska I, Green BR, Andersson B and Renger G (1995a) The nuclear-encoded chlorophyll-binding photosystem II-S protein is stable in the absence of pigments. J Biol Chem 270: 30141–30147

Funk C, Schröder WP, Napiwotzki A, Tjus SE, Renger G and Andersson B (1995b) The PSII-S protein of higher plants:

a new type of pigment-binding protein. Biochemistry 34: 11133–11141

Ganeteg U, Klimmek F and Jansson S (2004) Lhca5-an LHC-type protein associated with photosystem I. Plant Mol Biol 54: 641–651

Garczarek L, Poupon A and Partensky F (2003) Origin and evolution of transmembrane Chl-binding proteins: hydrophobic cluster analysis suggests a common one-helix ancestor for prokaryotic (Pcb) and eukaryotic (LHC) antenna protein superfamilies. FEMS Microbiol Lett 222: 59–68

Golan T, Müller-Moulé P and Niyogi KK (2006) Photoprotection mutants of *Arabidopsis thaliana* acclimate to high light by increasing photosynthesis and specific antioxidants. Plant Cell Environ 29: 879–887

Goss R, Ann Pinto E, Wilhelm C and Richter M (2006) The importance of a highly active and delta pH-regulated diatoxanthin epoxidase for the regulation of the PSII antenna function in diadinoxanthin cycle containing algae. J Plant Physiol 163: 1008–1021

Green BR (2003) The evolution of light-harvesting antennas. In: Green BR and Parson WW (eds) Light-Harvesting Antenna in Photosynthesis, pp. 129–168. Kluwer, Dordrecht

Green BR and Durnford DG (1996) The chlorophyll-carotenoid proteins of oxygenic photosynthesis. Annu Rev Plant Physiol Plant Mol Biol 47: 685–714

Green BR and Kühlbrandt W (1995) Sequence conservation of light-harvesting and stress-response proteins in relation to the three-dimensional molecular structure of LHCII. Photosynth Res 44: 149–162

Green BR and Pichersky E (1994) Hypothesis for the evolution of three-helix chlorophyll *a/b* and chlorophyll *a/c* light-harvesting antenna proteins from two-helix and four-helix ancestors. Photosynth Res 39: 149–162

Grimm B, Kruse E and Kloppstech K (1989) Transiently expressed early light-inducible thylakoid proteins share transmembrane domains with light-harvesting chlorophyll-binding proteins. Plant Mol Biol 13: 583–593

Haldrup A, Jensen PE, Lunde C and Scheller HV (2001) Balance of power: a view of the mechanism of photosynthetic state transitions. Trends Plant Sci 6: 301–301

Havaux M and Niyogi KK (1999) The violaxanthin cycle protects plants from photooxidative damage by more than one mechanism. Proc Natl Acad Sci USA 96: 8762–8767

Havaux M, Guedeney G, He Q and Grossman AR (2003) Elimination of high-light-inducible polypeptides related to eukaryotic chlorophyll *a/b*-binding proteins results in aberrant photoacclimation in *Synechocystis* PCC6803. Biochim Biophys Acta 1557: 21–33

He Q, Dolganov N, Björkman O and Grossman AR (2001) The high light-inducible polypeptides in *Synechocystis* PCC6803. Expression and function in high light. J Biol Chem 276: 306–314

Heddad M and Adamska I (2000) Light stress-regulated two-helix proteins in *Arabidopsis thaliana* related to the chlorophyll *a/b*-binding gene family. Proc Natl Acad Sci USA 97: 3741–3746

Heddad M and Adamska I (2002) The evolution of light stress proteins in photosynthetic organisms. Comp Funct Genom 3: 504–510

Heddad M, Norén H, Reiser V, Dunaeva M, Andersson B and Adamska I (2006) Differential expression and localization of early light-induced proteins in Arabidopsis. Plant Physiol 142: 75–87

Holt NE, Zigmantas D, Valkunas L, Li XP, Niyogi KK and Fleming GR (2005) Carotenoid cation formation and the regulation of photosynthetic light harvesting. Science 307: 433–436

Horton P and Ruban A (2005) Molecular design of the photosystem II light-harvesting antenna: photosynthesis and photoprotection. J Exp Bot 56: 365–373

Horton P, Johnson MP, Perez-Bueno ML, Kiss AZ and Ruban AV (2008) Photosynthetic acclimation: Does the dynamic structure and macro-organisation of photosysten II in higher plant grana membranes regulate light harvesting states? FEBS J 275: 1069–1079

Huson DH and Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23: 254–267

Hutin C, Nussaume L, Moise N, Moya I, Kloppstech K and Havaux M (2003) Early light-induced proteins protect *Arabidopsi*s from photooxidative stress. Proc Natl Acad Sci USA 100: 4921–4926

Iwai M, Takahashi Y and Minagawa J (2008) Molecular remodeling of photosystem II during state transitions in *Chlamydomonas reinhardtii*. Plant Cell 20: 2177–2189

Jahns P, Latowski D and Strzalka K (2009) Mechanism and regulation of the violaxanthin cycle: The role of antenna proteins and membrane lipids. Biochim Biophys Acta 1787: 3–14

Jakobsson M, Scholz SW, Scheet P, Gibbs JR, VanLiere JM, Fung HC, Szpiech ZA, Degnan JH, Wang K, Guerreiro R et al. (2008) Genotype, haplotype and copy-number variation in worldwide human populations. Nature 451: 998–1003

Jansson S (1999a) A guide to the identification of the *Lhc* genes and their relatives in Arabidopsis. Trends Plant Sci 4: 236–240

Jansson S (1999b) A proposal for extending the nomenclature of light-harvesting proteins of the three transmembrane helix type. Plant Mol Biol Rep 17: 221–224

Jansson S (2006) A protein family saga: From photoprotection to light-harvesting (and back?). In: Demmig-Adams B, William W, Adams III WW and Mattoo AK (eds) Photoprotection, photoinhibition, gene regulation and environment, pp. 145–153. Springer, Dordrecht

Jansson S, Andersson J, Kim SJ and Jackowski G (2000) An *Arabidopsis thaliana* protein homologous to cyanobacterial high-light-inducible protein. Plant Mol Biol 42: 345–351

Kereïche S, Kouril R, Oostergetel GT, Fusetti F, Boekema EJ, Doust AB, van der Weij-de Wit CD and Dekker JP (2008) Association of chlorophyll *a/c2* complexes to photosystem I and photosystem II in the cryptophyte *Rhodomonas* CS24. Biochim Biophys Acta 1777: 1122–1128

Kilian O, Steunou, AS, Grossman AR and Bhaya D (2008) A novel two-domain fusion protein in cyanobacteria with similarity to the CAB/ELIP/HLIP superfamily: Evolutionary implications and regulation. Mol Plant 1: 155–166

Kleine T, Kindgren P, Benedict C, Hendrickson L and Strand Å (2007) Genome-wide gene expression analysis reveals a critical role for CRYPTOCHROME1 in the response of Arabidopsis to high irradiance. Plant Physiol 144: 1391–1406

Klimmek F, Sjödin A, Noutsos C, Leister D and Jansson S (2006) Abundantly and rarely expressed *Lhc* protein genes exhibit distinct regulation patterns in plants. Plant Physiol 140: 793–804

Koziol AG and Durnford DG (2008) Euglena light-harvesting complexes are encoded by multifarious polyprotein mRNAs that evolve in concert. Mol Biol Evol 25: 92–100

Koziol AG, Borza T, Ishida K, Keeling P, Lee RW and Durnford DG (2007) Tracing the evolution of the light-harvesting antennae in chlorophyll *a/b*-containing organisms. Plant Physiol 143: 1802–1816

Kufryk G, Hernandez-Prieto MA, Kieselbach T, Miranda H, Vermaas W and Funk C (2008) Association of small CAB-like proteins (SCPs) of *Synechocystis* sp. PCC 6803 with photosystem II. Photosynth Res 95: 135–145

Kühlbrandt W, Wang DN and Fujiyoshi Y (1994) Atomic model of plant light-harvesting complex by electron crystallography. Nature 367: 614–621

Külheim C, Agren J and Jansson S (2002) Rapid regulation of light harvesting and plant fitness in the field. Science 297: 91–93

Lavaud J, Rousseau B and Etienne AL (2003) Enrichment of the light-harvesting complex in diadinoxanthin and implications for the nonphotochemical fluorescence quenching in diatoms. Biochemistry 42: 5802–5808

Lepetit B, Volke D, Szabó M, Hoffmann R, Garab G, Wilhelm C and Goss R (2007) Spectroscopic and molecular characterization of the oligomeric antenna of the diatom *Phaeodactylum tricornutum*. Biochemistry 46: 9813–9822

Li XP, Björkman O, Shih C, Grossman AR, Rosenquist M, Jansson S and Niyogi KK (2000) A pigment-binding protein essential for regulation of photosynthetic light harvesting. Nature 403: 391–395

Li XP, Muller-Moule P, Gilmore AM and Niyogi KK (2002) PsbS-dependent enhancement of feedback de-excitation protects photosystem II from photoinhibition. Proc Natl Acad Sci USA 99: 15222–15227

Li XP, Gilmore AM, Caffarri S, Bassi R, Golan T, Kramer D and Niyogi KK (2004) Regulation of photosynthetic light harvesting involves intrathylakoid lumen pH sensing by the PsbS protein. J Biol Chem 279: 22866–22874

Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F and Chisholm SW (2004) Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. Proc Natl Acad Sci USA 101: 11013–11018

Lindell D, Jaffe JD, Coleman ML, Futschik ME, Axmann IM, Rector T, Kettler G, Sullivan MB, Steen R, Hess WR et al. (2007) Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. Nature 449: 83–86

Liu Z, Yan H, Wang K, Kuang T, Zhang J, Gui L, An X and Chang W (2004) Crystal structure of spinach major light-harvesting complex at 2.72Å resolution. Nature 428: 287–292

Marquardt J and Rhiel E (1997) The membrane-intrinsinc light-harvesting complex of the red alga *Galdieria sulphuraria* (formely *Cyanidium caldarium*). Biochemical and immunological characterization. Biochim Biophys Acta 1320: 153–164

Marquardt J, Lutz B, Wans S, Rhiel E. and Krumbein WE (2001) The gene family coding for the light-harvesting polypeptides of photosystem I of the red alga *Galdieria sulphuraria*. Photosynth Res 68: 121–130

Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, Terry A, Salamov A, Fritz-Laylin LK, Maréchal-Drouard L et al. (2007) The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science 318: 245–250

Meyer G and Kloppstech K (1984) A rapidly light-induced chloroplast protein with a high turnover coded for by pea nuclear DNA. Eur J Biochem 138: 201–207

Mikami K, Kanesaki Y, Suzuki I and Murata N (2002) The histidine kinase Hik33 percives osmotic stress and cold stress in *Synechocystis* sp. PCC6803. Mol Microbiol 46: 905–915

Montané MH and Kloppstech K (2000) The family of light-harvesting-related proteins (LHCs, ELIPs, HLIPs): was the harvesting of light their primary function? Gene 258: 1–8

Naumann B, Stauber EJ, Busch A, Sommer F and Hippler M (2005) N-terminal processing of Lhca3 is a key step in remodeling of the photosystem I-light-harvesting complex under iron deficiency in *Chlamydomonas reinhardtii*. J Biol Chem 280: 20431–20441

Nield J and Barber J (2007) Refinement of the structural model for the photosystem II supercomplex of higher plants. Biophys Biochim Acta 1757: 353–361

Nield L, Kruse O, Ruprecht J, da Fonseca P, Büchel C and Barber J (2000) Three-dimensional structure of *Chlamydomonas reinhardtii* and *Synechococcus elongatus* photosystem II complexes allows for comparison of their oxygen-evolving complex organization. J Biol Chem 275: 27940–27946

Niyogi KK, Grossman AR and Björkman O (1998) Arabidopsis mutants define a central role for the xanthophyll

cycle in the regulation of photosynthetic energy conversion. Plant Cell 10: 1121–1134

Niyogi KK, Li XP, Rosenberg V and Jung HS (2005) Is PsbS the site of non-photochemical quenching in photosynthesis? J Exp Bot 56: 375–382

Oeltjen A, Krumbein WE and Rhiel E (2002) Investigations on transcript sizes, steady-state mRNA concentrations and diurnal expression of genes encoding fucoxanthin chlorophyll *a/c* light-harvesting polypeptides in the centric diatom *Cyclotella cryptica*. Plant Biol 4: 250–257

Olson JM (2006) Photosynthesis in the Archean era. Photosyn Res 88: 109–117

Palenik B, Grimwood J, Aerts A, Rouzé P, Salamov A, Putnam N, Dupont C, Jorgensen R, Derelle E, Rombauts S et al. (2007) The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation. Proc Natl Acad Sci USA 104: 7705–7710

Pascal AA, Liu Z, Broess K, van Oort B, van Amerongen H, Wang C, Horton P, Robert B, Chang W and Ruban A (2005) Molecular basis of photoprotection and control of photosynthetic light-harvesting. Nature 436: 134–137

Peers G, Truong T, Elrad D, Grossman A and Niyogi KK (2007) A non-photochemical quenching mutant of *Chlamydomonas* reveales a role for LI818/LHCSR proteins. Photosynth Res 91: 250

Pötter E and Kloppstech K (1993) Effect of light stress on the expression of early light-inducible proteins in barley. Eur J Biochem 214: 779–786

Promnares K, Komenda J, Bumba L, Nebesarova J, Vacha F and Tichy M (2006) Cyanobacterial small chlorophyll-binding protein ScpD (HliB) is located on the periphery of photosystem II in the vicinity of PsbH and CP47 subunits. J Biol Chem 281: 32705–32713

Rasmussen B, Fletcher IR, Brocks JJ and Kilburn MR (2008) Reassessing the first appearance of eukaryotes and cyanobacteria. Nature 455: 1101–1104

Reisinger V, Ploscher M and Eichacker LA (2008) Lil3 assembles as chlorophyll-binding protein complex during deetiolation. FEBS Lett 582: 1547–1551

Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud PF, Lindquist EA, Kamisugi Y et al. (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. Science 319: 64–69

Richard C, Ouellet H and Guertin M (2000) Characterization of the Li818 polypeptide from the green unicellular alga *Chlamydomonas reinhardtii*. Plant Mol Biol 42: 303–316

Rochaix JD (2007) Role of thylakoid protein kinases in photosynthetic acclimation. FEBS Lett 581: 2768–2775

Rossini S, Casazza AP, Engelmann EC, Havaux M, Jennings RC and Soave C (2006) Suppression of both ELIP1 and ELIP2 in *Arabidopsis* does not affect tolerance to photoinhibition and photooxidative stress. Plant Physiol 141: 1264–1273

Ruban AV, Lee PJ, Wentworth M, Young AJ and Horton P (1999) Determination of the stoichiometry and strength of binding of xanthophylls to the photosystem II light harvesting complexes. J Biol Chem 274: 10458–10465

Ruban AV, Berera R, Ilioaia C, van Stokkum IH, Kennis JT, Pascal AA, van Amerongen H, Robert B, Horton P and van Grondelle R (2007) Identification of a mechanism of photoprotective energy dissipation in higher plants. Nature 450: 575–578

Savard F, Richard C and Guertin M (1996) The *Chlamydomonas reinhardtii* LI818 gene represents a distant relative of the *cab I/II* genes that is regulated during the cell cycle and in response to illumination. Plant Mol Biol 32: 461–473

Sävenstrand H, Olofsson M, Samuelsson M and Strid Å (2004) Induction of early light-inducible protein gene expression in *Pisum sativum* after exposure to low levels of UV-B irradiation and other environmental stresses. Plant Cell Rep 22: 532–536

Six C, Worden AZ, Rodriguez F, Moreau H and Partensky F (2005) New insights into the nature and phylogeny of prasinophyte antenna proteins: *Ostreococcus tauri*, a case study. Mol Biol Evol 22: 2217–2230

Standfuss J, Terwisscha van Scheltinga AC, Lamborghini M and Kühlbrandt W (2005) Mechanisms of photoprotection and nonphotochemical quenching in pea light-harvesting complex at 2.5Å resolution. EMBO J 24: 919–928

Storm P, Hernandez-Prieto MA, Eggink LL, Hoober JK and Funk C (2008) The small CAB-like proteins of *Synechocystis* sp. PCC 6803 bind chlorophyll: In vitro pigment reconstitution studies on one-helix light-harvesting-like proteins. Photosynth Res 98: 479–488

Takahashi H, Iwai M, Takahashi Y and Minagawa J (2006) Identification of the mobile light-harvesting complex II polypeptides for state transitions in *Chlamydomonas reinhardtii*. Proc Natl Acad Sci USA 103: 477–482

Teramoto H, Itoh T and Ono T (2004) High-intensity-light-dependent and transient expression of new genes encoding distant relatives of light-harvesting chlorophyll-*a/b* proteins in *Chlamydomonas reinhardtii*. Plant Cell Physiol 45: 1221–1232

Teramoto H, Ishii A, Kimura Y, Hasegawa K, Nakazawa S, Nakamura T, Higashi S, Watanabe M, Ono TA (2006) Action spectrum for expression of the high intensity light-inducible Lhc-like gene *Lhl4* in the green alga *Chlamydomonas reinhardtii.* Plant Cell Physiol 47: 419–425

Thidholm E, Lindström V, Tissier C, Robinson C, Schröder W and Funk C (2002) Novel approach reveals localization and assembly pathway of the PsbS and PsbW proteins into the photosystem II dimer. FEBS Lett 513: 217–222

Tikkanen M, Piippo M, Suorsa M, Sirpiö S, Mulo P, Vainonen J, Vener AV, Allahverdiyeva Y and Aro EM (2006) State transitions revisited-a buffering system for dynamic low light acclimation of Arabidopsis. Plant Mol Biol 62: 779–793

Tokutsu R, Teramoto H, Takahashi J, Ono TA and Minagawa J (2004) The light-harvesting complex of photosystem

I in *Chlamydomonas reinhardtii*: protein composition, gene structures and phylogenic implications. Plant Cell Physiol 45: 138–145

Tokutsu R, Iwai M and Minagawa J (2009) CP29, a monomeric light-harvesting complex II protein, is essential for state transitions in *Chlamydomonas reinhardtii*. J Biol Chem 284: 7777–7782

Tzvetkova-Chevolleau T, Franck F, Alawady AE, Dall'Osto L, Carriére F, Bassi R, Grimm B, Nussaume L and Havaux M (2007) The light stress-induced protein ELIP2 is a regulator of chlorophyll synthesis in *Arabidopsis thaliana*. Plant J 50: 795–809.

Vavilin D, Yao D and Vermaas W (2007) Small Cab-like proteins retard degradation of photosystem II-associated chlorophyll in *Synechocystis* sp. PCC 6803: kinetic analysis of pigment labeling with ^{15}N and ^{13}C. J Biol Chem 282: 37660–37668

Wang Q, Jantaro S, Lu B, Majeed W, Bailey M and He Q (2008) The high light-inducible polypeptides stabilize trimeric photosystem I complex under high light conditions in *Synechocystis* PCC 6803. Plant Physiol 147: 1239–1250

Wehner A, Storf S, Jahns P and Schmid VH (2004) De-epoxidation of violaxanthin in light-harvesting complex I proteins. J Biol Chem 279: 26823–26829

Wehner A, Grasses T and Jahns P (2006) De-epoxidation of violaxanthin in the minor antenna proteins of photosystem II, LHCB4, LHCB5 and LHCB6. J Biol Chem 281: 21924–21933

Winter D, Vinegar B, Hardeep N, Ammar R, Wilson GV and Provart NJ (2007) An "electronic fluorescent pictograph" browser for exploring and analyzing large-scale biological data sets. PLOS One 2: e718

Wolfe GR, Cunningham FX, Durnford D, Green BR and Gantt E (1994) Evidence for a common origin of chloroplasts with light-harvesting complexes of different pigmentation. Nature 367: 566–568

Wollman FA (2001) State transitions reveal the dynamics and flexibility of the photosynthetic apparatus. EMBO J 20: 3623–3630

Xu H, Vavilin D, Funk C and Vermaas W (2002a) Small Cab-like proteins regulating tetrapyrrole biosynthesis in the cyanobacterium *Synechocystis* sp. PCC 6803. Plant Mol Biol 49: 149–160

Xu H, Vavilin D and Vermaas W (2002b) The presence of chlorophyll *b* in *Synechocystis* sp. PCC 6803 disturbs tetrapyrrole biosynthesis and enhances chlorophyll degradation. J Biol Chem 277: 42726–42732

Xu H, Vavilin D, Funk C and Vermaas W (2004) Multiple deletions of small Cab-like proteins in the cyanobacterium *Synechocystis sp*. PCC 6803: consequences for pigment biosynthesis and accumulation. J Biol Chem 272: 27971–27979

Yao D, Kieselbach T, Komenda J, Promnares K, Prieto MA, Tichy M, Vermaas W and Funk C. (2007) Localization of the small CAB-like proteins in photosystem II. J Biol Chem 282: 267–276

Zarter CR, Adams III WW, Ebbert V, Adamska I, Jansson S and Demmig-Adams B (2006) Winter acclimation of PsbS and related proteins in the evergreen *Arctostaphylos uva-ursi* as influenced by altitude and light environment. Plant Cell Environ 29: 869–878

Zhang J (2003) Evolution by gene duplication: an update. Trends Ecol Evol 18: 292–298

Zhang Y, Chen M, Zhou BB, Jermiin LS and Larkum AWD (2007) Evolution of the inner light-harvesting antenna protein family of cyanobacteria, algae and plants. J Mol Evol 64: 321–331

Zhu SH and Green BR (2008) Light-harvesting and photoprotection in diatoms: identification and expression of LI818-like proteins. In: Allen JF, Gantt E, Golbeck JH and Osmond B (eds) Photosynthesis. Energy from the Sun, 14th International Congress on Photosynthesis, Chapter 6, pp 261–264. Springer, Dordrecht

Chapter 12

The Evolution of Type I Reaction Centers: The Response to Oxygenic Photosynthesis

Bharat Jagannathan and Gaozhong Shen
Department of Biochemistry and Molecular Biology,The Pennsylvania State University, University Park, PA, USA

John H. Golbeck*
Department of Biochemistry and Molecular Biology and Department of Chemistry, The Pennsylvania State University, University Park, PA, USA

*Author for correspondence, e-mail: jhg5@psu.edu

R.L. Burnap and W.F.J. Vermaas (eds.), Functional Genomics and Evolution of Photosynthetic Systems, Advances in Photosynthesis and Respiration 33, pp. 285–316, DOI 10.1007/978-94-007-1533-2_12,

Summary

Recent advances in the understanding of homodimeric Type I reaction centers in contemporary anaerobic phototrophs are used to advance a scenario in the evolution of Photosystem I. The transition from anaerobic to aerobic photosynthesis 2.7 billion years ago was accompanied by profound changes on the acceptor side of the ancestral Type I reaction center. At the advent of oxygen evolution, the mobile bacterial dicluster ferredoxin, which initially served to transfer electrons from the F_X cluster to target redox proteins, became highly vulnerable to oxidative denaturation. In response, an exceedingly tight binding interface developed between the bacterial dicluster ferredoxin and the Type I reaction center core, thereby making it possible for the F_A and F_B iron-sulfur clusters to survive in the presence of oxygen. The need for an alternative, mobile electron carrier led to the evolution of an oxygen-insensitive [2Fe-2S] ferredoxin. The recruitment of a PsaD-like protein to generate a binding site for ferredoxin, by necessity, broke the perfect C_2-symmetry of the homodimeric reaction center and provided the selective pressure that led to its differentiation into separate PsaA and PsaB polypeptides. The Photosystem I reaction center nevertheless retained a high degree of C_2-symmetry surrounding the F_X cluster, resulting in a new problem of how to dock the bacterial dicluster ferredoxin in one of the two possible orientations on the PsaA/PsaB heterodimer. Ultimately, all of the information required for preferential docking was coded in the amino acid sequence of the dicluster ferredoxin that became PsaC. The differentiation and inversion of the redox potentials of F_A and F_B ensured that the electron was preferentially transferred to the [2Fe-2S] ferredoxin in the supersaturated solution of the oxygen-evolving cell.

I. Introduction

Life is abundant on Earth because of sunlight, the powerhouse for oxygenic photosynthesis in plants and cyanobacteria. In photosynthesis, the energy of a photon is converted into chemical bonds through the metabolic assimilation of carbon dioxide into carbohydrate, generating molecular oxygen as a byproduct. This process requires the serial participation of two multi-subunit membrane protein complexes, Photosystem I (PSI) and Photosystem II (PSII), which serve as the energy conversion machinery for driving electrons against a steep thermodynamic gradient at the expense of a photon. Indeed, the cooperation of a Type I (iron-sulfur containing) reaction center with a Type II (quinone containing) reaction center was the seminal event in the development of oxygenic photosynthesis. The ability to split the water molecule ultimately changed the redox balance on Earth and led to the appearance of advanced, eukaryotic forms of life. The distinctive feature of PSI is the low redox potential of the iron-sulfur clusters, which participate in the light-driven reduction of $NADP^+$ on the cytoplasmic (stromal) side of the thylakoid membrane (Golbeck, 1992, 1994). Because of the sensitivity of iron-sulfur clusters to molecular oxygen, the evolution of an oxygen-tolerant PSI reaction center needed to occur simultaneously with the onset of water splitting in PSII. How this may have occurred is the central topic of this article.

In recent years, significant progress has been made in understanding the functional genomics and biochemistry of photosynthetic organisms that contain Type I reaction centers. PSI has evolved from a simple one-gene product, homodimeric Type I reaction center into a complex heterodimeric Type I reaction center that generates the most negative redox potential known in biology. Achievements in obtaining high resolution crystal structures of PSI from a cyanobacterium (Jordan et al., 2001) and a higher plant (Amunts et al., 2007) have provided detailed structural information as well as snapshots on the progression from a prokaryotic phototroph to a eukaryotic phototroph (Blankenship, 2001). Much contemporary work on the evolution of photosynthesis

Abbreviations: A_0 – PSI primary electron acceptor; BChl – bacteriochlorophyll; Chl – chlorophyll; EPR – electron paramagnetic resonance; Fx – interpolypeptide [4Fe4S] cluster; P_{700} – PSI primary electron donor: PSI – photosystem I; PSII – photosystem II; PsaA/B – heterodimeric PSI reaction center core protein in cyanobacteria, algae, and plants; PsaC – FA/FB–containing protein; PscA – homodimeric reaction center core protein in green sulfur bacteria; PshA – homodimeric reaction center core protein in heliobacteria

adopts a bioinformatics approach, wherein inferences for structure and function are drawn from sequence comparisons and phylogenetic trees (Baymann et al., 2001; Lockhart et al., 1996; Mulkidjanian and Junge, 1997; Olson and Pierson, 1987; Raymond et al., 2002; Vermaas, 2002; Xiong et al., 2000). Other studies employ common structural themes and redox motifs that are present in all photochemical reaction centers as markers to provide insight into the evolution of photosynthesis (Allen and Williams, 1998; Golbeck, 1993; Heathcote et al., 2002; Nitschke and Rutherford, 1991). Here, we add a biochemical and biophysical perspective to the issue of evolutionary change. As a result, we hope to offer something new to the story of the evolution of Type I reaction centers. While many studies have dealt with the evolution of the membrane-bound reaction center core, particular attention is paid here to the membrane extrinsic subunits and their response to the advent of oxygenic photosynthesis.

A. From Chemotrophic Metabolism to Anaerobic Phototrophy

Before considering the evolution of PSI, it is instructive to describe the geological conditions that led to the onset of photosynthesis. According to current estimates, Earth is estimated to have formed at 4.6 Ga.[1] The ensuing Hadean Earth (4.6–3.8 Ga) was most certainly aqueous, reducing, and anaerobic. As early as 10–20 million years after the Moon-forming impact, an event that occurred 40–50 million years after the accretion of the solar system, the planet may already have been generating biochemical precursors (Zahnle et al., 2007). Through time, this led to a rich pre-biotic environment that gave rise to the first living organisms in a process that is entirely opaque. By exploiting every abundant redox pair in the environment, early life forms would have extracted Gibbs free energy from purely chemotrophic metabolic processes. The onset of photosynthesis would have occurred at the exhaustion of these redox pairs, leaving sunlight as the largest unexploited source of energy.

Because of the complexity of the machinery necessary to convert a photon into a chemical bond, and because evolution is largely incremental, a precursor to the photosynthetic reaction centers must have existed. Mulkidjanian and Junge (1997) have suggested that this precursor was a pigment-carrying protein whose function in primordial cells was to serve as a UV protector. The core idea is that in the absence of a UV-absorbing ozone layer, early prokaryotic cells would have been restricted to a sub-surface euphotic zone in an exclusively aqueous environment. The presence of a UV-shielding protein would have enabled these organisms to survive in a broad range of environments. This protein was proposed to have its origin in porphyrin binding subunits similar, perhaps, to high-light-inducible proteins (HLIPs) in contemporary cyanobacteria (Mulkidjanian et al., 2006). A larger protein containing multiple membrane-spanning subunits could have arisen from gene duplication events within these simpler pigment-binding proteins. In the absence of an ozone shield, the UV protecting protein would have disposed of harmful UV and visible light by converting an excited singlet state into thermal energy. However, an excited singlet state, although fleeting, does represent stored energy, and it is not difficult to imagine how a protein may have become altered so as to produce a charge-separated state between two adjacent pigment molecules. Protein dimerization may have been important in this transition (see Nitschke et al., 1996) because it would have allowed the juxtaposition of two chlorophylls, thereby producing a 'special pair' that could trap the resulting cation from the initial charge separated state (see Müller et al., 2003). A protein dimer is also reasonable in a geometric point of view because the simplest way to create a three-dimensional object is to abut two 2-dimensional surfaces (Jagannathan and Golbeck, 2009a). The product of a single gene could thus create a well-protected interface between two proteins in which photochemical charge separation would occur in a low dielectric medium. Subsequently, the problem would have become one of stabilizing the fleeting charge-separated state over longer periods of time – a necessary precondition for carrying out solution chemistry with already-existing redox proteins. According to Mulkidjanian and Junge (1997), the progressive loss of UV-absorbing amino acids and the subsequent acquisition of redox cofactors could have led to the means to stabilize the charge-separated state. Because of their ability to form spontaneously in the presence of iron, sulfide and

[1] Years before present (1950), as in Ga, giga-annum (10^9 years) and Ma, mega-annum (10^6 years).

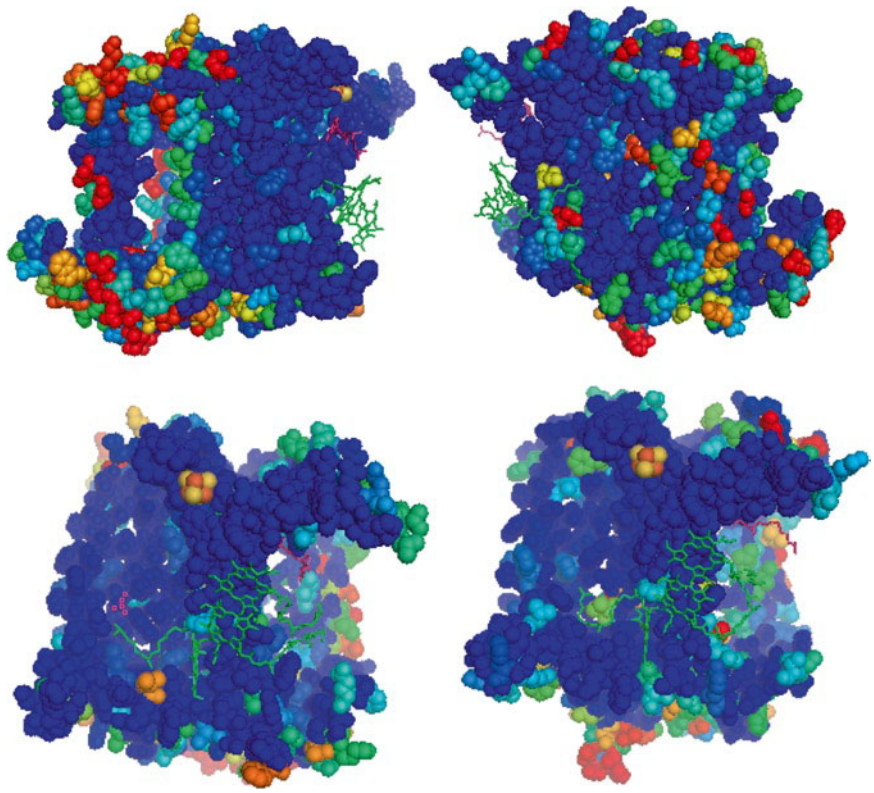


Fig. 12.1. Separated views of the PsaB protein (*left*; *top*, *side view*; *bottom*, *face-on view*) as spheres, showing the three chlorophyll molecules (P_{700}-B, A_B, A_{0B}) and phylloquinone (A_{1B}) in stick form; and the PsaA protein as spheres (*right*; *top*, *side view*; *bottom*, *face-on view*) showing the three chlorophyll molecules (P_{700}-A, A_A, A_{0A}) and phylloquinone (A_{1A}) in stick form. F_X is shown common to PsaB (*left*) and PsaA (*right*). The amino acids are color coded according to the degree of similarity in an amino acid sequence alignment using the PsaB sequences from 51 cyanobacterial, algal, and plant species: the colors represent a continuum between invariant residues (*blue*) and highly variable residues (*red*). Note that the residues closest to the electron transfer cofactors tend to be more highly conserved than the residues near the periphery of the reaction center; this is especially the case for the PsaB protein. Note also that when PsaA and PsaB form a heterodimer, a protected interface is formed between the two proteins wherein photochemical charge separation could occur in an exceedingly low dielectric medium. This also allows the juxtaposition of the two chlorophylls (P_{700}-A, P_{700}-B), thereby producing a 'special pair' that either initiates the charge separation or traps the resulting cation from the initial charge separated state (see Müller et al., 2003).

an external thiolate (DePamphilis et al., 1974; Martin and Russell, 2003; Wächterhäuser, 1992) iron-sulfur clusters could have been important in this process. Iron-sulfur cubanes can be exchanged into cysteine-containing proteins through a facile thiol ligand exchange reaction (Que et al., 1974). Notably, it would take only two appropriately positioned cysteines to participate in the ligation of an iron-sulfur cluster provided the cluster is located at the interface of the protein dimer, as is the F_X cluster in Type I reaction centers (see Fig. 12.1). The Gibbs free energy change for insertion of an iron-sulfur cluster into an apoprotein is favorable; although the enthalpic contribution is negligible due to the high degree of similarity in the Fe-thiol coordination bonds, the entropic contribution is significant due to the release of individual thiolate ligands and their substitution by the cysteine-containing polypeptide. The initial photochemical machinery need not have a high quantum yield because even a small supplementation of a cellular reductant such as reduced ferredoxin would have conferred a selective advantage to an organism ever on the hunt for increasingly scarce chemical reductants. Given the enormous evolutionary advantage offered by gaining the supplemental energy source of sunlight, selective pressure would have resulted in an inexorable increase in the quantum yield through the mechanisms of spontaneous mutation, gene duplication, domain swap-

ping, and natural selection. In this scenario, photosynthesis would have evolved rapidly after the development of the first prototypical charge separation machinery, with photochemistry carried out by a pigment and iron-sulfur cluster-containing reaction center.

B. The Primordial Photosynthetic Reaction Center

There exists today six bacterial phyla with members capable of chlorophyll-based phototrophy: Firmicutes (e.g., *Heliobacterium modesticaldum*), Chloroflexi (e.g., *Chloroflexis aurantiacus*), Chlorobi (e.g., *Chlorobium tepidum*), Proteobacteria, Cyanobacteria (e.g. *Synechocystis* sp. PCC 6803), and the recently added Acidobacteria (*Candidatus Chloracidobacterium thermophilum*) (Bryant et al., 2007). Proteobacteria can be sub-classified as α-Proteobacteria (e.g., *Rhodopseudomonas palustris*) β-Proteobacteria (e.g., *Rubrivivax gelatinosum*), and γ-Proteobacteria (e.g., *Chromatium vinosum*). Photosynthetic reaction centers within these bacteria come in two varieties: Type I, which contain an iron-sulfur cluster as the terminal acceptor, and Type II, which contain a mobile quinone as the terminal acceptor. Type I reaction centers are found in the phototrophic members of Firmicutes, Chlorobi, and Acidobacteria. Type II reaction centers are found in the phototrophic members of Chloroflexi, α-Proteobacteria, β-Proteobacteria, and γ-Proteobacteria. Both Type I and Type II reaction centers co-exist in cyanobacteria, which are the only prokaryotes capable of oxygenic photosynthesis. We address the question of whether Type I or Type II reaction centers evolved first by briefly summarizing the relevant literature[2] prior to describing the environmental composition of the early Earth.

Over the last 15 years, arguments have been put forward for the origin of photosynthesis among photosynthetic members of the Firmicutes (Gupta, 2003; Xiong et al., 2000), Chlorobi (Büttner et al., 1992), Chloroflexi (Pierson, 1994), Proteobacteria (Xiong et al., 2000), and Cyanobacteria (Mulkidjanian et al., 2006). A few of these studies are particularly noteworthy and are briefly mentioned here. Vermaas (1994) was among the first to point out that a 46 residue domain of the reaction center from *Heliobacillus mobilis* shows a strong resemblance (33% identity, 72% similarity), and that the N-terminal half of the of the reaction center from *H. mobilis* shows a reasonable similarity (33% identity over 232 residues) to regions of the PSII antenna complex, CP47. It was suggested that the heliobacterial reaction center and an antenna protein of PSII are related and hence likely to share a common ancestor. Xiong et al. (2000) analyzed a number of photosynthetic genes, including those involved in bacteriochlorophyll, carotenoid and electron transfer cofactor biosynthesis, from the green sulfur bacterium *Chlorobium tepidum* and the green non sulfur bacterium *Chloroflexus auranthacus*. They concluded that heliobacteria are common to the last common ancestor of all oxygenic photosynthetic lineages and that green sulfur bacteria and green nonsulfur bacteria are each other's closest relatives. They also proposed purple bacteria as the earliest emerging photosynthetic lineage as indicated by the phylogenetic tree constructed for the (bacterio)chlorophyll synthesis enzymes. However, the methodology employed by these authors has since been contested (see refs. Mix et al., 2005; Mulkidjanian et al., 2006), hence their derived evolutionary relationships should be evaluated with caution. Gupta and co-workers (Gupta (2003); Gupta et al., (1999)) used certain signature sequences such as conserved inserts or deletions (termed indels) as genetic markers for the analysis of the evolutionary relationships of photosynthetic taxa, and concluded that heliobacteria are the most ancestral of the photosynthetic lineages. Mix et al. (2005) limited their phylogenetic analysis to accessory antenna domains and, using maximum likelihood, parsimony, and neighbor joining methods, concluded that PSII core antenna proteins (PsbC, PsbB) arose within cyanobacteria from duplications of Type I reaction center-associated core antenna domains. Additionally, they proposed that the accessory antenna proteins (IsiA, PcbA, PcbC) arose from duplications of PsbB. Their evolutionary history of Type I reaction centers included an initially homodimeric reaction center core, which

[2]The literature survey presented here is not exhaustive; rather, we put forward an evolutionary scenario that is consistent with the biochemical and biophysical evidence from contemporary phototrophs that contain Type I reaction centers. As a result, we will not be delving into the controversies and opposing views of this active field.

was transmitted to green sulfur bacteria, heliobacteria, and the ancestor of cyanobacteria. Their phylogenetic tree is unrooted, but the favored scenario places the origin near the fork that separates heliobacteria and green sulfur bacteria. Considering the sum of the evidence, the original photosynthetic reaction center may have been a homodimeric Type I reaction center similar to that present in contemporary heliobacteria (Blankenship, 1992, 2002).

It is notable that Type II reaction centers do not appear on this tree (Mix et al., 2005); their relationship with Type I reaction centers and with PSII antenna and accessory antenna domains was left an open question. With the availability of X-ray crystal structures of cyanobacterial and plant PSI (Amunts et al., 2007; Ben-Shem et al., 2003; Jordan et al., 2001), and PSII (Ferreira et al., 2004; Müh et al., 2008; Zouni et al., 2001), a high degree of structural similarity is found in the reaction center cores, in which both PSI and PSII share a strikingly similar structural blueprint, especially in cofactor composition and arrangement (Fromme, 1996; Fromme et al., 1996; Schubert et al., 1998). This has led to speculation that Type I and Type II reaction centers are derived from a single ancestor (Blankenship, 1992; Raymond and Blankenship, 2006; Schubert et al., 1998).

Heliobacteria are members of the low G+C gram-positive Firmicutes group (Sattley et al., 2008). Considering their simple subunit composition and limited number of antenna chlorophyll molecules, heliobacteria contain the most rudimentary of the photochemical reaction centers that have been characterized to date (Heinnickel and Golbeck, 2007; Oh-oka, 2007; Sattley et al., 2008). Heliobacteria use a close evolutionary precursor to chlorophyll *a* (Chl *a*), bacteriochlorophyll *g* (BChl *g*), as the primary electron donor and light-harvesting pigment (Trost and Blankenship, 1989). At latest estimate, 22–30 BChl *g* molecules are bound to the PshA/PshA homodimer, which also harbors the F_X iron-sulfur cluster (Heinnickel et al., 2006). No small BChl *g*-binding proteins that function as the antenna system have yet been found. The light harvesting system in heliobacteria is therefore not particularly advanced, especially compared to that of green (sulfur or nonsulfur) bacteria and purple bacteria.

If Type I reaction centers are the most ancient of the photosystems, the question arises as to the identity of the earliest electron donor(s). Olson (2006) has cited evidence from the carbon isotope composition of the 3.8 Ga Isua Supercrustal Belt in Greenland (Tian et al., 2005) that the earliest reductant for photosynthesis may have been H_2. However, the supply of H_2 in the atmosphere would have eventually been depleted as a result of the generation of methane by methanogens. The presence of methane allows H_2 to escape into outer space at rates orders of magnitude faster than in its absence (Catling et al., 2001). H_2S represents another possible reductant (Olson and Pierson, 1987). A biological source for the mineral barite ($BaSO_4$) is implied in deposits of the 3.2 Ga Fig Tree Group of the Swaziland Sequence in South Africa due to the small difference in $\delta^{34}S$ between barite and sulfide. Oxidation of sulfide to sulfate by Chromatium has been shown to result in very little isotopic fractionation (summarized in Olson (2006)). Ferrous iron is a third possible reductant (Ehrenreich and Widdel, 1994; Heising et al., 1999; Heising and Schink, 1998; Olson (2001); its utilization around 3.0 Ga would have occurred as the strong reductants H_2 and H_2S would have faced depletion (Olson, 2006). Given the presence of reduced compounds such as H_2, H_2S or Fe^{2+}, a Type I rather than Type II reaction center makes for the most sense because of the need for linear rather than cyclic electron transfer to a low potential reductant such as ferredoxin.

Oxygenic photosynthesis has long been proposed to have evolved from the union of preexisting Type I and Type II reaction centers within a cyanobacterium (Blankenship, 1992; Olson and Blankenship, 2004; Raymond and Blankenship, 2006). More recently, Mulkidjanian and coworkers have put forth a counterproposal that cyanobacteria are not only the most ancient phototrophs but that they are the very origin of both Type I and Type II reaction centers (Mulkidjanian et al., 2006). Their approach was to analyze 15 cyanobacterial genomes for the purpose of identifying 'core cyanobacterial clusters of orthologous groups' of proteins, most of which turned out to be involved in basic cellular functions. They found that the vast majority of cyanobacterial photosynthetic genes had no detectable homologs in anoxygenic phototrophic bacteria, and conversely, that anoxygenic photosynthetic bacteria possessed very few photosynthetic genes shared by cyanobacteria and plants. From this evidence,

they proposed that both Type I and Type II reaction centers originated in an ancestral cyanobacterial lineage under the selective pressures of UV light and depletion of electron donors. Modern cyanobacteria would have inherited their tandem photosynthetic apparatus from these ancestral phototrophs. Lateral gene transfer might have been responsible for the distribution of the photosynthetic reaction center to other phyla (see also Blankenship, 1992; Mix et al., 2005; Olson and Blankenship, 2004). Heliobacteria and green sulfur bacteria would have obtained their Type I reaction center by way of lateral gene transfer well before the Type I reaction center became heterodimeric. In a similar manner, Proteobacteria and Chloroflexus would have acquired a Type II reaction center well before it was capable of oxidizing water.

This analysis, however, begs the larger question of why, in the absence of oxygen evolution, a protocyanobacterium would have contained both types of reaction center. Allen and Martin have proposed a solution by assigning independent tasks for the two types of reaction centers (Allen and Martin, 2007). The Type I reaction center would have carried out linear electron transfer, while the Type II reaction center would have carried out cyclic electron transfer. The former would resemble the Type I reaction center in contemporary green sulfur bacteria, which oxidizes H_2S and reduces ferredoxin, while the latter would resemble the Type II reaction center in contemporary purple non-sulfur bacteria, which generates a light-driven proton gradient via cyclic electron transfer. A regulatory redox switch was proposed, which would have selected between the operation of the Type I reaction center or the Type II reaction center as environmental conditions dictate (Allen, 2005). In an entirely new environment that lacks the necessary selective pressure, a loss of the genes for the Type I or Type II reaction center was proposed, resulting in organisms that are similar to contemporary green sulfur bacteria and purple non-sulfur bacteria, respectively. As the concentration of readily available redox pairs diminished, a mutation that led to the loss of the regulatory switch followed by the acquisition of the Mn_4Ca cluster would have resulted in the ability to oxidize water. As a consequence, the simultaneous expression of the two sets of genes for the Type I and Type II reaction centers would have allowed them to work together to both liberate oxygen and reduce ferredoxin.

C. The Onset of Oxygenic Photosynthesis at 2.7 Ga

Although the details of how the Mn_4Ca catalyst with its associated ligand structure came together is not at all clear, there is a growing consensus for the timing of the onset of oxygenic photosynthesis. The fossil record, in the form of stromatolites the Nauga Formation in South Africa, shows evidence for the presence of cyanobacteria at about 2.6 Ga. Chemical biomarkers in the form of 2α-methylhopanes and steranes are found in 2.7 Ga shales from the Pilbara Craton in Australia (Brocks et al., 1999; Summons et al., 1999), although their use as biomarkers of oxygenic photosynthesis has recently been questioned (Kirschvink and Kopp, 2008). The strongest geological evidence for the presence of an oxygen-containing atmosphere is the analysis of the sulfur isotope ratio of rocks. The presence of atmospheric oxygen is based on the finding that in rocks younger than at 2.3 Ga, a mass-dependent isotope fractionation of sulfur occurs as a result of a variety of aqueous chemical and biochemical reactions, whereas in older rocks, a mass-independent isotope fraction of sulfur occurs as the result of gas-phase photochemical reactions such as the photolysis of SO_2. This boundary agrees rather well with the dating of the Kalahari manganese fields in South Africa at 2.22 Ga, during which the conversion of soluble Mn^{2+} to Mn^{4+} occurred as a result of the presence of O_2. Thus, the so-called 'Great Oxidation Event' occurred at about 2.3 Ga, a good 400 million years after the first occurrence of oxygenic photosynthetic cyanobacteria (Kerr, 2005).

In this interval, a number of geological processes must have sequestered enough atmospheric O_2 for the Earth to remain anaerobic. Nevertheless, the onset of an oxygenic cellular environment around 2.7 Ga would have initiated a number of adaptations, including an intense evolutionary pressure to swap out oxygen-sensitive enzymes for oxygen-insensitive enzymes (Kirschvink and Kopp, 2008). Kirshvink and Kopp have pointed out that this scenario creates a bit of a 'chicken and egg' problem: a source of oxygen must have been present that predates biological oxygen

production so as to initially drive the evolution of oxygen tolerant enzymes. In their view, the generation of H_2O_2 due to the photolysis of water vapor by UV radiation resulted in the presence of trace amounts of O_2 as a result of chemical disproportionation. Even in small quantities, O_2 would be harmful, and enzymes would have evolved to deal with it as well as with the O_2^- radical generated by the action of ferrous iron and O_2. The formation of glacial H_2O_2 during the Makganyene Snowball Earth Event at 2.3 Ga has been proposed to be the trigger for the appearance of oxygen-mediating enzymes (Kirschvink and Kopp, 2008). This event would have set the stage for the development of water splitting by Type II reaction centers. Once developed, the onset of oxygenic photosynthesis would have led to necessary changes in Type I reaction centers, one of which was to adapt the protein containing the oxygen-sensitive F_A and F_B iron-sulfur clusters to function in an aerobic environment. A corollary of this hypothesis is that a remnant of a protein that contains the F_A and F_B clusters may still exist in photosynthetic anaerobes that contain Type I reaction centers. We shall return to this topic after we describe the genomics and biochemistry of the Type I reaction centers found in present-day heliobacteria, green sulfur bacteria and cyanobacteria.

II. Type I Reaction Center Cores in Contemporary Phototrophs

Type I reaction centers from anaerobic, aerobic and oxygenic phototrophs share the greatest degree of similarity in their membrane-embedded cores. The Type I reaction centers in the anaerobic green sulfur bacteria and heliobacteria (Büttner et al., 1992; Hauska et al., 2001) are composed of a homodimeric protein core consisting of two identical polypeptides (PscA/PscA in green sulfur bacteria and PshA/PshA in heliobacteria), and a loosely-bound bacterial dicluster ferredoxin (PscB in green sulfur bacteria (Hauska et al., 2001) and PshB in heliobacteria (Heinnickel and Golbeck, 2007). The reaction center in the aerobe Chloracidobacterium also appears to consist of a homodimeric protein core (PscA/PscA) (Bryant et al., 2007) and a small bacterial dicluster ferredoxin. PSI from cyanobacteria and eukaryotic oxygenic organisms are more highly differentiated and incorporate a heterodimeric core (PsaA/PsaB), a tightly-bound dicluster ferredoxin (PsaC), and a dozen or so additional protein subunits, not including the peripheral antenna complexes (Amunts et al., 2007; Jordan et al., 2001).

A. Functional Genomics of Type I Phototrophs

The evolutionary relationship between different Type I reaction centers is best illustrated by a phylogenetic analysis of the core proteins (Fig. 12.2). Regardless of the obviously branched monophyletic groups, PsaA, PsaB, PshA, PscA (green sulfur bacteria), and PscA (Chloracidobacterium) are rooted together, consistent with their close evolutionary kinship.[3] The PshA protein from heliobacteria is located closer to PsaB and PsaA in the phylogenetic tree than are the PscA proteins from green sulfur bacteria and Acidobacteria. This is also seen in Table 12.1 by lower identity scores (about 8–10% between PscA and PsaA/PsaB, and about 14–15% between PshA and PsaA/PsaB) and only moderate similarity (about 18–26% between PscA and PsaA/PsaB, and about 25–27% between PshA and PsaA/PsaB). Sequence alignment analysis suggests that the 11 transmembrane α-helices in PsaA/PsaB of PSI are also present in the PscA protein from green sulfur bacteria (Eisen et al., 2002; Hauska et al., 2001), in the PshA protein from heliobacteria (Liebl et al., 1993; Oh-oka, 2007; Sattley et al., 2008), and in the PscA protein from Chloracidobacterium (Bryant et al., 2007), each having a typical two-domain organization for antenna (helices I–VI) and electron transport cofactors (helices VII–XI).

The PsaA and PsaB proteins in oxygenic phototrophs are clustered as two phylogenetic clades (Fig. 12.2). Based on protein sequence alignment, a high degree of homology that averages 59% of the overall amino acid identity exists between the paralogous PsaA and PsaB proteins (Table 12.1). Their close evolutionary likelihood can be easily traced in the phylogenetic analysis tree by comparing their common, longer evolutionary branches to the PshA and PscA proteins (Fig. 12.2)

[3]Even though the PscA proteins from Chlorobi and Acidobacteria share the same 'A' suffix, they belong to two different monophyletic groups.

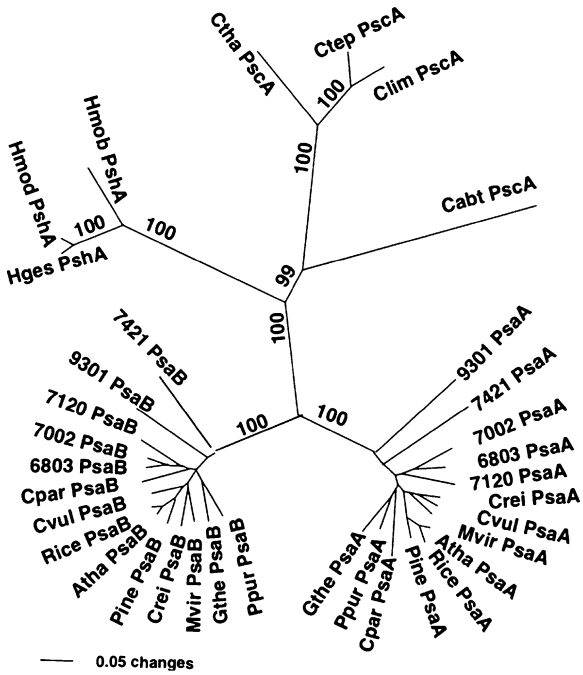


Fig. 12.2. Phylogenetic tree of core subunits of Type I reaction centers from several selected photosynthetic organisms. Amino acid sequences of the PsaA and PsaB proteins from cyanobacteria (*Synechococcus* sp. PCC 7002 (7002), *Synechocystis* sp. PCC 6803 (6803), *Anabaena* sp. PCC 7120 (7120) and *Prochlorococcus* MIT 9301 (9301)); non-green algae (*Guillardia theta* (Gthe), *Porphyra pururea* (Ppur) and *Cyanophora paradoxa* (Cpar)); green algae (*Chlamydomonas reinhardtii* (Crei), *Chlorella vulgaris* (Cvul), and *Mesostigma viride* (Mvir)); and land plants (*Arabidopsis thaliana* (Atha), *Oryza sativa* (Rice), and *Pinus koraiensis* (Pine)) were compared with PscA proteins from Chlorobi (*Chlorobium tepidum* (Ctep), *Chlorobium limicola* (Clim) and *Chloroperpeton thalassium* (Ctha)), the PscA protein from *Chloracidobacterium thermophilum* (Cabt), and the PshA proteins from Heliobacteria (*Heliobacterium modesticaldum* (Hmod), *Heliobacterium mobilis* (Hmob), and *Heliobacterium gestii* (Hges)). Sequence alignment was generated using ClustalW analysis in the MacVector program. The phylogenetic tree was generated using the PAUP phylogenetic analysis program. Branch lengths are derived from the maximum likelihood analysis. Numbers labeled on major branches represent boostrap percentages for 1,000 replicates each of maximum likelihood, parsimony and neighbor joining (NJ).

and their nearly identical twofold structural symmetry in the α-helical arrangement in cyanobacteria (Jordan et al., 2001) and plants (Amunts et al., 2007). Thus PsaA and PsaB are related to one or another by a gene duplication event that preceded their divergence in the evolution of heterodimeric Type I reaction centers. This is also supported by the fact that the *psaA* and *psaB* genes are organized in many organisms in tandem as a co-transcriptional bicistron.

As shown in Fig. 12.2, the PsaB clade is slightly closer to the PshA and PscA (green sulfur bacteria) clades than to the PsaA clade, suggesting different rates of evolutionary change. By comparing their amino acid sequences, the largest differences between PsaA and PsaB occur in the loops (except for the stromal loop that coordinates F_X), in sequence substitutions, in length variations, and in secondary structural elements. It appears that PsaB is slightly more conserved (alternately, more primitive) than PsaA, consistent with an accelerated rate of evolution of the PsaA protein (depicted visually in Fig. 12.1). In *Gleobacter violaceus*, a species that branched off at the earliest stage in the phylogenetic tree of cyanobacteria due to the absence of intracellular thylakoid membranes (Rippka et al., 1974), PsaB has a unusual 155 amino-acid C-terminal extension. This domain possesses similarities to a peptidoglycan-binding domain (Inoue et al., 2004), although how this relates to photosynthesis is not clear. While it is also not clear how the *psaB* gene developed in *G. violaceus*, the presence of the extension does support the proposal that PsaA and PsaB have evolved at different rates. Using the core and antenna domains for individual phylogenetic tree analyses, different rates of change can also be distinguished in PsaA and PsaB, especially in the antenna domain relative to the core domain (Raymond and Blankenship, 2006).

The greatest degree of similarity among the reaction center core proteins exists in the region that coordinates the electron transfer cofactors (see also Fig. 12.1), which suggests that the protein fold around the redox cofactors has been maintained over evolutionary time. We next describe differences in the organization of the electron transfer cofactors in cores from homodimeric and heterodimeric Type I reaction centers.

