ADVANCES IN
MOLECULAR AND
CELL BIOLOGY

Volume 22A • 1997

PROTEIN STRUCTURAL BIOLOGY
IN BIOMEDICAL RESEARCH

This Page Intentionally Left Blank

ADVANCES IN MOLECULAR AND CELL BIOLOGY

PROTEIN STRUCTURAL BIOLOGY IN BIOMEDICAL RESEARCH

Series Editor: E. EDWARD BITTAR
                 Department of Physiology
                 University of Wisconsin
                 Madison, Wisconsin

Guest Editors: NORMA M. ALLEWELL
                 CLARE WOODWARD
                 Department of Biochemistry
                 University of Minnesota
                 St. Paul, Minnesota

VOLUME 22A • 1997

CONTENTS (Volume 22A)

This Page Intentionally Left Blank

CONTENTS (Volume 22B)

This Page Intentionally Left Blank

LIST OF CONTRIBUTORS

Norma M. Allewell

Department of Biochemistry
University of Minnesota
St. Paul, Minnesota

Kim-Hung Chow

Department of Laboratory Medicine
and Pathology
University of Minnesota Medical School
Minneapolis, Minnesota

Valerie Daggett

Department of Medicinal Chemistry
University of Washington
Seattle, Washington

Michael L. Doyle

Department of Macromolecular
Sciences
SmithKline Beecham Pharmaceuticals
King of Prussia, Pennsylvania

Linda B.M. Ellis

Department of Laboratory Medicine
and Pathology
University of Minnesota Medical School
Minneapolis, Minnesota

Gregg B. Fields

Departments of Laboratory Medicine
and Pathology, Biochemistry, and the
Biomedical Engineering Center
University of Minnesota
Minneapolis, Minnesota

Bertrand García-Moreno E.

Department of Biophysics
Johns Hopkins University
Baltimore, Maryland

Tuomas Haltia   Helsinki Bioenergetics Group
                                        Institute of Biomedical Sciences
                                        Department of Medicinal Chemistry
                                        University of Helsinki
                                        Helsinki, Finland

Preston Hensley   Department of Macromolecular
                                        Sciences
                                        SmithKline Beecham Pharmaceuticals
                                        King of Prussia, Pennsylvania

Paul M. Horowitz   Department of Biochemistry
                                        University of Texas Health Science
                                        Center
                                        San Antonio, Texas

Steven L. Kazmirski   Department of Medicinal Chemistry
                                        University of Washington
                                        Seattle, Washington

Richard H. Lathrop   Artificial Intelligence Laboratory
                                        Massachusetts Institute of Technology
                                        Cambridge, Massachusetts

Kevin H. Mayo   Departments of Laboratory Medicine
                                        and Pathology, Biochemistry, and the
                                        Biomedical Engineering Center
                                        University of Minnesota
                                        Minneapolis, Minnesota

Michael P. McCarthy   MedImmune, Inc.
                                        Germantown, Maryland

Raman Nambudripad   BioMolecular Engineering Research
                                        Center
                                        Boston University
                                        Boston, Massachusetts

Himanshu Oberoi   Belmont Research Incorporated
                                        Cambridge, Massachusetts

Lynne Regan Department of Molecular Biophysics and Biochemistry
Yale University
New Haven, Connecticut

Irina M. Russu Department of Molecular Biology and Biochemistry
Wesleyan University
Middletown, Connecticut

Michael J. Rynkiewicz Department of Physiology
Boston University School of Medicine
Boston, Massachusetts

Barbara A. Seaton Department of Physiology
Boston University School of Medicine
Boston, Massachusetts

William Shalongo Department of Biochemistry
University of Iowa
Iowa City, Iowa

Earle Stellwagen Department of Biochemistry
University of Iowa
Iowa City, Iowa

Collin M. Stultz Committee on Higher Degrees on Biophysics
Harvard University
Cambridge, Massachusetts

James V. White TASC, Inc.
Reading, Massachusetts

This Page Intentionally Left Blank

PREFACE

Recent advances in protein structural biology, coupled with new developments in human genetics, have opened the door to understanding the molecular basis of many metabolic, physiological, and developmental processes in human biology. Medical pathologies, and their chemical therapies, are increasingly are being described at the molecular level. For single-gene diseases, and some multi-gene conditions, identification of highly correlated genes immediately leads to identification of covalent structures of the actual chemical agents of the disease, namely the protein gene products. Once the primary sequence of a protein is ascertained, structural biologists work to determine its three-dimensional, biologically active structure, or to predict its probable fold and/or function by comparison to the data base of known protein structures. Similarly, three-dimensional structures of proteins produced by microbiological pathogens are the subject of intense study, for example, the proteins necessary for maturation of the human HIV virus. Once the three-dimensional structure of a protein is known or predicted, its function, as well as potential binding sites for drugs that inhibit its function, become tractable questions. The medical ramifications of the burgeoning results of protein structural biology, from gene replacement therapy to "rational" drug design, are well recognized by researchers in biomedical areas, and by a significant proportion of the general population. The purpose of this book is to introduce biomedical scientists to important areas of protein structural biology, and to provide an insightful orientation to the primary literature that shapes the field in each subject.

The chapters in this volume cover aspects of protein structural biology which have led to the recognition of fundamental relationships between protein structure

and function. The first two chapters are concerned with identification and design of specific ligand binding sites. Chapter 1, "Protein Crystallography in Medicine" describes medically-relevant proteins for which crystal structures have provided targets for drug design. These include essential proteins of human HIV-1, influenza, rhino and polio viruses, as well as pathogenic bacteria. The chapter also covers protein crystal structures which explain the molecular action of drugs commonly used in chemical therapies, including specific proteins involved in blood coagulation, glaucoma, cancer, and inflammation. While protein-nucleic acid interactions and antibody-related structures are not included for lack of space, landmark structures of hormone and cytokine-receptor complexes, as well as calcineurin complexes, are noted. Chapter 2, "Engineering Metal-Binding Sites in Proteins," introduces strategies used in design of proteins with specified binding capabilities, and provides details of the approaches and results for one category of these, metal binding sites. New functions of designed metal binding sites include affinity purification tags, protein/peptide stabilization, and enzyme regulation. In design of metal binding sites, submolecular structural motifs associated with naturally occurring metal binding sites, for example modified zinc finger peptides, have been used as starting points. Alternatively, new metal binding sites have been designed into proteins as diverse as four-helix bundles and antibodies.

The development of computer based molecular graphics during the last decade has changed the way structural biologists think about proteins. In chapter 3, "Using Molecular Graphics to Analyze Protein Structures," the most useful depictions of protein structure are reviewed. Since proteins contain thousands of atoms, many components of protein structure are not readily comprehended by inspection of its full three-dimensional structure. To eliminate the visual and conceptual clutter of a full protein structure, individual aspects of protein structures can be emphasized by graphics techniques designed to highlight specific properties such as solvent accessible surface, electrostatic fields, packing of interior residues, and active site geometries. Perception of the three-dimensional relationship of important submolecular units is enhanced by the ability of the viewer to rotate the structures on the computer screen in real time. Graphics techniques also permit us to explore the four-dimensional structure, how the conformation changes with time, for example, in dynamics simulations. Advances in molecular graphics provide not only useful visual summaries of the results of calculations, but also the insight for new questions and experimental directions.

Electrostatic interactions are integral components of protein structure. Hydrogen bonds, one variety of electrostatic interactions, are essential features of secondary and tertiary structure. Interactions between charged groups are important in ligand binding specificity, catalytic activity, solubility, and stability of proteins. Chapter 4, "Electrostatics and Hydrogen Bonding in Proteins," discusses the physical character of electrostatic interactions in proteins, the various computational methods for calculation of electrostatic properties of proteins, the distribution of hydrogen bonds in proteins and their contribution to protein folding, and discussion of

experiments that have demonstrated of the role of electrostatic effects in protein function. Hydrophobic microenvironments, created within the matrix of the protein by close packing of apolar groups, have special effects on electrostatic interactions in proteins, and these are explained in Chapter 4.

Nuclear magnetic resonance techniques (NMR) provide the only means for determination of the atomic level structure of proteins in solution. Chapter 5, "Protein Structure Determination by Nuclear Magnetic Resonance Spectroscopy," describes the recent advances in NMR spectroscopy that underlie the surge of interest in, and output of, NMR protein structures. Basic concepts of NMR are introduced, and new applications of heteronuclear NMR spectroscopy are described. The development of heteronuclear spectroscopy, or three-dimensional NMR, is coupled with the availability of high field instrumentation, of ^{15}N- and ^{13}C-labeled proteins, and of increasingly standardized spectroscopic software. At present, size, solubility and aggregation state of the protein limit the application of the method to soluble monomers of about 25,000 D. However, the advent of even higher field spectrometers, along with advances in specific ^{15}N-, ^{13}C-, and ^{2}H-labeling of proteins are rapidly pushing the limits toward larger proteins. It is expected that NMR structure determination will become an application, rather than a dedicated field, for many in the next generation of protein structural biologists.

Membrane proteins are an important subclass of proteins that are traditionally refractory to structural techniques developed for soluble proteins. Recent breakthroughs in solving the structure of membrane proteins are discussed in Chapters 6 and 7, "Membrane Protein Structure" and "Structural Features of Membrane Proteins." Membrane proteins mediate many essential cellular functions, including energy transduction, cell signaling, hormone action, and membrane transport. The structural biology of membrane proteins, which underlie their specific interactions with phospholipid bilayers, also reduce their solubility in aqueous media and set up special problems for structure determination. Advances in purification and preparation of membrane proteins have permitted the determination of several canonical structures by X-ray crystallography techniques or by electron microscopic reconstruction methods. The structures of all membrane proteins known at the time of this writing are illustrated and discussed in Chapters 6 and 7.

Protein activity is often regulated by higher order interactions with other macromolecules, especially other proteins. Chapter 8, "Experimental Dissection of Protein-Protein Interactions in Solution" describes many of the most useful methods for characterizing energetics and kinetics of quaternary structural interactions in proteins. Experimental techniques for measurement of affinities and stoichiometries, and of quaternery effects on protein functional properties, are covered, along with methods for quantitative analysis of the data. The chapter includes a practical guide to the use of these methods, and illustrates the synergistic power of combinations of techniques.

In the mid- 1970's, molecular dynamics stimulated a conceptual shift in the general thinking of biologists and biochemists about protein structure. The remark-

able complexity of protein crystal structures, along with the stunning, complementary "fit" of closely packed interior atoms and the beautiful juxtaposition of active site moieties, tended to reinforce a rather static picture of protein structure. Although crude by today's standards, the first simulations of protein molecular dynamics trajectories tipped the generally held intuitive grasp of protein structure toward an appreciation of their four-dimensional structure. Proteins are flexible, and this property, no doubt, is associated with many of their unique functions. Internal motions within proteins occur on many time scales; molecular dynamics gives an atomic level depiction of the motions that occur in the nanosecond regime. Chapter 9, "Protein Dynamics: A Theoretical Perspective" includes the basic theory of molecular dynamics, along with a discussion of the practical considerations of how to choose the magnitude of adjustable parameters, and how to distinguish the properties and components of "good" simulations. This discussion is especially timely because many structural biologists wish to evaluate the utility of commercial packages of molecular dynamics programs, and to avoid mistakes that often accompany a "turn-key" approach to them. Chapter 9 includes a review of the applications of molecular dynamics simulations to problems of molecular recognition and conformational transitions in proteins.

The promise of a predictive algorithm for estimating protein tertiary structure from protein primary sequence has been a bright hope of protein chemists for decades, and has never been more intense than it is today with the flood of gene sequences flowing into the data banks. In Chapters 10 and 11, "Protein Structure Prediction from Primary Sequence" and "Predicting Protein Structure with Probabilistic Models," the presently available automatic and semi-automatic predictive methods are described. Although the ultimate goal of tertiary structure from covalent structure alone still remains elusive, there are a number of computerized methods for prediction of secondary structure, of fold "family," and of function-related submolecular motifs. Chapter 10 describes the theories on which various methods are based, examples of how protein primary/tertiary relationships are encoded in computer language, the relevant programs that are commercially available for protein structure prediction, and a critical analysis of the utility of the programs to structural biologists. A review of the recent applications and methodological advances is also presented in chapter 10, along with a highly useful glossary of terms in common usage in this field. Chapter 11 expands the discussion on the mathematical basis for probabilistic methods, and contains an in-depth discussion of "threading," perhaps the most useful method to date for identifying the tertiary fold of a protein from its primary sequence.

Numerous experimental methods have been designed to elucidate the mechanism of protein folding. The protein folding field is inherently interdisciplinary, and its approaches range broadly. They include: folding thermodynamics studies, and efforts directed to identification and quantification of partial processes that contribute to the relative stabilities of folded versus unfolded states; studies designed to identify structures that form early in folding, and whether they are

native-like or non-native; kinetics studies from which time constants and amplitudes of reaction phases are characterized; molecular biological studies designed to probe the tolerance of various three-dimensional positions to amino acid substitutions; and assisted folded, or chaperone, systems in which the correct folding of one protein requires the participation of another. In Chapters 13 and 14, "Model Simulations of the Kinetics of Protein Folding" and "Peptides as Models in Protein Folding Studies," two approaches to protein folding are discussed. Chapter 13 contains a detailed analysis of protein folding progress curves, traces of the kinetic record of any property that distinguishes folded from unfolded states. This analysis of kinetic phases in protein folding kinetic data is a basic question, since many researchers gather these kinds of data, and the extraction of specific rate constants and amplitudes involves a number of complexities and consequent problems of a fundamental nature. Chapter 14 covers more structural approaches to protein folding, specifically those that involve solid phase chemical synthesis of peptides and proteins. Oligopeptides prepared by chemical synthesis have been utilized to characterize the propensity of short stretches of amino acids to form secondary structure; the sequence specificity of the structure is probed by systematic replacement of residues all along the peptide. In a different approach, *de novo* designed proteins have been chemically synthesized in efforts to test whether principles gleaned from many types of protein folding studies can be incorporated into a consciously designed sequence. The test of the design is whether the protein in aqueous solution at physiological pH acquires a three-dimensional structure with the hallmarks of a naturally occurring protein, namely, a single low energy conformation with tight packing and a cooperative folding/unfolding transition. Chapter 14 reviews the efforts in this area to date, and gives a critical assessment of future directions.

The chapters in this book by no means represent comprehensive coverage of important topics in protein structural biology. Obvious omissions include protein crystal structure determination, the hydrophobic effect in protein structure/function, amide hydrogen exchange, and protein folding thermodynamics, to name a few. The selection of topics was guided by the limitations of space and a tendency not to include topics which already receive very high exposure in the review and essay literature. Special thanks are due to the authors who have contributed their ideas and work to this volume.

Clare Woodward
Professor of Biochemistry
University of Minnesota

This Page Intentionally Left Blank

PROTEIN CRYSTALLOGRAPHY IN MEDICINE

Michael J. Rynkiewicz and Barbara A. Seaton

I. INTRODUCTION

Visualization of molecular structure at the atomic level—a goal attainable through X-ray crystallography—already has proven its usefulness in addressing problems in clinical medicine. The crystal structures of biologically important "small molecules" were first steps toward fulfilling the promise of this technique. Dorothy Crowfoot Hodgkins earned a Nobel Prize in 1965 for determining the crystal structures of penicillin and cobalamin (vitamin B_{12}). Nowadays, such structures are routinely obtained for drugs and important endogenous compounds. These data help to explain the structural basis for their biological action and form a basis for identifying or designing molecules with more desirable characteristics. With the growing use of macromolecular crystallography, particularly of proteins and nucleic acid, obvious medical applications emerged. Two significant protein examples are those of hemoglobin and insulin, solved in the laboratories of Max Perutz and Dorothy Hodgkins, respectively. The molecular basis of sickle-cell anemia has been established largely through crystallographic comparison of normal and mutant forms of hemoglobin, and drug development strategies have arisen based on this knowledge. Similarly, improvements have been engineered into insulin to give it better pharmacological properties.

The number of examples is sufficiently great that Max Perutz, writing on the applications of protein structure to medicine, comments in the preface of his book that he set out to write a chapter and ended up with an entire volume (Perutz, 1992). His book offers a more comprehensive view of the subject than can be attempted here and is recommended for further reading. There are now many excellent reviews, specifically focused on structure-based or rational drug design, to which the interested reader is referred for detailed information. In this chapter, we have selected a few clinical areas—infectious disease, thrombosis, glaucoma, cancer, and inflammation—to illustrate how structural, biological, chemical, and clinical knowledge can be combined to produce often spectacular results in the understanding and treatment of human disease. Such synergy is not always attainable, however, and we have tried to give a sense of the biological or pathological conditions that preclude even the best crystallographic approach. Though the

subject matter is inherently interdisciplinary, this presentation is aimed at readers with familiarity of general biochemistry. In view of the huge number of publications involved, the literature cited herein is representative rather than comprehensive but should be helpful in referring readers to additional sources.

II. METHODS

The most direct use of X-ray crystallography in clinical medicine to date has been structure-based drug design. Although there are a variety of programs and techniques available, the underlying methodology is the same. There are several reviews of the procedure published (Navia and Murcko, 1992; Greer et al., 1994; Verlinde and Hol, 1994; Whittle and Blundell, 1994). This section presents an overview of the process. The value of structure-based drug design over random screening methods will be discussed, followed by an overview of the method, and finally a discussion of some of the techniques and programs used.

The random screening method entails the use of many compounds in the hope that at least one of them will act as an effective drug. There are problems inherent in this method. First, one is limited by pool bias. For instance, if a compound library consists primarily of nucleoside analogues, the resulting drug is likely to resemble a nucleoside analogue, thus missing a better drug that is not close in structure. This is becoming less of a problem as compound libraries are getting larger. The second problem is that the compound developed may have cross-reactivity with many other targets. Again, if one makes a nucleoside analogue, the analogue is likely to react with a number of nucleoside binding proteins and may not specifically hit the target. This leads to unwanted side effects. The last problem with random screening methods is that the number of compounds that need to be assayed to refine a lead is very large. Hundreds of compounds may need to be tested before one is found that acts as an effective drug for clinical studies.

Structure-based design avoids these problems, because more information is known. The compounds that are evolved are limited by the designer and by ease of synthesis, so that just about any structure can result. Since the compound made is designed specifically to fit one target, less cross-reactivity may be seen. And lastly, the number of compounds used is very small, because the target is well characterized and the method of binding understood so that less time is wasted trying to fit the wrong peg into the wrong hole.

Before a program of structure-based drug design can be initiated, certain pieces of information need to be gathered. The first is a target. This target comes from an understanding of the basic mechanism by which a disease is caused. Normally, this is a protein whose biological function could be controlled or blocked by a drug. The next piece is the three-dimensional structure of the target. This may be obtained by X-ray crystallography, nuclear magnetic resonance, or homology to other known structures. As much information about the target as possible should be gathered.

Since proteins are complex, dynamic molecules, determining what will bind to a target is a difficult question. The structure should be high resolution, showing as much fine detail as possible. In addition, the reaction of the target protein should be well characterized and the mechanism fairly well understood. Mechanistic information can be exploited to design a better ligand, e.g., one that is based on a known substrate or reaction intermediate. Finally, the structure of the target complexed with a substrate or known inhibitor should be solved, because proteins can undergo major structural rearrangements upon binding, and characterizing these rearrangements is crucial when designing a ligand that will best fit to the target.

The basic method of structure-based drug design is iterative, beginning with a lead compound. For a given protein target, an initial lead is generated from the literature, a computer-generated random screening, *de novo* design, or previously discovered drugs. Literature-derived leads may arise from mechanistic studies of the target or a closely-related protein. In random screening, the program DOCK can be used to test the binding of compounds from a library to the target surface, and the best binding compounds from the screening process are selected for potential leads. For *de novo* design, binding sites for small molecules can be calculated and then these can be linked together manually or by computer to generate hypothetical compounds for use as leads. Lastly, the lead can come from a previously discovered drug that may benefit from improvement in stability, efficacy, tolerance, or other parameters.

The lead compounds are then synthesized and tested for biological activity. If any compounds seem promising, further improvements can be guided by X-ray crystallographic data to show the drug-target interactions. This step is followed by analysis of the binding site and design of new ligands whose properties should improve upon the first by virtue of formation of better bonds or ability to fill the binding pocket better. This analysis can be done visually on a graphics system by the chemist and then a computer check of the binding of the new compound by molecular modeling can predict the effect of such a change on the binding. Estimation of solvent effects and entropy are major problems in the accurate ranking of compounds.

At each turn of the cycle, a few questions about the direction of the drug development can be easily answered. These are questions such as why a certain change caused an increase in binding, why an expected increase in binding did not occur, and what other improvements could be made that are likely to improve the efficacy of the compound as a drug.

Many computer programs have been developed or are available for use in structure- based drug design. There are programs designed to determine preferred binding sites on proteins for specific functional groups, e.g., GRID (Goodford, 1985), MCSS (Miranker and Karplus, 1990), and LUDI (Bohm, 1992). These programs are helpful when making improvements to an existing lead compound. The program DOCK (Leach and Kuntz, 1992) is useful for computer searches of

compound databases. There are also many programs for the *de novo* generation of ligands using a starting point (such as one atom of a known inhibitor-target structure) and building up a new ligand. The building procedure is varied. HOOK (Eisen et al., 1994), LUDI (Bohm, 1992), GenStar (Rotstein and Murcko, 1993a), GroupBuild (Rotstein and Murcko, 1993b), and LEGEND (Nishibata and Itai, 1993) all generate additional structures based on linking functional groups into the best chemical contacts. CONCEPTS (Pearlman and Murcko, 1993) and MCDNLG (Gehlhaar et al., 1995) use random methods to build on a structure using random variations in the added structures followed by selection of the lowest energy ligands. This list is not complete, there are many other programs available, but it offers a taste of what can be done. While software-based approaches can suggest potential lead compounds, a subsequent major problem can be choosing the best one. Too often, computer predictions do not match experimental results. As a consequence, if only the top-ranked computer-based compound is chosen, a vital lead may be missed. The ideal situation in which the designer plugs in a protein and gets out one best drug simply does not exist.

The drug design process requires input from several disciplines, and a potential drug must fulfill a daunting set of criteria. Even the most sophisticated structure-based design process cannot guarantee that the compounds being generated will be synthetically feasible on both a small and an industrial scale, will have no adverse side-effects, and will be biologically available as well as active. Relatively few drug candidates actually survive the substantial leap from in vitro to in vivo conditions.

III. DEVELOPMENT OF ANTIVIRAL THERAPEUTICS

Viruses are the smallest class of pathogens that represent a significant, widespread cause of human disease. They contain a central core of nucleic acid, the infectious component, and an outer covering of protein. Viruses can only reproduce in host cells, and their life cycles are intimately linked to those of host cells. Viral infection typically evokes an antibody response to the viral coat proteins, which is the basis of many vaccines. However, a number of viruses, particularly retroviruses, mutate their coat proteins sufficiently quickly to render a coat-directed vaccine ineffective very quickly. For these viruses, drugs that target a specific stage of the pathogen's life cycle present a more promising approach.

A. Human Immunodeficiency Virus, HIV-1

The spectrum of efforts led by pharmaceutical companies in combating HIV-1 (human immunodeficiency virus) give a excellent example of antiviral drug design. While the search for an effective acquired immune deficiency syndrome (AIDS) vaccine for the virus continues, research has focused on inhibiting enzymes that act at critical stages of the virus life cycle. AIDS is caused by HIV, which

preferentially infects T4 lymphocytes, a subset of T cells which contain CD4 cell surface antigens. These T4 cells, also called inducer/helper cells, act together with T8 lymphocytes or suppressor cells to respond to infection. However, upon encountering HIV-1 particles, the T4 cells are themselves infected as are monocyte/macrophages and other cells of the immune system. Infected cells fuse together to form a giant multinucleated mass, and the cells die. In time, HIV-1 infection leads to the breakdown of the entire immune system, principally through loss of function T4 lymphocytes.

HIV-1 infection (Figure 1) begins when contact is made between the viral envelope protein, gp120, and the T lymphocyte receptor, CD4. This event leads to fusion of the virus particle with the host cell membrane. Once this fusion occurs, the contents of the central core of the virus, mostly genetic material, enters the cytoplasm of the host cell and begins replication. Processes leading up to replication are referred to as early stage. Late stage events include viral RNA replication, gene transcription into mRNA, and translation of gene products. Replication may occur at extrachromosomal sites or in the nucleus, where the viral DNA becomes part of the host genome. Viral gene products are produced in the cytoplasm as polyproteins, transported to the cell surface, and cleaved into functional proteins by proteases. Viral enzymes are cleaved, and thus activated, by the viral-coded HIV protease, and envelope proteins are cleaved by host proteases. Without cleavage by the HIV protease, new viral particles are immature and noninfectious.

Several steps of the HIV life cycle have presented opportunities for drug development. The enzymes that have received the greatest attention are the HIV protease (HIV-PR) and reverse transcriptase. HIV and other retroviruses use the latter enzyme to convert viral RNA into DNA in the host cytoplasm. More recently, integrase, which integrates viral DNA into the host genome, has emerged as a potential target for drug development. CD4, the T cell surface target of HIV, is discussed in a later section on inflammation, though this protein is considered an anti-HIV drug target as well.

HIV Protease

The protease from HIV, like that of Rous sarcoma virus (Miller et al., 1989) is a homodimeric aspartic protease. As such it is structurally related to pepsin, with the same Asp-Thr-Gly active-site triad. The first HIV-PR structure, that of the recombinant enzyme, was solved by a group from Merck (Navia et al., 1989). The monomer is comprised mostly of β-sheet, and the dimer possesses C2 symmetry (Figure 2). The active site is located between the two monomers. A β-hairpin, or "flap" from each monomer folds over the active site and is involved in substrate/inhibitor binding. When no ligand is bound, these flaps are otherwise in an "open" conformation and highly flexible. Another notable structural feature is the so-called flap water molecule, which is completely solvent-inaccessible and which bridges the bound ligand and enzyme flaps.
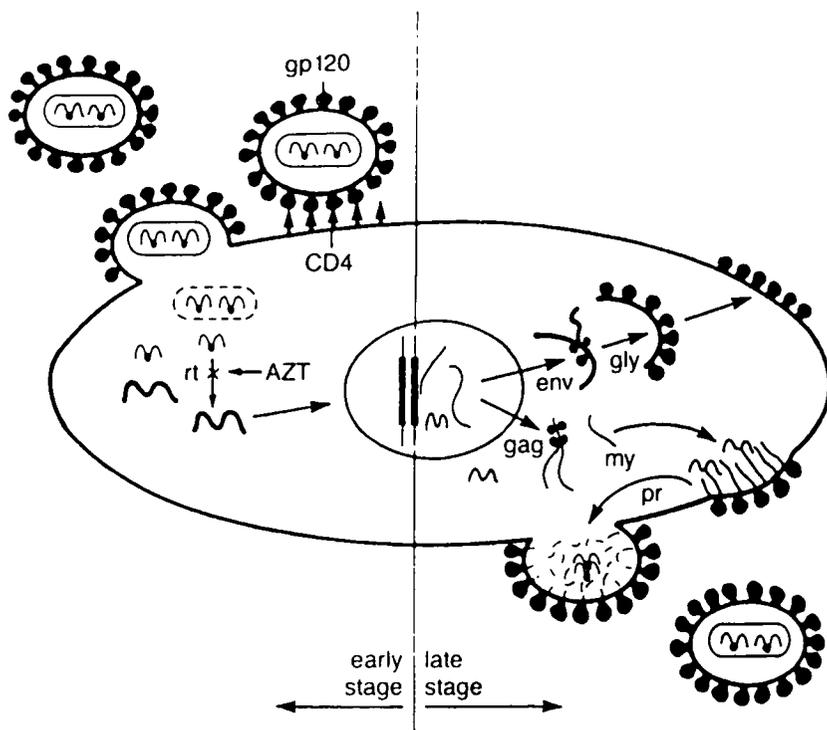
Figure 1. HIV life cycle. Early stage events: the viral life cycle is initiated when the viral envelope protein (gp120) binds to a receptor (CD4) on the host cell surface. A series of as yet undefined events leads to the fusion of the viral and cell membranes followed by the release of core components into the cytoplasm, where viral genomic RNA is converted into DNA by reverse transcriptase/RNaseH (rt) and integrated into the celluar DNA. Late-stage events: Replication of HIV genomic RNA. Transcription of viral genes into mRNA and translation of the mRNA into viral polyproteins. Env(gp160) maturation (env) involves membrane-associated synthesis with glycolsylation (gly) and host-mediated proteolysis resulting in the association of gp120/41 with the cellular membrane; gag (Pr55^{gag} and Pr160$^{gag-pol}$) maturation (gag) involves cytoplasmic synthesis with amino-terminal myristylation (my) resulting in association with the cellular membrane; gag polyprotein processing is catalyzed by a virus-encoded protein gp120. The two env surface proteins are noncovalently associated at the viral membrane. Reprinted with permission from Petteway et al., 1991.

At present, there are hundreds of HIV-PR crystal structures, predominantly of enzyme-inhibitor complexes and in pharmaceutical labs. The search for new lead compounds has drawn from several areas of structural knowledge about the enzyme

Figure 2.   HIV-1 protease. The active dimer is shown in the open conformation. Asp 25 is highlighted in each monomer to display the location of the active site. (PDB # 1HHP from Spinelli et al., 1991)

(an excellent review is presented by Wlodawer and Erickson, 1993). Inhibitors have been designed based on mechanistic considerations (transition-state mimetics), active-site C2 symmetry, and even the conserved flap water molecule (Erickson et al, 1990; Roberts et al., 1990; Norbeck and Kempf, 1991; Kim et al., 1995).

HIV Reverse Transcriptase

The most widely used target for drug therapy in HIV has been the HIV-reverse transcriptase (HIV-RT). In fact, the only clinically available drugs against HIV are targeted against HIV-RT. This enzyme is responsible for converting the viral genome from RNA into double stranded DNA for integration into the host DNA. Improvement of these existing drugs will be helped a great deal through X-ray crystallography. The structure of HIV-RT unliganded (Rodgers et al., 1995), complexed with double stranded DNA (Jacobo-Molina et al., 1993), and complexed with nevirapine (1) (Figure 3), an inhibitor (Kohlstaedt et al., 1992; Smerdon et al., 1994) have been determined. The structure of HIV-RT is a heterodimer consisting of a catalytic domain, p66, and a second subunit, p51. The p66 domain consists of a polymerase domain, a connection domain, and a RNase H domain. The polymerase domain has the shape of a right hand holding a rod, the fingers, palm and thumb representing the protein (this is how the respective domains are named) and

the rod representing the DNA strand. The p51 domain is a truncated version of the p66 domain, lacking the RNase H domain and having the remaining connection, fingers, palm, and thumb domains rearranged in relation to one another.

There are two main classes of HIV-RT inhibitors, nucleoside analogues and non-nucleoside analogues. AZT (3'-azido-3'-deoxythymidine) (3), ddI (dideoxyinosine) (4) and ddC (dideoxycytosine) (5) are three examples of nucleoside analogues (Figure 3). These molecules resemble normal nucleotides and are incorporated into the growing DNA strand, however, the strand can not be elongated past the inhibitor. From the structure of HIV-RT complexed to DNA and a mercurated nucleotide analogue, the binding site for these molecules can be modeled in at the 3' end of the primer strand (Arnold et al., 1992). The non-nucleoside analogues are typified by Nevirapine (1) and TIBO (2) (Figure 3). The crystal structure of nevirapine complexed with HIV-RT has been solved (Kohlstaedt et al., 1992; Smerdon et al., 1994). The inhibitor binds in a pocket formed by two tyrosines and a tryptophan located in the palm domain just under the catalytic site. This class of inhibitors is thought to work in a non-competitive fashion to inhibit DNA synthesis by HIV-RT, by displacing the residues in the hydrophobic pocket, causing a rotation of the tyrosines upon binding that leads to a distortion of the thumb domain (Rodgers et al., 1995). The non-nucleoside inhibitors tend to show a lower level of cytotoxicity than the nucleoside analogues, potentially due to less cross reactivity with endogenous polymerases. An interesting note to the binding sites for both classes of inhibitors is that in the unliganded structure, a shift in the thumb domain of 32-34° relative to its position in the complex with either Nevirapine or DNA (Rodgers et al., 1995) closes the cleft where the DNA binds.