B. Comparative Biochemistry of Type I Phototrophs

The 2.5 Å resolution X-ray crystal structure of cyanobacterial PSI depicts the three-dimensional arrangement of the electron transport cofactors in a heterodimeric Type I reaction center (Jordan et al., 2001). The PsaA and PsaB proteins, which

Table 12.1. Amino acid sequence similarity and identity matrix of the PsaA and PsaB proteins from *Oryza sativa* (Rice), *Chlamydomonas reinhardtii* (Crei), *Guillardia theta* (Gthe), and *Synechococcus* sp. PCC 7002 (7002), the PscA proteins from *Chloracidobacterium thermophilum* (Cabt), and *Chlorobium tepidum* (Ctep), and the PshA protein from *Heliobacterium modesticaldum* (Hmod). Multiple sequence alignments were generated using the MacVector program

Identity scores (%)

	Rice PsaA	Crei PsaA	Gthe PsaA	7002 PsaA	Rice PsaB	Crei PsaB	Gthe PsaB	7002 PsaB	Cabt PscA	Ctep PscA	Hmod PshA
Rice PsaA	100.0	84.5	79.4	76.4	42.2	42.2	43.6	42.3	7.9	10.9	14.2
Crei PsaA	92.4	100.0	83.9	80.3	41.2	41.4	42.6	40.7	8.2	11.2	14.3
Gthe PsaA	89.9	92.4	100.0	77.4	40.7	41.3	41.8	40.4	7.8	109	15.0
7002 PsaA	87.9	88.8	89.1	100.0	41.9	42.2	43.2	42.3	8.3	10.5	13.8
Rice RsaB	57.8	57.9	57.2	59.1	100.0	83.9	80.5	81.5	9.8	12.6	14.7
Crei PsaB	58.1	57.8	56.9	59.7	90.9	100.0	82.2	83.0	9.3	13.1	15.3
Gthe PsaB	59.5	59.2	58.2	60.7	90.2	90.6	100.0	84.5	10.5	12.3	14.0
7002 PsaB	58.5	57.8	57.3	59.5	90.9	90.6	91.6	100.0	10.3	13.0	14.8
Cabt PscA	17.9	18.3	18.3	19.0	20.8	20.1	21.3	21.1	100.0	19.2	10.0
Ctep PscA	22.7	22.9	23.2	23.7	25.1	25.2	25.2	25.8	30.3	100.0	12.3
Hmod PshA	27.3	27.0	26.6	27.3	26.6	26.3	25.2	26.4	21.0	26.6	100.0

Similarity scores (%)

form the membrane-embedded core of PSI, coordinate the majority of the electron transfer cofactors (P_{700}, A_0, A_1 and F_X) and 79 of the 98 chlorophyll molecules. The five transmembrane α-helices that comprise the C-terminal region of PsaA and PsaB harbor the electron transport cofactors, whereas the six transmembrane α-helices that comprise the N-terminal region harbor the antenna chlorophylls. The cofactors (P_{700}, A_0 and A_1) are located in two C_2-symmetric branches along the PsaA and PsaB subunits (Fig. 12.3). The two branches converge at the interpolypeptide F_X iron-sulfur cluster, culminating in linear electron transfer through the terminal F_A and F_B clusters. The C-terminal domains of PsaA and PsaB are oriented in a ying-yang (or handshake) configuration, complete with pseudo C_2-symmetry, which reveals the high degree of interconnectedness and interdependency of the two subunits (Jordan et al., 2001). A high-resolution X-ray crystal structure is not available for either the green sulfur bacterial reaction center or the heliobacterial reaction center. For structural information, we rely instead on biochemical and biophysical studies and on analogy with PSI.

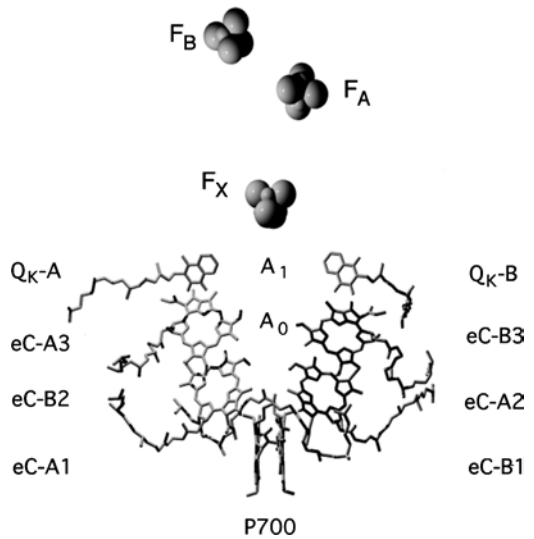


Fig. 12.3. Depiction of the electron transfer cofactors based on the 2.5 Å X-ray crystal structure of PSI from *Thermosynechococcus elongatus*. The crystallographic names of the six chlorophyll *a* molecules and the two phylloquinones on the A- and B-branches of cofactors are shown are shown in the *left* and *right* margins; the spectroscopic names are shown in the *middle*. Note that while the eC-B2 and eC-A2 chlorophyll molecules are ligated by PsaB and PsaA, respectively, they function in the A- and B-branches of electron transfer.

1. The Special Pair of (Bacterio)Chlorophylls

All Type I reaction centers employ two juxtaposed (bacterio)chlorophyll molecules, termed the special pair, which stabilizes the cation produced by light excitation. In PSI, the primary donor, termed P_{700} for its peak absorbance in the visible region, is comprised of a Chl *a*/Chl *a*′ special pair, wherein the Chl *a*′ is a 13^2 epimer of Chl *a*. The X-ray crystal structure of cyanobacterial PSI (Jordan et al., 2001) shows that Chl *a*′ (eC-A1) is coordinated by His_{680} on the PsaA protein[4] and that Chl *a* (eC-B1) is coordinated by His_{660} on the PsaB protein (Fig. 12.3). Even though both Chl *a* molecules are bound to His residues, an asymmetry in their environment is evident in the pattern of hydrogen bonds: three are present on the A-branch Chl *a*′ (eC-A1), whereas none are present on the B-branch Chl *a* (eC-B1) (Fromme et al., 2001). Although the role of these hydrogen bonds is not known, it has been postulated that they might be required for the proper insertion of a Chl *a*′ moiety on the A-branch (Webber and Lubitz, 2001).

The asymmetry in the hydrogen-bonding pattern might also have implications for the electronic and spectroscopic properties of P_{700}, which are measured typically in frozen solutions (Davis et al., 1993; Mac et al., 1998) or with single crystals of PSI (Käss et al. 1996, 2001). ESEEM and ENDOR data suggest that the spin density on P_{700}^+ is asymmetrically distributed such that the majority of the electron spin resides on the eC-B1 Chl *a*, leaving the minority of the electron spin to reside on the eC-A1 Chl *a*′ (Lubitz, 2006; Webber and Lubitz, 2001). It is not clear if the asymmetry at the level of P_{700} plays a role in determining whether electron transfer is favored along the A- or the B- branch of redox cofactors (see Srinivasan and Golbeck, 2009).

Because the reaction center cores in green sulfur bacteria and heliobacteria are homodimeric, it is reasonable to expect that the environment around each of the chlorophylls that constitute their special pairs will be perfectly symmetric.

[4]The amino acid numbering scheme used in this article is that of PS I from *Thermosynechococcus elongatus.*

Two of the bacteriochlorophyll *a* (BChl *a*) molecules bound to the PscA/PscA homodimeric core of the green sulfur bacterial reaction center are 13^2 epimers (BChl *a*′), and constitute the primary donor, P_{840} (Kobayashi et al., 2000). Similarly, two of the BChl *g* molecules bound to the PshA/PshA homodimeric core of the heliobacterial reaction center are BChl *g'* and constitute the primary donor, P_{798} (Kobayashi et al. 1991a, 1991b). Thus, in contrast to the heterodimeric Chl *a*/Chl *a*′ special pair in PSI, green sulfur bacteria, and heliobacteria employ a homodimeric BChl/BChl special pair. Spectroscopic studies show that the unpaired electron of P_{840}^{+} is nearly evenly distributed over the special pair of bacteriochlorophylls in the green sulfur bacterial reaction center (Rigby et al., 1994). This is consistent with two equivalent electron transfer branches as suggested by the homodimeric nature of these reaction center cores.

The His residues that coordinate P_{700} in PSI are also present in the amino acid sequences of the green sulfur bacterial and heliobacterial reaction centers (Fig. 12.4). The special pairs in *Chlorobium tepidum* and in *Heliobacterium modesticaldum* are therefore coordinated by axial His ligands (Hauska et al., 2001). Because the geometry and coordination motif has remained unchanged over evolutionary time, the juxtaposed arrangement of the special pair pigments was likely adopted very early during the evolution of photosynthetic reaction centers. Hence, the transition from a homodimeric to a heterodimeric Type I reaction center was probably not driven by any selective pressure exerted by the primary electron donor.

2. The Bridging Chlorophylls

In PSI, a Chl *a* molecule on each branch is considered to serve as an electron transfer bridge between P_{700} and A_0. The X-ray crystal structure of cyanobacterial PSI (Jordan et al., 2001) reveals that these bridging chlorophylls (denoted eC-B2 and eC-A2, see Fig. 12.3) are located ~12 Å (center-to-center distance) apart from P_{700}. The bridging Chl *a* molecules are not coordinated by an amino acid; instead two water molecules serve as the axial ligand to the Mg^{2+}. The water molecule that coordinates the bridging Chl *a* on the A-branch is hydrogen bonded to Asn_{591} on PsaB; hence the chlorophyll is denoted as eC-B2 even though it participates in electron transfer through the A-branch. Similarly, water molecule that coordinates the bridging Chl *a* on the B-branch is hydrogen bonded to Asn_{604} on PsaA; hence the chlorophyll is denoted as eC-A2 even though it participates in electron transfer through the B-branch.

The green sulfur bacterial reaction center contains four Chl *a* molecules, which are proposed to function in a manner similar to eC-A2 and eC-A3 (eC-B2 and eC-B3) in PSI based on spectroscopic

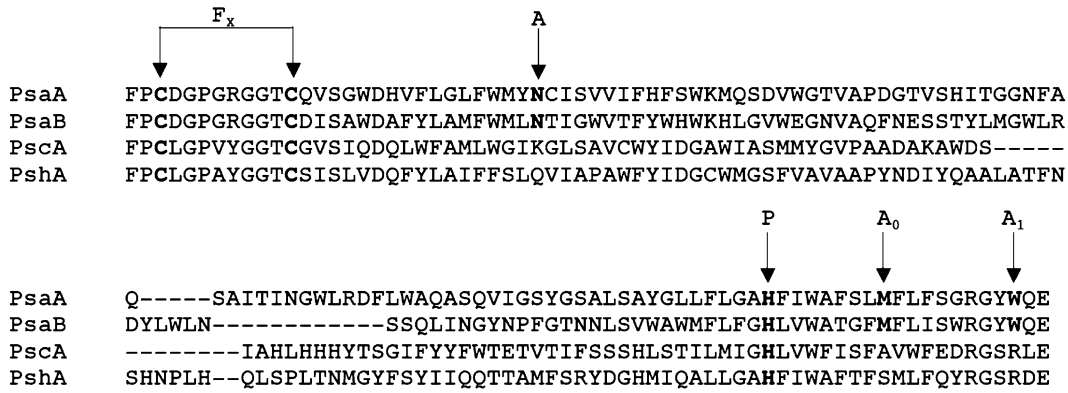


Fig. 12.4. Sequence alignment of PsaA and PsaB (*Thermosynechococcus elongatus*), PscA (*Chlorobium tepidum*) and PshA (*Heliobacterium modesticaldum*) in the region around the membrane-bound redox cofactors. The following residues are indicated by *arrows*: the His residues that ligate the 'special pair' (bacterio)chlorophylls eC-A1 and eC-B1 (denoted as P); the Asn residues that form H-bonds with the water molecule that coordinates the bridging chlorophylls eC-A2 and eC-B2 (denoted as A); the Met residues that coordinate the primary acceptor chlorophylls eC-A3 and eC-B3 (denoted as A_0); the Trp residues that π-stacks with the secondary acceptor phylloquinones Q_K-A and Q_K-B (denoted as A_1); the Cys residues that ligate the interpolypeptide [4Fe-4S] cluster (denoted as F_X).

studies that suggest a close spacing of the Chl *a* molecules (Hauska et al., 2001; Kjaer and Scheller, 1996). A sequence alignment of PsaA, PsaB, PscA and PshA (Fig. 12.4) shows that the Asn residue involved in hydrogen bonding with the coordinating water molecules is missing in PscA and PshA (Hauska et al., 2001; Sattley et al., 2008). Instead, the green sulfur bacterial reaction center contains a Lys residue, and the heliobacterial reaction center contains a Gln residue at this position. The side chains of both Lys and Gln are capable of forming hydrogen bonds with water and it is conceivable that these residues might be involved in coordinating similar bridging (bacterio)chlorophyll molecules. However, there is yet no direct experimental evidence for the existence of comparable chlorophylls in the homodimeric Type I reaction centers of green sulfur bacteria and heliobacteria. A high-resolution X-ray crystal structure is needed to show whether a comparable bridging chlorophyll exists in either of these homodimeric Type I reaction centers.

3. The Primary Acceptor Chlorophylls, A_0

The first stable charge-separated state in PSI is formed between the oxidized primary donor, P_{700}^+, and the reduced primary acceptor, A_0^- (Brettel and Leibl, 2001). A_0 is a monomeric Chl *a* molecule positioned ~8.5 Å (center-to-center distance) from the bridging Chl *a* molecule in the X-ray crystal structure of cyanobacterial PSI. The ligand structure of the A_0 chlorophylls is unusual in that the sulfur atoms from Met_{688} and Met_{668} coordinate the Mg^{2+} ions in the A-branch (eC-A3) and the B-branch (eC-B3) respectively (Fig. 12.3). This coordination pattern is unusual given that the interactions between the hard acid Mg^{2+} and the soft base sulfur should be weak according to the concept of hard and soft acids and bases (Fromme et al., 2001). It is unclear whether or not the sulfur ligand is responsible for the extremely low midpoint potential of A_0 (~ –1 V). Alterations in growth rates, spectroscopic properties, and forward electron transfer kinetics have been reported when the Met residues are replaced with Leu, Asn, or His in both cyanobacteria and algae (Cohen et al., 2004; Dashdorj et al., 2005; Savitsky et al., 2009; van der Est et al., 2009), underscoring the importance of this residue in PSI.

A Chl *a*-derivative (8^1-OH Chl *a*) has been identified as the primary electron acceptor in the heliobacterial reaction center (Amesz, 1995; van de Meent et al., 1991). Mass spectrometric studies on pigments extracted from the green sulfur bacterial reaction revealed that the A_0 Chl *a* molecule is esterified with 2,6-phytadienol (Kobayashi et al., 2000). In the absence of an X-ray crystal structure, the identity of the amino acid that coordinates the putative A_0 molecule in homodimeric Type I reaction centers remains unclear (see Fig. 12.4).

Although P_{700} has long been considered the 'primary' electron donor of PSI, an alternative mechanism of primary charge separation has been proposed by Holzwarth et al. (2006), wherein the initial charge separation occurs between the bridging Chl *a* molecule, eC-B2, and the primary acceptor eC-A3 (and between the bridging Chl *a* molecule, eC-A2, and the primary acceptor eC-B3). This is followed by rapid migration of the hole to P_{700} resulting in the quasi-stable $P_{700}^+A_{0A}^-$ ($P_{700}^+A_{0B}^-$) charge-separated state.

4. The A_{1A} and A_{1B} Phylloquinones

Quinones are fairly common redox cofactors in photosynthetic systems; the purple bacterial reaction center contains two ubiquinone molecules and PSII contains two plastoquinone molecules that participate in electron transfer (Jagannathan and Golbeck, 2009a). The adoption of a phylloquinone as a secondary electron acceptor, A_1, in PSI is however a surprising choice for an electron transport cofactor given the extremely low midpoint potential at which it must operate (i.e. as low as −800 mV (Iwaki and Itoh, 1994; Ptushenko et al., 2008; Vos and van Gorkom, 1990). The comparable redox potentials of the quinones in the purple bacterial reaction center and in PSII respectively are ca. −70 mV (Gunner et al., 1986) and ca. −100 mV (Krieger et al., 1995; Renger and Holzwarth, 2005).

The A_{1A} and A_{1B} phylloquinones (termed $Q_{K\text{-}A}$ and $Q_{K\text{-}B}$, respectively, in the X-ray crystal structure of cyanobacterial PSI; see Fig. 12.3) are held in their binding pockets via (i) a single hydrogen bond with the backbone amide of Leu_{722} on PsaA (Leu_{702} on PsaB); (ii) π-stacking interactions with Trp_{697} on PsaA (Trp_{677} on PsaB); (iii) hydrophobic interactions between the quinone tail and the

protein environment (Fromme et al., 2001; Jordan et al., 2001; Srinivasan and Golbeck, 2009). Functionally, the presence of the H-bond to the phylloquinone in PSI is counterintuitive in that it stabilizes the semiquinone anion relative to the ground state, and consequently raises (i.e. drives more positive) the midpoint potential. Because the A_1 quinones require a very low midpoint potential, the surrounding protein must drive the midpoint potential even more negative than would otherwise be necessary in the absence of the H-bond. Thus, it is within reason to assume that the H-bond serves a role that ensures conservation from cyanobacteria to plants. A recent study of an *in vivo* cyanobacterial variant that contained a Trp residue instead of Leu_{722} on PsaA suggests that its function may be to tie up the carbonyl group so that the semiquinone anion radical cannot become protonated, and hence susceptible to double reduction under high light conditions (Srinivasan et al., 2009).

The Trp residue that π-stacks with the phylloquinone is part of a highly conserved RGYWQE peptide segment on both PsaA and PsaB. The interaction of phylloquinone with the Trp residue may act to destabilize the negative charge on the semiquinone anion radical, thereby lowering its redox potential (reviewed in Srinivasan and Golbeck (2009)). Electrostatic calculations indicate that Trp_{697} on PsaA and Trp_{677} on PsaB contribute −27 mV to the midpoint potential of the respective quinone (Ishikita and Knapp, 2003) although quantum chemical calculations suggest a greater contribution of −50 to −150 mV (Kaupp, 2002). The significance of the π-stacking interactions is best demonstrated by magnetic resonance and time-resolved optical spectroscopic studies that were performed on a cyanobacterial PSI variant that contained a Phe residue instead of Trp (Xu et al., 2003a, b). The substitution led to an altered spin density distribution around the phylloquinone, and a slowing of electron transfer from A_1^- to F_X. The results indicate the important role of the π-stacking interactions in imparting functionality to the phylloquinone in PSI.

The participation of a quinone in forward electron transfer in the green sulfur bacterial and heliobacterial reaction center remains a subject of considerable debate. Unlike an iron-sulfur cluster, a quinone binding site does not have a consensus amino acid binding motif, making it difficult to draw inferences from sequence alignments of the reaction center core proteins. Nevertheless, an obvious difference can be identified in the putative A_1 binding pocket between heterodimeric and homodimeric Type I reaction centers. The Leu residue that forms a single H-bond with A_1 in PSI is not found at the corresponding position in PscA or PshA. Since only the NH backbone of Leu_{722} on PsaA (Leu_{702} on PsaB) is involved in the H-bonding with A_{1A} (A_{1B}), any amino acid, in principle, could substitute for Leu. The Trp residue in the conserved peptide RGYWQE of PsaA/PsaB, which plays a role in the function of the A_1 phylloquinone in PSI (Jordan et al., 2001; van der Est, 2006; Xu et al., 2003a, b), is not present either in PscA or in PshA (Fig. 12.4). This has led to the suggestion that a quinone, if present, is loosely bound in homodimeric Type I reaction centers (Hauska et al., 2001).

There is increasing experimental evidence for the participation of a quinone in forward electron transfer in homodimeric Type I reaction centers (Kjaer et al., 1998; Miyamoto et al., 2008; Muhiuddin et al., 1999). Originally, menaquinone-7 was considered to function as A_1 in the green sulfur bacterial reaction center (Hauska, 1988) based on an EPR signal similar to a semiquinone radical in photoaccumulated membranes (Muhiuddin et al., 1999; Nitschke et al., 1987) and isolated reaction centers (Kjaer et al., 1998). Subsequent studies have shown that electron transfer occurs to the F_A/F_B clusters, even when the reaction centers are devoid of menaquinone (Frankenberg et al., 1996; Hager-Braun et al., 1997; Permentier et al., 2000; Schmidt et al., 2000). An alternate mechanism in which menaquinone accepts electrons, albeit via a side pathway has been proposed (Hauska et al., 2001). The sparsity of evidence supporting (or opposing) the role of the menaquinone as the A_1 acceptor in green sulfur bacteria thus precludes a final verdict on this issue.

The situation is similar in the heliobacterial RC wherein there is biochemical and spectroscopic evidence for and against the involvement of menaquinone-9 in electron transfer (reviewed in Heinnickel and Golbeck (2007)). A recent report on the detection of a spin correlated $P_{798}^+A_1^-$ radical pair in the heliobacterial reaction center

by transient EPR (Miyamoto et al., 2008) may be the first solid evidence for its involvement in forward electron transfer. Nevertheless, the absence of unequivocal kinetic or functional data for a functional quinone in green sulfur bacteria and heliobacteria prevents us from saying anything definitive about the role of quinones in the transition from a homodimeric to heterodimeric core.

5. The Interpolypeptide F_X Iron-Sulfur Cluster

Electron transfer from the primary donor, P_{700}, through A, A_0 and A_1 bifurcates along the PsaA and PsaB branches in PSI (reviewed in Redding and van der Est (2006); Srinivasan and Golbeck (2009); van der Est (2006)), but the branches converge at a unique interpolypeptide [4Fe-4S] cluster termed F_X (Fig. 12.3). F_X is thought to be the most reducing, functional iron-sulfur cluster in biology, with a reported midpoint potential of ~ −650 to −700 mV (Chamorovsky and Cammack, 1982; Parrett et al., 1989; Ptushenko et al., 2008). The four iron atoms are coordinated by two highly conserved Cys residues from PsaA (Cys_{578} and Cys_{587}) and two highly conserved Cys residues from PsaB (Cys_{565} and Cys_{574}), as confirmed initially by mutagenesis studies (Hallahan et al., 1995; Vassiliev et al., 1995) and later by the 2.5 Å X-ray crystal structure of cyanobacterial PSI (Jordan et al., 2001). The ligating Cys residues are part of highly conserved loop regions that connect the transmembrane α-helices h and i in both PsaA and PsaB. The loop regions have multiple roles; in addition to coordinating F_X, they appear to enclose the [4Fe-4S] cluster in a protective cage (Jagannathan and Golbeck, 2009b).

The F_X-coordinating loops and the stromal loops of helices h and i contain negatively charged Asp residues that form strong ionic bonds with positively charged Arg/Lys residues on PsaC, thereby providing a stable binding surface for the stromal protein (Antonkine et al., 2003; Jagannathan and Golbeck, 2009b) (see Section III.A.3). The high degree of C_2-symmetry around F_X is remarkable; the Cys residues that ligate the [4Fe-4S] cluster and the Asp residues that form ionic bonds with PsaC are located at identical positions in the loop regions on the PsaA and PsaB subunits. The near-perfect C_2 symmetry at the interface of PsaA and PsaB is considered to be an evolutionary remnant of the transition from a homodimeric to a heterodimeric core.

F_X is the defining cofactor in both prokaryotic and eukaryotic photosynthetic organisms. A close examination of the sequences of Type I reaction centers in the region binding the redox components indicates that the highest degree of conservation exists in the loop that incorporates the two cysteine residues that harbor the [4Fe-4S] cluster F_X. The F_X-binding motif (FPCXGPXXGGTC) is found in the PscA protein of green sulfur bacteria, the PscA protein of *Chloracidobacteria* and the PshA protein of heliobacteria (Fig. 12.4) (Bryant et al., 2007; Hauska et al., 2001). There is every expectation that these two cysteine residues coordinate F_X in the homodimeric reaction center cores of PscA/PscA and PshA/PshA in a manner analogous to the heterodimeric PsaA/PsaB core of PSI. One significant difference is that the F_X-coordinating loops in the PscA and PshA proteins do not contain the negatively charged Asp residues that are involved in PsaC-PsaA/PsaB interactions in PSI. It appears that homodimeric Type I reaction centers adopt a different mode of interaction between the membrane core and the F_A and F_B-containing protein. Interestingly, the F_X cluster appears to be stable to oxygen in the reaction centers of heliobacteria and green sulfur bacteria.

The magnetic properties of F_X are different in homodimeric and heterodimeric Type I reaction centers. The F_X cluster in PSI exists in a pure S = ½ ground spin state (Vassiliev et al., 2001), whereas the F_X cluster in the green sulfur bacterial (Jagannathan and Golbeck, 2008; Vassiliev et al., 2000) and the heliobacterial reaction center (Heinnickel et al., 2006; Miyamoto et al., 2008) is reported to exist in S = ½ and S = 3/2 ground spin states. Although the factors that force the coupling of the ferrous and ferric ions in cubane [4Fe-4S] clusters into a particular spin state are not completely understood (see Noodleman et al. 1985, 1995; for reviews), a highly symmetrical environment surrounding an interpolypeptide [4Fe-4S] cluster is a common feature in proteins that fail to form an exclusive S = ½ ground spin state (Lindahl et al., 1985). In spite of these differences, the F_X cluster in Type I reaction centers is highly conserved, making it unlikely that it provided the selective pressure for the transition from a homodimeric core to a heterodimeric core. We therefore turn to the properties of the stromal

polypeptides, including PsaC, PsaD and PsaE to fill this role.

III. The Terminal Iron Sulfur Clusters in Type I Reaction Centers

In all known Type I reaction centers, two [4Fe-4S] clusters named F_A and F_B and located on a membrane-extrinsic polypeptide, serve as the terminal electron acceptors. In this section we will examine the similarities and differences between the F_A/F_B-containing proteins in heterodimeric and homodimeric Type I reaction centers. A number of discoveries involving these polypeptides have been made in the last few years that have influenced our thinking of how Type I reaction centers evolved. In particular, we will focus on the modifications that have occurred on the acceptor side of PSI in response to the onset of oxygenic photosynthesis.

A. The F_A- and F_B-Containing Protein in Type I Reaction Centers

The F_A/F_B-containing protein in PSI (PsaC), in green sulfur bacteria (PscB), in *chloracidobacteria* (PscB), and in heliobacteria (PshB) share several features in common with bacterial dicluster ferredoxins. Firstly, PsaC is highly conserved from prokaryotic cyanobacteria through eukaryotic algae and plants, clustering together as a close clade in the phylogenetic tree (Fig. 12.5). Secondly, an alignment of the amino acid sequences shows that, with the exception of a unique motif in PscB, the iron-sulfur binding motif consists of a CxxCxxCxxxCP sequence typical of bacterial ferredoxins. Thirdly, the three-dimensional structure of PsaC (and likely PscB and PshB) surrounding the [4Fe-4S] clusters are similar to those of low molecular mass, bacterial dicluster ferredoxins, which contain just enough polypeptide to encapsulate the clusters. Nevertheless, there are significant differences in the sizes and amino acid sequences of the N-terminal and C-terminal regions of PsaC, PscB and PshB. These differences will be discussed in the context of their structure and function, with particular attention paid on how these changes enabled the F_A/F_B clusters in PSI to function in an aerobic environment.

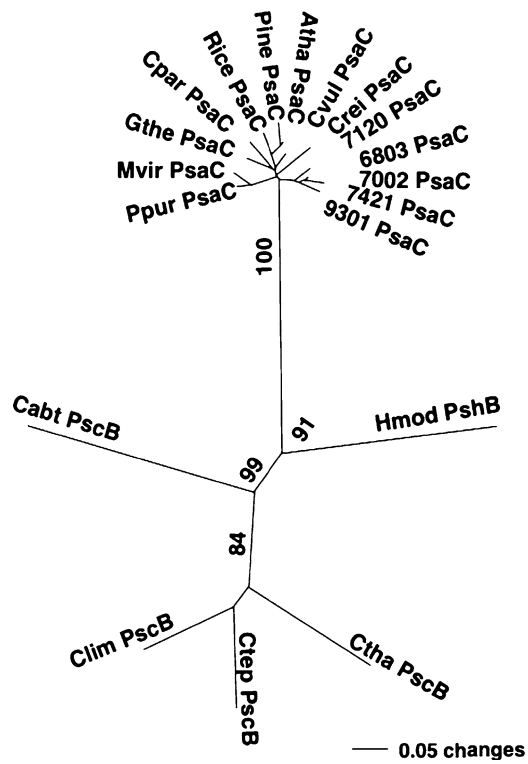


Fig. 12.5. Phylogenetic tree of the PsaC proteins from selected oxygenic photosynthetic organisms, the PscB proteins from selected Chlorobi species and *Chloracidobacterium thermophilum* and the PshB protein from *Heliobacterium modesticaldum*. The designated simplified names for different species are same as used in Fig. 12.3.

1. The PshB Polypeptide in the Heliobacterial Reaction Center

Although it had been shown as early as 1990 that the F_A/F_B clusters in heliobacterial reaction centers could be reduced by light at cryogenic temperatures (Nitschke et al., 1990), the identity of the protein that carried the Fe/S clusters remained elusive. Heinnickel and co-workers made a key discovery in 2005 that the F_A/F_B-containing polypeptide (named PshB) could be dissociated from the HbRC at low ionic strengths, and that it could be recovered intact after passage through a 30-kDa cutoff ultrafiltration membrane (Heinnickel et al., 2005). Prior to the identification of the salt-dissociable PshB subunit, Hatano and co-workers reported the isolation of two soluble ferredoxins, Fd1 and Fd2, from *Heliobacterium mobilis* (Hatano et al., 2005). Biochemical and spectroscopic studies showed that the recombinant Fd2 protein was

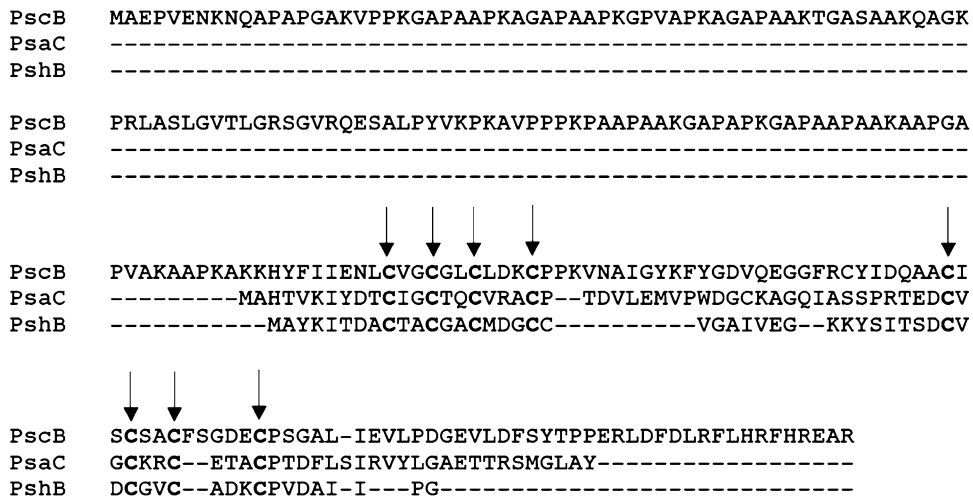


Fig. 12.6. Alignment of the amino acid sequences of PscB (*Chlorobium tepidum*), PsaC (*Synechococcus elongatus*) and PshB (*Heliobacterium modesticaldum*). The Cys residues that ligate the [4Fe-4S] clusters are indicated by *arrows*.

equivalent to the native PshB protein (Heinnickel et al., 2007). The amino acid sequence of Fd2/PshB (Fig. 12.6) contains two [4Fe-4S] cluster-binding motifs, one of which is a traditional CxxCxxCxxxCP motif and the other is a novel CxxCxxCxxxCC motif (Heinnickel and Golbeck, 2007). It is not known whether the presence of a Cys residue instead of a Pro adjacent to the fourth Cys has functional or structural significance. More recently, it was shown that the Fd1 (since renamed PshB2) holoprotein from *H. modesticaldum* is capable of binding to heliobacterial reaction center cores and of supporting low temperature electron transfer to the F_A/F_B clusters (Romberger et al., 2010).

2. The PscB Polypeptide in the Green Sulfur Bacterial Reaction Center

The PscB polypeptide in the green sulfur bacterial reaction center is significantly different from its homologs, PshB and PsaC. Firstly, the 23-kDa polypeptide has a unique, positively charged N-terminal extension (Fig. 12.6), which is highly enriched in Lys, Pro and Ala residues (Büttner et al., 1992; Figueras et al., 2002). Although several roles have been proposed for this extension (Büttner et al., 1992; Figueras et al., 2002), none of them involve the functioning of the F_A or F_B clusters. Secondly, although the C-terminal region of PscB is similar to that of most bacterial dicluster ferredoxins, it contains a novel CxxCxxCxxxxxCP Fe/S cluster-binding motif that ligates the putative F_A cluster (Büttner et al., 1992; Figueras et al., 2002). This motif is not found in any other known dicluster ferredoxins and may represent a unique binding motif for a [4Fe-4S] cluster. EPR and Mössbauer spectroscopic studies reveal that unlike recombinant PshB and PsaC, the two [4Fe-4S] clusters cannot be properly inserted into recombinant PscB *in vitro* (Hager-Braun et al., 1999; Jagannathan and Golbeck, 2008). Thus, PscB represents one of the few instances in which [4Fe-4S] clusters fail to become properly exchanged into an *E. coli*-expressed apo-ferredoxin.

Similar to PshB in the heliobacterial reaction center, the PscB polypeptide in the green sulfur bacterial reaction center can also be dissociated from the membrane core under conditions of moderate ionic strength (Jagannathan and Golbeck, 2008). Thus, a loosely-bound F_A/F_B-containing protein may be a common feature of homodimeric Type I reaction centers. This finding indicates that in spite of the differences in the amino acid composition of PscB and PshB, their mode of binding to homodimeric Type I reaction share more similarities than previously thought.

3. The PsaC Polypeptide in Photosystem I

PsaC is the best characterized of the F_A/F_B-containing proteins in Type I reaction centers. The amino acid sequence of the 9-kDa polypeptide is

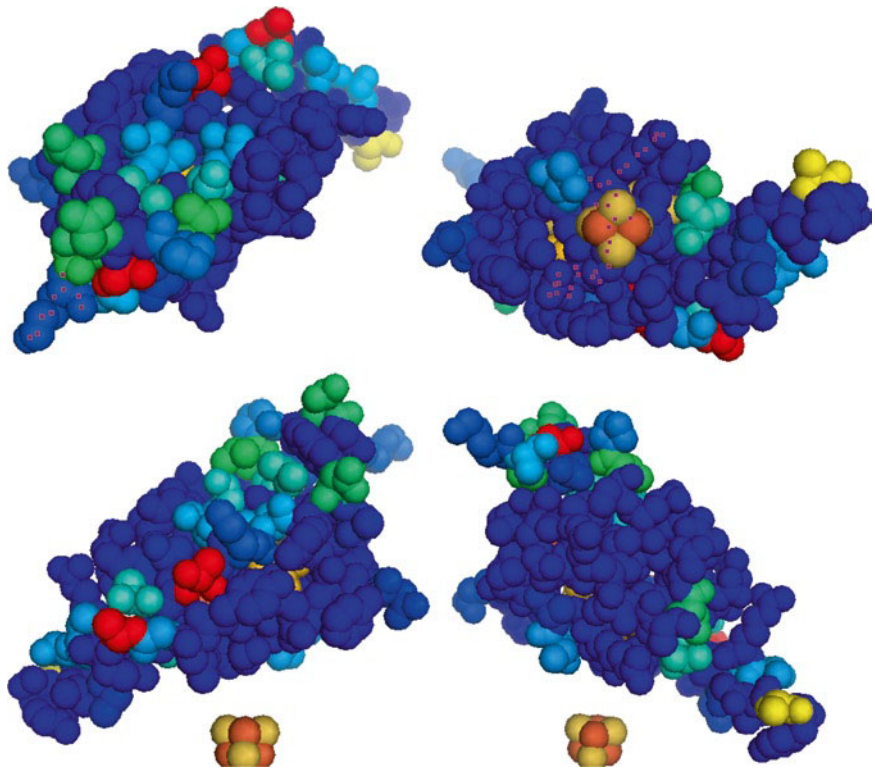


Fig. 12.7. Four views of the PsaC protein (*top*: view toward the stromal surface (*left*) and view from the stromal surface (*right*), the latter showing also the contact points (*magenta squares*) with Asp_{555}/Asp_{566} from PsaB and with Asp_{568}/Asp_{579} from PsaA; *bottom*: two *side views* also showing the F_X cluster. The amino acids are color coded according to the degree of similarity in an amino acid sequence alignment using the PsaC sequences from 51 cyanobacterial, algal, and plant species: the colors represent a continuum between invariant residues (*blue*) and highly variable residues (*red*).

typical of bacterial dicluster ferredoxins in that it contains two traditional CxxCxxCxxxCP cluster-ligating motifs (Fig. 12.6) and an amino acid sequence that is highly conserved among all cyanobacterial, algal and plant species. The most conserved residues are located at the PsaC-PsaA/PsaB binding interface, whereas the variable residues are usually located on the stromal face (Fig. 12.7). The coordinating Cys ligands for the F_A and F_B clusters in PsaC were initially assigned by site-specific mutagenesis and EPR spectroscopy (Golbeck, 1999; Zhao et al., 1992). F_A was shown to be coordinated by Cys_{20}, Cys_{47}, Cys_{50} and Cys_{53}, and the F_B cluster was shown to be coordinated by Cys_{10}, Cys_{13}, Cys_{16} and Cys_{57}, a pattern of ligation similar to that of well-characterized bacterial dicluster ferredoxins. Spectroscopic studies indicate that PsaC is oriented on PSI such that the F_A cluster accepts the electron from the membrane-intrinsic F_X cluster, and the F_B cluster donates the electron to a soluble acceptor protein such as ferredoxin (DiazQuintana et al., 1998; Vassiliev et al., 1998).

The 2.5 Å resolution X-ray crystal structure of cyanobacterial PSI (PDB ID 1JB0) provides precise details on the interactions between PsaC and the PsaA/PsaB heterodimeric membrane core (Fig. 12.8). The binding contacts include an intricate network of ionic bonds consisting of one Lys and two Arg residues on PsaC, and four Asp residues on the PsaA/PsaB heterodimer (Antonkine et al., 2003; Jagannathan and Golbeck, 2009b), two of which are part of an extrinsic loop on PsaA and on PsaB that also provides the Cys residues that coordinate the F_X cluster. The network of Asp-Arg/Lys ionic contacts is C_2-symmetric; Arg_{52} on PsaC forms five ionic bonds with Asp_{568}/Asp_{579} on PsaA, and Lys_{51}/Arg_{65} on PsaC form five ionic bonds with Asp_{555}/Asp_{566} on PsaB (Antonkine et al., 2003). The equal division of the ten ionic contacts between

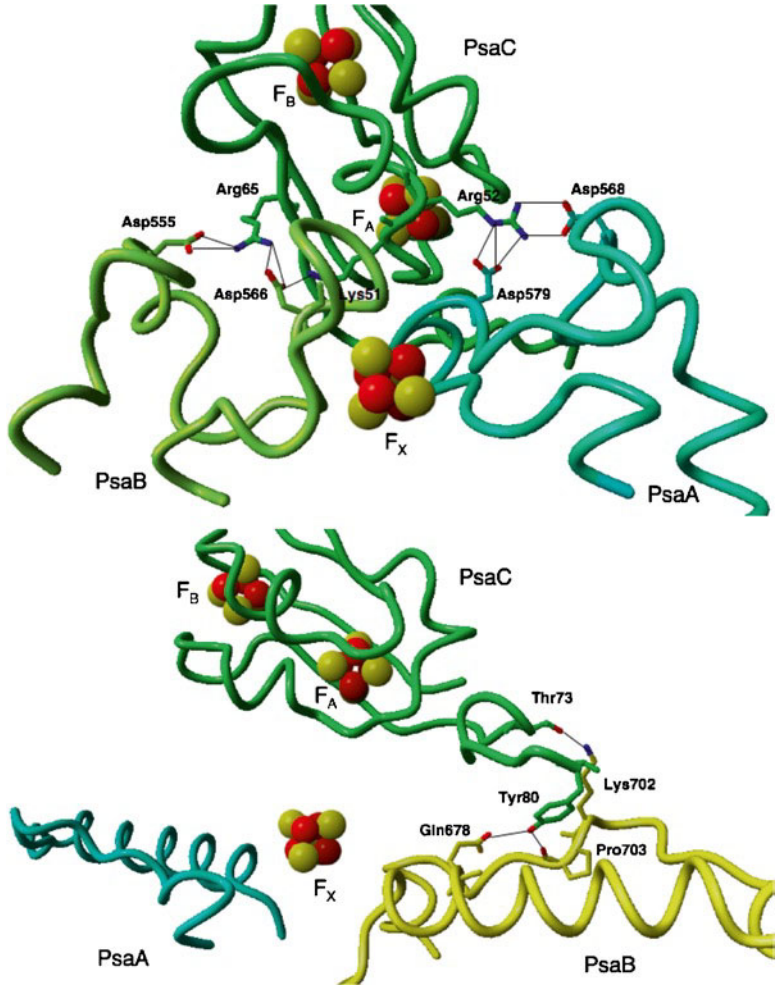


Fig. 12.8. The network of ionic (*top*) and H-bonds (*bottom*) between PsaC and the PsaA/PsaB heterodimer in PSI. Note the five ionic bonds between Arg_{65}/Lys_{51} on PsaC and Asp_{555}/Asp_{566} on PsaB and the five ionic bonds between Arg_{52} on PsaC and Asp_{568}/Asp_{579} on PsaA. Rotating PsaC about the 180° C_2 symmetry axis that passes through F_X and P_{700} generates five similar ionic bonds, but between Arg_{52} on PsaC and Asp_{555}/Asp_{566} on PsaB and five similar ionic bonds between Arg_{65}/Lys_{51} on PsaC and Asp_{568}/Asp_{579} on PsaA. The H-bonds between Thr_{73} of PsaC and Lys_{702} of PsaB and between Tyr_{80} of PsaC and Gln_{678}/Pro_{703} of PsaB have no counterpart between PsaC and PsaA when PsaC is rotated around the C_2 symmetry axis.

PsaA and PsaB represents an extraordinarily tight binding surface for the PsaC protein.

PsaC also forms three H-bonds in a region of the P_{700}-F_X core distant from the ionic contacts (Fig. 12.8). These bonds are formed between Thr_{73}/Tyr_{80} on the C-terminus of PsaC and $Gln_{678}/Lys_{702}/Pro_{703}$ on a surface-located segment of PsaB (Antonkine et al., 2003). Unlike the symmetric ionic contacts, a symmetry-related binding pocket for the C-terminus of PsaC is not present on the comparable surface-located segment of PsaA. Thus, the three H-bonds between the C-terminus of PsaC and PsaB break the otherwise-perfect C_2-symmetry between PsaC and the PsaA/PsaB heterodimer.

4. The Function of the Tight PsaC-PsaA/PsaB Binding Interface

The relatively loose binding of PshB to the PshA/PshA homodimer (Heinnickel et al., 2005) and PscB to the PscA/PscA homodimer (Jagannathan and Golbeck, 2008) is in striking contrast to the tight binding of PsaC to the PsaA/PsaB heterodimer in PSI (Antonkine et al., 2003; Parrett et al., 1989). The reason for the binding interface in PSI

was probed in a study that changed Arg_{52}, Arg_{65} and Tyr_{80} in PsaC to Ala_{52}, Ala_{65} and Ala_{80} respectively, thus generating variants that lack one or more of the key binding contacts with the PsaA/PsaB heterodimer (Jagannathan and Golbeck, 2009c). All of the reconstituted P_{700}-F_X/variant PsaC/PsaD complexes showed slower room temperature charge recombination kinetics between P_{700}^+ and $[F_A/F_B]^-$ and altered light-induced low temperature EPR spectra of the F_A/F_B clusters (Jagannathan and Golbeck, 2009c). All supported light-induced flavodoxin reduction, albeit at lower rates than native PSI complexes. However, the rates decreased significantly in the presence of atmospheric oxygen due to the degradation of F_A and F_B. The decrease in rates was more profound in the absence of the symmetric ionic contacts, which is not surprising given that they are in the vicinity of the F_A cluster (Jagannathan and Golbeck, 2009c). Differences in the EPR spectral line shapes and the rotational correlation times of a PsaC-bound spin label during the different stages of assembly further demonstrated the altered binding of the PsaC variants (Jagannathan et al., 2010). The lack of a single ionic bond or a symmetry-breaking H-bond forming contact on PsaC is clearly enough to disrupt the shielding needed to protect the F_A/F_B clusters against oxidative denaturation. The extensive network of contacts between PsaC and the PsaA/PsaB heterodimer is therefore critical to the retention of the F_A and F_B clusters, especially in a photosynthesizing cell that is supersaturated with molecular oxygen. Although it had previously been proposed that the close positioning of the N-terminus with the F_A cluster was the reason for the oxygen stability of PSI-bound PsaC (Antonkine et al. 2002, 2003), it appears that the tight binding interface between PsaC and the PsaA/PsaB heterodimer is instead the main reason for the stability of the F_A/F_B clusters to molecular oxygen (Jagannathan and Golbeck, 2009c). This hypothesis is consistent with the observation that the terminal iron-sulfur clusters in the green sulfur bacterial reaction center degrade under aerobic conditions, as indicated by the loss of the light-induced, low-temperature EPR resonances from F_A^- and F_B^- in whole cells of *Chlorobium tepidum* (Vassiliev et al., 2000).

B. The Evolution of the Acceptor Side of Type I Reaction Centers

The onset of oxygenic photosynthesis would have necessitated several changes in the organization of the ancestral Type I reaction center. In particular, the iron-sulfur clusters in the mobile dicluster ferredoxin would have become sensitive to denaturation due to the oxidation of the bridging μ-sulfido atoms by molecular oxygen to sulfur-zero (Golbeck et al., 1977). Here, we provide a possible scenario in the development of the electron acceptor side of Type I reaction centers, starting with the premise that the first photosynthetic organisms were anoxygenic prokaryotes that contained a mobile bacterial dicluster ferredoxin carrying the F_A and F_B clusters (Blankenship, 1992; Olson and Blankenship, 2004).

1. Modification of the Acceptor Side in Response to Oxygenic Photosynthesis

Because heliobacteria and green sulfur bacteria exist in a highly anaerobic environment, their terminal iron-sulfur clusters do not need to be protected against oxidative denaturation. In these organisms, the primary function of the binding interface between the membrane core and the F_A/F_B subunit is to ensure the shortest possible distance between the F_X and the F_A clusters, thereby guaranteeing that the rate of forward electron transfer exceeds the rate of charge recombination by at least several orders of magnitude. It appears that the close positioning of the F_X and F_A clusters can be achieved using very few contacts between the iron-sulfur protein and the membrane core of the homodimeric reaction center, which is likely responsible for the loose binding of PshB and PscB.

The emergence of oxygenic photosynthesis at ~2.7 Ga (Kasting, 2001; Olson, 2006) would have required significant modifications to protect the F_A/F_B clusters against oxidative denaturation. We propose that the advent of oxygenic photosynthesis forced the otherwise loosely-bound F_A/F_B polypeptide to become tightly bound so as to protect the F_A/F_B clusters against denaturation. The tightening of the binding interface would have been achieved by the development of an array of oppositely charged residues around the F_X and F_A clusters, respectively, thereby facilitating the

formation of strong ionic contacts and an oxygen-impenetrable interface (Fig. 12.7). Soluble bacterial dicluster ferredoxins do not contain charged amino acids around the cluster-ligating motif that coordinates the F_A-like cluster, a finding consistent with a highly specific role for these residues in PsaC. The charged residues around F_X and F_A are also missing in the corresponding regions of the heliobacterial and the green sulfur bacterial reaction centers, further highlighting the importance of these residues in establishing a tight binding interface.

2. The Issue of C_2-Symmetry in the Core Binding Region

The development of a tightly bound F_A/F_B protein had an unintended consequence: it broke the perfect C_2 symmetry of the ancestral Type I reaction center. In contemporary PSI, an axis of pseudo C_2-symmetry occurs along the PsaA/PsaB interface and passes through the center of the F_X cluster and P_{700}. The network of Arg/Lys-Asp ionic contacts between PsaC and PsaA/PsaB is highly C_2-symmetric, i.e. the ionic PsaC-PsaA contacts appear very similar to the ionic PsaC-PsaB contacts. Arg_{52} on PsaC forms five ionic bonds with Asp_{568}/Asp_{579} on PsaA, and Lys_{51}/Arg_{65} on PsaC form five ionic bonds with Asp_{555}/Asp_{566} on PsaB (Fig. 12.8, top) (Antonkine et al., 2003). If PsaC were rotated (*in silico*) 180° about the C_2 axis of symmetry, Arg_{52} on PsaC would still form five ionic bonds, but with Asp_{555}/Asp_{566} on PsaB, and Lys_{51}/Arg_{65} would still form five ionic bonds, but with Asp_{568}/Asp_{579} on PsaA (Antonkine et al., 2003). Given that the ζ-carbons of Arg_{52} and Arg_{65} in contemporary PsaC are superimposed when the C_2-symmetry operation is carried out (Antonkine et al., 2003), the distances between F_X, F_A and F_B would remain identical in the two symmetry-related orientations.

3. Recruitment of an Oxygen-Insensitive [2Fe-2S] Ferredoxin

The immobilization of the bacterial dicluster ferredoxin would have meant that an alternative, mobile protein would be required to shuttle the electrons from the terminal iron-sulfur cluster to ferredoxin: $NADP^+$ oxidoreductase and other target redox proteins. This led to the evolution of an oxygen-tolerant [2Fe-2S] ferredoxin (which, interestingly enough, was one of the very first electron transfer proteins to be discovered in photosynthetic tissue). The 1.7 Å NMR solution structure of ferredoxin from *Thermosynechococcus elongatus* shows that the protein consists of extended four-stranded β-sheet, a short two-stranded β-sheet, and three short α-helices (Baumann et al., 1996). The [2Fe-2S] cluster is located in a loop segment towards the outer edge of the protein in an appropriate location to carry out electron transfer (Baumann et al., 1996). Unlike typical bacterial dicluster ferredoxins, wherein a bridging μ-sulfido atom of the N-terminal [4Fe-4S] cluster is partially solvent exposed despite the proximity of the N-terminus (Jagannathan and Golbeck, 2009c), the sulfur atoms of the [2Fe-2S] cluster are well shielded by the protein.

The ensuing evolutionary challenge on the acceptor side of Type I reaction centers involved the development of a docking site for the newly-evolved, oxygen-tolerant [2Fe-2S] ferredoxin. Spectroscopic studies have shown that the [2Fe-2S] ferredoxin is reduced within 500 ns after the electron reaches the F_X cluster (Sétif, 2001). The [2Fe-2S] cluster in ferredoxin must be positioned close to the terminal [4Fe-4S] cluster in PsaC to support such a high rate of electron transfer. The sequence insertion of eight residues in the middle of the loop connecting the two cluster binding motifs in PsaC (see Fig. 12.6) constitutes part of the ferredoxin-docking domain. This insertion is not found in bacterial dicluster ferredoxins and hence is highly specific to PsaC. Also, Lys35 in the internal loop of PsaC has been found to be essential for fast electron transfer from PsaC to ferredoxin, as shown by an *in vivo psaC* mutant of *Chlamydomonas reinhardtii* (Meimberg et al., 1999). This Lys residue is strictly conserved in the PsaC proteins of PSI (Fig. 12.6), but it is not present in PshB or PscB.

There is also no indication of an *fdx* gene in the genomes of *Heliobacterium modesticaldum* or *Chlorobium tepidum* (Eisen et al., 2002; Sattley et al., 2008). Although the chlorosomes of green sulfur bacteria contain [2Fe-2S] cluster proteins (Vassilieva et al., 2001), they are significantly larger in mass than the plant-type [2Fe-2S] ferredoxins and likely adopt a different protein fold, consistent with a different function.

4. The Role of PsaD in the Transition to a Heterodimeric Core

The ferredoxin docking site would have been completed by the acquisition of the PsaD and PsaE subunits, which flank PsaC on either side on the membrane surface in contemporary photosynthetic organisms. PsaD is a 15-kDa hydrophilic polypeptide that fastens PsaC on the PsaA/PsaB heterodimer via a long stretch of amino acids on its C-terminus known as the C-clamp (Antonkine et al., 2003). This C-terminal domain plays a role in ferredoxin binding, in PSI assembly, and in efficient electron transport from F_A and F_B to the [2Fe-2S] clusters of ferredoxin (Lagoutte et al., 2001). Prior to the appearance of the X-ray crystal structure, chemical cross-linking assays had indicated that ferredoxin interacts with all three stromal proteins (Lelong et al., 1994, 1996). The X-ray crystal structure of cyanobacterial PSI shows a binding pocket for ferredoxin that is lined with positively charge residues formed by PsaC, PsaD, and PsaE (Fromme et al., 2001), which attract the complementary negative charges on ferredoxin. The importance of the involvement of PsaD in the docking of ferredoxin is highlighted by studies of PSI complexes isolated from an *in vivo psaD* deletion mutant of *Synechocystis* sp. PCC 6803. The variant PSI complex showed a lower binding affinity for ferredoxin and supports only ~10% of the rate of light-induced flavodoxin-mediated $NADP^+$ reduction compared to native PSI (Barth et al., 1998; Chitnis et al., 1996).

A polypeptide termed PscD, has been found in the green sulfur bacterial reaction center and was initially proposed play a role in ferredoxin-mediated $NADP^+$ reduction and/or in the transfer of light energy from the antenna pigments to the reaction center (Tsukatani et al., 2004). However, PscD from *Chlorobium tepidum* shows only about 10% identity and 25% similarity to the cyanobacterial PsaD. Furthermore, *pscD* deletion mutants support ~80% of the steady-state rate of ferredoxin mediated $NADP^+$ reduction compared to the wild-type (Tsukatani et al., 2004). Given the putative location of PscD, close to the Fenna-Mathews-Olsen protein on the cytoplasmic side (Hauska et al., 2001), it is possible that the primary role of PscD is to facilitate energy transfer from the chlorosomes to the reaction center. A *psaD*-like (or *pscD*-like) gene is not present in the genome of *H. modesticaldum*, further highlighting the rudimentary design of the heliobacterial reaction center (Sattley et al., 2008).

PsaE is an 8-kDa hydrophilic polypeptide present in plant and algal PSI. PsaE may have multiple roles; it assists in the docking of ferredoxin and flavodoxin (Mühlenhoff et al., 1996; Sétif et al., 2002), and it is involved in a cyclic electron transport pathway around PSI (Yu et al., 1993). There have been suggestions that PsaE acts to stabilize PSI polypeptides such as PsaF and PsaJ (Cohen et al., 1993). The *psaE* gene is not essential for the photoautotrophic growth of cyanobacteria (Zhao et al., 1993) and plants (Ihnatowicz et al., 2007). Nevertheless, a *psaE* deletion mutant in the cyanobacterium *Synechocystis* sp. PCC 6803 leads to increased oxidative stress and susceptibility to light-induced damage (Jeanjean et al., 2008). Considering the greater sensitivity of a *psaE* deletion mutant in *Arabidopsis thaliana* to oxidative stress (Ihnatowicz et al., 2007; Varotto et al., 2000), the PsaE protein may play a more significant role in higher plants than in cyanobacteria.

There is no *psaE*-like gene in either *C. tepidum* or *H. modesticaldum*, making it unlikely that a PsaE like protein exists in green sulfur bacteria or heliobacteria. Because the PsaE protein appears to play a protective role as a response to an oxygenated environment, it may not be necessary in the absence of oxygen.

C. Scenarios for the Development of Contemporary PSI

Based on the existing biochemical, biophysical and phylogenetic information on the electron acceptor side of Type I reaction centers, we envision two scenarios for the development of contemporary PSI over evolutionary time. In the first, the bacterial dicluster ferredoxin was recruited onto a pre-existing heterodimeric core to form PsaC. This scenario presupposes that the primordial reaction center consisted of a heterodimeric core or the transition from a homodimer to the heterodimer occurred before the evolution of modern PsaC. However, there is no evidence to support this model, and it is not clear what evolutionary pressure might have existed for the membrane core to undergo the transition from a simple, one-gene product into a heterodimer.