A main problem thus far with treatment of AIDS with these HIV-RT inhibitors is that resistant strains of the virus arise quite quickly. However, many of these mutations have been mapped out and a structural analysis done (for a review, see Tantillo et al., 1994). As one might suspect, most of these mutations line up in the corresponding binding sites for the inhibitor, directly affecting the binding. With the crystal structure of the DNA/HIV-RT complex and the mutational map, the door is open for rational design of better inhibitors of HIV-RT that avoid the mutational inactivation faced by AZT, nevirapine, and the others.

HIV Integrase

The viral integrase of HIV-1 represents another potential drug target for combatting the AIDS infection. The crystal structure of the catalytically active core domain, residues 50- 212, of HIV-1 integrase has been determined (Dyda et al., 1994). The overall topology of this domain, that of a five-stranded β-sheet flanked by α-helices, relates it into the superfamily of polynucleotidyl transferases that includes both the ribonuclease H and connection domain of reverse transcriptase. Three enzymatic reactions are catalyzed by integrase, 3' processing viral strand transfer, and disintegration. During processing, the two strands of viral DNA made

Nevirapine (1)

ddC (5)

Clavulanic acid (8)

Pyrimethamine (12)

TIBO (2)

Sialic acid (6)

Sulbactam (9)

Isoniazid (13)

AZT (3)

2-Deoxy-2,3-didehydro-N-acetyl neuraminic acid (7)

Tazobactam (10)

Acetazolamide (14)
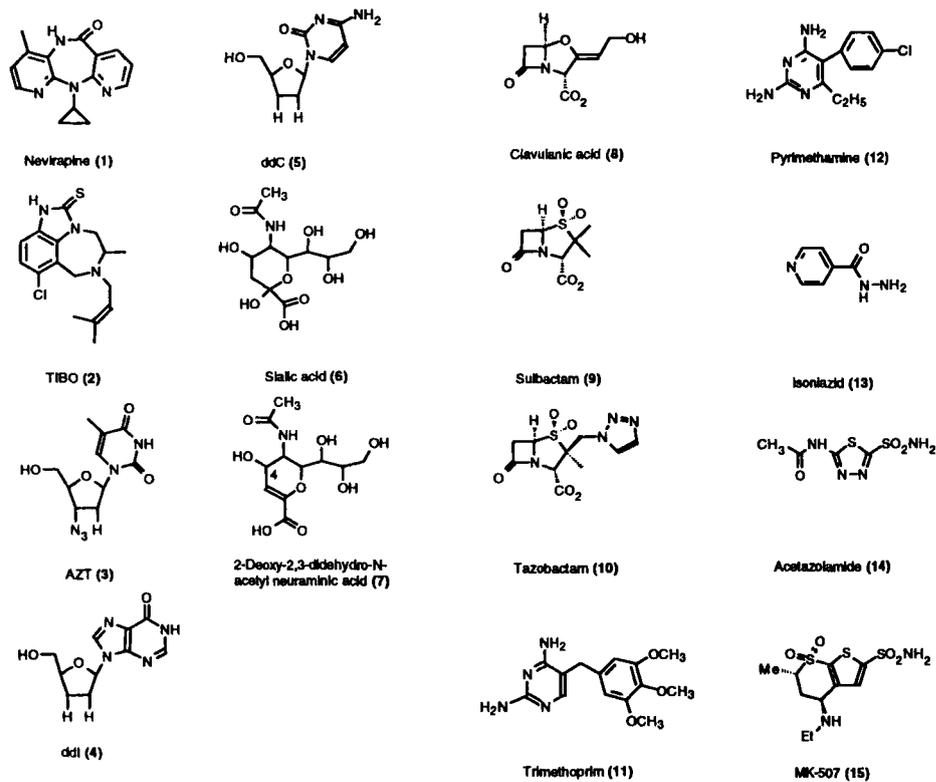
ddI (4)

Trimethoprim (11)

MK-507 (15)

Figure 3.   Drug compounds I. (1)-(7) are antiviral drugs, (8)-(13) are antibacterial compounds, and (14) and (15) are glaucoma drugs.

by reverse transcription are each shortened by two nucleotides. These are then joined to the 5′ ends of the host DNA. In cases where the DNA substrate mimics the strand transfer product, integrase can undo the joining of the viral and host DNA. The disintegration reaction involving polynucleotidyl transfer requires only the core domain, while the other two reactions require the intact 288-residue integrase molecule.

Though their sequences are not similar, the retroviral integrase core domain and the transcriptase ribonuclease H domain share common active site features. Two of three highly conserved active site residues involved in all three integrase-catalyzed reactions are also found in the ribonuclease H domain. In both cases, a glutamate and two aspartate residues in similar positions are found to be essential for catalysis. Another similarity between the two enzymes is a cofactor requirement for divalent cations such as Mg^{2+} or Mn^{2+}, which are observed near the active site in the ribonuclease H domain.

B. Influenza Virus

Influenza (flu) is caused not by a retrovirus but a myxovirus. These RNA viruses are about 1000Å in diameter, and consist of strands of RNA surrounded by phospholipid membrane. Embedded in the membrane are soluble nucleoprotein antigens from which it derives a classification of A, B, or C. Influenza type A, which includes several distinct serotypes, is responsible for most clinical cases of flu. Types A and B can be associated with epidemics and occasionally pandemics with considerable mortality. Those at greatest risk are the very young, the elderly, or patients with compromised immune, cardiovascular, or respiratory systems or other weakened conditions.

The influenza virus coat glycoproteins, hemagglutinin and sialidase (or neuraminidase), are strain-specific and mutate frequently. Cell surfaces contain glycoprotein receptors for hemagglutinin and sialidase, and myxoviruses attach to these. This recognition process leads to fusion between cell and virion, followed by introduction of viral genetic material into the cell. Intracellular replication, which is achieved differently than in the retroviruses, is followed by viral assembly at the cell surface and budding of new virions.

Pharmacological approaches to influenza include both prevention and treatment. Prevalent strains of influenza, which are monitored epidemiologically, are included in mixed flu vaccines that protect against infection by these strains for 1-2 years. Inactivated virus particles or components or even semi-purified viral hemagglutinins elicit reasonably good immunity. Amantidine, or its analogue rimantadine, can be used in the prevention or treatment of infection by influenza A. Both inhibit the ion channel function of a viral protein that is present in the type A, but lacking in the type B, influenza virus. Side effects and development of drug-resistant viral strains have been problems associated with these agents.

Sialidase (Neuraminidase) and Hemagglutinin

Sialidase is a glycohydrolase found in animals and in many pathogenic micro-organisms. The enzyme cleaves sialic acids (6) (Figure 3) with terminal α-ketoside linkages and acts upon many glycoproteins, oligosaccharides, and glycolipids. Its pharmacologically important target is the cellular receptor for the influenza virus. During infection, influenza virus sialidase may assist the release of new virus particles from infected cells and help the virus move through the respiratory tract.

Crystal structures of sialidase were solved by Colman and co-workers (Colman et al., 1983; Varghese et al., 1983; Varghese and Colman, 1991). The enzyme is a homotetramer, and its tertiary structure has been termed a β-propeller, i.e., six groups of four-stranded β-sheet spread out radially like propeller blades. The active site is located in a deep cavity on the protein surface and is lined with invariant residues. Some key interactions with the substrate molecule have been worked out, and these are three arginine residues that form hydrogen bonds to a carboxyl of the substrate and a tyrosine, a glutamate, and an aspartate that form hydrogen bonds to hydroxyls (Varghese et al., 1992). All these interactions are thought to force strain in the six-membered ring of the substrate, and enable the formation of an oxocarbonium ion that drives the forward reaction (Janakiraman et al., 1994). In addition, several studies have been performed on escape mutants of sialidase. Escape mutants are those viruses, grown in the presence of monoclonal antibodies that inhibit wild-type sialidase, that are infectious and resistant to the antibody. These studies have helped to determine the structural basis for protein-antibody interactions (Tulip et al., 1991; Colman, 1992; Colman, 1994).

Inhibitors aimed at conserved active site features should be effective regardless of influenza strain (Varghese et al., 1992). Computer-assisted analysis and modeling of the active site bound with known substrate analogues led to the prediction that replacement of a four-hydroxyl group in a known inhibitor (7) (Figures 3) with an amino or guanidinium group would create a new salt bridge and thus lead to tighter binding (von Itzstein et al., 1993). Although these compounds showed markedly enhanced inhibition of sialidase *in vitro*, the drugs were rapidly expelled *in vivo*, and thus not promising as curative agents. In addition, the derivatives were difficult to synthesize, and thus further studies with this class of compounds would be difficult. Recently, some aromatic derivatives of sialic acid were studied (Jedrzejas et al., 1995). These compounds were found to bind in the same site as sialic acid, but were not very good inhibitors. However, the aromatic moiety lends itself well to synthesis, so that further studies with this class may prove fruitful.

The crystal structure of hemagglutinin (for review, see Wiley and Skehel, 1987), which also binds sialic acid on the host cell surface, shows that the molecule is a homotrimer. Each monomer is anchored to the membrane by a hydrophobic piece that was cleaved off before crystallization. The remaining soluble portion includes a partly helical stalk that holds the monomers together and a globular head region. The head, a β-barrel structure, contains the recognition apparatus, i.e., surface
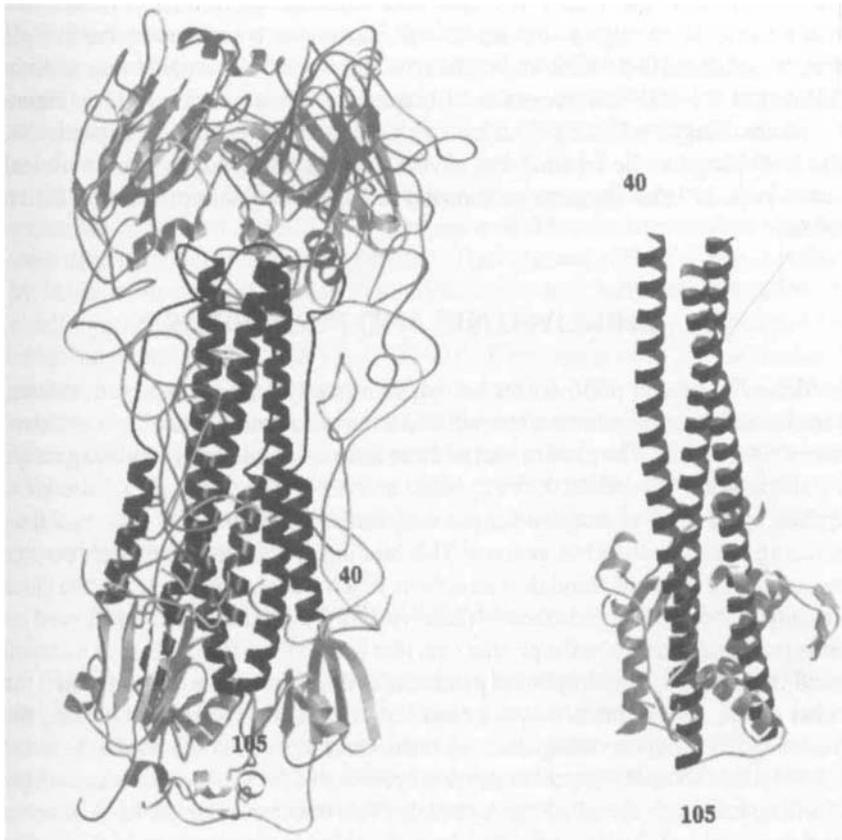
Figure 4. Hemagglutinin in high and low pH. On the left is the structure of hemagglutinin in its normal membrane bound state. Sialic acid has been removed from the strucuture for clarity (PDB # 1HGF from Sauter et al., 1992). Upon lowering of the pH, the conformation shifts to the structure on the right (PDB # 1HTM from Bullough et al., 1994). In both structures, residues 40-105 of the HA_2 chains are highlighted to show the important changes that occur upon lowering of the pH, most notably formation of a coiled coil and the motion of the "fusion peptide" (located at the N-terminus at the top of the coiled coil) from the bottom of the high pH structure to the top of the low pH structure.

antigenic loops and the sialic acid binding site. Development of drugs to block sialic acid binding would prevent infection by interfering with the virus-receptor inter-actions. But hemagglutinin actually plays a double role in virus replication, and it leads to a second potential approach to therapeutic intervention. After internaliza-

tion of the virus-receptor complex, hemagglutinin activates fusion of the viral and endosomal membranes so that the vesicle contents, including viral enzymes and genetic materials, can empty into the host cell. Hemagglutinin activates this membrane fusion through a conformational change that is induced by the low pH environment found in the endosome. The crystal structure of low pH-hemagglutinin (Bullough et al., 1994) differs considerably from that of the neutral-pH form (Figure 4), with shuffling or refolding of major regions of secondary structure. In particular, there is a change in the location and environment of the hydrophobic N-terminal "fusion peptide" that suggests an induced readiness to participate in the fusion process.

IV. RHINOVIRUSES AND POLIOVIRUSES

The rhinoviruses and polioviruses belong to a family called the picornoviruses, characterized by a protein coat shaped like an icosahedron containing a positive-sense RNA genome. The protein coat of these viruses is built up by repeating arrays of four proteins, termed VP1, VP2, VP3, and VP4. The life-cycle of the virus requires binding to a receptor on the cell surface, entering the cell, and then uncoating to expose the RNA genome. This binding process to the cellular receptor was suggested to occur through interactions in a hydrophobic pocket in the floor of a canyon located in the surface of these viruses (Rossmann, 1989), followed by conformational changes in the protein coat that lead to uncoating. There is a natural ligand that binds to a hydrophobic patch under the floor of the canyon called the pocket factor, whose function is not clear but may be important in stabilizing the virus particle so that uncoating does not occur until the receptor is bound (Grant et al., 1994). Once bound, the pocket factor may leave and the virus undergoes changes to form an intermediate called the A particle. This then can insert its RNA genome into a new host cell. There are several drugs that bind in the same site as the pocket factor and, for the most part, these canyon binders inhibit uncoating because they have higher affinity for the canyon binding site than the natural ligand. Most of the work to date has been targeted towards a better understanding of this uncoating process.

A. Rhinoviruses

Rhinoviruses are the cause of about 90% of all common cold strains. There are over 100 different strains of rhinovirus identified so far, and a drug to cure the common cold should therefore have cross-reactivity with a variety of viral strains. There are drugs that inhibit the uncoating process, and the design of these has been reviewed recently (Whittle and Blundell, 1994). Briefly, these drugs were found to bind in the hydrophobic surface canyon. By using the crystal structure and some analogues, the properties of the drugs that limit binding were determined (Diana et

al., 1992). For instance, there are portions of the drug molecules that are sterically sensitive and hydrophobic interactions dominate the binding. Using this information, better rhinovirus inhibitors can be designed.

Subsequently, the interaction of the human rhinovirus-16 (HRV16) with its cellular receptor, intercellular adhesion molecule-1 (ICAM-1), was determined to a molecular level using a combination of X-ray crystallography, cryoelectron microscopy, and homology mapping (Olson et al., 1993). Cryoelectron microscopy was first performed on a complex of HRV16 and ICAM-1, determining the relative binding position of ICAM-1 on HRV to a resolution of about 28 Å. Since the structures of HRV16 and ICAM-1 were not available, two homologous structures were used to model in atomic interactions. The structure of HRV16 was substituted by human rhinovirus-14 (HRV14) and ICAM-1 was substituted by CD4. The resulting structural data agreed well with the electron microscopy, and showed some of the interactions of ICAM-1 and HRV16. There are several interactions in the surface canyon, mostly between ICAM-1 and the VP1 protein. The ICAM-1 overlaps the same site in the canyon as residues affected by anti-viral drug binding (Smith et al., 1986) and residues that were known to be involved in receptor binding (Colonno et al., 1988). As we grow to further understand the life-cycle of this virus and the interactions that drive infection, we may be able to develop better strategies for curing the common cold.

B. Poliovirus

The virus that causes the disease polio is a picornavirus with some similarity to the rhinoviruses. The structure of poliovirus is known (Hogle et al., 1985). Interest in drugs that bind to poliovirus is twofold. First, these drugs stabilize the virus particles, so that vaccines treated with the drugs can be stored without refrigeration. Second, its similarity to rhinovirus means that knowledge learned from poliovirus may help cure the common cold. The receptor for poliovirus is known, but its function is unclear. The receptor is a member of the immunoglobulin family, and some work has been done to determine the amino acids involved in binding (Bernhardt et al., 1994), but no X-ray structure is available. Much has been done on the uncoating process, however.

A mechanism for drug inhibition of uncoating was recently proposed from work done with several virus-drug structures (Grant et al., 1994). The drug binding site was well characterized, and few structural changes occur upon binding beyond the altered conformation of Met 132 of the VP1 protein. The drug is proposed to act through inhibition of conformational changes occurring later in the viral life cycle, rather than by receptor antagonism or by altering the receptor binding site. Models have also been proposed from this mechanism for drug entry and exit into the binding site and the switch from an intact virus to the A particle. In addition, poliovirus strains that are drug resistant can provide some clue as to the amino acids on the virus that are involved in this change (Mosser et al., 1994). There are three

main patches that give rise to infective strains: one is the site of drug binding: one is near the fivefold axis in an area thought to be important for receptor binding; and the third is along the inner surface of the capsid, forming a path for the RNA that is to be extruded. As understanding of the structural basis for virus binding and the mechanism of transition from virus to A particle to newly infected cell is increased, new drugs that are better inhibitors should result.

V. ANTIBACTERIAL THERAPEUTICS

Infection by pathogenic bacteria is a very common problem in medicine. Because there is an increasing resistance of bacteria to established antibiotics, many groups are working on the design of new antibiotics. There are multiple modes of bacterial resistance, modifying enzymes that inactivate the antibiotic, "pumping" mechanisms that remove the antibiotic from the cell, and modification of the target enzyme so that the antibiotic no longer binds (Spratt, 1994; Nikaido, 1994). In the search for new antibiotics, the focus has been on discovery of new natural products or modification of existing natural products. However, since natural products come from nature, nature probably has a way to fight the antibiotic, merely by virtue of the fact that the cell that produces the antibiotic needs a defense against it (Spratt, 1994). Therefore, some have suggested that the best course of action is structure-based drug design (Spratt, 1994). Here, advances will be outlined in the targeting of β-lactamases, InhA, and dihydrofolate reductase as antibacterials.

A. β-Lactamases

The most commonly used antibiotics are the β-lactams, such as penicillin. These act by inhibiting enzymes in the cell that are responsible for cell wall synthesis. The mechanism of this inhibition by the β-lactams has been elucidated (Ghuysen, 1991). Cell walls consist of disaccharides linked together through a pentapeptide. The peptides are cross-linked together through a peptide bond between the c-terminus of one pentapeptide and a side-chain amino group of a second pentapeptide. To create this cross-link, the cell utilizes transpeptidase enzymes that catalyze a reaction analogous to a serine protease reaction. The first step is the formation of an acyl-enzyme intermediate through an active site serine residue. This intermediate is then displaced from the enzyme by the attack on the carbonyl of the amino group of the second pentapeptide. This reaction is inhibited by β-lactams. The β-lactams can bind to the transpeptidase and form the acyl-enzyme intermediate, but this intermediate is stable to the next step because the site of attack is hindered. In addition, hydrolysis does not occur at a fast rate in the active site of the enzyme, so the lactam cannot be released as the carboxylic acid derivative. Therefore, the enzyme is stuck in a dead-end complex, cell wall synthesis is halted, and eventually the cell dies. However, some bacteria have evolved a mechanism of defense against the action of β-lactams.

The β-lactamases hydrolyze the lactam ring of the β-lactams, thereby destroying their ability to inhibit cell wall synthesis, and are the most common defense mechanism bacteria use against antibiotics. A possible way to avoid the deleterious effects β-lactamases have on antibiotics is to use an inhibitor of lactamases in combination with a normal antibiotic, allowing normal antibiotic function. Design of these inhibitors can be substantially helped by X-ray crystallography. There are several X-ray structures for α-lactamases published for class A lactamases (penicillinases) (Dideberg et al., 1987; Herzberg and Moult, 1987; Moews et al., 1990; Strynadka et al., 1992; Jelsh et al., 1993) and class C lactamases (cephalosporinases) (Oefner et al., 1990; Lobovsky et al., 1993).

Mechanistically, β-lactamases react in much the same way as a serine protease does. However, the lactamases are distinct from the proteases in that the histidine residue acting as a general base is replaced by lysine or tyrosine, there is formation of only one molecule as a product and not two, and the hydrolysis step is different in that the angle of attack of water in the protease mechanism is sterically hindered in lactamase (Chen et al., 1993). There are well known sequence motifs, Ser-Xxx-Xxx-Lys, Ser or Tyr-Xxx-Asn, and Lys or His-Thr or Ser-Gly (for a review, see Ghuysen, 1991), where Xxx is any amino acid, that form a common active site. The active site serine is located in the Ser-Xxx-Xxx-Lys sequence. There are differences in distances in the active site and in the surrounding loops that could conceivably account for the change in specificity and class (Lobovsky et al., 1993). Details of the interactions of the acyl-enzyme complex (Oefner et al., 1990; Strynadka et al., 1992) and the transition state (Chen et al., 1993; Lobovsky et al., 1994) have been determined by co-crystal structures. There is, therefore, a wealth of structural information available about the β-lactamases and their mechanism of action and the idea of performing rational drug design to improve inhibitors is feasible.

Three examples of such lactamase inhibitors are clavulanic acid (8), sulbactam (9) and tazobactam (10) (Figure 3). These compounds resemble β-lactams except for substitution of a sulfur atom or a sulfonate in a five-membered ring for oxygen and different substituents attached elsewhere on the five-membered ring. The mechanism of inhibition of these appears to be the formation of a long-lived acyl-enzyme intermediate, which can then have several chemical fates other than hydrolysis (see Chen and Herzberg, 1992 for a summary). A structure has been solved for a complex of clavulanate and the β-lactamase from *Staphylococcus aureus* (Chen and Herzberg, 1992) and shows the reacted clavulanate in two forms, *cis* and *trans* enamines (Figure 5). Molecular modeling has been done on both clavulanate (Imtiaz et al., 1993) and sulbactam (Imtiaz et al., 1994) to determine the possible reactions each undergoes, and the results are that sulbactam may undergo different rearrangements in the active site from the enamine form to an imino form that can then form a second covalent bond to the enzyme. This is a model, however, and there is no structural information available. Similar mechanisms were proposed for clavulanate, but the X-ray studies show only enamine forms in the active site.
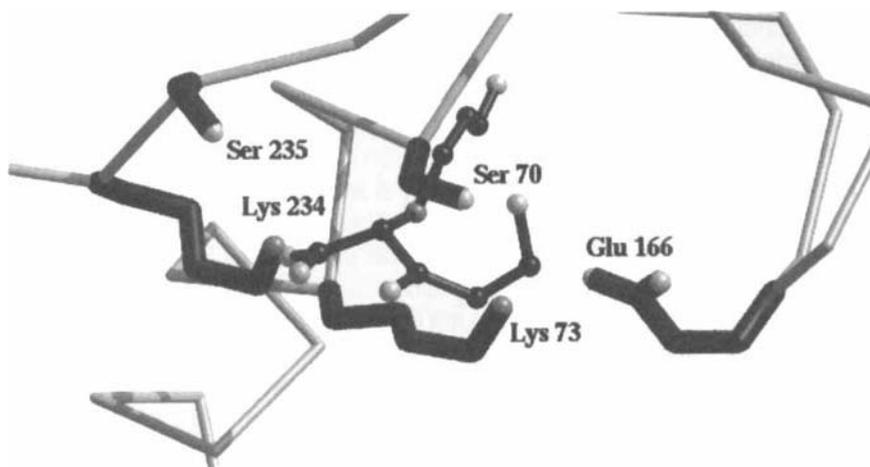
Figure 5. The active site of β-lactamase from *S. aureus*. The active site is shown complexed with the *cis* form of clavulanic acid (PDB # 1BLC from Chen and Herzberg, 1992). The *trans* form of clavulanate has been removed for clarity. The active site residues are highlighted. Ser 70 is the site of acylation; Lys 234, Ser 235, Lys 73, and Glu 166 all form consensus sequences with other β-lactamases.

A second avenue of study of inhibition of β-lactamases has appeared recently, but is only a blip on the horizon. The recent discovery and characterization of the β-lactamase inhibitor protein (BLIP) from *Streptomyces clavuligerus* (Strynadka et al., 1994) provides an interesting alternative to small molecule inhibitors. BLIP was shown to be a potent lactamase inhibitor in the subnanomolar range of K_i. The structure of BLIP shows a helix-loop-helix motif packed against a four-strand anti-parallel β-sheet. How BLIP interacts with lactamases is not yet known. As we grow to understand the differences in the various lactamases and the intricacies of their reactions, we may be able to design better inhibitors to help in combination therapies and fighting infections.

B.   Dihydrofolate Reductase

The enzyme dihydrofolate reductase (DHFR) catalyzes the conversion of 7,8-dihydrofolate to 5,6,7,8-tetrahydofolate using NADPH as a cofactor. The role of folate in the cell is a supplier of one-carbon groups important in the biosynthesis of thymidine and certain amino acids. Thus, inhibition of DHFR leads to cell death. This fact has prompted interest in DHFR inhibitors as treatments for bacterial

infection and cancer. As far as use of crystallography in drug design is concerned, DHFR is one of the oldest and most widely studied systems. There is a wealth of information about the enzymatic mechanism, binding of inhibitors, and design of drugs (Roth, 1986; Kraut and Matthews, 1987; Kuyper, 1989).

A problem with targeting an enzyme that is utilized by both the host and the target is crossreactivity. X-ray crystallography can help find the subtle differences that allow specific drugs to be designed. The structural differences between the human and the *Escherichia coli* DHFRs have been worked out by the crystallization of the human (Davies et al., 1990) and *E. coli* (Reyes et al., 1995) DHFRs complexed with folate and other folate analogues. The structures are basically the same, where the folate binding site is a hydrophobic pocket formed by two parallel β-sheets flanked by two α-helices. The pteridine ring in both cases forms hydrogen bonds to a carboxyl side chain and a water molecule on a helix that forms one side of the cleft. The NADPH cofactor was found to be in an optimal position to donate a hydride to the C6 of the substrate in the reduction, and causes some shifts in the binding site for folate when seen in a ternary complex. However, the N8 position on the pteridine ring is in a different environment (Reyes et al., 1995), and may indicate a different protonation state.

Trimethoprim (11) and pyrimethamine (12) (Figure 3) are well-established antibacterial agents whose target molecule is DHFR. The drug binds well to *E. coli* DHFR, but shows significantly lower binding to human DHFR (Burchall and Hitchings, 1965). The structural reason for this selectivity has been worked out through crystal structures of *E. coli* DHFR and DHFR from mouse (Stammers et al., 1987) and chicken (Matthews et al., 1985) complexed with trimethoprim. The trimethoprim binds in the same site as the folate cofactor, except that the carboxyl group now binds to the pterine ring at the 4-amino position, essentially rotating the pterine ring 180°. The differential binding appears to arise from a small shift in one portion of the binding site. In the vertebrate DHFR, the active site cleft is about 1.5 Å wider. This has the consequence of shifting the position of the inhibitor in the binding cleft and precluding formation of a hydrogen bond to the backbone carbonyl of Val-115. Thus, the complex of trimethoprim and vertebrate DHFR is missing one hydrogen bond at the 4-amino position, and this is thought to provide the specificity for the bacterial enzyme. Thus, subtle differences in structure can be exploited to provide specificity, and the possibility exists that one could improve DHFR inhibitors and create new, specific antibacterial agents using structure-based design.

C. InhA

A drug called isoniazid (isonicotinic acid hydrazide) (13) (Figure 3) has been known to be effective against tuberculosis since 1952 (Middlebrook, 1952). Isoniazid is believed to function by inhibition of the bacterial biosynthesis of β-hydroxy fatty acids that are key components of the cell wall of *Mycobacterium*

tuberculosis (Quemard et al., 1991). Thus, isoniazid acts in a manner similar to β-lactams, but with a different target, and is specific for mycobacteria. The protein target of isoniazid has been identified as InhA, an enzyme that catalyzes the reduction of 2-*trans*-enoyl-acyl carrier protein using NADH as a cofactor (Banerjee et al., 1994; Dessen et al., 1995).

Bacterial strains have arisen that are immune to the effect of isoniazid, mostly in AIDS patients who are afflicted with tuberculosis (Snider and Roper, 1992). Some of the mutations that confer this resistance have been mapped to genes encoding catalases, and the mechanism of resistance is thought to be the inability of the cell to process the drug into an active form (Zhang et al., 1992). However, there are other mutations that have recently been mapped to the InhA gene, specifically substitution of a serine residue with an alanine (Banerjee et al., 1994).

An important step has been taken in the rational design of tuberculosis antibacterials when the structure of the InhA protein was solved (Dessen et al., 1995). The overall topology of the enzyme is that of a chair. The location of the cofactor binding site was determined by cocrystallization, and the binding site of the lipid was deduced to be in a hydrophobic patch formed from three β-sheets and three α-helices. The mutated serine residue turns out to be in the cofactor binding site, coordinating to an important water molecule that is bound to the P_N phosphate of NAD. In the mutant structure, the alanine can no longer coordinate the water molecule, and this results in a shift of a loop away from the NAD and a lower affinity for the cofactor. A second mutant, an isoleucine to threonine, was also studied. The isoleucine is located near the ribose binding site of NAD. The authors theorize that the mechanism of resistance has to do with a decreased affinity for NAD, slowing the enzyme reaction, but drastically affecting the isoniazid binding. With the structure, the door is now open for rational design of better drugs to combat tuberculosis.

VI.   BLOOD COAGULATION

Blood platelets normally circulate in a nonadherent or "nonactivated" form. When blood vessel injury exposes the endothelial lining, platelets adhere, become activated, and begin to aggregate. Blood coagulation helps to seal off bleeding by forming clotting material, fibrin, around a hemostatic plug. Two or more biochemical pathways converge to form a macromolecular complex on the surface of endothelial cells or activated platelets. This complex activates prothrombin, the otherwise inactive precursor of the serine protease thrombin. Thrombin, in turn, acts upon the precursor fibrinogen and thus initiates events that lead to a cross-linked fibrin clot. Thrombin also targets other protein components of the blood coagulation pathway (for review, see Furie and Furie, 1992).

Normally, hemostatic regulatory mechanisms exist to keep coagulation processes under control and avoid either excessive clot formation (thrombosis) or

bleeding (hemorrhage). In pathological situations, the clot (thrombus) blocks blood flow through large blood vessels and thus leads to oxygen starvation of tissues. Venous thrombosis can lead to pulmonary embolism, a potential medical emergency. When a clot forms in a vein, it may be carried through the venous circulation through the right heart and into the pulmonary artery, where it becomes lodged. Similarly, thrombotic blockage of a coronary or cerebral artery can lead to myocardial infarction (heart attack) or stroke. Whether venous or arterial, a sufficiently large blockage has high risk for fatality. One approach to treating thrombosis is the introduction of tPA (tissue plasminogen activator), urokinase, or streptokinase, which enhances fibrolysis, the breakdown of fibrin clots. tPA and urokinase are human physiologic activators, while streptokinase is a bacterial product. Prevention of thrombosis will also prevent pulmonary embolism. However, new avenues of treatments, or improvements in currently used therapies, are being sought because of the prevalence and seriousness of cardiac and circulatory diseases.