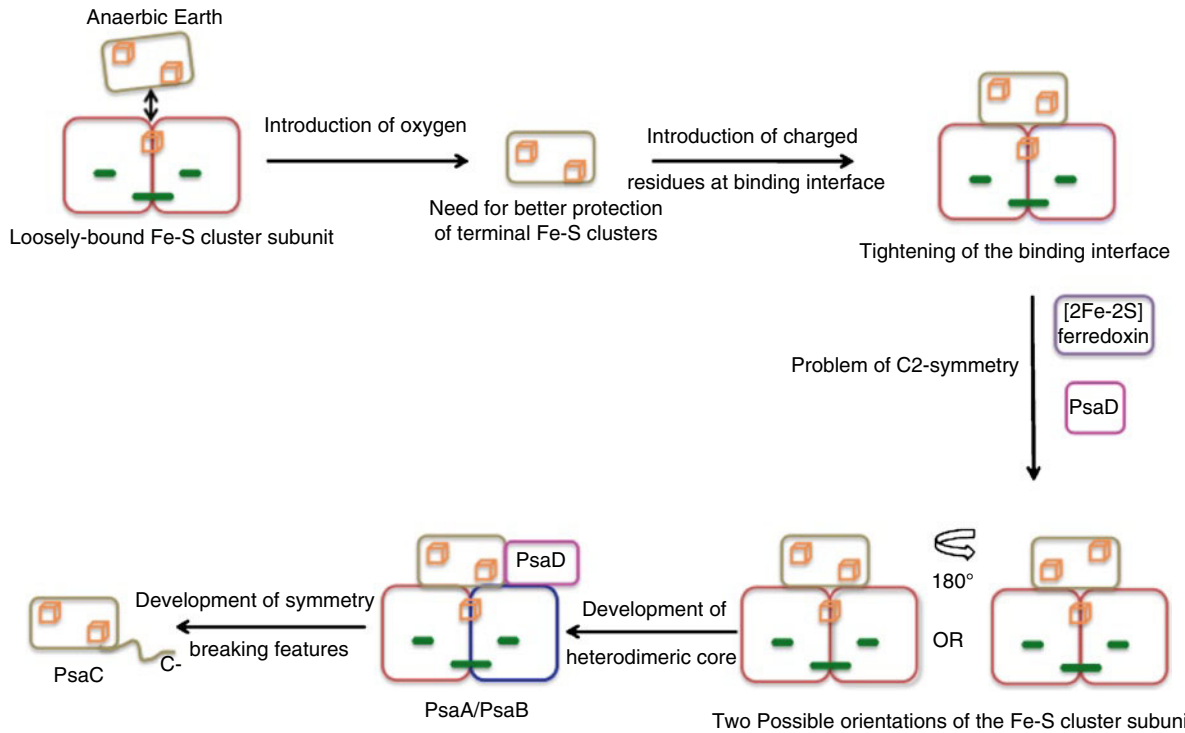


Fig. 12.9. Cartoon depicting the transition from a homodimeric Type I reaction center to a heterodimeric PSI reaction center under the selective pressure of oxygen.

In the second scenario, the bacterial dicluster ferredoxin was recruited onto a homodimeric core, thereby breaking the C_2-symmetry of the reaction center (Fig. 12.9). This model is in line with the proposal that a homodimer preceded a heterodimer during the evolution of photosynthetic reaction centers (Blankenship, 1992; Olson and Blankenship, 2004). As shown in the X-ray crystal structure of PSI, PsaC is positioned on the PsaA/PsaB heterodimer such that the axis through the two [4Fe-4S] clusters is tilted 62° from the membrane normal. The distance from F_X to F_A is 14.9 Å and the distance from F_X to F_B is 22 Å (Antonkine et al., 2003; Jordan et al., 2001). This configuration ensures linear electron transfer from $F_X \rightarrow F_A \rightarrow F_B \rightarrow$ ferredoxin. It should be noted that this binding configuration is equally possible on a homodimeric core as well as on a heterodimeric core. If the ancestor of PsaC were to bind to a homodimeric core at the onset of oxygenic photosynthesis, the overall C_2-symmetry of the reaction center would similarly have been broken.

While the tight binding of PsaC caused an initial disruption of C_2-symmetry, the subsequent recruitment of PsaD to complete the ferredoxin docking site may have generated the selective pressure that led to the differentiation of the reaction center core into separate PsaA and PsaB polypeptides (Fig. 12.9). The reason for this is that although the precursor to PsaC could function in either orientation on the homodimer, there is no corresponding symmetry element for PsaD. As described in detail in (Antonkine et al., 2003), the N-terminus of PsaD forms 9 H-bonds and five ionic bonds with PsaA and 2 H-bonds and two ionic bonds with PsaB. The C-terminus of PsaD forms 5 H-bonds and a single ionic bond with PsaB and is also involved in a network of contacts with PsaC that comprise the so-called C-clamp. If PsaD were rotated 180° about the PSI symmetry axis, only a few of these extensive contacts would form. Hence, PsaD requires a differentiation of binding sites on the two proteins that comprise the reaction center core. This may have provided the selective pressure to differentiate the reaction center core into two proteins after a gene duplication event. Thus, in this scenario, the tight binding of PsaC acted as the trigger and the recruitment of PsaD provided

the incentive for the membrane core to undergo the transition from a homodimer to a heterodimer.

Because the body of PsaD binds largely to the PsaB protein, the counterpart position became available on the PsaA protein to provide a docking site for PsaE. Similar to PsaD, PsaE has no symmetry relationship with the PsaA/PsaB heterodimer; PsaE forms eight H-bonds with PsaA and four H-bond with PsaB (Antonkine et al., 2003). The presence of a heterodimeric core with asymmetrically bound PsaD and PsaE would then have put intense selective pressure on PsaC to orient in one of two possible orientations on the PsaA/PsaB heterodimer. To be sure, PsaD would lose most of its contacts with the PsaA/PsaB heterodimer if it were to bind to the PSI core that contained PsaC bound in the rotated orientation (Antonkine et al., 2003). However, as we have seen, the central binding region of PsaC above the F_X cluster is C_2-symmetric. So, how did PsaC come to bind asymmetrically on the PSI heterodimer?

1. The Binding of PsaC in a Single Orientation on the PsaA/PsaB Heterodimer

As shown in the 2.5 Å X-ray crystal structure, the PSI heterodimer retains a very high degree of C_2-symmetry surrounding the F_X cluster. In the absence of a symmetry breaking element, the presence of a perfectly symmetric binding surface would lead to a 50% probability that PsaC would bind to the PSI core in one orientation, and a 50% probability that PsaC would bind to the PSI core in the 180°-rotated orientation. Yet, the X-ray crystal structure shows that PsaC is bound in only one of the two possible orientations (i.e. with the F_A cluster closer to Leu_{706} on PsaB than to Leu_{722} on PsaA). Furthermore, PsaD and PsaE can assemble on their binding sites on the PSI core only if PsaC is bound in the proper orientation (i.e. with the F_A cluster closer to Leu_{706} on PsaB) (Antonkine et al., 2003), making it highly unlikely that a second isoform of PSI remains to be discovered with PsaC oriented 180° opposite of that already found. Because PsaC assembles prior to PsaD and PsaE (Chitnis et al., 1996; Yu et al., 1995; Zhao et al., 1993), the problem becomes one of how PsaC can distinguish between the two otherwise-identical orientations on the PsaA/PsaB interface so as to dock in the correct orientation for PsaD and PsaE to properly bind.

Ultimately, the C-terminal domain on PsaC is the feature that breaks the otherwise perfect C_2-symmetry. A hydrophobic binding pocket exists on the PsaB protein (Fig. 12.8, bottom) in which three H-bonds are formed between Thr_{73}/Tyr_{80} on the C-terminus of the PsaC and $Gln_{678}/Lys_{702}/Pro_{703}$ on PsaB (Antonkine et al., 2003). The binding pocket on PsaB includes the sequence $Lys_{702}/Pro_{703}/Val_{704}$, whereas the equivalent segment on PsaA includes the sequence $Gln_{718}/Pro_{719}/Arg_{720}$. The presence of Gln_{718} and Arg_{720} on PsaA results in steric hindrance with Thr_{73}/Tyr_{80} on the C-terminus of the PsaC, thereby disallowing binding in the 180° orientation. While this solves the problem of breaking symmetry, it opens yet another question: which contacts are established first, the ionic contacts in the C_2-symmetric region or the H-bonds on the C-terminus. To put it explicitly, something must prevent the ionic contacts from forming first so that the protein does not bind equally in both orientations. Certainly, after the initial contacts are established between PsaC and PsaA/PsaB, the conformational dynamics of PsaC will become restricted and there will be a limited number of routes to establish the remainder of the contacts.

A comparison of the structures of unbound PsaC (Antonkine et al., 2002) and PSI-bound PsaC reveals significant structural differences, especially in the N- and C-terminal regions of the protein (Antonkine et al., 2003). The NMR solution structure of unbound PsaC shows that the C-terminus adopts a helically coiled conformation, and that the N-terminus, in contrast to its position in bacterial dicluster ferredoxins, is positioned perpendicular to the pre-C-terminal region away from the F_A cluster (Antonkine et al., 2002). In the PSI-bound state, the C-terminal region of PsaC assumes an extended conformation and is immobilized on the heterodimeric membrane core of PSI (Antonkine et al., 2003; Jordan et al., 2001). The N-terminus moves physically closer to F_A and forms an antiparallel β-sheet with the pre C-terminus (Antonkine et al., 2003), an arrangement typical of dicluster ferredoxins (Sticht and Rösch, 1998).

The N-terminus of PsaC contains an extension of two amino acids that is not present in bacterial dicluster ferredoxins (Sticht and Rösch, 1998).

The positioning of the N-terminus (in particular the two amino acid extension) in unbound PsaC pushes the strands of polypeptide containing the binding residues Lys_{51}, Arg_{52}, and Arg_{65} away from their equilibrium positions relative to that in bound PsaC. This has the (intended) consequence that Lys_{51}, Arg_{52}, and Arg_{65} are no longer oriented so as to form the ten ionic bonds with Asp_{555}/Asp_{566} on PsaB and Asp_{568}/Asp_{579} on PsaA (Jagannathan and Golbeck, 2009b). The misaligning of the ionic bond forming neatly precludes the possibility that the symmetric region of PsaC bind first.

Because binding in the symmetric region is prevented by the bent N-terminus, the symmetry-breaking H-bonds between the C-terminus of PsaC and the hydrophobic binding pocket on PsaB will be the first to be formed. This guarantees that symmetry is broken immediately and the correct orientation on the PsaA/PsaB heterodimer is established at the onset of PsaC binding. The binding of PsaC relies on the ability of the C-terminal Tyr_{80} residue to locate its H-bonding partners Gln_{678} and Pro_{703} on PsaB. The flexibility of the extended C-terminus allows it to explore a large conformational space in order to locate the H-bonding residues on PsaB. The C-terminus of PsaC contains a Gly_{77}/Leu_{78}/Ala_{79}/Tyr_{80} sequence that could participate in hydrophobic interactions with the Pro_{703}/Val_{704}/Ala_{705}/Leu_{706} sequence on PsaB. We suspect that the H-bonds are not formed instantly and that the hydrophobic amino acids aggregate and undergo minor realignments to maximize the hydrophobic interactions (Jagannathan and Golbeck, 2009b). The H-bonds between Tyr_{80}/Thr_{73} on PsaC and Gln_{678}/Lys_{702}/Pro_{703} would subsequently form to lock the system into a thermodynamic minimum. The loss of structured water molecules due to the fusion of the large hydrophobic surface on PsaB with the C-terminal region of PsaC may provide a significant entropic contribution to the driving force in the initial steps of the binding process.

The uncoiling and binding of the C-terminus may impose considerable strain on the pre C-terminus, given that the entire PsaC protein will trail behind the tethered C-terminus. The energy stored in the pre C-terminal region of PsaC must be released for the system to attain thermodynamic stability. The pre C-terminus (amino acids 64–68) would need to twist to release the strain caused by the extended C-terminus. This twisting action would position the pre C-terminal backbone to form H-bonds with the N-terminal backbone. The formation of these bonds would require the N-terminus (amino acids 2–5) to be pulled towards the F_A cluster and straighten, thereby placing it in a position to form the antiparallel β-sheet with the pre-C-terminus. The establishment of the antiparallel β-sheet between the N- and pre-C-terminus would be accompanied by the formation of multiple H-bonds, thereby providing a strong enthalpic contribution to the driving force for binding (Jagannathan and Golbeck, 2009b).

In the 30 minimized NMR structures of unbound PsaC, one shows the N-terminus of PsaC positioned behind F_A. This may be evidence for the presence of a dynamic equilibrium between the bent and straightened forms of the N-terminus, which would readily allow PsaC to realign upon the binding of the C-terminus to its specific pocket on PsaB. The movement of the N-terminus of PsaC would relax the strands of polypeptide containing the binding residues Lys_{51}, Arg_{52}, and Arg_{65} and position them to form the ionic contacts with Asp_{555}/Asp_{566} on PsaB and Asp_{568}/Asp_{579} on PsaA. It is tempting to speculate that the ionic bonds are formed in a sequential manner, with the first contact being established between the residue closest to the PsaC C-terminus, Arg_{65}, and Asp_{555}/Asp_{566} on PsaB. The final docking of PsaC would then occur through ionic bond formation between residues Arg_{52} of PsaC and Asp_{568}/Asp_{579} of PsaA, and between Lys_{51} of PsaC and Asp_{566} of PsaB. The free energy of docking would be driven largely by the enthalpic contribution to the free energy of ionic bond formation. It could also be predicted that as PsaC presses down on PsaA/PsaB, it displaces water molecules that are structured near the hydrophobic regions of all three proteins, thereby driving the binding through an additional entropic contribution to the free energy change.

2. Inversion of the Redox Potentials of F_A and F_B

There is one additional consequence of a tightly bound PsaC in terms of a second adaptation to oxygenic photosynthesis. Consensus values for

the experimental (Srinivasan and Golbeck, 2009) and calculated (Ptushenko et al., 2008) redox potentials of F_X, F_A, F_B and ferredoxin are ca. −620 (−654) mV, −530 (−481) mV, −580 (−585) mV, and −420 mV. Because the sequence of electron transfer in PSI is $F_X \rightarrow F_A \rightarrow F_B \rightarrow$ ferredoxin (DiazQuintana et al., 1998; Vassiliev et al., 1998), there exists a thermodynamically unfavorable electron transfer step from $F_A \rightarrow F_B$. Now, this would ordinarily not be a problem because the following $F_B \rightarrow$ ferredoxin electron transfer step is downhill, making the overall $F_A \rightarrow$ ferredoxin electron transfer thermodynamically favorable. Nevertheless, the uphill electron transfer step has two consequences, one minor and one major. The minor consequence is that a Boltzmann penalty would be incurred which would slow the $F_A \rightarrow F_B$ electron transfer. However, because the (slower) forward electron transfer is still many orders of magnitude greater than the charge recombination with P_{700}^+ (Vassiliev et al., 1997), a continued high quantum yield is guaranteed. The major consequence is that the electron would reside on F_A ~90% of the time and on F_B ~10% of the time given that their equilibrium midpoint potentials differ by ~60 mV. At first glance, it appears counterintuitive that the electron would not be available most of the time on F_B for further transfer to the [2Fe-2S] ferredoxin. However, because F_A is located deep in the protein, it is inaccessible to oxygen, whereas F_B is located at the surface near the ferredoxin docking site. In a saturated oxygen environment, surface-exposed F_B cluster could donate an electron to molecular oxygen by simple diffusion. Because the collisional residence time is short, the majority of the time the oxygen molecule will not find an electron on F_B. Ferredoxin, on the other hand, occupies a binding site on the PsaC/PsaD/PsaE surface, thereby providing enough time for the electron on F_A to pass via F_B to the [2Fe-2S] cluster of ferredoxin. The binding of ferredoxin can, in principle, alter the midpoint potential of the F_B cluster, further driving electron transfer from $F_A \rightarrow$ ferredoxin. It therefore appears that the purpose for the inverted redox potentials of F_A and F_B is to prevent wasteful (and potentially harmful) electron transfer to oxygen, thereby ensuring a very high quantum yield of electron transfer to ferredoxin, and ultimately, to $NADP^+$.

IV. Concluding Remarks

The evolution of the acceptor side of PSI may be an example of how an altered set of conditions resulted in a retrofitting of an existing electron transfer protein to continue a function that would otherwise not be possible in a new environment. The evolution of PsaC from a soluble dicluster bacterial ferredoxin may have involved the following steps: (i) The introduction of positively charged residues in the vicinity of the F_A cluster and negatively charged residues in the vicinity of the F_X cluster that would eventually form a tight network of ionic bonds; (ii) The modification and extension of the N-terminus by two amino acids to splay the polypeptide strands containing the symmetric ionic bond forming contacts, thus precluding the binding of PsaC in the 180º-rotated orientation; (iii) The introduction of a hydrophobic C-terminus as the symmetry-breaking feature and a specific binding pocket on PsaB (but not PsaA) that allows PsaC to initially recognize one of two possible orientations; (iv) The insertion of eight amino acids in the loop region connecting the two cluster binding motifs to provide a portion of the docking surface for the oxygen-tolerant [2Fe-2S] ferredoxin. By incorporating these relatively minor changes in the amino acid sequence of the existing mobile, bacterial dicluster ferredoxin, the protection of the terminal F_A/F_B clusters against oxidative degradation was ensured by a tight binding interface to the membrane core of the reaction center.

The trigger for the transition from a homodimeric core to a heterodimeric core would have been the recruitment of the tightly bound bacterial dicluster ferredoxin, which by definition broke the perfect C_2-symmetry of the reaction center (Fig. 12.9). The recruitment of PsaD to construct the binding site for the oxygen-insensitive [2Fe-2S] ferredoxin would have provided the selective pressure to differentiate the reaction center into separate PsaA and PsaB polypeptides. Finally, the inversion of the redox potentials of F_A and F_B ensured that the electron would be preferentially transferred to the soluble [2Fe-2S] ferredoxin rather than to oxygen in the supersaturated solution.

These changes opened the door to further elaboration and differentiation of the PSI reaction center, such as adding small, membrane-intrinsic

polypeptides to bind additional chlorophylls, and generating binding sites for the LHCI and the LHCII antenna complexes that allowed plants and algae to cope with changing light conditions through the mechanism of state transitions (see Boekema et al., 2006; Croce et al., 2006; for reviews). Thus, through a progression of gene evolution and structural improvements starting from simple Type I reaction centers in phototropic anaerobes and proceeding to PSI supercomplexes in higher plants, evolution has tuned the machinery of photoconversion to maximize excitation energy utilization and to improve electron transfer efficiency in a wide variety of cellular and physical environments.

Acknowledgments

This work was supported by grants from the NSF (MCB-1021725) and the DOE (DE-FG-02-98-ER20314).

References

Allen JF (2005) A redox switch hypothesis for the origin of two light reactions in photosynthesis. FEBS Lett 579: 963–968

Allen JF and Martin W (2007) Evolutionary biology: out of thin air. Nature 445: 610–612

Allen JP and Williams JC (1998) Photosynthetic reaction centers. FEBS Lett 438: 5–9

Amesz J (1995) The heliobacteria, a new group of photosynthetic bacteria. J Photochem Photobiol 30: 89–96

Amunts A, Drory O and Nelson N (2007) The structure of a plant photosystem I supercomplex at 3.4 Å resolution. Nature 447: 58–63

Antonkine ML, Liu G, Bentrop D, Bryant DA, Bertini I, Luchinat C, Golbeck JH and Stehlik D (2002) Solution structure of the unbound, oxidized photosystem I subunit PsaC, containing [4Fe-4S] clusters F_A and F_B: a conformational change occurs upon binding to photosystem I. J Biol Inorg Chem 7: 461–472

Antonkine ML, Jordan P, Fromme P, Krauß N, Golbeck JH and Stehlik D (2003) Assembly of protein subunits within the stromal ridge of photosystem I. Structural changes between unbound and sequentially PSI-bound polypeptides and correlated changes of the magnetic properties of the terminal iron sulfur clusters. J Mol Biol 327: 671–697

Barth P, Lagoutte B and Sétif P (1998) Ferredoxin reduction by photosystem I from *Synechocystis* sp. PCC 6803: toward an understanding of the respective roles of subunits PsaD and PsaE in ferredoxin binding. Biochemistry 37: 16233–16241

Baumann B, Sticht H, Scharpf M, Sutter M, Haehnel W and Rösch P (1996) Structure of *Synechococcus elongatus* [Fe_2S_2] ferredoxin in solution. Biochemistry 35: 12831–12841

Baymann F, Brugna M, Mühlenhoff U and Nitschke W (2001) Daddy, where did (PS)I come from? Biochim Biophys Acta 1507: 291–310

Ben-Shem A, Frolow F and Nelson N (2003) Crystal structure of plant photosystem I. Nature 426: 630–5

Blankenship R (1992) Origin and early evolution of photosynthesis. Photosynth Res 33: 91–111

Blankenship RE (2001) Molecular evidence for the evolution of photosynthesis. Trends Plant Sci 6: 4–6

Blankenship R (2002) Molecular Mechanisms of Photosynthesis. Blackwell Science, Ltd, London

Boekema EJ, Kouril R, Dekker JP and Jensen PE (2006) Association of photosystem I and light-harvesting complex II during state transitions. In: Golbeck JH (ed.) Photosystem I: The Light-Induced Plastocyanin:Ferredoxin Oxidoreductase, pp 41–46. Springer, Dordrecht

Brettel K and Leibl W (2001) Electron transfer in photosystem I. Biochim Biophys Acta 1507: 100–114

Brocks JJ, Logan GA, Buick R and Summons RE (1999) Archean molecular fossils and the early rise of eukaryotes. Science 285: 1033–1036

Bryant DA, Costas AM, Maresca JA, Chew AG, Klatt CG, Bateson MM, Tallon LJ, Hostetler J, Nelson WC, Heidelberg JF and Ward DM (2007) Candidatus *Chloracidobacterium thermophilum*: an aerobic phototrophic Acidobacterium. Science 317: 523–526

Büttner M, Xie DL, Nelson H, Pinther W, Hauska G and Nelson N (1992) Photosynthetic reaction center genes in green sulfur bacteria and in photosystem I are related. Proc Natl Acad Sci USA 89: 8135–8139

Catling DC, Zahnle KJ and McKay C (2001) Biogenic methane, hydrogen escape, and the irreversible oxidation of early Earth. Science 293: 839–843

Chamorovsky SK and Cammack R (1982) Direct determination of the midpoint potential of the acceptor X in chloroplast photosystem I by electrochemical reduction and ESR spectroscopy. Photochem Photobiol 4: 195–200

Chitnis VP, Jung YS, Albee L, Golbeck JH and Chitnis PR (1996) Mutational analysis of photosystem I polypeptides - role of PsaD and the lysyl 106 residue in the reductase activity of photosystem I. J Biol Chem 271: 11772–11780

Cohen Y, Chitnis VP, Nechushtai R and Chitnis PR (1993) Stable assembly of PsaE into cyanobacterial photosynthetic membranes is dependent on the presence of other accessory subunits of photosystem I. Plant Mol Biol 23: 895–900

Cohen RO, Shen G, Golbeck JH, Xu W, Chitnis PR, Valieva AI, van der Est A, Pushkar Y and Stehlik D (2004)

Evidence for asymmetric electron transfer in cyanobacterial photosystem I: analysis of a methionine-to-leucine mutation of the ligand to the primary electron acceptor A_0. Biochemistry 43: 4741–4754

Croce R, Morosinotto T and Bassi R (2006) LHC1: The antenna complex of photosystem I in plants and green algae. In: Golbeck JH (ed) Photosystem I: The Light-Induced Plastocyanin:Ferredoxin Oxidoreductase, pp 119–137. Springer, Dordrecht

Dashdorj N, Xu W, Cohen RO, Golbeck JH and Savikhin S (2005) Asymmetric electron transfer in cyanobacterial photosystem I: charge separation and secondary electron transfer dynamics of mutations near the primary electron acceptor A_0. Biophys J 88: 1238–1249

Davis IH, Heathcote P, MacLachlan DJ and Evans MCW (1993) Modulation analysis of the electron spin echo signals of *in vivo* oxidized primary donor ^{14}N chlorophyll centers in bacterial, P870, P960, and plant photosystem I, P700 reaction centers. Biochim Biophys Acta 1143: 183–189

DePamphilis BV, Averill BA, Herskovitz T, Que LJ and Holm R (1974) Synthetic analogs of the active sites of iron-sulfur proteins. VI. Spectral and redox characteristics of the tetranuclear clusters $[Fe_4S_4(SR)_4]^{2-}$. J Am Chem Soc 96: 4159–4167

DiazQuintana A, Leibl W, Bottin H and Sétif P (1998) Electron transfer in photosystem I reaction centers follows a linear pathway in which iron-sulfur cluster F_B is the immediate electron donor to soluble ferredoxin. Biochemistry 37: 3429–3439

Ehrenreich A and Widdel F (1994) Anaerobic oxidation of ferrous iron by purple bacteria, a new type of phototrophic metabolism. Appl Environ Microbiol 60: 4517–4526

Eisen JA, Nelson KE, Paulsen IT, Heidelberg JF, Wu M, Dodson RJ, Deboy R, Gwinn ML, Nelson WC, Haft DH, Hickey EK, Peterson JD, Durkin AS, Kolonay JL, Yang F, Holt I, Umayam LA, Mason T, Brenner M, Shea TP, Parksey D, Nierman WC, Feldblyum TV, Hansen CL, Craven MB, Radune D, Vamathevan J, Khouri H, White O, Gruber TM, Ketchum KA, Venter JC, Tettelin H, Bryant DA and Fraser CM (2002) The complete genome sequence of *Chlorobium tepidum* TLS, a photosynthetic, anaerobic, green-sulfur bacterium. Proc Natl Acad Sci USA 99: 9509–9514

Ferreira KN, Iverson TM, Maghlaoui K, Barber J and Iwata S (2004) Architecture of the photosynthetic oxygen-evolving center. Science 303: 1831–1838

Figueras JB, Cox RP, Hojrup P, Permentier HP and Miller M (2002) Phylogeny of the PscB reaction center protein from green sulfur bacteria. Photosynth Res 71: 155–164

Frankenberg N, Hager-Braun C, Feiler U, Fuhrmann M, Rogl H, Schneebauer N, Nelson N and Hauska G (1996) P840-reaction centers from *Chlorobium tepidum*-quinone analysis and functional reconstitution into lipid vesicles. Photochem Photobiol 64: 14–19

Fromme P (1996) Structure and function of photosystem I. Curr Opin Struct Biol 6: 473–484

Fromme P, Witt HT, Schubert WD, Klukas O, Saenger W and Krauß N (1996) Structure of photosystem I at 4.5 Å resolution: a short review including evolutionary aspects. Biochim Biophys Acta 1275: 76–83

Fromme P, Jordan P and Krauß N (2001) Structure of photosystem I. Biochim Biophys Acta 1507: 5–31

Golbeck J (1992) Structure and function of photosystem I. Annu Rev Plant Physiol Plant Mol Biol 43: 293–324

Golbeck JH (1993) Shared thematic elements in photochemical reaction centers. Proc Natl Acad Sci. USA 90: 1642–1646

Golbeck JH (1994) Photosystem I in Cyanobacteria. In: Bryant DA (ed) The Molecular Biology of Cyanobacteria, pp 179–220. Kluwer Academic Publishers, Dordrecht

Golbeck JH (1999) A comparative analysis of the spin state distribution of *in vivo* and *in vitro* mutants of PsaC. A biochemical argument for the sequence of electron transfer in Photosystem I as F_X -> F_A -> F_B -> ferredoxin/flavodoxin. Photosynth Res 61: 107–149

Golbeck JH, Lien S and San Pietro A (1977) Isolation and characterization of a subchloroplast particle enriched in iron-sulfur protein and P700. Arch Biochem Biophys 178: 140–150

Gunner MR, Robertson DE and Dutton PL (1986) Kinetic studies on the reaction center protein from *Rhodopseudomonas sphaeroides*: the temperature and free energy dependence of electron transfer between various quinones in the Q_A site and the oxidized bacteriochlorophyll dimer. J Phys Chem 90: 3783–3795

Gupta RS (2003) Evolutionary relationships among photosynthetic bacteria. Photosynth Res 76: 173–183

Gupta RS, Mukhtar T and Singh B (1999) Evolutionary relationships among photosynthetic prokaryotes (*Heliobacterium chlorum*, *Chloroflexus aurantiacus*, cyanobacteria, *Chlorobium tepidum* and proteobacteria): implications regarding the origin of photosynthesis. Mol Microbiol 32: 893–906

Hager-Braun C, Zimmermann R and Hauska G (1999) The homodimeric reaction center of Chlorobium. In: Peschek G, Loeffelhardt W, and Schmetterer G (eds) Phototrophic Prokaryotes, pp 169–181. Kluwer Academic Publishers, New York

Hager-Braun C, Jarosch U, Hauska G, Nitschke W and Riedel A (1997) EPR studies of the terminal electron acceptors of the green sulfur bacterial reaction centre. Revisited. Photosynth Res 51: 127–136

Hallahan BJ, Purton S, Ivison A, Wright D and Evans MCW (1995) Analysis of the proposed Fe-Sx binding region of photosystem I by site-directed mutation of PsaA in *Chlamydomonas reinhardtii*. Photosynth Res 46: 257–264

Hatano A, Seo D, Kitashima M, Sakurai H and Inoue K (2005) Purification of two ferredoxins and cloning of these genes from the photosynthetic bacterium *Heliobacillus mobilis*. In: van der Est A and Bruce D (eds) 13th International

Congress on Photosynthesis, pp 83–85. Allen Press Inc., Lawrence, Kansas

Hauska G (1988) Phylloquinone in photosystem I: are quinones the secondary electron acceptors in all types of photosynthetic reaction centers? Trends Biochem Sci 13: 415–416

Hauska G, Schoedl T, Remigy H and Tsiotis G (2001) The reaction center of green sulfur bacteria. Biochim Biophys Acta 1507: 260–277

Heathcote P, Fyfe PK and Jones MR (2002) Reaction centres: the structure and evolution of biological solar power. Trends Biochem Sci 27: 79–87

Heinnickel M and Golbeck JH (2007) Heliobacterial photosynthesis. Photosynth Res 92: 35–53

Heinnickel M, Shen G, Agalarov R and Golbeck JH (2005) Resolution and reconstitution of a bound Fe-S protein from the photosynthetic reaction center of *Heliobacterium modesticaldum*. Biochemistry 44: 9950–9960

Heinnickel M, Agalarov R, Svensen N, Krebs C and Golbeck JH (2006) Identification of F_X in the heliobacterial reaction center as a [4Fe-4S] cluster with an S = 3/2 ground spin state. Biochemistry 45: 6756–6764

Heinnickel M, Shen, G and Golbeck JH (2007) Identification and characterization of PshB, the dicluster ferredoxin that harbors the terminal electron acceptors F_A and F_B in *Heliobacterium modesticaldum*. Biochemistry 46: 2530–2536

Heising S and Schink B (1998) Phototrophic oxidation of ferrous iron by a *Rhodomicrobium vannielii* strain. Microbiology 144: 2263–2269

Heising S, Richter L, Ludwig W and Schink B (1999) *Chlorobium ferrooxidan*s sp. nov., a phototrophic green sulfur bacterium that oxidizes ferrous iron in coculture with a "*Geospirillum*" sp. strain. Arch Microbiol 172: 116–124

Holzwarth AR, Müller MG, Niklas J and Lubitz W (2006) Ultrafast transient absorption studies on photosystem I reaction centers from *Chlamydomonas reinhardtii*. 2: mutations near the P700 reaction center chlorophylls provide new insight into the nature of the primary electron donor. Biophys J 90: 552–565

Ihnatowicz A, Pesaresi P and Leister D (2007) The E subunit of photosystem I is not essential for linear electron flow and photoautotrophic growth in *Arabidopsis thaliana*. Planta 226: 889–895

Inoue H, Tsuchiya T, Satoh S, Miyashita H, Kaneko T, Tabata S, Tanaka A and Mimuro M (2004) Unique constitution of photosystem I with a novel subunit in the cyanobacterium *Gloeobacter violaceus* PCC 7421. FEBS Lett 578: 275–279

Ishikita H and Knapp EW (2003) Redox potential of quinones in both electron transfer branches of photosystem I. J Biol Chem 278: 52002–52011

Iwaki M and Itoh S (1994) Reaction of reconstituted acceptor quinone and dynamic equilibration of electron transfer in the photosystem I reaction center. Plant Cell Physiol 35: 983–993

Jagannathan B and Golbeck JH (2008) Unifying principles in homodimeric type I photosynthetic reaction centers: Properties of PscB and the F_A, F_B and F_X iron-sulfur clusters in green sulfur bacteria. Biochim Biophys Acta 1777: 1535–1544

Jagannathan B and Golbeck JH (2009a) Photosynthesis: Microbial. In: Schaechter M (ed) Encyclopedia of Microbiology, 3rd Edition, pp 325–341. Elsevier, London

Jagannathan B and Golbeck JH (2009b) Breaking biological symmetry in membrane proteins: the asymmetrical orientation of PsaC on the pseudo-C_2 symmetric photosystem I core. Cell Mol Life Sci 66: 1257–1270

Jagannathan B and Golbeck JH (2009c) Understanding of the binding interface between PsaC and the PsaA/PsaB heterodimer in photosystem I. Biochemistry 48: 5405–5416

Jagannathan B, Dekat S, Golbeck JH and Lakshmi KV (2010) The assembly of a multi-subunit photosynthetic membrane protein complex: A site-specific spin labeling EPR spectroscopic study of the PsaC subunit in photosystem I. Biochemistry 49: 2398–2408

Jeanjean R, Latifi A, Matthijs HCP and Havaux M (2008) The PsaE subunit of photosystem I prevents light-induced formation of reduced oxygen species in the cyanobacterium *Synechocystis* sp. PCC 6803. Biochim Biophys Acta 1777: 308–316

Jordan P, Fromme P, Witt HT, Klukas O, Saenger W and Krauß N (2001) Three dimensional structure of photosystem I at 2.5 Å resolution. Nature 411: 909–917

Käss H, Fromme P and Lubitz W (1996) Quadrupole parameters of nitrogen nuclei in the cation radical $P700^{\bullet+}$ determined by ESEEM of single crystals of photosystem I. Chem Phys Lett 257: 197–206

Käss H, Fromme P, Witt HT and Lubitz W (2001) Orientation and electronic structure of the primary donor cation radical $P700^+$ in photosystem I: A single crystals EPR and ENDOR study. J Phys Chem 105: 1225–1239

Kasting JF (2001) Earth history. The rise of atmospheric oxygen. Science 293: 819–820

Kaupp M (2002) The function of photosystem I. Quantum chemical insight into the role of tryptophan quinone interactions. Biochemistry 41: 2895–2900

Kerr RA (2005) Earth science. The story of O_2. Science 308: 1730–1732

Kirschvink JL and Kopp RE (2008) Palaeoproterozoic ice houses and the evolution of oxygen-mediating enzymes: the case for a late origin of photosystem II. Philos Trans R Soc Lond B Biol Sci 363: 2755–2765

Kjaer B and Scheller HV (1996) An isolated reaction center complex from the green sulfur bacterium *Chlorobium vibioforme* can photoreduce ferredoxin at high rates. Photosynth Res 47: 33–39

Kjaer B, Frigaard NU, Yang F, Zybailov B, Miller M, Golbeck JH and Scheller HV (1998) Menaquinone-7 in the reaction center complex of the green sulfur bacterium *Chlorobium vibrioforme* functions as the electron acceptor A_1. Biochemistry 37: 3237–3242

Kobayashi M, van de Meent EJ, Erkelens C, Amesz J, Ikegami I and Watanabe T (1991a) Bacteriochlorophyll *g* epimer as a possible reaction center component of heliobacteria. Biochim Biophys Acta 1057: 89–96

Kobayashi M, Watanabe T, Ikegami I, van de Meent EJ and Amesz J (1991b) Enrichment of bacteriochlorophyll *g*' in membranes of *Heliobacterium chlorum* by ether extraction. Unequivocal evidence for its existence *in vivo*. FEBS Lett 284: 129–31

Kobayashi M, Oh-Oka H, Akutsu S, Akiyama M, Tominaga K, Kise H, Nishida F, Watanabe T, Amesz J, Koizumi M, Ishida N and Kano H (2000) The primary electron acceptor of green sulfur bacteria, bacteriochlorophyll 663, is chlorophyll a esterified with Δ2,6-phytadienol. Photosynth Res 63: 269–280

Krieger A, Rutherford AW and Johnson GN (1995) On the determination of redox midpoint potential of the primary quinone electron acceptor, Q_A, in photosystem II. Biochim Biophys Acta 1229: 193–201

Lagoutte B, Hanley J and Bottin H (2001) Multiple functions for the C terminus of the PsaD subunit in the cyanobacterial photosystem I complex. Plant Physiol 126: 307–316

Lelong C, Sétif P, Lagoutte B and Bottin H (1994) Identification of the amino acids involved in the functional interaction between photosystem I and ferredoxin from *Synechocystis* sp. PCC 6803 by chemical cross-linking. J Biol Chem 269: 10034–10039

Lelong C, Boekema EJ, Kruip J, Bottin H, Rögner M and Sétif P (1996) Characterization of a redox active cross-linked complex between cyanobacterial photosystem I and soluble ferredoxin. EMBO J 15: 2160–2168

Liebl U, Mockensturm-Wilson M, Trost JT, Brune DC, Blankenship RE and Vermaas W (1993) Single core polypeptide in the reaction center of the photosynthetic bacterium *Heliobacillus mobilis* - structural implications and relations to other photosystems. Proc Natl Acad Sci USA 90: 7124–7128

Lindahl, P. A., Day, E. P., Kent, T. A., Orme-Johnson, W. H. and Munck, E. (1985) Mossbauer, EPR, and magnetization studies of the *Azotobacter vinelandii* Fe protein. Evidence for a $[4Fe\text{-}4S]^{1+}$ cluster with spin $S = 3/2$. J Biol Chem 260: 11160–73

Lockhart PJ, Steel MA and Larkum AW (1996) Gene duplication and the evolution of photosynthetic reaction center proteins. FEBS Lett 385: 193–196

Lubitz W (2006) EPR studies of the primary electron donor P700 in photosystem I. In: Golbeck JH (ed.) Advances in Photosynthesis (Series editor Govindjee). Photosystem I. The Plastocyanin: Ferredoxin Oxidoreductase in Photosynthesis, pp 245–269 Kluwer Academic Publishers, Dordrecht

Mac M, Bowlby NR, Babcock GT and McCracken J (1998) Monomeric spin density distrubution in the primary donor of photosystem I as determined by electron magnetic resonance: functional and thermodynamic implications. J Am Chem Soc 120: 13215–13223

Martin W and Russell MJ (2003) On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. Philos Trans Biol Sci 358: 59–85

Meimberg K, Fischer N, Rochaix JD and Mühlenhoff U (1999) Lys35 of PsaC is required for the efficient photoreduction of flavodoxin by photosystem I from *Chlamydomonas reinhardtii*. Eur J Biochem. 263: 137–144

Mix LJ, Haig D and Cavanaugh CM (2005) Phylogenetic analyses of the core antenna domain: investigating the origin of photosystem I. J Mol Evol 60: 153–163

Miyamoto R, Mino H, Kondo T, Itoh S and Oh-Oka H (2008) An electron spin-polarized signal of the $P800^+A_1(Q)^-$ state in the homodimeric reaction center core complex of *Heliobacterium modesticaldum*. Biochemistry 47: 4386–4393

Müh F, Renger T and Zouni A (2008) Crystal structure of cyanobacterial photosystem II at 3.0 Å resolution: A closer look at the antenna system and the small membrane-intrinsic subunits. Plant Physiol Biochem 46: 238–264

Muhiuddin IP, Rigby SE, Evans MC, Amesz J and Heathcote P (1999) ENDOR and special TRIPLE resonance spectroscopy of photoaccumulated semiquinone electron acceptors in the reaction centers of green sulfur bacteria and heliobacteria. Biochemistry 38: 7159–7167

Mühlenhoff U, Zhao JD and Bryant DA (1996) Interaction between photosystem I and flavodoxin from the cyanobacterium *Synechococcus* sp. PCC 7002 as revealed by chemical cross-linking. Eur J Biochem. 235: 324–331

Mulkidjanian AY and Junge W (1997) On the origin of photosynthesis as inferred from sequence analysis - A primordial UV-protector as common ancestor of reaction centers and antenna proteins. Photosynth Res 51: 27–42

Mulkidjanian AY, Koonin EV, Makarova KS, Mekhedov SL, Sorokin A, Wolf YI, Dufresne A, Partensky F, Burd H, Kaznadzey D, Haselkorn R and Galperin MY (2006) The cyanobacterial genome core and the origin of photosynthesis. Proc Natl Acad Sci USA 103: 13126–13131

Müller MG, Niklas J, Lubitz W and Holzwarth AR (2003) Ultrafast transient absorption studies on photosystem I reaction centers from *Chlamydomonas reinhardtii*. 1. A new interpretation of the energy trapping and early electron transfer steps in photosystem I. Biophys J 85: 3899–3922

Nitschke W and Rutherford AW (1991) Photosynthetic reaction centres: variations on a common structural theme? Trends Biochem Sci 16: 241–245

Nitschke W, Feiler U, Lockau W and Hauska G (1987) The photosystem of the green sulfur bacterium *Chlorobium limicola* contains two early electron acceptors similar to photosystem I. FEBS Lett 218: 283–286

Nitschke W, Setif P, Liebl U, Feiler U and Rutherford AW (1990) Reaction center photochemistry of *Heliobacterium chlorum*. Biochemistry 29: 11079–11088

Nitschke W, Mattioli T and Rutherford AW (1996) The Fe-S type photosystems and the evolution of photosynthetic

reaction centers. In: Baltscheffsky H (ed) Origin and Evolution of Biological Energy Conservation, pp 177–203. VCH, New York

Noodleman L, Norman Jr JG, Osborne JH, Aizman A and Case DA (1985) Models for ferredoxins: Electronic structures of iron-sulfur clusters with one, two and four iron atoms. J Am Chem Soc 107: 3418–3426

Noodleman L, Peng CY, Case DA and Mouesca JM (1995) Orbital interactions, electron delocalization and spin coupling in iron-sulfur clusters. Coord Chem Rev 144: 199–244

Oh-oka H (2007) Type I reaction center of photosynthetic heliobacteria. Photochem Photobiol 83: 177–186

Olson JM (2001) Evolution of Photosynthesis (1970), re-examined thirty years later. Photosynth Res 68: 95–112

Olson JM (2006) Photosynthesis in the Archean era. Photosynth Res 88: 109–117

Olson JM and Blankenship RE (2004) Thinking about the evolution of photosynthesis. Photosynth Res 80: 373–386

Olson JM and Pierson BK (1987) Evolution of reaction centers in photosynthetic prokaryotes. Int Rev Cytol 108: 209–248

Parrett KG, Mehari T, Warren PG and Golbeck JH (1989) Purification and properties of the intact P700 and F_X-containing photosystem I core protein. Biochim Biophys Acta 973: 324–332

Permentier HP, Schmidt KA, Kobayashi M, Akiyama M, Hager-Braun C, Neerken S, Miller M and Amesz J (2000) Composition and optical properties of reaction centre core complexes from the green sulfur bacteria *Prosthecochloris aestuarii* and *Chlorobium tepidum*. Photosynth Res 64: 27–39

Pierson K (1994) Early life on earth. In: Bengtson S (ed) Nobel Symposium No. 84. Columbia Univ Press, New York

Ptushenko VV, Cherepanov DA, Krishtalik LI and Semenov AY (2008) Semi-continuum electrostatic calculations of redox potentials in photosystem I. Photosynth Res 97: 55–74

Que L, Jr., Bobrick MA, Ibers JA and Holm R (1974) Synthetic analogs of active sites of iron-sulfur proteins. VII. Ligand substitution reactions of the tetranuclear clusters $[Fe_4S_4(SR)_4]^{2-}$ and the structure of $[(CH_3)_4N]_2[Fe_4S_4(SC_6H_6)_4]$. J Am Chem Soc 96: 4168–4177

Raymond J and Blankenship RE (2006) Evolutionary relationships among Type I reaction centers. In: Golbeck J (ed) Photosystem I: The Light-Driven Plastocyanin:Ferredoxin $NADP^+$ Oxidoreductase, pp 413–437. Springer, Dordrecht

Raymond J, Zhaxybayeva O, Gogarten JP, Gerdes SY and Blankenship RE (2002) Whole-genome analysis of photosynthetic prokaryotes. Science 298: 1616–1620

Redding K and van der Est A (2006) The directionality of electron transport in photosystem I. In: Golbeck J (ed) Photosystem I: The Light-Induced Plastocyanin:Ferredoxin Oxidoreductase, pp 413–437. Springer, Dordrecht

Renger G and Holzwarth AR (2005) Primary electron transfer. In: Wydrzynski TJ and Satoh K (eds) Photosystem II: The Light-Driven Water: Plastoquinone Oxidoreductase, pp 139–175. Springer, Dordrecht

Rigby SE, Thapar R, Evans MC and Heathcote P (1994) The electronic structure of $P840^+$. The primary donor of the *Chlorobium limicola* f. sp. thiosulphatophilum photosynthetic reaction centre. FEBS Lett 350: 24–28

Rippka R, Waterbury J and Cohen-Bazire G (1974) A cyanobacterium which lacks thylakoids. Arch Microbiol 100: 419–436

Romberger S, Castro C, Sun Y and Golbeck J (2010) Identification and characterization of PshBII, a second F_A/F_B-containing polypeptide in the photosynthetic reaction center of *Heliobacterium modesticaldum*. Photosynth Res 104: 293–303

Sattley WM, Madigan MT, Swingley WD, Cheung PC, Clocksin KM, Conrad AL, Dejesa LC, Honchak BM, Jung DO, Karbach LE, Kurdoglu A, Lahiri S, Mastrian SD, Page LE, Taylor HL, Wang ZT, Raymond J, Chen M, Blankenship RE and Touchman JW (2008) The genome of *Heliobacterium modesticaldum*, a phototrophic representative of the Firmicutes containing the simplest photosynthetic apparatus. J Bacteriol 190: 4687–4696

Savitsky A, Gopta O, Mamedov M, Golbeck J, Tikhonov A, Möbius K and Semenov AY (2009) Alteration of the axial Met ligand to electron acceptor A_0 in photosystem I: Effect on the generation of $P700^+A_1^-$ radical pairs as studied by W-band transient EPR. App Mag Res 37: 85–102

Schmidt KA, Neerken S, Permentier HP, Hager-Braun C and Amesz J (2000) Electron transfer in reaction center core complexes from the green sulfur bacteria *Prosthecochloris aestuarii* and *Chlorobium tepidum*. Biochemistry 39: 7212–7220

Schubert WD, Klukas O, Saenger W, Witt HT, Fromme P and Krauß N (1998) A common ancestor for oxygenic and anoxygenic photosynthetic systems: A comparison based on the structural model of photosystem I. J Mol Biol 280: 297–314

Sétif P (2001) Ferredoxin and flavodoxin reduction by photosystem I. Biochim Biophys Acta 1507: 161–179

Sétif P, Fischer N, Lagoutte B, Bottin H and Rochaix JD (2002) The ferredoxin docking site of photosystem I. Biochim Biophys Acta 1555: 204–209

Srinivasan N and Golbeck J (2009) Protein-cofactor interactions in bioenergetic systems. Biochim Biophys Acta 1787: 1057–1088

Srinivasan N, Karyagina I, Bittl R, van der Est A and Golbeck JH (2009) Role of the hydrogen bond from Leu722 to the A_{1A} phylloquinone in photosystem I. Biochemistry 48: 3315–3324

Sticht H and Rösch P (1998) The structure of iron-sulfur proteins. Prog Biophys Mol Biol 70: 95–136

Summons RE, Jahnke LL, Hope JM and Logan GA (1999) 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. Nature 400: 554–557

Tian F, Toon OB, Pavlov AA and De Sterck H (2005) A hydrogen-rich early Earth atmosphere. Science 308: 1014–1017

Trost JT and Blankenship RE (1989) Isolation of a photoactive photosynthetic reaction center-core antenna complex from *Heliobacillus mobilis*. Biochemistry 28: 9898–9904

Tsukatani Y, Miyamoto R, Itoh S and Oh-Oka H (2004) Function of a PscD subunit in a homodimeric reaction center complex of the photosynthetic green sulfur bacterium *Chlorobium tepidum* studied by insertional gene inactivation. Regulation of energy transfer and ferredoxin-mediated $NADP^+$ reduction on the cytoplasmic side. J Biol Chem 279: 51122–51130

van de Meent EJ, Kobayashi M, Erkelens C, van Veelen PA, Amesz J and Watanabe T (1991) Identification of 8^1-hydroxychlorophyll *a* as a functional reaction center pigment in heliobacteria. Biochim Biophys Acta 1058: 356–362

van der Est A (2006) Electron transfer involving phylloquinone in photosystem I. In: Golbeck J (ed) Phototosystem I: The Light-Induced Plastocyanin:Ferredoxin Oxidoreductase, pp 387–411. Springer, Dordrecht

van der Est A, Chirico S, Karyagina I, Cohen R, Shen G and Golbeck J (2009) Alteration of the axial Met ligand to electron acceptor A_0 in Photosystem I: An investigation of low temperature electron transfer by multifrequency time -resolved and CW EPR. App Mag Res 37: 103–121

Varotto C, Pesaresi P, Meurer J, Oelmuller R, Steiner-Lange S, Salamini F and Leister D (2000) Disruption of the Arabidopsis photosystem I gene *psaE1* affects photosynthesis and impairs growth. Plant J 22: 115–124

Vassiliev IR, Jung YS, Smart LB, Schulz R, Mcintosh L and Golbeck JH (1995) A mixed-ligand iron-sulfur cluster ($C556S_{PsaB}$ or $C565S_{PsaB}$) in the F_X-binding site leads to a decreased quantum efficiency of electron transfer in photosystem I. Biophys J 69: 1544–1553

Vassiliev IR, Jung YS, Mamedov MD, Semenov AY and Golbeck JH (1997) Near-IR absorbance changes and electrogenic reactions in the microsecond-to-second time domain in photosystem I. Biophys J 72: 301–315

Vassiliev IR, Jung YS, Yang F and Golbeck JH (1998) PsaC subunit of photosystem I is oriented with iron-sulfur cluster F_B as the immediate electron donor to ferredoxin and flavodoxin. Biophys J 74: 2029–2035

Vassiliev IR, Ronan MT, Hauska G and Golbeck JH (2000) The bound electron acceptors in green sulfur bacteria: resolution of the g-tensor for the F_X iron-sulfur cluster in *Chlorobium tepidum*. Biophys J 78: 3160–3169

Vassiliev IR, Antonkine ML and Golbeck JH (2001) Iron-sulfur clusters in Type I reaction centers. Biochim Biophys Acta 1507: 139–160

Vassilieva EV, Antonkine ML, Zybailov BL, Yang F, Jakobs CU, Golbeck JH and Bryant DA (2001) Electron transfer may occur in the chlorosome envelope: the CsmI and CsmJ proteins of chlorosomes are 2Fe-2S ferredoxins. Biochemistry 40: 464–473

Vermaas WFJ (1994) Evolution of heliobacteria: implications for photosynthetic reaction center complexes. Photosynth Res 41: 285–294

Vermaas WFJ (2002) Evolution of photosynthesis. In: Daly AO (ed) Encyclopedia of Life Sciences, pp 1–18. Nature Publishing Group, London

Vos MH and van Gorkom HJ (1990) Thermodynamical and structural information on photosynthetic systems obtained from electroluminescence kinetics. Biophys J 58: 1547–1555

Wächterhäuser G (1992) Groundworks for an evolutionary biochemistry: the iron-sulfur world. Prog Biophys Mol Biol 58: 85–201

Webber AN and Lubitz W (2001) P700: The primary electron donor of photosystem I. Biochim Biophys Acta 1507: 61–79

Xiong J, Fischer WM, Inoue K, Nakahara M and Bauer CE (2000) Molecular evidence for the early evolution of photosynthesis. Science 289: 1724–1730

Xu W, Chitnis P, Valieva A, van der Est A, Pushkar YN, Krzystyniak M, Teutloff C, Zech SG, Bittl R, Stehlik D, Zybailov B, Shen G and Golbeck JH (2003a) Electron transfer in cyanobacterial photosystem I: I. Physiological and spectroscopic characterization of site-directed mutants in a putative electron transfer pathway from A_0 through A_1 to F_X. J Biol Chem 278: 27864–27875

Xu W, Chitnis PR, Valieva A, van der Est A, Brettel K, Guergova-Kuras M, Pushkar YN, Zech SG, Stehlik D, Shen G, Zybailov B and Golbeck JH (2003b) Electron transfer in cyanobacterial photosystem I: II. Determination of forward electron transfer rates of site-directed mutants in a putative electron transfer pathway from A_0 through A_1 to F_X. J Biol Chem 278: 27876–27887

Yu L, Zhao JD, Mühlenhoff U, Bryant DA and Golbeck JH (1993) PsaE is required for *in vivo* cyclic electron flow around photosystem I in the cyanobacterium *Synechococcus* sp. PCC 7002. Plant Physiol 103: 171–180

Yu J, Smart LB, Jung YS, Golbeck J and McIntosh L (1995) Absence of PsaC subunit allows assembly of Photosystem I core but prevents the binding of PsaD and PsaE in *Synechocystis* sp. PCC 6803. Plant Mol. Biol. 29: 331–42

Zahnle K, Arndt N, Cockell C, Halliday A, Nisbet E, Selsis F and Sleep NH (2007) Emergence of a habitable planet. Space Science Reviews 129: 35–78

Zhao J, Li N, Warren PV, Golbeck JH and Bryant DA (1992) Site-directed conversion of a cysteine to aspartate leads to the assembly of a [3Fe-4S] cluster in PsaC of photosystem I. The photoreduction of F_A is independent of F_B. Biochemistry 31: 139–160

Zhao J, Snyder WB, Muhlenhoff U, Rhiel E, Warren PV, Golbeck JH and Bryant DA (1993) Cloning and characterization of the *psaE* gene of the cyanobacterium *Synechococcus* sp. PCC 7002 - Characterization of a *psaE* mutant and overproduction of the protein in *Escherichia coli*. Mol Microbiol 9: 183–194

Zouni A, Witt HT, Kern J, Fromme P, Krauß N, Saenger W and Orth P (2001) Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8 Å resolution. Nature 409: 739–743

Chapter 13

Patterns of Conservation and Divergence of the Photosystem II Complex

Aparna Nagarajan and Robert L. Burnap*
Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK 74078, USA

*Author for correspondence, e-mail: rob.burnap@okstate.edu

R.L. Burnap and W.F.J. Vermaas (eds.), Functional Genomics and Evolution of Photosynthetic Systems, Advances in Photosynthesis and Respiration 33, pp. 317–344, DOI 10.1007/978-94-007-1533-2_13,

Summary

The photosystem II (PSII) complex catalyzes the crucial and highly labile photosynthetic water-splitting reaction that uses solar energy to reduce plastoquinone using electrons derived from water. This chapter presents a brief summary of the structure, function, and damage repair cycle of the PSII complex. A bioinformatic analysis of the PSII proteins, utilizing the amino acid sequence conservation patterns projected onto the three dimensional structure, is presented. The analysis illustrates structural domains of the PSII complex where amino acid substitutions appear to have been subjected to selective constraint and areas that have had freedom to mutate. In addition to the expected conservation at known functional sites within the proteins, the analysis shows conserved surface features that likely form the interfaces with other proteins, such as antenna complexes, that are not present in the x-ray crystal structure. This chapter also discusses how patterns of structural divergence can be attributed to functional variations that have been selected on the basis of different environmental regimes, such as high versus low light intensity conditions. Finally, a model of extrinsic protein evolution involving the duplication and divergence of assembly factor proteins is proposed to account for the phyletic variations in the protein composition surrounding the manganese cluster of the H_2O-oxidation domain of PSII. In this model, PsbP is proposed to spatially replace PsbV, whereas PsbQ is proposed to bind to the outward-facing surface of the e-loop of CP43.

I. Molecular Structure and Function of the Photosystem II (PS II) Complex

A. Photosystem II: The H_2O-Plastoquinol Photo-Oxidoreductase

The light-driven extraction of hydrogen atoms from substrate water molecules is the signature catalytic feature of oxygenic photosynthesis and is catalyzed by the water oxidation complex (WOC) of photosystem II (PSII). A number of excellent reviews of the structure and function of PSII have been published over the last decade (Rappaport and Diner, 2002; McEvoy and Brudvig, 2006; Nelson and Yocum, 2006; Kern et al., 2007; Kern and Renger, 2007; Renger and Renger, 2008). Electrons derived from the endergonic splitting of H_2O are transferred to the electron acceptor plastoquinone (PQ), thereby forming plastoquinol (PQH_2). In turn, PQH_2 diffuses from its substrate-binding site within PSII into the membrane bilayer, thereby mediating the transfer of electrons to the remainder of the electron transport chain. These electrons are transferred to the remainder of the electron transport chain and used to form NADPH, which can be used for biosynthesis. PSII thus occupies the basal position of the oxygenic photosynthetic mechanism as it supplies reductant for biosynthetic metabolism using water as the source of electrons.

B. Water Oxidation

The H_2O-oxidation reaction is a thermodynamically difficult process since the extraction of tightly bound electrons from the highly electronegative oxygen atom of the water molecule requires a very powerful oxidant. This situation is further complicated by the fact that the full oxidation of water with the concomitant liberation of molecular oxygen is a four-electron, four-proton process, whereas oxidizing power

Abbreviations: bRC2 – type II anoxygenic bacterial reaction center; Chl – chlorophyll; ^{cy}PSII – cyanobacterial PSII; ^{dt}PSII – diatomic PSII; ^{gr}PSII – green algal and plant PSII; rdPSII – red algal (rhodophytic) PSII; HMM – hidden Markov model; MSA – multiple sequence alignment; Pheo – pheophytin; PQ – plastoquinone; PQH_2 – plastoquinol; PsbA – PSII reaction center protein, D1; PsbB – PSII proximal antenna protein, CP47; PsbC – PSII proximal antenna protein, CP43; PsbD – PSII reaction center protein, D2; PsbE/PsbF – cytochrome b_{559} of PSII; PsbO – extrinsic manganese-stabilizing protein of PSII, also known as the 33-kDa protein; PsbP – extrinsic 23-kDa protein of higher plant and green algal PSII; PsbQ – extrinsic 17-kDa protein of higher plant and green algal PSII; PsbV – cytochrome c_{550} of cyanobacterial and some algal PSII; PSI – photosystem I; PSII – photosystem II; Q_A – non-exchangeable plastoquinone acting as an early acceptor of electrons; Q_B – exchangeable plastoquinone; RC – reaction center; TMH – transmembrane helix; WOC – water oxidation complex; Y_z – redox-active tyrosine 161 of the D1 protein

produced at the photochemical reaction center is essentially univalent. Therefore, the WOC must not only catalyze an energetically expensive removal of electrons from water, but also needs to do this four times during a single catalytic cycle. To accomplish this feat, the photochemical reaction center (RC) successively turns over four times, each time generating a powerful oxidant, which is used to accumulate the four necessary oxidizing equivalents within the WOC (Kok et al., 1970). A metal cluster consisting of four Mn and one Ca (Mn_4-Ca) is involved in this accumulation of oxidizing equivalents. Each successive turnover of the RC involves a charge separation that results in the reduction of the non-exchangeable plastoquinone, Q_A, on the acceptor side of the photochemical reaction center and the oxidation of P_{680}, a chlorophyll species (see e.g., Holzwarth et al., 2006) on the donor side of the PSII complex, thus forming the charge-separated state: $P_{680}^{+}Q_A^{-}$. This process involves the transient formation of the primary reductant, pheophytin ($Pheo^{-}$), which reduces Q_A in less than 500 picoseconds. It is P_{680}^{+} that provides the oxidizing power to remove tightly bound electrons from substrate water. However, the oxidation of water is catalyzed by the Mn_4-Ca within the WOC. The oxidation of the WOC by P_{680}^{+} is mediated by the redox active tyrosine 161 (Y_z) of the reaction center D1 protein (Barry and Babcock, 1987; Debus et al., 1988; Noren and Barry, 1992). The oxidation of Y_z by P_{680}^{+} appears to be coupled to the movement of its phenolic proton to its H-bonding partner, D1-His190 (Diner et al., 2004), resulting in the neutral radical $Y_z^{\cdot}$, thereby facilitating the oxidation of Y_z yet retaining the positive charge in that region of the WOC (Haumann et al., 1997; Rappaport and Lavergne, 1997; Christen and Renger, 1999). The sequential oxidation of the H_2O-splitting enzyme causes it to cycle through a series of intermediate redox states (Joliot et al., 1969; Kok et al., 1970; Forbush et al., 1971) termed S-states (S_i, where i = 0–4), corresponding, at least in part, to the step-wise oxidation of the Mn_4-Ca. The mechanism of H_2O-oxidation remains to be resolved with a variety of models involving proton-coupled electron transfer (PCET) under consideration (McEvoy and Brudvig, 2004; Haumann et al., 2005; Dau and Haumann, 2008; Rappaport and Diner, 2008; Sproviero et al., 2008).

C. Photosystem II Structure

1. Subunit Composition of PSII

Detailed biochemical analyses of the PSII complex have shown that it contains a common core of intrinsic proteins found in all PSII-containing organisms examined thus far: the homologous reaction center D1 (PsbA) and D2 (PsbD) proteins forming a heterodimer coordinating the cofactors involved in primary charge-separation and stabilization; the homologous CP43 (PsbC) and CP47 (PsbB) proteins, which coordinate the majority of chlorophyll (Chl) in the PSII complex and serve as proximal antennae for the D1/D2 reaction center; and a number of small intrinsic subunits. Among the small common intrinsic proteins is a *b*-type cytochrome, cyt b_{559} (PsbE/F), which appears to have a protective role returning electrons from the acceptor side to the donor side of the reaction center, thereby avoiding the accumulation of potentially dangerous oxidizing species such as P_{680}^{+} (Thompson and Brudvig, 1988; Magnuson et al., 1999; Hung et al., 2010). Besides the common core of intrinsic subunits, one extrinsic protein, PsbO (the 33 kDa manganese-stabilizing protein) is found in all phyletic versions of PSII. This protein is associated with the WOC of the PSII complex and serves to stabilize the structure of the WOC and modulate the accessibility of the Mn_4-Ca to the aqueous phase. In addition to PsbO, at least two other extrinsic proteins are associated with the WOC. These additional extrinsic subunits are different among the different phyla, and these differences may be explained in evolutionary terms as discussed later.

2. Three-Dimensional Structure of PSII

Considerable progress has been made towards resolving the three-dimensional (3D) atomic structure of the PSII complex (Fig. 13.1). Structure models, currently all of thermophilic cyanobacterial PSII, have been published from several research groups (Zouni et al., 2001; Kamiya and Shen, 2003; Ferreira et al., 2004; Guskov et al., 2009) (Protein Data Bank entries 1FE1, 1IZL, 1S5L, 3BZ1). The most recent structure model, that of *Thermosynechococcus elongatus*, diffracts at up to 1.9 Å resolution and has assignments for virtually all amino acids and cofactors (Umena et al., 2011). While some of the assignments of

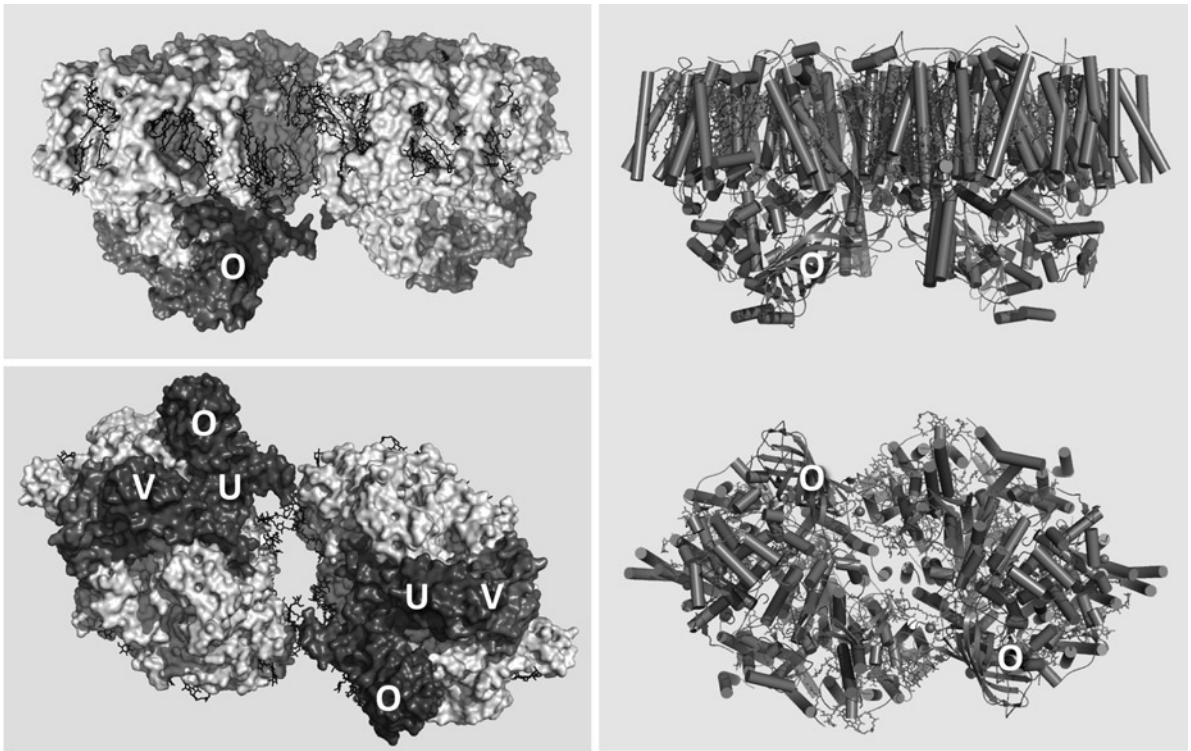


Fig. 13.1. The overall structure of PSII dimer rendered in space-filling (*left panels*) and cartoon representations (*right panels*). The *top panels* depict side views looking parallel to the plane of the membrane bilayer that the PSII complex is embedded in with the bottom oriented towards the lumen and the top oriented towards the cytoplasm. The *lower panels* depict views perpendicular to the plane of the membrane looking at the portions of the PSII complex that are comprised of the extrinsic proteins overlying the WOC. The extrinsic subunits are rendered in *dark gray* and the intrinsic subunits are rendered in *light gray* in the space-filling representations of the PSII dimer (*left panels*). Note that in the space-filling illustration, the two monomeric halves of the dimer have been separated by approximately 15 Å to highlight the monomer boundaries and to show the manner in which the extrinsic PsbO subunits interact with the other half of the dimer resulting in the interlocking of the two monomers. The figures were prepared using the Protein Data Bank structural coordinates 3BZ1 and 3BZ2 as manipulated in the PyMOL Molecular Graphics System (Schrödinger, LLC).

the structural model remain contentious especially regarding the arrangement of the Mn_4-Ca (cf. Chu et al., 2004; Mizusawa et al., 2004), many of the assignments are consistent with conclusions drawn from earlier chemical cross-linking and site-directed mutagenesis studies (Odom and Bricker, 1992; Frankel and Bricker, 1995; Miura et al., 1997; Enami et al., 1998; Frankel et al., 1999). Overall, the crystallized PSII complex is dimeric with each monomer containing a single copy of the intrinsic (e.g., D1) and extrinsic (e.g., PsbO) components and with the D1/D2 heterodimer flanked by the CP43 and CP47 proteins. As discussed below, the D1/D2 heterodimer retains striking structural similarity to anoxygenic bacterial reaction centers, which are comprised of the homologous L and M subunits that are similarly arranged as a heterodimer coordinating the pigments and cofactors involved in primary charge separation and stabilization.

While the thermophilic cyanobacterial reaction center is currently the most highly resolved PSII structure, other lower resolution techniques besides x-ray crystallography have provided important structural information on PSII, especially for higher plants and algae where these other techniques have yielded the best 3D structural information now available. These alternative approaches include the use of electron microscopic or electron diffraction analysis of PSII preparations with the former relying on image averaging techniques (Boekema et al., 1998; Hankamer et al., 1999; Kuhl et al., 2000; Nield et al., 2000). PSII complexes isolated from plants and green algae are larger and more complex, principally due to their association with

intrinsic membrane light-harvesting complexes containing Chl *a* and Chl *b*. This has allowed tentative assignment of the positions of the extrinsic subunits. Also solved using crystallographic or NMR methods are the structures of the poorly understood, low potential, *c*-type cytochrome, cyt c_{550} (PsbV), associated with the H_2O-oxidation complex at a resolution of 1.21 Å (Frazao et al., 2001) and of the D1 protein **c**arboxy **t**erminal **p**rotease (CtpA) with a resolution of 1.8 Å (Liao et al., 2000). While not a component of the PSII complex, CtpA is necessary for activity since it is dedicated to the post-translational processing of the D1 protein rendering it competent to assemble the active site Mn_4-Ca (Diner et al., 1988; Seibert et al., 1989; Nixon et al., 1992; Taguchi et al., 1995; Trost et al., 1997; Ivleva et al., 2000). Likewise, Psb27 (discussed below) is a lipoprotein that appears to be important for Mn_4-Ca assembly (Roose and Pakrasi, 2007) and it too has been structurally resolved (Cormann et al., 2009; Mabbitt et al., 2009).

II. Evolution of the PSII Reaction Center (RC)

A. Early Bioinformatic Insights on the Evolution of PSII

The photosynthetic reaction center probably evolved more than three billion years ago and a relatively comprehensive account of the evidence and thinking on the topic is given in this volume (see Jagannathan et al., Chapter 12). Here, we summarize some of this information and focus on the possible origins of PSII. The first far-reaching bioinformatic insight into the evolution of PSII complex came with the recognition of amino acid sequence similarities between the rhodobacterial (purple anoxygenic photosynthetic bacterial) reaction center polypeptides L and M, and the PSII proteins, D1 and D2 (PsbA and PsbD) (Williams et al., 1984; Youvan et al., 1984; Rutherford, 1986). While similarities in the cofactors of primary charge separation between RCs from *Rhodobacteria* and PSII had already been observed in the 1970s, the initial amino acid sequence comparisons between the anoxygenic bacterial and oxygenic RC polypeptides provided the first definitive clues to a common ancestry between these two kinds of RC (reviewed in Sauer, 1979). The technical advance at the time was the cloning and DNA sequencing of these genes, which allowed deduction of the amino acid sequences and this led to the performance of the first multiple sequence alignments (MSAs) of PSII and rhodobacterial reaction center amino acid sequences (Williams et al., 1984; Youvan et al., 1984). These alignments showed a low but detectable sequence similarity between the rhodobacterial and PSII reaction center subunit proteins. Furthermore, hydropathy analysis correctly predicted that each protein was likely to have five transmembrane helix (TMH) segments. About this time, the atomic structure of the rhodobacterial reaction center was determined, revealing that the L and M polypeptides are arranged as a pseudo-symmetric heterodimer with each subunit having five TMH segments that together coordinate the cofactors involved in primary charge separation (Deisenhofer et al., 1984).

With biochemical proof that D1 and D2 form the heterodimeric core of the PSII RC (Nanba and Satoh, 1987), it then became clear that the Rhodobacterial RC and PSII RC share a common core structure, and, by inference, a common evolutionary ancestry. Phylogenetic trees were constructed on the basis of the pairwise difference scores of sequences in the early sequence alignments and these did not reveal an obvious line of descent from either one of the rhodobacterial sequences into one or the other of the D1 or D2 sequences. In fact, the D1 and D2 polypeptides are more similar to each other than to either the L or M polypeptides of the bacterial RC. Reciprocally, the L and M polypeptides are more similar to each other than to either the D1 or D2 polypeptides of PSII. Therefore, it was concluded that two independent gene duplication events occurred with one giving rise to the L and M subunits and the other duplication event giving rise to the D1 and D2 subunits. Thus, it is possible that a similar selective driving force operated to break symmetry in the RC structure, perhaps associated with accompanying advantages in asymmetry in charge separation and quinone acceptor function: an example of the latter is the evolution of an exchangeable secondary electron acceptor to convey electrons to other bioenergetic complexes such as the cytochrome bc_1 complex (Blankenship, 1992; Schubert et al., 1998). It is worth noting that despite the small sample size (a combined total of 11 L, M, D1, and D2 amino acid sequences) used in those early analyses, the

original phylogenetic tree remains consistent with trees developed from much larger datasets and more recently developed computational tools for structural alignment and phylogenetic inference (cf. Sadekar et al., 2006). What is at the unresolved root of this phylogenetic tree? The discovery and analysis of Heliobacteria seem to provide an important clue (see Chapters 3 and 12 in this volume for related discussion). The most likely scenario now appears to be a homodimeric ancestral RC, the gene for which independently duplicated in several phyletic groups followed by the independent diversification of the duplicated genes to form different types of heterodimeric reaction centers such as the rhodobacterial, PSII and PSI RCs.

B. All Reaction Centers Share a Common Ancestor

Whereas from weak similarity between photosynthetic reaction center protein sequences, the evolution from a common ancestor was apparent and the location of functionally important residues in PSI had been correctly predicted already over 15 years ago (Vermaas, 1994), the depth of the similarity among photochemical reaction centers was more fully realized upon the structural determination of the PSI RC from the cyanobacterium, *Thermosynechococcus elongatus* (Schubert et al., 1997). Again, biophysical analysis had shown important similarities between the different types of RC regarding some of the cofactors mediating photochemistry, notably, the existence of a primary donor comprised of a 'special pair' of Chl. However, the acceptor system of PSI also contains several iron sulfur centers and these centers have no counterparts among the rhodobacterial and PSII RCs. Furthermore, biochemical and molecular genetic analysis of the PSI reaction center genes and polypeptides showed that the main PSI RC proteins are much larger than any of the bacterial and PSII RCs proteins: the PSI subunits, PsaA and PsaB, each have 11 transmembrane spanning segments, whereas the rhodobacterial and PSII reaction center subunits each contain 5 transmembrane segments. Therefore, until the PSI crystal structure was resolved, the possibility remained that PSI might have a fundamentally different structure. Instead, the crystal structure revealed strong structural similarity between PSI and the rhodobacterial RC: PsaA and PsaB subunits of the PSI complex, while much larger than the *rhodobacterial* RC core subunits, were shown to adopt a transmembrane structural fold of the last five transmembrane segments towards the carboxy-terminus that is highly similar to the rhodobacterial RC core. Each half of the heterodimeric PSI RC core is fused to a polypeptide that constitutes a large, Chl-binding, six-TMH domain that serves as the proximal light-harvesting antenna system and that is functionally and structurally homologous to CP47 and CP43. Thus, the known reaction center varieties adopt a common protein fold. Further, the antenna system domains of PSI and PSII are also remarkably similar, although in PSI the antenna domain is fused to the RC core polypeptides, whereas is PSII these domains are synthesized as separately encoded polypeptides that associate with the core RC after expression.

Despite the common core, a fundamental dichotomy is observed in the types of RC according to the composition of the primary electron acceptor. The rhodobacterial RC and PSII RC discussed above utilize a pair of quinone electron acceptors (Q_A and Q_B) and are classified as Type II reaction centers. Other anoxygenic phototrophs possessing Type II reaction centers include green non-sulfur bacteria such as *Chloroflexus* (for a review, see Bryant and Frigaard, 2006). On the other hand, Type I reaction centers are distinguished by the presence of iron-sulfur clusters that function as the initial acceptors, as mentioned above. Type I reaction centers include PSI and the RCs found in the anoxygenic phototrophs, the heliobacteria and green sulfur bacteria. The Type I reaction centers are comprised of either a dimer (anoxygenic bacteria) or a heterodimer (PSI of oxygenic phototrophs) of 11 TMH. Based upon analyses of sequence and structure alignments (see next section), it is likely that the ancestral Type I RCs were homodimeric, as represented in extant heliobacteria and green sulfur bacteria (Mix et al., 2005; Sadekar et al., 2006). The homodimeric ancestral Type I RCs are likely to have given rise to the PSI heterodimeric configuration of PSI in cyanobacteria, algae, and plants (see Chapter 12 for full discussion). The analyses are also consistent with the occurrence of three separate gene duplication events giving rise to the heterodimeric RCs: PSI, the Type II RC of rhodobacteria and green non-sulfur bacteria, and

the D1/D2 heterodimer of PSII. Furthermore, it is probable that the heterodimeric Type II RCs are comprised of two 5 TMH polypeptides (*i.e.* the Type II RC of rhodobacteria and green non-sulfur bacteria, and the D1/D2 polypeptides of PSII), evolved by gene-splitting of 11 TMH RCs, thereby yielding a 5 TMH RC protein 'un-fused' from the larger Chl-binding 6 TMH domain that serves as the proximal light-harvesting antenna system. It is not clear whether this gene-splitting (fission of antenna from RC domains) would have occurred prior or after the duplication/divergence events leading to heterodimeric RCs.

C. Structure Based Alignments of Reaction Centers

As discussed, the Type I and Type II RCs share a common structural core that has been retained through the approximately three billion years of evolution and diversification involving multiple gene duplication and gene fission events, all of which are obscure, but must have occurred in the distant past. Despite the clear 3D structural similarities among the different classes of RC, sequence similarity is very low, sometimes less than 10%, due to the extreme antiquity of the original divergence events that led to the different classes of RC (Schubert et al., 1998). This low level of sequence similarity thus makes phylogenetic inferences regarding the early stages of RC evolution problematic. However, is possible to use the 3D structural similarity quantitatively as the basis for building a phylogenetic tree, which can be used in cases where most sequence similarity is already lost (Johnson et al., 1990; Sadekar et al., 2006). The x-ray crystal structure coordinates of all the known reaction centers were aligned by minimizing the spatial deviations of the alpha-carbon backbone during an iterative computational superpositioning of each of the structures of the different RC proteins being aligned (Sadekar et al., 2006). Such a multiple structural alignment permits the calculation of the pair-wise root mean square distance (RMSD) values for any two proteins in the structural alignment. The corresponding set of pairwise RMSD values was used to calculate phylogenetic trees using a neighbor-joining algorithm. The tree derived from the structure similarities proved to have a topology identical to the tree derived from sequence similarities (Sadekar et al., 2006). Importantly, the amino acid sequence alignment obtained from the structural alignment was not identical to that obtained using sequence-based alignment methods underscoring the utility of the structural alignment approach to producing accurate alignments of homologous amino acid sequences as opposed to alignment methods based purely on sequence comparisons.[1] Overall, both the structure-based and the sequence-based phylogenetic analyses indicate that the ancestral reaction center was homodimeric and incapable of oxygen evolution.

D. Anoxygenic Type II RC Versus PSII RC: Evolution of H_2O Oxidation

Perhaps the most interesting, yet least understood, evolutionary aspect of the PSII complex is the acquisition of the Mn_4-Ca and the evolution of H_2O-oxidation activity. The evolutionary appearance of oxygenic photosynthesis is estimated to have occurred approximately 2.7 billion years ago (Ga). This evolutionary milestone eventually transformed the Earth's atmosphere from anoxic to oxic (for full discussion of the early evolution of

[1] In this regard it is worth noting that the original sequence alignment scoring matrices made by Margaret Dayhoff and colleagues for the purpose of calculating sequence optimal amino acid alignment were actually derived by making extensive use of the 3D protein structural information available at the time (Dayhoff et al., 1978). Although large protein structure databases and sophisticated computation tools for aligning protein structures were not yet available, these researchers were able to use data from soluble proteins in several protein families to produce 'trusted alignments' of amino acid sequences. In other words the multiple sequence alignments (MSAs) were validated by analysis of the 3D structures available. These trusted MSAs were then used to derive statistical information concerning frequencies of 'point accepted mutations' (PAMs)—that is the frequencies at which a given amino acid is observed to have been replaced by each of the other 19 amino acids. These accepted mutation frequencies are compiled into the Dayhoff scoring matrices still in use today, over 30 years later. What is different about the discussed approach and its related antecedents is that the structural alignment from a single protein family is being used to: (1) show remarkable structural conservation despite astonishingly low sequence similarity and (2) generate phylogenetic trees from a purely structural deviation analysis; these trees are topologically consistent with the best sequence-based phylogenetic trees.

photosynthesis, see Jagannathan et al., Chapter 12). Whereas the acceptor side of PSII appears to have retained the main ancestral features of the Q_A and Q_B binding sites in anoxygenic Type II RCs, the donor side has undergone drastic structural changes to increase the oxidizing potential of the primary donor chlorophyll, allow the introduction of the redox active tyrosine, Y_Z, and accommodate the binding and encapsulation of the Mn_4-Ca, (Allen and Williams, 2011).

A potentially robust bioinformatic approach to defining functionally significant amino acid residues in RCs employs a combination of structure-based MSAs and Hidden Markov Modeling (HMM). As discussed above, the 3D structure of reaction center proteins is maintained even when long periods of evolution have tended to scramble the primary amino acid sequence so that the level of sequence identity is very low (<10%). While certain key amino acids are strongly conserved in the related protein structures, the typical amino sequence alignment algorithms such as ClustalW often have difficulty identifying these key amino acids. The utilization of 3D structural information into the sequence alignment strategy has the potential to improve this situation. To implement this, 3D structural alignments were used to produce a trusted 'core MSA' for the amino acids of the RC proteins (Krammer et al., 2009a). Once constructed, the trusted core MSA was then used to produce a profile hidden Markov model (pHMM) containing sequences with high similarity to the sequences in the core MSA. The pHMM represents the conservation properties abstracted from the core MSA using HMM methods. This pHMM was then used as a scoring tool to align many more additional RC sequences allowing a more comprehensive, yet reliable, view of the amino acid sequence conservation in a large number of sequences in the RC protein family.

Using this profile HMM approach, a number of amino acid conservation features relating to function were identified and included commonalities between the bacterial and PSII RCs as well as some differences. Differences included the definition of a somewhat more polar environment around the Q_A binding site in PSII compared to the bacterial RC. The more polar Q_A binding pocket in PSII could, in principle, account for the observation that Q_A in PSII can become doubly reduced and protonated during photoinhibition (Vass et al., 1992) (see below for discussion of photoinhibition), whereas doubly reduced Q_A has never been observed in bacterial RCs. Furthermore, this approach has been extended to obtain a unique perspective on the proton pathway networks associated with the quinone acceptors at the Q_B binding site in photosynthetic reaction centers (Krammer et al., 2009b). However, this bioinformatic approach did not prove highly informative regarding the factors modulating the very positive redox potential of the special pair, P_{680} of PSII, which is poised at about +1,100 mV, compared to the homologous special pair of *rhodobacteria*, P865, which is poised at about +500 mV—a difference that cannot be explained by the nature of the chromophore (bacteriochlorophyll *a* versus chlorophyll *a*) alone. It also could not address the structural features determining what may be a fundamentally different charge separation mechanism that is initiated not at the special pair Chl, but rather at the adjacent Chl (Holzwarth et al., 2006). These differences are likely due to a complex combination of a variety different protein environment factors (see Allen and Williams, 2011 for discussion) and thus difficult to approach using current sequence-oriented bioinformatic methods. For example, if the local dielectric or other ensemble properties of the chromophore-protein complex are important for P_{680} function, these may be difficult to identify using sequence-oriented methods in the absence of more explicit mapping of the sequence conservation features to the 3D structure.

The evolutionary origin of the Mn_4-Ca responsible for H_2O-oxidation is especially enigmatic and presently there are no unambiguous clues to provide direction on this point. Given the thermodynamic difficulty and kinetic complexity of the H_2O-oxidation mechanism, it seems likely that evolutionary path to the Mn_4-Ca center involved intermediate forms of metal complexes functioning as secondary electron donors. These evolutionary intermediates of the catalyst may have taken the form of transition metal complexes of lower nuclearity. One possible scenario is that the ancestral proto-PSII obtained electrons to the primary donor from H_2O_2, which is thought to have been present, albeit probably not abundant, due to abiotic processes in the Archean Earth (Blankenship and Hartman, 1998). Another possibility, the partial, two electron oxidation of H_2O to H_2O_2, seems

unlikely because the midpoint potential for this is approximately +1,300 mV at pH 7 as compared to about +815 mV for the four electron oxidation of water to O_2 (Wood, 1988). Since the oxidizing potential of present day P_{680}^+ is approximately +1,100 mV, it is hard to imagine how a precursor to this would have the potential to achieve the required oxidation of H_2O to H_2O_2.

III. Structural Conservation in PSII: A Bioinformatic Perspective

Excellent reviews and bioinformatic analysis on PSII proteins are available, but these analyses are largely restricted to the extrinsic proteins (De Las and Barber, 2004; De Las et al., 2007; Enami et al., 2008; Williamson, 2008). In this section, an extension of this analysis is presented with special emphasis on the patterns of amino acid residue conservation making extensive use of the PSII 3D structure.

A. Projection of Amino Acid Sequence Conservation onto the 3D Structure of PSII

The construction of an MSA remains one of the most basic and useful methods for revealing variability and therefore inferring the functional importance of amino acid residues. Amino acid residues that are most highly conserved in a protein are generally involved in specific functions of the protein. These include residues involved in the formation and stabilization of the protein structure, the binding of ligands, involvement in protein-protein interactions, or participation in catalysis. An important application of the MSA is to map the patterns of residue conservation to the 3D structure of the protein. This allows for the identification of functional regions of the protein such as an active site or the site of important protein-protein interactions. This approach has been useful, for example, in defining receptor binding domains on the surface of proteins (Lichtarge et al., 1996b). Such patterns of conservation may not be readily apparent upon inspection of the MSA because the spatially adjacent residues in a conserved region of the 3D structure are often widely separated in the linear amino acid sequence. This contrasts with the usual illustration of molecular 3D structures, where amino acids are rendered using coloring schemes that depict the chemical properties of the residues. Visualizing the patterns of conservation in the 3D structure involves the projection of the conservation patterns by coloring residues according to the degree of conservation determined by MSA and phylogenetic methods. This has been done for PSII for many of the illustrations in this chapter: residues at highly conserved sites are colored blue to represent strong conservation, whereas sites that are more variable may be colored red as discussed more fully below. This approach has been combined with state-of-the-art phylogenetic tools to develop color-coding schemes that reflect underlying models for the evolution of a family of homologous proteins (Lichtarge et al., 1996a; Landau et al., 2005). This section describes the application of one such computational tool, Consurf, for the analysis of the PSII reaction center proteins presented in this chapter.

Consurf, which stands for 'conserved surface', is a flexible web-based application that contains a database of 3D structures that have been analyzed using evolutionary methods and presented using a molecular graphics structure viewer with the amino acid residues of the protein colored according to the degree of estimated evolutionary conservation (Landau et al., 2005). Importantly, the Consurf web-server allows the user to upload one or all of the key ingredients for producing this type of analysis: an MSA, a PDB file of the structure, and a phylogenetic tree. To implement the coloring according to residue conservation, the server algorithm substitutes the conservation scores for each of the amino acid residues into the PDB file by modifying the original PDB file. The output from the server includes an altered PDB file that contains the calculated conservation values in the PDB text file. One output option is to substitute the original b-values (reflecting the temperature factor in the crystal structure) with the conservation values. In this way, the residues can be colored according to conservation by coloring the protein molecule using the 'color by b-value' command found in most molecular graphics viewers. Thus, the residues will be colored according to conservation using the *faux* b-values. Another output option provides the conservation information as a header for the modified PDB file that can be interpreted by the molecular viewer Protein Explorer.

The current crystal structures of the PSII complex from the cyanobacterium *Thermosynechoccus elongatus* resolve as many as 20 protein subunits per monomer. Except for two small (<5 kDa) single TMH subunits around the periphery of the complex, all subunits now have amino acid assignments. Furthermore, over 70 cofactors including chlorophylls, quinones, carotenoids, and metal ions are resolved. Therefore, a relatively comprehensive analysis of the spatial distribution of amino acid conservation is possible for the PSII complex. Sequences were retrieved for each of the proteins defined in the crystal structure and MSAs were constructed. The automated sequence retrieval at the Consurf website uses the iterative Psi-Blast algorithm (Altschul et al., 1997) and was utilized for most of the PSII proteins. The use of the iterative Psi-Blast algorithm for sequence retrieval, as opposed to a standard Blast search (Altschul et al., 1997), ensures a diverse phyletic distribution so that the MSA of each protein used for the analysis contains plant, algal, and cyanobacterial homologs. The Consurf web server automatically aligned the retrieved sequences using the Muscle program (Edgar, 2004). During the construction of the MSAs used here, all taxonomic groups are represented for each of the common PSII proteins, such as the D1, D2, CP43, and CP47 proteins. For proteins only found in the cyanobacteria and red algae, for example PsbV, the MSAs were naturally restricted to sequences from those taxa. Since the crystallized cyanobacterial PSII complex does not contain major extrinsic proteins found in green plants and algae, for example, PsbP and PsbQ, the proteins are absent in the current analysis using the cyanobacterial PSII crystal structure. Future analysis using separate alignments for each of the major oxygenic taxa will provide a more discerning perspective that might reflect interesting taxon specific functional differences, for example binding sites for different light-harvesting antennae, that may exist among the PSII complexes from different taxa.

B. Overall Patterns of Structural Conservation of the PSII Complex

Surface exposed residues involved in functional interactions, such as the specific binding of a ligand or the interaction sites between subunits, generally have less freedom to accept mutations without deleterious consequences and are conserved during evolution. Exposed residues not involved in functional interactions can, on the other hand, accept mutations without deleterious effects and should exhibit more variability. These patterns of conservation and variability are readily evident in the PSII structure. To visualize the patterns of structural conservation in PSII, Consurf analysis was performed as described in the previous section. Figure 13.2 depicts the PSII dimer viewed from different angles relative to the plane of the thylakoid membrane (see Fig. 13.2 legend for details). The structure of PSII is depicted so that one half of the PSII dimer is rendered as a solid surface and colored according to the degree of amino acid conservation and the other half is rendered in sticks and colored to delineate the various subunits. The amino acid conservation is displayed in a color spectrum: red represents highly mutable residues, yellow and green indicate intermediate levels of conservation, whereas cyan and blue indicate the most highly conserved protein sites. The findings are, in general, as expected: the exterior protein surfaces exposed the aqueous phase of the thylakoid lumen exhibit a comparatively high degree of variability as indicated by the preponderance of red and yellow colored amino acid sites on the external surface. Specifically, the aqueous exposed surface of the extrinsic proteins and the aqueous exposed portions of the lumenal domains of CP43 and CP47 are rather poorly conserved as indicated by the many sites that are rendered in red, yellow and green (Fig. 13.2). Thus, the lumenal surface represented in the cyanobacterial crystal structure is relatively free to mutate and it is probably safe to conclude that these surfaces are not engaged in important protein-protein interactions. As shown below, this is in striking contrast to the internal sites, such as the contacts between subunits (Fig. 13.2 and later, in Fig. 13.7). On the other hand, the cytoplasmic face of the PSII complex is very conserved as indicated by the higher density of cyan and blue-colored residues (Fig. 13.2). The higher degree of conservation on the cytoplasmic face of PSII may reflect important *in vivo* interactions such as thylakoid stacking interactions in the grana region and, in the case of red algae and cyanobacteria, interactions between PSII and its the light-harvesting phycobilisome.

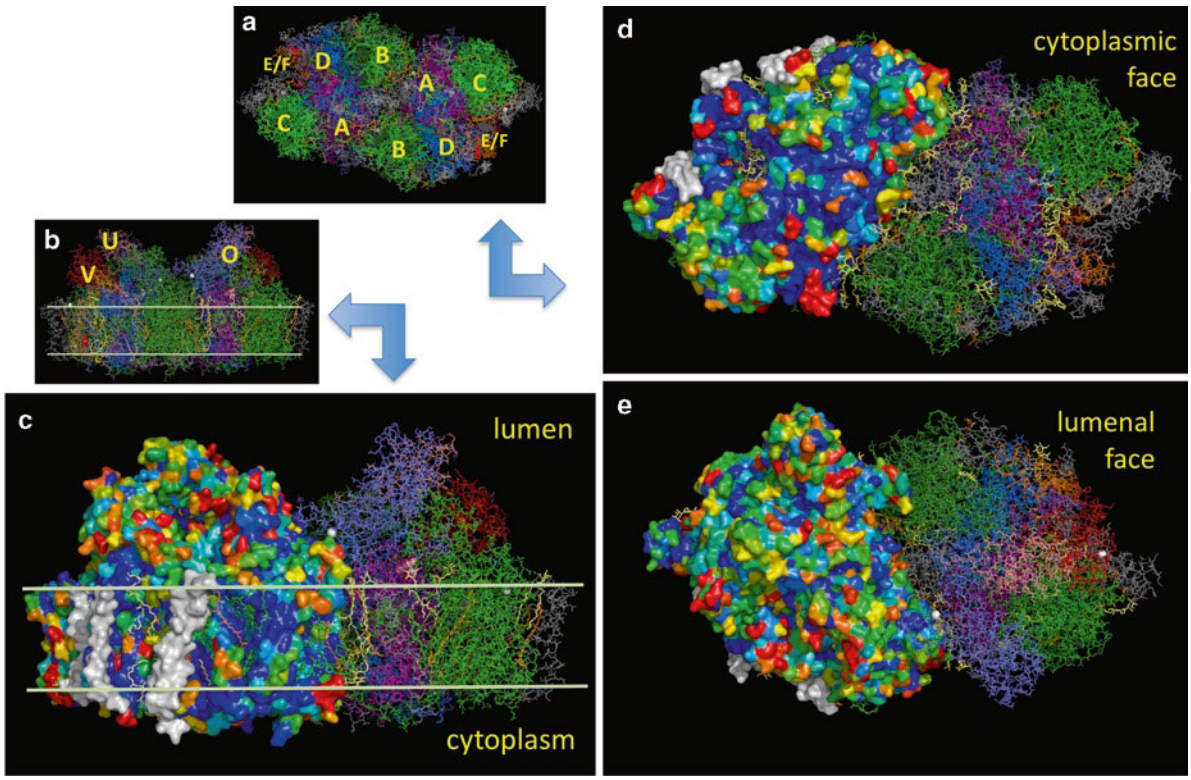


Fig. 13.2. Overall patterns of residue conservation in the PSII complex as projected onto the 3D crystal structure of *Thermosynechococcus elongatus*. *Upper right* shows the PSII dimer (3BZ1 and 3BZ2) viewed towards the cytoplasmic face and *lower right* shows the dimer viewed towards the lumenal face. In the *left panels* the PSII dimer is viewed from the side along the plane of the membrane, which is indicated by the *parallel gray lines*. The labels show the positions of some of the major subunits and with molecular components rendered in lines and colored according to subunit. Note that there are two coloring schemes here: Except for the two views in the *upper left* which are entirely rendered as lines, the right half of each dimer is rendered as lines and colored according to subunit: PsbA, *purple*; PsbB, *dark green*; PsbC, *light green*; PsbD, *blue*; PsbE/F, *orange*; PsbO, *slate blue*; PsbU, *tan*; PsbV, *red*. The *left half* of the dimer in each view is rendered as surface and colored in a spectrum according to the degree of conservation calculated from the MSAs of the PSII proteins: *red* represents highly mutable residues, *yellow* and *green* indicate intermediate levels of conservation, whereas *cyan* and *blue* indicate the most highly conserved protein sites. The aqueous exposed surface of the extrinsic proteins and the aqueous exposed portions of the lumenal domains of CP43 and CP47 are rather poorly conserved as indicated by the many sites that are rendered in *red*, *yellow* and *green*. Note that the *gray* regions correspond to the small single TMH subunits, which could not be assigned in the crystal structure. The figures and color assignments were prepared using multiple sequence alignments and conservation calculation in the Consurf suite of structural analysis tools (Landau et al., 2005) as discussed in the text. Final figures were prepared in the PyMOL Molecular Graphics System (Schrödinger, LLC).

C. Patterns of Conservation and Variation in the Proximity of Cofactors

The polypeptides of the PSII complex organize a large set of different cofactors including 36 Chl, 2 Pheo, 9 carotenoids, 3 plastoquinones, and an assortment of metal ions, most notably, the Mn_4-Ca of the H_2O-oxidation complex and the non-heme Fe situated between the quinone acceptors. Detailed accounts of these cofactors in the cyanobacterial complex are given in a number of articles (Ferreira et al., 2004; Loll et al., 2005; Kargul et al., 2007; Kern et al., 2007; Kern and Renger, 2007; Krivanek et al., 2007). Patterns of amino acid conservation of the reaction center and the protein-cofactor interactions modulating charge separation have been analyzed using the HMM methods outlined above. The ligation of Chl typically involves an imidazole nitrogen of a histidine coordinated to the Mg ion of the Chl.

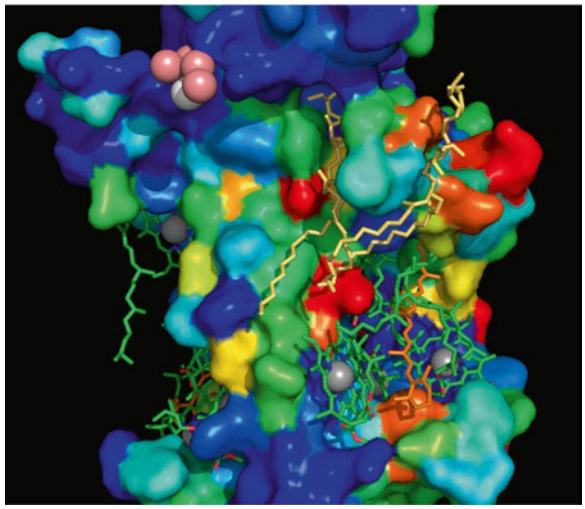


Fig. 13.3. Interior view of the patterns of conservation at the interface between CP43 and the D1/D2 heterodimer. Note the very strong conservation of residues (indicated by *blue* and *cyan* rendering) in the coordination environment of the Mn_4-Ca (Mn, *pale red spheres*; Ca, *white sphere*). The histidine residues coordinating the Mg atom (*gray spheres*) of chlorophyll (*green sticks*) are also generally conserved (*blue* rendering), whereas the regions surrounding these conserved ligands exhibit much less conservation (*red*, *orange*, *yellow*). Similarly, the surfaces surrounding the lipids (*pale yellow sticks*) and carotenoids (*orange sticks*) are not well conserved. This could either indicate that the presence or arrangement of these cofactors is less well-conserved among the various phyletic variants of PSII or that the structural determinants for binding the cofactors are not as stringent as other cofactors such as quinones and metals. The figures and the corresponding color assignments for each amino acid were prepared as in Fig. 13.2.

These histidines tend to be very strongly conserved (Fig. 13.3), as noted during the initial cloning of the corresponding genes for the chlorophyll-proteins. The majority of chlorophyll of the proximal antennae proteins, CP43 and CP47, are bound this way. Similarly, the reaction center, the special pair Chl, P_{D1} and P_{D2} are ligated by D1-His198 and D2-His197, respectively, which are strictly conserved (not shown here). On the other hand, Chl_{D1} and Chl_{D2} of the reaction center, which are the so-called 'accessory chlorophyll' of the reaction center, are probably coordinated via a water molecule to D1-Thr179 and the carbonyl backbone oxygen of D2-Gly174, respectively. These D1 and D2 residues thus act as second sphere ligands rather than direct ligands, and yet, are also strictly conserved (not shown here). Several residues adjacent to Chl in the CP43 and CP47 proteins also appear to act as second sphere ligands. In some of these cases, the nearby amino acid side chains are not conserved. This is consistent with the possibility that these nearby amino acids are acting as second sphere ligands with less stringent conservation (Fig. 13.3). It is also striking that considerable amino acid variability exists in the cofactor dense regions at the interface between subunits within the PSII interior. This variation includes sites around lipids and phytyl chains of chlorophyll (Fig. 13.3). The actual positions of the cofactors in PSII of other species are unknown; however, the present results anticipate that there may be significant variation in the binding of these cofactors and/or that the structural binding determinants are relatively relaxed for the retention of function—this variation seems to be especially apparent for the lipids and the phytyl chains of chlorophyll molecules. At the same time nearby areas critical for enzyme function, such as the coordination environment of the Mn_4-Ca and the quinone H-bonding sites, are very strongly conserved attesting to the low degree of tolerance for mutations in residues that evolution has selected as critical for function (Fig. 13.4).

D. Variations for Adaptive Specialization: High and Low Light Intensity Forms of D1

There are alternative forms of the D1 reaction center protein that are associated with high and low light adaptation. The redox midpoint potential of the Phe_{D1} is modulated by a hydrogen bond between the 9-keto group of the chlorine ring of Phe_{D1} and the amide nitrogen of glutamine (Gln) side chain at position 130 of D1 (PsbA). Experimental mutation of this residue to leucine (Leu) abolishes this H-bond and thereby destabilizes $Pheo^-$ making its midpoint potential more negative. Substitution with glutamate (Glu) makes the H-bond stronger and the midpoint potential more positive. In nature, both the Glu and Gln variants are found. The strength of H-bonding at the alternative side chains correlates with different charge recombination characteristics in the reaction center. In turn, reaction centers of the variants exhibit differing tendencies to generate reactive oxygen species, which results in different tendencies towards photoinhibition due to the formation of triplet Chl at the reaction center (Cser and Vass, 2007). The D1-Glu130 form is

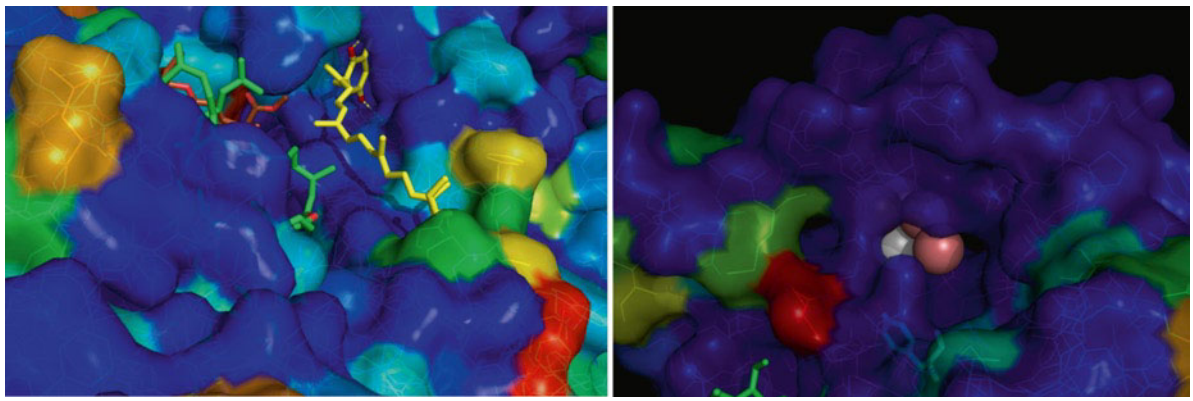


Fig. 13.4. Ligation environment of Q_A and the manganese cluster. *Left Panel*: The non-exchangeable plastoquinone, Q_A (*yellow sticks*), contains keto oxygens (*red atoms*) of the quinone ring structure, which H-bond to adjacent amino acid residues that are highly conserved. *Right Panel*: Very strong conservation of residues in the coordination environment of the Mn_4-Ca. The four Mn ions (*salmon color*) and one Ca ion (*white*) are embedded in a highly conserved ligation environment indicated by the surrounding amino acid surface rendered *blue*. The figures and the corresponding color assignments for each amino acid were prepared as in Fig. 13.2.

expressed at high light intensity in some cyanobacteria, whereas the D1-Gln130 form is expressed in low light (Kulkarni and Golden, 1994; Campbell et al., 1998). Higher plants and algae only possess the high light, D1-Glu130, form of the protein. These variations are thus not due to random variation and lack of selective constraint, but instead are due to positive selection that has led to adaptive specialization to specific environmental conditions. When colored according to conservation characteristics, the D1-130 site is seen as being quite variable (orange rendering), yet in contrast to the variation due to lack of selective constraint for the examples already discussed, this variation reflects functional diversity and lends a cautionary message in the interpretation of sequence/structure conservation results (Fig. 13.5).

IV. Extrinsic Subunits Associated with the Water Oxidation Complex

The large body of literature regarding the extrinsic proteins includes a number of up-to-date reviews that the reader is referred to (Enami et al., 2005, 2008; Roose et al., 2007b; Ifuku et al., 2008). These reviews include a description of the functional roles of the proteins and recent findings showing that there is a growing set of loosely or transiently associated PSII proteins that are only beginning to be understood. Here we highlight certain facets for the purpose of interpreting structural conservation features and presenting scenarios for the evolution of the WOC in different phyletic groups. In each phyletic version of PSII, there are extrinsic proteins that function to maintain the ion composition in the vicinity of the Mn_4-Ca and to stabilize the structure of the WOC. Evolutionary divergence of PSII has resulted in variations in the composition of the extrinsic proteins associated with the WOC. As noted above, the PSII complex exhibits an invariant core set of polypeptides found in all phylogenetic versions of the complex distributed among higher plants, algae, and cyanobacteria. This invariant core consists of the PsbA, PsbB and PsbC, PsbD, PsbE and PsbF, PsbO, and several small intrinsic polypeptides. In terms of the more traditional nomenclature, this corresponds to the D1-D2 heterodimeric reaction center, the proximal intrinsic antennae proteins CP43 and CP47, the enigmatic cyt b_{559}, and the extrinsic manganese stabilizing protein. Besides this invariant core, there are additional extrinsic subunits, the identity of which depends on the phylum. The most extensive structure-function analysis has been performed on the extrinsic proteins from higher plant PSII, although the function of the cyanobacterial and algal extrinsic proteins has

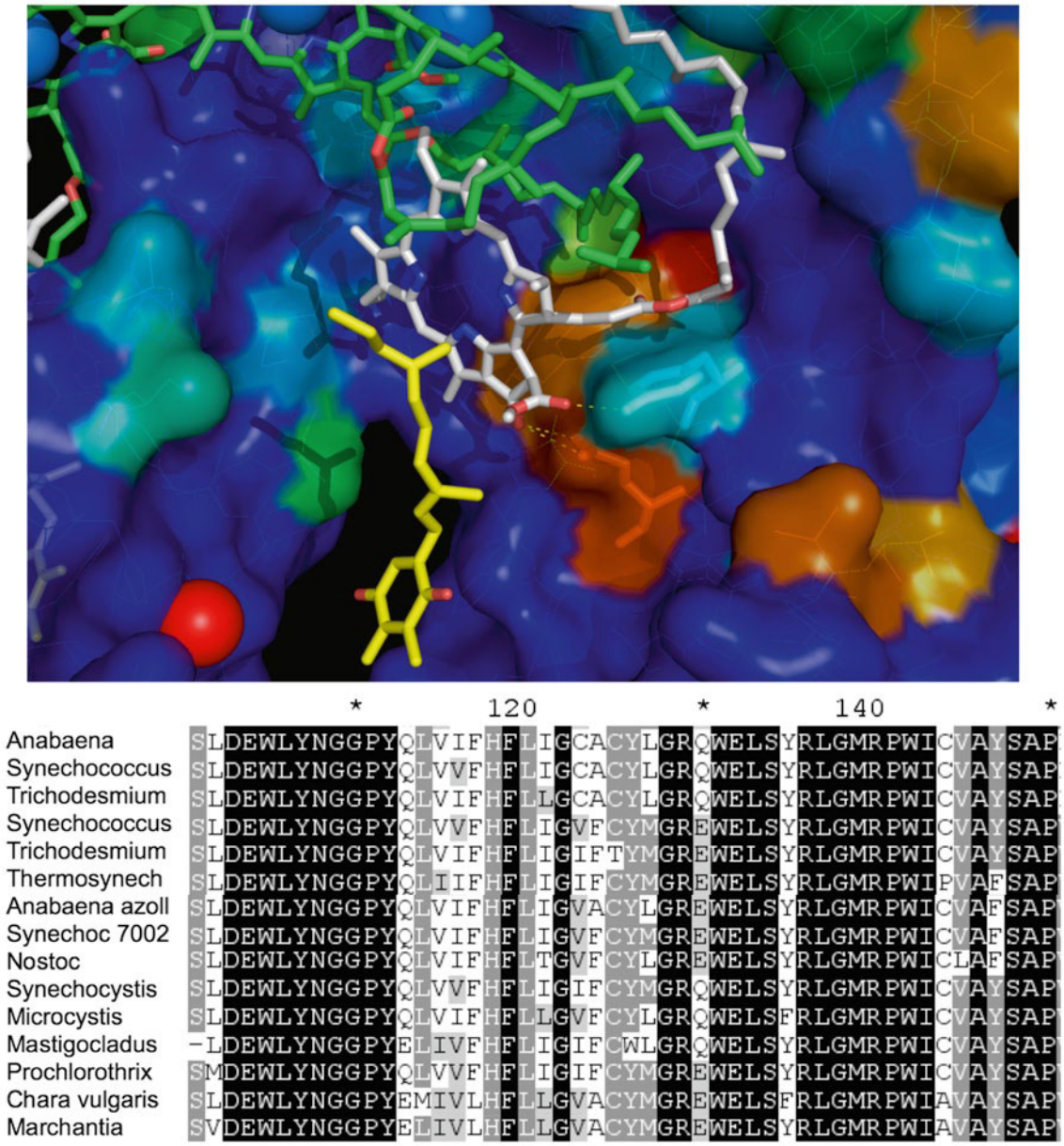


Fig. 13.5. Patterns of amino acid conservation of the D1 protein. *Top panel* shows the conservation of amino acids of D1 protein in the vicinity of the main electron pathway (A-branch). Note the residues involved in H-bonding (indicated as *yellow dashed lines*) the pheophytin (*gray sticks* rendering) are less strictly conserved. This 'systematic' variability (orange and red) is mainly associated with alternative forms of the D1 protein with either glutamine or glutamate at position 130 (D1-Q/E130). This variation is associated with different charge separation properties associated with growth at high and low light intensity (see text). *Bottom panel* shows the amino acid sequence alignment of the glutamine/glutamate 130 (D1-Q/E130) region. The figures and the corresponding color assignments for each amino acid were prepared as in Fig. 13.2.

also been well-studied. For the purpose of referring to the different phyletic types of PSII and its constituent proteins, we adopt a nomenclature where a superscripted appellation on the upper left designates whether it is cyanobacterial (cy), red algal (rd), diatom (dt), or green plants/green algae (gr). For example, green plant PSII is referred to as '^{gr}PSII'.

A. Extrinsic Subunit Variations in Phyletic Groups

1. Green Plants and Algae

The WOC of higher plants and green algae contains ^{gr}PsbP and ^{gr}PsbQ (a.k.a. 23 and 18 kDa proteins, respectively). Together with ^{gr}PsbO, these form the 'canonical' complement of extrinsic proteins. PsbO stabilizes the structure of the WOC, aids in the retention of Cl^-, and may participate in ejection of protons produced during the oxidation of water. ^{gr}PsbP and ^{gr}PsbQ help in the retention of Ca^{2+} and Cl^- at the Mn cluster. The ^{gr}PsbP and ^{gr}PsbQ proteins have long been known to reduce the demand for Ca^{2+} of the H_2O-oxidation reaction and minimize exchange of ions between the Mn active site and the external medium. The biochemical basis for the role of ^{gr}PsbP and ^{gr}PsbQ in the retention of Ca^{2+} has been elegantly demonstrated in experiments showing the rapid loss of Ca^{2+} from the Mn_4-Ca in the higher S-states, but only when these proteins were removed (Miqyass et al., 2008). This correlates with observation that light, which obviously can drive S-state cycling, promotes the loss of PSII Ca^{2+} in polypeptide-depleted samples (Boussac and Rutherford, 1988). While the precise structural basis for the S-state dependent lability of the Ca^{2+} is not understood, it is now clear that the extrinsic proteins serve to maintain high local concentrations of Ca^{2+} and Cl^- in the vicinity of the active site of H_2O-oxidation. An additional small subunit, PsbR, has also been identified and it appears to be unique to green plants and algae (Suorsa et al., 2006; Allahverdiyeva et al., 2007). PsbR appears to be important in assembly of the ^{gr}PsbP and ^{gr}PsbQ proteins. Whereas, the ^{gr}PsbO protein is assembled into the PSII complex in the stromal thylakoid lamellae, the ^{gr}PsbP and ^{gr}PsbQ proteins are incorporated into the plant PSII complex only after the migration of the complex into the grana regions of the thylakoids (Suorsa et al., 2006). In *Arabidopsis* mutants lacking PsbR, PSII activity is depressed and the ^{gr}PsbP and ^{gr}PsbQ are not assembled properly. PsbR is therefore proposed to facilitate the binding of ^{gr}PsbP protein to the complex at this later stage of PSII assembly. PsbR, which has a single TMH, appears also to be associated with another intrinsic single TMH protein, PsbJ. This is interesting since PsbJ, which is found in all phyletic versions of PSII, provides a contact site for the binding of ^{cy}PsbV seen in the cyanobacterial crystal structure. This might suggest that PsbP or PsbQ occupy a location in plants that corresponds to the location of PsbV in the cyanobacterial complex.

2. Cyanobacteria

The cyanobacterial PSII crystal structure contains neither PsbP nor PsbQ (Zouni et al., 2001; Kamiya and Shen, 2003; Ferreira et al., 2004; Guskov et al., 2009). Before the crystal structure was solved, it was already evident that highly purified cyanobacterial PSII has an extrinsic protein complement that consists of ^{cy}PsbO, the *c*-type cytochrome, ^{cy}PsbV (a.k.a. cyt. c_{550}) and PsbU. Yet no hint of PsbP and PsbQ in these highly purified preparations was observed. The PsbV and PsbU proteins serve a similar function to ^{gr}PsbP and ^{gr}PsbQ WOC proteins in that they reduce the demand for the Ca^{2+} and Cl^- cofactors of the H_2O-oxidation reaction. However, this picture of ^{cy}PSII containing only the ^{cy}PsbO, ^{cy}PsbV, and ^{cy}PsbU subunits is an oversimplification of the actual subunit composition of the in vivo ^{cy}PSII (Thornton et al., 2004; Kashino et al., 2006; Roose et al., 2007a) as demonstrated by the use of a genetically modified PSII containing a poly-histidine 'tag' engineered into CP47, which allows purification of highly active PSII from detergent solubilized cyanobacterial thylakoids that contains additional polypeptides (Bricker et al., 1998; Kashino et al., 2002). Proteomic analysis indicates that the cyanobacterial WOC is associated with additional polypeptides, including stoichiometric amounts of a PsbQ-like subunit (^{cy}PsbQ). When the ^{cy}PsbQ is reciprocally engineered to have a poly-histidine tag, highly active PSII complexes can be affinity purified using PsbQ alone, giving very strong evidence that this otherwise neglected cyanobacterial subunit is in fact a component of the native ^{cy}PSII (Roose et al., 2007a). It appears, therefore, that this subunit is loosely

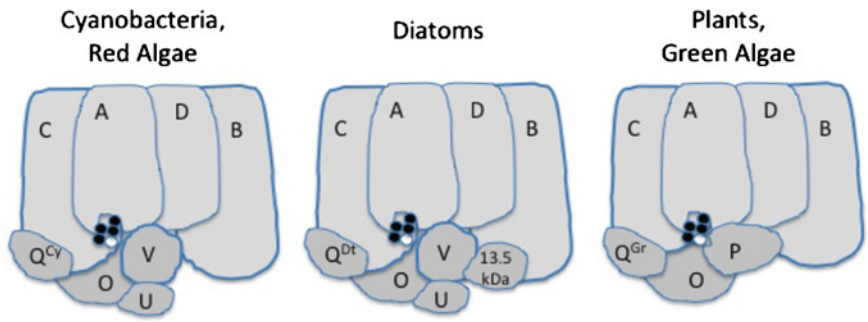


Fig. 13.6. Three major phyletic types of polypeptide configuration of the WOC in PSII.

bound and tends to be lost during biochemical isolation procedures, including those used to prepare solubilized complexes for crystallization and structural analysis. Unlike the ^{gr}PsbQ associated with ^{gr}PSII, ^{cy}PsbQ is likely a lipoprotein, anchored in the membrane by cysteine-linked lipid, although biochemical proof of the lipid moiety has yet to be established. Additionally, gently extracted cyanobacterial PSII preparations also contain sub-stoichiometric amounts of a PsbP homolog, ^{cy}PsbP, which is also predicted to be a cysteine-linked lipoprotein. ^{cy}PsbP is more closely related to PPL proteins than the PsbP associated with the WOC in ^{gr}PSII. This could mean that ^{cy}PsbP may also be involved in PSII assembly, rather than serving a structural role maintaining the ionic composition within the WOC as with the originally defined plant PsbP that is tightly associated with all ^{gr}PSII. Similarly, sub-stoichiometric amounts of another small protein, Psb27, is detected in gently purified PSII (Kashino et al., 2002) and in stoichiometric amounts of purified, partially assembled PSII monomers (Nowaczyk et al., 2006). As discussed below, Psb27 is one of several transiently associated PSII protein involved in PSII biogenesis.