A.   Thrombin Inhibitors as Anticoagulants

Several approaches to preventing or limiting coagulation target thrombin (Fenton et al., 1991). Administration of heparin, a highly sulfated polysaccharide, prevents prothrombin conversion to thrombin by activating antithrombin III, a plasma protease inhibitor of coagulation enzymes. This inhibitor neutralizes blood-clotting proteases that are not associated with the developing clot, thereby serving as a regulatory protein. However, once thrombin has been converted to its proteolytically active form and the blood coagulation cascade has been initiated, heparin is no longer an effective anticoagulant because the heparin-antithrombin III complex cannot access its target proteases that remain associated with the clot. It has been shown that individuals with genetic defects involving antithrombin III or other blood coagulation regulatory factors have increased risk for thrombosis. Antithrombin III, one of the so-called serpins (for serine protease inhibitors), has been studied extensively with X-ray crystallography (for review, see Goldsmith and Mottonen, 1994). A more potent protein thrombin inhibitor ($K_i=10^{-14}$; Stone and Hofsteenge, 1986) is hirudin, derived from leeches. Hirudin prevents coagulation of the host's blood, which the leech feeds on. Crystal structures of the thrombin-hirudin complex (Grütter et al., 1990; Rydel et al., 1991) have provided many molecular details pertaining to the high specificity of this interaction. Despite the availability of these macromolecular inhibitors, lower molecular weight thrombin inhibitors are preferable from a therapeutic point of view. Therefore attention has focused on development of smaller, synthetic compounds such as active-site directed inhibitors or mimetics of hirudin.

α-Thrombin is a disulfide-linked heterodimer, with a 36 (human) or 49 (bovine) A-chain and a 259 residue B-chain. Thrombin exhibits a trypsinlike specificity, though it prefers arginine over lysine. Unlike trypsin, thrombin is highly selective for fibrinogen and its other macromolecular targets. Several topological features

contribute to this unusual specificity, including: (1) a protruding 60-residue loop that partially blocks the active-site cleft, (2) three specific binding pockets in the active site region, and (3) a separate substrate recognition site distant from the active site. These constraints limit the number of bonds that are cleaved: only four out of almost 400 possible cleavage sites in fibrinogen. The active site with its Ser-His-Asp catalytic triad is a deep canyonlike cleft into which fibrinogen fits.

At the bottom of the first complementary subsite on thrombin (the S1 pocket, into which protein substrate fits) is an aspartate residue that forms a salt bridge with the fibrinogen arginine residue (Figure 6A). This feature is similar to other trypsinlike proteases. An apolar binding site near the S1 pocket accommodates hydrophobic residues from the substrate. Aryl-aryl interactions particularly stabi-
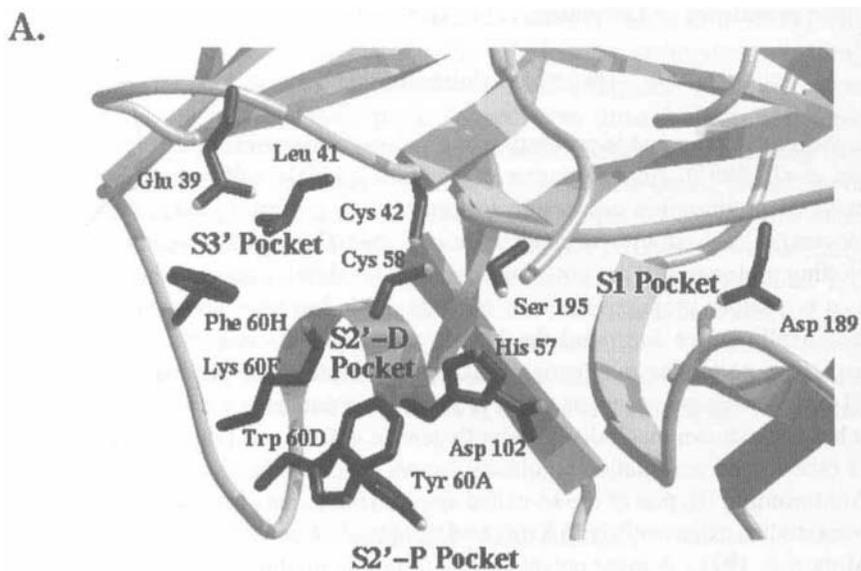

Figure 6. Bovine thrombin. **A.** The active site. The catalytic triad is formed from residues Ser 195, His 57, and Asp 102. To the right is the S1 pocket; at the bottom of the pocket is residue Asp 189, which is important for selectivity. For drug design, the S2' and S3' pockets are potential targets. The S2' pocket is divided into two portions: the distal or D pocket (Cys 42, Cys 58, Lys 60F, and His 57), and the proximal or P pocket (Tyr 60A, Lys 60F, Trp 60D, and His 57). The residues in the S2'-P pocket are on a loop that can move to accommodate certain substrates or inhibitors. The S3' pocket consists of residues Glu39, Leu 41, and Phe 60H. **B.** Thrombin complexed with hirudin. Alpha thrombin is shown in a ribbon diagram. Hirudin is shown in black as an α-carbon trace. The hirudun binding can be divided into three parts: the active site blocker (residues 1-48), a linker (residues 49-54), and exosite binding (residues 55-65) (PDB # 1HRT from Vitali et al., 1992).

lize aromatic side-chains in a binding site proximal to the apolar binding site. Banner and Hadváry (1991) use the terms P and D pockets, referring to proximal or distal from the active site, for the apolar (S2) pocket and the aryl-binding site (Bode et al., 1990), respectively. A tyr-pro-pro-trp loop that lines the hydrophobic double pocket moves if spatial accommodation is needed for substrate or inhibitor residues (Banner and Hadváry, 1991). Further recognition of fibrinogen is provided by a post-cleavage "exosite" (i.e., distinct from the active site) that contains both ionic and hydrophobic features. This region has a basic electrostatic field that contributes to anionic substrate binding.

Extensive crystallographic studies of peptidic and non-peptidic thrombin inhibitors that are active site-directed utilize the specificity and hydrophobic binding pockets found in this serine protease. Non-hydrolyzable peptide derivatives such as the chloromethyl ketone D-Phe-Pro-Arg-CH$_2$Cl (PPACK), and modified fibrinopeptides (e.g., Stubbs et al., 1992) tend to resemble substrate more closely. PPACK, which also inactivates trypsin, binds covalently to the active site serine and histidine residues of the catalytic triad in thrombin (Bode et al., 1989; Banner and Hadváry, 1991). The L-arginine residue occupies the S1 pocket, the normal substrate recognition pocket found in both enzymes. The L-Pro of PPACK is
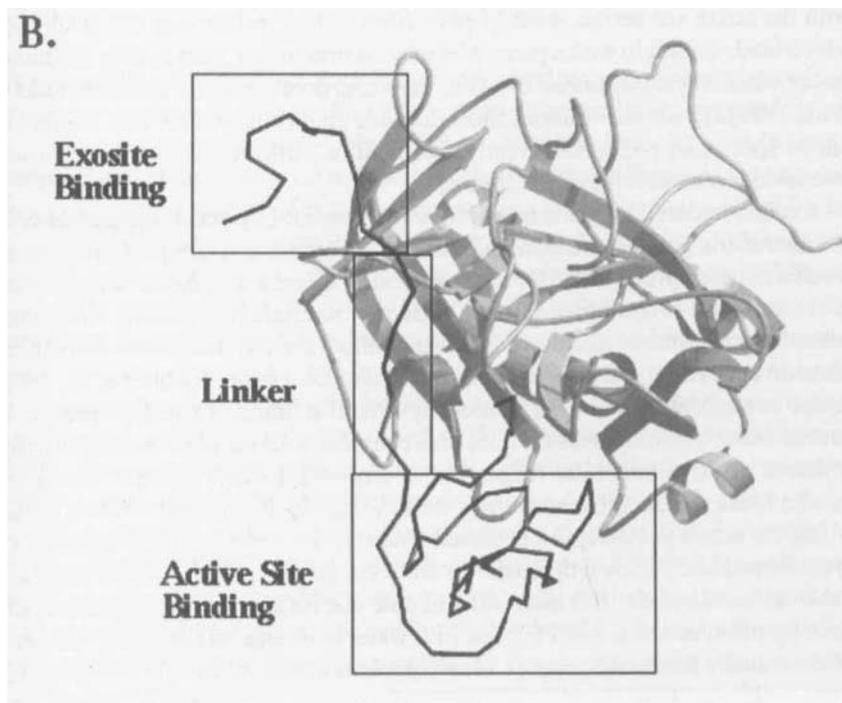

Figure 6. Continued.

comfortably accommodated by the hydrophobic P-pocket of thrombin, and the D-Phe side- chain fits the D-pocket (with even greater complementarity than natural L-Phe). Benzamidine binds to the S1 pocket of trypsin or thrombin (Banner and Hadváry, 1991; Brandstetter et al., 1992), but its affinity for thrombin is enhanced when moieties are added that will bind in the thrombin specificity pockets. Effective non-peptidic thrombin inhibitors have been developed that are based on benzamide or arginine, including NAPAP (Nα-(2-naphthyl-sulfonyl-glycyl)-D-*p*-amidino-phenylalanine-piperidine) and MQPA ((2R,4R)-4-methyl-1-[Nα-((RS)-3-methyl-1,2,3,4-tetrahydro-8-quinolenesulfonyl)-L-arginyl]-2-piperidine carboxylic acid (Banner and Hadváry, 1991; Brandstetter et al., 1992). In contrast to the peptidic inhibitors, which tend to bind in an extended conformation to the thrombin active site cleft, these non-peptidic inhibitors bind in a compact, U-shaped conformation. However, these still take advantage of complementarity provided by thrombin recognition pockets. NAPAP and MQPA both bind with their piperidine moiety positioned in the thrombin P- pocket and packed firmly against the naphthyl or quinoline group in the adjacent D-pocket. The p-amidino-phenylalanine or L-arginine moiety in each inhibitor binds at the S1 recognition site, albeit in different conformations. MQPA, also known as MD-805 or Argatroban, is already in clinical use as an anticoagulant. Boronic acid-containing peptide analogues have been shown to mimic the serine protease transition state by forming a tetrahedral adduct with the active site serine. Analogues of Ac-(D)Phe-Pro-boroArg-OH (DuP714), which binds thrombin with a picomolar inhibition constant, vary in their inhibitory power when Arg is substituted by lysine, amidine, homolysine, or ornithine (Weber et al., 1995). These substitutions show that side-chain interactions with the thrombin S1 specificity pocket can greatly alter inhibitor affinity, even while active site interactions remain relatively constant.

Another important approach to designing thrombin-targeted drugs derives from the interaction between hirudin and thrombin. Hirudin is a compact, 65-residue, disulfide-linked protein that forms an extremely tight, 1:1 complex with fibrinogenic thrombin (Figure 6B). Hirudin differs in structure and binding from other natural serine protease inhibitors. Binding of hirudin to thrombin involves both the protease catalytic site and the fibrinogen recognition exosite (Grütter et al., 1990; Rydel et al., 1991). A posttranslationally-modified sulfated tyrosine, present in natural but not recombinant hirudin, also contributes to the high binding affinity, which is in the femtomolar range. The N-terminal domain residues 1-48 from hirudin block the thrombin active site, with the hirudin N-terminal residues buried within the active site itself. An extended C-terminal domain (residues 55-65), rich in acidic residues, binds at the basic "exosite" on thrombin. A six-residue fragment links the two domains. It is interesting to note that hirudin does not penetrate into the S1 pocket, which is filled instead with water molecules in the crystal structure of the hirudin-thrombin complex. The complementarity of the hirudin-thrombin complex appears to be based on extensive, albeit weak, stabilizing interactions, and a conformational change occurs upon formation of the complex. In contrast, the

trypsin-BPTI complexation involves fewer binding interactions and essentially no induced-fit character.

Drug development approaches based on the thrombin-hirudin interaction include different types of analogues. Hirutonins are also bifunctional hirudin-derived inhibitors with nonhydrolyzable moieties that mimic the scissile bond (Zdanov et al., 1993). Hirugen, a C-terminal peptide analogue, is N-acetylhirudin 53-64 and contains the sulfated tyrosine 63. This inhibitor, along with other synthetic hirudin C-terminal analogues, blocks the thrombin exosite but leaves the active site empty. Hirulog-3 (D-Phe-Pro-β-homoArg-(Gly)$_5$-desulfato-Tyr63-hirugen) is a bivalent peptide that interacts with both the thrombin active site and the exosite, with a dissociation constant in the nanomolar range. The β-Arg-Gly is not hydrolyzable, and the five-glycine spacer links the two binding sites. The crystal structure of the hirulog 3-thrombin complex also reveals the nature of the S′ subsites in thrombin. With modeling based on this knowledge, binding could be improved by judicious replacement of the spacer glycines for residues with complementary side-chains (Qiu et al., 1992). Hirulog was in clinical trials until recently, when it was withdrawn by its maker Biogen; though more effective as an antithrombotic, hirulog was not as cost-effective as its competing product, heparin.

VII.   GLAUCOMA

Glaucoma refers to a group of disorders characterized by increased pressure within the eye (ocular hypertension) and/or optic nerve damage leading to impaired vision. It is a leading cause of irreversible blindness, affecting millions of individuals. Several forms of the disease exist, and these tend to afflict different age groups and may arise from a number of causes. Chronic open-angle glaucoma (COAG), for example, is the most common form after age 30; risk factors for COAG include diabetes, extreme near-sightedness (myopia), or hereditary influences. While the intraocular pressure (IOP) rise in COAG may be gradual, acute forms of glaucoma may occur as precipitous attacks, accompanied by severe headaches and, in some cases, nausea and vomiting. While surgery is indicated in some cases, it is not always an option. Frequently, the treatment of choice is either long-term topical (eyedrops) or systemic medication to reduce the pressure within the eye. Unfortunately, many systemic drugs have serious side effects, leading to cardiovascular, hematological, psychiatric, and other symptoms.

IOP is directly related to the balance between production and drainage of aqueous humor within the eye (for review, see Hurvitz et al., 1991). Drugs that act on the nervous system, whether parasympathetic (e.g., pilocarpine) or sympathetic (e.g., epinephrinelike α-adrenergic agents or β-blockers), act through either mechanism, while prostaglandins are being investigated for their usefulness in increasing outflow by mediating inflammation processes. Carbonic anhydrase inhibitors act by decreasing aqueous production. The reaction catalyzed by the enzyme is:

$$H_2O + CO_2 \leftrightarrow H_2CO_3$$
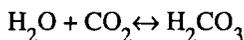
The subsequent acid-base reaction is:

$$H_2CO_3 \leftrightarrow H^+ + HCO_3^-$$
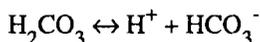
In secretory cells, like those in the eye which produce carbonic anhydrase, the enzyme provides protons or bicarbonate anions as products of the hydration reaction. These in turn may affect the local pH or osmotic pressure and reduce Na^+ and fluid efflux into the aqueous humor. The precise mechanism is not clearly understood. However, as antiglaucoma agents, carbonic inhibitors are well-established, primarily as orally-administered systemic agents. Unfortunately, a substantial proportion of patients are unable to tolerate these drugs due to side effects, since the drugs also affect carbonic anhydrase action elsewhere in the body. Topical application reduces these adverse effects, and much attention has been paid to development of new, effective carbonic anhydrase inhibitors that can be administered locally and in lower concentrations. There are practical constraints specific to topical drugs: they must be hydrophilic yet lipophilic enough to penetrate the cornea, and possess sufficient potency and contact time with the cornea to effectively decrease hypertension.

Human carbonic anhydrase II (HCA II) is a 29.3 kDa zinc metalloenzyme and one of the fastest enzymes known. This high catalytic efficiency means that inhibition must be close to 100% to avoid a significant amount of product generated by the inhibited enzyme. The crystal structure shows the active site to be a deep, conical cleft, at the bottom of which lies the zinc ion (Eriksson et al., 1988). The metal is tetrahedrally coordinated by three histidine residues and a hydroxyl ion from solvent. This ion reacts with CO_2 during catalysis, and one means of inhibiting the enzyme is to replace this water. For example, the aromatic inhibitor 1,2,4-triazole, an analogue of imidazole, binds directly to the zinc ion of HCA II and replaces the water ligand with a nitrogen atom from the inhibitor (Mangani and Liljas, 1993). Crystallographic studies of enzyme-inhibitor complexes have played a key role in drug development. Aromatic sulfonamides act as transition state mimetics and can displace the hydroxide ligand from zinc. The crystal structure of HCA II with bound acetazolamide (14) (Figure 3), a major antiglaucoma drug, shows this type of interaction (Vidgren et al., 1990). Unfortunately, acetazolamide has low lipid solubility and has not been found suitable for topical application (Hurvitz et al., 1991).

A newer class of compounds, based on a thiophenesulfonamide segment instead of the thiadiazolesulfonamide of azetazolamide, is being evaluated for its potential as a topical agent. In one study, the HCA II inhibition properties of a number of derivatives of this newer class were compared. For one of the best inhibitors, a racemic mixture called MK- 927, it was found that the S enantiomer was more a effective inhibitor than the R enantiomer. Crystal structures were obtained of each bound to HCA II. The structural basis for the preference was apparent: the geometry

was less favorable for the R enantiomer, specifically, the relationship between the thiophene ring and the substituent that was varied in the derivative series (Baldwin et al., 1989). The S enantiomer, known as MK-417 or sezolamide hydrochloride, became the basis for another drug, MK-507 (15) (Figure 3) or dorzolamide hydrochloride. All three drugs have demonstrated significant clinical efficacy for topical treatment of glaucoma or ocular hypertension (Wilkerson et al., 1993).

VIII. CANCER

In the search for a cure to cancer, there has been much interest in structure-based design as an avenue of approach. Our knowledge of the mechanism of cancer has increased, and with this knowledge, new compounds can be designed that would never have been considered in the past. Although the exact mechanisms vary, one common feature to all cancers is the rapid, uncontrolled growth of cells. Therapies are designed to kill rapidly growing cells. Three examples of how X-ray crystallography can help in the cure of cancer are the design of new inhibitors of purine nucleoside phosporylase, thymidylate synthase, and dihydrofolate reductase.

A. Purine Nucleoside Phosphorylase

A well-known target for rational drug design is the enzyme purine nucleoside phosphorylase (PNP). PNP functions in the purine salvage pathway by converting purine ribo- or deoxyribo- nucleosides (guanine and inosine) into the purine base and the corresponding 1-phosphoribose or 1-phosphodeoxyribose (Friedkin and Kalckar, 1961; Parks and Agarwal, 1972). The clinical interest in PNP is that PNP deficiency has been shown to affect T cell proliferation, while B cell function remains normal. PNP inhibitors are possible drugs for the treatment of T cell leukemia, T cell lymphoma, T cell-mediated autoimmune diseases such as rheumatoid arthritis and lupus, and T cell mediated tissue rejection from transplantation (St. Georgiev, 1993). Further interest in PNP inhibitors lies in the fact that PNP is somewhat nonspecific and will react with an anti-AIDS drug such as dideoxyinosine (Erion, 1990) or anticancer drugs such as 2'-deoxy-6-thioguanosine (LePage et al., 1964), rendering them useless *in vivo*. Therefore, PNP inhibitors could be of potential use in combination therapies in AIDS and other cancers. There was a problem, however, in that a successful PNP inhibitor should be very potent ($K_i <$ 10 nM) due to high competition with substrate *in vivo* (Ealick et al., 1991). Therefore, there was a need for a rational approach.

The structure of PNP has been solved in the absence of any ligand (Ealick et al. 1990). Crystals of PNP were known to possess enzymatic activity (Krenitsky, 1967) and the crystals contained 76% solvent, so PNP was attractive as a system for rational drug design because the ability to soak in a variety of inhibitors seemed very feasible. The enzyme is a trimer, with the active site located near the interface

between subunits. The active site was characterized by cocrystallization with 8-iodoguanine and Iodoformycin B (Ealick et al., 1990) and consists of polar residues for formation of hydrogen bonds to the purine nucleoside (Glu 201 and Lys 244), a hydrophobic pocket (Phe 200, Tyr 88, and Phe 159 from the neighboring subunit), and a phosphate binding site (Figure 7). An important feature of the enzyme is the movement of a loop (residues 241-260) upon binding of inhibitors (Ealick et al., 1991) which brings the phosphate site closer to the nucleoside binding site and needed to be taken into account in later stages of design. The best inhibitor
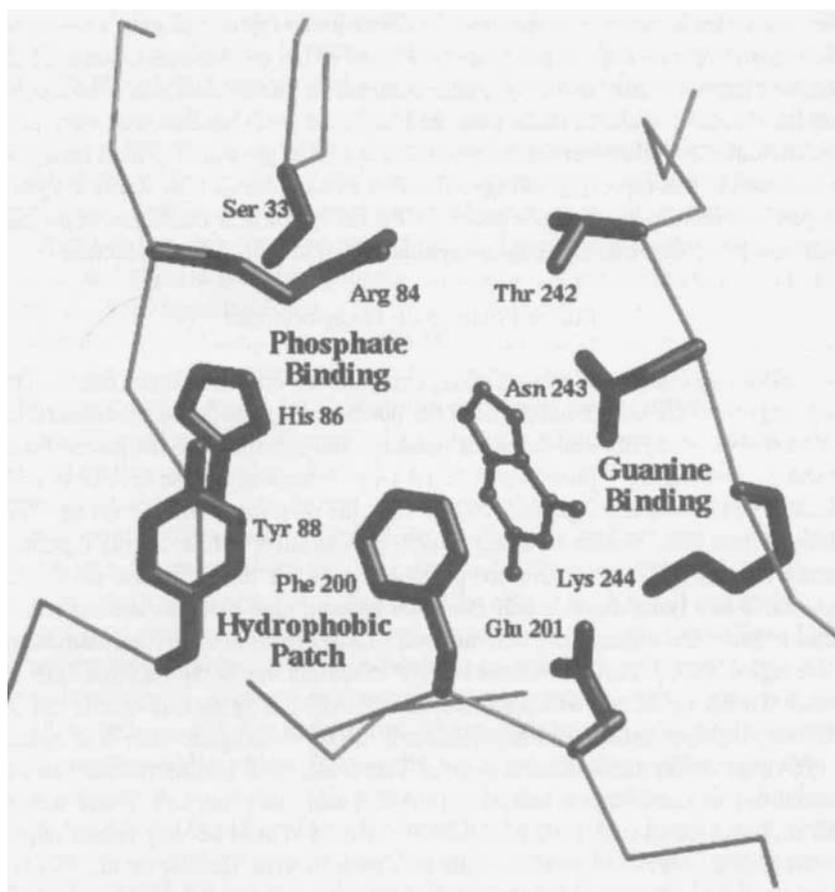

Figure 7. The active site of purine nucleoside phosphorylase. The active site is shown complexed with guanine (PDB # 1ULB from Ealick et al., 1991). There are three sites targeted for design of ligands: a guanine binding site (Glu 201, Thr 242, Asn 243, and Lys 244), a hydrophobic patch (Tyr 88, Phe 200, and Phe 159 from a neighboring subunit which is not shown), and a phosphate binding site (His 86, Arg 84, and Ser 33).

tested initially was acyclovir diphosphate (Ealick et al., 1991), but this was not a viable drug candidate since its availability *in vivo* was questionable (Erion et al., 1993).

The structure-based design was carried out by repeated cycles of modeling, synthesis, crystallography, and inhibition (reviewed in Bugg et al., 1993; Montgomery and Secrist, 1994; Walsh et al., 1994). In the first series of experiments, the difference between guanine and the 8-amino (16) (Figure 8) and 9-deaza (17) derivatives were discovered to be a result of preferential association with Thr 242 (Ealick et al., 1991). The derivatives showed better binding than guanosine as a result of hydrogen bonding to Thr 242, but the 8-amino-9-deaza derivative showed worse binding due to steric strain caused by a shift in the position of Thr242 when bound to a deazaguanine. This result lead to the creation of 9-deaza compounds that had various ring substituents at the 9 position (Montgomery et al., 1993; Secrist et al., 1993) in an attempt to take advantage of the hydrophobic pocket. The best compound was the 3-chlorobenzyl derivative (Montgomery et al., 1993). The aromatic compounds showed better binding than the alicyclic derivatives due to stabilizing interactions with the aromatic rings in the pocket. From this, an attempt was made to see if the phosphate site could be utilized in the binding of derivatives (Erion et al., 1993; Guida et al., 1994) using a series of compounds with substituents attached to the 9-methylene group. A comparison of 9-deazaguanosines and 9-deazahypoxanthines (Niwas et al., 1994) was also performed. The best compounds were the 2-carboxyethyl and the 2-cyanoethyl derivatives, with Ki values of 6 and 11 nM, respectively (Erion et al., 1993), however, due to the difficulty in synthesis, neither of these were continued into clinic. A PNP inhibitor (18) (Figure 8) from this study has recently completed Phase II clinical trials (Peugeot et al., 1995).

B. Thymidylate Synthase

Thymidylate synthase (TS) is an enzyme that catalyzes the conversion of deoxyuridylate (dUMP) to deoxythymidyalte (dTMP) using 5,10-methylenedihydrofolate (CH_2H_2-folate) as a cofactor. TS is responsible for the creation of the thymidylate pool in the cell, which is directly related to the amount of DNA synthesis. Since stopping DNA synthesis is an attractive method of chemotherapy, TS inhibitors have received much attention. Some have even been used in clinical trials as cancer treatments. Examples are CB3717 (reviewed in Clarke et al., 1993), ICI D1694 (reviewed in Clarke et al., 1993), and AG-331 (19) (Figure 8) (O'Connor et al., 1994).

TS has served as a basis for structure-based drug design, since a great deal of structural information is available. Most of the work has been done with the TS from *E. coli* or *Lactobacillus casei*, since the protein is readily available in large quantities. High homology allows extension of the structural studies with the bacterial proteins to the human form. The overall topology of TS is that of a clockwise twisted β-sandwich of two monomers to form the active dimeric unit
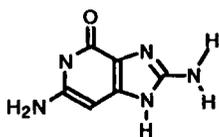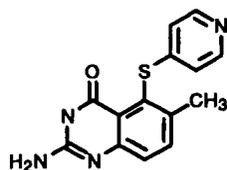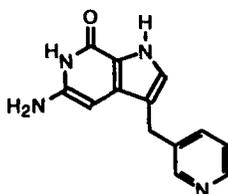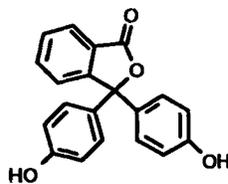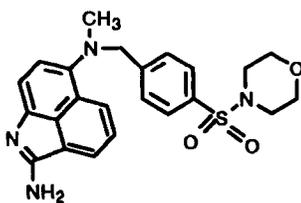
8-Aminoguanine (16)
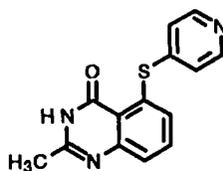
(20)
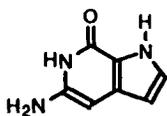
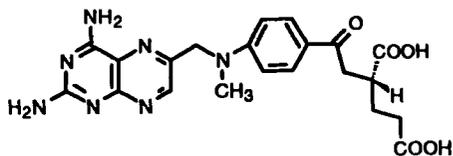9-Deazaguanine (17)

(21)

BCX-34 (18)

Phenolphthalein (22)

AG-331 (19)

Methotrexate (23)

Figure 8.   Drug compounds II. Anticancer drugs.

30

(Matthews et al., 1990a). There are residues from both monomers in each active site, and thus two active sites per molecule.

TS has been crystallized in six important steps along its reaction pathway: the native enzyme (Hardy et al., 1987; Perry et al., 1990); with substrate bound (Finer-Moore et al., 1993); with a cofactor analogue (N^{10}-propargyl-5,8-dideaza-folate or CB3717) that is also an inhibitor bound (Kamb et al., 1992); with CB3717 and a substrate analog (5-fluoro-2'deoxyuridylate or FdUMP) covalently bound to the enzyme showing the reactive complex (Matthews et al., 1990a), a transition state complex showing FdUMP and cofactor (CH_2H_2-folate) covalently bound (Matthews et al., 1990b); and a postreaction complex with a reduced cofactor (H_2-folate) and thymidylate in the active site (Fauman et al., 1994). The substrate binding does not cause many changes from the native structure. Upon binding the cofactor, however, there are many structural changes that occur. Among these changes are a motion of the C-terminus to close down on the cofactor closing off the entrance to the active site and sequestering the reactants from bulk solvent (Kamb et al., 1992). The dUMP-CB3717-TS complex is important for drug design since it shows the interactions between dUMP and TS, cofactor and TS, and cofactor and dUMP (Matthews et al., 1990a) such as: the stacking of the pteridine ring of folate and the pyrimidine of dUMP, the coordination sites of several water mole-cules thought to play key roles in catalysis, a nucleoside binding domain similar to other nucleoside binding proteins, and hydrophobic contacts between TS and the folate cofactor. Thus, TS is a well understood and well characterized system, therefore design programs should be successful using this information.

The first group to perform structure-based design on TS was Appelt et al. (1991), who created AG-331. Starting with the structure of TS and the program GRID, several small molecule sites were found in the active site. The most interesting of these was a napthalene, which was found in a deep hydrophobic pocket near the position where the pterine ring of the cofactor normally binds, interacting with Trp 83. From there, various other substituents were added to form a complete compound. The proposed interactions of a lactam ring linked to the naphthene moeity with a nearby carbonyl of Ala262 did not come out as predicted, rather the two are linked by an intervening water molecule. This observation actually indicates good agreement between prediction and experimental finding. Using this lead, several cycles of design, synthesis, testing, and crystallographic analysis were then carried out (Applet et al., 1991; Varney et al., 1992) to arrive at the final compound, AG-331 (19) (Figure 8). AG-331 is currently in clinical trials (O'Connor et al., 1994).

Webber et al. (1993) used crystallography to help in the design of better TS inhibitors using a combination of modeling and the crystal structure to design a substituted quinazolinone compound. The idea was to create novel lipophilic compounds that resemble the folate cofactor to obtain a drug with fewer side effects than previous folate analogues. The design process started with four quinazolines. Of these, one (20) (Figure 8) showed promise, and the crystal structure of the

inhibitor-TS complex was determined. A methyl group was added to that structure to take advantage of nearby hydrophobic interactions with Trp80. Next, to make the final inhibitor (21) (Figure 8), an amino group was added at the C-2 position of the quinazoline ring to make a favorable hydrogen bond with the carbonyl of Ala263. In these two steps, the potency of the drug was increased from 0.96 μM to 15 nM. In addition, other structure-activity relationships, such as the addition of gluamates or aliphatic groups to the pyridine ring, were discovered so that lipophilicity can be easily altered in future experiments.

Shoichet et al. (1993) used another approach to the problem, this one exploring the possibility of doing a random screening test via a computer instead of at a lab bench. Using a computer docking program (DOCK) to screen compounds in the Fine Chemical Directory for binding to *L. casei* TS, the compounds were positioned on the enzyme surface and were then ranked by steric and electronic considerations. The substrate and some known inhibitors received good scores, as well as some previously unknown compounds. One of these, sulisobenzone, was further characterized by crystallization. Interestingly, the sulisobenzone bound in a spot that was different from the DOCK-predicted model. Sulisobenzone was found in a pocket that was distinct from any found so far, with elements of both the dUMP and folate binding sites, and the mechanism of inhibition may have to do with a solvent effect. Since this was a new lead, a second and third screening was done by finding a list of compounds with structure similar to sulisobenzone, docking them into the binding site, and then testing the inhibitory activity of the top compounds. From this analysis, derivatives of phenolphthalein (22) (Figure 8) were found to be the most promising leads. More research and design is needed before these compounds are ready for clinical trials, but they provide a completely new method of binding and inhibiting TS.

C.   Dihydrofolate Reductase and Cancer

The most famous inhibitor of DHFR is methotrexate (MTX) (23) (Figure 8), which has been around since the 1950s (Seeger et al., 1949) as an anticancer drug. DHFR catalyzes a reaction that is key in the recovery of folate, which is a cofactor in many important processes and whose inhibition is fatal to the cell. The structure and differences between vertebrate and bacterial DHFR are discussed in the section on antibacterial agents; here the focus is only on recent developments in the targeting of human DHFR as a cancer treatment.

There are some problems inherent in the use of inhibitors of DHFR to treat cancer. The first is that the need for folate in the cell is widespread, and even cells that are not growing need folate for function. Therefore, DHFR inhibitors are somewhat toxic to normal cells as well as those that need to be targeted, and pursuing an aggressive therapy would be fatal to the patient. A second problem is that some tumor cell lines show resistance to treatment with MTX, limiting the use of the drug. Two mechanisms of resistance have been found. The first is changes

in the cellular transport system that moves the drug into the cell such as inhibition of polyglutamation of the inhibitor by other enzymes (Li et al., 1992). The second is mutations in DHFR itself that lower binding of MTX, while leaving catalytic functions unimpaired (Blakley et al., 1993).