3. Red Algae and Diatoms

Current results obtained from sequence analysis support the monophyletic origin of all modern chloroplasts with the ancestral chloroplast deriving from an endosymbiotic cyanobacterium residing in a vacuole of a eukaryotic host of uncertain taxonomic derivation (see Chapter 8 and references therein). Red algae (Rhodophytes) occupy a key taxonomic position in the deduced evolutionary history of oxygenic photosynthesis because they closely resemble the postulated symbiotic consortium between an endosymbiotic cyanobacterium and a eukaryotic host.[2] Like cyanobacteria, red algae use phycobilisomes as the major light harvesting antenna and, most relevant to the present discussion, the composition of the extrinsic proteins associated with the WOC is quite similar to that of cyanobacteria. In fact, given that cyanobacterial PSII contains PsbQ (Kashino et al., 2006; Roose et al., 2007a), there appears to be a one-to-one correspondence between the two in terms of the composition of the extrinsic proteins associated with the WOC (Fig. 13.6). Diatoms appear to have a variant composition: in addition to the common PsbO subunit, they contain PsbV, PsbQ, PsbU, and a novel 13.5 kDa soluble protein, termed Psb31. Also, the PsbQ variant in red algae and diatoms has many amino acid sequence differences compared to the higher plant and green algal forms. The nuclear genome of the recently sequenced red alga has a homolog to Psb31, although the PSII complexes purified from red algae do not appear to have this protein. Nevertheless, it is quite possible that, like the case

[2] A group of eukaryotic phototrophs, the *Glaucocystophytes*, retain even more primitive features of the original ensymbiotic consortium. One *Glycocystophyte*, *Cyanophora paradoxa*, has chloroplast-like inclusions, termed cyanelles, which closely resemble a free-living (Suorsa et al., 2006) coccoid cyanobacterium in terms of ultrastructure and biochemistry. For example, the cyanelles possess a bacterial-type peptidoglycan cell wall. However, the cyanelles are incapable of independent growth since their genome is the size and structure of a chloroplast genome rendering the cyanelle genetically dependent upon their unicellular host (Burnap and Trench, 1989a, b, c).

with highly resolved biochemical preparations of PSII in cyanobacteria, some subunits may be lost during purification of rdPSII complexes. Regardless of whether or not Psb31 is present, the red algal configuration of extrinsic subunits appears highly similar to cyanobacteria and thus it is likely that the red algal PSII most closely resembles the ancestral arrangement that existed in the early chloroplasts soon after the initial cyanobacterial endosymbiosis event, which led to the progenitor chloroplast, and before the divergence of eukaryotic species into the modern red, green, higher plant and other oxygenic eukaryotic genera. Thus, one of the more interesting recent findings is that homologs of PsbQ have been identified in all the major phyletic variants of PSII, albeit exceptions in certain species may exist (see Roose et al., 2007b). Recently two homologues of PsbQ have been observed in *Arabidopsis* and these might be involved in the assembly or stabilization of the chloroplast NAD(P)H dehydrogenase complex (Suorsa et al., 2006). The functions of these homologues are yet to be characterized. These homologues of PsbQ might bear some similarity to the PPL2 (PsbP like) proteins in *Arabidopsis* that are also involved in the NAD(P)H dehydrogenase complex (Ishihara et al., 2007).

A schematic representation of the extrinsic protein of the WOC in different phyletic groups is suggested in Fig. 13.6.

B. Conservation of Binding Sites of Extrinsic Subunits

As noted above, the exposed surfaces of the lumenal domain of PSII exhibit a high degree of sequence diversity as indicated by the paucity of amino acid residues rendered blue (in the lower right panel of Fig. 13.2, recalling that blue indicates conservation at the site). This indicates that the aqueous exposed lumenal region is dominated by amino acid sites that have tended to accept point mutations during the evolution of divergent lineages of oxygenic photosynthesis. However, the internal residues are much more conserved. The binding sites for the extrinsic proteins have been especially well conserved. As shown in Fig. 13.7, removal of the extrinsic proteins exposes a large conserved surface that is situated between the large protrusions of the e-loops of CP43 and CP47. This corresponds to contact regions between the intrinsic subunits (mainly CP43, CP47, D1, and D2) and the extrinsic subunits as indicated by the exposure of a large blue patch of surface on the

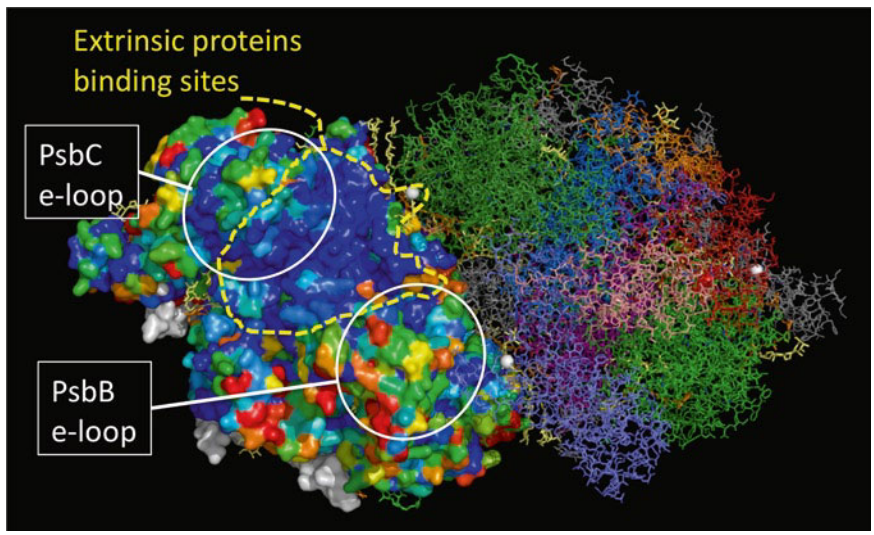


Fig. 13.7. Exposure of the conserved extrinsic protein-binding surface viewed from the side of the membrane. The extrinsic proteins (PsbO, PsbU, and PsbV) have been removed to expose the surface enclosed within the perimeter approximately delineated by the *dashed yellow line*. The extrinsic protein-binding surface is sandwiched between the protruding e-loops of CP43 and CP47, which are indicated by the white oval lines approximately enclosing these domains. Note that part of the PsbO protein extends to form contacts with the other monomer as indicated by the protrusion to the CP47 subunit (rendered as *fine green lines*). The figures and the corresponding color assignments for each amino acid were prepared as in Fig. 13.2.

intrinsic portion of the complex. Each of the extrinsic proteins is in contact with more than one of the other PSII proteins and these contact sites are largely conserved on both the contact surface of the extrinsic protein and its cognate binding surface on the intrinsic protein complex. Furthermore, PsbO forms a highly conserved contact with the opposite dimer: an ionic bond between aspartate 49 of the CP47 protein on one half of the dimer and lysine 85 of the PsbO protein bound to the other half of the dimer. Clearly, there has been strong selective pressure to maintain the amino acid interaction sites stabilizing the binding of the extrinsic proteins.

A particularly intriguing aspect of the conserved interfacial region is that sites interacting with the PsbV protein, a subunit not found in green plants and algae, are also conserved in the plant structures. Presently, there is only a crystal structure for the cyanobacterial PSII complex and information regarding the binding sites for the PsbP and PsbQ proteins in green algal and higher plant PSII is fragmentary. Specifically, the conserved contact sites on the CP43, D1 and D2 proteins, which mediate the interaction with the PsbV and PsbU proteins, are also conserved in higher plants and green algae. It is therefore likely that one or more of the alternative extrinsic proteins found in these taxa (*i.e.*, PsbP, PsbQ, or PsbR) bind to the corresponding locations since these are also not free to mutate in ^{gr}PSII. This could mean that PsbV was not simply lost from the PSII complex during the evolution of ^{gr}PSII, but was probably replaced at the same site by other protein(s), such as PsbP or PsbQ, as discussed at the end of this chapter.

There is an interesting apparent exception to the low degree of conservation on the lumenal surface of the PSII complex: there is a patch of conserved residues on the e-loop of CP43, but not in the symmetrically related region of CP47 (Fig. 13.8). Since cyanobacterial PSII is known to stoichiometrically bind a homolog of PsbQ that is lost during crystallization, this conserved patch on CP43 may possibly be the binding site for

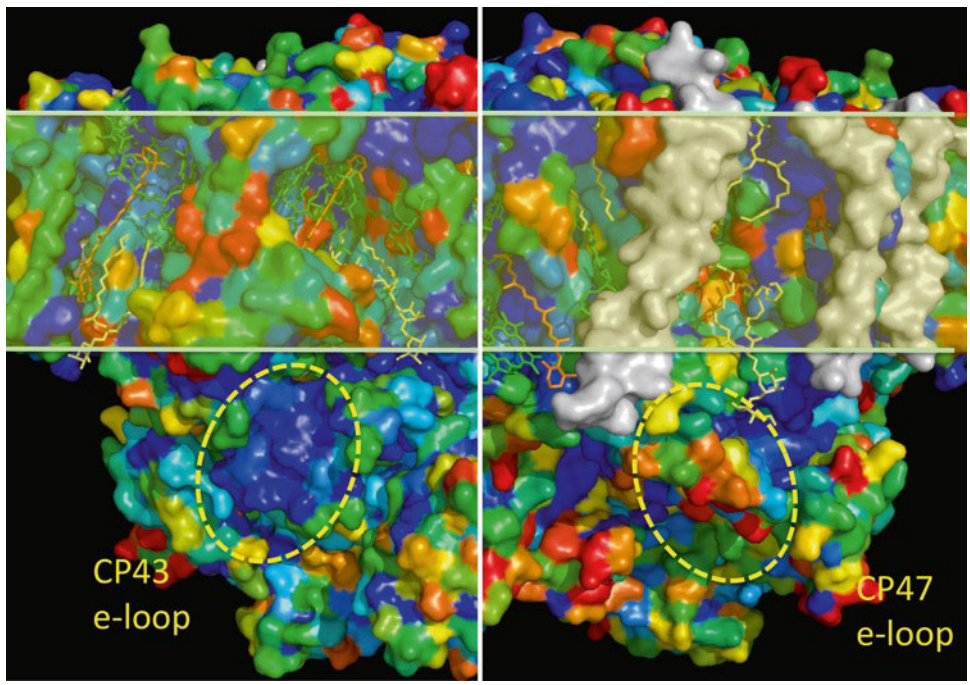


Fig. 13.8. Apparent exception to the low degree of conservation on the lumenal surface of the PSII complex. A patch of conserved residues is present on the e-loop of CP43, but not in the symmetrically related region of CP47. Since cyanobacterial PSII is known to stoichiometrically bind a homolog of PsbQ that is lost during crystallization, this conserved patch on CP43 may possibly be the binding site for ^{cy}PsbQ. However, it is presently impossible to rule out that it might be the binding site of another protein or simply reflects some other constraint. The figures and the corresponding color assignments for each amino acid were prepared as in Fig. 13.2.

^{cy}PsbQ. However, it is presently impossible to rule out that it might be the binding site of another protein or simply reflects some other constraint. However, it is unlikely that such a large degree of apparent conservation occurred by chance in the MSAs that were used to generate the projection of sequence similarity onto the protein surface. This location would be consistent with the finding that ^{cy}PsbQ is likely a lipoprotein because the conserved patch on the CP43 e-loop is situated on outside portion of CP43 exposed to the lipid bilayer. Based upon these considerations, it is tempting to speculate that ^{cy}PsbQ is actually bound to the outside face of the CP43 e-loop and that it exerts its stabilizing effects upon the WOC by reducing the mobility of the e-loop. This may be where ^{gr}PsbQ binds to the PSII complex in higher plants and green algae. Based upon considerations discussed below, PsbP in plants and green algae may occupy the homologous position of PsbV in cyanobacteria and red algae.

V. Biogenesis of PSII

With over 20 different protein subunits and nearly 80 cofactors per monomer, plus multiple protein co-factors mediating assembly, the biogenesis of PSII remains poorly understood despite considerable progress. The process is certainly more complex when compared to the anoxygenic photosynthetic organisms like purple bacteria. Much of this complexity can be attributed to increased structural and enzymatic complexity that accompanied the evolution of the H_2O-oxidation capacity. The PSII complex is also susceptible to photodamage and it undergoes repeated cycles of subunit replacement in response to this photodamage. The availability of structural information on PSII, combined with molecular genetic and biochemical analysis, is allowing an increasingly detailed understanding of the process (see Nixon et al., 2010, for recent review). Structure-function studies using mutants lacking specific subunits have led to models for a stepwise assembly process involving partially assembled PSII complexes as intermediates (Komenda et al., 2004, 2005; Sobotka et al., 2005). A scheme of the assembly process has been proposed based on the analysis of these mutants. The incorporation of cyt b_{559} in the membrane may initiate the assembly process by associating with D2 to form the D2-cytochrome b_{559} sub-complex. This is followed by the addition of preD1 and PsbI subunit to form an RC-like complex. The D1 protein is synthesized in a precursor form, termed preD1, which has a C-terminal extension of 16 residues in *Synechocystis*, but this length varies according to the species (Nixon et al., 1992). This extension is cleaved by a specific C-terminal processing protease (CtpA) to allow the assembly of the Mn_4-Ca (Anbudurai et al., 1994; Liao et al., 2000; Inagaki et al., 2001) since the C-terminal b-carboxyl serves as a ligand for the Mn_4-Ca. Binding of CP47 to this RC-like complex forms another assembly intermediate called the RC47 complex (Komenda et al., 2004). Attachment of CP43 to this RC47 complex enables the formation of the monomeric core PSII complex (RCC1); this being the starting point for the light-driven assembly of Mn_4-Ca to allow water oxidation to occur. Metal ions of the Mn_4-Ca are ligated to PSII by a light-driven process called photoactivation (Ananyev et al., 2001; Ono, 2001; Burnap, 2004). During photoactivation, Mn^{2+} ions are oxidized and their valency increases from Mn^{2+} to $Mn^{\geq 3+}$ as these oxidized ions become coordinated into the active site. The assembly of Mn_4-Ca occurs both during *de novo* synthesis of PSII and in mature photosynthetic membranes that sustained photodamage. Assembly of Mn_4-Ca is by itself a complex process and much of the research is still underway to better understand this mechanism. Interestingly, the transfer of excitation energy from the phycobilisome to the PSII RC is inefficient prior to the assembly of the Mn_4-Ca and excitation energy transfer becomes efficient only after the assembly of the Mn_4-Ca (Hwang et al., 2008). This is probably a protective mechanism that prevents photodamage to RCs in the process of assembly, which are not catalytically competent. The association of lumenally located extrinsic subunits PsbO, PsbV, PsbU and PsbQ stabilizes the assembled Mn_4-Ca cluster to form an active PSII capable of oxygen evolution in cyanobacteria. The last step in the assembly process appears to be the dimerization of PSII. Both the PSII dimer and monomer are found to be oxygen evolving, although dimers are considerably more active than monomers, at least after biochemical isolation (Nowaczyk et al., 2006). Recent studies argue the dimeric form of PSII to be a detergent induced artifact and to be absent

in vivo (Takahashi et al., 2009; Watanabe et al., 2009). However, the preponderance of evidence, notably, electron micrographic analysis of intact thylakoids, indicates that the dimer is indeed the main functional unit found in vivo (see *e.g.* Seibert et al., 1987; Bumba and Vacha, 2003). Several small, single TMH subunits, PsbM, PsbTc and PsbL, appear to help mediate the association of PSII monomers into the dimeric configuration. For example, native gel electrophoresis has been used to show that dimer formation is dependent on the presence of PsbTc (Bentley et al., 2008). Also, there is a decrease in the amount of PSII dimers in the PsbM deletion mutant (Bentley et al., 2008). PsbL is closely associated with PsbM and PsbTc based on the crystal structure of PSII (Guskov et al., 2009). This protein subunit has a very conserved C-terminus and transmembrane region. The C-terminus is believed to be interacting with PsbM and PsbTc (Luo and Eaton-Rye, 2008). A tyrosine residue in a TMH has been shown to interact with a lipid moiety (Luo and Eaton-Rye, 2008). Based on these observations, PsbL is believed to stabilize the PSII dimers along with PsbM and PsbTc.

As discussed below, there are several other low molecular weight protein subunits, including the membrane-anchored lipoprotein Psb27, that have been detected in gently purified PSII complexes. Some of these proteins act as cofactors in the assembly process and transiently associate with the PSII intermediates during biogenesis and repair (Nowaczyk et al., 2006; Mamedov et al., 2007; Roose and Pakrasi, 2007). Functions of several of these subunits have been described and information on these can be found in recent reviews (Roose et al., 2007b; Nixon et al., 2010).

A. Photodamage to PSII

Light, the driver of photosynthesis, is also a major stress factor and can cause irreversible damage to the components of the photosynthetic apparatus. Damage is due to reactive oxygen species (Lupinkova and Komenda, 2004) created by, for example, oxygen interacting with chlorophyll triplets formed during charge recombination (Keren et al., 1997; Cser and Vass, 2007). Because of this damage, a major pathway for the biogenesis of PSII is a repair pathway that involves the replacement of damaged components. Thus, PSII repair does not imply the chemical reversal of oxidative damage to proteins, but instead, PSII repair refers to the removal of damaged protein (*e.g.* D1) and the insertion of a replacement protein. A decline in the photosynthetic activity due to light stress is called photoinhibition (Vass et al., 1992; Aro et al., 1993; Adir et al., 2003; Murata et al., 2007; Vass and Cser, 2009). PSII is the most susceptible component of the photosynthetic mechanism and the D1 protein subunit of PSII sustains the highest rate of damage and exhibits the highest metabolic turnover rate of all the photosynthetic proteins. To cope with photodamage, a complex and efficient repair process operates to replace the irreversibly damaged D1 protein with a newly synthesized copy of the D1 protein (Nixon et al., 2005, 2010). Interestingly, radioactive pulse-chase experiments in vivo have shown the de novo synthesis of replacement D1 and degradation of damaged D1 are coupled (Komenda and Barber, 1995; Komenda et al., 2000). Therefore, it appears that biosynthesis and repair are linked, although the mechanism for the interesting coordination of protein degradation and synthesis is not understood. Models for the damage-repair cycle (*e.g.* Nixon et al., 2005) begins with damage to PSII, a sensing mechanism that elicits the repair mechanism, the monomerization of the PSII dimer and the dissociation or loss of certain other subunits. It is tempting to hypothesize that a sensing mechanism exists to detect damaged PSII complexes and specifically target them for repair; however, evidence for such a mechanism has yet to be obtained. An alternative mechanism could be the induction of increased specific D1 proteolysis and nascent D1 synthesis activity to ensure the population of PSII in the thylakoids is relatively young and thus unlikely to have become inactivated. Further work would be required to distinguish between these alternatives. Regardless of the initiation mechanism, the D1 replacement process represents a structural challenge: much of the D1 protein and the entirety of the Mn_4-Ca are deeply buried within the dimeric PSII complex and it is likely that the repair of D1 requires significant structural rearrangements that allow proteases to access the damaged D1. Recent work indicates that the cytoplasmically exposed amino terminus of the D1 protein is critical for the process (Komenda et al., 2007). The reassembly

of functional PSII could be very similar to steps involved in the de novo biogenesis of PSII.

VI. Accessory Proteins Associated with PSII Biogenesis

A. Proteases Associated with Repair of PSII

FtsH and Deg proteases have long been identified as important candidates for the D1 degradation. FtsH belongs to a family of ATP dependent proteases consisting of an N-terminal transmembrane domain, a hydrophilic ATPase domain and a Zn^{2+} metalloprotease domain (Silva et al., 2003; Ito and Akiyama, 2005). There are four FtsH proteases encoded in the *Synechocystis* genome, of which FtsH2 seems to play the dominant role in D1 degradation (Silva et al., 2003; Komenda et al., 2006). These proteases bear a moderate sequence similarity to the FtsH protease from *E. coli*. The role of FtsH in the degradation of damaged D1 seems to be evolutionarily conserved from cyanobacteria through plants. FtsH has been shown to be involved during the early stages of D1 degradation and is not restricted to the removal of D1 breakdown products (Silva et al., 2003). The observation that FtsH mutant strains in both *Synechocystis* and *Arabidopsis* do not completely block the D1 degradation suggests that there are additional proteases involved in this process, but operate with much lower efficiency with respect to D1 degradation. FtsH is known to have an ATP-dependent unfolding capacity, which is likely to be important for the extraction of D1 from the membrane for processive proteolytic degradation. As noted above, it was recently shown that an exposed N-terminus of D1 is essential for its rapid degradation by FtsH (Komenda et al., 2007). Presumably, FtsH recognizes the N-terminus of the damaged (destabilized) D1 protein subunit to engage in degradation. Destabilization in D1 might be either due to oxidative damage during photoinhibition or loss of cofactors, metal ions or protein subunits (Nixon et al., 2005; Komenda et al., 2010). However, as noted above, a specific damage-sensing mechanism has not been established and the possibility of a general, indiscriminate turnover of PSII under damage-prone conditions, while seemingly unlikely, has not been eliminated.

The other important family, the HtrA/Deg proteases, consists of membrane associated serine proteases that are present in both cyanobacteria and green plants. There are several members of the Deg protease gene family in *Synechocystis* (3 members) (Huesgen et al., 2009) and *Arabidopsis* (16 members) (Huesgen et al., 2009). DegP2 from *Arabidopsis* has been shown to be required for the initial D1 cleavage event *in vitro* to produce breakdown products (Hauszuhl et al., 2001). In general, Deg proteases that are located in chloroplasts respond to light stress and have an important, though perhaps not essential, role in D1 degradation. The members of Deg proteases from *Synechocystis*, on the other hand, do not seem to be involved in D1 degradation. A mutation of all three Deg proteases in *Synechocystis* did not affect the repair process (Nixon et al., 2005). Thus, in cyanobacteria Deg proteases are not obviously involved D1 degradation and their function is remains unknown. Therefore, the role of DegP protease during D1 degradation in chloroplasts might have occurred after the divergence of cyanobacteria and green algae.

B. Psb27

Psb27 is a lipoprotein transiently associated with PSII complexes (Nowaczyk et al., 2006; Mamedov et al., 2007). It is considered to prevent the premature binding of extrinsic subunits PsbO, PsbV, and PsbU, which appears to allow the efficient incorporation of Mn into the WOC (Roose and Pakrasi, 2007). This blocking of extrinsic subunits is presumed to occur primarily by preventing the attachment of PsbO. The 3D structure of Psb27 was resolved using NMR techniques (Cormann et al., 2009; Mabbitt et al., 2009) and its docking site onto the lumenal side of PSII was modeled (Cormann et al., 2009). Based on this model, two helices III and IV of Psb27 are the most probable sites of interaction with PSII. This site overlaps with the binding site of PsbO and explains the prevention of binding of PsbO in the presence of Psb27 (Nowaczyk et al., 2006; Mamedov et al., 2007; Roose and Pakrasi, 2007). According to this model, the N-terminus of Psb27 containing the lipid modification is located at the monomer-monomer interface. This would explain the release of Psb27 from monomeric PSII without the action of lipase or other cofactors. Psb27

shows an interesting pattern of conservation. The pattern of structural conservation deduced from the sequence analysis of Psb27 shows that helices III and IV are very conserved however, the actual modes of interaction with PSII awaits experimental interrogation.

VII. Are PSII Assembly Proteins the Evolutionary Source of PsbP and PsbQ in Green Algae and Plants?

One of the more remarkable findings of plant genome sequencing and proteomics projects has been the identification of multiple paralogs for several of the ^{gr}PSII extrinsic proteins (e.g. see Schubert et al., 2002). Though similar to the proteins of the ^{gr}PSII WOC, these paralogs do not appear have a structural role in PSII, but instead seem to be involved in assembly. This includes at least two additional variants of the PsbP protein. These paralogous proteins have been annotated as PsbP-like (PPL) proteins. They show greater sequence similarity to cyanobacterial PsbP than the bona fide PsbP associated with the WOC of ^{gr}PSII. Interestingly, two of these paralogs, PPL1 and PPL2, of *Arabidopsis* appear to facilitate assembly and/or maintenance of the PSII and Type-1 dehydrogenase complexes, respectively (Ishihara et al., 2007). The ^{gr}PsbP of the WOC has been shown to bind Mn^{2+} and facilitate photoactivation (Bondarava et al., 2007). It remains to be determined whether or not the PPL proteins have similar properties. However, the presence of these additional 'assembly variants' of the PsbP and the existence of close homologs in the cyanobacteria seem to provide an important clue to the evolution of PSII. It is likely that ^{cy}PsbP was the ancestral type of PsbP that underwent multiple gene duplication events during the course of evolution (Sato, 2010). These duplication events led to multiple copies of ^{gr}PsbP, some retained the function as an assembly factor and, possibly, the other paralog replaced the function of PsbV (see Fig. 13.9).

Of course, it is presently impossible to exclude alternative scenarios for the evolutionary replacement of PsbV by PsbP in ^{gr}PSII. For example, PsbP may have first directly replaced the function of PsbV as structural component of the WOC and then, following *psbP* gene duplication, evolved to acquire assembly functions such as now seen in PPL proteins associated with the assembly of PSII and Type-1 dehydrogenase (Ishihara et al., 2007). In this regard, increased efforts to determine whether or not ^{cy}PsbP is involved in ^{cy}PSII assembly may be worthwhile since current results on this

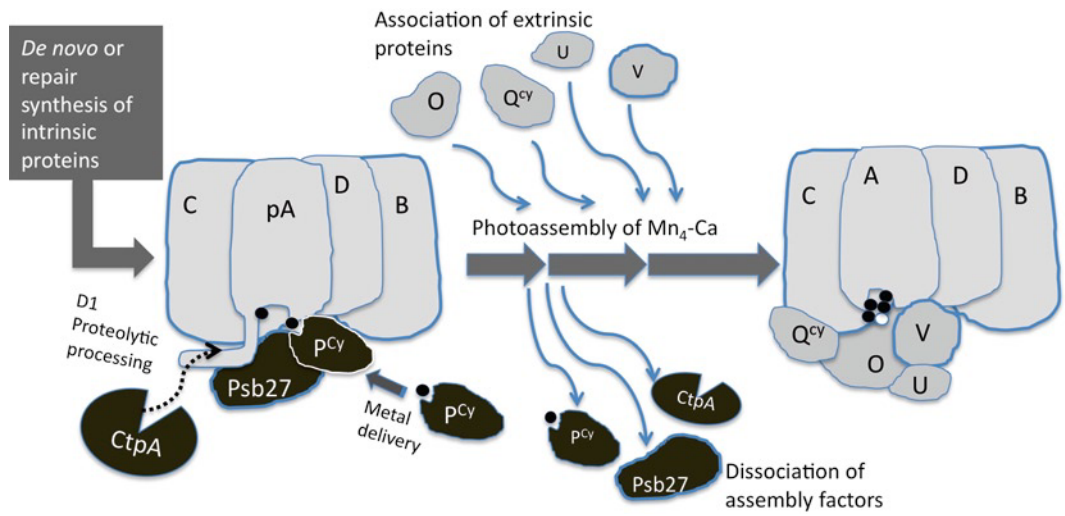


Fig. 13.9. Speculative model: Accessory proteins facilitate the later steps in the assembly of PSII.

point remain equivocal (Ishikawa et al., 2005; Summerfield et al., 2005). If ^{cy}PsbP acts as an Mn^{2+} carrier during Mn_4-Ca assembly as suggested with ^{gr}PsbP (Bondarava et al., 2007), it would lend support to the idea that ^{gr}PsbP evolved from an assembly factor replacing PsbV. A substitution also seems likely since the simple loss of PsbV has quite severe negative phenotypic consequences in extant cyanobacteria (Shen et al., 1995, 1998). Interestingly, the genomes of some marine cyanobacteria in the *Prochlorococcus* genus, such as *Prochlorococcus marinus* MED4, lack PsbV, yet they do have a copy of PsbP. Since *Prochlorococcus* derives from a *Synechococcus*-like ancestor (Zhaxybayeva et al., 2009), it is possible that the original evolution of PsbP in ^{gr}PSII as component of the WOC has been recapitulated during the evolution of some, but not all, of the *Prochlorococcus* strains. However, isolation and subunit characterization of PSII from *Prochlorococcus marinus* MED4 still needs to be performed to evaluate whether or not PsbP is actually found in the WOC of this organism.

Acknowledgments

The authors with to thank Prof. Jian-Ren Shen for discussions regarding unpublished structural data collected by his group. This work was supported by the National Science Foundation under Grant No. 0818371 and US Department of Energy Grant No. DE-FG02-08ER15968.

References

Adir N, Zer H, Shochat S and Ohad I (2003) Photoinhibition – A historical perspective. Photosyn Res 76: 343–370

Allahverdiyeva Y, Mamedov F, Suorsa M, Styring S, Vass I and Aro EM (2007) Insights into the function of PsbR protein in *Arabidopsis thaliana*. Biochim Biophys Acta 1767: 677–685

Allen JP and Williams JC (2011) The evolutionary pathway from anoxygenic to oxygenic photosynthesis examined by comparison of the properties of photosystem II and bacterial reaction centers. Photosynth Res 107: 59–69

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402

Ananyev GM, Zaltsman L, Vasko C and Dismukes GC (2001) The inorganic biochemistry of photosynthetic oxygen evolution/water oxidation. Biochim Biophys Acta 1503: 52–68

Anbudurai PR, Mor TS, Ohad I, Shestakov SV and Pakrasi HB (1994) The *ctpA* gene encodes the C-terminal processing protease for the D1 protein of the photosystem II reaction center complex. Proc Natl Acad Sci USA 91: 8082–8086

Aro EM, Virgin I and Andersson B (1993) Photoinhibition of Photosystem II. Inactivation, protein damage and turnover. Biochim Biophys Acta 1143: 113–134

Barry B and Babcock G (1987) Tyrosine radicals are involved in the photosynthetic oxygen-evolving system. Proc Natl Acad Sci USA 84: 7099–7103

Bentley FK, Luo H, Dilbeck P, Burnap RL and Eaton-Rye JJ (2008) Effects of inactivating *psbM* and *psbT* on photodamage and assembly of photosystem II in *Synechocystis* sp. PCC 6803. Biochemistry 47: 11637–11646

Blankenship RE (1992) Origin and early evolution of photosynthesis. Photosynth Res 33: 91–111

Blankenship RE and Hartman H (1998) The origin and evolution of oxygenic photosynthesis. Trends Biochem Sci 23: 94–97

Boekema EJ, Nield J, Hankamer B and Barber J (1998) Localization of the 23-kDa subunit of the oxygen-evolving complex of photosystem II by electron microscopy. Eur J Biochem 252: 268–276

Bondarava N, Un S and Krieger-Liszkay A (2007) Manganese binding to the 23 kDa extrinsic protein of Photosystem II. Biochim Biophys Acta 1767: 583–588

Boussac A and Rutherford AW (1988) Ca^{2+} binding to the oxygen evolving enzyme varies with the redox state of the Mn enzyme. Biochim Biophys Acta 236: 432–436

Bricker TM, Morvant J, Masri N, Sutton HM and Frankel LK (1998) Isolation of a highly active photosystem II preparation from *Synechocystis* 6803 using a histidine-tagged mutant of CP 47. Biochim Biophys Acta 1409: 50–57

Bryant DA and Frigaard NU (2006) Prokaryotic photosynthesis and phototrophy illuminated. Trends in Microbiol 14: 488–496

Bumba L and Vacha F (2003) Electron microscopy in structural studies of Photosystem II. Photosyn Res 77: 1–19

Burnap RL (2004) D1 protein processing and Mn cluster assembly in light of the emerging photosystem II structure. Phys Chem Chem Phys 6: 4803–4809

Burnap RL and Trench RK (1989a) Sites of synthesis of polypeptides of the cyanelles of *Cyanophora paradoxa*: I. Characterization of the photosynthetic apparatus. Proc R Soc Lond B 238: 53–72

Burnap RL and Trench RK (1989b) Sites of synthesis of polypeptides of the cyanelles of *Cyanophora paradoxa*: II. In vivo labeling in the presence of translational and transcriptional inhibitors. Proc R Soc Lond B 238: 73–87

Burnap RL and Trench RK (1989c) Sites of synthesis of polypeptides of the cyanelles of *Cyanophora paradoxa*:III. In vitro synthesis of cyanellar polypeptides. Proc R Soc Lond B 238: 89–102

Campbell D, Eriksson MJ, Oquist G, Gustafsson P and Clarke AK (1998) The cyanobacterium *Synechococcus* resists UV-B by exchanging photosystem II reaction-center D1 proteins. Proc Natl Acad Sci USA 95: 364–369

Christen G and Renger G (1999) The role of hydrogen bonds for the multiphasic $P680^{+\bullet}$ reduction by Y_Z in photosystem II with intact oxygen evolution capacity. Analysis of kinetic H/D isotope exchange effects. Biochemistry 38: 2068–2077

Chu HA, Hillier W and Debus RJ (2004) Evidence that the C-terminus of the D1 polypeptide of photosystem II is ligated to the manganese ion that undergoes oxidation during the S_1 to S_2 transition: an isotope-edited FTIR study. Biochemistry 43: 3152–3166

Cormann KU, Bangert JA, Ikeuchi M, Rogner M, Stoll R and Nowaczyk MM (2009) Structure of Psb27 in solution: Implications for transient binding to photosystem II during biogenesis and repair. Biochemistry 48: 8768–8770

Cser K and Vass I (2007) Radiative and non-radiative charge recombination pathways in Photosystem II studied by thermoluminescence and chlorophyll fluorescence in the cyanobacterium *Synechocystis* 6803. Biochim Biophys Acta 1767: 233–243

Dau H and Haumann M (2008) The manganese complex of photosystem II in its reaction cycle – Basic framework and possible realization at the atomic level. Coordination Chem Rev 252: 273–295

Dayhoff MO, Schwartz RM and Orcutt BC (1978) A model of evolutionary change in proteins. In: Dayhoff MO (ed) Atlas of Protein Sequence and Structure, Supplement 3, Vol 5 pp 345–352. National Biomedical Research Foundation, Washington, DC

De Las Rivas J and Barber J (2004) Analysis of the structure of the PsbO protein and its implications. Photosynth Res 81: 329–343

De Las Rivas J, Heredia P and Roman A (2007) Oxygen-evolving extrinsic proteins (PsbO,P,Q,R): Bioinformatic and functional analysis. Biochim Biophys Acta 1767: 575–582

Debus RJ, Barry BS, Sithole I, Babcock GT and McIntosh L (1988) Directed mutagenesis indicates that the donor to $P680^+$ in photosystem II is Tyr-161 of the D1 polypeptide. Biochemistry 27: 9071–9074

Deisenhofer J, Epp O, Miki K, Huber R and Michel H (1984) X-ray structure analysis of a membrane protein complex electron density map at Ångstrom resolution and a model of the chromophores of the photosynthetic reaction center from *Rhodopseudomonas viridis*. J Mol Biol 180: 385–398

Diner BA, Ries DF, Cohen BN and Metz JG (1988) COOH-terminal processing of polypeptide D1 of the photosystem II reaction center of *Scenedesmus obliquus* is necessary for the assembly of the oxygen-evolving complex. J Biol Chem 263: 8972–8980

Diner BA, Bautista JA, Nixon PJ, Berthomieu C, Hienerwadel R, Britt RD, Vermaas WFJ and Chisholm DA (2004) Coordination of proton and electron transfer from the redox-active tyrosine, Y_Z, of Photosystem II and examination of the electrostatic influence of oxidized tyrosine, $Y_D^{\bullet}(H^+)$. Phys Chem Chem Phys 6: 4844–4850

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. Bmc Bioinformatics 5: 1–19

Enami I, Kamo M, Ohta H, Takahashi S, Miura T, Kusayanagi M, Tanabe S, Kamei A, Motoki A, Hirano M, Tomo T and Satoh K (1998) Intramolecular cross linking of the extrinsic 33 kDa protein leads to loss of oxygen evolution but not its ability of binding to photosystem. J Biol Chem 273: 4629–4634

Enami I, Suzuki T, Tada O, Nakada Y, Nakamura K, Tohri A, Ohta H, Inoue I and Shen JR (2005) Distribution of the extrinsic proteins as a potential marker for the evolution of photosynthetic oxygen-evolving photosystem II. FEBS J 272: 5020–5030

Enami I, Okumura A, Nagao R, Suzuki T, Iwai M and Shen JR (2008) Structures and functions of the extrinsic proteins of photosystem II from different species. Photosynth Res 98: 349–363

Ferreira KN, Iverson TM, Maghlaoui K, Barber J and Iwata S (2004) Architecture of the photosynthetic oxygen-evolving center. Science 303: 1831–1838

Forbush B, Kok B and McGloin MP (1971) Cooperation of charges in photosynthetic O_2 evolution: II Damping of flash yield oscillation, deactivation. Photochem Photobiol 14: 307–321

Frankel LK and Bricker TM (1995) Interaction of the 33 kDa extrinsic protein with photosystem II: identification of domains on the 33 kDa protein that are shielded from NHS biotinylation by photosystem II. Biochemistry 34: 7492–7497

Frankel LK, Cruz JA and Bricker TM (1999) Carboxylate groups on the manganese-stabilizing protein are required for its efficient binding to photosystem II. Biochemistry 38: 14271–14278

Frazao C, Enguita FJ, Coelho R, Sheldrick GM, Navarro JA, Hervas M, De la Rosa MA and Carrondo MA (2001) Crystal structure of low-potential cytochrome c_{549} from *Synechocystis* sp. PCC 6803 at 1.21 Å resolution. J Biol Inorg Chem 6: 324–332

Guskov A, Kern J, Gabdulkhakov A, Broser M, Zouni A and Saenger W (2009) Cyanobacterial photosystem II at 2.9-Ångstrom resolution and the role of quinones, lipids, channels and chloride. Nature Structural & Molecular Biology 16: 334–342

Hankamer B, Morris EP and Barber J (1999) Revealing the structure of the oxygen-evolving core dimer of photosystem II by cryoelectron crystallography. Nat Struct Biol 6: 560–564

Haumann M, Mulkidjanian A and Junge W (1997) Electrogenicity of electron and proton transfer at the oxidizing side of photosystem II. Biochemistry 36: 9304–9315

Haumann M, Liebisch P, Muller C, Barra M, Grabolle M and Dau H (2005) Photosynthetic O_2 formation tracked by time-resolved X-ray experiments. Science 310: 1019–1021

Hauszuhl K, Andersson B and Adamska I (2001) A chloroplast DegP2 protease performs the primary cleavage of the photodamaged D1 protein in plant photosystem II. EMBO J 20: 713–722

Holzwarth AR, Muller MG, Reus M, Nowaczyk M, Sander J and Rogner M (2006) Kinetics and mechanism of electron transfer in intact photosystem II and in the isolated reaction center: Pheophytin is the primary electron acceptor. Proc Natl Acad Sci USA 103: 6895–6900

Huesgen PF, Schuhmann H and Adamska I (2009) Deg/HtrA proteases as components of a network for photosystem II quality control in chloroplasts and cyanobacteria. Res Microbiol 160: 726–732

Hung CH, Hwang HJ, Chen YH, Chiu YF, Ke SC, Burnap RL and Chu HA (2010) Spectroscopic and functional characterizations of cyanobacterium *Synechocystis* PCC 6803 mutants on and near the heme axial ligand of cytochrome b_{559} in photosystem II. J Biol Chem 285: 5653–5663

Hwang HJ, Nagarajan A, McLain A and Burnap RL (2008) Assembly and disassembly of the photosystem II manganese cluster reversibly alters the coupling of the reaction center with the light-harvesting phycobilisome. Biochemistry 47: 9747–9755

Ifuku K, Ishihara S, Shimamoto R, Ido K and Sato F (2008) Structure, function, and evolution of the PsbP protein family in higher plants. Photosynth Res 98: 427–437

Inagaki N, Maitra R, Satoh K and Pakrasi HB (2001) Amino acid residues that are critical for in vivo catalytic activity of CtpA, the carboxyl-terminal processing protease for the D1 protein of photosystem II. J Biol Chem 276: 30099–30105

Ishihara S, Takabayashi A, Ido K, Endo T, Ifuku K and Sato F (2007) Distinct functions for the two PsbP-like proteins PPL1 and PPL2 in the chloroplast thylakoid lumen of Arabidopsis. Plant Physiol 145: 668–679

Ishikawa Y, Schrode RW and Funk C (2005) Functional analysis of the PsbP-like protein (sll1418) in *Synechocystis* sp. PCC 6803. Photosyn Res 84: 257–262

Ito K and Akiyama Y (2005) Cellular functions, mechanism of action, and regulation of the FtsH protease. Ann Rev Microbiol 59: 211–231

Ivleva NB, Shestakov SV and Pakrasi HB (2000) The carboxyl-terminal extension of the precursor D1 protein of photosystem II is required for optimal photosynthetic performance of the cyanobacterium *Synechocystis* sp. PCC 6803. Plant Physiol 124: 1403–1412

Johnson MS, Sutcliffe MJ and Blundell TL (1990) Molecular anatomy: phyletic relationships derived from three-dimensional structures of proteins. J Mol Evol 30: 43–59

Joliot P, Barbieri G and Chabaud R (1969) Un nouveau modele des centres photochimique du systeme II. Photochem Photobiol 10: 309–329

Kamiya N and Shen JR (2003) Crystal structure of oxygen-evolving photosystem II from *Thermosynechococcus vulcanus* at 3.7 Å resolution. Proc Natl Acad Sci USA 100: 98–103

Kargul J, Maghlaoui K, Murray JW, Deak Z, Boussac A, Rutherford AW, Vass I and Barber J (2007) Purification, crystallization and X-ray diffraction analyses of the *T. elongatus* PSII core dimer with strontium replacing calcium in the oxygen-evolving complex. Biochim Biophys Acta 1767: 404–413

Kashino Y, Lauber WM, Carroll JA, Wang Q, Whitmarsh J, Satoh K and Pakrasi HB (2002) Proteomic analysis of a highly active photosystem II preparation from the cyanobacterium *Synechocystis* sp. PCC 6803 reveals the presence of novel polypeptides. Biochemistry 41: 8004–8012

Kashino Y, Inoue-Kashino N, Roose JL and Pakrasi HB (2006) Absence of the PsbQ protein results in destabilization of the PsbV protein and decreased oxygen evolution activity in cyanobacterial photosystem II. J Biol Chem 281: 20834–20841

Keren N, Berg A, van Kan PJ, Levanon H and Ohad II (1997) Mechanism of photosystem II photoinactivation and D1 protein degradation at low light: The role of back electron flow. Proc Natl Acad Sci USA 94: 1579–1584

Kern J and Renger G (2007) Photosystem II: structure and mechanism of the water:plastoquinone oxidoreductase. Photosynth Res 98: 53–80

Kern J, Biesiadka J, Loll B, Saenger W and Zouni A (2007) Structure of the Mn_4-Ca cluster as derived from X-ray diffraction. Photosynth Res 92: 389–405

Kok B, Forbush B and McGloin M (1970) Cooperation of charges in photosynthetic oxygen evolution: I. A linear four step mechanism. Photochem Photobiol 11: 457–475

Komenda J and Barber J (1995) Comparison of *psbO* and *psbH* deletion mutants of *Synechocystis* PCC 6803 indicates that degradation of D1 protein is regulated by the Q_B site and dependent on protein synthesis. Biochemistry 34: 9625–9631

Komenda J, Hassan HAG, Diner BA, Debus RJ, Barber J and Nixon PJ (2000) Degradation of the Photosystem II D1 and D2 proteins in different strains of the cyanobacterium *Synechocystis* PCC 6803 varying with respect to the type and level of *psbA* transcript. Plant Mol Biol 42: 635–645

Komenda J, Reisinger V, Muller BC, Dobakova M, Granvogl B and Eichacker LA (2004) Accumulation of the D2 protein is a key regulatory step for assembly of the photosystem II reaction center complex in *Synechocystis* PCC 6803. J Biol Chem 279: 48620–48629

Komenda J, Tichy M and Eichacker LA (2005) The PsbH protein is associated with the inner antenna CP47 and facilitates D1 processing and incorporation into PSII in the cyanobacterium *Synechocystis* PCC 6803. Plant Cell Physiol 46: 1477–1483

Komenda J, Barker M, Kuvikova S, de Vries R, Mullineaux CW, Tichy M and Nixon PJ (2006) The FtsH protease Slr0228 is important for quality control of photosystem II in the thylakoid membrane of *Synechocystis* sp. PCC 6803. J Biol Chem 281: 1145–1151

Komenda J, Tichy M, Prasil O, Knoppova J, Kuvikova S, de Vries R and Nixon PJ (2007) The exposed N-terminal tail of the D1 subunit is required for rapid D1 degradation during photosystem II repair in *Synechocystis* sp PCC 6803. Plant Cell 19: 2839–2854

Komenda J, Knoppova J, Krynicka V, Nixon PJ and Tichy M (2010) Role of FtsH2 in the repair of Photosystem II in mutants of the cyanobacterium *Synechocystis* PCC 6803 with impaired assembly or stability of the $CaMn_4$ cluster. Biochim Biophys Acta 1797: 566–575

Krammer EM, Sebban P and Ullmann GM (2009a) Profile hidden Markov models for analyzing similarities and dissimilarities in the bacterial reaction center and photosystem II. Biochemistry 48: 1230–1243

Krammer EM, Till MS, Sebban P and Ullmann GM (2009b) Proton-transfer pathways in photosynthetic reaction centers analyzed by profile hidden Markov models and network calculations. J Mol Biol 388: 631–643

Krivanek R, Kern J, Zouni A, Dau H and Haumann M (2007) Spare quinones in the Q_B cavity of crystallized photosystem II from *Thermosynechococcus elongatus*. Biochim Biophys Acta 1767: 520–527

Kuhl H, Kruip J, Seidler A, Krieger-Liszkay A, Bunker M, Bald D, Scheidig AJ and Rogner M (2000) Towards structural determination of the water-splitting enzyme – Purification, crystallization, and preliminary crystallographic studies of photosystem II from a thermophilic cyanobacterium. J Biol Chem 275: 20652–20659

Kulkarni RD and Golden SS (1994) Adaptation to high light intensity in *Synechococcus* sp. strain PCC 7942: regulation of three psbA genes and two forms of the D1 protein. J Bacteriol 176: 959–965

Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T and Ben-Tal N (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res 33: W299–W302

Liao DI, Qian J, Chisholm DA, Jordan DB and Diner BA (2000) Crystal structures of the photosystem II D1 C-terminal processing protease. Nat Struct Biol 7: 749–753

Lichtarge O, Bourne HR and Cohen FE (1996a) An evolutionary trace method defines binding surfaces common to protein families. J Mol Biol 257: 342–358

Lichtarge O, Bourne HR and Cohen FE (1996b) Evolutionarily conserved G(alpha beta gamma) binding surfaces support a model of the G protein-receptor complex. Proc Natl Acad Sci USA 93: 7507–7511

Loll B, Kern J, Saenger W, Zouni A and Biesiadka J (2005) Towards complete cofactor arrangement in the 3.0 Å resolution structure of photosystem II. Nature 438: 1040–1044

Luo H and Eaton-Rye J (2008) Directed mutagenesis of the transmembrane domain of the PsbL subunit of photosystem II in *Synechocystis* sp. PCC 6803. Photosyn Res 98: 337–347

Lupinkova L and Komenda J (2004) Oxidative modifications of the Photosystem II D1 protein by reactive oxygen species: from isolated protein to cyanobacterial cells. Photochem Photobiol 79: 152–162

Mabbitt PD, Rautureau GJP, Day CL, Wilbanks SM, Eaton-Rye JJ and Hinds MG (2009) Solution structure of Psb27 from cyanobacterial photosystem II. Biochemistry 48: 8771–8773

Magnuson A, Rova M, Mamedov F, Fredriksson PO and Styring S (1999) The role of cytochrome b_{559} and tyrosine D in protection against photoinhibition during in vivo photoactivation of photosystem II. Biochim Biophys Acta 1411: 180–191

Mamedov F, Nowaczyk MM, Thapper A, Rogner M and Styring S (2007) Functional characterization of monomeric photosystem II core preparations from *Thermosynechococcus elongatus* with or without the Psb27 protein. Biochemistry 46: 5542–5551

McEvoy JP and Brudvig GW (2004) Structure-based mechanism of photosynthetic water oxidation. Phys Chem Chem Phys 6: 4754–4763

McEvoy JP and Brudvig GW (2006) Water-splitting chemistry of photosystem II. Chem Rev 106: 4455–4483

Miqyass M, Marosvolgyi MA, Nagel Z, Yocum CF and van Gorkom HJ (2008) S-state dependence of the calcium requirement and binding characteristics in the oxygen-evolving complex of photosystem II. Biochemistry 47: 7915–7924

Miura T, Shen JR, Takahashi S, Kamo M, Nakamura E, Ohta H, Kamei A, Inoue Y, Domae N, Takio K, Nakazato K and Enami I (1997) Identification of domains on the extrinsic 33 kDa protein possibly involved in electrostatic interaction with photosystem II complex by means of chemical modification. J Biol Chem 272: 3788–3798

Mix LJ, Haig D and Cavanaugh CM (2005) Phylogenetic analyses of the core antenna domain: investigating the origin of photosystem I. J Mol Evol 60: 153–163

Mizusawa N, Kimura Y, Ishii A, Yamanari T, Nakazawa S, Teramoto H and Ono TA (2004) Impact of replacement of D1 C-terminal alanine with glycine on structure and function of photosynthetic oxygen evolving complex. J Biol Chem 279: 29622–29627

Murata N, Takahashi S, Nishiyama Y and Allakhverdiev SI (2007) Photoinhibition of photosystem II under environmental stress. Biochim Biophys Acta 1767: 414–421

Nanba O and Satoh K (1987) Isolation of a photosystem II reaction center consisting of D1 and D2 polypeptides and cytochrome b-559. Proc Natl Acad Sci USA 84: 109–112

Nield J, Kruse O, Ruprecht J, da Fonseca P, Buchel C and Barber J (2000) Three-dimensional structure of *Chlamydomonas reinhardtii* and *Synechococcus elongatus* photosystem II complexes allows for comparison of

their oxygen-evolving complex organization. J Biol Chem 275: 27940–27946

Nixon PJ, Trost JT and Diner BA (1992) Role of the carboxy terminus of polypeptide D1 in the assembly of a functional water oxidizing manganese cluster in photosystem II of the cyanobacterium *Synechocystis* sp. PCC 6803: assembly requires a free carboxyl group at C terminal position 344. Biochemistry 31: 10859–10871

Nixon PJ, Barker M, Boehm M, de Vries R and Komenda J (2005) FtsH-mediated repair of the photosystem II complex in response to light stress. J Exp Bot 56: 357–363

Nixon PJ, Michoux F, Yu J, Boehm M and Komenda J (2010) Recent advances in understanding the assembly and repair of photosystem II. Ann Bot 106: 1–16

Noren GH and Barry BA (1992) The YF161D1 mutant of *Synechocystis* 6803 exhibits an EPR signal from a light-induced photosystem II radical. Biochemistry 31: 3335–3342

Nowaczyk MM, Hebeler R, Schlodder E, Meyer HE, Warscheid B and Rogner M (2006) Psb27, a cyanobacterial lipoprotein, is involved in the repair cycle of photosystem II. Plant Cell 18: 3121–3131

Odom W and Bricker TM (1992) Interaction of CPa-1 with the manganese-stabilizing protein of photosystem II: Identification of domains cross-linked by 1-ethyl-3-[3-(dimethylamino)propyl]carbodiimide. Biochemistry 31: 5616–5620

Ono T (2001) Metallo-radical hypothesis for photoassembly of $(Mn)_4$-cluster of photosynthetic oxygen evolving complex. Biochim Biophys Acta 1503: 40–51

Rappaport F and Diner BA (2002) Structure, dynamics, and energetics of the primary photochemistry of Photosystem II of oxygenic photosynthesis. Annu Rev Plant Physiol Plant Mol Biol 53: 551–580

Rappaport F and Diner BA (2008) Primary photochemistry and energetics leading to the oxidation of the (Mn)4Ca cluster and to the evolution of molecular oxygen in Photosystem II. Coordination Chem Rev 252: 259–272

Rappaport F and Lavergne J (1997) Charge recombination and proton transfer in manganese-depleted photosystem II. Biochemistry 36: 15294–15302

Renger G and Renger T (2008) Photosystem II: The machinery of photosynthetic water splitting. Photosyn Res 98: 53–80

Roose JL, Kashino Y and Pakrasi HB (2007a) The PsbQ protein defines cyanobacterial Photosystem II complexes with highest activity and stability. Proc Natl Acad Sci USA 104: 2548–2553

Roose JL, Wegener KM and Pakrasi HB (2007b) The extrinsic proteins of Photosystem II. Photosynth Res 92: 369–387

Roose JL and Pakrasi HB (2008) The Psb27 protein facilitates manganese cluster assembly in photosystem II. J Biol Chem: 283: 4044–4050

Rutherford AW (1986) How close is the analogy between the reaction center of photosystem II and that of the purple bacteria? Biochem Soc Trans 14: 15–17

Sadekar S, Raymond J and Blankenship RE (2006) Conservation of distantly related membrane proteins: photosynthetic reaction centers share a common structural core. Mol Biol Evol 23: 2001–2007

Sato N (2010) Phylogenomic and structural modeling analyses of the PsbP superfamily reveal multiple small segment additions in the evolution of photosystem II-associated PsbP protein in green plants. Mol Phylogenet Evol 56: 176–186

Sauer K (1979) Photosynthesis-Light Reactions. Annu Rev Phys Chem 30: 155–178

Schubert WD, Klukas O, Krauss N, Saenger W, Fromme P and Witt HT (1997) Photosystem I of *Synechococcus elongatus* at 4 Å resolution: comprehensive structure analysis. J Mol Biol 272: 741–769

Schubert WD, Klukas O, Saenger W, Witt HT, Fromme P and Krauss N (1998) A common ancestor for oxygenic and anoxygenic photosynthetic systems: A comparison based on the structural model of photosystem I. J Mol Biol 280: 297–314

Schubert M, Petersson UA, Haas BJ, Funk C, Schroder WP and Kieselbach T (2002) Proteome map of the chloroplast lumen of *Arabidopsis thaliana*. J Biol Chem 277: 8354–8365

Seibert M, Dewit M and Staehelin LA (1987) Structural localization of the O_2-evolving apparatus to multimeric (tetrameric) particles on the lumenal surface of freeze-etched photosynthetic membranes. J Cell Biol 105: 2257–2265

Seibert M, Tamura N and Inoue Y (1989) Lack of photoactivation capacity in Scenedesmus obliquus LF-1 results from loss of half the high-affinity manganese binding site. Relationship to the unprocessed D1 protein. Biochim Biophys Acta 974: 185–191

Shen JR, Vermaas W and Inoue Y (1995) The role of cytochrome c_{550} as studied through reverse genetics and mutant characterization in *Synechocystis* sp. PCC 6803. J Biol Chem 270: 6901–6907

Shen JR, Qian M, Inoue Y and Burnap RL (1998) Functional characterization of *Synechocystis* sp. PCC 6803 delta psbU and delta psbV mutants reveals important roles of cytochrome c_{550} in cyanobacterial oxygen evolution. Biochemistry 37: 1551–1558

Silva P, Thompson E, Bailey S, Kruse O, Mullineaux CW, Robinson C, Mann NH and Nixon PJ (2003) FtsH is involved in the early stages of repair of photosystem II in *Synechocystis* sp PCC 6803. Plant Cell 15: 2152–2164

Sobotka R, Komenda J, Bumba L and Tichy M (2005) Photosystem II assembly in CP47 mutant of *Synechocystis* sp. PCC 6803 is dependent on the level of chlorophyll precursors regulated by ferrochelatase. J Biol Chem 280: 31595–31602

Sproviero EM, Gascon JA, McEvoy JP, Brudvig GW and Batista VS (2008) Quantum mechanics/molecular mechanics study of the catalytic cycle of water splitting in photosystem II. J Am Chem Soc 130: 3428–3442

Summerfield TC, Winter RT and Eaton-Rye JJ (2005) Investigation of a requirement for the PsbP-like protein in *Synechocystis* sp. PCC 6803. Photosynth Res 84: 263–268

Suorsa M, Sirpio S, Allahverdiyeva Y, Paakkarinen V, Mamedov F, Styring S and Aro EM (2006) PsbR, a missing link in the assembly of the oxygen-evolving complex of plant photosystem II. J Biol Chem 281: 145–150

Taguchi F, Yamamoto Y and Satoh K (1995) Recognition of the structure around the site of cleavage by the carboxyl terminal processing protease for D1 precursor protein of the photosystem II reaction center. J Biol Chem 270: 10711–10716

Takahashi T, Inoue-Kashino N, Ozawa S, Takahashi Y, Kashino Y and Satoh K (2009) Photosystem II complex in vivo is a monomer. J Biol Chem 284: 15598–15606

Thompson LK and Brudvig GW (1988) Cytochrome b_{559} may protect photosystem II from photoinhibition. Biochemistry 27: 6653–6658

Thornton LE, Ohkawa H, Roose JL, Kashino Y, Keren N and Pakrasi HB (2004) Homologs of plant PsbP and PsbQ proteins are necessary for regulation of photosystem ii activity in the cyanobacterium *Synechocystis* 6803. Plant Cell 16: 2164–2175

Trost JT, Chisholm DA, Jordan DB and Diner BA (1997) The D1 C-terminal processing protease of photosystem II from *Scenedesmus obliquus*. Protein purification and gene characterization in wild type and processing mutants. J Biol Chem 272: 20348–20356

Umena Y, Kawakami K, Shen JR and Kamiya N (2011) Crystal structure of oxygen evolving photosystem II at a resolution of 1.9 angstrom. Nature 473 (7345): 55–60

Vass I and Cser K (2009) Janus-faced charge recombinations in photosystem II photoinhibition. Trends Plant Sci 14: 200–205

Vass I, Styring S, Hundal T, Koivuniemi A, Aro E and Andersson B (1992) Reversible and irreversible intermediates during photoinhibition of photosystem II: stable reduced Q_A species promote chlorophyll triplet formation. Proc Natl Acad Sci USA 89: 1408–1412

Vermaas WFJ (1994) Evolution of Heliobacteria-Implications for photosynthetic reaction center complexes. Photosyn Res 41: 285–294

Watanabe M, Iwai M, Narikawa R and Ikeuchi M (2009) Is the Photosystem II complex a monomer or a dimer? Plant Cell Physiol 50: 1674–1680

Williams JC, Steiner LA, Feher G and Simon MI (1984) Primary structure of the L-subunit of the reaction center from *Rhodopseudomonas sphaeroides*. Proc Natl Acad Sci USA 81: 7303–7307

Williamson AK (2008) Structural and functional aspects of the MSP (PsbO) and study of its differences in thermophilic versus mesophilic organisms. Photosynth Res 98: 365–389

Wood PM (1988) The potential diagram for oxygen at pH 7. Biochem J 253: 287–289

Youvan DC, Bylina EJ, Alberti M, Begusch H and Hearst JE (1984) Nucleotide and deduced polypeptide sequences of the photosynthetic reaction-center, B870 antenna, and flanking polypeptides from *R. capsulata*. Cell 37: 949–957

Zhaxybayeva O, Doolittle WF, Papke RT and Gogarten JP (2009) Intertwined evolutionary histories of marine *Synechococcus* and *Prochlorococcus marinus*. Genome Biol Evol 2009: 325–339

Zouni A, Witt HT, Kern J, Fromme P, Krauss N, Saenger W and Orth P (2001) Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8 Å resolution. Nature 409: 739–743

Chapter 14

Bioinformatic Identification and Structural Characterization of a New Carboxysome Shell Protein

Fei Cai and Cheryl A. Kerfeld*
US Department of Energy, Joint Genome Institute, Walnut Creek, CA 94598, USA;
Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA

Gustaf Sandh
US Department of Energy, Joint Genome Institute, Walnut Creek, CA 94598, USA

Summary

Bacterial Microcompartments (BMCs) are organelles composed of a polyhedral protein shell that encapsulates metabolically related enzymes. The best characterized BMC, the carboxysome, which functions to enhance CO_2 fixation by D-ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO), is found in all cyanobacteria. It is an essential part of the cyanobacterial CO_2 concentrating mechanism. The shell of BMCs is composed of small (~100 amino acids) proteins with a conserved primary structure known as the BMC domain. Proteins that contain BMC domains were shown to form hexamers that assemble in layers to form the facets of BMC shells. Previous structural models of the carboxysome shell were built from proteins which contain a single BMC domain. Recently, a new carboxysome shell protein was detected bioinformatically in *Prochlorococcus* and *Synechococcus* species. The crystal structure of this protein, CsoS1D, unexpectedly was the first tandem BMC domain protein structurally characterized. These data, together with transcriptomic evidence suggested that CsoS1D is a novel alpha-carboxysome shell protein with functionally important features. Here we used bioinformatic and

*Author for correspondence, e-mail: ckerfeld@lbl.gov

R.L. Burnap and W.F.J. Vermaas (eds.), Functional Genomics and Evolution of Photosynthetic Systems, Advances in Photosynthesis and Respiration 33, pp. 345–356, DOI 10.1007/978-94-007-1533-2_14,

comparative structural modeling to show that a hypothetical protein found in all beta cyanobacterial genomes is the ortholog of CsoS1D. We also discuss observations of other tandem BMC domain proteins, and we propose the hypothesis that the carboxysome shell may be a dynamic structure that responds to the environmental conditions within the cell.

I. Introduction

Cyanobacteria and many chemoautotrophs contain a polyhedral organelle for CO_2 fixation known as the carboxysome (Fig. 14.1). Carboxysomes were first described, in 1969, in transmission electron micrographs of cyanobacteria (Gantt and Conti 1969). They were later shown to be composed entirely of protein: a thin protein shell encapsulates most, if not all of the cellular complement of ribulose1,5-bisphosphate carboxylase/oxygenase (RubisCO) (Shively et al. 1973). RubisCO does not have a particularly high affinity for its substrate, CO_2, nor is it a kinetically "fast" enzyme, and its efficiency is further compromised because it catalyzes the apparently unproductive fixation of O_2 in competition with CO_2. Several structural and functional properties of carboxysomes compensate for these shortcomings (Fig. 14.2). CO_2 is concentrated near the active site of RuBisCO through the diffusion of HCO_3^- into the carboxysome and its subsequent conversion by the carboxysomal carbonic anhydrase into CO_2. Further, maintenance of a high local concentration of CO_2 is possible because the shell limits leakage of CO_2 from the carboxysome interior and, possibly, provides a barrier to oxygen so as to minimize the wasteful oxygenation reaction that competes with carboxylation of RuBP.

Comparative genomic analysis indicates that two different types of carboxysomes are present in cyanobacteria (Badger and Price 2003; Badger et al. 2006). The carboxysomes of the dominant open ocean cyanobacteria, *Synechococcus* and *Prochlorococcus,* contain *cso*- or alpha-type carboxysomes that are encoded by an operon (Fig. 14.3a). In contrast, *Synechococcus elongatus* PCC 7942, *Synechocystis* sp. PCC 6803 and many other cyanobacteria contain *ccm*- or beta-type carboxysomes, which are encoded by several small, scattered gene clusters that do not obviously share a common promoter (Fig. 14.3b).

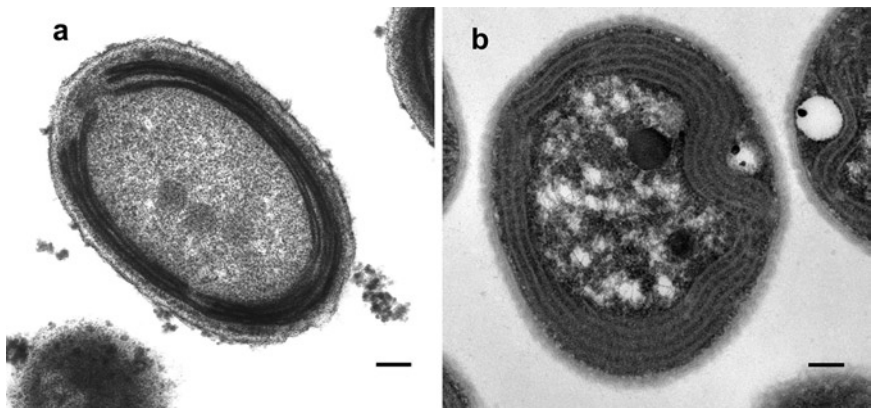


Fig. 14.1. Transmission electron micrographs of negatively stained thin sections of (**a**) *Prochlorococcus marinus* str. MIT9313 and (**b**) *Synechocystis* sp. PCC6803 containing three and one carboxysomes, respectively. Bars represent 0.1 μm (**a**) and 0.2 μm (**b**). Transmission electron micrograph courtesy of Claire Ting, Department of Biology, Williams College (**a**) and Patrick Shih and Cheryl Kerfeld, University of California, Berkeley (**b**).

Abbreviations: BMC – bacterial microcompartment; CA – carbonic anhydrase; CCM – CO_2 concentrating mechanism; HL – high light; HMM – hidden Markov model; LL – low light; MSA – multiple sequence alignment; MV – methyl viologen; PG – phosphoglycolate; 3-PGA – 3-phosphoglyceric acid; RMSD – root mean square deviation; RubisCO – ribulose 1,5-bisphosphate carboxylase/oxygenase; RuBP – ribulose 1,5-bisphosphate

Alpha- and beta- carboxysomes are distinct with respect to the form of RubisCO (Form 1A or 1B, respectively), and the type of carbonic anhydrases (CsoSCA or CcmM and CcaA, respectively) they contain. In addition there is no sequence homology between CsoS2 and CcmN, two other components of the carboxysome whose function is currently unknown.

In contrast, the major components of both the alpha- and beta-carboxysome shell are recognizably similar, and contain a domain now known as the Bacterial Micro Compartment domain (BMC; Pfam00936). Structural studies have shown that these small proteins (93–115 amino acids) assemble into hexamers that form the facets of the carboxysome shell (Kerfeld et al. 2005; Tsai et al. 2007). Absolutely conserved residues converging at the six-fold axis of symmetry form narrow (~4–7 Å), positively charged pores (Fig. 14.4) that appear to provide conduits for metabolite flow across the shell (Kerfeld et al. 2005; Tsai et al. 2007).

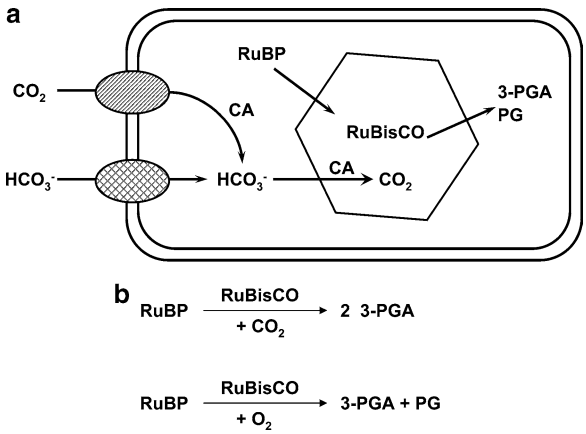


Fig. 14.2. (**a**) A schematic model of the cyanobacterial CO_2 concentrating mechanism (*CCM*). The accumulation of cytoplasmic HCO_3^- is achieved by the CO_2 and HCO_3^- uptake systems in the cell membrane (*shown as ovals*) and cytoplasmic carbonic anhydrase (*CA*). A carboxysomal CA functions to convert HCO_3^- to CO_2 within the carboxysome, which results in a localized high CO_2 concentration around the encapsulated RubisCO. (**b**) RubisCO uses either CO_2 or O_2 as substrate. (*PG* phosphoglycolate, *3-PGA* 3- phosphoglyceric acid, *RubisCO* ribulose1,5-bisphosphate carboxylase/oxygenase, *RuBP* ribulose 1,5-bisphosphate).

In addition, carboxysome gene clusters encode another small protein, CsoS4A/B and CcmL (Pfam 03319; Fig. 14.3). These proteins are now known to form pentamers providing the requisite vertices for modeling the carboxysome as an icosahedron (Tanaka et al. 2008). However, it's been shown that the deletion of both *csoS4A* and *csoS4B* does not seem to affect carboxysome morphology (Cai et al. 2009).

In the 1990s it was discovered that some enteric bacteria contain genes encoding BMC domain proteins (Shively et al. 1998; Bobik et al. 1999; Kofoid et al. 1999); these are clustered with Pfam 03319 domain-containing genes and genes encoding enzymes for the B_{12}-dependent degradation of 1,2-propanediol (*pdu* operon) or ethanolamine (*eut* operon). Furthermore, bioinformatic analysis of microbial genome sequences shows that the potential to form bacterial microcompartments is widespread (Fig. 14.5); BMC-domain containing gene clusters have been identified in nearly 600 bacterial genomes. The precise function for most of these BMCs is unknown, but there appears to be at least ten distinct types (Kerfeld et al. 2010) based on the putative enzymatic function of the genes clustered with these shell proteins. For example, *Rhodospirillium rubrum* and *Rhodopseudomonas*

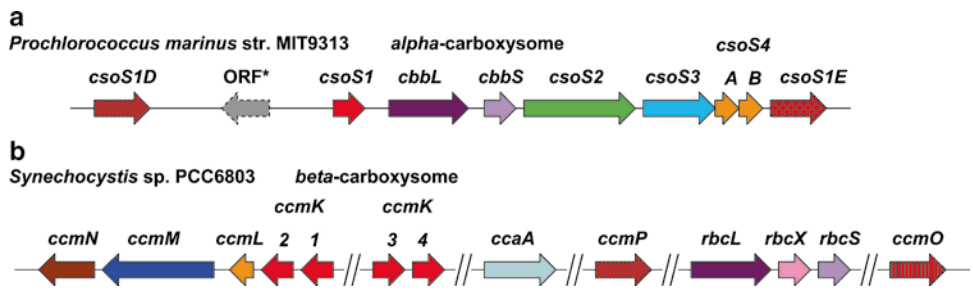


Fig. 14.3. Examples of genomic organization of the alpha- or beta-carboxysome gene cluster(s) in cyanobacteria, with Pfam00936 (BMC domain) genes shown in red and Pfam03319 genes in orange. Tandem BMC-domain proteins have hatching. (**a**) *Prochlorococcus marinus* str. MIT9313; (**b**) *Synechocystis* sp. PCC 6803.

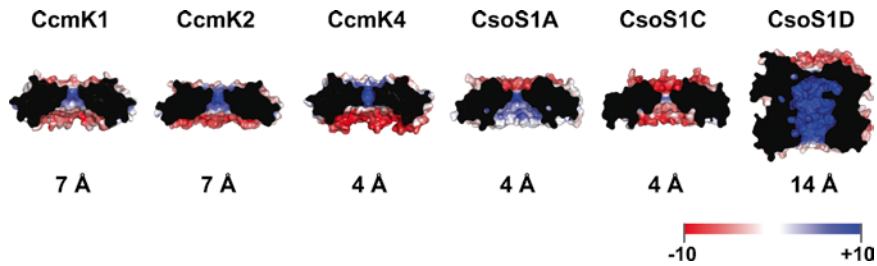


Fig. 14.4. Carboxysomal BMC domain shell protein hexamers visualized from the side, slabbed to show a cross-section of the pore and colored by electrostatic potential (*blue, positive; red negative*). The diameter of the pore is given below each shell protein. The CsoS1D dimer of trimers is shown with the closed trimer uppermost. Figures prepared with GRASP2 (Petrey and Honig 2003) using PDB codes: 3BN4 (CcmK1), 2A1B (CcmK2), 2A10 (CcmK4), 2A13 (CsoS1A), 3H8Y (CsoS1C), and 3F56 (CsoS1D).

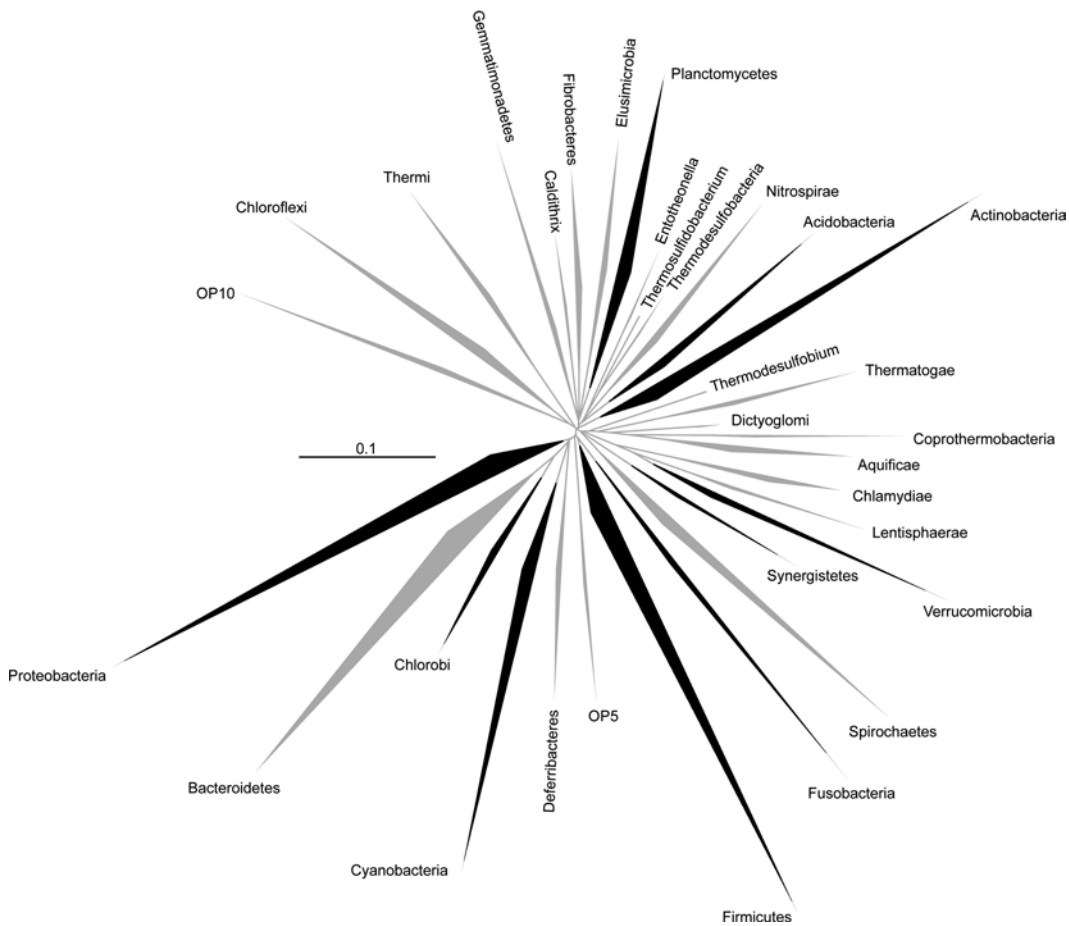


Fig. 14.5. Occurrence of bacterial microcompartment gene clusters among bacterial phyla. An un-rooted maximum-likelihood tree of bacterial phyla, in which the wedge size reflects the number representatives with genomes sequenced. Phyla with members containing BMC gene clusters are shown in black (Phylum tree adapted by Philip Hugenholtz from Herleman et al., 2009).

palustris strain BisB18 contain a BMC gene cluster that encodes a putative B_{12}-independent diol dehydratase, possibly for 1,2-propanediol metabolism (Kerfeld et al. 2010). This reaction involves a toxic intermediate that must be sequestered within the cell. The presence of the bacterial microcompartment operon in *Rps. palustris* BisB18 but not in other closely related *Rps. palustris* strains,

also exemplifies another trend discernible from comparative genomic analysis of BMC gene clusters: they appear to be frequently horizontally transferred (Kerfeld et al. 2010).

II. CsoS1D: A Novel Alpha-Carboxysome Shell Protein

A. Detection of the CsoS1D Gene

Comparative analyses of BMC gene clusters, including carboxysomes, across bacterial genomes can be used to identify some characteristic features of their composition that are presumably important for their assembly, structure and function. All BMC gene clusters contain at least one copy of a gene for the Pfam03319 domain and at least two copies of a Pfam00936 (BMC) domain. The reason for this redundancy in number of BMC domains is unknown. Nearly all BMC clusters encode at least one tandem BMC domain-containing protein, like *csoS1E*, the last gene in the alpha-carboxysome operon of low light adapted *Prochlorococcus* strains (Fig. 14.3a), and *ccmO* of the beta-carboxysome (Fig. 14.3b). Only the carboxysome operon of *Prochlorococcus* sp. MED4, and other high-light adapted *Prochlorococcus* strains (MIT9301, MIT9215, MIT9311, MIT 9515) appeared to be an exception to the rule that there is always more than one BMC domain present per genome. The absence of a second annotated BMC domain-containing gene in the high-light adapted *Prochlorococcus* carboxysome operon prompted us to look more closely for additional BMC-domain containing genes in this group of organisms. A gene encoding an ~250 amino acid protein that was predicted by sequence homology to contain a BMC domain in its C-terminal half was identified upstream, separated by one ORF encoded on the opposite strand, from the rest of the carboxysome operon (Fig. 14.3a). This gene, now known as *csoS1D* (Klein et al. 2009), is found in all of cyanobacteria that contain alpha-carboxysomes and in nearly all of the carboxysome-containing chemoautotrophs.[1] The primary structure of the *csoS1D* gene product in cyanobacteria (referred as CsoS1D-cyano henceforth) is highly conserved (88% identical) among marine *Synechococcus* and *Prochlorococcus* strains. Notably, the first 50 amino acids of CsoS1D-cyano are missing among all CsoS1D orthologs found in chemoautotrophs (referred as CsoS1D-chemo henceforth) (Fig. 14.6). This absence suggests the N-terminal ~50 amino acids of CsoS1D-cyano must be specific to metabolic needs of alpha-cyanobacteria (Fig. 14.6). In addition to partial sequence homology of CsoS1D to known carboxysome shell proteins, expression and structural data (discussed below) support a role for it in the carboxysome (Klein et al. 2009). The lack of continuity of its gene with the rest of the *cso* operon may reflect its differential regulation.

B. Structural Genomics: CsoS1D is a Component of the Alpha-Carboxysome

Structural analysis is an often overlooked method of assigning function to a protein, however there are numerous examples of its successful deployment; classic examples include Hwang et al. (1999), Watson et al. (2006), Allen (2007), and Hermann et al. (2007). To test the hypothesis that CsoS1D was a component of the carboxysome, Klein et al. (2009) solved the crystal structure of CsoS1D from *Prochlorococcus* strain MED4 (PDB codes 3F56 and 3FCH, Fig. 14.7). Surprisingly, CsoS1D proved to be the first structurally characterized tandem BMC domain protein: residues 50–150 which lack significant homology to any other known BMC domain also contained a BMC fold. Despite the low sequence homology (18% identity) between two domains (N-BMC and C-BMC), the alpha-carbon can be superimposed on each other with an RMSD of 1.27 Å. Interestingly, CsoS1D forms a trimer, which recapitulates the hexameric shape of single BMC domain proteins and suggested that the CsoS1D trimer would readily fit into existing models for the facets of the carboxysome shell (Fig. 14.7c). Moreover, a lysine residue (Lys108 in the N-BMC and Lys212 in the C-BMC; Fig. 14.6), which is structurally conserved in previously characterized BMC structures and is proposed to be essential for the interaction between two adjacent hexamers of single BMC domain proteins (Klein et al. 2009), is also observed at the edges of CsoS1D pseudohexamer.

Another unique feature of the CsoS1D pseudohexamer is the gated pore at its threefold symmetry

[1]The exceptions are *Bradyrhizobium* sp. BTAi1 and three *Acidithiobacillus ferrooxidans* strains (ATCC 53993, ATCC 23270, and DSM 10331) sequenced to-date.

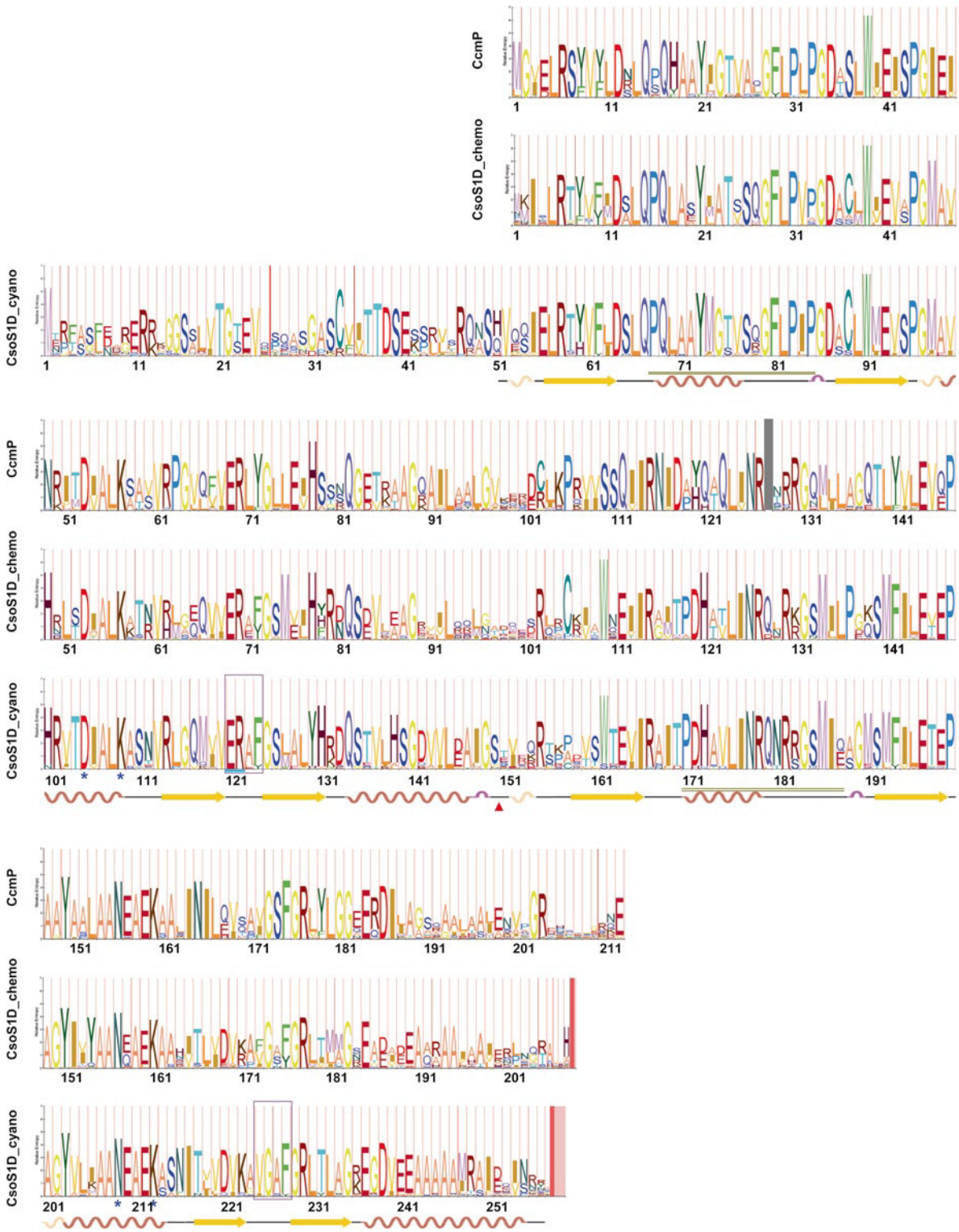


Fig. 14.6. HMM sequence logos of CcmP, CsoS1D of chemoautotrophs (CsoS1D-chemo), and CsoS1D of alpha- cyanobacteria (CsoS1D-cyano), and the secondary structure of CsoS1D. The multiple sequence alignment (*MSA*) was built with MUSCLE (Robert 2004), using 28 CcmP, 12 CsoS1D-chemo, and 27 CsoS1D-cyano orthologs. HMMs were generated and visualized by HMMBUILD (Eddy 1998, 2008) and LogoMat-M (Schuster-Böckler, et al. 2004). A red arrow head marks the separation of the N-BMC and C-BMC domains according to the crystal structure of CsoS1D. Residues of the loops that gate the pore are boxed, and the conserved Glu120 and Arg121 residues are underlined. The conserved residues that are proposed to be important for interaction of adjacent shell proteins in the shell (Asp104, Lys 108, Asn208, and Lys 212) are marked with asterisks. Residues involved in the trimer-trimer interface are double underlined. One amino acid residue deletion is found in the canonical sequence of CcmP (*shown as a gray block*), comparing to its ortholog CsoS1D.

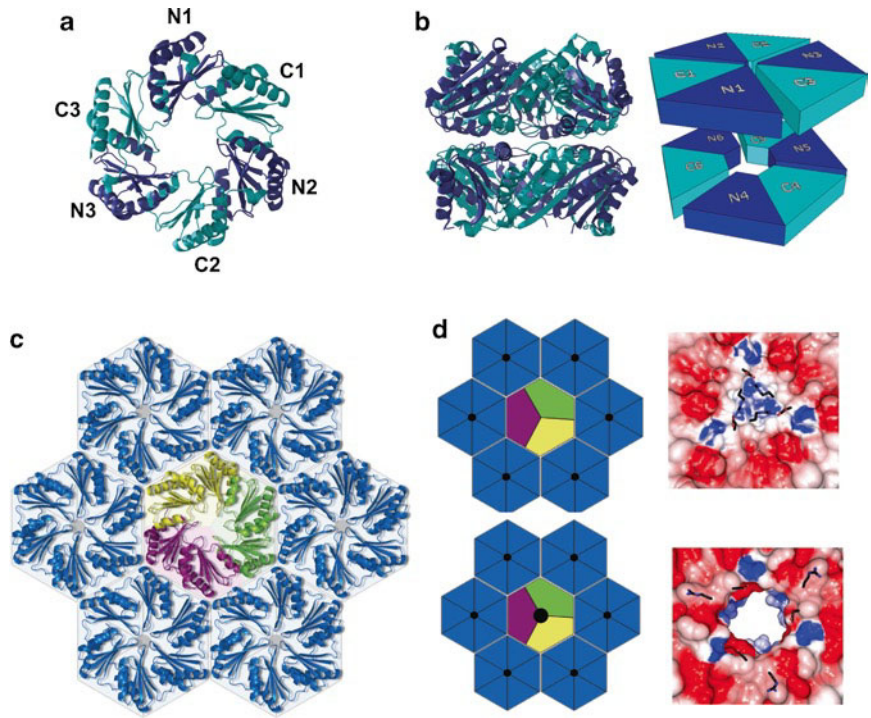


Fig. 14.7. (**a**) Tandem BMC-domain protein CsoS1D forms pseudohexamer. N-BMC and C-BMC are colored in blue and cyan, respectively. (**b**) Face-to-face dimerization of CsoS1D trimers. N-BMC and C-BMC are colored as in (**a**), and the schematic model of this architecture reveals that two trimers are offset by a ~60º rotation. (**c**) The three-dimensional model of CsoS1D trimer (individual monomers are colored with *yellow*, *green*, and *purple*) interacting with CsoS1 hexamers (*blue*). (**d**) Simplified CsoS1D-CsoS1 interaction model (*left*) and a close-up view of the gated pore of the CsoS1D trimer (*right*). The electrostatic potential is colored as in Fig. 14.4, and the conserved residues (Glu120 and Arg121) are shown as sticks. The sizes of pores are scaled relatively. Figures of crystal structure are prepared with PyMol (DeLano 2002).

axis (Fig. 14.7d). This pore can be either "open" (~14 Å in diameter) or "closed" by the side chains of residues Glu120 and Arg121 of the N-BMC domain, which are also absolutely conserved residues among all identified CsoS1D orthologs (Fig. 14.6). This unusually large, gated pore was the first evidence that conformational changes in shell proteins could regulate flux of metabolites across a BMC shell. Intriguingly, two CsoS1D crystal structures, determined from unrelated crystallization conditions, both showed a face-to-face dimerization of one open and one closed trimer. This results in a large (13,613 $Å^3$) cavity between the two trimers (Fig. 14.4). The non-polar nature of the trimer-trimer interface, the amount of buried surface area, and the shape complementarity of the interface between two trimers all suggest this dimerization of Csos1D trimers is physiologically relevant (Klein et al. 2009).

The CsoS1D structure provides a solution to a problem implicit in all previous models of the carboxysome shell: how did the shell allow the larger metabolites (ex. RuBP, 3-PGA) to cross while providing a barrier to the loss of smaller metabolites? A plausible model for flux of larger metabolites such as RuBP across the shell could involve CsoS1D functioning analogously to an air-lock. RuBP could pass through an open pore of the CsoS1D stacked trimers, reside in the cavity created by the dimerization of the two trimers while the cytosolic-facing pore closes and the pore facing the interior of the carboxysome opens, permitting the RuBP to pass from the central cavity to the interior of the carboxysome.

III. Identification of the CsoS1D Ortholog in Beta-Carboxysomes: CcmP

A. A CsoS1D Ortholog is Found in All Beta-Cyanobacteria

A *csoS1D* ortholog (best bi-directional BLAST hit; Fig. 14.6) is also found in all genomes of beta-carboxysome containing cyanobacteria sequenced to-date. For example, in *Synechocystis* PCC 6803 it is *slr0169*, annotated as a hypothetical protein. Like other known beta-carboxysome genes (e.g. *ccmK3* and *ccmK4*), this conserved gene is not encoded by the main cluster of carboxysome genes (*ccmK-ccmN*) in the genome (Fig. 14.3b). The beta-carboxysome CsoS1D ortholog is ~200 amino acids long, similar to CsoS1D-chemo, and is predicted to have both N- and C-BMC domains. Sequence comparison using the method of Hidden Markov Models (HMMs) (Fig. 14.6) suggests that the gated pore is also present in the presumptive gene product: in this hypothetical protein residues Glu69 and Arg70, and residues Lys57 and Lys160 are the counterparts of gating residues and conserved lysine residues (Lys108 and Lys212) of CsoS1D, respectively. Collectively, these features as well as the ability to model a trimer that would fit into a layer of CcmK hexamers (not shown), suggest that this hypothetical protein is like CsoS1D, a carboxysome component, and we propose it be named CcmP.

B. Predicted Similarities and Differences Between CcmP and CsoS1D

A notable difference between CsoS1D and a model of CcmP based on the CsoS1D structure is revealed when comparing the relative levels of sequence conservation mapped onto the tertiary structure. Residues of CsoS1D-cyano fall into extremes (Fig. 14.8a), they are either very strongly conserved or poorly conserved, while the conservation of predicted CcmP structure is more gradual (Fig. 14.8b). By using ConSurf (Glaser et al. 2003) to quantitate conservation we examined whether CcmP is likely to form dimers of trimers, like CsoS1D. The average ConSurf scores (which range between 1 and 9; with 9 being the most conserved and 1 being the least) of the residues at the trimer-trimer interface and of the overall structure of CsoS1D-cyano are 6.5 and 6.2, respectively. However, the corresponding scores for CcmP are 5.4 and 5.7, which suggests the region of the putative trimer-trimer interface (residues 16–33 and 119–135) is relatively less conserved than the overall primary structure of CcmP.

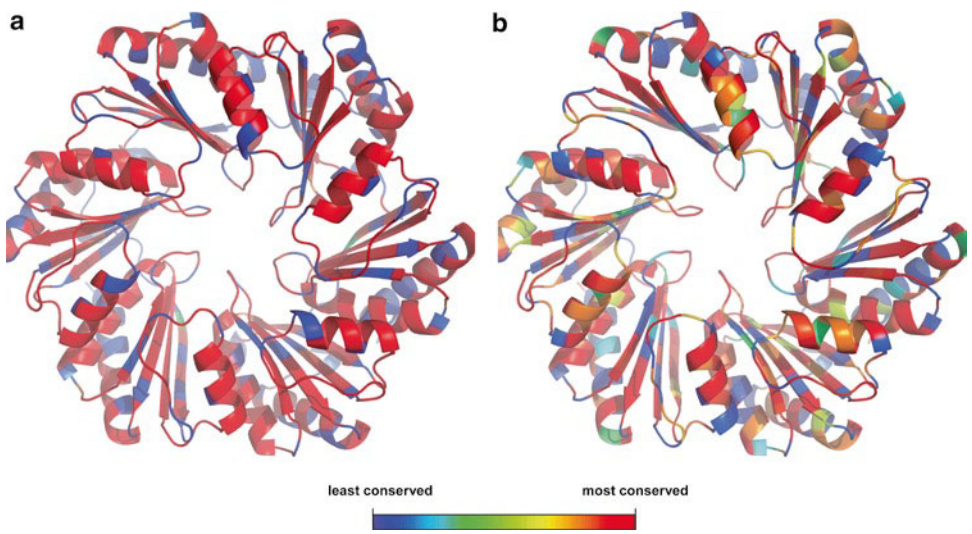


Fig. 14.8. Sequence conservation mapped on the CsoS1D structure (**a**) and predicted CcmP structure (**b**) The CcmP structure was generated using Swiss-Model (Peitsch 1995; Arnold et al. 2006; Kiefer et al. 2009). Conservation scores are calculated with ConSurf (Glaser et al. 2003) using MSA files prepared for Fig. 14.6; figures are prepared with PyMol. Each amino-acid position is assigned a grade from 1 (most variable) to 9 (most conserved) based on their evolutionary rates and then colored according to the grade with blue variable and red conserved.

Therefore, whether or not CcmP trimers dimerize is a particularly interesting question to be addressed with X-ray crystallographic studies.

C. Expression Data Provides Complementary Evidence to Structural Studies for a Role for CcmP in the Carboxysome

The functional role of *csoS1D* in the α-carboxysome is supported by its co-expression with the genes of the *cso* operon over the diurnal light/dark cycle in *Prochlorococcus* (Klein et al. 2009). Interestingly, a similar pattern can be detected in the beta-carboxysome containing cyanobacterium *Synechococcus elongatus* PCC 7942, in which all of the structural components of the carboxysome (including *ccmP*), *ccmM* & *N*, the large and small subunit of RubisCO (*rbcLS*) and the carboxysome-specific carbonic anhydrase, *ccaA*, co-express over the diurnal light/dark cycle (Ito et al. 2009; Vijayan et al. 2009). In both experiments the expression of *ccmP* and the other carboxysome genes peaked around dawn together with many genes involved in the light reactions of photosynthesis. In addition, the diurnally oscillating expression pattern for *ccmK-P*, *rbcLS* and *ccaA* persisted in continuous light, hence fulfilling two of the criteria for being under the control of the cyanobacterial circadian clock (Dong and Golden 2008). This tight diurnal co-regulation of *ccmP* with other carboxysome genes as well as most photosynthetic genes substantiates that it plays a role in photosynthesis.

To investigate the robustness of the transcriptional co-regulation of all of the *ccm* genes and their correlation with *ccmP,* we examined their expression patterns in six available microarray data sets (Fig. 14.9). Under limited inorganic carbon concentrations, all of the carboxysome related genes investigated, with the exception of *ccmO*, were down-regulated over time (Fig. 14.9a and b; Wang et al. 2004; Eisenhut et al. 2007). However, Eisenhut et al. (2007) showed that the down regulation on the transcriptional level does not correspond to lowered protein levels; a constant amount of CcmK proteins (all paralogs) was present throughout the experiment. Moreover, an almost twofold increase in the number of carboxysomes was detected in cells grown under lower CO_2 concentrations. Using the complete time series provided in Wang et al. (2004) we analyzed the expression behavior of *ccmP* using the Mfuzz "fuzzy" clustering software (Futschik and Carlisle 2005; Kumar and Futschik 2007). In these data, independent of the physiological effect of carbon limitation, *ccmP* clearly behaves as a *ccm* gene, clustering with *ccmM* and *ccmN*.

In contrast to the response to reduced CO_2 concentrations increased light intensity (Fig. 14.9d; Hihara et al. 2001) stimulated the transcription of all carboxysome related genes investigated. A lower induction was however evident in *ccaA*, *ccmL* and all the *ccm* genes encoded for outside of the *ccm* operon (e.g. *ccmK4* and *ccmP*). The separation in expression pattern between *ccmP* and the *ccm* operon was also clear when performing fuzzy clustering, as above, of the whole time course experiment provided by Hihara et al. (2001). While the entire *ccm* operon and the RubisCO subunits (*rbcLS*) cluster strongly together *ccmP* and the other *ccm* genes do not. This pattern is very similar to the response to nitrogen limitation (Fig. 14.9c; Osanai et al. 2006) where *rbcLS* and the *ccm* operon is strongly down regulated while the other *ccm* genes are almost unchanged.

Induction of oxidative stress in low light adapted cells (Fig. 14.9e; Kobayashi et al. 2004) resulted in an up regulation of all carboxysome related genes. A different response was triggered when oxidative stress was induced in high light acclimated cells. The two subunits of RubisCO (*rbcLS*) and the members of the *ccm* operon were strongly repressed while the rest shown only a slight or no change (Fig. 14.9f; Kobayashi et al. 2004). Kobayashi et al. (2004) also showed that the expression of the *ccm* operon is possibly regulated by a Fur-type transcriptional regulator (*slr1738*) under stress conditions as a *slr1738* mutant gave increased transcription of *ccmL*, *ccmM* and *ccmN*.

In summary the transcriptional regulation of *ccmP* like that of *csoS1D* follows that of all other *ccm* genes over the diurnal light/dark cycle. However regulatory differences are seen between *ccm* operon encoded genes and genes transcribed outside of the operon (i.e. *ccmK3*, *ccmK4*, *ccmO* and *ccmP*) in response various environmental conditions (e.g. nitrogen limitation and oxidative stress under HL). Differential regulation of the different subunits of the carboxysome shell could be one way to alter the composition of the carboxysome shell in response to the cells physiological needs.

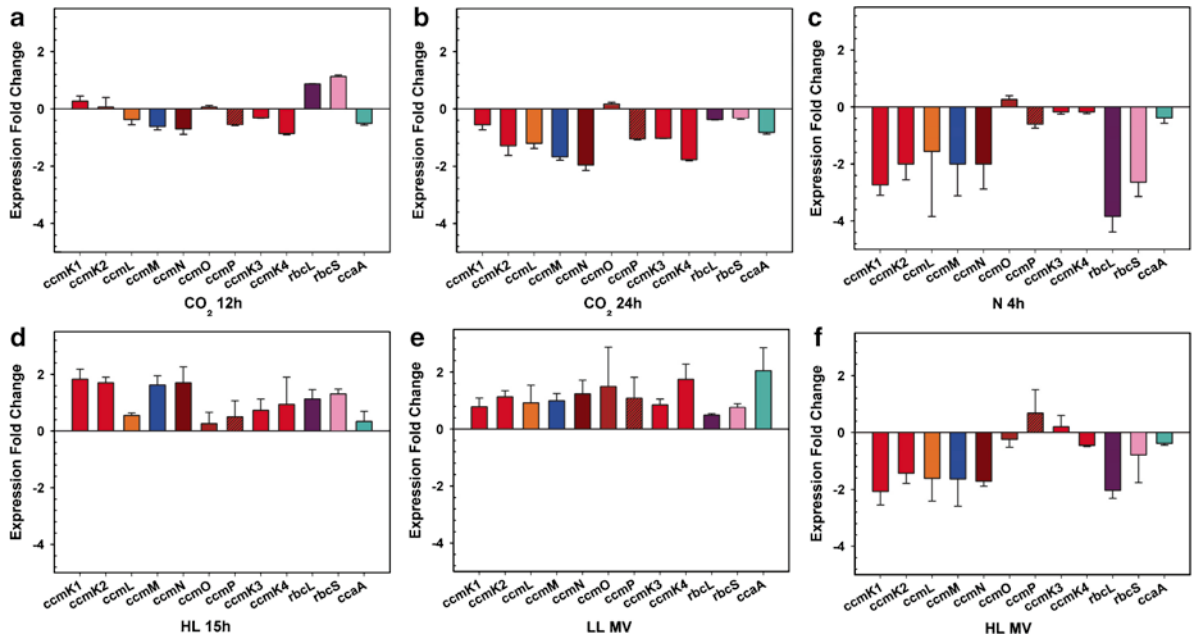


Fig. 14.9. Transcriptional regulation of carboxysome related genes taken from published microarray data for *Synechocystis* PCC 6803 under a range of physiological conditions. (**a**) 12 h after switching cells from 3% CO_2 to air (Wang et al. 2004); (**b**) 24 h after switching cells from 5% CO_2 to air (Eisenhut et al. 2007); (**c**) 4 h after removing combined nitrogen from the growth media (Osanai et al. 2006); (**d**) 15 h after switching cells from 20 to 300 μE (*HL* = high light) (Hihara et al. 2001); (**e**) 15 min after addition of methyl viologen (*MV*) to low light adapted cells (Kobayashi et al. 2004); (**f**) 15 min after addition of methyl viologen (*MV*) to high light adapted cells (Kobayashi et al. 2004). All values are given as log2 ratios of the given treatment compared to the expression at time 0. The locus tags for the genes shown are *sll1029* (*ccmK1*), *sll1028* (*ccmK2*), *sll1030* (*ccmL*), *sll1031* (*ccmM*), *sll1032* (*ccmN*), *slr0436* (*ccmO*), *slr0169* (*ccmP*), *slr1838* (*ccmK3*), *slr1839* (*ccmK4*), *slr0009* (*rbcL*), *slr0012* (*rbcS*), *slr1347* (*ccaA*). Each individual gene is colored as in Fig. 14.3b.

Changing the composition of the carboxysome shell, using different shell proteins with different types of pores (CcmP in particular) could alter metabolic fluxes across the carboxysome shell.

IV. Prospects for Understanding the Function of Tandem Bacterial Microcompartment Domain Proteins in Carboxysome Physiology

Since the publication of the CsoS1D structure (Klein et al. 2009), the structures of two additional tandem BMC domain proteins have been reported (Heldt et al. 2009; Sagermann et al. 2009; Tanaka et al. 2010). Unlike CsoS1D, these proteins were predicted by sequence homology to be tandem-BMC domain proteins. Like CsoS1D, each by trimerization forms a pseudohexamer. EtuB and EutL likewise provide examples of the increased structural and functional versatility that results from the fusion of two BMC domains: in both EtuB (Heldt et al. 2009) and the so-called closed (Tanaka et al. 2010) conformation of EutL, a pore is formed between the two domains in the monomer, resulting in three pores in the pseudohexamer. Because the pores are not formed at a symmetry axis, the resulting channels are asymmetric; this may confer additional selectivity through the channel. Moreover, a second structure of EutL (Tanaka et al. 2010) demonstrated that EutL, like CsoS1D, has an open form, in which one large central pore (triangular in shape, with an edge of approximately 11 Å) presumably to permit the passage of larger metabolites across the Eut BMC shell.

Among the carboxysome shell proteins, the structures of the predicted tandem domain shell proteins CcmO and CcmP remain to be elucidated. A comparison of these structures, in conjunction with analysis of expression data will provide insight into the specific roles of tandem BMC domain proteins in the carboxysome shell

and provide a test of the hypothesis that the shell may be a dynamic structure that responds to the environmental conditions within the cell.

Acknowledgments

We thank Gordon C. Cannon, Sabine Heinhorst and Claire Ting for helpful discussions, Seth Axen for preparing Fig. 14.4, Phil Hugenholtz for providing a template for Fig. 14.5 and Edwin Kim for preparing Fig. 14.7c and d.

FC and CAK's current carboxysome research is supported by NSF grant and MCB-085170. The work of CAK and GS is performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract number DE-AC02-05CH11231, and Lawrence Livermore National Laboratory under contract number DE-AC52-07NA27344.

References

Allen KN (2007) Form finds function. Nat Chem Biol 3: 452–453

Arnold K, Bordoli L, Kopp J and Schwede T (2006) The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics 22: 195–201

Badger MR and Price GD (2003) CO_2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. J Exp Bot 54: 609–622

Badger MR, Price GD, Long BM and Woodger FJ (2006) The environmental plasticity and ecological genomics of the cyanobacterial CO_2 concentrating mechanism. J Exp Bot 57: 249–265

Bobik TA, Havemann GD, Busch RJ, Williams DS and Aldrich HC (1999) The propanediol utilization (*pdu*) operon of *Salmonella enterica* serovar Typhimurium LT2 includes genes necessary for formation of polyhedral organelles involved in coenzyme B_{12}-dependent 1,2-propanediol degradation. J Bacteriol 181: 5967–5975

Cai F, Menon BB, Cannon GC, Curry KJ, Shively JM and Heinhorst S (2009) The pentameric vertex proteins are necessary for the icosahedral carboxysome shell to function as a CO_2 leakage barrier. PLoS One 4: e7521

DeLano WL (2002) DeLano Scientific Palo Alto, CA

Dong G and Golden SS (2008) How a cyanobacterium tells time. Curr Opin Microbiol 11: 541–546

Eddy SR (1998) Profile hidden Markov models. Bioinformatics 14: 755–763

Eddy SR (2008) A probabilistic model of local sequence alignment that simplifies statistical significance estimation. PLoS Comput Biol 4: e1000069

Eisenhut M, von Wobeser EA, Jonas L, Schubert H, Ibelings BW, Bauwe H, Matthijs HCP and Hagemann M (2007) Long-term response toward inorganic carbon limitation in wild type and glycolate turnover mutants of the cyanobacterium *Synechocystis* sp. strain PCC 6803. Plant Physiology 144: 1946–1959

Futschik ME and Carlisle B (2005) Noise-robust soft clustering of gene expression time-course data. J Bioinform Comput Biol 3: 965–988

Gantt E and Conti SF (1969) Ultrastructure of blue-green algae. J Bacteriol 97 (3): 1486–1493

Glaser F, Pupko T, Paz I, Bell RE, Bechor D, Martz E and Ben-Tal N (2003) ConSurf: Identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics 19: 163–164

Heldt D, Frank S, Seyedarabi A, Ladikis D, Parsons JB, Warren MJ and Pickersgill RW (2009) Structure of a trimeric bacterial microcompartment shell protein, EtuB, associated with ethanol utilization in *Clostridium kluyveri*. Biochem J 423: 199–207

Herlemann DPR, Geissinger O, Ikeda-Ohtsubo W, Kunin V, Sun H, Lapidus A, Hugenholtz P and Brune A (2009) Genomic analysis of *Elusimicrobium minutum*, the first cultivated representative of the phylum *Elusimicrobia* (formerly termite group 1). Appl Environ Microbiol 75: 2841–2849

Hermann JC, Marti-Arbona R, Fedorov AA, Fedorov E, Almo SC, Shoichet BK and Raushel FM (2007) Structure-based activity prediction for an enzyme of unknown function. Nature 448: 775–781

Hihara Y, Kamei A, Kanehisa M, Kaplan A and Ikeuchi M (2001) DNA microarray analysis of cyanobacterial gene expression during acclimation to high light. Plant Cell 13: 793–806

Hwang KY, Chung JH, Kim S, Han YS and Cho Y (1999) Structure-based identification of a novel NTPase from *Methanococcus jannaschii*. Nat Struct Biol 6: 691–696

Ito H, Mutsuda M, Murayama Y, Tomita J, Hosokawa N, Terauchi K, Sugita C, Sutita M, Terauchi K, Sugita C, Sugita M, Kondo T, Iwasaki H (2009) Cyanobacterial daily life with Kai-based circadian and diurnal genome-wide transcriptional control in *Synechococcus elongatus*. Proc Nat Acad Sci USA 106: 14168–14173

Kerfeld CA, Sawaya MR, Tanaka S, Nguyen CV, Phillips M, Beeby M and Yeates TO (2005) Protein structures forming the shell of primitive bacterial organelles. Science 309: 936–938

Kerfeld CA, Heinhorst S and Cannon GC (2010) Bacterial microcompartments. Annu Rev Microbiol 64: 391–408

Kiefer F, Arnold K, Künzli M, Bordoli L and Schwede T (2009) The SWISS-MODEL Repository and associated resources. Nucleic Acids Res 37: D387–D392

Klein MG, Zwart P, Bagby SC, Cai F, Chisholm SW, Heinhorst S, Cannon GC and Kerfeld CA (2009) Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. J Mol Biol 392: 319–333

Kobayashi M, Ishizuka T, Katayama M, Kanehisa M, Bhattacharyya-Pakrasi M, Pakrasi HB and Ikeuchi M (2004) Response to oxidative stress involves a novel peroxiredoxin gene in the unicellular cyanobacterium *Synechocystis* sp. PCC 6803. Plant Cell Physiol 45: 290–299

Kofoid E, Rappleye C, Stojiljkovic I and Roth J (1999) The 17-gene ethanolamine (*eut*) operon of *Salmonella typhimurium* encodes five homologues of carboxysome shell proteins. J Bacterial 181: 5317–5329

Kumar L and Futschik ME (2007) Mfuzz: A software package for soft clustering of microarray data. Bioinformation 2: 5–7

Osanai T, Imamura S, Asayama M, Shirai M, Suzuki I, Murata N and Tanaka K (2006) Nitrogen induction of sugar catabolic gene expression in *Synechocystis* sp. PCC 6803. DNA Res 13: 185–195

Peitsch MC (1995) Protein modeling by E-mail. Nature Biotechnol 13: 658–660

Petrey D and Honig B (2003) GRASP2: visualization, surface properties, and electrostatics of macromolecular structures and sequences. Methods Enzymol 374: 492–509

Robert CE (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797

Sagermann M, Ohtaki A and Nikolakakis K (2009) Crystal structure of the EutL shell protein of the ethanolamine ammonia lyase microcompatment. Proc Natl Acad Sci USA 106: 8883–8887

Schuster-Böckler B, Schultz J and Rahmann S (2004) HMM Logos for visualization of protein families. BMC Bioinformatics 5: 7

Shively JM, Ball F, Brown DH and Saunders RE (1973) Functional organelles in prokaryotes: polyhedral inclusions (carboxysomes) of *Thiobacillus neapolitanus*. Science 182: 584–586

Shively JM, Bradburne CE, Aldrich HC, Bobik TA, Mehlman JL, Jin S and Baker SH (1998) Sequence homologs of the carboxysomal polypeptide CsoS1 of the thiobacilli are present in cyanobacteria and enteric bacteria that form carboxysomes-polyhedral bodies. Can J Bot 76: 906–916

Tanaka S, Kerfeld CA, Sawaya MR, Cai F, Heinhorst S, Cannon GC and Yeates TO (2008) Atomic-level models of the bacterial carboxysome shell. Science 319: 1083–1086

Tanaka S, Sawaya MR and Yeates TO (2010) Structure and mechanisms of a protein-based organelle in *Escherichia coli*. Science 327: 81–84

Tsai Y, Sawaya MR, Cannon GC, Cai F, Williams EB, Heinhorst S, Kerfeld CA and Yeates TO (2007) Structural analysis of CsoS1A and the protein shell of the *Halothiobacillus neapolitanus* carboxysome. PLoS Biology 5: e144

Vijayan V, Zuzow R and O'Shea EK (2009) Oscillations in supercoiling drive circadian gene expression in cyanobacteria. Proc Natl Acad Sci USA 106: 22564–22568

Wang H, Postier BL and Burnap RL (2004) Alterations in global patterns of gene expression in *Synechocystis* sp. PCC 6803 in response to inorganic carbon limitation and the inactivation of *ndhR*, a LysR family regulator. J Biol Chem 279: 5739–5751

Watson JD, Sanderson S, Ezersky A, Savchenko A, Edwards A, Orengo C, Joachimiak A, Laskowshi RA and Thornton JM (2006) Towards fully automated structure-based function prediction in structural genomics: a case study. J Mol Biol 367: 1511–1522

Chapter 15

The Physiology and Functional Genomics of Cyanobacterial Hydrogenases and Approaches Towards Biohydrogen Production

Jens Appel*
Botanisches Institut, Christian-Albrechts-Universität, Am Botanischen Garten 1-9, 24118 Kiel, Germany

Summary

Three different enzymes, the nitrogenase and two different hydrogenases, an uptake and a bidirectional enzyme, are involved in cyanobacterial hydrogen metabolism. In strains containing nitrogenase, H_2 is produced as a byproduct during nitrogen fixation. Many cyanobacterial strains additionally express an uptake hydrogenase that recycles these reducing equivalents. Since not every nitrogen-fixing strain encodes the genes of the uptake hydrogenase, it seems to be dispensable under some environmental conditions. Genome comparisons suggest that the cyanobacterial uptake hydrogenase requires the presence of five additional accessory genes for its maturation.

The primary function of the bidirectional hydrogenase is to increase energetic yield during fermentation and light induced H_2 production in transition states, when cells shift from dark anaerobic conditions to those where light is present. Under oxidizing conditions, the bidirectional enzyme can also catalyze hydrogen uptake. Comparative genomics and database searches reveal the specific association of the pyruvate:ferredoxin/flavodoxin oxidoreductase with the bidirectional hydrogenase, indicating a functional

*Author for correspondence, e-mail: jappel@bot.uni-kiel.de

R.L. Burnap and W.F.J. Vermaas (eds.), Functional Genomics and Evolution of Photosynthetic Systems, Advances in Photosynthesis and Respiration 33, pp. 357–381, DOI 10.1007/978-94-007-1533-2_15,

association. This chapter summarizes what is known about the physiological function of the hydrogenases, how they are integrated in the overall metabolism, their phylogenetic ancestry, and their distribution in cyanobacterial genomes.