Several inhibitor-human DHFR complexes (Cody et al., 1993) have been solved. The compounds tested are more lipophilic than MTX, and therefore should have better transport into the cell, since they can pass through the membrane independent of cellular transport systems, avoiding this method of resistance. These compounds, 10-ethyl-10-deazaminopterin, piritrexim, and trimetrexate, all bind into the same site as MTX (Figure 9), and may be useful in cases where MTX fails.

If one transfects healthy cells with a MTX-resistant DHFR, those cells can tolerate the drug and survive while the cancer cells are killed off. This sort of system has worked when murine bone marrow cells infected by recombinant retroviruses containing a MTX- insensitive DHFR *in vitro* were transplanted into cancerous mice (Williams et al., 1987; Corey et al., 1990). The expressed cells showed a better survival rate in the presence of otherwise lethal doses of MTX, supporting the hypothesis. Crystallography has also helped here in the search for the best DHFR mutant to use, i.e., design of a protein that will not fit a ligand molecule. Two series of mutants were studied, at Leu 22 (Lewis et al., 1995) and at Phe 31 (Chunduru
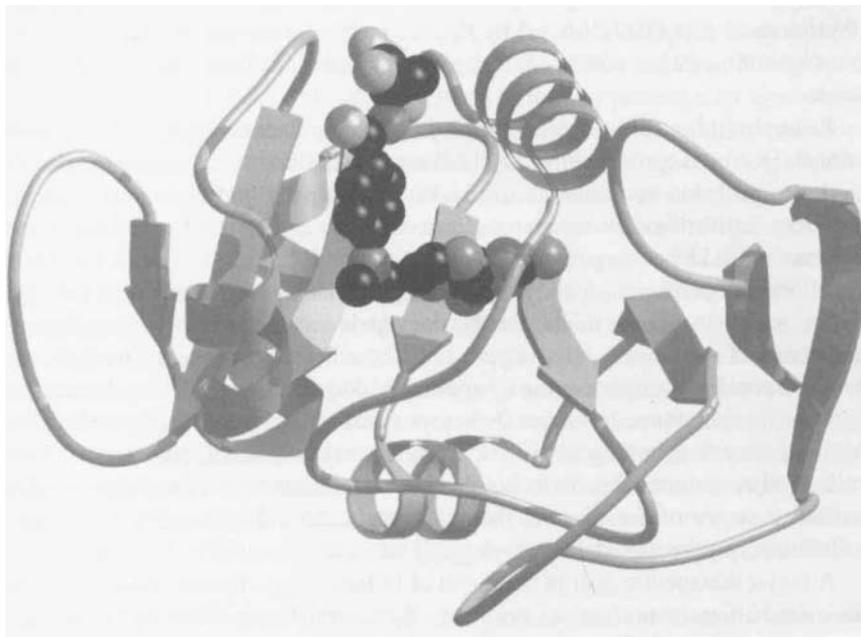

Figure 9.   Dihydrofolate reductase complexed with methotrexate (PDB # 4DFR from Bolin et al., 1982).

et al., 1994). In each series, mutants were found with higher k_i for MTX than wild-type and k_{cat} was comparable. The differential binding in the case of Phe 31 was easily rationalized, as Phe 31 is close to the binding site of the pteridine ring of MTX. The best of these mutants was the Gly 31. The decreased binding of the Leu 22 mutants was attributed to an increase in k_{off} caused by the mutations. The best of these was a Tyr 22. The Leu mutants showed more favorable k_{cat} values than the Phe mutants, so these appear to be the ones to try in conferring resistance to cells during aggressive treatment with MTX.

IX.   INFLAMMATION

The inflammatory response to local tissue injury is marked by several physiological processes that facilitate healing. Immediately surrounding the injury site, capillary permeability increases, which tends to produce edema. Leukocytes brought to the area adhere to blood vessel walls, and damaged cells are removed and destroyed. Vasodilation produces redness and local temperature rises. Various effector proteins, such as growth factors, stimulate anabolic reactions that aid wound healing. Systemic inflammatory reactions include fever, which raises the body temperature to make a less hospitable environment for growth of invading pathogens. The biology of the inflammatory response is extremely complex as it involves many physiological processes directed by diverse molecular mediators. These include prostaglandins, thromboxanes, leukotrienes, pyrogens, and cytokines, to name but a few.

Pathophysiological inflammation plays a role in diseases such as rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE, commonly known as lupus). Both are probable autoimmune disorders in which the body's immune system produces autoantibodies against its own tissues, leading to tissue injury and inflammation. Depending upon the autoantigen and the organ systems involved, symptoms range from fever and malaise to significant, or even fatal, organ damage. In RA, which attacks the joints, chronic damage is accompanied by pain, stiffness, and in more severe cases, deformity. SLE affects connective tissue, so damage may be widespread and cause varying symptoms. Milder forms of SLE may be marked by arthritic symptoms, fever, headaches, or rash, while severe forms, which may be life-threatening, may seriously compromise renal, central nervous system, blood cell, or other systems. Osteoarthritis also has an inflammatory component in more advanced stages of the disease, though it begins as a degenerative rather than inflammatory joint disorder.

A major therapeutic goal in treatment of inflammatory disease is to control the undesired inflammation and its symptoms by administering either corticosteroids or non- steroid anti-inflammatory drugs (NSAIDs), and other avenues of intervention are being sought to increase effectiveness or reduce side effects. Dose regimens of systemic (as opposed to topical) corticosteroids must be carefully monitored to

minimize adverse effects and avoid compromising the normal immune response and adrenal gland function. However, their use is indicated in severe disease to suppress inflammation. NSAIDs, a heterogeneous group that includes aspirin, ibuprofin, and indomethicin (but not acetaminophen or paracetamol), are used more often for milder cases. While disease symptoms may recede with NSAID administration, tissue damage may continue. Thus a regimen based on more than one drug may be advised, since the drugs may affect different aspects of the inflammatory process.

A. Prostaglandin Inhibitors

The prostaglandins (PGs) act on different organ systems throughout the body and exhibit diverse cellular activities, including roles in inflammation, regulation of blood vessel tone, renin release, uterine contraction, and platelet aggregation. These fatty acids are derived from the parent compound, prostanoic acid, which has a cyclopentane ring and a 20-carbon chain. Substitutions on the ring and hydroxyl groups and double bonds on the hydrocarbon chain distinguish different prostaglandins chemically and functionally. Prostaglandins, thromboxanes, and leukotrienes are biologically synthesized from arachidonic acid in a series of steps. In the first, arachidonic acid is converted to prostaglandin H_2 (PGH_2) by prostaglandin H_2 synthase (PGHS). PGHS is a bifunctional enzyme that catalyzes a cyclooxygenase reaction to produce prostaglandin G_2 (PGG_2), followed by a peroxidase reaction that converts PGG_2 to PGH_2. Other PGs and thromboxanes are biosynthesized from PGH_2 by other enzymes including prostacyclin synthetase ($PGH_2 \rightarrow PGI_2$) and thromboxane synthetase ($PGH_2 \rightarrow TXA_2$). Lipoxygenase converts arachidonic acid to leukotrienes (LTs) through an intermediate, H-PETE.

NSAIDs frequently target enzymes in the PG biosynthetic pathway, though they may have other anti-inflammatory mechanisms such as direct action upon the membrane (Kitsis et al., 1991). The best known PG biosynthetic enzyme, and the only one for which a crystal structure exists, is PGHS (also known as COX or cyclooxygenase). PGHS is inhibited by salicylates such as aspirin and by propionic acid derivatives such as ibuprofin. Prostaglandin biosynthesis can be inhibited through other enzymes as well: prostacyclin synthase activity by certain hydroperoxyarachidonic or -linoleic acid derivatives, and thromboxane synthase activity by indole derivatives. As can be seen, different NSAIDs may not block the same PG enzyme target, which provides a rationale for combination drug therapy. Leukotrienes may be even more potent mediators of the inflammatory response than PGs, and LT biosynthesis inhibitors are being developed as alternative anti-inflammatory drugs.

The enzyme PGHS is also one of the very few integral membrane proteins to yield to crystallographic analysis. The enzyme is not membrane-spanning but contains one membrane-inserted domain; hence PGHS has been termed a "monotopic" (one-sided) membrane protein. The crystal structure of PGHS-1 (Picot et al.,

1994) reveals a mostly a-helical protein, a homodimer of 70 kDa subunits. Each subunit has three domains. The first is an epidermal growth factor domain, a structural unit found in viral coat and other proteins, that forms disulfide bonds to the rest of the PGHS molecule. The second domain contains several amphipathic helices and appears to be responsible for anchoring the enzyme to the membrane. The third, the globular body of PGHS, is the catalytic subunit (Figure 10) and contains two distinct active sites, one for each catalytic function. The peroxidase active site contains a heme group. The cyclooxygenase site, the NSAID target, is a long (25Å), narrow (6Å) channel that extends from the membrane-binding domain to the center of the catalytic domain. The catalytic center is believed to be at the apex of this channel, near the edge of the heme. A tyrosine residue in the active site has been implicated in cyclooxygenase catalysis. As PGHS is crystallized with the inhibitor flurbiprofen, a propionic acid NSAID, the exact substrate binding site is not known. However, the position of flurbiprofen in the channel clearly blocks substrate access to the upper channel region and to the proposed catalytic tyrosine.

Figure 10. Prostaglandin H2 synthase. The view is of the cyclooxygenase active site looking down the channel toward the heme molecule at the back. Important residues are highlighted, including Ser 530, which is the target for aspirin (PDB # 1PRH from Picot et al., 1994).

PGHS occurs as two isozymes (COX-1 and COX-2) that differ in expression, tissue distribution, and probably physiological function, though both are found within the endoplasmic reticulum. The two isozymes are very similar and share 60% sequence identity. PGHS-1 (COX-1) is the ubiquitous, constitutive isozyme, while PGHS-2 (COX-2) appears to be induced by inflammatory or mitogenic stimili, and only in a few tissues. While only PGHS-1 has been characterized crystal-lographically, sequence homology between the two isozymes suggests that active site features, including a channel leading to the catalytic center, are conserved.

While PGHS-1 provides a convenient structural model for PGHS-2, the two isozymes are not always inhibited by drugs in the same way. Flurbiprofen and ibuprofen inhibit both isozymes equally well, but other NSAID discriminate between them. Indomethacin prefers PGHS-1 (10-30-fold selection), while 6-methoxy-2-naphthylacetic acid prefers PGHS-2 (sevenfold selection). Aspirin (acetylsalicylate) binds both isozymes but inhibits their prostaglandin H2 produc-tion through different mechanisms. In the case of PGHS-1, aspirin, unlike flurbipro-fen, forms a covalent attachment to (acetylates) a serine residue in the PGHS-1 channel near the active site. The crystal structure of a complex between PGHS-1 and bromylated aspirin (Loll et al., 1995) shows no significant conformational differences between flurbiprofen or aspririn complexes with PGHS-1, and common features of the complexes suggest that both drugs share the same inhibitory mechanism—preventing prostaglandin access to the catalytic tyrosine residue. However, in the PGHS-1/bromoaspirin crystal structure, an alternate position is observed for the acetylated serine sidechain. This more open position allows some access to the active site. This second conformation may correspond to the site occupied by aspirin in PGHS-2, which has not been observed crystallographically. Such a conformational difference might explain why aspirin inactivates PGHS-1 but catalyzes a slightly different reaction in PGHS-2 to produce an alternate product, 15-R-hydroxyeicosatetranoic acid (15-R-HETE), instead of prostaglandin H2 (Allen, 1995).

B. CD4

CD4 is an important cellular receptor that is involved in T cell activation by interaction with the major histocompatibility complex (MHC) of an antigen-pre-senting cell in concert with the T cell receptor. CD4 has been implicated in some autoimmune diseases, like multiple sclerosis, and is also known to be the receptor for the binding of the HIV virus through the gp120 viral coat protein. The structure, an N-terminal fragment of CD4, has been solved, and the overall fold is disulfide-linked β-sheets like an immunoglobulin (Ryu et al., 1990; Wang et al., 1990). The regions of the structure that are important for HIV binding by a protein called gp120 and MHC recognition have been mapped out to various surface loops. Therefore, there is some interest in developing peptide mimetics of these surface loops for use as CD4 antagonists.

There are problems with peptide mimetics, namely, that they have poor biological half- lives due to endogenous proteases and antibodies generated against the peptide. An interesting solution to this problem with CD4 has recently come to light (Jameson et al., 1994) through structural knowledge of CD4. The CDR3-like (complementarity determining region 3 from immunoglobulins) of CD4 forms a loop that was known to be therapeutic in cases of experimental allergic encephalomyelitis in mice, which is used as a model for multiple sclerosis. Replacement of the normal L-amino acids with D-amino acids gives a better half life, but the structure of the loop is now the mirror image of the one desired for activity. By inspection, the workers noticed that if the sequence is reversed, so that the C-terminus becomes the N-terminus and vice-versa, the result is a structure that looks like the original surface, only the carbonyl and amino groups of the backbone are swapped (Figure 11). This new peptide, backwards and composed of L- amino acids, showed good activity in mice. Although this technique is limited by the sequence of the peptide (for example, β-branched side chains would have the wrong stereochemistry) and folding considerations, it is an interesting
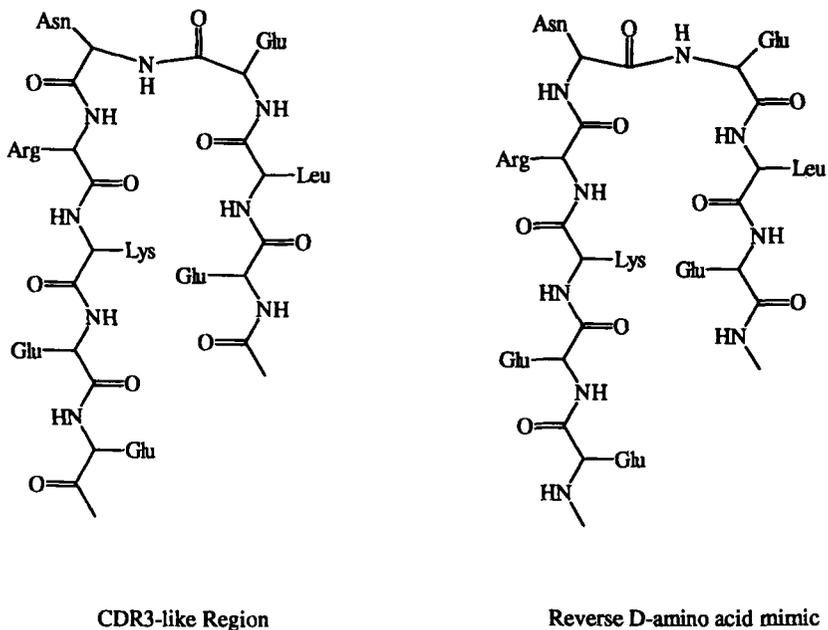

CDR3-like Region   Reverse D-amino acid mimic

Figure 11.   Designing CD4 mimics. On the right is the sequence of a loop on CD4 that was known to be therapeutic in allergic encephalomyelitis. The sequence on the left is the same sequence, only the C-terminus and N-terminus are swapped. If one uses D-amino acids instead of L-amino acids, the surface of the two loops is identical (Jameson et al., 1994).

answer to the problem of peptides as drugs, and a breakthrough that would not have been possible without prior structural knowledge.

C. Interleukin-1β and Interleukin-1β-Converting Enzyme

The cytokine family, which includes interleukins, growth factors, colony stimulating factors, erythropoietin, tumor necrosis factor, and interferons, are proteins that mediate cellular processes such as growth, differentiation, and the immune and inflammatory responses (see, for example, Arai et al., 1990). Interleukin-1β (IL-1β) has been implicated in the pathophysiology of diseases such as RA, septic shock, and even insulin-dependent diabetes mellitus (Dinarello and Wolff, 1993).

IL-1β is expressed primarily by activated monocytes and acts upon a wide range of cell types and tissues. Several proinflammatory roles are associated with this cytokine, which is found in synovial fluid of RA patients and has been shown to promote arthritis in animal models. IL-1β is associated with several activities central to the inflammatory response, e.g., it induces prostaglandin and collagenase production and promotes cartilage degradation and bone resorption. The crystal structure of this cytokine (Finzel et al., 1989; Priestle et al., 1989) shows a β-barrel protein with 12 antiparallel β-strands and no α-helix. A very similar three-dimensional topology is observed for human basic fibroblast growth factor, a mitogen involved in cell proliferation and differentiation, despite very low (10%) sequence identity (Zhang et al., 1991). Both proteins may belong to a larger, structurally homologous family of mitogenic factors.

IL-1β antagonists and monoclonals have not been very successful as potential anti-inflammatory drugs. A better approach has been through interleukin-1b-converting enzyme (ICE), which produces the active form of IL-1β by proteolytic cleavage of a precursor molecule pro-IL-1β. ICE is a cysteine protease, but has a novel, α/β topology not found in other cysteine proteases such as papain and liver cathepsin B. The cytosolic enzyme consists of a homodimer, where each monomer contains a p20 (20 kDa) and a p10 (10 kDa) subunit. The two subunits are themselves produced by cleavage from a proenzyme. Both subunits are required for activity. The crystal structure of ICE complexed with a tetrapeptide chloromethylketone (Walker et al., 1994) shows that the inhibitor binds in a substrate-binding groove on the enzyme surface, while the phenyl ring of the tetrapeptide P4-tyr fits into a hydrophobic channel. This crystal structure, and others likely to follow, show considerable promise for targeted drug design. Interest in ICE as a possible drug target has been increased by its sequence homology to the *Caenorhabitis elegans* cell death gene product, CED3, which is essential for apoptosis.

D. Phospholipase A$_2$

The synovial membrane in joints typically becomes inflamed in diseases such as RA. Several proteins that are overproduced in synovial fluid during such

inflammatory conditions include phospholipases and interleukins. Both of these families are under investigation as mediators of the inflammatory response and therefore as possible targets for anti- inflammatory drug design.

Phospholipase A_2 (PLA$_2$) is a phospholipid hydrolase that initiates the inflammatory response by releasing from cell membranes the arachidonic acid that is subsequently converted into prostaglandins, leukotrienes, and thromboxanes. There are two structurally distinct PLA$_2$ families: secretory enzymes from pancreatic and venom sources, and cytoplasmic PLA$_2$. Several secretory PLA$_2$ have been characterized crystallographically, including one from human rheumatoid arthritis synovial fluid (Wery et al., 1991). These enzymes are small (~14 kDa), disulfide-bonded, calcium-binding proteins. Crystallographic complexes of several enzyme-inhibitor complexes have elucidated much of the cataytic mechanism (e.g., Thunnissen et al., 1990; Scott and Sigler, 1994). In contrast, cytoplasmic PLA$_2$ have not been well characterized structurally. These enzymes are several times larger than their extracellular counterparts and contain multiple domains. A major difficulty with targeting approaches against inflammation through phospholipase A_2, aside from lack of structural data on cytoplasmic PLA$_2$, is that it is not clear which form is primarily responsible for inflammatory disease. Until this biology is better understood, other approaches to treating inflammation promise better results in the near term.

X.   ON THE HORIZON

The future prospects for crystallography-based approaches to clinical problems are increasingly bright. Advances in computation, structure determination methods, and molecular biology have opened up the field for new advances in molecular medicine. At present, there are far more case histories of interest than could have been included in this chapter. Apologies are begged for the inevitable omissions and oversights; there simply is not space to adequately cover protein-nucleic acid interactions or antibody-related structures, but these are reviewed elsewhere. However, we end with a partial listing of what are some of the more interesting recent crystal structures that have some potential for use in understanding and treating human disease. The landmark structure of the human growth hormone complexed with its receptor (De Vos et al., 1992) has been followed by several other hormone or cytokine-receptor complexes. Studies of human growth hormone relate to hematopoietins (Wells and De Vos, 1993). Work in this area could lead to methods of improved generation of blood cells after chemotherapy, for example. Studies conducted with the tumor necrosis factor (TNF), a useful drug target in its own right, shed light on cytokine-receptor interactions. There are crystal structures of TNFs alone (Eck and Sprang, 1989; Jones et al. 1989), a TNF-TNF receptor complex (Banner et al., 1993), and the unliganded TNF receptor extracellular domain (Naismith et al., 1995). The structure of the tyrosine kinase domain of the

human insulin receptor (Hubbard et al., 1994) adds another key element to the understanding of this clinically important system.

The recently-characterized ternary complex (Griffith et al., 1995) between calcineurin (protein phosphatase 2B), FKBP12, and FK506 is another important and exciting structure. Calcineurin, which is activated by Ca^{2+}/calmodulin, plays an essential role in the T cell activation pathway. Cyclosporin A (CsA) and FK506 are two macrocyclic natural products that, in association with their binding proteins cyclophilin (CyP) and FKBP12, respectively, inhibit calcineurin. This inhibition blocks a critical step in T cell proliferation. Because of this biological role, CsA and FK506 are both used as immunosuppresant drugs to prevent posttransplantation graft rejection and to treat some autoimmune diseases. The complex between phorbol ester and the Cys2 activator-binding domain of protein kinase Cδ has also been determined recently (Zhang et al., 1995), an important advance relevant to the cancer field and others. The crystal structure of the synaptotagmin C2 domain (Sutton et al., 1995), a calcium-binding domain with homology to protein kinase C and cytoplasmic PLA_2 contributes new structural data about these important signal transduction enzymes. This very brief list shows that crystallographic problems long considered intractable experimentally are now yielding to more advanced technology. These and other discoveries will help speed progress in preventing, understanding, and treating human disease.

ACKNOWLEDGEMENTS

REFERENCES

Abola, E.E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F., & Weng, J. (1987). In: "Protein Data Bank" in Crystallographic Databases—Information Content, Software Systems, Scientific Applications. (Allen, F.H., Bergerhoff, G., & Sievers, R. eds.), pp. 107-132. Data Commission of the International Union of Crystallography, Bonn/Cambridge/Chester.

Allen, K.N. (1995). Aspirin—now we can see it. Nature Medicine 1, 882-883.

Appelt, K., Bacquet, R.J., Bartlett, C.A., Booth, C.L.J., Freer, S.T., Fuhry, M.A., Gehring, M.R., Herrmann, S.M., Howland, E.F., Janson, C.A., Jones, T.R., Kan, C., Kathardekar, V., Lewis, K.K., Marzoni, G.P., Matthews, D.A., Mohr, C., Moomaw, E.W., Morse, C.A., Oatley, S.J., Ogden, R.C., Reddy, M.R., Reich, S.H., Schoettlin, W.S., Smith, W.W., Varney, M.D., Villafranca, J.E., Ward, R.W., Webber, S., Webber, S.E., Welsh, K.M., & White, J. (1991).

Design of enzyme inhibitors using iterative protein crystallographic analysis. J. Med. Chem. 34, 1925-1934.

Arai, K., Lee, F., Miyajima, A., Miyatake, S., Arai, N., & Yokota, T. (1990). Cytokines: Coordinators of immune and inflammatory responses. Ann. Rev. Biochem. 59, 783-836.

Arnold, E., Jacobo-Molina, A., Nanni, R.G., Williams, R.L., Lu, X., Ding, J., Clark, A.D., Jr., Zhang, A., Ferris A.L., Clark, P., Hizi, A., & Hughes, S.H. (1992). Structure of HIV-1 reverse transcriptase/DNA complex at 7 Å resolution showing active site locations. Nature 357, 85-90.

Baldwin, J.J., Ponticello, G.S., Anderson, P.S., Christy, M.E., Murcko, M.A., Randall, W.C., Schwam, H., Sugrue, M.F., Springer, J.P., Gautheron, P., Grove, J., Mallorga, P., Viader, M.-P., McKeever, B.M., & Navia, M.A. (1989). Thienopyran-2-sulfonamides: Novel topically active carbonic anhydrase inhibitors for the treatment of glaucoma. J. Med. Chem. 32, 2510-2513.

Banerjee, A., Dubnau, E., Quemard, A., Balasubramanian, V., Um, K.S., Wilson, T., Collins, D., de Lisle, G., & Jacobs, W.R., Jr. (1994). inhA, a gene encoding a target for isoniazid and ethionamide in Mycobacterium tuberculosis. Science 263, 227-230.

Banner, D.W., D-Arcy, A., Janes, W., Gentz, R., Schoenfeld, H.-J., Broger, C., Loetscher, H., & Lesslauer, W. (1993). Crystal structure of the soluble human 55kd TNF receptor-human TNFβ complex: Implications for TNF receptor activation. Cell 73, 431-445.

Banner, D.W., & Hadváry, P. (1991). Crystallographic analysis at 3.0-Å resolution of the binding to human thrombin of four active site-directed inhibitors. J. Biol. Chem. 266, 20085-20093.

Bernhardt, G., Harber, J., Zibert, A., deCrombrugghe, M., & Wimmer, E. (1994) The poliovirus receptor: Identification of domains and amino acid residues critical for virus binding. Virology 203, 344-356.

Blakley, R.L., Appleman, J.R., Chunduru, S.K., Nakano, T., Lewis, W.S., & Harris, S.E. (1993). In: Chemistry and Biology of Pteridines and Folates. (Ayling, J.E., Nair, M.G., and Baugh, C.M., eds.), pp. 473-478, Plenum Press, New York.

Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S.R. & Hofsteenge, J. (1989). The refined 1.9Å crystal structure of human α-thrombin: Interaction with D-Phe-Pro-Arg chloromethyl-ketone and significance of the Tyr-Pro-Pro-Trp insertion segment. EMBO J. 8, 3467-3475.

Bode, W., Turk, D. & Stürzebecher, J. (1990) Geometry of binding of the nezamidine- and arginine-based inhibitors NAPAP and MQPA to human α-thrombin: X-ray crystallographic determination of the NAPAP-trypsin complex and the modelling of NAPAP-thrombin and MQPA-thrombin. Eur. J. Biochem. 193, 175-182.

Bohm, H.-J. (1992). LUDI: Rule-based automatic design of new substituents for enzyme inhibitor leads. J. Comput. Aided Mol. Des. 6, 593-606.

Bolin, J.T., Filman, D.J., Matthews, D.A., Hamlin, R.C., & Kraut, J. (1982). Crystal structures of Escherichia coli and Lactobacillus casei dihydrofolate reductase refined at 1.7 Å resolution. I. general features and binding of methotrexate. J. Biol. Chem. 257, 13650-13662.

Brandstetter, H., Turk, D., Hoeffken, H.W., Grosse, D., Stürzebecher, J., Martin, P.D., Edwards, B.F.P., & Bode, W. (1992). Refined 2.3 Å x-ray crystal structure of bovine thrombin complexes formed with the benzamidine and arginine-based thrombin inhibitors NAPAP, 4-TAPAP and MQPA: A starting point for improving antithrombotics. J. Mol. Biol. 226, 1085-1099.

Bugg, C.E., Carson, W.M., & Montgomery, J.A. (1993). Drugs by design. Sci. Amer. 6, 92-98.

Bullough, P.A., Hughson, F.M., Skehel, J.J., & Wiley, D.C. (1994). Structure of influenza haemagglutinin at the pH of membrane fusion. Nature 371, 37-43.

Burchall, J.J., & Hitchings, G.H. (1965). Inhibitor binding analysis of dihydrofolate reductases from various species. Mol. Pharmacol. 1, 126-136.

Chen, C.C.H., & Herzberg, O. (1992). Inhibition of β-lactamase by clavulanate: Trapped intermediates in cryocrystallographic studies. J. Mol. Biol. 224, 1103-1113.

Chen, C.C.H., Rahil, J., Pratt, R.F., & Herzberg, O. (1993). Structure of a phosphonate-inhibited β-lactamase: An analog of the tetrahedral transition state/intermediate of β-lactam hydrolysis. J. Mol. Biol. 234, 165-178.

Chunduru, S.K., Cody, V., Luft, J.R., Pangborn, W., Appleman, J.R., & Blakley, R.L. (1994). Methotrexate-resistant variants of human dihydrofolate reductase: Effects of Phe31 substitutions. J. Biol. Chem. 269, 9547-9555.

Clarke, S.J., Jackman, A.L., & Judson, I.R. (1993). In: Novel Approaches to Selective Treatments of Solid Human Tumors: Laboratory and Clinical Correlation. (Rustum, Y.M., ed.), pp. 277-287. Plenum Press, New York.

Cody, V., Wotjtczak, A., Kalman, T.I., Friesheim, J.H., & Blakley, R.L. (1993). In: Chemistry and Biology of Pteridines and Folates. (Ayling, J.E., Nair, M.G., & Baugh, C.M., eds.), pp. 481-486. Plenum Press, New York.

Colman, P.M. (1992). Structural basis of antigenic variation: Studies of influenza virus neuraminidase. Immun. Cell Biol. 70 (pt 3), 209-214.

Colman, P.M. (1994). Effects of amino acids sequence changes on antibody-antigen interactions. Res. Immun. 145, 33-36.

Colman, P.M., Varghese, J.N., & Laver, W.G. (1983). Structure of the catalytic and antigenic sites in influenza virus neuraminidase. Nature 303, 41-44.

Colonno, R.J., Condra, J.H., Mizutani, S., Callahan, P.L., Davies, M.E., & Murcko, M.A. (1988). Evidence for the direct involvement of the rhinovirus canyon in receptor binding. Proc. Natl. Acad. Sci. USA 85, 5449-5453.

Corey, C.A., DeSilva, A.D., Holland, C.A., & Williams, D.A. (1990). Serial transplantation of methotrexate-resistant bone marrow: Protection of murine recipients from drug toxicity by progeny of transduced stem cells. Blood 75, 337-343.

Davies, J.F., II, Delcamp, T.J., Prendergast, N.J., Ashford, V.A., Freisheim, J.H., & Kraut, J. (1990). Crystal structures of recombinant human dihydrofolate reductase complexed with folate and 5-deazafolate. Biochemistry 29, 9467-9479.

Dessen, A., Quemard, A., Blanchard, J.S., Jacobs, W.R. Jr., & Sacchettini, J.C. (1995). Crystal structure and function of the isoniazid target of *Mycobacterium tuberculosis*. Science 267, 1638-1641.

De Vos, A.M., Ultsch, M., & Kossiakoff, A.A. (1992). Human growth hormone and the extracellular domain of its receptor: Crystal structure of the complex. Science 255, 306-312.

Diana, G.D., Kowalczyk, P., Treasurywala, A.M., Oglesby, R.C., Pevear, V.C., & Dutko, F.J. (1992). CoMFA analysis of the interactions of antipicornavirus compounds in the binding pocket of human rhinovirus-14 [erratum appears in J. Med. Chem. (1992) Dec 11; 35(25):4768]. J. Med. Chem. 35, 1002-1008.

Dideburg, O., Charlier, P., Wery, J.-P., Dehottay, P., Dusart, J., Erpicum, T., Frere, J.-M., & Ghuysen, J.-M. (1987). The crystal structure of the α-lactamase of *Streptomyces albus* G at 0.3 nm resolution. Biochem. J. 245, 911-913.

Dinarello, C.A., & Wolff, S.M. (1993). The role of interleukin-1 in disease. N. Engl. J. Med. 328, 106-113.

Dyda, F., Hickman, A.B., Jenkins, T.M., Engelman, A., Craigie, R., & Davies, D.R. (1994). Crystal structure of the catalytic domain of HIV-1 integrase: Similarity to other polynucleotidyl transferases. Science 266, 1981-1986.

Ealick, S.E., Babu, Y.S., Bugg, C.E., Erion, M.D., Guida, W.C., Montgomery, J.A., & Secrist, J.A., 3rd (1991). Application of crystallographic and modeling methods in the design of purine nucleoside phosphorylase inhibitors. Proc. Natl. Acad. Sci. USA 88, 11540-11544.

Ealick, S.E., Rule, S.A., Carter, D.C., Greenhough, T.J., Babu, Y.S., Cook, W.J., Habash, J., Helliwell, J.R., Stoeckler, J.D., Parks, R.E., Jr., Chen, S., & Bugg, C.E. (1990). Three- dimensional structure of human erythrocytic purine nucleoside phosphorylase at 3.2 Å resolution. J. Biol. Chem. 265, 1812-1820.

Eck, M.J., & Sprang, S.R. (1989). The structure of tumor necrosis factor-α at 2.6Å resolution: Implications for receptor binding. J. Biol. Chem. 264, 17595-17605.