Oxygenic phototrophs provide the framework that is needed for biological hydrogen production from sunlight and water, and they could be used as a blueprint for biomimetic systems for hydrogen generation. The genetic modifications that have been made to achieve higher production rates are described and future strategies for the use of cyanobacteria as rewarding H_2 producers are outlined.

I. Introduction to Hydrogenases

Hydrogen is a valuable energy source. This could be true for the provisioning of the human society at its present day level of development, but it is and was certainly so in the microbial world when the first prokaryotes evolved. Since Earth's primordial atmosphere was reducing (Dietrich et al., 2006), hydrogen was probably used by the early unicellular organisms to supply their metabolic and energetic needs.

The large number of extant hydrogenases, the enzymes splitting or making hydrogen, that occur in many different groups of bacteria, archaea and some eukaryotes, attest to their metabolic significance. Three different classes are distinguished on the basis of their metal content (Vignais et al., 2001; Vignais and Billoud, 2007).

The Fe-hydrogenases are confined to methanogens and contain one iron ion in their active site (Korbas et al., 2006; Shima et al., 2008). This enzyme class actually can also be considered as a dehydrogenase since a cofactor (methenyltetrahydromethanopterin) is needed for the catalysis and is hydrogenated/dehydrogenated (Vogt et al., 2007).

The second class of hydrogenases, the FeFe-hydrogenases, are found in bacteria but not in archaea. They also occur in algae and in anaerobic eukaryotes. Some of the latter contain specialized organelles called hydrogenosomes that are closely related to mitochondria. If present, the hydrogenases reside in these organelles (Hackstein et al., 2006; Mentel and Martin, 2008). The active site of the FeFe-hydrogenases is made up of a 4Fe4S-cluster that is bridged via a cysteine thiolate to an unusual dimer of iron atoms. Each of these iron ions has one carbon monoxide and one cyanide as diatomic ligands. There is an additional CO that might swing between a bridging position between the two irons and a semi-bridging position at the distal iron depending on the redox state of the active site (Nicolet et al., 2001; Roseboom et al., 2006; Zilberman et al., 2006). Two non-proteinaceous thiolates provide two bridging ligands in the iron pair. Although di(thiomethyl)amin was favored as the dithiolate ligand, since it would also provide a base that could accommodate a proton during heterolytic cleavage of the H_2 molecule (Thomann et al., 1991; van Dam et al., 1997; Nicolet et al., 2001; Fontecilla-Camps et al., 2007; Lubitz et al., 2007), recent structural refinement that it is 1,3-propanedithiol instead (Pandey et al., 2008). The characterized FeFe-hydrogenases show widely differing sensitivities to oxygen. Most, like the algal enzyme (Happe and Naber, 1993), are rapidly and irreversibly inactivated in the presence of oxygen, but there are enzymes that can be aerobically isolated and reactivated in the absence of O_2 (Fontecilla-Camps et al., 2007).

NiFe-hydrogenases are the most widespread class in prokaryotes and are distributed in bacteria and archaea. Up to now, no eukaryotes have been found to contain a NiFe-enzyme. This type of hydrogenase can be classified into four different families with a number of subfamilies that are distinguishable by signature amino acid sequences involved in binding the NiFe binuclear center (Vignais et al., 2001; Vignais and Billoud, 2007, see Fig. 15.1). Current enzymological evidence favors a mechanism involving heterolytic splitting of the hydrogen molecule and the formation of a hydride that is transiently bound to the nickel. During catalysis, the nickel changes its redox state whereas the oxidation state of the iron does not change (Fontecilla-Camps et al., 2007; Lubitz et al., 2007). The significance of this observation is not yet clear.

Most of the NiFe-hydrogenases are inactivated in the presence of oxygen, but all can be reversibly reactivated upon return to anaerobic conditions. The only hydrogenases known to be oxygen tolerant, that is to function at ambient levels of O_2,

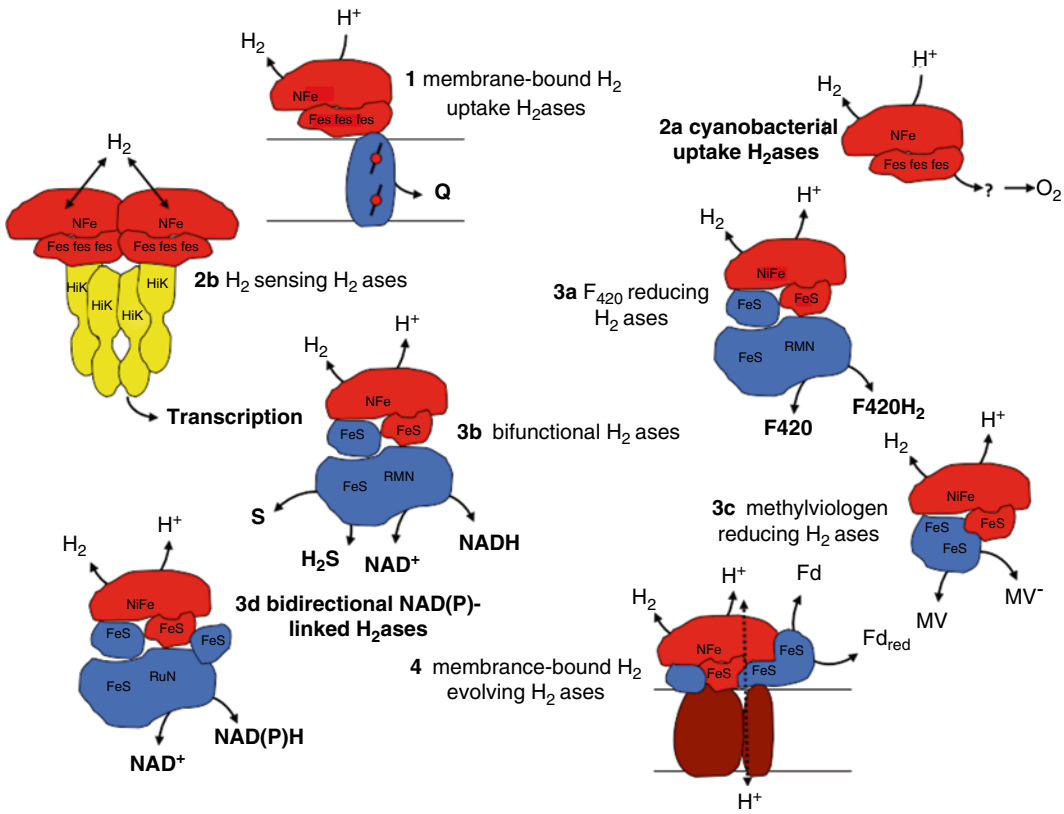


Fig. 15.1. Schematic representation of the structure of the different NiFe-hydrogenase. Groups present in cyanobacteria are *in bold*. The subunit structure is exemplary and in some cases subject to even more variations. Every enzyme consists of a hydrogenase module (*red*) that is made up of a large hydrogenase subunit with the NiFe-active site and a small subunit that either contains one or three FeS-clusters. This module is connected to another module (*blue, brown* or *yellow*) that works as an electron acceptor/donor, proton-transferring device or in signal transduction. Group 1 is connected to a transmembrane cytochrome b that is able to reduce a quinone in the membrane, group 2a are the cyanobacterial-type uptake hydrogenase with an unknown electron acceptor, group 2b are the receptor hydrogenases that are connected to a tetramer of a specific histidine kinase (Buhrke et al. 2004), group 3a are the F_{420}-reducing hydrogenases of methanogens, group 3b are the bifunctional NAD(P) linked hydrogenases first discovered in extremophiles and formerly called sulfhydrogenases, group 3c are the methylviologen-reducing hydrogenases of methanogens, group 3d are the bidirectional NAD(P)-linked hydrogenases. In cyanobacteria this enzyme is a heteropentamer (HoxEFUYH) (Schmitz et al. 2002; Germer et al. 2009) and in oxyhydrogen bacteria a heterohexamer (HoxFUYHI$_2$) (Burgdorf et al. 2005). Concerning the subunit composition group 4 might be most versatile containing probably between 6 and 14. Only the minimal version of this enzyme is shown, that is the energy-converting hydrogenase of *Methanosarcina barkeri*. The CO-oxidizing:H_2-forming enzyme, hydrogenase 3 and 4 of *E. coli* and the membrane bound hydrogenases of *Pyrococcus* also belong to group 4. In some cases a ferredoxin is the electron donor/acceptor for these hydrogenases.

belong to this class (Schneider and Schlegel, 1977; King, 2003; Burgdorf et al., 2005a; Vincent et al., 2006; Cracknell et al., 2008; Goldet et al., 2008).

It is important to note that all three classes, Fe-, FeFe-, and NiFe-hydrogenases, contain a low spin Fe (II) ion bound to the protein by one or two thiolates coordinated with CO/CN ligands, or ligands with similar properties (like the 2-pyridinol compound of the Fe-hydrogenase). Since these three enzyme classes are phylogenetically unrelated, it is a beautiful example of convergent evolution. It might indicate that an iron with this specific coordination sphere is the first and perhaps only choice for a living organism to split or form hydrogen.

Metabolically, hydrogen is very versatile. Either its reducing power can be used to generate cell constituents, or its energy is directly used to drive the generation of a proton gradient. Additionally, when hydrogen is produced under fermentative conditions, its evolution is accompanied by an additional step of substrate level

phosphorylation (see Fig. 15.5). In the absence of oxygen, pyruvate can be converted to CO_2 and acetyl-CoA. When the latter is converted to acetylphosphate, its high phosphate group transfer potential can be used for the production of an additional ATP. The whole reaction also generates two low potential electrons and would not be feasible if it were not for the hydrogenase that uses these electrons and produces H_2.

The versatility of FeFe- and NiFe-hydrogenases makes them intriguing examples how protein modules can be shuffled to invent new traits. Accordingly, hydrogen oxidation/reduction is linked to the reduction/oxidation of an array of electron acceptors/donors that serve a variety metabolic needs of different organisms. The NiFe-hydrogenases in particular show a wide variety of combinations with other redox mediating modules (Fig. 15.1). In these enzymes, a large hydrogenase subunit that contains the NiFe active site always occurs with a small hydrogenase subunit with either one or three FeS clusters. In the different enzyme classes, this hydrogenase module forms a complex with other redox modules. Examples include a transmembrane cytochrome *b* in the periplasmic membrane bound H_2 uptake hydrogenases (group 1), a diaphorase in case of the bidirectional NAD(P)-linked hydrogenases, and the formate dehydrogenase in case of some of the energy converting hydrogenases (group 4).

Many metalloenzymes need multiple accessory proteins for the insertion of metals into their active sites (see i.e. Kuchar and Hausinger, 2004). The elaborate metal centers of hydrogenases are no exception. Three additional genes are necessary for the posttranslational assembly of an active FeFe-hydrogenase (Posewitz et al., 2004; King et al., 2006). A minimum of six so-called *hyp*-genes (hydrogenase pleitropic, *hypA-F*, s. Figs. 15.2 and 15.4) are needed to insert iron in conjunction with its diatomic ligands, (Reissmann et al., 2003) and nickel into the catalytic center of the NiFe-hydrogenases. All the respective homologs of these *hyp*-genes have also been described for cyanobacteria (Hoffmann et al., 2006). Some of the membrane bound H_2-uptake hydrogenases, like the oxygen tolerant enzyme of *Ralstonia eutropha,* need at least 13 processing proteins for their maturation (Kortlüke and Friedrich, 1992; Ludwig et al., 2009). At least some of the additional assembly proteins are needed for the proper translocation of the enzyme across the cytoplasmic membrane, which occurs via a TAT-type mechanism. In the case of the periplasmic NiFe-hydrogenases, only the small subunits contain a TAT-signal, and the transfer is catalyzed after the assembly of the entire enzyme complex (Manyani et al., 2005; Schubert et al., 2007). For a detailed review of the maturation process of FeFe- and NiFe-hydrogenases, see Böck et al. (2006).

II. Cyanobacterial Hydrogenases

Cyanobacteria can be used for the production of hydrogen using minerals, water, and sunlight. Due to their potential biotechnological applications, cyanobacterial hydrogen metabolism and the biology of cyanobacterial hydrogenases have been a matter of interest for some time (Frenkel et al., 1950). Nevertheless, biochemical characteristics and much of their physiology have remained elusive and are only beginning to be unraveled.

Cyanobacteria harbor two different NiFe-hydrogenases. Searches for FeFe-hydrogenases have been unsuccessful until now (Ludwig et al., 2006). Depending upon their physiological behavior, they have been named as either uptake or bidirectional hydrogenases. In the classification scheme developed by Vignais et al. (2001) and Vignais and Billoud (2007), they belong to the subfamilies of cyanobacterial uptake hydrogenases (group 2a) and bidirectional NAD(P)-linked hydrogenases (group 3d). Phylogenetically, both cyanobacterial enzymes share common ancestors with the enzymes of the filamentous anoxygenic phototrophic bacteria (FAPB), *Chloroflexi*, which also posses both types (Ludwig et al., 2006). Therefore, the current distribution found in cyanobacteria is the result of the retention of both, or differential loss of one, or both classes, depending upon the ecological niche the particular cyanobacterial strain occupies (Table 15.1).

In recent years a number of reviews have appeared about cyanobacterial hydrogenases. For the most recent the reader is referred to Tamagnini et al. (2007).

A. Uptake Hydrogenase

Up to now, the cyanobacterial uptake hydrogenase encoded by *hupS* and *hupL* (Fig. 15.2) has been exclusively found in strains harboring nitrogenase

Table 15.1. Distribution of hydrogenase genes in the sequenced cyanobacterial genomes. *hupL* is the gene of the large subunit of the uptake hydrogenase, *hoxH* the gene of the large subunit of the bidirectional hydrogenase, *nifD* encodes the α-subunit of the nitrogenase, *nifH* the Fe-protein of the nitrogenase and *nifJ* the pyruvate:ferredoxin/flavodoxin oxidoreductase. The accession numbers of the deduced protein sequences are given

Strain	HupL[a]	NifD	NifH	HoxH	NifJ
Acaryochloris marina MBIC11017				YP_001521996	YP_001522063
Anabaena variabilis ATCC 29413	YP_325087	YP_324742	YP_324416	YP_325153	YP_323551* YP_321599
Arthrospira maxima CS-328				ZP_03273562	ZP_03273569*
Arthrospira platensis str. Paraca				ZP_06380718	ZP_06381891
Crocosphaera watsonii WH 8501	ZP_00519188	ZP_00516387	ZP_00516386		ZP_00518015
Cyanobium sp. PCC 7001				YP_002597857	YP_002597848
Cyanothece sp. ATCC 51142	YP_001802481	YP_001801977	YP_001801976	YP_001803731	YP_001802370
Cyanothece sp. CCY 0110	ZP_01728928	ZP_01727766	ZP_01727765	ZP_01727423	ZP_01730229
Cyanothece sp. PCC 7424	YP_002377118	YP_002377414	YP_002377415	ZP_02972728	YP_002376576*
Cyanothece sp. PCC 7425	YP_002483374[b]	YP_002483084	YP_002483083	YP_002484718	YP_002485040
Cyanothece sp. PCC 7822	ZP_03153783	ZP_03154128	ZP_03154129	ZP_03154336	ZP_03157112*
Cyanothece sp. PCC 8801	YP_002373345	YP_002371988	YP_002371987	YP_002370357	YP_002374020
Cyanothece sp. PCC 8802	ZP_03142797	ZP_03144923	ZP_03144922	ZP_03143669	ZP_03141892
Cylindrospermopsis raciborskii CS-505	ZP_06309263	ZP_06309411	ZP_06309412	ZP_06307638	ZP_06307770 ZP_06309351
Gloeobacter violaceus PCC 7421					
Lyngbya sp. PCC 8106	ZP_01619041	ZP_01620767	ZP_01620768	ZP_01622077	ZP_01622083*
Microcoleus chthonoplastes PCC 7420				YP_002619903	YP_002620835
Microcystis aeruginosa NIES-843				YP_001656435	YP_001658828
Microcystis aeruginosa PCC 7806				CAO89286	CAO88863.
Nodularia spumigena CCY 9414	ZP_01628406	ZP_01628430	ZP_01629115	ZP_01629499	ZP_01630855
Nostoc azollae 0708	ZP_03768004	ZP_03768758	ZP_03768757 ZP_03767991 ZP_03769194	ZP_03765204	
Nostoc punctiforme PCC 73102	ZP_00112356	ZP_00112319	YP_001869139		ZP_00111796
Nostoc sp. PCC 7120	NP_484720	NP_485484	NP_485497	NP_484809	NP_485951 NP_486843
Prochlorococcus marinus str. AS9601					
Prochlorococcus marinus str. MIT 9211					
Prochlorococcus marinus str. MIT 9215					
Prochlorococcus marinus str. MIT 9301					
Prochlorococcus marinus str. MIT 9303					
Prochlorococcus marinus str. MIT 9312					
Prochlorococcus marinus str. MIT 9313					
Prochlorococcus marinus str. MIT 9515					

(continued)

Table 15.1. (continued)

Strain	HupL[a]	NifD	NifH	HoxH	NifJ
Prochlorococcus marinus str. NATL1A					
Prochlorococcus marinus str. NATL2A					
Prochlorococcus marinus subsp. *marinus* str. CCMP1375					
Prochlorococcus marinus subsp. *pastoris* str. CCMP1986					
Raphidiopsis brookii D9				ZP_06303751	ZP_06305273
Synechococcus elongatus sp. PCC 6301				YP_172265	YP_172431*
Synechococcus elongatus sp. PCC 7942				YP_401572	YP_401401*
Synechococcus sp. BL107					
Synechococcus sp. CC9311					
Synechococcus sp. CC9605					
Synechococcus sp. CC9902					
Synechococcus sp. JA-2-3B'a(2–13)		YP_476681	YP_476680		
Synechococcus sp. JA-3-3Ab		YP_475237	YP_475238		
Synechococcus sp. PCC 7002				YP_001733469	YP_001734690
Synechococcus sp. PCC 7335		YP_002711054	YP_002711055	YP_002710310	YP_002711016
Synechococcus sp. RCC307					
Synechococcus sp. RS9916					
Synechococcus sp. RS9917					
Synechococcus sp. WH 5701				ZP_01085930	ZP_01085923*
Synechococcus sp. WH 7803					
Synechococcus sp. WH 7805					
Synechococcus sp. WH 8102					
Synechocystis sp. PCC 6803				NP_441411	NP_442703
Thermosynechococcus elongatus BP-1					
Trichodesmium erythraeum IMS101	YP_722943	YP_723618	YP_723617		
Cyanobacterium UCYN-A	YP_003421184	YP_003421697	YP_003421696		

*The respective *nifJ* gene is found in a gene cluster with *hyp*-genes or close to the *hox*-gene cluster

[a]The BLAST search for the orthologs was made with the respective protein sequence

[b]This hydrogenase does not belong to the cyanobacterial-like uptake hydrogenases but has highest homologies to the bifunctional NAD(P) linked hydrogenases

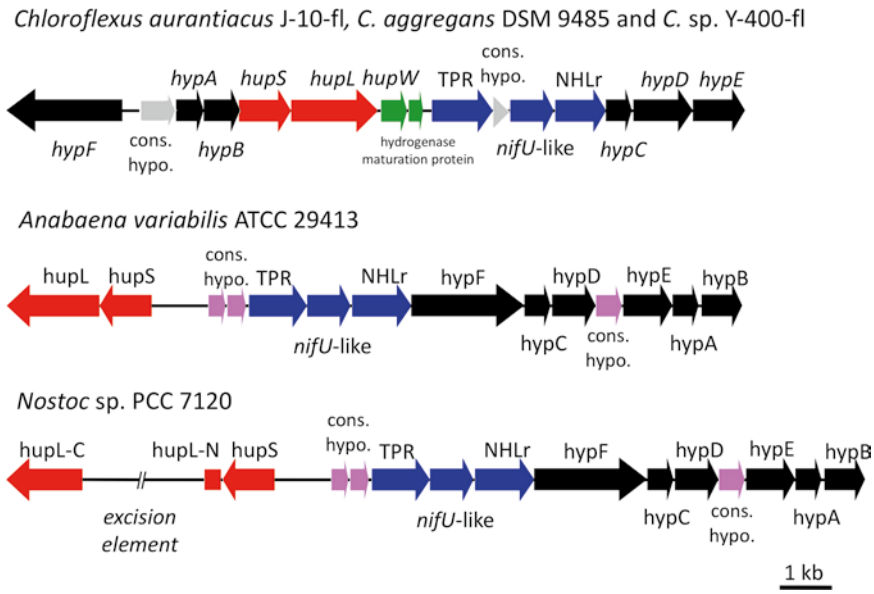


Fig. 15.2. Structure of the gene clusters of the uptake hydrogenase genes in filamentous anoxygenic photosynthetic bacteria (*Chloroflexi*) and some cyanobacteria. In the *Chloroflexi* the structural genes *hupS* and *hupL* (*red*) are tightly linked to a putative hydrogenase specific protease (*hupW*, *green*) and all *hyp*-genes (*black*) necessary for the post translational processing of the hydrogenase are found in direct vicinity. All these gene clusters contain two conserved hypothetical genes (cons. hypo., *grey*). In the cyanobacteria *hupS* and *hupL* are encoded in the inverse direction compared to all the *hyp*-genes. All the cyanoabacterial gene clusters share all the hyp-genes and three conserved hypothetical genes (cons. hypo., *magenta*). All the *Chloroflexi* and the cyanobacteria also encode for a tetratrico peptide repeat protein, a NifU-like protein and one containing NHL-repeats. In some cyanobacteria like *Nostoc* sp. PCC 7120 an excision element (≈10 kb) inside the *hupL*-gene is removed upon heterocyst development by a specific excisase.

genes. The transcription of its genes is activated only when the cells are fixing nitrogen, and is dependent on NtcA (Oliveira et al., 2004; Leitão et al., 2005; Weyman et al., 2008), a global regulator of cyanobacterial nitrogen metabolism (Herrero et al., 2001, 2004). In some N_2-fixing heterocystous strains, the expression of the uptake hydrogenase depends on the excision of a large DNA fragment during heterocyst development (Carrasco et al., 1995, 2005). Considering that all nitrogenases have an inherent hydrogen producing activity, evolving at least one H_2 per dinitrogen reduced to ammonia, it seems straightforward to assume that these hydrogenases function to recycle this hydrogen so that its reducing power is not lost for the cells.

Surprisingly, several N_2-fixing cyanobacteria do not harbor an uptake hydrogenase. These include the marine N_2-fixing unicellular *Synechococcus* sp. Miami BG 043511 (Ludwig et al., 2006), which has been the subject of detailed physiological studies on this matter and has been shown to be unable to take up the H_2 produced by its nitrogenase (Mitsui and Suda, 1995). Likewise, two *Synechococcus* strains (JA-2-3B'a(2–13) and JA-3-3Ab) isolated from a hot spring, *Cyanothece* sp. PCC 7425 and *Synechococcus* sp. PCC 7335, have the *nif*-genes encoding the nitrogenase but have no uptake hydrogenase genes (Table 15.1). Although some of these strains (*Synechococcus* sp. Miami BG 043511, *Cyanothece* sp. PCC 7425) contain the genes for the bidirectional hydrogenase, it is known that this enzyme is not used to oxidize hydrogen under N_2-fixing conditions (Masukawa et al., 2002). Of all the 57 sequenced cyanobacterial strains (Table 15.1), 20 have the genes for nitrogen fixation, but only 16 of these have the *hupL*-gene. Additionally, competition experiments in *Nostoc* sp. PCC 7120 did not result in the wild type cells overgrowing the uptake hydrogenase-free mutant. At low light intensities, the mutant grew as well as the wild type and at higher light intensities, its fraction decreased but was never eliminated (Lindblad et al., 2002).

Therefore, a selective advantage provided by expressing an uptake hydrogenase seems to be dependent upon specific environmental conditions, such as high light, which are not necessarily encountered by all N_2-fixing cyanobacteria.

From a biotechnological perspective aimed at high hydrogen production this enzyme diminishes photobiological yields in N_2-fixing cyanobacteria. Consequently, a number of studies examined deletion mutants of this hydrogenase in terms of hydrogen production under nitrogen fixing conditions via the nitrogenase (Tsygankov et al., 1999; Masukawa et al., 2002; Yoshino et al., 2007). All these studies found a considerable increase of H_2 output and will be further discussed in Section III.B.

Nevertheless, the uptake hydrogenase itself might be of special interest because of its high similarity to the H_2-sensing hydrogenases (Vignais et al., 2001). The latter class is able to catalyze H/D exchange even at ambient oxygen concentrations but does not catalyze net hydrogen evolution or uptake (Vignais et al., 1997, 2000). In contrast, the cyanobacterial counterpart actively takes up hydrogen and was found to be very oxygen sensitive. Although older literature suggests that this enzyme is irreversibly inhibited by O_2 when isolated (Houchins and Burris, 1981a), recent studies on a cyanobacterial-like uptake hydrogenase of the proteobacterium *Acidithiobacillus ferrooxidans* show that it can be reductively reactivated by H_2 (Schröder et al., 2007). Comparative studies on the mechanisms controlling the O_2-resistance of the H_2-sensing hydrogenases and the readily poisoned cyanobacterial uptake hydrogenases could further our understanding of how H_2 is split and illuminate the principles of oxygen tolerance.

The protein module that transfers the electrons of this type of NiFe-hydrogenase has not yet been described. Since the enzyme is membrane bound (Tel-Or et al., 1978; Houchins and Burris, 1981b; Lindblad and Sellstedt, 1990; Seabra et al., 2009), it may be linked to a transmembrane cytochrome b-like subunit by analogy to the H_2-uptake hydrogenases of *Wolinella succinogenes* or *Ralstonia eutropha* (Dross et al., 1992; Bernhard et al., 1997; Gross et al., 1998, 2004). However, clear homologs of these subunits are missing in cyanobacteria and their small subunits do not contain the TAT-signal of these periplasmic hydrogenases. Therefore, the cyanobacterial uptake hydrogenase resides on the cytoplasmic side of the membrane with an up to now unknown electron acceptor.

Hydrogen uptake is activated in the light and might be regulated by thioredoxin (Papen et al., 1986). In the absence of nitrogen and oxygen hydrogen evolution in heterocystous cyanobacteria is maximal. Under these conditions acetylene is reduced indicating an active nitrogenase, but since its substrate is lacking, a large proportion of its electrons are funneled into H_2 production (Benemann and Weare, 1974; Jones and Bishop, 1976; Lambert et al., 1979; Lambert and Smith, 1980; Tsygankov et al., 1998). The uptake of H_2 was found to be inhibited by CO, which is an inhibitor of NiFe-hydrogenases (Adams et al., 1980; Adams, 1990). If oxygen is added acetylene reduction continues at the same level but hydrogen output is reduced, indicating that at least part of the electrons of the uptake hydrogenase are passed into the respiratory chain and ultimately to oxygen.

1. Phylogenetic Origin of the Cyanobacterial Uptake Hydrogenase and Its Implications

The phylogenetic relationship of the uptake hydrogenase is shown in Fig. 15.3. All sequences of cyanobacterial-like uptake hydrogenases available to date have been included in the analysis. Four large clusters are distinguishable. One formed by the actinobacteria and *Chloroflexi*, one of the cyanobacteria together with some proteobacteria, one of the H_2-sensing hydrogenases, and one with the membrane bound H_2-uptake hydrogenases. The most deeply branching photosynthetic bacteria in the cluster of the cyanobacterial-like uptake hydrogenases are the *Chloroflexi*. This confirms the previously described pattern and the shared common ancestor of this enzyme for *Chloroflexi* and cyanobacteria (Ludwig et al., 2006).

In all the cyanobacteria containing the uptake hydrogenase the gene clusters of the *hyp*-genes (the hydrogenase maturation genes) contain five additional genes (Agerwald et al., 2008). All the proteobacteria branching from the cyanobacterial clade also share these same putative additional accessory genes (*asr0689, asr0690, alr0691, alr0692* and *alr0693* according to the gene numbering in *Nostoc* sp. 7120). This is strong support for them encoding accessory proteins. Since one encodes a NifU-like protein (Alr0692),

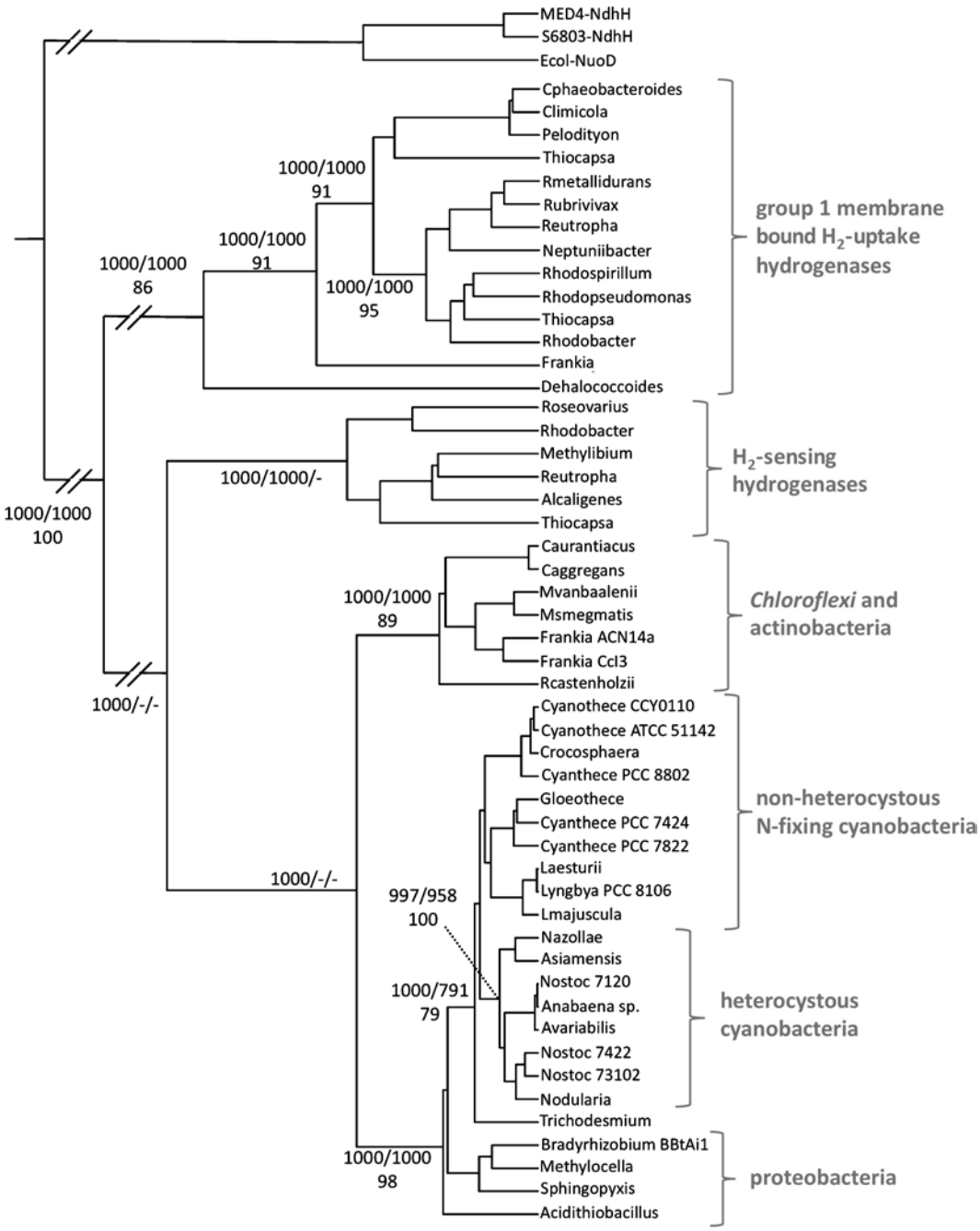


Fig. 15.3. Phylogenetic tree of the cyanobacterial-like uptake hydrogenases and H_2-sensing hydrogenases. The tree has been generated as described in Ludwig et al. (2006). The significant bootstrap values are given at the most important nodes and are for neighbor-joining, parsimony and maximum likelihood, respectively. The NuoD (NdhH in cyanobacteria) subunit has been used as an outgroup. The used sequences have been taken from the following bacteria *Synechocystis* sp. PCC 6803, *Prochlorococcus* sp. MED4, *Escherichia coli*, *Chlorobium phaeobacteroides* DSM 266, *Chlorobium limicola* DSM 245, *Pelodictyon phaeoclathratiforme* BU-1, *Thiocapsa roseopersicina* ,*Ralstonia metallidurans* CH34, *Rubrivivax gelatinosus*, *Ralstonia eutropha* H16, *Neptuniibacter caesariensis*, *Rhodospirillum rubrum* ATCC 11170, *Rhodopseudomonas palustris* CGA009, *Rhodobacter sphaeroides* 2.4.1, *Frankia* sp. EAN1pec, *Dehalococcoides ethenogenes* 195, *Roseovarius* sp. HTCC2601, *Methylibium petroleiphilum* PM1 *Alcaligenes hydrogenophilus, Chloroflexus aurantiacus* J-10-fl, *Chloroflexus aggregans* DSM 9485, *Mycobacterium vanabalenii* PYR-1, *Mycobacterium smegmatis* str. MC2 155, *Frankia* ACN14a*, Frankia* CcI3, *Roseiflexus castenholzii*, *Cyanothece* sp. CCY0110, *Cyanothece* sp. ATCC 51142, *Crocosphaera watsonii* WH 8501, *Cyanothece* sp. PCC 8802, *Gloeothece* sp. PCC 6909, *Cyanothece* sp. PCC 7424, *Cyanothece* sp. PCC 7822, *Lyngbya aestuarii, Lyngbya* sp. PCC 8106, *Lyngbya majuscula* CCAP 1446/4, *Nostoc* '*azollae*' 0708, *Anabaena siamensis, Nostoc* sp. PCC 7120, *Anabaena* sp., *Anabaena variabilis* ATCC 29413, *Nostoc* sp. PCC 7422, *Nostoc punctiforme* sp. PCC 73102, *Nodularia spumigena* CCY9414, *Trichodesmium erythraeum* IMS101, *Bradyrhizobium* sp. BTAi1, *Methylocella silvestris* BL2, *Sphingopyxis alaskensis* RB2256, *Acidithiobacillus ferrooxidans* ATCC 53993.

they might be necessary for the insertion of FeS-clusters (Johnson et al., 2005) into the small hydrogenase subunit.

A current analysis, including 31 proteins that have not been subject to horizontal gene transfer, confirms that all the proteobacterial lineages are the most recent additions to the bacterial kingdom (Ciccarelli et al., 2006). The same conclusion was already reached by Gupta and Griffiths (2002), who analyzed patterns of insertions/deletions in housekeeping proteins. Therefore, the proteobacterial hydrogenases resembling the cyanobacterial uptake hydrogenases must have been a later development and most likely originated from a lateral gene transfer from the cyanobacteria.

B. Bidirectional Hydrogenase

From a biotechnological perspective, the bidirectional hydrogenase encoded by the *hox*-genes (*hoxEFUYH*) (Fig. 15.4) is probably the most interesting in cyanobacteria. This is because it can be rapidly reactivated after exposure to oxygen and it is the native hydrogenase in cyanobacteria best suited for the generation of hydrogen from sunlight and water. Although oxygen tolerant hydrogenases are available (Burgdorf et al., 2005a) that are probably capable of photosynthetic hydrogen production in parallel to oxygen evolution, it might be difficult to derive a good expression system for them in cyanobacteria. Furthermore, not all the details of the maturation process of the NiFe-hydrogenases are known in detail. For example, the metabolic origin of the CO ligands is not yet clear (Forzi et al., 2007; Lenz et al., 2007) and might not be easily transferable to photosynthetic organisms.

The native bidirectional hydrogenase of cyanobacteria is involved in fermentative hydrogen production (Troshina et al., 2002; Gutthann et al., 2007; Serebryakova and Tsygankov, 2007; Ananyev et al., 2008) as well as photoinduced hydrogen evolution and hydrogen uptake (Abdel-Basset and Bader, 1997, 1998; Abdel-Basset et al., 1998; Appel et al., 2000; Cournac et al., 2002, 2004; Gutthann et al., 2007). Its physiological behavior depends on the redox conditions in the cell. If low potential electrons are plentiful,

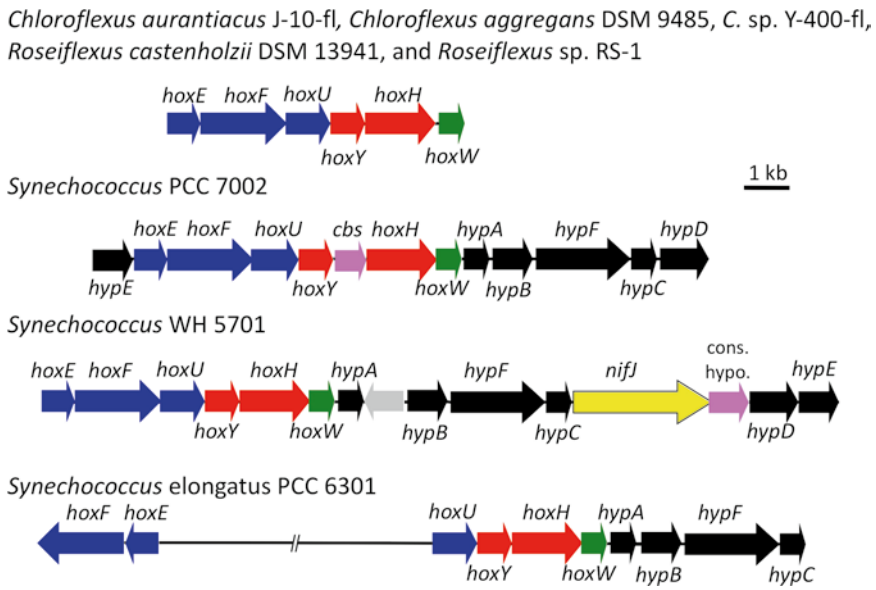


Fig. 15.4. Organization of the genes of the bidirectional hydrogenase in *Chloroflexi* and cyanobacteria. The small and large hydrogenase subunits encoded by *hoxY* and *hoxH* (*red*) and the diaphorase subunits encoded by *hoxE*, *F* and *U* (*blue*) make up the pentameric enzyme. In some cyanobacteria like *Synechococcus* sp. PCC 7002 a cbs domain protein is encoded between the structural genes and the gene of a putative protease is directly downstream of the structural genes. The gene cluster of *Synechococcus* WH 5701 not only contains all the *hyp*-genes (*black*) but also *nifJ* (*yellow*) encoding a pyruvate flavodoxin/ferredoxin oxidoreductase. In some cyanobacteria like *Synechococcus elongatus* PCC 6301 *hoxE* and *hoxF* are separated by more than 300 kb from all the other *hox*-genes.

this enzyme is used to pass them on to protons to generate H_2. If reducing power is low and oxidizing conditions prevail, H_2 is split to protons and electrons and the latter are transferred back into the cellular metabolism.

Recent biochemical and spectroscopic investigations confirm that catalytic amounts of NADH or NADPH in the absence of oxygen are sufficient to reactivate the bidirectional enzyme (Germer et al., 2009). The catalytic center was found to contain two cyanide ligands and one carbon monoxide, as in the standard NiFe-hydrogenases. Most surprisingly, only four different redox states of the NiFe active site could be distinguished by FT-IR (Fourier transform infrared spectroscopy), and the nickel was EPR silent in the as isolated, most oxidized, inactive state. The latter finding is especially remarkable since the EPR detectable Ni_u-A and Ni_r-B inactive states of the standard NiFe-hydrogenases transiently shift to an EPR-undetectable state before shifting to another active state of the NiFe-center. It is appealing to speculate that the cyanobacterial bidirectional hydrogenase might be trapped in such a latent EPR-silent state when inactivated and switches back to an active state in a single step. Obviously, this would be much faster than any of the two activation steps needed in the standard enzymes. Putting it another way, it seems that cyanobacteria found their own way to activate or deactivate NiFe-hydrogenases.

1. Fermentative Hydrogen Production by the Bidirectional Hydrogenase

Under anoxic conditions, cyanobacteria switch from their normal metabolism to fermentation. A number of different pathways have been described such as homo- and heterolactic acid fermentation, mixed acid and homoacetate fermentation, and are reviewed by Stal and Moezelaar (1997). Under these conditions a wide variety of cyanobacteria are known to produce hydrogen (Lambert and Smith, 1980; Howarth and Codd, 1985; Almon and Böger, 1988; van der Oost and Cox, 1989; Troshina et al., 2002; Gutthann et al., 2007). The available information strongly suggests that hydrogen is evolved due to a phosphoroclastic reaction involving a pyruvate: ferredoxin/flavodoxin oxidoreductase encoded by *nifJ* (van der Oost et al., 1989; Schmitz et al., 2001b; Troshina et al., 2002) (Fig. 15.5).

This is consistent with the distribution of *nifJ* in the different cyanobacterial genomes currently sequenced (Table 15.1). There are only three exceptions to the rule that every genome containing the genes of the bidirectional hydrogenase also harbors *nifJ* and vice versa. These are the nitrogen fixing strains *Crocosphaera watsonii, Nostoc azollae,* and *Nostoc punctiforme* that do not harbor the *hox*-genes but do harbor *nifJ*. This is easily explained since the pyruvate: ferredoxin/flavodoxin oxidoreductase is also used during nitrogen fixation for the production of reducing equivalents for the nitrogenase (Neuer and Bothe, 1982; Bauer et al., 1993). Moreover, there are several cyanobacterial strains where *nifJ* is encoded in the same gene cluster as some or even all of the *hyp*-genes and close to the *hox*-genes (Table 15.1). In contrast to experimental results that found very low or undetectable rates of H_2-production in vitro with plant ferredoxin or the major photosynthetic ferredoxin (Serebryakova et al., 1996; Schmitz et al., 2002), the observed genetic linkage between *nifJ* and the bidirectional hydrogenase suggests that ferredoxin/flavodoxin might be direct electron donors to hydrogen production. This apparent contradiction might reflect limitations in the choice of the ferredoxins in the in vitro assays. Cyanobacteria harbor a variety of ferredoxins and nothing is known about those used when oxygen is absent.

In summary, one primary role of this type of hydrogenase is to use protons as final electron acceptors while cells ferment. This function is consistent with the observation that there is a considerable increase of dark anaerobic hydrogen evolution when nitrate reduction is impaired and cannot be used as an alternative electron sink (Gutthann et al., 2007) (Fig. 15.6).

2. Photoinduced Hydrogen Production and Uptake by the Bidirectional Hydrogenase

In a number of cyanobacteria, light-induced hydrogen evolution has been shown to be due to the activity of the bidirectional hydrogenase (Abdel-Basset and Bader, 1997, 1998; Abdel-Basset et al., 1998; Cournac et al., 2002, 2004; Gutthann et al., 2007). The hydrogenase is activated under anoxic conditions and if cells are then illuminated they produce hydrogen for a

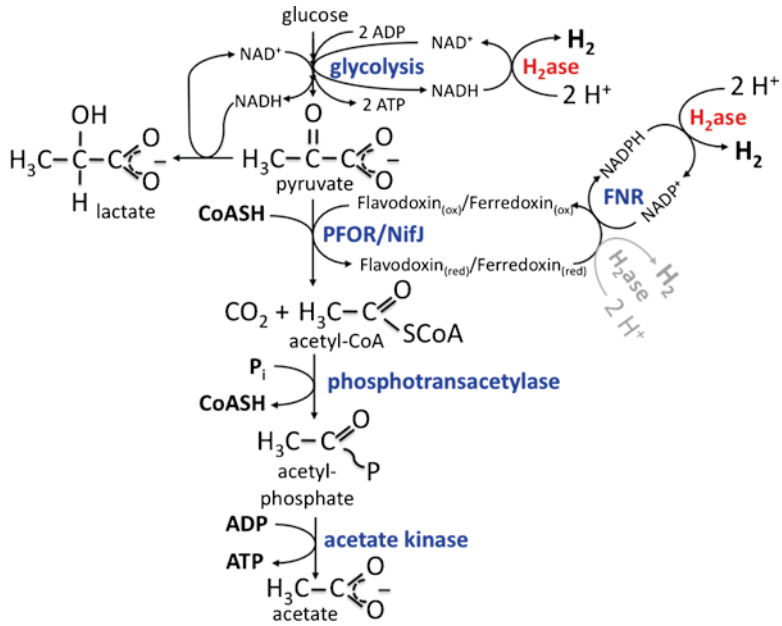


Fig. 15.5. Fermentative hydrogen evolution in cyanobacteria by the bidirectional hydrogenase (H_2ase). Lactate fermentation is shown here. An additional ATP generating step could be added by using the phosphoroclastic reaction that oxidizes pyruvate to acetate. The first step is catalyzed by the PFOR encoded by *nifJ*. In this case reduced ferredoxin or flavodoxin are generated. Both can be regenerated by the FNR that reduces in turn $NADP^+$. The bidirectional hydrogenase could then oxidize NADPH by producing H_2, although this pathway would be thermodynamically less favorable, these reactions are all feasible on the basis of current experimental evidence. If the hydrogenase can also accept reduced ferredoxin/flavodoxin as electron donor is not yet clear, but would be more appropriate for applied hydrogen production (see text). It has been shown that this hydrogenase is able to accept NADPH as well as NADH (Schmitz and Bothe, 1996b) and could therefore also regenerate part of the NADH produced in glycolysis.

transient period (i.e. Fig. 15.6). Because other processes start to compete and conditions become more and more oxidizing ultimately the hydrogen production phase gives way to hydrogen uptake. The duration of production can be expanded in mutants of the respiratory oxidases (Gutthann et al., 2007), and especially in a mutant without the NDH-1 complex that is heavily impaired in respiration as well as cyclic electron transfer (Cournac et al., 2004).

The electrons used to produce hydrogen reside in the electron transfer chain, but if hydrogen production is prolonged, such as in the mutant M55 that lacks the NDH-1 complex (Ogawa, 1991; Zhang et al., 2004), the electrons are clearly derived from water-splitting (Cournac et al., 2002, 2004). They can also be delivered by organic compounds if provided externally and if photosynthetic electron transfer is blocked by DCMU (Cournac et al., 2004). If light-induced hydrogen evolution is a by-product of activation of the hydrogenase in the absence of oxygen or fulfills an essential function as a redox buffer remains to be shown.

However, it has to be taken into account that under normal aerobic conditions oxygen represents an important electron acceptor of low potential electrons generated by photosystem I. This reaction is called the Mehler-reaction (Mehler, 1951; Mehler and Brown, 1952). It is also reffered to as the water-water cycle, since water is split at one end of the electron transfer chain (the oxygen evolving complex of PSII) and produced at the other end (the acceptor side of PSI) by reducing oxygen (Asada, 1999).

Several enzymes of the Calvin-Benson-Bassham cycle are inactive in darkness and need a reductive activation via the ferredoxin/thioredoxin system (Dai et al., 1996; Buchanan and Balmer, 2005). Therefore, in transition states when photosynthesis resumes and CO_2-fixation is not yet active, the pools of the electron acceptors of photosystem I are quickly filled. The resulting traffic jam inhibits the safe delivery of

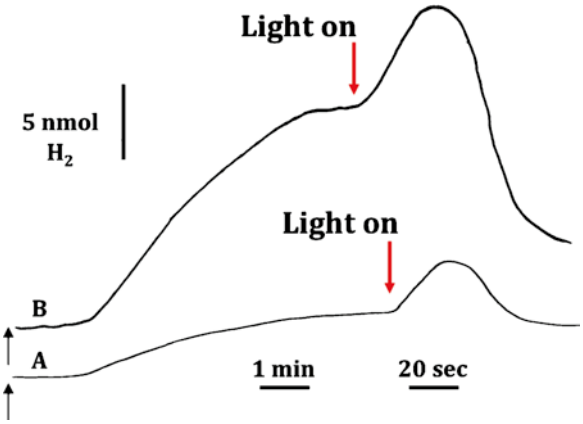


Fig. 15.6. Original recorder traces from measurements at the hydrogen electrode of *Synechocystis* sp. PCC 6803 cultures either grown in BG-11 (*trace A*) or in BG-11_0 with tungstate instead of molybdate and ammonia (*trace B*) (figure taken from Gutthann et al. 2007). The measurements were performed in BG-11. Due to the addition of tungstate, tungsten is inserted into the active site of the nitrate reductase resulting in an inactive enzyme. Even after addition of nitrate it cannot be used as electron acceptor (*trace B*). The upward pointing arrows indicate the addition of glucose, glucose oxidase and catalase at the start of the experiment. Reprinted from BBA-Bioenergetis, 1767, F. Gutthann, M. Egert, A. Marques and J. Appel, Inhibition of respiration and nitrate assimilation enhances photohydrogen evolution under low oxygen concentrations in Synechocystis sp. PCC 6803, 161–169, 2007, with permission from Elsevier.

the energy absorbed by the photosystems and causes photodamage. It can only be overcome when CO_2-fixation has already been activated by previous light treatments (Harbinson and Hedley, 1993; Schansker et al., 2003, 2005, 2006). Therefore, the Mehler-reaction is an important valve to avoid overexcitation. In cyanobacteria, a similar reaction has been described that is catalyzed by A-type flavoproteins by direct reduction of oxygen to water (Helman et al., 2003, 2005). Consequently, the production of superoxide and the activity of the superoxide dismutase were found to be very low in cyanobacteria (Tichy and Vermaas, 1999). The deletion of the genes encoding the respective flavoproteins results in severe acceptor side limitation at photosystem I when cells are shifted from dark to light (Helman et al., 2003). Triple isotope fractionation of oxygen was used to determine the amount of photoreduction and respiration in *Synechocystis* sp. PCC 6803. By using mutants of the respiratory oxidases (Howitt and Vermaas, 1998) and the flavoproteins, it was estimated that 6% of the electrons leaving photosystem II are fed into the respiratory oxidases whereas 40% are used to photoreduce oxygen (Helman et al., 2005) and confirms previous estimations (Badger et al., 2000).

These results underscore the importance of electron sinks other than CO_2-fixation beyond photosystem I. If photosynthesis has to resume from a dark oxygen-limited, or even an anaerobic stage, these alternatives might be scarce. Comparing photoproduction of hydrogen in the presence and absence of nitrate reduction shows that nitrate reduction is used as such an alternative and that protons are used more so when it is absent (Fig. 15.6). Consequently, studies showed that photosystem I of a hydrogenase-free mutant is acceptor side limited compared to wild type cells (Appel et al., 2000). This difference becomes especially obvious when compared to cells kept under aerobic conditions (Fig. 15.7).

It is also remarkable that the cyanobacterial bidirectional hydrogenase is unusually rapidly activated in the absence of oxygen (Appel et al., 2000; Cournac et al., 2004; Gutthann et al., 2007). Many other NiFe-hydrogenases of anaerobic bacteria need a time span on the order of an hour to be reactivated once inactivated and need a special treatment under reducing conditions (Fernandez et al., 1984; Lissolo et al., 1984). Therefore, the bidirectional hydrogenase seems to be well adapted as a valve for low potential electrons in rapidly alternating redox conditions that might be unique to cyanobacteria.

3. Transcriptional Regulation of the Bidirectional Hydrogenase

Different cyanobacterial strains show different expression patterns of this type of hydrogenase. A number of unicellular strains like *Synechocystis* sp. PCC 6803, *Gloeocapsa alpicola,* and *Aphanothece halophytica* already express this hydrogenase under aerobic autotrophic growth conditions (Serebryakova et al., 1998; Appel et al., 2000; Ludwig et al., 2006; Serebryakova and Sheremetieva, 2006) and induce transcription further when shifted to microaerobic conditions (Summerfield et al., 2008; Kiss et al., 2009). Other

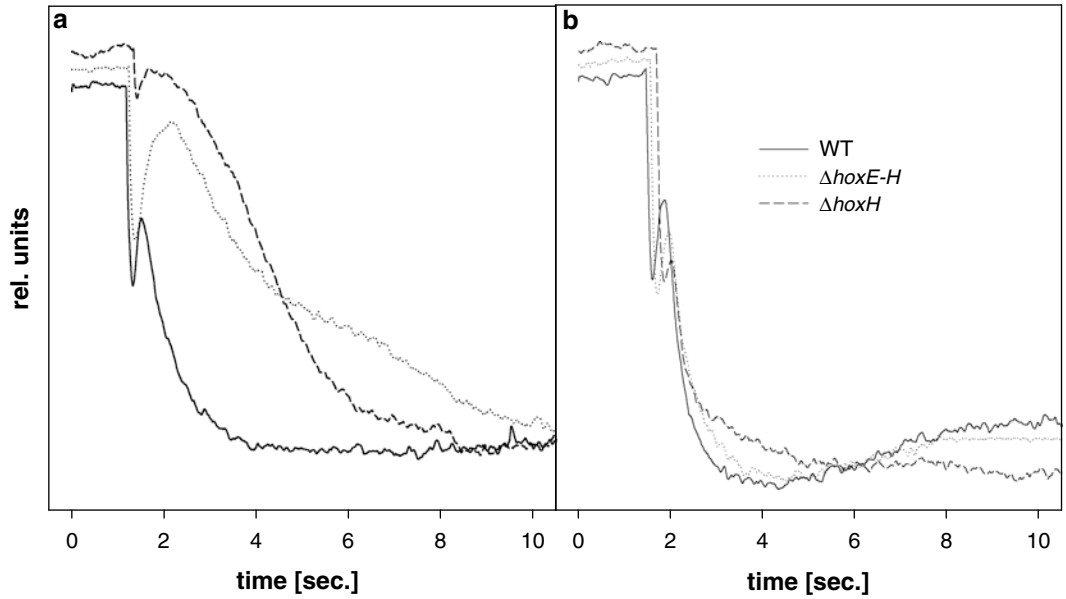


Fig. 15.7. P700 oxidation kinetics of wild type cells of *Synechocystis* sp. PCC 6803 in comparison to two mutants without the whole operon of the bidirectional hydrogenase (Δ*hoxE-H*) or the gene of the large subunit (Δ*hoxH*). At the steep decline around 1–1.5 s after the start of measurements, light was switched on. The delay until the oxidation level of P700 reaches a steady state under anoxic conditions (**a**) when compared to aerobic conditions (**b**) is obvious in the strains missing a functional bidirectional hydrogenase.

strains like the heterocystous strains *Nostoc* sp. PCC 7120, *Anabaena cylindrica,* and *Anabaena variabilis* induce this hydrogenase only under anaerobic conditions or in the presence of DCMU in the light (Houchins and Burris, 1981a; Serebryakova et al., 1992, 1994). These differences suggest that there is more to learn about the fine-tuning of this enzyme, but certainly also more about the ecological niches that different cyanobacterial strains have adapted to.

Two different patterns concerning the genetic organization of the *hox*-genes of the bidirectional hydrogenase are found. In some cyanobacteria like *Synechocystis* sp. PCC 6803 (Appel and Schulz, 1996), *Nodularia spumigena*, *Synechococcus* sp. PCC 7002, and *Synechococcus* sp. WH 5701 all the genes are in a single gene cluster. In *Synechocystis* it was shown that the all the genes are transcribed in a single polycistronic transcript (Gutekunst et al., 2005; Oliveira and Lindblad, 2005). In other cyanobacteria such as *Synechococcus* sp. PCC 7942, *Nostoc* sp. PCC 7120, and *Anabaena variabilis hoxEF* are separated from *hoxUYH*. For *Synechococcus* sp. PCC 6301 and sp. PCC 7942 it has been shown that these gene clusters are transcribed as single transcripts and are under the control of different promoters (Boison et al., 2000; Schmitz et al., 2001a). Interestingly, in the latter *Synechococcus* strain both *hox*-operons are expressed with a circadian rhythm in synchronized cultures with a maximum at subjective dusk. The transcript level of *hoxU* was approximately 3.5-fold higher compared to *hoxE* and, because of putative internal termination sites in the *hoxUYH* operon 6.4 × higher compared to *hoxH* (Schmitz et al., 2001a). Whether or not these transcript stoichiometries are relevant for the expression of the hydrogenase subunits remains to be seen.

Detailed investigations of the promoter region of the *hox*-operon of *Synechocystis* revealed that LexA is a transcription activator (Gutekunst et al., 2005). In contrast to enteric bacteria where it is a repressor of the SOS regulon (Butala et al., 2009), it seems to have a different function in cyanobacteria. Mutational investigations suggest that it is not involved in the regulation of DNA damage repair but in the regulation of the carbon metabolism (Domain et al., 2004) and working as a regulator in the redox sensitive expression of the RNA helicase crhR (Patterson-Fortin et al., 2006; Patterson-Fortin and Owttrim, 2008).

Subsequently, a number of studies showed that LexA also binds upstream of *hyp*- and *hox*-genes in other cyanobacteria (Ferreira et al., 2007; Sjöholm et al., 2007). In addition to LexA, an AbrB-like transcription factor was shown to be involved in transcriptional activation of the *hox*-operon in *Synechocystis* (Oliveira and Lindblad, 2008). Since LexA and the two AbrB-like proteins (Sll0359 and Sll0822) in the *Synechocystis* genome have also been implicated in regulating the bicarbonate transporter encoded by *sbtA* (Lieman-Hurwitz et al., 2008), there might be an elaborate interplay of different transcription factors to adjust the expression of a number of cyanobacterial genes related to carbon and energy metabolism.

4. Putative Interaction of the Bidirectional Hydrogenase with the NDH-1 Complex

As the first sequences of the cyanobacterial bidirectional NAD(P)-linked hydrogenases became available, a strong similarity of the diaphorase part of the enzyme to part of the subunits of the respiratory complex I (NDH-1) was recognized. This led to the hypothesis that the hydrogenase might form the peripheral part of this complex oxidizing NAD(P)H or H_2 and delivering electrons into the respiratory electron transfer (Appel and Schulz, 1996; Schmitz and Bothe, 1996a, b). When the complete sequence of *Synechocystis* sp. PCC 6803 was published (Kaneko et al., 1996) it was obvious that other homologs of these three subunits could not be found elsewhere in the genome. This is confirmed by all other completely sequenced cyanobacteria available to date.

The existence of a functional complex of these subunits was challenged by studies showing that cyanobacterial respiration neither relies on the presence of the bidirectional hydrogenase (Boison et al., 1999) nor is impaired if *hoxF* is deleted (Howitt and Vermaas, 1999). Nonetheless, the hypothesis of the hydrogenase being part of the NDH-1 complex still needs to be conclusively disproven. Since it was shown that even the deletion of two of the three terminal respiratory oxidases in *Synechocystis* sp. PCC 6803 does not result in a measurable decrease of respiratory activity (Howitt and Vermaas, 1998), there seems to be a high metabolic flexibility and redundancy concerning the electron transfer chain. This is especially true for the high number of different NDH-1 complexes in cyanobacteria involved in respiration, cyclic electron transfer, and the CO_2-concentrating mechanism (Ohkawa et al., 2000; Shibata et al., 2001; Maeda et al., 2002; Zhang et al., 2004; Battchikova and Aro, 2007). A number of investigations showed that the complex oxidizes NADPH (Cooley and Vermaas, 2001; Ma et al., 2006). It was hypothesized that the FNR might replace these subunits in chloroplasts (Guedeney et al., 1996), but recent investigations could not confirm its presence in fractions of the cyanobacterial NDH-1 that were actively oxidizing NADPH (Ma et al., 2006).

Interestingly, the absence of any NDH-1 complexes, such as described for the M55 mutant (Zhang et al., 2004), induces a low hydrogenase activity (Gutthann et al., 2007), indicating a relationship between the two. If this is due to a regulatory effect because of the altered redox state of the plastoquinone pool of this mutant (Schreiber et al., 1995), or due to a lack of interaction partners weakening the structural integrity of the hydrogenase remains to be shown. Therefore, it is still elusive if this enzyme is purely soluble or part of one of several respiratory NDH-1 complexes involved in respiration that can be substituted by others.

5. Ecological Distribution of the Bidirectional Hydrogenase

As summarized in the previous sections, the bidirectional hydrogenase is activated only in the absence of oxygen. During daylight hours it is unlikely for oxygenic phototrophs to become anaerobic but in eutrophic lakes cyanobacterial and algal blooms are known to induce such heavy respiratory activity that oxygen is completely consumed, especially during the night. Analysis of the ecological affiliation of the bidirectional hydrogenase in cyanobacteria reveals that all freshwater, all microbial mat associated, and many coastal strains sequenced to date harbor the *hox*-genes. Conversely those found in the open ocean, including those with large genomes like *Crocosphaera watsonii* and *Trichodesmium erythraeum* do not. For marine cyanobacteria living in the oligotrophic ocean, anoxic conditions are a rare encounter but for many other ecological niches there seems to be a selective force to be

able to switch to a more efficient fermentation and to restart photosynthesis from anaerobiosis (Barz et al., 2010).

This certainly applies to cyanobacterial mats where the oxygen concentrations change from supersaturating to completely anaerobic each day (Jørgensen et al., 1979, 1986; Epping et al., 1999). In fact, these mats have been found to produce hydrogen at maximum values around dawn (Hoehler et al., 2001). Since the microbial community of these mats is highly diverse (Ley et al., 2006), the hydrogen is not necessarily originating from cyanobacteria or cyanobacteria alone, but the evolution characteristics seem to fit into the patterns found for pure cyanobacterial cultures in the laboratory (Fig. 15.6).

Under anaerobiosis, in the presence of sulfide, some cyanobacterial strains switch from oxygenic to anoxygenic photosynthesis (Padan, 1979; Arieli et al., 1994). Under these conditions *Aphanothece halophytica* and *Oscillatoria limnetica* are either able to fix CO_2 at the expense of hydrogen (Belkin and Padan, 1978a), or to produce hydrogen if CO_2 is absent (Belkin and Padan, 1978b). More recent investigations showed that the bidirectional hydrogenase is the only hydrogenase in these two strains and therefore supports these two processes (Ludwig et al., 2006).

III. Manipulation of Pathways to Enhance Hydrogen Production

Since cyanobacteria possess the machinery needed for light-driven water-splitting, their photosynthetic complexes are of prime interest for the development of artificial photosynthetic or biomimetic systems. In addition, their hydrogenases might also be of technological interest because of their peculiar traits such as rapid reactivation or similarity to H_2-sensors. Although exciting, the investigations of these aspects (Vincent et al., 2006; Alonso-Lomillo et al., 2007) are outside the scope of this review and are not discussed any further. The remainder of this chapter focuses on direct hydrogen production from cyanobacterial cells. For an overview also including photobiohydrogen production by algae see the review of Ghirardi et al. (2007).

It is certain that cyanobacteria will not produce large amounts of H_2 voluntarily. Since it is a waste of energy, except under anaerobic conditions, there has been a strong natural selection against it and in favor of a tight regulation of the involved enzymes. If biotechnologically rewarding production is to be achieved, these competitive processes and regulatory mechanisms have to be bypassed and pathways have to be manipulated, diverting more reducing equivalents to the hydrogenase. Three options are imaginable and might be worthy pursuit.

1. In a two-stage process autotrophic growth in the light could be coupled to dark anaerobic fermentation to produce H_2. This has been suggested for algae (Benemann, 2000; Hallenbeck and Benemann, 2002) and also has been investigated in cyanobacteria (Serebryakova and Tsygankov, 2007).
2. The nitrogen-fixing heterocysts of some filamentous cyanobacteria provide excellent conditions for the production of hydrogen and are in fact the sole biological system that has been used in continuous long-term studies to produce hydrogen (Tsygankov et al., 2002).
3. In terms of efficiency, direct coupling of photolysis of water and hydrogen evolution is the most promising. Although many different cumbersome modifications are necessary to achieve this goal, it is the silver bullet of biohydrogen production if energetic yield is considered.

A. Fermentative Hydrogen Production

Fermentative hydrogen production occurs in many habitats where organic debris is degraded under anoxic conditions. Under normal conditions microbial communities are established in these environments that recycle hydrogen very efficiently so that its production is accompanied by direct consumption by nitrate-, iron-, manganese- and sulfate-reduction and methanogenesis and acetogenesis. Because of the thermodynamic driving force the respective electron acceptors and processes are used in the order given (Lovely and Godwin, 1988; Hoehler et al., 1998). The latter process produces methane and is desirable if biogas is to be produced but it is a strong competitor if high yields of hydrogen have to be achieved. Actually many efforts have been made and are actually underway to increase the yield of hydrogen while biomass is decomposed (Westermann et al., 2007).

Cyanobacteria and other microalgae can be used to generate biomass that is consecutively

used for hydrogen production (Benemann, 2000). The unrivaled growth yield of these organisms (Richmond, 2000) is a strong argument for their use as biofuel producers. However, if the maximum efficiency of 10% for the conversion of solar energy to biomass (Long et al., 2006) is taken into account, and assuming that 60% of this is glucose and all the glucose is converted to two acetate and 4 H_2 (Fig. 15.5), fermentative hydrogen evolution would have a maximal attainable efficiency of around 2% of the originally absorbed sunlight. In fact there are in vitro systems using the oxidative pentosephosphate pathway near equilibrium that obtained stoichiometric levels of 12 H_2 per glucose (Woodward et al., 2000), but these conditions are not feasible for living cells.

From this calculation it is clear that fermentative hydrogen production from cyanobacterial biomass will only play a minor role on the energy market. Nevertheless, calculations based on the economic value might change if coupled to the production of other interesting and valuable compounds like lipids, carotenoids, vitamins or other biofuels so that hydrogen might become an interesting byproduct.

B. Heterocystous Hydrogen Production

Under nitrogen depletion, some filamentous cyanobacteria develop specialized cells. In these heterocysts, an oxygen free environment is provided for the extremely oxygen sensitive nitrogenase, which reduces dinitrogen to ammonia (for a review see Wolk et al., 1994). Because of the strong triple bond of N_2, a high reductive power has to be generated to cleave it. For each electron transferred from the Fe-protein to the MoFe-protein of the nitrogenase two ATP are consumed. In the Mo-dependent nitrogenases, one H_2 is inevitably produced as a byproduct for each N_2 reduced.

The ratio of H_2 produced per N_2 split is even higher in the alternative nitrogenases that either use vanadium or iron instead of molybdenum (Eady, 1996; Henderson, 2005; Peters and Szilagyi, 2006). This is particularly evident in a study on deletion mutants of the three different nitrogenases of the purple bacterium *Rhodopseudomonas palustris* (Oda et al., 2005). Although the nitrogenase activity of the strains only expressing the V-nitrogenase or the Fe-nitrogenase was only around 10% and 1%, their hydrogen evolution was 1.5-fold or more than 4-fold higher than the wild type cells, respectively (Oda et al., 2005).

In general, two options are available for heterocystous hydrogen production. Hydrogen could either be produced by a nitrogenase, or it is taken advantage of the reducing oxygen free environment and a hydrogenase is introduced that is able to accept electrons from the reduced ferredoxin generated in these cells (Schrautemeier and Böhme, 1985).

Increased photohydrogen evolution via the nitrogenase has been accomplished in a number of attempts by either sparging the cultures with N_2-free gas (Benemann and Weare, 1974; Jones and Bishop, 1976; Lambert et al., 1979) or by using mutants of the uptake hydrogenase (Tsygankov et. al., 1999; Masukawa et al., 2002; Yoshino et al., 2007). Some of these studies showed hydrogen production of these cultures for 40 days under outdoor conditions (Tsygankov et al., 2002) and are therefore probably the longest ever conducted under natural conditions for in vivo hydrogen production. Similar studies also investigated the influence of the V-dependent nitrogenase or the disruption of the homocitrate synthase (NifV) on photohydrogen production (Masukawa et al., 2007). Homocitrate is a ligand of the molybdenum ion in the active site of the nitrogenase (Hoover et al., 1987), and it was shown to have an influence on the proportion of produced H_2 and to aid in discriminating H_2 against N_2 (McLean and Dixon, 1981). However, the dissociation of the Fe-protein from the MoFe-protein has a very low rate constant of 6.4 s^{-1} (Burgess and Lowe, 1996) and the high-energy demand of 2 ATP per electron entering the process renders this strategy very inefficient. The energy conversion efficiency was found to be below 1% (Tsygankov et al., 2002; Yoon et al., 2006) and therefore is still far from technical applicability.

An alternative, more energy efficient pathway could be opened by introducing a hydrogenase in the heterocysts that is able to withdraw electrons from the ferredoxin that normally reduces the Fe-protein. Since many of the FeFe-hydrogenases are known to interact with ferredoxin, and since only three accessory genes seem to be necessary for their maturation (Posewitz et al., 2004; King et al., 2006), they are ideal candidates for a heterologous expression in these cells. Thermodynamics favor hydrogen production via the hydrogenase

compared to delivering the low potential electrons to the nitrogenase. Therefore, countermeasures have to be taken such that the cells are either supplied with sufficient nitrogen sources, or are still able to produce enough fixed nitrogen to support low cell growth or minimal cell maintenance.

C. Hydrogen Production Directly Coupled to Photolysis of Water

As was already discussed in Section II.B, there are a number of rivaling processes that compete with direct hydrogen production using the bidirectional hydrogenase. First of all, the Calvin-Bassham-Benson cycle, after its reductive activation, withdraws a large proportion of the reducing equivalents, and photoreduction of oxygen is another strong competitor. The most direct and also promising approach to overcome this competition is coupling of the hydrogenase to the peripheral subunits of photosystem I (PsaC, D and E) at the stromal face of the membrane. Thereby, all the other alternative routes could be blocked and the majority of low potential electrons diverted to hydrogen.

In vitro attempts already have been made with isolated photosystem I incubated in the presence of the FeFe-hydrogenases of *Clostridium pasteurianum* and the NiFe-hydrogenase of *Rhodococcus* sp. MR11 (McTavish, 1998), and most recently with the membrane bound hydrogenase of *Ralstonia eutropha* that has been fused to PsaE of a cyanobacterial photosystem I (Ihara et al., 2006). The latter construct yielded 0.014 mol H_2/mol PSI/s, but was unable to compete with a mixture of ferredoxin, FNR, and $NADP^+$, when added to the solution. Since the hydrogenase was fused to the C-terminus of PsaE, its FeS clusters were quite remote from the F_A and F_B clusters of photosystem I. Nevertheless, since electron transfer was possible, these are encouraging findings for future attempts. The same is true for experiments that used Pt- or Au-nanoparticles tethered to the F_B cluster of PsaC via a thiol. The amount of produced hydrogen was even higher, and amounted to 1.2 mol H_2/mol PSI/s when cytochrome c_6 was used as the electron donor (Grimme et al., 2008; Lubner et al., 2010).

In future attempts, especially the putative docking site of ferredoxin to the stromal face of photosystem I (Fromme et al., 2003) should be accounted for. Thereby, it should be possible to construct hydrogenase-photosystem fusions with a more efficient electron transfer that effectively compete with other electron acceptors in the cell and achieve hydrogen production rates comparable to the maximum turnover (~8000 s^{-1} (Rousseau et al., 1993)) of photosystem I. After transferring these constructs into cyanobacterial cells, biohydrogen evolution might become more attainable, but still a number of other obstacles will need to be addressed for it to become sizeable.

A continuous hydrogen production system using the low potential electrons generated in the light reaction without consuming the ATP will build up a proton gradient that will ultimately impair all further electron transport. Therefore, the proton transfer has to be decoupled from electron transfer by using protonophores or uncoupling proteins (Nedergaard et al., 2001).

In addition, it would be necessary to funnel part of the available excitation energy into the minimum core functions, like repair of existing cellular components. Another option could be to cycle the cells between a production phase and regeneration phase by the use of inducible promoters that switch the expression of the fusion proteins on and off.

Whether or not oxygen inhibition of the NiFe-hydrogenases will still be an issue if the enzyme is connected to photosystem I remains to be seen. The constant supply of reducing equivalents might protect its active site against oxidation. Such a protective role has been suggested for NADPH reducing the active site of the soluble hydrogenase of *R. eutropha* (Burgdorf et al., 2005b).

IV. Summary and Perspective

A large body of new insights has been gained in recent years into the functioning of hydrogenases. Considerable progress has been made in the understanding of the cyanobacterial enzymes and cyanobacterial hydrogen metabolism. Recent biochemical and spectroscopical investigations suggest that the cyanobacterial bidirectional hydrogenase has peculiar properties as compared to other NiFe-hydrogenases and might be a good candidate for biotechnological applications. More undiscovered surprises are certainly ahead, but

the general framework is becoming clear and can now be used for directed and thoughtful improvements. In principle, all biochemical, structural, and functional knowledge in combination with the available genomic information should allow us to construct cellular nanofactories that produce hydrogen for our own use.

Acknowledgments

I am especially grateful for the exceptional support I received from Rüdiger Schulz during the major part of my scientific life and I like to thank all the Ph.D. and diploma students I have been working with in the lab and who were part of exciting and many fruitful discussions. My special thanks are to Horst Senger, Reto Strasser, Georg Schmetterer, Eva-Mari Aro, Laurent Cournac, Teruo Ogawa, Aaron Kaplan, Petra Fromme, Anne Jones and Wim Vermaas. Personally I like to thank Kirstin Gutekunst who added a warm and invaluable character to my life.

Financial aid from DFG, COST 841, Innovationsfond des Landes Schleswig-Holstein, Linde AG, Arizona State University and the Intellectual Fusion Investment Fund for ASU by Brian and Kelly Swette is gratefully acknowledged.

References

Abdel-Basset R and Bader KP (1997) Characterization of hydrogen photoevolution in *Oscillatoria chalybea* detected by means of mass spectrometry. Z. Naturforsch C 52: 775–781

Abdel-Basset R and Bader KP (1998) Physiological analyses of the hydrogen gas exchange in cyanobacteria. J Photochem Photobiol B 43: 146–151

Abdel-Basset R, Spiegel S and Bader KP (1998) Saturation of cyanobacterial photoevolution of molecular hydrogen by photosynthetic redox components. J Photochem Photobiol B 47: 31–38

Adams MWW, Mortenson LE and Chen JS (1980) Hydrogenase. Biochim Biophys Acta 594: 105–176.

Adams MWW (1990) The structure and mechanism of iron-hydrogenases. Biochim Biophys Acta 1020: 115–145

Agervald A, Stensjö K, Holmqvist M and Lindblad P (2008) Transcription of the extended *hyp*-operon in *Nostoc* sp. strain PCC 7120. BMC Microbiol 8: 69

Almon H and Böger P (1988) Hydrogen metabolism of the unicellular cyanobacterium *Chroococcidiopsis thermalis* ATCC 29380. FEMS Microbiol Lett. 49: 445–449

Alonso-Lomillo MA, Rudiger O, Maroto-Valiente A, Velez M, Rodriguez-Ramos I, Munoz FJ, Fernandez VM and De Lacey AL (2007) Hydrogenase-coated carbon nanotubes for efficient H_2 oxidation. Nano Lett 7: 1603–1608

Ananyev G, Carrieri D and Dismukes GC (2008) Optimization of metabolic capacity and flux through environmental cues to maximize hydrogen production by the cyanobacterium *Arthrospira* (*Spirulina*) *maxima*. Appl Environ Microbiol 74: 6102–6113

Appel J and Schulz R (1996) Sequence analysis of an operon of a NAD(P)-reducing nickel hydrogenase from the cyanobacterium *Synechocystis* sp. PCC 6803 gives additional evidence for direct coupling of the enzyme to NAD(P)H-dehydrogenase (complex I). Biochim Biophys Acta 1298: 141–147

Appel J, Phunpruch S, Steinmüller K and Schulz, R (2000) The bidirectional hydrogenase of *Synechocystis* sp. PCC 6803 works as an electron valve during photosynthesis. Arch Microbiol 173: 333–338

Asada K (1999) The water-water cycle in chloroplasts: Scavenging of active oxygens and dissipation of excess photons. Annu Rev Plant Physiol Plant Mol Biol 50: 601–639

Arieli B, Shahak Y, Taglicht D, Hauska G and Padan E (1994) Purification and characterization of sulfide-quinone reductase, a novel enzyme driving anoxygenic photosynthesis in *Oscillatoria limnetica*. J Biol Chem 269: 5705–5711

Badger MR, von Caemmerer S, Ruuska S and Nakano H (2000) Electron flow to oxygen in higher plants and algae: rates and control of direct photoreduction (Mehler reaction) and RubisCO oxygenase. Philos Trans R Soc Lond B Biol Sci 355: 1433–1446

Battchikova N and Aro EM (2007) Cyanobacterial NDH-1 complexes: multiplicity in function and subunit composition. Physiol Plant 131: 22–32

Barz M, Beimgraben C, Staller T, Germer F, Opitz F, Marquardt C, Schwarz C, Gutekunst K, Vanselow KH, Schmitz R, LaRoche J, Schulz R and Appel J (2010) Distribution analysis of hydrogenases in surface waters of marine and freshwater environments. PLoS One 5: e13846

Bauer CC, Scappino L and Haselkorn R (1993) Growth of the cyanobacterium Anabaena on molecular nitrogen: NifJ is required when iron is limited. Proc Natl Acad Sci USA 90: 8812–8816

Belkin S and Padan E (1978a) Hydrogen metabolism in the facultative anoxygenic cyanobacteria (blue-green algae) *Oscillatoria limnetica* and *Aphanothece halophytica*. Arch Microbiol 116: 109–111

Belkin S and Padan E (1978b) Sulfide-dependent hydrogen evolution in the cyanobacterium *Oscillatoria limnetica*. FEBS Lett 94: 291–294

Benemann JR and Weare NM (1974) Hydrogen Evolution by nitrogen-fixing *Anabaena cylindrica* cultures. Science 184: 174–175

Benemann JR (2000) Hydrogen production by microalgae. J Appl Phycol 12: 291–300

Bernhard M, Benelli B, Hochkoeppler A, Zannoni D and Friedrich B (1997) Functional and structural role of the cytochrome b subunit of the membrane-bound hydrogenase complex of *Alcaligenes eutrophus* H16. Eur J Biochem 248: 179–186

Böck A, King PW, Bloeksch M and Posewitz MC (2006) Maturation of hydrogenases. Adv Microb Physiol 51: 1–71

Boison, G; Bothe, H; Hansel, A and Lindblad P (1999) Evidence against a common use of the diaphorase subunits by the bidirectional hydrogenase and by the respiratory complex I in cyanobacteria. FEMS Microbiol Lett 174: 159–165

Boison G, Bothe H and Schmitz O (2000) Transcriptional analysis of hydrogenase genes in the Cyanobacteria *Anacystis nidulans* and *Anabaena variabilis* monitored by RT-PCR. Curr Microbiol 40: 315–321

Buchanan BB and Balmer Y (2005) Redox regulation: A broadening horizon. Annu Rev Plant Biol 56: 187–220

Burgdorf T, Lenz O, Buhrke T, van der Linden E, Jones AK, Albracht SP and Friedrich B (2005a) [NiFe]-hydrogenases of *Ralstonia eutropha* H16: modular enzymes for oxygen-tolerant biological hydrogen oxidation. J Mol Microbiol Biotechnol 10: 181–196

Burgdorf T, van der Linden E, Bernhard M, Yin QY, Back JW, Hartog AF, Muijsers AO, de Koster CG, Albracht SP and Friedrich B (2005b) The soluble NAD^+–Reducing [NiFe]-hydrogenase from *Ralstonia eutropha* H16 consists of six subunits and can be specifically activated by NADPH. J Bacteriol 187: 3122–3132

Burgess BK and Lowe DJ (1996) Mechanism of molybdenum nitrogenase. Chem Rev 96: 2983–3012

Butala M, Zgur-Bertok D and Busby SJ (2009) The bacterial LexA transcriptional repressor. Cell Mol Life Sci 66: 82–93

Carrasco CD, Buettner JA and Golden JW (1995) Programmed DNA rearrangement of a cyanobacterial *hupL* gene in heterocysts. Proc Natl Acad Sci USA 92: 791–795

Carrasco CD, Holliday SD, Hansel A, Lindblad P and Golden JW (2005) Heterocyst-specific excision of the *Anabaena* sp. strain PCC 7120 *hupL* element requires *xisC*. J Bacteriol 187: 6031–6038

Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B and Bork P (2006) Toward automatic reconstruction of a highly resolved tree of life. Science 311: 1283–1287

Cooley JW and Vermaas WFJ (2001) Succinate dehydrogenase and other respiratory pathways in thylakoid membranes of *Synechocystis* sp. strain PCC 6803: Capacity comparisons and physiological function. J Bacteriol 183: 4251–4258

Cournac, L; Mus, F; Bernard, L, Guedeney G, Vignais P and Peltier G (2002) Limiting steps of hydrogen production in *Chlamydomonas reinhardtii* and *Synechocystis* PCC 6803 as analysed by light-induced gas exchange transients. Int J Hydr Energ 27: 1229–1237

Cournac L, Guedeney G, Peltier G and Vignais PM (2004) Sustained photoevolution of molecular hydrogen in a mutant of *Synechocystis* sp. strain PCC 6803 deficient in the type I NADPH-dehydrogenase complex. J Bacteriol 186: 1737–1746

Cracknell JA, Vincent KA, Ludwig M, Lenz O, Friedrich B and Armstrong FA (2008) Enzymatic oxidation of H_2 in atmospheric O_2: the electrochemistry of energy generation from trace H_2 by aerobic microorganisms. J Am Chem Soc 130: 424–425

Dai S, Schwendtmayer C, Schürmann P, Ramaswamy S and Eklund H (1996) Redox signaling in chloroplasts: cleavage of disulfides by an iron-sulfur cluster. Science 287: 655–658

Dietrich LE, Tice MM and Newman DK (2006) The co-evolution of life and Earth. Curr Biol 16: R395–400

Domain F, Houot L, Chauvat F and Cassier-Chauvat C (2004) Function and regulation of the cyanobacterial genes *lexA*, *recA* and *ruvB*: LexA is critical to the survival of cells facing inorganic carbon starvation. Mol Microbiol 53: 65–80

Dross F, Geisler V, Lenger R, Theis F, Krafft T, Fahrenholz F, Kojro E, Duchêne A, Tripier D, Juvenal K and Kröger A (1992) The quinone-reactive Ni/Fe-hydrogenase of *Wolinella succinogenes*. Eur J Biochem 206: 93–102

Eady RR (1996) Structure-function relationships of alternative nitrogenases. Chem Rev 96: 3013–3030

Epping EHG, Khalili A and Thar R (1999) Photosynthesis and the dynamics of oxygen consumption in a microbial mat as calculated from transient oxygen microprofiles. Limnol Ocean 44: 1936–1948

Fernandez VM, Aguirre R and Hatchikian EC (1984) Reductive activation and redox properties of hydrogenase from *Desulfovibrio gigas*. Biochim Biophys Acta 790: 1–-7

Ferreira D, Leitão E, Sjöholm J, Oliveira P, Lindblad P, Moradas-Ferreira P and Tamagnini P (2007) Transcription and regulation of the hydrogenase(s) accessory genes, hypFCDEAB, in the cyanobacterium *Lyngbya majuscula* CCAP 1446/4. Arch Microbiol 188: 609–617

Fromme P, Melkozernov A, Jordan P and Krauss N (2003) Structure and function of photosystem I: interaction with its soluble electron carriers and external antenna systems. FEBS Lett 555: 40–44

Forzi L, Hellwig P, Thauer RK, Sawers RG (2007) The CO and CN^- ligands to the active site Fe in [NiFe]-hydrogenase of *Escherichia coli* have different metabolic origins. FEBS Lett 581: 3317–3321

Frenkel A, Gaffron H and Battley EH (1950) Photosynthesis and photoreduction by the blue green alga, Synechococcus elongatus, Näg. Biol Bull 99: 157–162

Fontecilla-Camps JC, Volbeda A, Cavazza C and Nicolet Y (2007) Structure/function relationships of [NiFe]- and [FeFe]-hydrogenases. Chem Rev 107: 4273–4303

Germer F, Zebger I, Saggu M, Lendzian F, Schulz R and Appel J (2009) Overexpression, isolation and spectroscopic characterization of the bidirectional [NiFe] hydro-

genase from *Synechocystis* sp. PCC 6803. J Biol Chem 284: 36462–36472

Ghirardi ML, Posewitz MC, Maness PC, Dubini A, Yu J and Seibert M (2007) Hydrogenases and hydrogen photoproduction in oxygenic photosynthetic organisms. Annu Rev Plant Biol 58:71–91

Goldet G, Wait AF, Cracknell JA, Vincent KA, Ludwig M, Lenz O, Friedrich B and Armstrong FA (2008) Hydrogen production under aerobic conditions by membrane-bound hydrogenases from *Ralstonia* species. J Am Chem Soc 130: 11106–11113

Grimme, RA; Lubner, CE; Bryant, DA and Golbeck JH (2008) Photosystem I/molecular wire/metal nanoparticle bioconjugates for the photocatalytic production of H_2. J Amer Chem Soc 130: 6308–6309

Gross R, Simon J, Lancaster CR and Kröger A (1998) Identification of histidine residues in *Wolinella succinogenes* hydrogenase that are essential for menaquinone reduction by H_2. Mol Microbiol 30: 639–646

Gross R, Pisa R, Sänger M, Lancaster CR and Simon J (2004) Characterization of the menaquinone reduction site in the diheme cytochrome b membrane anchor of *Wolinella succinogenes* NiFe-hydrogenase. J Biol Chem 279: 274–281

Guedeney G, Corneille S, Cuiné S and Peltier G (1996) Evidence for an association of *ndhB*, *ndhJ* gene products and ferredoxin-NADP-reductase as components of a chloroplastic NAD(P)H dehydrogenase complex. FEBS Lett 378: 277–280

Gupta RS and Griffiths E (2002) Critical issues in bacterial phylogeny. Theor Popul Biol 61: 423–434

Gutekunst K, Phunpruch S, Schwarz C, Schuchardt S, Schulz-Friedrich R and Appel J (2005) LexA regulates the bidirectional hydrogenase in the cyanobacterium *Synechocystis* sp. PCC 6803 as a transcription activator. Mol Microbiol. 58: 810–823

Gutthann F, Egert M, Marques A and Appel J (2007) Inhibition of respiration and nitrate assimilation enhances photohydrogen evolution under low oxygen concentrations in *Synechocystis* sp. PCC 6803. Biochim Biophys Acta 1767: 161–169

Hackstein JH, Tjaden J and Huynen M (2006) Mitochondria, hydrogenosomes and mitosomes: products of evolutionary tinkering! Curr Genet 50: 225–245

Hallenbeck PC and Benemann JR (2002) Biological hydrogen production; fundamentals and limiting processes. Int J Hydr Energ 27: 1185–1193

Happe T and Naber JD (1993) Isolation, characterization and N-terminal amino acid sequence of hydrogenase from the green alga *Chlamydomonas reinhardtii*. Eur J Biochem 214: 475–481

Harbinson J and Hedley CL (1993) Changes in P-700 oxidation during the early stages of the induction of photosynthesis. Plant Physiol 103: 649–660

Helman Y, Tchernov D, Reinhold L, Shibata M, Ogawa T, Schwarz R, Ohad I and Kaplan A (2003) Genes encoding A-type flavoproteins are essential for photoreduction of O_2 in cyanobacteria. Curr Biol 13: 230–235

Helman Y, Barkan E, Eisenstadt D, Luz B and Kaplan A (2005) Fractionation of the three stable oxygen isotopes by oxygen-producing and oxygen-consuming reactions in photosynthetic organisms. Plant Physiol 138: 2292–2298

Henderson RA (2005) Mechanistic studies on synthetic Fe-S-based clusters and their relevance to the action of nitrogenases. Chem Rev 105: 2365–2437

Herrero A, Muro-Pastor AM, Valladares A and Flores E (2004) Cellular differentiation and the NtcA transcription factor in filamentous cyanobacteria. FEMS Microbiol Rev 28: 469–487

Herrero A, Muro-Pastor AM, Flores E (2001) Nitrogen control in cyanobacteria. J Bacteriol. 183: 411–425

Hoehler, TM; Alperin, MJ; Albert, DB, Martens CS (1998) Thermodynamic control on hydrogen concentrations in anoxic sediments. Geochim Cosmochim Acta 62: 1745–1756

Hoehler TM, Bebout B and Des Marais DJ (2001) The role of microbial mats in the production of reduced gases on the early Earth. Nature 412: 324–327

Hoffmann D, Gutekunst K, Klissenbauer M, Schulz-Friedrich R and Appel J (2006) Mutagenesis of hydrogenase accessory genes of *Synechocystis* sp. PCC 6803. Additional homologues of *hypA* and *hypB* are not active in hydrogenase maturation. FEBS J 273: 4516–4527

Hoover TR, Robertson AD, Cerny RL, Hayes RN, Imperial J, Shah VK and Ludden PW (1987) Identification of the V factor needed for synthesis of the iron-molybdenum cofactor of nitrogenase as homocitrate. Nature 329: 855–857

Houchins JP and Burris RH (1981a) Physiological reactions of the reversible hydrogenase from *Anabaena* 7120. Plant Physiol 68: 717–721

Houchins JP, Burris RH (1981b) Comparative characterization of two distinct hydrogenases from *Anabaena* sp. strain 7120. J Bacteriol 146: 215–221

Howarth DC and Codd GA (1985) The uptake and production of molecular hydrogen by unicellular cyanobacteria. J Gen Microbiol 131: 1561–1569

Howitt CA and Vermaas WF (1998) Quinol and cytochrome oxidases in the cyanobacterium *Synechocystis* sp. PCC 6803. Biochemistry 37: 17944–17951

Howitt CA and Vermaas WFJ (1999) The subunits of the NAD(P)-reducing Ni-containing hydrogenase do not act as part of the type-1 NAD(P)H-dehydrogenase in the cyanobacterium *Synechocystis* sp. PCC 6803. In: Peschek GA, Loffelhardt W, Schmetterer G (eds) Phototrophic Prokaryotes, pp 595–608, Kluwer Academic Publishers, New York

Ihara M, Nishihara H, Yoon KS, Lenz O, Friedrich B, Nakamoto H, Kojima K, Honma D, Kamachi T and Okura I (2006) Light-driven hydrogen production by a hybrid complex of a [NiFe]-hydrogenase and the cyanobacterial photosystem I. Photochem Photobiol 82: 676–682

Johnson DC, Dean DR, Smith AD and Johnson MK (2005) Structure, function, and formation of biological iron-sulfur clusters. Annu Rev Biochem 74: 247–281

Jones LW and Bishop NI (1976) Simultaneous Measurement of Oxygen and Hydrogen Exchange from the Blue-Green Alga *Anabaena*. Plant Physiol 57: 659–665

Jørgensen BB, Revsbech NP, Blackburn TH and Cohen Y (1979) Diurnal Cycle of Oxygen and Sulfide Microgradients and Microbial Photosynthesis in a Cyanobacterial Mat Sediment. Appl Environ Microbiol 38: 46–58

Jørgensen BB, Cohen Y and Revsbech NP (1986) Transition from Anoxygenic to Oxygenic Photosynthesis in a *Microcoleus chthonoplastes* Cyanobacterial Mat. Appl Environ Microbiol 51: 408–417

Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirosawa M, Sugiura M, Sasamoto S, Kimura T, Hosouchi T, Matsuno A, Muraki A, Nakazaki N, Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M and Tabata S (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. DNA Res 3: 109–136

King GM (2003) Uptake of carbon monoxide and hydrogen at environmentally relevant concentrations by mycobacteria. Appl Environ Microbiol 69: 7266–7272

King PW, Posewitz MC, Ghirardi ML and Seibert M (2006) Functional studies of [FeFe] hydrogenase maturation in an *Escherichia coli* biosynthetic system. J Bacteriol 188: 2163–2172

Kiss E, Kós PB and Vass I (2009) Transcriptional regulation of the bidirectional hydrogenase in the cyanobacterium *Synechocystis* 6803. J Biotechnol 142: 31–37

Korbas M, Vogt S, Meyer-Klaucke W, Bill E, Lyon EJ, Thauer RK and Shima S (2006) The iron-sulfur cluster-free hydrogenase (Hmd) is a metalloenzyme with a novel iron binding motif. J Biol Chem 281: 30804–30813

Kortlüke C and Friedrich B (1992) Maturation of membrane-bound hydrogenase of *Alcaligenes eutrophus* H16. J Bacteriol 174: 6290–6293

Kuchar J and Hausinger RP (2004) Biosynthesis of metal sites. Chem Rev 104: 509–525

Lambert GR, Daday A and Smith GD (1979) Duration of Hydrogen Formation by *Anabaena cylindrica* B629 in Atmospheres of Argon, Air, and Nitrogen. Appl Environ Microbiol 38: 530–536

Lambert GR and Smith GD (1980) Hydrogen metabolism by filamentous cyanobacteria. Arch Biochem Biophys 205: 36–50

Leitão E, Oxelfelt F, Oliveira P, Moradas-Ferreira P and Tamagnini P (2005) Analysis of the hupSL operon of the nonheterocystous cyanobacterium Lyngbya majuscula CCAP 1446/4: regulation of transcription and expression under a light-dark regimen. Appl Environ Microbiol 71: 4567–4576

Lenz O, Zebger I, Hamann J, Hildebrandt P and Friedrich B (2007) Carbamoylphosphate serves as the source of CN^-, but not of the intrinsic CO in the active site of the regulatory [NiFe]-hydrogenase from *Ralstonia eutropha*. FEBS Lett 581: 3322–3326

Ley RE, Harris JK, Wilcox J, Spear JR, Miller SR, Bebout BM, Maresca JA, Bryant DA, Sogin ML and Pace NR. (2006) Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. Appl Environ Microbiol 72: 3685–3695

Lieman-Hurwitz J, Haimovich M, Shalev-Malul G, Ishii A, Hihara Y, Gaathon A, Lebendiker M and Kaplan A (2008) A cyanobacterial AbrB-like protein affects the apparent photosynthetic affinity for CO_2 by modulating low-CO_2-induced gene expression. Environ Microbiol 11: 927–936

Lindblad P and Sellstedt A (1990) Occurrence and localization of an uptake hydrogenase in the filamentous heterocystous cyanobacterium *Nostoc* PCC 73102. Protoplasma 159: 9–15

Lindblad P, Christensson K, Lindberg P, Fedorov A, Pinto F and Tsygankov A (2002) Photoproduction of H_2 by wildtype *Anabaena* sp. PCC 7120 and a hydrogen uptake deficient mutant: from laboratory experiments to outdoor culture. Int J Hydr Energ 27: 1271–1281

Lissolo, T., S. Pulvin and D. Thomas. (1984) Reactivation of the hydrogenase from *Desulfovibrio gigas* by hydrogen. Influence of redox potential. J Biol Chem 259: 11725–11729

Long, SP; Zhu, XG; Naidu, SL and Ort DR (2006) Can improvement in photosynthesis increase crop yields? Plant Cell Environ 29: 315–330

Lovley DR, and Goodwin S (1988) Hydrogen concentrations as an indicator of the predominant terminal electronaccepting reactions in aquatic sediments. Geochim Cosmochim Acta 52: 2993–3003

Lubitz W, Reijerse E and van Gastel M (2007) [NiFe] and [FeFe] hydrogenases studied by advanced magnetic resonance techniques. Chem Rev 107: 4331–4365

Lubner CE, Knörzer P, Silva PJ, Vincent KA, Happe T, Bryant DA and Golbeck JH (2010) Wiring an [FeFe]-hydrogenase with photosystem I for light-induced hydrogen production. Biochem 49: 10264–10266

Ludwig M, Schulz-Friedrich R and Appel J (2006) Occurrence of hydrogenases in cyanobacteria and anoxygenic photosynthetic bacteria: implications for the phylogenetic origin of cyanobacterial and algal hydrogenases. J Mol Evol 63: 758–768

Ludwig M, Schubert T, Zebger I, Wisitruangsakul N, Saggu M, Strack A, Lenz O, Hildebrandt P and Friedrich B (2009) Concerted action of two novel auxiliary proteins in assembly of the active site in a membrane-bound [NiFe] hydrogenase. J Biol Chem 284: 2159–2168

Ma WM, Deng Y, Ogawa T and Mi HL (2006) Active NDH-1 complexes from the cyanobacterium *Synechocystis* sp strain PCC 6803. Plant Cell Phys 47: 1432–1436

Maeda S, Badger MR and Price GD (2002) Novel gene products associated with NdhD3/D4-containing NDH-1 complexes are involved in photosynthetic CO_2 hydration in the cyanobacterium, *Synechococcus* sp. PCC7942. Mol Microbiol 43: 425–435

Manyani H, Rey L, Palacios JM, Imperial J and Ruiz-Argüeso T (2005) Gene products of the *hupGHIJ* operon are involved in maturation of the iron-sulfur subunit of the [NiFe] hydrogenase from *Rhizobium leguminosarum* bv. viciae. J Bacteriol 187: 7018–7026

Masukawa H, Mochimaru M and Sakurai H (2002) Disruption of the uptake hydrogenase gene, but not of the bidirectional hydrogenase gene, leads to enhanced photobiological hydrogen production by the nitrogen-fixing cyanobacterium *Anabaena* sp. PCC 7120. Appl Microbiol Biotechnol 58: 618–624

Masukawa H, Inoue K and Sakurai H (2007) Effects of disruption of homocitrate synthase genes on *Nostoc* sp. strain PCC 7120 photobiological hydrogen production and nitrogenase. Appl Environ Microbiol 73: 7562–7570

McLean PA and Dixon RA (1981) Requirement of nifV gene for production of wild-type nitrogenase enzyme in *Klebsiella pneumoniae*. Nature 292: 655–656

McTavish H (1998) Hydrogen evolution by direct electron transfer from photosystem I to hydrogenases. J Biochem 123: 644–649

Mehler AH (1951) Studies on reactivities of illuminated chloroplasts. I. Mechanism of the reduction of oxygen and other Hill reagents. Arch Biochem Biophys 33: 65–77

Mehler AH and Brown AH (1952) Studies on reactivities of illuminated chloroplasts. III. Simultaneous photoproduction of and consumption of oxygen studied with oxygen isotopes. Arch Biochem Biophy 38: 365–370

Mentel M and Martin W (2008) Energy metabolism among eukaryotic anaerobes in light of Proterozoic ocean chemistry. Philos Trans R Soc Lond B Biol Sci 363: 2717–2729

Mitsui A and Suda S (1995) Alternative and cyclic appearance of H_2 and O_2 photoproduction activities under nongrowing conditions in an aerobic nitrogen-fixing unicellular cyanobacterium *Synechococcus* sp. Curr Microbiol 30: 1–6

Nedergaard J, Golozoubova V, Matthias A, Asadi A, Jacobsson A and Cannon B (2001) UCP1: the only protein able to mediate adaptive non-shivering thermogenesis and metabolic inefficiency. Biochim Biophys Acta 1504: 82–106

Neuer G and Bothe H (1982) The pyruvate: ferredoxin oxidoreductase in heterocysts of the cyanobacterium *Anabaena cylindrica*. Biochim Biophys Acta 716: 358–365

Nicolet Y, de Lacey AL, Vernède X, Fernandez VM, Hatchikian EC and Fontecilla-Camps JC (2001) Crystallographic and FTIR spectroscopic evidence of changes in Fe coordination upon reduction of the active site of the Fe-only hydrogenase from *Desulfovibrio desulfuricans*. J Am Chem Soc 123: 1596–1601

Oda Y, Samanta SK, Rey FE, Wu L, Liu X, Yan T, Zhou J and Harwood CS (2005) Functional genomic analysis of three nitrogenase isozymes in the photosynthetic bacterium *Rhodopseudomonas palustris*. J Bacteriol 187: 7784–7794

Oliveira P, Leitão E, Tamagnini P, Moradas-Ferreira P and Oxelfelt F (2004) Characterization and transcriptional analysis of *hupSLW* in *Gloeothece* sp. ATCC 27152: an uptake hydrogenase from a unicellular cyanobacterium. Microbiology 150: 3647–3655

Oliveira P and Lindblad P (2005) LexA, a transcription regulator binding in the promoter region of the bidirectional hydrogenase in the cyanobacterium *Synechocystis* sp. PCC 6803. FEMS Microbiol Lett 251: 59–66

Oliveira P and Lindblad P (2008) An AbrB-Like protein regulates the expression of the bidirectional hydrogenase in *Synechocystis* sp. strain PCC 6803. J Bacteriol 190: 1011–1019

Ogawa T (1991) A gene homologous to the subunit-2 gene of NADH dehydrogenase is essential to inorganic carbon transport of *Synechocystis* PCC 6803. Proc Natl Acad Sci USA 88: 4275–4279

Ohkawa H, Pakrasi HB and Ogawa T (2000) Two types of functionally distinct NAD(P)H dehydrogenases in *Synechocystis* sp. strain PCC6803. J Biol Chem 275: 31630–31634

Padan E (1979) Facutative anoxygenic photosynthesis in cyanobacteria. Annu Rev Plant Physiol Plant Mol Biol 30: 27–40

Pandey AS, Harris TV, Giles LJ, Peters JW and Szilagyi RK (2008) Dithiomethylether as a ligand in the hydrogenase h-cluster. J Am Chem Soc 130: 4533–4540

Papen H, Kentemich T, Schmülling T and Bothe H (1986) Hydrogenase activities in cyanobacteria. Biochimie 68: 121–132

Patterson-Fortin LM, Colvin KR and Owttrim GW (2006) A LexA-related protein regulates redox-sensitive expression of the cyanobacterial RNA helicase, *crhR*. Nucleic Acids Res 34: 3446–3454

Patterson-Fortin LM and Owttrim GW (2008) A *Synechocystis* LexA-orthologue binds direct repeats in target genes. FEBS Lett 582: 2424–2430

Peters JW and Szilagyi RK (2006) Exploring new frontiers of nitrogenase structure and mechanism. Curr Opin Chem Biol 10: 101–108

Posewitz MC, King PW, Smolinski SL, Zhang L, Seibert M and Ghirardi ML (2004) Discovery of two novel radical S-adenosylmethionine proteins required for the assembly of an active [Fe] hydrogenase. J Biol Chem 279: 25711–25720

Reissmann S, Hochleitner E, Wang H, Paschos A, Lottspeich F, Glass RS and Böck A (2003) Taming of a poison: biosynthesis of the NiFe-hydrogenase cyanide ligands. Science 299 1067–1070

Richmond A (2000) Microalgal biotechnology at the turn of the millenium: a personal view. J Appl Phycol 12: 441–451

Roseboom W, De Lacey AL, Fernandez VM, Hatchikian EC and Albracht SP (2006) The active site of the [FeFe]-hydrogenase from *Desulfovibrio desulfuricans*. II. Redox properties, light sensitivity and CO-ligand exchange as observed by infrared spectroscopy. J Biol Inorg Chem 11: 102–118

Rousseau F, Sétif P and Lagoutte B (1993) Evidence for the involvement of PSI-E subunit in the reduction of ferredoxin by photosystem I. EMBO J 12: 1755–1765

Schansker G, Srivastava A, Govindjee and Strasser RJ (2003) Characterization of the 820-nm transmission signal paralleling the chlorophyll a fluorescence rise (OJIP) in pea leaves. Funct Plant Biol 30: 785–796

Schansker G, Tóth SZ and Strasser RJ (2005) Methylviologen and dibromothymoquinone treatments of pea leaves reveal the role of photosystem I in the Chl a fluorescence rise OJIP. Biochim Biophys Acta 1706: 250–261

Schansker G, Tóth SZ and Strasser RJ (2006) Dark recovery of the Chl a fluorescence transient (OJIP) after light adaptation: the qT-component of non-photochemical quenching is related to an activated photosystem I acceptor side. Biochim Biophys Acta 1757: 787–797

Schmitz O and Bothe H (1996a) The diaphorase subunit HoxU of the bidirectional hydrogenase as electron transferring protein in cyanobacterial respiration? Naturwissenschaften 83: 525–527

Schmitz O and Bothe H (1996b) $NAD(P)^+$-dependent hydrogenase activity in extracts from the cyanobacterium *Anacystis nidulans*, FEMS Microbiol Lett 135: 97–101

Schmitz O, Boison G and Bothe H (2001a) Quantitative analysis of expression of two circadian clock-controlled gene clusters coding for the bidirectional hydrogenase in the cyanobacterium *Synechococcus* sp. PCC7942. Mol Microbiol 41: 1409–1417

Schmitz O, Gurke J and Bothe H (2001b) Molecular evidence for the aerobic expression of *nifJ*, encoding pyruvate:ferredoxin oxidoreductase, in cyanobacteria. FEMS Microbiol Lett 195: 97–102

Schmitz O, Boison G, Salzmann H, Bothe H, Schütz K, Wang SH and Happe T (2002) HoxE - a subunit specific for the pentameric bidirectional hydrogenase complex (HoxEFUYH) of cyanobacteria. Biochim Biophys Acta 1554: 66–74

Schneider K and Schlegel HG (1977) Localization and stability of hydrogenases from aerobic hydrogen bacteria. Arch Microbiol 112: 229–238

Schrautemeier B and Böhme H (1985) Distinct ferredoxin for nitrogen-fixation isolated from heterocysts of the cyanobacterium *Anabaena variabilis*. FEBS Lett 184, 304–308

Schubert T, Lenz O, Krause E, Volkmer R and Friedrich B (2007) Chaperones specific for the membrane-bound [NiFe]-hydrogenase interact with the Tat signal peptide of the small subunit precursor in *Ralstonia eutropha* H16. Mol Microbiol 66: 453–467

Schreiber U, Endo T, Mi HL and Asada K (1995) Quenching analysis of chlorophyll fluorescence by the saturation pulse method – particular aspects relating to the study of eukaryotic algae and cyanobacteria. Plant Cell Physiol 36: 873–882

Schröder O, Bleijlevens B, de Jongh TE, Chen Z, Li T, Fischer J, Förster J, Friedrich CG, Bagley KA, Albracht SP and Lubitz W (2007) Characterization of a cyanobacterial-like uptake [NiFe] hydrogenase: EPR and FTIR spectroscopic studies of the enzyme from *Acidithiobacillus ferrooxidans*. J Biol Inorg Chem 12: 212–233

Seabra R, Santos A, Pereira S, Moradas-Ferreira P and Tamagnini P (2009) Immunolocalization of the uptake hydrogenase in the marine cyanobacterium *Lyngbya majuscula* CCAP 1446/4 and two *Nostoc* strains. FEMS Microbiol Lett 292: 57–62

Serebryakova LT, Zorin NA and Gogotov IN (1992) Hydrogenase activity in filamentous cyanobacteria. Microbiology 61: 107–112

Serebryakova LT, Zorin NA and Lindblad P (1994) Reversible hydrogenase in *Anabaena variabilis* ATCC 29413 – presence and localization in non-N_2 fixing cells. Arch Microbiol 161: 140–144

Serebryakova LT, Medina M, Zorin NA, Gogotov IN, Cammack R (1996) Reversible hydrogenase of *Anabaena variabilis* ATCC 29413: catalytic properties and characterization of redox centres. FEBS Lett 383: 79–82

Serebryakova LT, Sheremetieva M, Tsygankov AA (1998) Reversible hydrogenase activity of *Gloeocapsa alpicola* in continuous culture. FEMS Microbiol Lett 166: 89–94

Serebryakova LT and Sheremetieva ME (2006) Characterization of catalytic properties of hydrogenase isolated from the unicellular cyanobacterium *Gloeocapsa alpicola* CALU 743. Biochem Moscow 71: 1370–1376

Serebryakova LT and Tsygankov AA (2007) Two-stage system for hydrogen production by immobilized cyanobacterium *Gloeocapsa alpicola* CALU 743. Biotechnol Prog 23: 1106–1110

Shibata M, Ohkawa H, Kaneko T, Fukuzawa H, Tabata S, Kaplan A and Ogawa T (2001) Distinct constitutive and low-CO_2-induced CO_2 uptake systems in cyanobacteria: genes involved and their phylogenetic relationship with homologous genes in other organisms. Proc Natl Acad Sci USA 98: 11789–11794

Shima S, Pilak O, Vogt S, Schick M, Stagni MS, Meyer-Klaucke W, Warkentin E, Thauer RK and Ermler U (2008) The crystal structure of [Fe]-hydrogenase reveals the geometry of the active site. Science 321: 572–575

Sjöholm J, Oliveira P and Lindblad P (2007) Transcription and regulation of the bidirectional hydrogenase in the cyanobacterium *Nostoc* sp. strain PCC 7120. Appl Environ Microbiol 73: 5435–5446

Stal LJ and Moezelaar R (1997) Fermentation in cyanobacteria. FEMS Microbiol Rev 21: 179–211

Summerfield TC, Toepel J and Sherman LA (2008) Low-oxygen induction of normally cryptic *psbA* genes in cyanobacteria. Biochemistry 47: 12939–12941

Tamagnini P, Leitão E, Oliveira P, Ferreira D, Pinto F, Harris DJ, Heidorn T and Lindblad P (2007) Cyanobacterial hydrogenases: diversity, regulation and applications. FEMS Microbiol Rev 31: 692–720

Tel-Or E, Luijk LW and Packer L (1978) Hydrogenase in N_2-fixing cyanobacteria. Arch Biochem Biophys 185: 185–194

Thomann H, Bernardo M and Adams MWW (1991) Pulsed ENDOR and ESEEM spectroscopic evidence for unusual nitrogen coordinaton to the novel H_2-activating Fe-S center in hydrogenase. J Amer Chem Soc 113: 7044–7046

Tichy M and Vermaas W (1999) In vivo role of catalase-peroxidase in *Synechocystis* sp. strain PCC 6803. J Bacteriol 181: 1875–1882

Troshina O, Serebryakova L, Sheremetieva M and Lindblad P (2002) Production of H_2 by the unicellular cyanobacterium *Gloeocapsa alpicola* CALU 743 during fermentation. Int J Hydr Energ 27: 1283–1289

Tsygankov AA, Serebryakova LT, Rao KK and Hall DO (1998) Acetylene reduction and hydrogen photoproduction by wild-type and mutant strains of *Anabaena* at different CO_2 and O_2 concentrations. FEMS Microbiol Lett 167: 13–17

Tsygankov AA, Borodin VB, Rao KK and Hall DO (1999) H_2 Photoproduction by batch culture of *Anabaeana variabilis* ATCC 29413 and its mutant PK84 in a photobioreactor. Biotech Bioeng 64: 709–715

Tsygankov AA, Fedorov AS, Kosourov SN and Rao KK (2002) Hydrogen production by cyanobacteria in an automated outdoor photobioreactor under aerobic conditions. Biotechnol Bioeng 80: 777–783

van Dam PJ, Reijerse EJ and Hagen WR (1997) Identification of a putative histidine base and of a non-protein nitrogen ligand in the active site of Fe-hydrogenases by one-dimensional and two-dimensional electron spin-echo envelope-modulation spectroscopy. Eur J Biochem 248: 355–361

Van der Oost J and Cox RP (1989) Hydrogenase activity in nitrate-grown cells of the unicellular cyanobacterium *Cyanothece* PCC 7822. Arch Microbiol 151: 40–43

Van der Oost J, Bulthuis BA, Feitz S, Krab K and Krayenhof R (1989) Fermentation of the unicellular cyanobacterium *Cyanothece* PCC 7822. Arch Microbiol 152: 415–419

Vignais PM, Dimon B, Zorin NA, Colbeau A and Elsen S (1997) HupUV proteins of *Rhodobacter capsulatus* can bind H_2: evidence from the H-D exchange reaction. J Bacteriol 179: 290–292

Vignais PM, Dimon B, Zorin NA, Tomiyama M and Colbeau A (2000) Characterization of the hydrogen-deuterium exchange activities of the energy-transducing HupSL hydrogenase and H_2-signaling HupUV hydrogenase in *Rhodobacter capsulatus*. J Bacteriol 182: 5997–6004

Vignais PM, Billoud B, Meyer J (2001) Classification and phylogeny of hydrogenases. FEMS Microbiol Rev 25: 455–501

Vignais PM and Billoud B (2007) Occurrence, classification and biological function of hydrogenases: an overview. Chem Rev 107: 4206–4272

Vincent KA, Cracknell JA, Clark JR, Ludwig M, Lenz O, Friedrich B and Armstrong FA (2006) Electricity from low-level H_2 in still air–an ultimate test for an oxygen tolerant hydrogenase. Chem Commun (Camb) 28: 5033–5035

Vogt S, Lyon EJ, Shima S and Thauer RK (2007) The exchange activities of [Fe] hydrogenase (iron-sulfur-cluster-free hydrogenase) from methanogenic archaea in comparison with the exchange activities of [FeFe] and [NiFe] hydrogenases. J Biol Inorg Chem 13: 97–106

Westermann P, Jorgensen B, Lange L, Ahring BK and Christensen CH (2007) Maximizing renewable hydrogen production from biomass in a bio/catalytic refinery int. J Hydr Energ 32: 4135–4141

Weyman PD, Pratte B and Thiel T (2008) Transcription of *hupSL* in *Anabaena variabilis* ATCC 29413 is regulated by NtcA and not by hydrogen. Appl Environ Microbiol 74: 2103–2110

Wolk CP, Ernst A, Elhai J (1994) Heterocyst metabolism and development. In: Bryant DA (ed) The Molecular Biology of Cyanobacteria, pp 69–823. Kluwer Academic Publishers, Dordrecht

Woodward J, Orr M, Cordray K and Greenbaum E (2000) Enzymatic production of biohydrogen. Nature 405: 1014–1015

Yoon JH, Shin JH, Kim MS, Sim SJ and Park TH (2006) Evaluation of conversion efficiency of light to hydrogen energy by *Anabaena variabilis*. Int J Hydr Energ 31: 721–727

Yoshino F, Ikeda H, Masukawa H and Sakurai H (2007) High photobiological hydrogen production activity of a *Nostoc* sp. PCC 7422 uptake hydrogenase-deficient mutant with high nitrogenase activity. Mar Biotechnol (NY) 9: 101–112

Zhang P, Battchikova N, Jansen T, Appel J, Ogawa T and Aro EM (2004) Expression and functional roles of the two distinct NDH-1 complexes and the carbon acquisition complex NdhD3/NdhF3/CupA/Sll1735 in *Synechocystis* sp PCC 6803. Plant Cell 16: 3326–3340

Zilberman S, Stiefel EI, Cohen MH and Car R (2006) Resolving the CO/CN ligand arrangement in CO-inactivated [FeFe] hydrogenase by first principles density functional theory calculations. Inorg Chem 45: 5715–5717

Subject Index

R.L. Burnap and W.F.J. Vermaas (eds.), Functional Genomics and Evolution of Photosynthetic Systems, Advances in Photosynthesis and Respiration 33, pp. 383–393, DOI 10.1007/978-94-007-1533-2,