Eisen, M.B., Wiley, D.C., Karplus, M., & Hubbard, R.E. (1994). HOOK: a program for finding novel molecular achitechtures that satisfy the chemical and steric requirements of a macromolecule binding site. Proteins 19, 199-221.

Erickson, J., Neidhart, D.J., VanDrie, J., Kempf, D.J., Wang, X.C., Norbeck, D.W., Plattner, J.J., Rittenhouse, J.W., Turon, M., Wideburg, N., Kohlbrenner, W.E., Simmer, R., Helfrich, R., Paul, D.A., & Knigge, M. (1990). Design, activity, and 2.8Å crystal structure of a C_2 symmetric inhibitor complexed to HIV-1 protease. Science 249, 527-533.

Eriksson, A.E., Jones, T.A., & Liljas, A. (1988). Refined structure of human carbonic anhydrase II at 2.0Å resolution. Prot. Struct. Funct. Genet. 4, 274-282.

Erion, M.D. (1990). European Patent Application 374,096.

Erion, M.D., Niwas, S., Rose, J.D., Ananthan, S., Allen, M., Secrist, J.A., 3d, Babu, Y.S., Bugg, C.E., Guida, W.C., & Ealick S.E. (1993). Structure-based design of inhibitors of purine nucleoside phosphorylase. 3. 9-Arylmethyl derivatives of 9-deazaguanine substituted on the methylene group [erratum appears in J. Med. Chem. (1994) Apr 1; 37(7):1034]. J. Med. Chem. 36, 3771-3783.

Fauman, E.B., Rutenber, E.E., Maley, G.F., Maley, F., & Stroud, R.M. (1994). Water-mediated substrate/product discrimination: The product complex of thymidylate synthase at 1.83 Å. Biochemistry 33, 1502-1511.

Fenton, J.W. II, Ofosu, F.A., Moon, D.G., & Maragonore, J.M. (1991). Thrombin structure and function: Why thrombin is the primary target for antithrombotics. Blood Coagul. Fibrin 2, 69-75.

Finer-Moore, J.S., Fauman, E.B., Foster, P.G., Perry, K.M., Santi, D.V., & Stroud, R.M. (1993). Refined structures of substrate-bound and phosphate-bound thymidylate synthase from *Lactobacillus casei*. J. Mol. Biol. 232, 1101-1116.

Finzel, B.C., Clancy, L.L., Holland, D.R., Muchmore, S.W., Watenpaugh, K.D., & Einspahr, H.M. (1989). Crystal structure of recombinant human interleukin-1β at 2.0Å resolution. J. Mol. Biol. 209, 779-791.

Friedkin, M., & Kalckar, H. (1961). In: The Enzymes 2nd Ed. (Boyer, P.D., Lardy, H., & Myrback, K., eds.), pp. 237-256. Academic Press, New York.

Furie, B. & Furie, B.C. (1992). Molecular and cellular biology of blood coagulation. N. Engl. J. Med. 326, 800-806.

Gehlhaar, D.K., Moerder, K.E., Zichi, D., Sherman, C.J., Ogden, R.C., & Freer, S.T. (1995). *De novo* design of enzyme inhibitors by monte carlo ligand generation. J. Med. Chem. 38, 466-472.

Ghuysen, J.-M. (1991). Serine β-lactamases and penicillin-binding proteins. Annu. Rev. Microbiol. 45, 37-67.

Goldsmith, E.J., & Mottonen, J. (1994). Serpins: The uncut version. Structure 2, 241-244.

Goodford, P.J. (1985). A computational procedure for determining energetically favorable binding sites on biologically important molecules. J. Med. Chem. 28, 849-857.

Grant, R.A., Hiremath, C.N., Filman, D.J., Syed, R., Andries, K., & Hogle, J.M. (1994). Structures of poliovirus complexes with anti-viral drugs: Implications for viral stability and drug design. Curr. Biol. 4, 784-797.

Greer, J., Erickson, J.W., Baldwin, J.J., & Varney, M.D. (1994). Application of the three- dimensional structures of protein target molecules in structure-based drug design. J. Med. Chem. 37, 1035-1054.

Griffith, J.P., Kim, J.L., Kim, E.E., Sintchak, M.D., Thomson, J.A., Fitzgibbon, M.J., Fleming, M.A., Caron, P.R., Hsiao, K., & Navia, M.A. (1995). X-ray structure of calcineurin inhibited by the immunophilin-immunosupressant FKBP12-FK506 complex. Cell 82, 507-522.

Grütter, M.G., Priestle, J.P., Rahuel, J., Grossenbacher, H., Bode, W., Hofsteenge, J. & Stone, S.R. (1990). Crystal structure of the thrombin-hirudin complex: A novel mode of serine protease inhibition. EMBO J. 9, 2361-2365.

Guida, W.C., Elliot, R.D., Thomas, H.J., Secrist, J.A., 3d, Babu, Y.S., Bugg, C.E., Erion, M.D., Ealick, S.E., & Montgomery, J.A. (1994). Structure-based design of inhibitors of purine nucleoside phosphorylase. 4. A study of phosphate mimics. J. Med. Chem. 37, 1109-1114.

Hardy, L.W., Finer-Moore, J.S., Montfort, W.R., Jones, M.O., Santi, D.V., & Stroud, R.M. (1987). Atomic structure of thymidylate synthase: Target for rational drug design. Science 235, 448-455.

Herzberg, O., & Moult, J. (1987). Bacterial resistance to β-lactam antibiotics: Crystal structure of α-lactamase from *Staphylococcus aureus* PC1 at 2.5 Å resolution. Science 236, 694-701.

Hogle, J.M., Chow, M., & Filman, D.J. (1985). Three-dimensional structure of poliovirus at 2.9 Å resolution. Science 229, 1358-1365.

Hubbard, S.R., Wei, L., Ellis, L., & Hendrickson, W.A. (1994). Crystal structure of the tyrosine kinase domain of the human insulin receptor. Nature 372, 746-754.

Hurvitz, L.M., Kaufman, P.L., Robin, A.L., Weinreb, R.N., Crawford, K., & Shaw, B. (1991). New developments in the drug treatment of glaucoma. Drugs 41, 514-532.

Imtiaz, U., Billings, E., Knox, J.R., Manavathu, E.K., Lerner, S.A., & Mobashery, S. (1993). Inactivation of class A β-lactamases by clavulanic acid: The role of arginine-244 in a proposed nonconcerted sequence of events. J. Am. Chem. Soc. 115, 4435-4442.

Imtiaz, U., Billings, E.M., Knox, J.R., & Mobashery, S. (1994). A structure-based analysis of the inhibition of class A β-lactamases by sulbactam. Biochemistry 33, 5728-5738.

Jacobo-Molina, A., Ding, J., Nanni, R.G., Clark, A.D., Lu, X., Tantillo, C., Williams, R.L., Kamer, G., Ferris, A.L., Clark, P., Hizi, A., Hughes, S.H., & Arnold, E. (1993). Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. Proc. Natl. Acad. Sci. USA 90, 6320-6324.

Jameson, B.A., McDonnell, J.M., Marini, J.C., & Korngold, R. (1994). A rationally designed CD4 analogue inhibits experimental allergic encephalomyelitis. Nature 368, 744-746.

Janakiraman, M.N., White, C.L., Laver, W.G., Air, G.M., & Luo, M. (1994). Structure of influenza virus neuraminidase B/Lee/40 complexed with sialic acid and a dehydro analog at 1.8 Å resolution: Implications for the catalytic mechanism. Biochemistry 33, 8172-8179.

Jedrzejas, M.J., Singh, S., Brouillette, W.J., Laver, W.G., Air, G.M., & Luo, M. (1995). Structures of aromatic inhibitors of influenza virus neuraminidase. Biochemistry 34, 3144-3151.

Jelsh, C., Mourey, L., Masson, J.M., & Samama, J.P. (1993). Crystal structure of *Escherichia coli* TEM-1 β-lactamase at 1.8 Å resolution. Proteins: Struct. Funct. Genet. 16, 364-383.

Jones, E.Y., Stuart, D.I., & Walker, N.P.C. (1989). Structure of tumour necrosis factor. Nature 338, 225-228.

Kamb, A., Finer-Moore, J.S., & Stroud, R.M. (1992). Cofactor triggers the conformational change in thymidylate synthase: Implications for an ordered binding mechanism. Biochemistry 31, 12876-12884.

Kim, E.E., Baker, C.T., Dwyer, M.D., Murcko, M.A., Rao, B.G., Tung, R.D., & Navia, M.A. (1995). Crystal structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme. J. Am. Chem. Soc. 117, 1181-1182.

Kitsis, E.A., Weissmann, G., & Abramson, S.B. (1991). The prostaglandin paradox: Additive inhibition of neutrophil function by aspirin-like drugs and the prostaglandin E1 analog misoprostal. J. Rheumatology. 18, 1461-1465.

Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A., & Steitz, T.A. (1992). Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. Science 256, 1783-1790.

Kraulis, P.J. (1991). MOLSCRIPT: A program to produce both detailed and schematic plots of protein structure. J. Appl. Cryst. 24, 946-950.

Kraut, J., & Matthews, D.A. (1987). In: Biological Macromolecules and Assemblies, Vol. III (Jurnak, F. & McPherson, A., eds.). John Wiley & Sons, New York.

Krenitsky, T.A. (1967). Purine nucleoside phosphorylase: Kinetics, mechanism, and specificity. Mol. Pharmacol. 3, 526-536.

Kuyper, L.F. (1989). In: Computer Aided Drug Design: Methods and Applications (Perun, T.J., & Propst, C.L., eds.), pp. 327-367. Marcel Dekker, Inc., New York.

Leach A.R., & Kuntz, I.D. (1992). Conformational analysis of flexible ligands in macromolecular receptor sites. J. Comput. Chem. 13, 730-748.

LePage, G.A., Junga, I.G., & Bowman, B. (1964). Biochemical and carcinostatic effects of 2′-thioguanosine. Cancer Res. 24, 835-840.

Lewis, W.S., Cody, V., Galitsky, N., Luft, J.R., Pangborn, W., Chunduru, S.K., Spencer, H.T., Appleman, J.R., & Blakley, R.L. (1995). Methotrexate-resistant variants of human dihydrofolate reductase with substitutions of leucine 22: Kinetics, crystallography, and potential as selectable markers. J. Biol. Chem. 270, 5057-5064.

Li, W-W., Lin, J.T., Tong, W.P., Trippet, T.M., Brennan, M.F., & Bertino, J.R. (1992). Mechanisms of natural resistance to antifolates in human soft tissue sarcomas. Cancer Res. 52, 1434-1438.

Lobovsky, E., Moews, P.C., Liu, H., Zhao, H., Frere, J.-M., & Knox, J.R. (1993). Evolution of an enzyme activity: Crystallographic structure at 2 Å resolution of cephalosporinase from the *ampC* gene of *Enterobacter cloacae* P99 and comparison with a class A penicillinase. Proc. Natl. Acad. Sci. USA 90, 11257-11261.

Lobovsky, E., Billings, E.M., Moews, P.C., Rahil, J., Pratt, R.F., & Knox, J.R. (1994). Crystallographic structure of a phosphonate derivative of the Enterobacter cloacae P99 cephalosporinase: Mechanistic interpretation of a β-lactamase transition state analog. Biochemistry 33, 6762-6772.

Loll, P.J., Picot, D., & Garavito, R.M. (1995). The structural basis of aspirin activity inferred from the crystal structure of inactivated prostaglandin H_2 synthase. Nature Structural Biol. 2, 637-643.

Mangani, S., & Liljas, A. (1993). Crystal structure of the complex between human carbonic anhydrase II and the aromatic inhibitor 1,2,4-triazole. J. Mol. Biol. 232, 9-14.

Matthews, D.A., Appelt, K., Oatley, S.J., & Xuong, N.H. (1990a). Crystal structure of Escherichia coli thymidylate synthase containing bound 5-fluoro-2'-deoxyuridylate and 10-propargyl-5,8-dide-azafolate. J. Mol. Biol. 214, 923-936.

Matthews, D.A., Bolin, J.T., Burridge, J.M., Filman, D.J., Volz, K.W., & Kraut, J. (1985). Dihydrofolate reductase. The stereochemistry of inhibitor selectivity. J. Biol. Chem. 260, 392-399.

Matthews, D.A., Villafranca, J.E., Janson, C.A., Smith, W.W., Welsh, K., & Freer, S. (1990b). Stereochemical mechanism of action for thymidylate synthase based on the x-ray structure of the covalent inhibitory ternary complex with 5-fluoro-2'-deoxyuridylate and 5,10-methylenetetrahydrofolate. J. Mol. Biol. 214, 937-948.

Middlebrook, G. (1952). Sterilization of tubercule bacilli by isonicotinic acid hydrazide and the incidence of variants resistant to the drug *in vitro*. Am. Rev. Tuberc. Pulm. Dis. 65, 765-767.

Miller, M., Jaskólski, M., Rao, J.K.M., Leis, J., & Wlodawer, A. (1989). Crystal structure of a retroviral protease proves relationship to aspartic protease family. Nature 337, 576-579.

Miranker, A. & Karplus, M. (1991). Functionality maps of binding sites: A multiple copy simultaneous search method. Proteins 11, 29-34.

Moews, P.C., Knox, J.R., Dideburg, O., Charlier, P., & Frere, J.-M. (1990). α-lactamase of *Bacillius licheniformis* 749/C at 2 Å resolution. Prot. Struct. Funct. Genet. 7, 156-171.

Montgomery, J.A., Niwas, S., Rose, J.D., Secrist, J.A., 3rd, Babu, Y.S., Bugg, C.E., Erion, M.D., Guida, W.C., & Ealick, S.E. (1993). Structure-based design of inhibitors of purine nucleoside phosphorylase. 1. 9-(Arylmethyl) derivatives of 9-deazaguanine. J. Med. Chem. 36, 55-69.

Montgomery, J.A. & Secrist, J.A., 3rd (1994). PNP inhibitors. Prespec. in Drug Discovery Design 2, 205-220.

Mosser, A.G., Sgro, J.-Y., & Rueckert, R.R. (1994). Distribution of drug resistance mutations in type 3 poliovirus identifies three regions involved in uncoating functions. J. Virology 68, 8193-8201.

Naismith, J.H., Devine, T.Q., Brandhuber, B.J., & Sprang, S.R. (in press). Crystallographic evidence for dimerization of unliganded tumor necrosis factor receptor. J. Biol. Chem.

Navia, M.A., Fitzgerald, P.M.D., McKeever, B.M., Leu, C.-T., Heimbach, J.C., Herber, W.K., Sigal, I.S., Darke, P.L., & Springer, J.P. (1989). Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. Nature 337, 615-620.

Navia, M.A., & Murcko, M.A. (1992). Use of structural information in drug design. Curr. Opin. Struct. Biol. 2, 202-210.

Nikaido, H. (1994). Prevention of drug access to bacterial targets: Permeability barriers and active efflux. Science 264, 382-388.

Nishibata, Y., & Itai, A. (1993). Confirmation of usefulness of structure construction program based on three-dimensional receptor structure for rational lead generation. J. Med. Chem. 36, 2921-2928.

Niwas, S., Chand, P., Pathak, V.P., & Montgomery, J.A. (1994). Structure-based design of inhibitors of purine nucleoside phosphorylase. 5. 9-Deazahypoxanthines. J. Med. Chem. 37, 2477-2480.

Norbeck, D.W., & Kempf, D.J. (1991). HIV protease inhibitors. Ann. Rep. Med. Chem. 26, 141-150.

O'Connor, B.M., Webber, S., Jackson, R.C., Galivan, J., & Rhee, M.S. (1994). Biological activity of a novel rationally designed lipophilic thymidylate synthase inhibitor. Cancer Chemother. Pharmacol. 34, 225-229.

Oefner, C., D'Arcy, A., Daly, J.J., Gubernator, K., & Charnas, R.L., Heinze, I., Hubschwerlen, C., & Winkler, F.K. (1990). Refined crystal structure of β-lactamase from *Citrobacter freundii* indicates a mechanism for β-lactam hydrolysis. Nature 343, 284-288.

Olson, N.H., Kolatkar, P.R., Oliveira, M.A., Cheng, R.H., Greve, J.M., McClelland, A., Baker, T.S., & Rossmann, M.G. (1993). Structure of human rhinovirus complexed with its receptor molecule. Proc. Natl. Acad. Sci. USA 90, 507-511.

Parks, R.E., Jr., & Agarwal, R.P. (1972). In: The Enzymes, 3rd Ed. (Boyer, P.D, ed.), pp. 483-495, Academic Press, New York.

Pearlman, D.A., & Murcko, M.A. (1993) CONCEPTS: New dynamic algorithm for *de novo* drug suggestion. J. Compt. Chem. 14, 1184-1193.

Perry, K.M., Fauman E.B., Finer-Moore, J.S., Montfort, W.R., Maley, G.F., Maley, F., & Stroud, R.M. (1990). Plastic adaption toward mutations in proteins: Structural comparison of thymidylate synthases. Proteins: Struct. Funct. Genet. 8, 315-333.

Perutz, M. (1992). In: Protein Structure: New Approaches to Disease and Therapy. W.H. Freeman & Company, New York.

Petteway, S.R., Jr., Lambert, D.M., & Metcalf, B.W. (1991). The chronically infected cell as a target for the treatment of HIV infection and AIDS. Trends Pharm. Sci. 12, 28-34.

Peugeot, R.L., Sams, W.M. Jr., Omura, E.F., Cook, W.J., & Snyder, H.W. Jr. (1995). Clinical efficacy of a novel purine nucleoside phosphorylase inhibitor (BCX-34) in the treatment of cutaneous T-cell lymphoma. Presented at the Meeting of the Society for Investigative Dermatology, Chicago IL, May 25, 1995.

Picot, D., Loll, P.J., & Garavito, R.M. (1994). The x-ray crystal structure of the membrane protein prostaglandin H_2 synthase-1. Nature 367, 243-249.

Priestle, J.P., Schär, H.-P., & Grütter, M.G. (1989). Crystallographic refinement of interleukin 1β at 2.0Å resolution. Proc. Natl. Acad. Sci. USA 86, 9667-9671.

Quemard, A., Lacave, C., & Laneelle, G. (1991). Isoniazid inhibition of mycolic acid synthesis by cell extracts of sensitive and resistant strains of *Mycobacterium aurum*. Antimicrob. Agents Chemother. 35, 1035-1039.

Qiu, X., Padmanabhan, K.P., Carperos, V.E., Tulinsky, A., Kline, T., Maraganore, J.M., & Fenton, J.W. (1992). Structure of the hirulog 3-thrombin complex and nature of the S' subsites of substrates and inhibitors. Biochemistry 31, 11689-11697.

Reyes, V.M., Sawaya, M.R., Brown, K.A., & Kraut, J. (1995). Isomorphous crystal structures of *Escherichia coli* dihydrofolate reductase complexed with folate, 5-deazafolate, and 5,10-dideazatetrahydrofolate: Mechanistic implications. Biochemistry 34, 2710-2723.

Roberts, N.A., Martin, J.A., Kinchington, D., Broadhurst, A.V., Craig, J.C., Duncan, I.B., Galpin, S.A., Handa, B.K., Kay, J., Kröhn, Lambert, R.W., Merrett, J.H., Mills, J.S., Parkes, K.E.B., Redshaw, S., Ritchie, A.J., Taylor, D.L., Thomas, G.J., & Machin, P.J. (1990). Rational design of peptide-based HIV proteinase inhibitors. Science 248, 358-361.

Rodgers, D.W., Gamblin, S.J., Harris, B.A., Ray, S., Culp, J.S., Hellmig, B., Woolf, D.J., Debouck, C., & Harrison, S.C. (1995). The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1. Proc. Natl. Acad. Sci. USA 92, 1222-1226.

Rossmann, M.G. (1989). The canyon hypothesis: Hiding the host cell receptor attachment site on a viral surface from immune surveillance. J. Biol. Chem. 264, 14587-14590.

Roth, B. (1986). Design of dihydrofolate reductase inhibitors from x-ray crystal structures. Fed. Proc. 45, 2765-2772.

Rotstein, S.H., & Murcko, M.A. (1993a). GenStar: A method for *de novo* drug design. J. Comput. Aided Mol. Des. 7, 23-43.

Rotstein, S.H., & Murcko, M.A. (1993b). GroupBuild: A fragment-based method for *de novo* drug design. J. Med. Chem. 36, 1700-1710.

Rydel, T.J., Tulinsky, A., Bode, W., & Huber, R. (1991). Refined structure of the hirudin- thrombin complex. J. Mol. Biol. 221, 583-601.

Ryu, S.-E., Kwong, P.D., Truneh, A., Porter, T.G., Arthos, J., Rosenberg, M., Dai, X., Xuong, N., Axel, R., Sweet, R.W., & Hendrickson, W.A. (1990). Crystal structure of an HIV-binding recombinant fragment of human CD4. Nature 348, 419-426.

Sauter, N.K., Hanson, J.E., Glick, G.D., Brown, J.H., Crowther, R.L., Park, S.J., Skehel, J.J., & Wiley, D.C. (1992). Binding of hemaggutinin to analogs of its cell-surface receptor, sialic acid: Analysis by proton nuclear magnetic resonance spectroscopy and X-ray crystallography. Biochem. 31, 9609-9621.

Scott, D.L., & Sigler, P.B. (1994). Structure and catalytic mechanism of secretory phospholipases A_2. Adv. Prot. Chem. 45, 53-88.

Secrist, J.A., 3rd, Niwas, S., Rose, J.D., Babu, Y.S., Bugg, C.E., Erion, M.D., Guida, W.C., Ealick, S.E., & Montgomery, J.A. (1993). Structure-based design of inhibitors of purine nucleoside phosphorylase. 2. 9-Alicyclic and 9-heteroalicyclic derivatives of 9-deazaguanine. J. Med. Chem. 36, 1847-1854.

Seeger, D.R., Cosulich, D.B., Smith, J.R., Jr., & Hultquist, M.E. (1949). Analogs of pteroylglutamic acid. III. 4-Amino derivatives. J. Am. Chem. Soc. 71, 1753-1758.

Shoichet, B.K., Stroud, R.M., Santi, D.V., Kuntz, I.D., & Perry, K.M. (1993). Structure-based discovery of inhibitors of thymidylate synthase. Science 259, 1445-1450.

Smerdon, S.J., Jager, J., Wang, J., Kohlstaedt, L.A., Chirino, A.J., Friedman, J.M., Rice, P.A., & Steitz, T.A. (1994). Structure of the binding site for non-nucleoside inhibitors of the reverse transcriptase of human immunodeficiency virus type 1. Proc. Natl. Acad. Sci. USA 91, 3911- 3915.

Smith, T.J., Kremer, M.J., Luo, M., Vriend, G., Arnold, E., Kamer, G., Rossmann, M.G., McKinlay, M.A., Diana, G.D., & Otto, M.J. (1986). The site of attachment in human rhinovirus- 14 for antiviral agents that inhibit uncoating. Science 233, 1286-1293.

Snider, D.E., Jr. & Roper, W.L. (1992). The new tuberculosis. N. Engl. J. Med. 326, 703-705.

Spinelli, S., Liu, Q.Z., Alzari, P.M., Hirel, P.H., & Poljak, R.J. (1991). The three-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU. Biochimie. 73, 1391-1396.

Spratt, B.G. (1994). Resistance to antibiotics mediated by target alterations. Science 264, 388- 393.

Stammers, D.K., Champness, J.N., Bedell, C.R., Dann, J.G., Eliopoulos, E., Geddes, A.J., Ogg, D., & North, A.C.T. (1987). The structure of mouse L1210 dihydrofolate reductase-drug complexes and the construction of a model of human enzyme. FEBS Lett. 218, 178-184.

St. Georgiev, V. (1993). Enzymes of purine metabolism: Inhibition and therapeutic potential. Ann. N.Y. Acad. Sci. 685, 207-216.

Stone, R.S. & Hofsteenge, J. (1986) Kinetics of the inhibition of thrombin by hirudin. Biochemistry 25, 4622-4628.

Strynadka, N.C.J., Adachi, H., Jensen, S.E, Johns, K., Sielecki, A., Betzel, C., Sutoh, K., & James, M.N.G. (1992). Molecular structure of the acyl-enzyme intermediate in β-lactam hydrolysis at 1.7 Å resolution. Nature 359, 700-705.

Strynadka, N.C.J., Jensen, S.E., Johns, K., Blanchard, H., Page, M., Matagne, A., Frere, J.-M., & James, M.N.G. (1994). Structural and kinetic characterization of a β-lactamase inhibitor protein. Nature 368, 657-660.

Stubbs, M.T., Oschkinat, H., Mayr, I., Huber, R., Angliker, H., Stone, S.R., & Bode, W. (1992). The interaction of thrombin with fibrinogen: A structural basis for its specificity. Eur. J. Biochem. 206, 187-195.

Sutton, R.B., Daveletov, B.A., Berghuis, A.M., Südhof, T.C. & Sprang, S.R. (1995). Structure of the first C2 domain of synaptotagmin I: A novel Ca^{2+}/phospholipid-binding fold. Cell 80, 929-938.

Tantillo, C., Ding, J., Jacobo-Molina, A., Nanni, R.G., Boyer, P.L., Hughes, S.H., Pauwels, R., Andries, K., Janssen, P.A.J., & Arnold, E. (1994). Locations of anti-AIDS drug binding sites and resistance mutations in the three-dimensional structure of HIV-1 reverse transcriptase. J. Mol. Biol. 243, 369-387.

Thunnissen, M.M.G.M., Elso, A.B., Lakl, K.H., Drenth, J., Dijkstra, B.W., Kulpers, O.P., Dijkman, R., de Haas, G.H., & Verheij, H.M. (1990). X-ray structure of phospholipase A_2 complexed with a substrate-derived inhibitor. Nature 347, 689-691.

Tulip, W.R., Varghese, J.N., Baker, A.T., van Donkelaar, A., Laver, W.G., Webster, R.G., & Colman P.M. (1991). Refined atomic structures of N9 subtype influenza virus neuraminidase and escape mutants. J. Mol. Biol. 221, 487-497.

Varney, M.D., Marzoni, G.P., Palmer, C.L., Deal, J.G., Webber, S., Welsh, K.M., Bacquet, R.J., Bartlett, C.A., Morse, C.A., Booth, C.L.J., Herrmann, S.M., Howland, E.F., Ward, R.W., & White, J. (1992). Crystal structure-based design and synthesis of benz[cd]indole-containing inhibitors of thymidylate synthase. J. Med. Chem. 35, 663-676.

Varghese, J.N., Laver, W.G., & Colman, P.M. (1983). Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. Nature 303, 35-40.

Varghese, J.N., & Colman, P.M. (1991). Three-dimensional structure of the neuraminidase of influenza virus A/Tokyo/3/67 at 2.2 Å resolution. J. Mol. Biol. 221, 473-486.

Varghese, J.N., McKimm-Beschkin, J.L., Caldwell, J.B., Kortt, A.A., & Colman, P.M. (1992). The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor. Proteins 14, 327-332.

Verlinde, C.L.M.J., & Hol, W.G.J. (1994) Structure-based drug design: Progress, results, and challenges. Structure 2, 577-587.

Vidgren, J., Liljas, A., & Walker, N.P.C. (1990). Refined structure of acetazolamide complex of human carbonic anhydrase II at 1.9 Å. Intl. J. Biol. Macromol. 12, 342-344.

Vitali, J., Martin, P.D., Malkowski, M.G., Robertson, W.D., Lazar, J.B., Winant, R.C., Johnson, P.H., & Edwards, B.F. (1992). The structure of a complex of bovine alpha thrombin and recombinant hirudin at 2.8 Å resolution. J. Biol. Chem. 267, 17670-17678.

von Itzstein, M., Wu, W.-Y., Kok, G.B., Pegg, M.S., Dyason, J.C., Jin, B., Phan, T.V., Smythe, M.L., White, H.F., Oliver, S.W., Colman, P.M., Varghese, J.N., Ryan, D.M., Woods, J.M., Bethell, R.C., Hotham, V.J., Cameron, J.M., & Penn, C.R. (1993). Rational design of potent sialidase-based inhibitors of influenza virus replication. Nature 363, 418-423.

Walker, N.P.C., Talanian, R.V., Brady, K.D., Dang, L.C., Bump, N.P., Ferenz, C.R., Franklin, S., Ghayur, T., Hackett, M.C., Hammill, L.D., Herzog, L., Hugunin, M., Houy, W., Mankovich, J.A., McGuinness, L., Orlewicz, E., Paskind, M., Pratt, C.A., Reis, P., Summani, A., Terranova, M., Welch, J.P., Xiong, L., Möller, A., Tracey, D.E., Kamen, R., & Wong, W.W. (1994). Crystal structure of the cysteine protease interleukin-1α-converting enzyme: A (p20/p10)$_2$ homodimer. Cell 78, 343-352.

Walsh, G.M., Reddy, N.S., Bantia, S., Babu, Y.S., & Montgonery, J.A. (1994). Development of inhibitors of purine nucleoside phosphorylase. Hematol. Rev. 8, 87-97.

Wang, J., Yan, Y., Garrett, T.P.J., Liu, J., Rodgers, D.W., Garlick, R.L., Tarr, G.E., Husain, Y., Reinherz, E.L., & Harrison, S.C. (1990). Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains. Nature 348, 411-418.

Webber, S.E., Bleckman, T.M., Attard, J., Deal, J.G., Kathardekar, V., Welsh, K.M., Webber, S., Janson, C.A., Matthews, D.A., Smith, W.W., Freer, S.T., Jordan, S.R., Bacquet, R.J., Howland, E.F., Booth, C.L.J., Ward, R.W., Hermann, S.M., White, J., Morse, C.A., Hilliard, J.A., & Bartlett, C.A. (1993). Design of thymidylate synthase inhibitors using protein crystal structures: The synthesis and biological evaluation of a novel class of 5-substituted quinazolinones. J. Med. Chem. 36, 733-746.

Weber, P.C., Lee, S.-L., Lewandowski, F.A., Schadt, M.C., Chang, C.-H., & Kettner, C.A. (1995). Kinetic and crystallographic studies of thrombin with Ac-(D)Phe-Pro-boroArg-OH and its lysine, amidine, homolysine, and ornithine analogs. Biochemistry 34, 3750-3757.

Wells, J.A., & De Vos, A.M. (1993). Structure and function of human growth hormone: Implications for the hematopoietins. Ann. Rev. Biophys. Biomol. Struct. 22, 329-351.

Wery, J.-P., Schevitz, R.W., Clawson, D.K., Bobbitt, J.L., Dow, E.R., Gamboa, G., Goodson, T., Jr., Hermann, R.B., Kramer, R.M., McClure, D.B., Mihelich, E.D., Putnam, J.E., Sharp, J.D., Stark, D.H., Teater, C., Warrick, M.W., & Jones, N.D. (1991). Structure of recombinant human rheumatoid arthritic synovial fluid phospholipase A$_2$ at 2.2Å resolution. Nature 352, 79-82.

Whittle, P.J., & Blundell, T.L. (1994). Protein structure-based drug design. Ann. Rev. Biophys. Biomol. Struct. 23, 349-375.

Wiley, D.C., & Skehel, J.J. (1987). The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. Ann. Rev. Biochem. 56, 365-394.

Wilkerson, M., Cyrlin, M., Lippa, E.A., Esposito, D., Deasy, D., Panebianco, D., Fazio, R., Yablonski, M., & Shields, B. (1993). Four-week safety and efficacy study of dorzolamide, a novel, active topical carbonic anhydrase inhibitor. Arch. Ophthalmol. 111, 1343-1350.

Williams, D.A., Hsieh, K., DeSilva, A., & Mulligan, R.C. (1987). Protection of bone marrow transplant recipients from lethal doses of methotrexate by the generation of methotrexate-resistant bone marrow. J. Exp. Med. 166, 210-218.

Wlodawer, A., & Erickson, J.W. (1993). Structure-based inhibitors of HIV-1 protease. Ann. Rev. Biochem. 62, 543-585.

Zdanov, A., Wu, S., DiMaio, J., Konishi, Y., Li, Y., Wu, X., Edwards, B.F.P., Martin, P.D., & Cygler, M. (1993). Crystal structure of the complex of human α-thrombin and nonhydrolyzable bifunctional inhibitors, hirutonin-2 and hirutonin-6. Prot. Struct. Funct. Genet. 17, 252-265.

Zhang, G., Kazanietz, M.G., Blumberg, P.M., & Hurley, J.H. (1995) Crystal structure of the Cys2 activator-binding domain of protein kinase Cδ in complex with phorbol ester. Cell 81, 917-924.

Zhang, J., Cousens, L.S., Barr, P.J., & Sprang, S.R. (1991). Three-dimensional structure of human basic fibroblast growth factor, a structural homolog of interleukin 1β. Proc. Natl. Acad. Sci. USA 88, 3446-3450.

Zhang, Y., Heym, B., Allen, B., Young, D., & Cole, S. (1992) The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*. Nature 358, 591-593.

ENGINEERING METAL BINDING SITES IN PROTEINS

Lynne Regan

I. INTRODUCTION AND OVERVIEW

There have been several reports of the *de novo* design of proteins to fold into specified structures (for example Regan and DeGrado 1988; Goraj et al., 1990; Hecht et al., 1990; Quinn et al., 1994). The first designs concentrated upon reproducing simple structural motifs that are found in natural proteins. More recent advances include designs in which ligand binding activities are introduced onto preexisting protein frameworks.

The *de novo* design of specific ligand binding sites is an important route to proteins with novel properties and activities. Recent attention has focused upon the design of binding sites for metal ions (Tainer et al., 1991; Regan, 1993). Why are metal ions an attractive design target? First, metal ions are a simple ligand; the minimal requirement for a binding site is the correct positioning of two, three, or four amino acid side-chains. Second, the roles played by metal ions in natural proteins encompass a wide range, including nucleophilic catalysis, electron transfer, and stabilization of protein structure. (For excellent reviews on natural metalloproteins the reader is referred to the following articles: (Berg, 1990; Vallee and Auld, 1990; Christianson, 1991; Glusker, 1991; Vallee et al., 1991.) The design of a metal-binding site may therefore provide a route to novel proteins with a host of interesting properties. Third, the introduction of a ligand binding site into a new context allows one to attempt to delineate the factors responsible for high-affinity ligand binding. Such studies are complementary to those which start with a natural site and seek to identify the important features by directed mutagenesis. Finally, a practical reason for choosing metal binding site design is that a number of powerful spectroscopic techniques can be employed to report in detail on the number and identity of the metal ligands. The application of these methods allows the success of a design to be assessed, even before high resolution structure determination is undertaken.

The majority of designs concentrate on achieving correct positioning of the two, three, or four residues that will form the metal binding site. This is an apparently simple route to generate proteins which have the potential to display an array of interesting properties. As will become evident, however, achieving functional binding sites is not a trivial accomplishment.

The overall organization of this review is to discuss different metal site designs together in sections, according to the number of protein-to-metal ligands that are

involved. It is also important to note that all the designs that are discussed can also be grouped in two main classes, based on the intent of the design. Under this scheme, the first class includes designs in which metal binding sites have been introduced into proteins to impart specific properties, but which do not necessarily seek to mimic natural sites. The second class includes sites designed with the aim of reproducing the geometry and associated structural and catalytic properties of natural sites.

In natural Zn(II)-proteins, for example, coordination geometry is typically tetrahedral or distorted tetrahedral. Either four ligands to the metal are from protein, or three ligands are from protein, with a molecule of solvent at the fourth position. To successfully mimic such a site a design must not only correctly position three or four protein ligands, but must also exclude undesirable additional ligands from neighboring protein side-chains or solvent. Finally, it is important for both types of designed metal binding sites to avoid energetically unfavorable interactions between the introduced metal binding residues and the rest of the protein. For each example I will discuss the specific aims of the design and the extent to which characterizations of the metal binding properties of the protein have provided support for the successful realization of the design.

II.   METAL BINDING SITES AS DISCRETE UNITS

In the majority of metalloproteins, the residues involved in metal binding come close together in the tertiary structure to form the binding site, but are dispersed along the amino acid sequence. Consequently, the metal site cannot generally be regarded as a discrete, movable unit. In this section two interesting exceptions are described.

A.   The Square Planar Cu(II) Binding Site of Human Albumin

The Cu(II) transport site of human albumin is located on a short linear sequence at the N-terminus of the protein. The sequence is NH_2-Asp-Ala-His... The square-planar Cu(II) site is formed by the alpha-amino nitrogen, the two intervening peptide nitrogens, and the imidazole nitrogen of the histidine residue at the third position. Sarcar and colleagues synthesized a short peptide, NH_2-Gly-Gly-His-N-methylamide, to mimic the site and to test the importance of the identity of the first two residues (Kruck et al., 1976). Their peptide binds metal and displays optical absorption spectra that are identical to those of human albumin with Cu(II) bound (see Figure 1). Equilibrium dialysis measurements using $^{67}Cu(II)$ show that the peptide binds metal with remarkable affinity; the dissociation constant is 2.07 x 10^{-17} M.

The crystal structure of the peptide-Cu(II) complex has been solved and confirms that the peptide binds metal with expected geometry. A single atom of Cu(II)
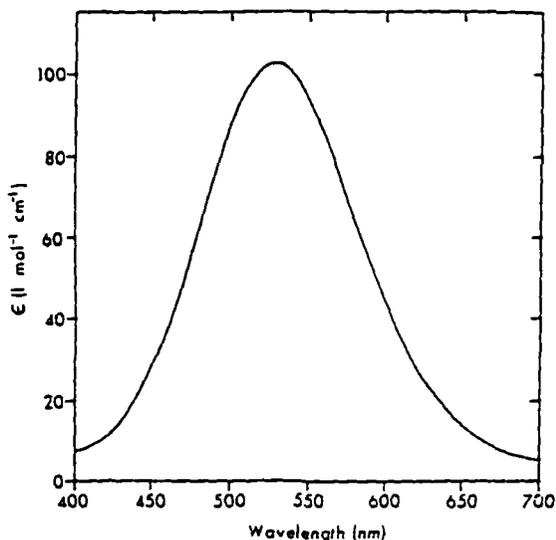
Figure 1. The visible spectrum of the complex of GlyGlyHis-methyl-amide with Cu(II) at pH 8.
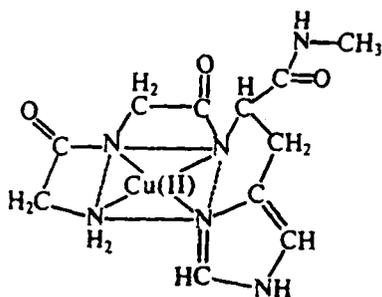
Figure 2. Illustration of the proposed structure of the Cu(II)-GlyGlyHis-methyl amide complex.

is tetradentately chelated by the amino terminal nitrogen, the next two peptide nitrogens, and the histidyl nitrogen of a single peptide in a slightly distorted, square planar arrangement, as shown in Figure 2. As might be anticipated from the behavior of the isolated peptide, the same sequence may be added to the N-terminus of an existing protein as a convenient means by which to introduce Cu(II) binding activity.

B. The Ca^{2+} Binding Loop of Thermolysin

A related example of a discrete, transportable, metal binding site is the Ca(II) binding site of thermolysin. Thermolysin has a surface exposed loop which includes all the residues that are involved in Ca(II) binding. The loop is called an omega loop, with the name intended to describe its "pinched in" shape (Leszcznski and Rose, 1986). To determine if the loop can indeed be considered a discrete functional unit, it was substituted for an existing surface loop in neutral protease, by Toma and colleagues (Toma et al., 1991). The transplant was successful and the loop retains Ca(II) binding activity in its new context.

III. BINDING SITES FORMED BY TWO HISTIDINE RESIDUES

The simplest metal binding sites which have been incorporated into proteins are those where only two of the ligands are provided by the protein; the remaining co-ordination sites are filled by exogenous ligands. In this case the goal has not been to mimic natural metalloproteins, but rather the metal binding sites have been introduced with a view to applications in protein purification, stabilization, and the control of enzymatic activity. The designs realized to date are extremely simple with two histidine residues placed in appropriate positions, for example in an α helix (HisX$_3$His) or a β sheet (HisXHis) (Arnold and Haymore, 1991).

A. Affinity Purification by the Introduction of Two Histidine Sites

The metal binding affinity of the two-histidine sites varies between 2×10^{-4} M and 2×10^{-6} M, depending on the exact details of the site. The factors which are thought to increase the binding affinity include a significant site rigidity and high solvent accessibility. Energetically unfavorable steric interactions between the liganding residues, or the metal itself, and the rest of the protein will result in lower affinity sites (Arnold and Haymore, 1991).

For proteins with sites that bind metal sufficiently tightly, metal affinity columns can be used as a single step purification of the engineered protein from a crude lysate. Iminodiacetate (IDA) attached to a solid support is commonly used, with Cu(II) as the metal. The IDA provides three ligands to the metal ion and the two histidine residues from the protein and solvent molecules complete the coordination shell.

In fact, for protein purification it is not even necessary to have a "designed" site. A string of six histidine residues attached to the N-or C-terminus of a protein is often sufficient to allow the protein to be purified from crude extracts on metal affinity columns. In a commercially available system (The QIAexpressionist (2nd Ed.), QIAGEN Inc., Chatsworth, California), Ni(II)-NDA (nitrilotriacetic acid) columns are chosen in preference to IDA as they provide four matrix ligands to metal, resulting in less leaking of metal from the support (see Figure 3).
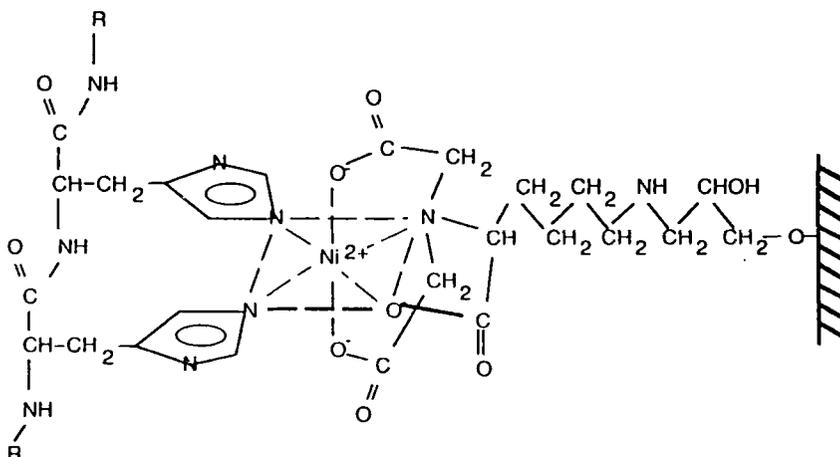
Figure 3. Schematic illustration of the binding of a His containing protein to a Ni(II)NTA resin.

B.   Protein and Peptide Stabilization by Two Histidine Sites

Metal binding and protein denaturation are coupled chemical equilibria. As a consequence, metal binding at the two histidine sites stabilizes the protein towards denaturation, with the tightest binding sites showing the greatest effect on protein stability (Kellis et al., 1991; Todd et al., 1991). A possible practical application of the introduction of such sites is to create more resilient proteins for use in industrial or medicinal applications.

Histidine and cysteine residues appropriately positioned as the i and i+4 residues on an α helix can also stabilize helix formation in short peptides. Ghadiri and colleagues (Ghadiri and Choi, 1990; Ghadiri et al., 1992) have synthesized peptides with the sequences

> Ac-Ala-Glu-Ala-Ala-Ala-Lys-Glu-Ala-Ala-Ala-Lys-X_1-Ala-Ala-Ala-
> X_2-Ala-CONH$_2$
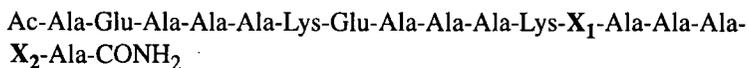
where either X_1 is cysteine and X_2 is histidine, or both X_1 and X_2 are histidine. In both examples, the addition of millimolar Cu(II), Zn(II), or Cd(II) results in an almost two-fold increase in helicity of the peptide in aqueous solution.

C.   A Two Histidine Design To Regulate Enzymatic Activity

An inventive example of the use of a two histidine metal binding site may be found in the work of Craik and colleagues, who used metal binding as a means to

modulate the activity of trypsin (Higaki et al., 1990). A histidine residue was introduced into trypsin such that it could form a two ligand metal binding site, in conjunction with the histidine residue of the active site catalytic triad. A key feature of the design is that the side-chain of the catalytic triad histidine must rotate out of its position in the active site in order to bind a metal ion. It follows that the addition of metal ions to the solution should inhibit the enzyme. The expected movement of the histidine residue is illustrated in Figure 4.

The properties of the mutant enzyme are as follows: even in the absence of metal ions, the enzyme is somewhat less active than wild-type trypsin. The reason for this is not known. It may be due to steric or electrostatic perturbations in the vicinity of the active site resulting from the Arg to His substitution. In the presence of Cu(II) or Zn(II), however, the activity of the modified enzyme is dramatically inhibited. The inhibition is non-competitive with respect to substrate, with a Ki for Cu(II) of 2.1×10^{-5} M and for Zn(II) of 1.3×10^{-4} M. These affinities are rather poor and originally suggested either that binding induces deformation of the protein's structure or that the metal may not be binding exactly as designed. More recently the X-ray crystal structure of the modified trypsin in the presence of bound Cu(II) was solved and reveals the molecular details of the inhibition (McGrath et al., 1993). As specified in the original design, the active site His57 rotates out of position to chelate Cu(II) along with the newly introduced histidine. Although not anticipated in the original design, the Cu(II) ion is additionally coordinated in a bidentate

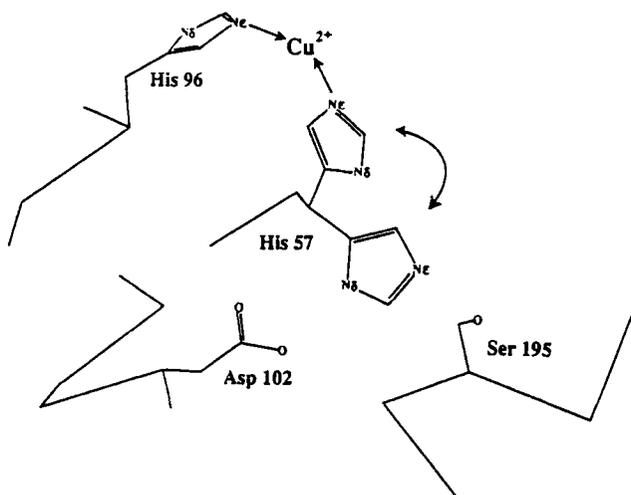*Figure 4.* A schematic illustration of the proposed metal-induced switch in trypsinR96H. The C_α positions of the active site region are shown along with the modeled coordination complex formed between a single Cu(II) ion, His96 and His57 of trypsinR96H.

fashion by a molecule of Tris (present as a Cu(II) carrier in the crystallizations) t
give a distorted square planar coordination geometry.

IV.   DESIGNED SITES WITH THREE PROTEIN LIGANDS

Several natural proteins contain three coordinated Zn(II) sites with tetrahedral c
distorted tetrahedral geometry. The metal ions bound at such sites play an essentia
role in the enzymes' catalytic mechanisms, acting as electrophillic catalysts an
stabilizing negative charges. Carboxypeptidase and carbonic anhydrase are tw
interesting examples. In carboxypeptidase, Zn(II)-bound water is activated fc
nucleophilic attack on the carbonyl of the scissile peptide bond (Christiansor
1991). Similarly, in carbonic anhydrase, metal-bound water provides a source o
nucleophilic hydroxyl groups at neutral pH which can attack either the natura
substrate CO_2 or hydrolyze exogenously added esters (Sasaki and Kaiser, 1989).

To date, the goal of the designs for three ligand sites has been to achiev
activation of metal-bound water for nucleophilic attack. With this aim, it is essentia
that the designed sites allow only a single exogenous ligand to bind. Otherwise, i
two or three ligands bind, with trigonal bipyrimide or octahedral geometry, the pK
of each will most likely not be reduced sufficiently to achieve efficient catalysis a
neutral pH (the pKa of water in $Zn(H_2O)_6^{2+}$ is about 10).

Model compounds give hope that such simple sites will display at least som
level of enzymatic activity. The first such compound was the macrocycle shown ii
Figure 5a (Wolley, 1975) which was synthesized to determine whether a simpl
organic chelate of Zn(II) could mimic carbonic anhydrase activity by activating i
molecule of water for nucleophilic attack. The intention was that the compoun
would bind Zn(II) and the metal would in turn bind a molecule of water, a
illustrated in Figure 5b.

The catalytic activity of the Zn(II)-ligand complex was studied using tw
classical carbonic anhydrase assays: the hydration of acetaldehyde and the hydra
tion of CO_2. The complex displayed high activity in the aldehyde hydration assay
which was comparable to that of carbonic anhydrase. In the carbon dioxid
hydration assay, the model compound showed about one fourhundreth the activit
of carbonic anhydrase, which is still remarkably high for such a simple enzym
model. More recent small compound mimetics have proven even more successfu
(Kimura et al., 1990).

The   compound   [Tris[(4,5-dimethyl-2-imidazolyl)methyl]phospheneox
ide]Zn(II) is also a mimetic of carbonic anhydrase and is a weak catalyst in botl
the CO_2 hydration and p-nitrophenyl acetate hydrolysis assays (Armitage et al.
1978). Similarly, model compounds that mimic carboxypeptidase and are capabl
of catalysis have been synthesized (Groves and Olson, 1985). No engineered meta
binding site in a protein has yet been demonstrated to have associated catalyti
activity of any kind.
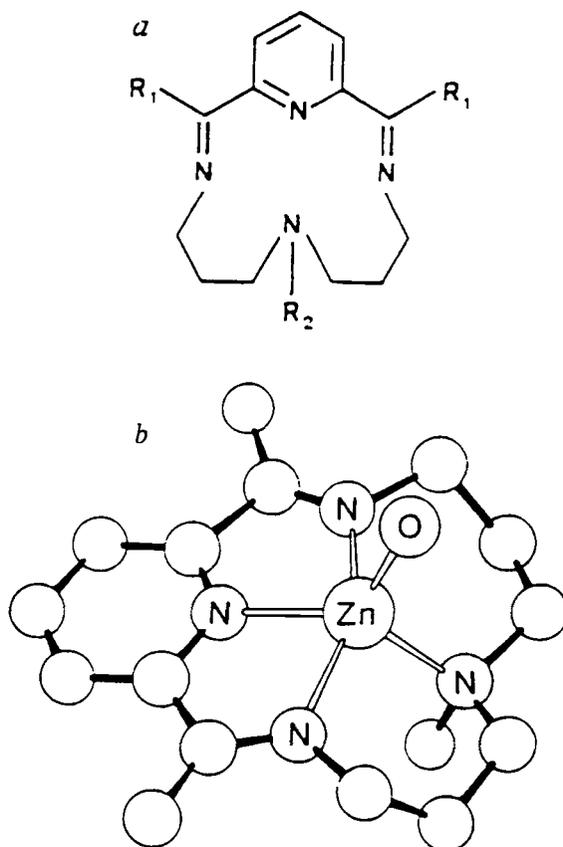
Figure 5. (a) Structure of the free macrocyclic ligand. (b) The complex of the macrocycle with Zn(II). The four nitrogen atoms bind the metal ion which also binds water. The skeleton of the $Zn(N-MeCR)(H_2O)^{2+}$ complex is shown (H omitted and unlabeled atoms are C). $CR:R_1=CH_3$, $R_2=H:N-MeCR;R_1=R_2=CH_3$; $desdiMeCR:R_1=R_2=H$.

A. Metalloantibodies

Tainer and colleagues have reported their attempts to introduce metal binding activity onto an antibody framework (Iverson et al., 1986, Roberts et al., 1990). The construction of such "metalloantibodies" is an attractive strategy because antibodies can potentially be raised against a desired substrate, or transition state analogue, with the Zn(II)-activated water positioned close by for nucleophilic attack. Underlying their design was the recognition that the antiparallel β-structure

Figure 6. Schematic illustration of the main-chain folding of carbonic anhydrase, with the Zn(II) chelating His side-chains shown. Cylinders represent α helices and arrows β sheets.

around the three histidine ligands of carbonic anhydrase B (see Figure 6) is similar to that around two of the complementarity determining regions in antibody light-chains (Christianson, 1991). On the basis of this comparison, three histidine residues were introduced at appro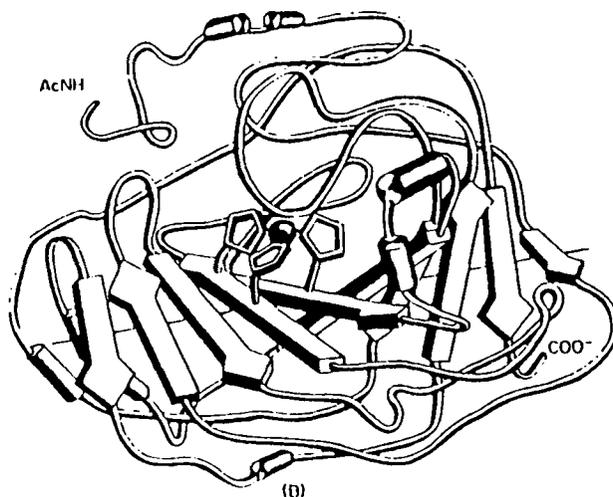priate positions in an antifluorescein antibody. The choice of this antibody allows fluorescence to be used as a convenient means to detect hapten binding in solution.

The interaction of protein with metal was monitored by detecting the quenching of the protein's intrinsic tryptophan fluorescence that occurred when Cu(II) bound. When metal was added to an antibody-fluorescein complex, the fluorescence of both the protein's tryptophan residues, and of the hapten, was quenched, suggesting that it is possible to achieve the ternary complex of antibody, fluorescein, and metal.

Quantitative fluorescence quenching studies were used to determine a binding constant for the Cu(II)-antibody complex of about 10^{-6} M. This value is several orders of magnitude poorer than that of a carbonic anhydrase-Cu(II) complex. It is closer to the values of dissociation constants measured for two histidine binding sites. This could potentially be caused by local structural perturbations or by incomplete histidine coordination.

A practical problem in using an antibody as a framework is the extreme difficulty in obtaining significant quantities of the purified protein. This precludes the use of informative spectroscopic techniques to analyze the nature of the binding site and to determine whether all three histidine residues are involved in binding and

whether the binding site is tetrahedral. As an alternative approach towards answering the first of these questions, site-directed mutagenesis, is being used to mutate each of the proposed liganding residues to alanine. It will then be possible to compare the metal binding properties of this family of mutant proteins and to determine which of the designed ligands are actually involved in chelating metal.

B. A His$_3$ Site in a Designed Four Helix Bundle Protein

A three histidine site has been introduced into a designed protein, α_4. The α_4 protein framework is an idealized four helix bundle in which four identical helices are connected by three identical loops and there is minimal sequence complexity (Regan and DeGrado, 1988). Three surface histidine residues were introduced into this protein, two on one helix at HisX$_3$His (recall the two histidine sites with this arrangement) and the third on a neighboring helix (Handel and DeGrado, 1990). The protein contains no aromatic residues, which facilitates the assignment of the histidine CH2 and CH4 resonances in ^{1}H nuclear magnetic resonance (NMR) spectra. Spectra were recorded as increasing amounts of Zn(II) were added to the protein. The resonances of the CH2 and CH4 histidine ring protons of all three chelating residues were observed to shift until a 1:1 stoichiometry of metal to protein was reached. These observations support the conclusion that all three of the designed ligands are involved in interactions with the metal in a 1:1 protein:Zn(II) complex (see Figure 7). In this protein an additional effect of metal binding was observed; it was associated with a "tightening" of the somewhat molten structure of the protein, as evidenced by an increase in the dispersion of the 1-D ^{1}H NMR spectrum (Handel et al., 1993). Experiments to reveal the geometry of the site and *its affinity for metal have not yet been reported.*

C. A Modified Zinc Finger Peptide

Berg and colleagues reported an interesting example of a three liganded site (Merkle et al., 1991). The approach differed from those discussed above, in that their site was created by the modification of an existing four-coordinate site in a 26 residue zinc finger peptide. The peptide sequence is a consensus sequence for a prototypical TFIIIA-like zinc finger. The natural site is formed by a His and a Cys residue on an α helix and a His and a Cys residue on an adjacent β sheet. The sequence of the peptide is:

ProTyrLysCysProGluCysGlyLysSerPheSerGlnLysSerAspLeuValLys
HisGlnArgThrHisThrGly

The first attempt to create a Cys(His)$_2$ coordination site by the removal of the first five amino acids of the peptide was unsuccessful and the peptide showed no indication of Co(II) binding. By contrast, deletion of the last four amino acids, to
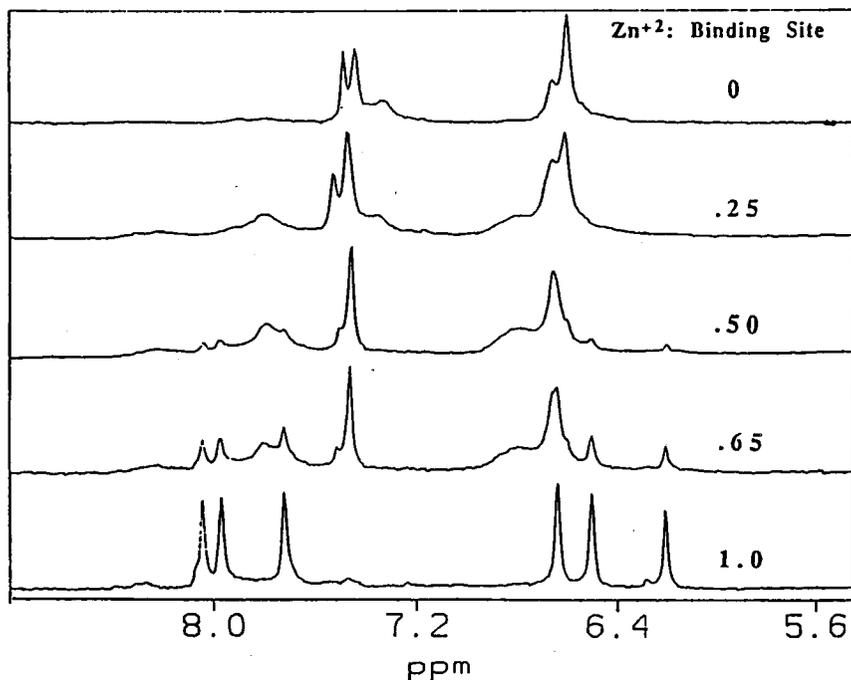
Figure 7. Down Field region of the ^{1}H NMR spectra of the designed Zn(II)-binding protein as a function of Zn(II) concentration. The peptide was in ^{2}H$_2$O with 50 mM NaOAc and was titrated with zinc acetate.

create a (Cys)$_2$His site, resulted in a peptide that bound cobalt with an affinity that was within an order of magnitude of that of the parent peptide.

The optical absorption spectrum of the protein-Co(II) complex shows a long wavelength absorption envelope with an intensity characteristic of tetrahedral coordination, ruling out higher coordination numbers (Bertini and Luchinat, 1984) (see Figure 8a). This indicates that the new site successfully maintains the tetrahedral geometry of the original site, presumably with a molecule of water now bound as the fourth ligand. Moreover, titration of this complex with β-mercaptoethanol resulted in a series of spectra with the final spectrum displaying wavelength maxima that are indicative of a three thiolate one imidazole coordination sphere (see Figure 8b). This final spectrum represents the conditions in which the thiolate of β-mercaptoethanol is coordinating Co(II) in the vacant ligand position.

Disappointingly, attempts to demonstrate carbonic anhydrase-like catalysis by peptide-Zn(II) complexes, in acetaldehyde hydration and p-nitrophenyl acetate hydrolysis assays, were unsuccessful. One reason for this may be that the pKa of
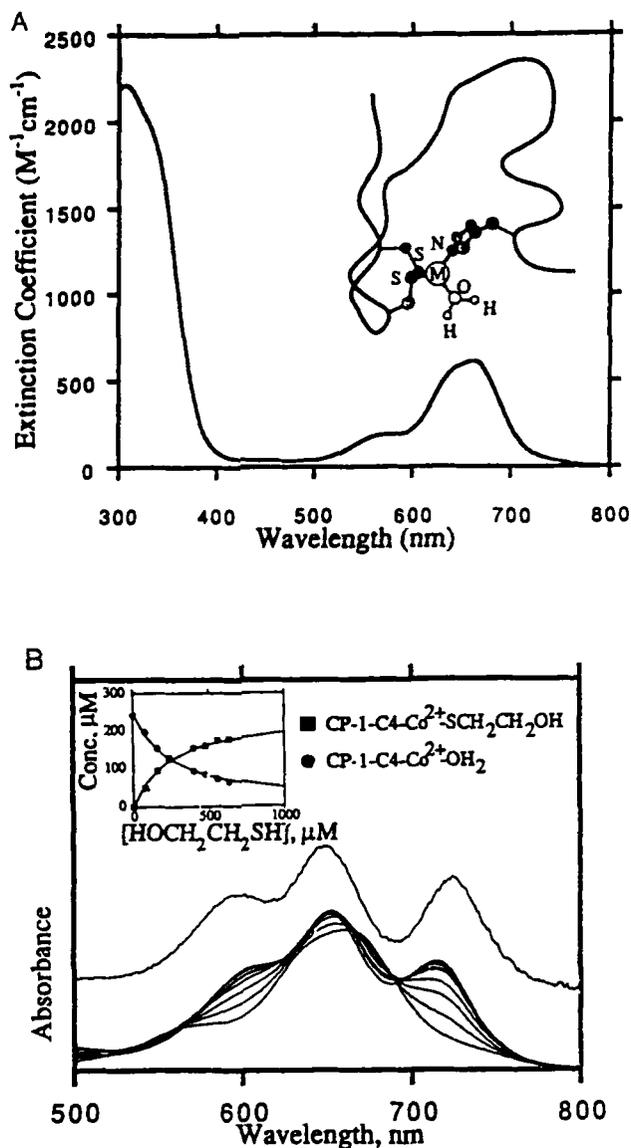
Figure 8. (A) Absorption spectrum of the 1:1 complex of the modified zinc-finger peptide with Co(II) bound. The inset shows a schematic of the proposed structure. (B) Titration of the peptide-Co(II) complex with b-mercaptoethanol. The offset is a derivative of the peptide with $(Cys)_3His$ ligands. This spectrum is similar to that at the end of the titration with mercaptoethanol. The inset shows the experimental data from spectral deconvolution and the fit using a β-mercaptoethanol dissociation constant of 2×10^{-4} M.

63

the metal-bound water at this two-thiolate-one-imidazole site is greater than nine, which compares unfavorably with the pKa of seven in the three imidazole site of carbonic anhydrase.

D.   The Zn(II) Binding "Minibody"

A three histidine has been introduced into a protein whose design is based on the fold of an immunoglobulin. The "minibody" is a trimmed down version of an antibody which uses a portion of the heavy-chain variable domain as a template to create a protein with a novel β-sheet scaffold and two loop connections corresponding to the hypervariable loops H1 and H2. A metal binding site was designed between one histidine in loop H1 and two histidines in loop H2 (Pessi et al., 1990) (Figure 9). Circular dichroism (CD) spectra of the minibody confirmed the presence

Figure 9.   Ribbon representation of the minibody structural model. The side-chains of the metal chelating His residues are shown and the position of the metal is indicated by a sphere. N- and C- termini are indicated.

of β-structure and an interaction between the minibody and metal ions was demonstrated by co-elution of radioactive ^{65}Zn(II) and protein on a size exclusion column. Unfortunately, the extremely low solubility of the minibody precluded more detailed characterizations. Subsequent redesigns have focused upon increasing the minibody's solubility with a view to more detailed structural studies using NMR (Bianchi et al., 1994).

E.   A Three-Coordinate Zn(II) Site in Retinol Binding Protein

One of the highest affinity three histidine Zn(II) binding site yet created is in the β-barrel protein mammalian serum retinol binding protein (RBP) (Muller and Skerra, 1994). Two different three His sites (A and B in Figure 10) were designed on the surface of this protein. Both involve positioning of three histidine residues, two as His-X-His in a β-strand and the third from an adjacent β-strand in an arrangement similar to that of the metal binding residues in carbonic anhydrase. Equilibrium dialysis using radioactive ^{65}Zn(II) was used to measure the affinity of each site. Sites A and B bound Zn(II) with dissociation constants of 3.6 x 10^{-8} M and 4.4 x 10^{-7} M, respectively, and metal binding stabilized the protein against chemical denaturation. Interestingly, the sites do not bind metal with tetrahedral coordination geometry, but rather two or more ligands in addition to the chelating His residues can be accommodated. The non-tetrahedral coordination of Zn(II) is in this example an asset because the sites were introduced to allow the proteins to bind to immobilized metal-ion columns for purification purposes.

F.   Three-Coordinate Zn(II) Sites in a Designed Four Helix Bundle Protein

Zα_4, a four helix bundle protein with a designed tetrahedral (Cys)$_2$(His)$_2$ Zn(II) binding site is described in detail in the following section (Regan and Clarke, 1990). This protein was used as the starting point for the design of three coordinate Zn(II) binding derivatives. Each metal ligand was independently mutated to alanine, the small non-reactive side-chain being chosen to enable a molecule of water to bind at the vacant coordination position.

Three of the four changes produced proteins which behave as anticipated. They bind Zn(II) with affinities that are reduced relative to Zα_4 by up to an order of magnitude. The most important observation is that the optical absorption spectra of the protein-cobalt complexes suggest that one or two solvent molecules are bound in the vacant coordination position. Specifically, the spectra display absorption envelopes at long wavelengths in which the wavelength maxima are shifted as expected for sulfur to oxygen or nitrogen to oxygen substitutions of a single coordinating ligand. The proteins are currently being tested for carbonic anhydrase-like activities, although as was observed with the modified Zn(II) finger design, it is possible that the pKa of the metal-bound water may be too high (Klemba and Regan, unpublished observations).
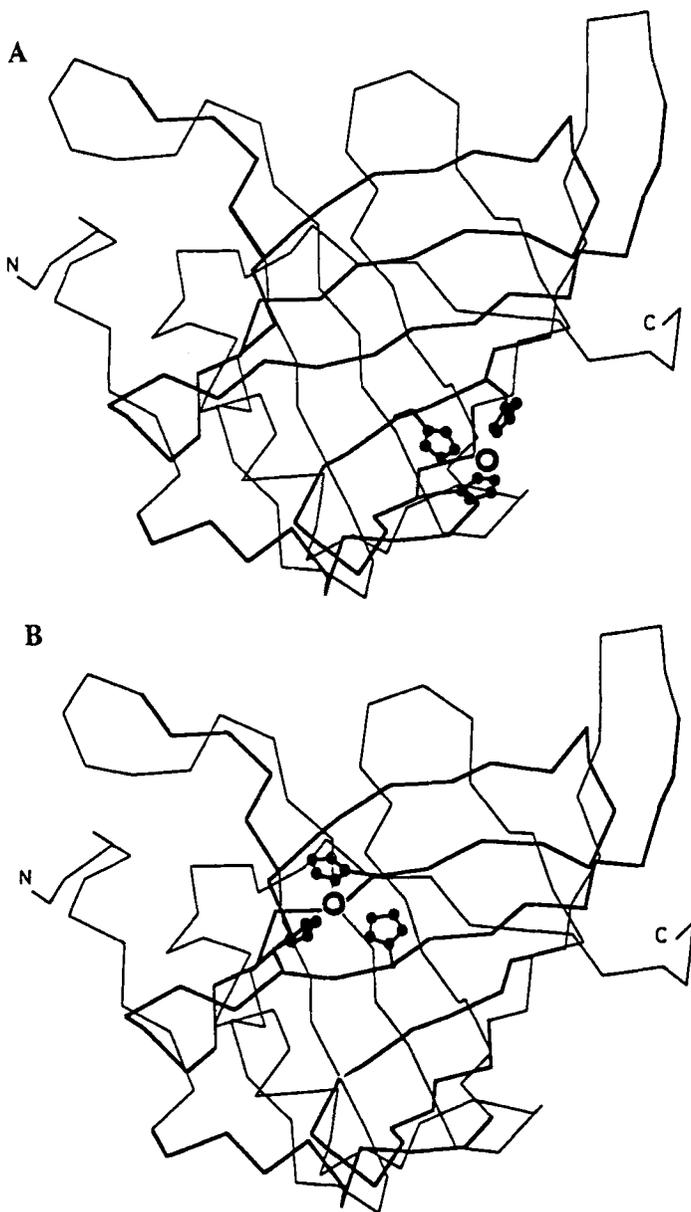
Figure 10. Cα drawing of RBP residues 1-174 with the side-chains of the metal chelating His residues and the metal ion position of site **A** and **B** shown.

V. DESIGNED SITES WITH FOUR PROTEIN LIGANDS

A. A Tetrahedral Binding Site in a Designed Four Helix Bundle Protein

A tetrahedral Zn(II) binding site was introduced into the designed four helix bundle protein, α_4, to create the protein $Z\alpha_4$ (Regan and Clarke, 1990) (Figure 11). The design was the result of a computer search to identify all potential positions in the protein where such a site could be introduced. The program takes as input the coordinates for the model of α_4. For each residue the program generates a list of positions that could be occupied by a Zn(II) ion, if that residue were Cys or His. The possible positions of a metal ion are dictated by the ligand involved (S γ of Cys; N δ or N ϵ N of His) and the appropriate dihedral angles for the side chain.

The program uses subroutines from the PDB_PROTEUS library (Pabo and Suchaneck, 1986) to generate the initial list of possible positions. The sets of possible metal positions for each residue in the protein are compared to one another, and sets of four residues with potential metal sites in fairly close proximity are analyzed further. At each site the program iteratively attempts to find a consensus position for a metal, while allowing a more complete range of chi angles than was allowed in the initial screen. Finally, the site is assessed for tetrahedrality. The best site in α_4 was chosen and the geometry of the model for it improved by performing energy minimization and simulated annealing with the restraints of fixed main-chain atoms and the metal site side-chains. The final model differed little from the starting model, unfavorable steric contacts were absent and the geometry of the site was good.

Optical absorption spectroscopy was used to obtain detailed information on the geometry of the binding site. Figure 12 shows the absorption spectrum of the Co(II) substituted protein. The important features of the spectrum are the absorption envelope at long wavelength and the charge-transfer band at 300 nm. The intensity of the absorption maximum at long wavelength is fully consistent with tetrahedral coordination of the Co(II) and rules out higher coordination numbers. The charge-transfer band, with maximum at 300 nm, is clearly indicative of cysteinate coordination. In addition, the wavelength maxima of the long wavelength transitions is consistent with Co(II) coordination in a $(thiolate)_2(imidazole)_2$ environment.

Optical absorption spectroscopy was also used to directly measure the protein's affinity for Co(II) and to measure its affinity for Zn(II) by competition. The dissociation constant for Co(II) is 1.6×10^{-5} M and for Zn(II) 2.5×10^{-8} M. The lower affinity for Co(II) is consistent with binding at a tetrahedral site. Co(II) is expected to bind less tightly, primarily due to the change from octahedral $(Co(H_2O)_6)^{2+}$ to tetrahedral coordination upon binding which will be accompanied by a loss of ligand field stabilization energy for Co(II) but not for Zn(II).

As a final precaution, gel filtration chromatography with radioactive Zn^{65} was used to demonstrate that the protein binds Zn(II) as a monomer. It was important to confirm this point as there are precedents for metal-induced dimerization in insulin (His ligands) (Blundell et al., 1972); Tat, the transactivating protein of HIV
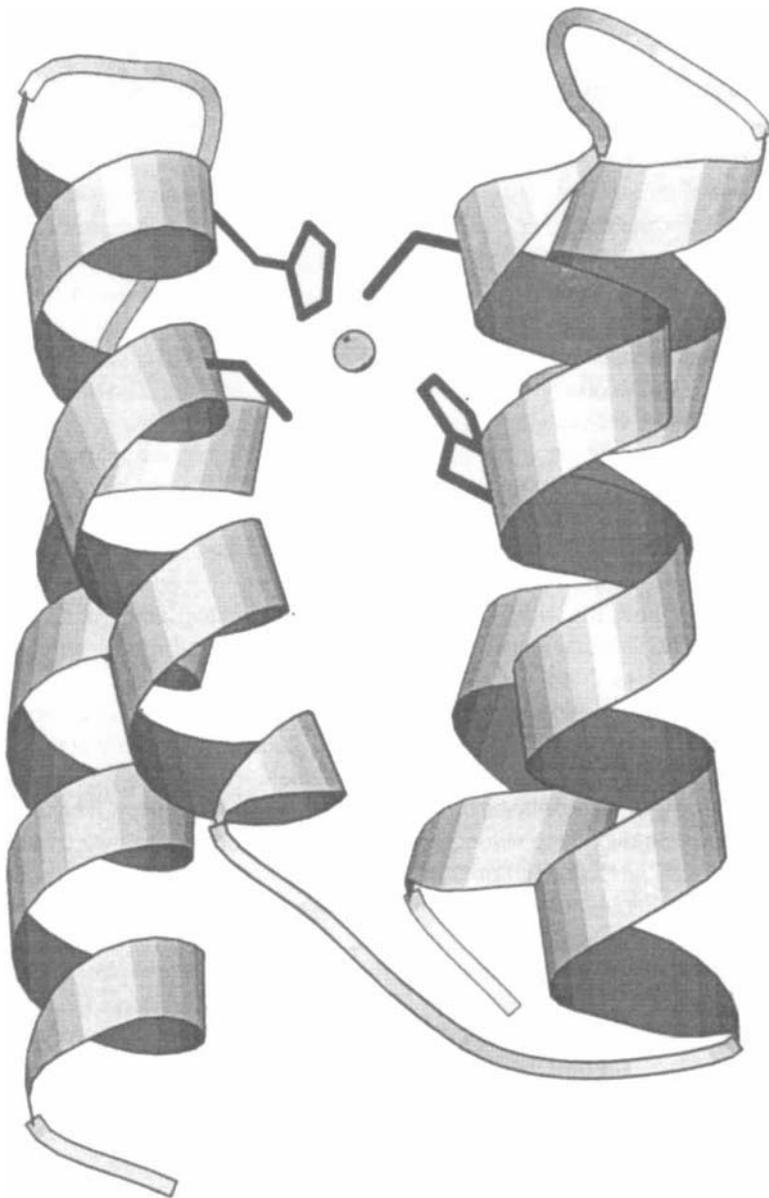
Figure 11. Ribbon representation of the model for Zα₄ with the metal chelating side-chains shown.
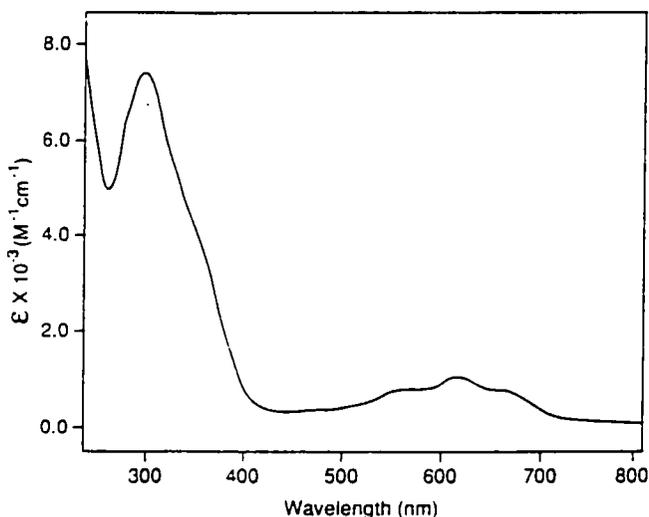
Figure 12. Optical absorption spectrum of the $Z\alpha_4$-Co(II) complex.

(Cys ligands) (Frankel et al., 1988), and human growth hormone (His and Glu ligands) (Cunningham et al., 1991). The results of the characterizations suggest that the design was successful and that protein binds Zn(II) in a 1:1 complex, with high affinity, at a site with tetrahedral coordination geometry.

B. A Tetrahedral Binding Site in a Structurally Well-Characterized Small Protein

Although the design discussed in the preceding section successfully achieved tetrahedral metal binding, a high resolution structure of the model protein onto which the metal binding site was introduced has not yet been determined. Therefore, with a view to the possibility of more detailed structural characterizations of a designed metal binding proteins, a similar strategy was used to introduce a metal binding site into a protein of known structure: the B1 domain of IgG-binding protein G.

The B1 domain of IgG-binding protein G is a 56 residue domain which is composed of four strands of β-sheet crossed by a single α-helix (Gronenborn et al., 1991). This protein has a number of properties that are ideal for metal-site design: it is stable, structurally well characterized, easy to produce in large quantities, and free of pre-existing cysteine or histidine residues that may act as potential metal ligands.

To design a metal binding site, the Metal_Search program was used to identify all possible positions in the protein at which a Zn(II) ion could be bound with
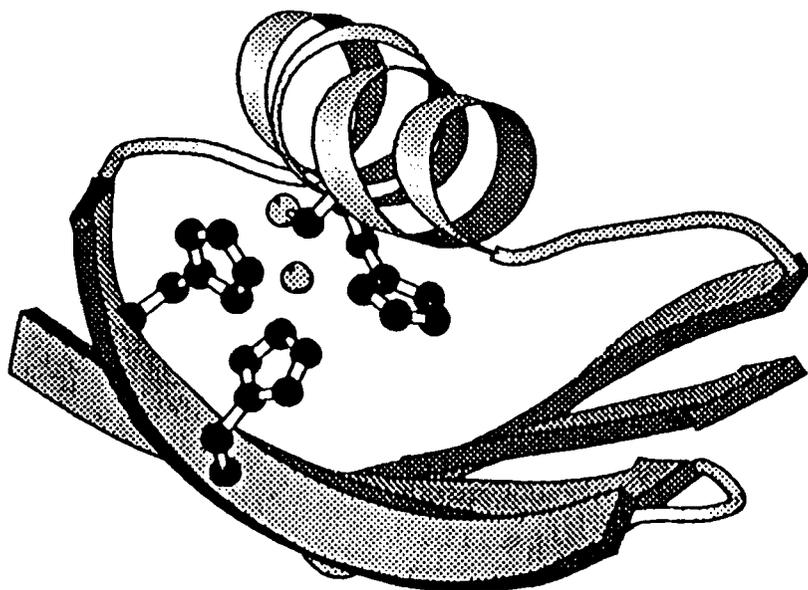
Figure 13. Ribbon representation of Zβ1M, with the metal chelating side-chains shown.

tetrahedral geometry. The program used the NMR-derived coordinates of the wild-type B1 domain as input and produced a list of 10 geometrically reasonable sites, each of which was evaluated further. A major concern was that introduction of the metal binding residues might perturb protein structure. Therefore all sites in which more than one buried residue would be replaced with a metal ligand were rejected, and this screen reduced the list from 10 possible sites to one. The metal-binding site is formed by introducing two His residues on a β-strand (His-X-His) and one His and one Cys residue on the α-helix (His-X_2-Cys) (Figure 13). Molecular modeling identified an unfavorable interaction between the side-chain of His30 and the side-chain of Leu5 (on β strand 1). Two additional variants were therefore considered in which Leu5 was replaced by either alanine or methionine. The designed metal-binding proteins are designated Zβ1L, Zβ1M, and Zβ1A to indicate the identity of the residue at position 5.

Optical absorption spectra of all three variants substituted with Co(II) showed an envelope of d-d transitions at long wavelength, with an extinction co-efficient
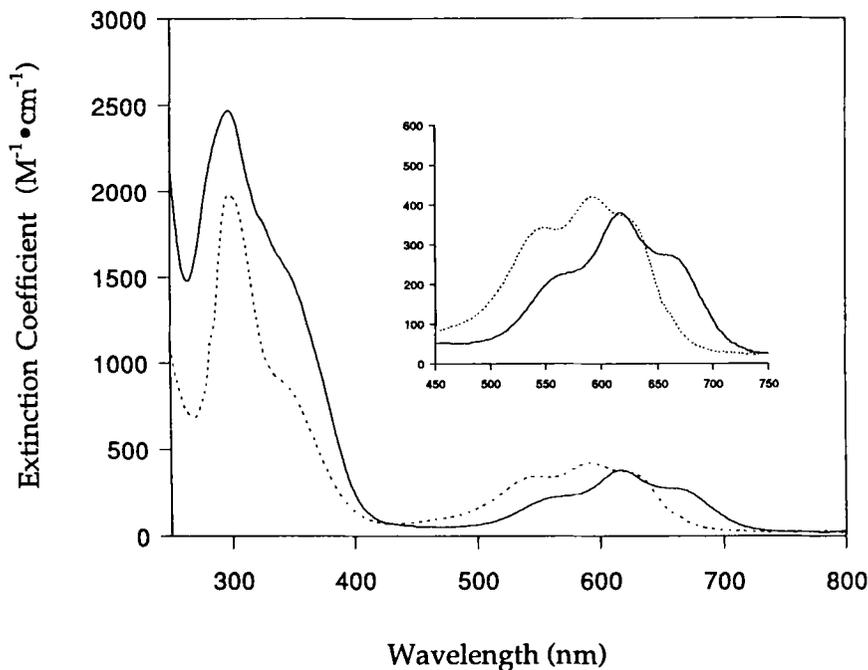
Figure 14. Optical absorption spectra of Co(II)-substituted $Z\alpha_4$ (solid line) and $Z\beta1L$ (dashed line). The spectra of $Z\beta1A$ and $Z\beta1M$ are essentially the same as that of $Z\beta1L$ and have been omitted for clarity. The inset shows an enlargement of the long wavelength region of the spectra.

at the absorption maximum indicative of tetrahedral binding and a wavelength maximum consistent with coordination by nitrogen ligands and one sulfur ligand (Figure 14). The sensitivity of such spectra to the identity of the coordinating ligands is illustrated by the inset to Figure 14, which compares the long wavelength spectra of Co(II) substituted $Z\alpha_4$ (His_2Cys_2) ligands, with that of $Z\beta1L$ (His_2Cys_2) ligands.

All three $Z\beta1$ variants bind metals tightly, with dissociation constants for Zn(II) of about 10^{-9} M. The structural consequences of the identity of residue 5 were, however, dramatic. In the absence of metal $Z\beta1A$ is completely unfolded, $Z\beta1M$ is partially folded and $Z\beta1L$ is fully folded. Upon metal binding, all three variants adopt a folded structure which CD and NMR studies suggest is closely similar to that of the wild-type B1 domain. The severe destabilizing effects of the alanine and methionine substitutions were unexpected, but it appears that metal binding is sufficiently energetically favorable to compensate. $Z\beta1M$ binds metal ions the most

tightly of all three variants because it represents the optimal combination of removal of the steric clash, while not being so severely compromised in stability as the ZβlA variant.

C.   A Design for a Blue Copper Site

A similar approach to the design of a four coordinate metal site was taken by Hellinga and Richards (Hellinga and Richards, 1991; Hellinga et al., 1991). Here, the target is a "type I" copper site, as found in blue copper proteins including azurin and plastacyanin. In these sites copper is completely inaccessible to solvent and is chelated by S γ of Cys, two N δ of His and the S δ of methionine. The four coordinating atoms are arranged in a distorted tetrahedron. (see Figure 15). As their name suggests, blue copper binding sites have a number of characteristic spectroscopic properties which reflect the detailed geometry of the copper coordination sphere.

Thioredoxin from *Escherichia. coli*, a protein whose X-ray crystal structure is known, was chosen as the framework on which to introduce the new site. A computer program was used which takes the protein coordinates, keeping the main-chain atoms in their original positions, and substitutes each side-chain with each of the four potential ligands in positions dictated by preferred rotamer dihedral libraries (Ponder and Richards, 1988). The program then searches for a set of backbone positions that are arranged such that if appropriate side chains were attached they would form a Cu(II) binding site. The new model is subjected to energy minimization and checked for both steric clashes, which would result from
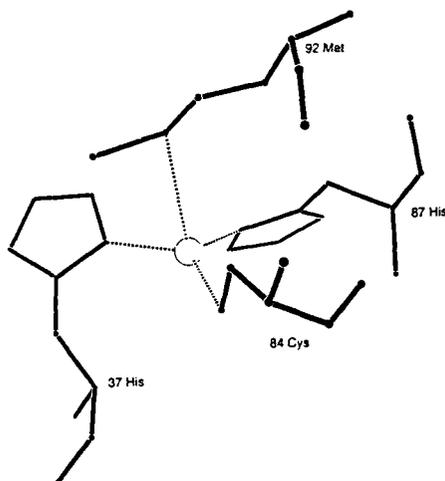

Figure 15.   The natural copper-binding site in plastocyanin, from the Brookhaven PDE entry 1PCY YY.

the protein now being "overpacked" and solvent accessibility changes (which are taken as an indication that the protein is now "underpacked"). If either problem exists, the protein is repacked around the site using a combinatorial repacking algorithm (Ponder and Richards, 1988). The best site is chosen by considering the goodness of the geometry, by minimizing the number of additional mutations that must be made and by checking the integrity of the protein structure. The first site required only four amino acid changes at the binding site. A complication was encountered when characterization of the protein was initiated: wild-type thioredoxin displays metal binding activity and binds one atom of copper or two of mercury. The binding sites are on the surface of the protein and were first observed when heavy atom derivatives were obtained in crystallographic studies on the protein.

Changes in the protein's intrinsic tryptophan fluorescence induced by metal binding were used to determine the affinity and stoichiometry of the interaction. The Scatchard plots of this data show classical two site binding and indicate that an additional binding site for metal has been introduced that binds Cu(II) with an affinity of 4×10^{-5} M compared with the surface site which binds Cu(II) with a dissociation constant of 4×10^{-6} M. (see Figure 16). Optical absorption spectroscopy of protein-Cu(II) complexes was used to determine the coordination of the new metal binding site, blue copper sites having a characteristic intense band at 600 mn corresponding to charge transfer interactions between copper and the
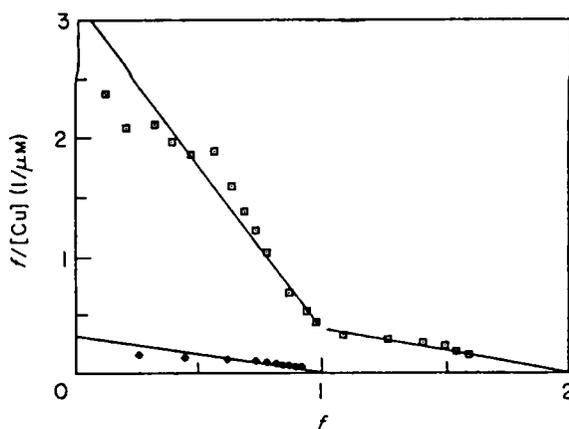
Figure 16. Scatchard analysis of Cu(II)-binding. f is the mole fraction of bound metal $f = n\Delta F/\Delta F_{max}$, (where ΔF is the change in fluorescence at a given metal concentration and ΔF_{max} is the change in fluorescence at infinite concentration of metal), n is the number of metal binding sites, and [Cu(II)] is the concentration of Cu(II). The filled diamonds are data for wild-type thioredoxin, the open squares are data for thioredoxin with the additional designed site.
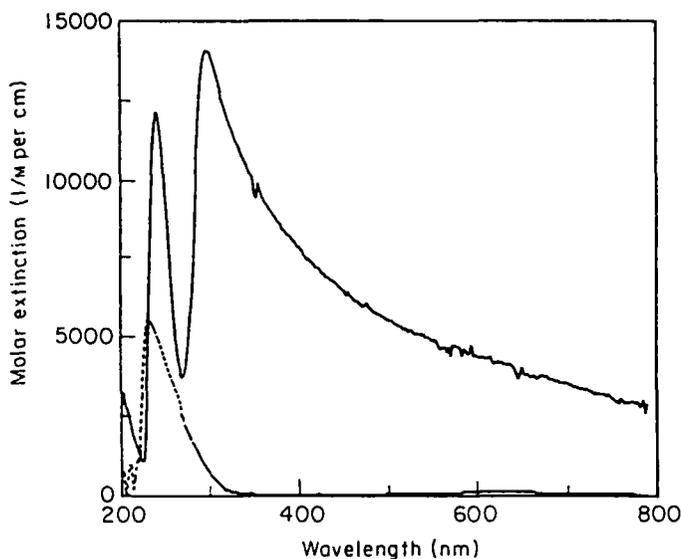
Figure 17. Optical absorption spectra of the thioredoxin-Cu(II) complex (broken line) and the designed protein-Cu(II) complex (solid line).

cysteine thiolate and methionine thioether. The absorption spectrum of the protein-Cu(II) complex does not have this band. It shows an unusual broad absorbance over the entire visible region. (See Figure 17). This indicates that either the copper-thiolate and copper-thioether bonds are both missing or that there are a number of different forms of the protein-metal complex, with the broad spectrum representing transitions with an overlapping range of energies. The spectrum does show strong bands at 240 and 300 nm that are characteristic of Cu-imidazole interactions. Electron Paramagnetic Resonance (EPR) spectra of the protein-Cu(III) complexes suggest that the copper may be coordinated by two nitrogen and two oxygen ligands. The oxygen ligands may come from nearby acidic groups or from solvent.

In an attempt to make the designed ligands the only potential chelators in the vicinity of the site, several neighboring acidic side-chains have been removed. Absorption spectra of protein-cobalt complexes of this improved version of the design show charge transfer bands in the near UV that are indicative of thiol coordination. Therefore, it appears that the metal is now coordinated by at least three of the four designed ligands. Further characterizations and redesigns are in progress towards the aim of achieving a true type-I site (Hellinga, personal communication).

D. A Simplified $(Cys)_2(His)_2$ Zinc Finger Peptide

The following discussion the modification of a natural zinc-finger peptide is included as a caution for protein design: its behavior is quite unexpected. A 26 residue peptide, MZF (minimalist zinc finger), based on the TFIIIA metal binding sequence, was synthesized (Michael et al., 1992). It contains the conserved metal binding and hydrophobic residues that are characteristic of this class of protein domain, with 16 of the remaining 19 residues being changed to alanine:

Lys-**Tyr**-Ala-<u>Cys</u>-Ala-Ala-<u>Cys</u>-Ala-Ala-Ala-**Phe**-Ala-Ala-Lys-Ala-Ala-**Leu**-Ala-Ala-<u>His</u>-Ala-Ala-Ala-<u>His</u>-Ala-Lys

The unusual properties of this peptide were first revealed when the optical absorption spectrum was recorded during the titration of the protein with cobalt. The shape of the spectrum was observed to change during the course of the titration. A spectrum in which the ligand-field bands were relatively red-shifted was observed early in the titration. As increasing amounts of cobalt were added, the initial spectrum decayed and a new spectrum appeared and increased until saturation.

These unusual observations can be explained as follows. The spectrum at low peptide concentrations resembles the spectrum of Co(II) complexed by four cysteinate ligands. It is therefore likely that early in the titration, when the ratio of peptide:Co is high, a dimer is formed in which cobalt is coordinated by two cysteine ligands from each of two peptides. Later in the titration, when the peptide:Co ratio is low, the "correct" species is formed, with cobalt coordinated by two histidine and two cysteine residues from a single peptide. Figure 18A shows a superposition of the spectra of the two different species and Figure 18B shows the amounts of each complex as a function of concentration. One-and two-dimensional NMR studies indicate that the peptide 1:1 complex has a structure that is very similar to that of an authentic TFIIIA Zn finger peptide.

Additional mutations were made in MZF to disrupt its hydrophobic core by substituting the conserved hydrophobic amino acids (shown in bold) with alanine. These peptides form the 2:1 species to a much greater extent than the MZF peptide. Also, the absorption spectra of their 1:1 complexes with cobalt are much broader than that of MZF, which suggests that they may not form a unique structure, but rather an ensemble of structures. These results are remarkable as they demonstrate how little structure is required to generate a peptide that can assemble and bind to a metal ion. The formation of the unexpected 2:1 complexes also demonstrates, as was observed in the design for a blue copper site, that if alternative liganding schemes are possible, the design must incorporate features to destabilize these structures relative to the intended structure.

A

B
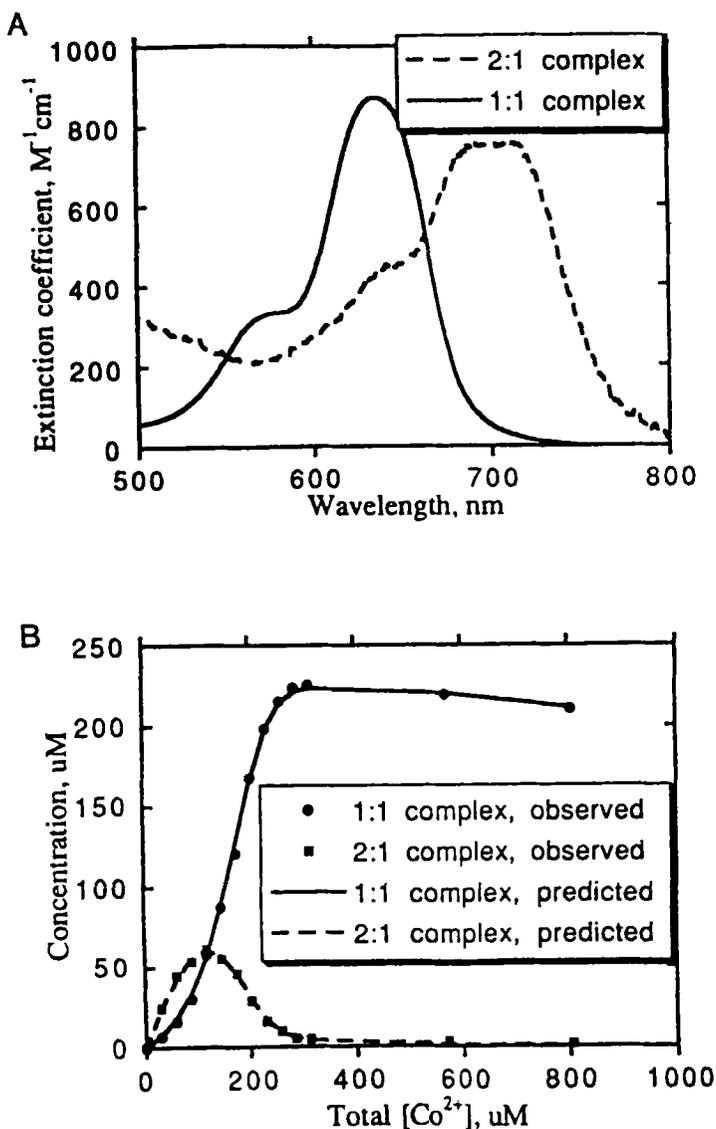
Figure 18. (A) Derived spectra for the 2:1 and 1:1 complexes of the MZF peptide wit Co(II). The spectra were calculated as linear combinations of the initial and final titratio spectra. (**B**) Plot of the observed concentrations of the 1:1 peptide:Co(II) (solid circle and 2:1 peptide:Co(II) (solid square) complexes as a function of total Co(I concentration. The dashed and solid lines represent fits to the data with dissociatio constants of 5×10^{-7} M for the 1:1 species and 4×10^{-11} M^2 for the 2:1 species. The 1: complex concentrations decrease at high Co(II) concentrations due to the increase volume of the peptide sample.

VI. DISCUSSION AND SUMMARY

It is clear that certain of the metal site designs have been extremely successful and that the new proteins display the anticipated properties. A high resolution structure has not yet been determined for any of the designed proteins, so that the exact details await confirmation. The designed sites incorporate the minimum features required to obtain metal-binding. They are based primarily on geometrical considerations, with additional checks to avoid unfavorable steric interactions. By comparison, natural metal sites display additional features, described below, that result in affinities for metal that are higher than those of the designed sites.

In many natural sites the side-chains of the metal binding residues are additionally hydrogen-bonded to other side-chains or to backbone carbonyl and amide groups to position them correctly (Christianson, 1991; Glusker, 1991) (see Figure 19). Some natural metal binding sites have evolved electrostatic surfaces with high negative potential in the region of the binding site. This is an additional feature which will enhance the affinity of the site for positively charged metal ions (Gilson and Honig, 1988).

Finally, Eisenberg and colleagues have noted that metal ions appear to bind in regions of proteins that show "high hydrophobicity contrast" (Yamashita et al., 1990). That is, the metal sites are centered in a shell of hydrophilic ligands that is surrounded by a shell of carbon-containing groups. This could reflect the nature of the metal-ligand interaction: the metal-binding residues chelate through the electron pair donors oxygen, nitrogen,. and sulfur, which themselves are bound to the
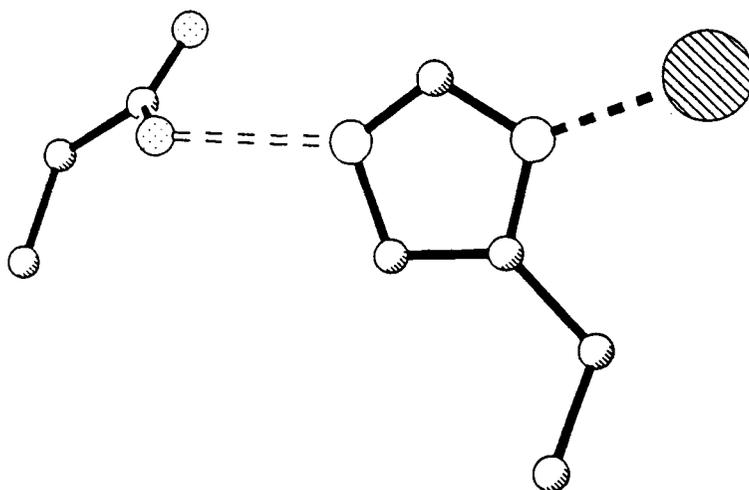

Figure 19. The aspartate-histidine-zinc triad of the active site of carboxypeptidase that may contribute to enhanced Zn(II) affinity at the metalloenzyme active site.

less polar carbon atoms of the side-chain. There must necessarily be a large difference in hydrophobicity at this point. However, the nature of the second shell of residues around the site can serve to enhance that degree of hydrophobic contrast. How could a hydrophobic shell around the site lead to enhanced metal binding affinity? It is possible that the hydrophobic sphere restricts the flexibility of the site "pre-organizing" it for metal binding. Alternatively, the low dielectric of the hydrophobic sphere could enhance electrostatic interactions between groups within it.

A final consideration in the design of high affinity sites is the potential for interaction of the metal with unanticipated additional ligands. These may either be groups on the protein or molecules of solvent. A close inspection and possible modification of potential ligating residues in the vicinity of the designed site would be a step towards addressing this problem. The participation of unwanted solvent ligands is a somewhat more difficult problem to address. One precaution is to ensure that interaction of the protein ligands with the metal does not require unfavorable side-chain dihedral angles, or other energetically unfavorable distortions of the structure.

From the discussions above, it is clear that there are additional features that can be incorporated into future designs to generate sites that more closely reproduce the affinities of natural sites. However, it is most encouraging that there are already designed sites that bind metal. These can be the starting point for a systematic investigation of the additional factors that contribute to high affinity binding. In addition to higher binding activities, it is to be expected that future designs and redesigns will soon achieve catalysis at the metal site.

ACKNOWLEDGMENTS

REFERENCES

Armitage, I.M., Schoot-Uiterkamp, A.J.M., Chlebowski, J.F., & Coleman, J.E. (1978). Characterization of a small molecule carbonic anhydrase mimetic. J. Magn. Reson. 29, 375-383.

Arnold, F.H., & Haymore, B.L. (1991). Engineered metal-binding proteins: Purification to protein folding. Science 252, 1796-1797.

Berg, J. M. (1990). Zinc finger domains: Hypotheses and current knowledge. Ann. Rev. Biophys. Biophysical Chem. 19, 405-427.

Bertini, I., & Luchinat, C. (1984). High spin cobalt (II) as a probe for the investigation of metalloproteins. Adv. Inog. Chem. 6, 71-111.

Bianchi, E., Venturini, S., Pessi, A., Tramontano, A., & Sollazzo, M. (1994). High level expression and rational mutagenesis of a designed protein, the minibody. From an insoluble to a soluble molecule. J. Mol. Biol. 236, 649-659.

Blundell, T., Dodson, G., Hodgkin, D., & Mercola, D. (1972). Insulin: The structure in the crystal and its reflection in chemistry and biology. Adv. Protein Chem. 42, 279-303.

Christianson, D.W. (1991). Structural biology of zinc. Adv. .Prot. Chem. 42, 224-244.

Cunningham, B.C., Mulkernin, M.G., & Wells, J. A. (1991). Dimerization of human growth hormone by zinc. Science 253, 545-548.

Frankel, A. D., Bredt, D. S., & Pabo, C. O. (1988). Tat protein from human immunodeficiency virus forms a metal-linked dimer. Science 240, 70-73.

Ghadiri, M. R., & Choi, C. (1990). Secondary structure nucleation in peptides. Transition metal ion stabilized α-helices. J. Am. Chem. Soc. 112, 1630-1632.

Ghadiri, M.R., Soares, C., & Choi, C. (1992). Design of an artificial four-helix bundle metalloprotein via a novel ruthenium (II)-assisted self-assembly process. J. Am. Chem. Soc. 114, 4000-4002.

Gilson, M.K., & Honig, B. (1988). Calculation of the total electrostatic energy of a macromolecular system: Solvation enery, binding energies and conformational annalysis. Proteins 4, 7-17.

Glusker, J. P. (1991). Structural aspects of metal liganding to functional groups in proteins. Adv. Prot. Chem. 42, 1-27.

Goraj, K., Renard, A., & Martial, J.A. (1990). Synthesis, purification and initial structural characterization of octarellin, a *de novo* polypeptide modelled on the α/β barell proteins. Protein Eng. 3, 259-266.

Gronenbom, A. M. et al., (1991). The immunoglobulin binding domain of *Streptococcal* Protein G has a novel and highly stable polypeptide fold. Science 253, 657-661.

Handel, T.M., & DeGrado, W.F. (1990). *De novo* design of a Zn(II)-binding protein. J. Am. Chem. Soc. 112, 6710-6711.

Handel, T. M., Williams, S. A., & DeGrado, W. F. (1993). Metal ion-dependent modulation of the dynamics of a designed protein. Science 241, 879-885.

Hecht, M.H., Richardson, J.S., Richardson, D.C., & Ogden, R.C. (1990). *De novo* design, expression and characterization of Felix: A four-helix bundle protein of native-like sequence. Science 249, 884-891.

Hellinga, H. W., & Richards, F. M. (1991). Construction of new ligand-binding sites in proteins. I. Computer aided modelling of sites with pre-defined geometry. J. Mol. Biol. 222, 763-786.

Hellinga, H. W., Caragonna, J. P., & Richards, F. M. (1991). Construction of new ligand-binding sites in proteins. II. Grafting of a burried transition metal-binding site into *E. coli* thioredoxin. J. Mol. Biol. 222, 787-793.

Higaki, J., Fletterick, R.J., & Craik, C.S. (1990). Regulation of serine protease activity by an engineered metal switch. Biochemistry 29, 8582-8586.

Iverson, B.L., et al., (1986). Metalloantibodies. Science 249, 659-662.

Kellis, J. T., Jr., Todd, R. J., & Arnold, F. H. (1991). Protein stabilization by engineered metal chelation. Bio/Technology 9, 994-995.

Kimura, et al., (1990). A zinc(II) complex of 1,5,9-triazacyclododecane ([12]aneN$_3$) as a model for carbonic anhydrase. J. Am. Chem. Soc. 112, 5805-5811.

Klemba, M., Gardner, K. H., Marino, S., Clarke, N. D., & Regan, L. (1995). Novel metal-binding proteins by design. Nature Structural Biology. In Press.

Kruck, T.P.A., Lau, S.-J., & Sarkar, B. (1976). Cu(II)-binding by NH$_2$-Gly-Gly-His-N-methylamide. Can. J. Chem. 8, 1300-1315.

Leszcznski, J. F., & Rose, G.D. (1986). Loops in globular proteins, a novel category of secondary structure. Science 234, 849-852.

McGrath, M. E., Haymore, B. L., Summers, N. L., Craik, C. S., & Fletterick, R. J. (1993). Biochemistry structure of an engineered, metal-attenuated switch in trypsin 32, 1914-1919.

Merkle, D. L., Schmidt, M. H., & Berg, J. M. (1991). Design and characterization of a ligand-binding metallopeptide. J. Am. Chem. Soc. 113, 5450-5451.

Michael, S. F., Kilfoil, V.J., Dchmidt, M. H., Amann, B.T., & Berg, J. M. (1992). Metal-binding and folding of a minimalist Cys$_2$His$_2$ zinc finger peptide. Proc. Natl. Acd. Sci. USA 89, 4796-4800.

Muller, H. N., & Skerre, A. (1994). Grafting of a high affinity Zn(II)-binding site on the β-barrel of retinol-binding protein results in enhanced folding stability and enables simplified purification. Biochemistry 33, 14126-14135.

Pabo, C.O., & Suchaneck, E.G. (1986). Computer-aided model building strategies for protein design. Biochemistry 25, 5987-5993.

Pessi, A., Bianchi, E., Crameri, A., Venturini, S., & Tramontano, A. (1990). A designed metal-binding protein with a novel fold. Nature 362, 367-369.

Ponder, J. W., & Richards, F. R. (1988). Tertiary tempates for proteins: Use of packing criteria in the ennumeration of allowed sequences for different structural classes. J. Mol. Biol. 193, 775-791.

Quinn, T. P., et al., (1994). Betadoublet: *De novo* design, synthesis, and characterization of a β-sandwich protein. Proc. Natl. Acad. Sci. USA 91, 8747-8751.

Regan, L. (1993). The design of metal-binding sites in proteins. Ann. Rev. Biophys. Biomol. Struct. 22, 257-281.

Regan, L., & Clarke, N.D. (1990). A tetrahedral Zn(II) binding site introduced into a designed protein. Biochemistry 29, 10879-10883.

Regan, L., & DeGrado, W.F. (1988). Characterization of a helical protein designed from first principles. Science 241, 976-981.

Roberts, V.A., Iverson, B.L., Iverson, S.A., Benkovic, S.J., Lerner, R.A., et al. (1990). Antibody remodeling: A general solution to the design of a metal-coordination site in an antibody binding pocket. Proc. Natl. Acad. Sci. USA 87, 6654-6658.

Sasaki, T., & Kaiser, E.T. (1989). Helichrome: Synthesis and enzymatic activity of a designed hemeprotein. J. Am. Chem. Soc. 111, 380-381.

Tainer, J.A., Roberts, V.A., & Getzoff, E.D. (1991). Designed metalloproteins. Curr. Opin. Biotechnology 2, 582-590

Todd, R. J., Van Dam, M. E., Casimiro, D., Haymore, B. L., & Arnold, F. H. (1991). Cu(II)-binding properties of a cytochrome c with a synthetic metal-binding site: His-X3-His in an α-helix. Proteins: Structure, Function Genetics 10, 156-61.

Toma, S., Campagnoli, S., Margarit, I., Gianna, R.Grandi, G. et al., (1991). Grafting of a calcium-binding loop of thermolysin to *Bacillus subtilis* neutral protease. Biochemistry 30, 97-106.

Vallee, B.L., & Auld, D.S. (1990). Zinc coordination, function and structure in zinc enzymes and other proteins. Biochemistry 29, 5647-5659.

Vallee, B.L., Coleman, J.E., & Auld, D.S. (1991). Zinc fingers, zinc clusters and zinc twists in DNA-binding proteins. Proc. Natl. Acad. Sci. USA 88, 999-1003.

Woolley, P. (1975). Models for metal ion function in carbonic anhydrase. Nature 258, 677-682.

Yamashita, M.M., Wesson, L., Eisenman, G., & Eisenberg, D. (1990). Proc. Natl. Acad. Sci. USA 87:5648.

USING MOLECULAR GRAPHICS TO ANALYZE PROTEIN STRUCTURES

Himanshu Oberoi and Norma M. Allewell

I. INTRODUCTION

Of the many developments in computer technology in the past decades, computer graphics has perhaps had the greatest impact in recent years. Computer graphics is gradually changing the way in which we perceive and interact with the world, in areas as disparate as children's television and protein folding. Computer graphics were first used to visualize molecules in 1964, with the development of the MAC program by Langridge and Levinthal at the National Institutes of Health (Langridge, 1995). With the ever increasing hardware capabilities of computers since that time, real-time interactive modeling of complex macromolecules and their properties has become increasingly accessible. Simultaneous advances in mathematical approaches for visualizing and analyzing molecular structures have facilitated the elucidation of new principles of macromolecular structure. These computational developments have, in turn, generated new experimental questions and stimulated the development of new experimental approaches. With the rapid increase in the number of known structures that has resulted from advances in recombinant DNA technology, X-ray crystallography, and nuclear magnetic resonance (NMR) spectroscopy combined with the development of more powerful methods for visualization and significant advances in experimental approaches, it is possible that most protein structural motifs will soon be known and that methods for predicting the structural motif of a newly discovered protein with certainty will soon be in place. On another front, the explosive growth of the Internet has prompted the development of Web browsers that allow the interactive visualization of three-dimensional molecular structures located anywhere on the World Wide Web (WWW or the Web) (Berners-Lee et al., 1992; http://www.w3.org) and offer the first glimpse into a not-too-distant future where it will be possible to perceive these structures in their own virtual world.

The development of molecular graphics software and hardware has largely been driven by X-ray crystallography and, more recently, by NMR spectroscopy. The first three-dimensional structures of protein that were solved by X-ray crystal-

lography some four decades ago were modeled with balsa wood models at low resolution and wire models constructed to scale at high resolution. The most commonly used optical method for fitting skeletal models to electron density maps was the Richards' Optical Comparator, which used a half-silvered mirror to superimpose the model and the map (Richards, 1968). Since then more than a thousand protein structures have been solved and more portable, less cumbersome graphics-based methods have become essential for solution, analysis, and display of crystal structures. The development of computers that can display three-dimensional structures in real time has met this need and the development of various types of theoretical models and software to make them readily accessible continues to be a very active area of research.

The approaches that are used to address problems in computational biology often involve intensive numerical calculations that produce large amounts of data. For example, more than half a million structures may be produced in a molecular dynamics run with picosecond time intervals. The resulting datasets are both large and complex. Typically several different visual representations are used to analyze these data, since no single representation allows all aspects of the calculated structures to be analyzed. Graphics are also beginning to be used extensively in the theoretical modeling of protein folding, driven in no small part by renewed interest in protein structure prediction and folding which itself is due to the challenges of the human genome project and renewed optimism that the protein folding problem can be solved.

Computational approaches are used not only to address fundamental questions in protein biophysics but also in a number of important applications. One area in which molecular modeling and advanced visualization techniques already have a well-established track record is in the design of pharmaceuticals through rational drug design (Olson and Morris, 1993). A well known example is the development of inhibitors of HIV protease (Gustchina et al., 1994). Important advances in drug design include the development of sophisticated graphic and algorithmic docking techniques (cf. Meng et al., 1994), image analysis algorithms for the rapid analysis and classification of macromolecular complexes (Peitsch, et al., 1995), interactive techniques using simulated annealing approaches to dock flexible substrates into protein active sites (Surles et al., 1994) and methods for the accurate calculation of binding energies (reviewed in Lesyng and McCammon, 1993; Whittle and Blundell, 1994). Graphics are also widely used in molecular biology to help predict the three-dimensional structure and function of the gene products of new DNA sequences. Protein structure predictions generally depend upon comparing the unknown sequence with a large library of sequences in the databases. However, more sophisticated approaches are also under development; for example, the common protein motif database attempts to identify structural elements that are common in proteins and to use the consensus motifs to probe the structural database to identify functionally similar structural components (Murzin et al., 1995).

Most molecular graphics packages support the verification, querying, and browsing of macromolecular structures read from files with atom identification and

positions defined by Cartesian coordinates. These packages allow browsing by structural components such as polypeptide or polynucleotide chains, individual residues, and heterogroups (e.g., QUANTA, INSIGHT, and MI-DASPlus).

The most popular file format, the Protein Data Bank (PDB) at Brookhaven, also includes the residue name and number, chain identification (in multisubunit proteins), and crystallographic temperature factors and occupancies. A number of other formats are also in use and packages that translate from one format to another are available; one of the most widely used is BABEL (Shah et al. 1994).

Sophisticated packages include a variety of tools for analyzing the structure and comparing it with idealized geometry parameters. Various parameters can be plotted: for example, backbone torsional angles (Ramachandran plots) and temperature factors, while other geometric parameters such as bond distances and bond angles can be displayed in tables. Values selected in these displays can be highlighted in the structure, with separate displays for main-chain and side-chain geometry (bond lengths and dihedral angles), temperature factors, disulfide linkages, and non-bonded contacts. Specific amino acids or atom types can be selected. The current releases of most software tools also include facilities for validating the geometry of the structure.

Other molecular graphics packages can calculate more complex molecular properties such as fractal atomic density (Anvir et al., 1984), and maximal sphere accessibility (Lewis and Rees, 1985). Together these properties can be used to analyze structure and chemistry at the molecular surface and to identify ligand binding sites. Some of these programs can also be used easily as front ends for other more general purpose molecular modeling packages; for example, GEMM (Lee, unpublished observations). A few graphics programs accept input from another program or data passed over a network socket connection and dynamically update the displayed picture; for example QMView (Baldridge and Greenberg, 1995). These programs are particularly useful for monitoring the progress of large-scale simulations. Some graphics packages are designed primarily for generating publication quality, line-based black and white drawings (MolScript, ORTEP and SETOR).

In this chapter, we present an overview of the approaches that have been developed for visualizing and analyzing protein structures and describe important applications. Section II describes widely used representations of molecular structures while section III describes key techniques including animation and virtual reality approaches. Section IV discusses the application of molecular graphics in a variety of fields: molecular recognition and drug design, molecular dynamics, electrostatic modeling, crystallography, and electron microscopy. Unsolved problems are presented in section V. Future prospects are discussed in section VI while section VII describes the resources available, and widely used packages and databases are catalogued in section VIII.

II. REPRESENTATIONS

A. Wire Frame Models

Wire frame representations are a good first step in visualizing three-dimensional objects but quickly become unsatisfactory at moderate levels of complexity. Depthcuing, in which darker shades are used to represent more distant parts of the structure, is often used to give the viewer a better sense of the three-dimensional nature of the structure (Figure 1). With standard wire frame models, the viewer generally has a choice of combining color and the depthcue gradient to alter the appearance of the image. Wire frame representations are especially useful when examining changes in the backbone and side-chain positions between several
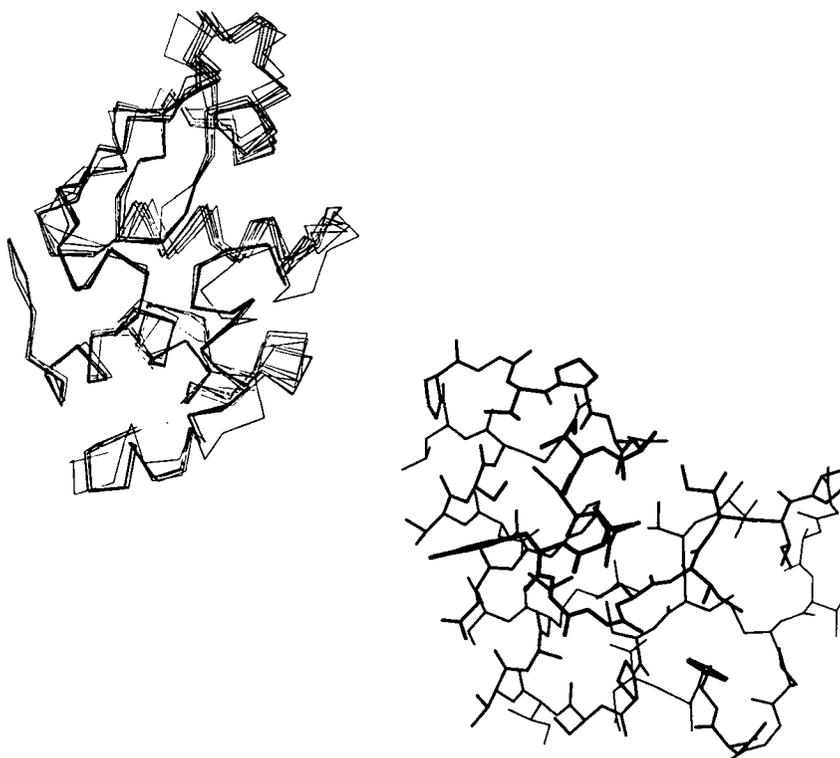
Figure 1. Wire frame drawings of crambin (PDB entry 1CRB) with depth cueing showing interatom connectivities.

superimposed structures (e.g., Trikha et al., 1995), packing of molecules in a crystal, and interactions between side-chains at an active site.

B. Schematic Models

Schematic diagrams are useful when emphasizing a specific structural feature; for example the topology, the arrangement of subunits and domains, and molecular or crystal symmetry. One of the most popular representations is the ribbon diagram in which helices are represented as ribbons and sheets as arrows showing the direction of the polypeptide chain (Figure 2a) (Richardson, 1981). Other representations include Artek plots, in which the polypeptide backbone is drawn using
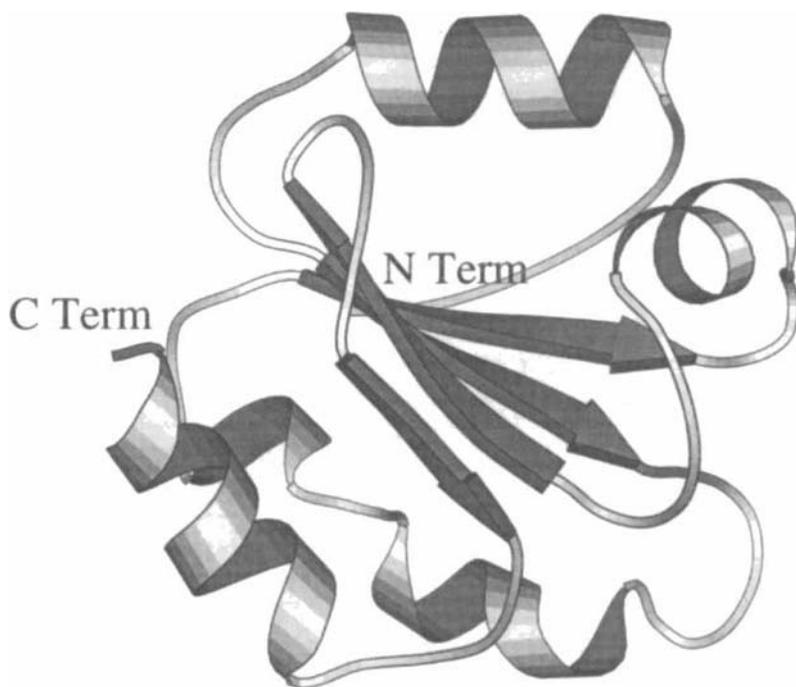
Figure 2.   (a) Ribbon drawing of thioredoxin (PDB entry 3TRX), illustrating the principal secondary structural elements. (b) The four helix bundle of amphibian ferritin (PBD entry 1RCE) shown with alpha carbon atoms connected by a smooth curve. The overall topology of the molecule and the connectivity are clearly seen. (c) Schematic representation of L. ferritin viewed down the 4-fold axis illustrating its 432 symmetry. Each subunit is represented by a cylinder with radius 13 Å.
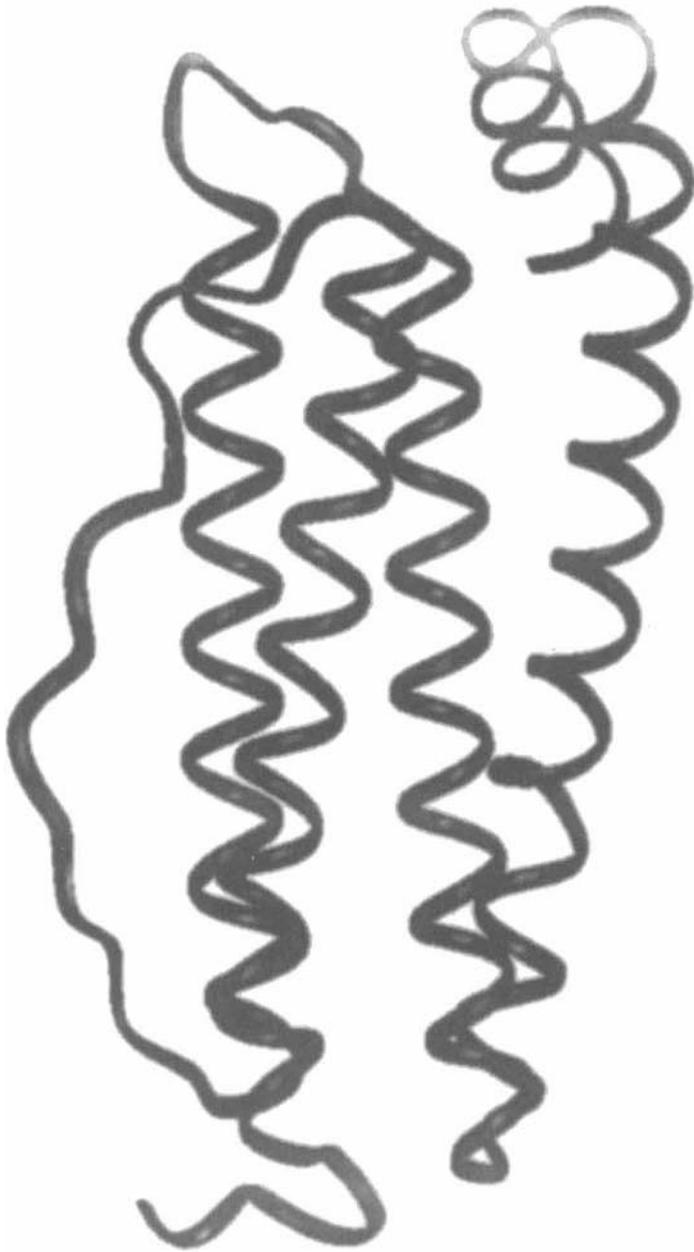
Figure 2. Continued.
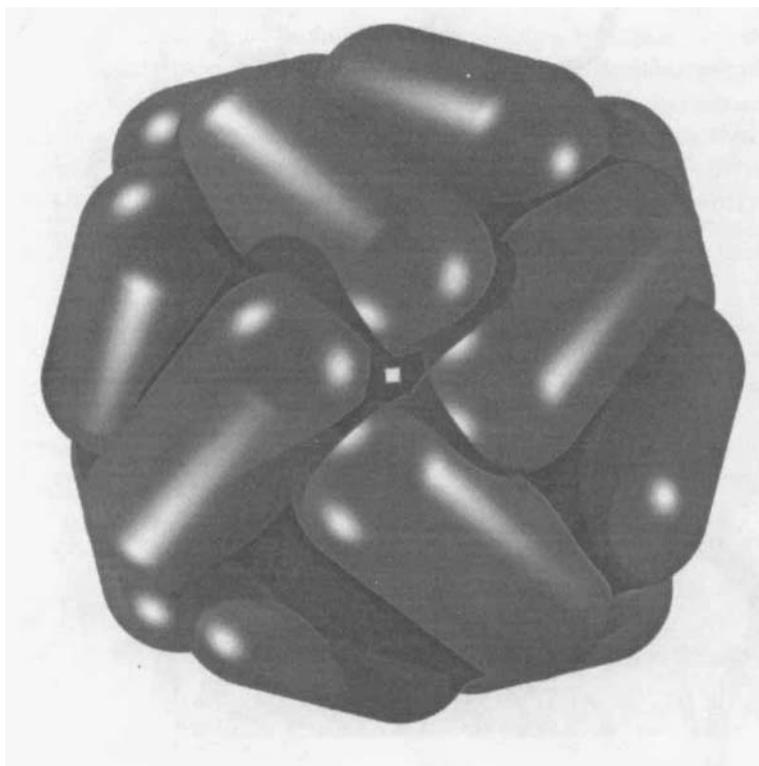
Figure 2. Continued.

arrows and cylinders (Lesk and Hardman, 1982), or as a smoothed curve through the alpha carbon atoms (Figure 2b). Cylinders having the same dimension and orientation as the polypeptide chain can also be used to illustrate the packing of subunits (Figure 2c).

C. Space Filling Models

Space filling models represent each atom as a sphere with a radius proportional to its Van der Waals radius and with the identity of the atom color coded; for example, carbon atoms in black, nitrogen atoms in blue, oxygen atoms in red, hydrogen atoms in white, and sulfur atoms in yellow in the CPK representation

(Figure 3). The CPK color scheme is based upon the colors of the popular plastic space filling models developed by Corey and Pauling and later improved by Kultun. The first space filling models were built to scale by hand. They are especially useful for illustrating in a realistic way the three-dimensional molecular shape, steric complementarity, and channels and cavities. Space filling models are often used to explore the chemical behavior of macromolecules; for example, the relationship between the free energy of protein folding and the solvent accessible surface area, or the relationship between ligand affinity and steric complementarity at a binding site.

D.   Molecular Surfaces

The computation of surface area, pioneered by Lee and Richards (1971), is still a very active area of research, since the accurate representation and rapid calculation of molecular surfaces are critical in calculating hydrophobic and electrostatic properties which, in turn, are the key to developing accurate models of protein folding.
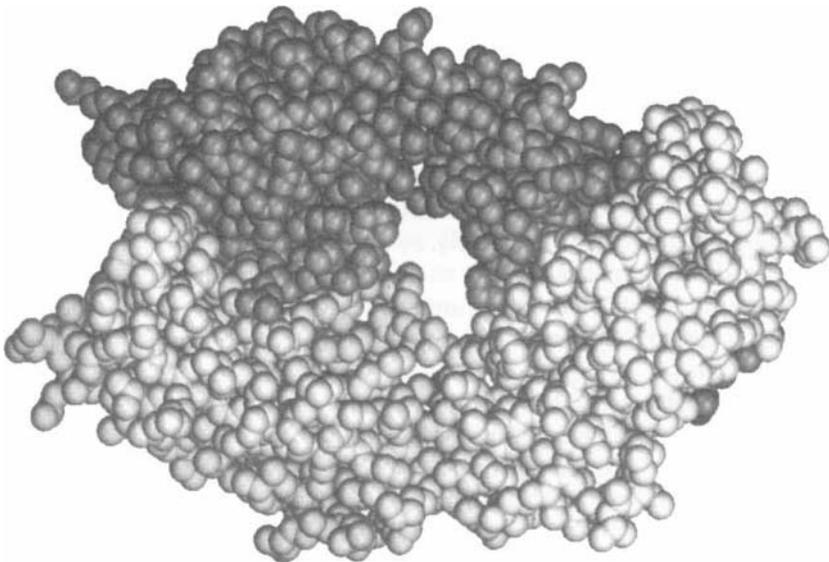
Figure 3. Space filling drawing of the structure of the phosphocholine binding immunoglobulin FAB fragment (PDB entry 1MCP; Satow et al., 1987). One subunit is shown in dark gray, the second in light gray. The principal use of a space filling drawing is to illustrate the packing of domains and side-chains within a structure.

A detailed discussion of the issues that must be considered in calculating and representing the molecular surface can be found in Richards (1977). One approach to calculating the surface is to treat the molecule as a collection of interlocking spheres representing individual atoms and to consider the molecular surface as the union of the exposed regions of these spheres. Calculating the solvent accessible molecular surface is more challenging. The general approach is to roll a solvent molecule or a cluster of solvent molecules, with a defined radius, around the surface. B-spline and density of surface neighborhood techniques are particularly useful in studying protein interactions at solvent interfaces; the spline functions are used to highlight significant features of the molecular surface by smoothing out local irregularities (Colloc'h and Mornon, 1990). The use of spherical harmonic functions to represent protein surfaces and properties (Duncan and Olson, 1993) allows the resolution of surface geometry and surface properties to be controlled and is particularly well-suited to analyzing protein-protein and protein-ligand interactions. Accurate representations of the molecular surface are also essential in continuum electrostatic calculations. A technique using curvilinear three-sided elements to calculate the molecular surface as a smooth triangulated manifold has recently been applied to the boundary element solution of the Poisson-Boltzmann equation (Zauhar, 1995).

E.   Mapping Molecular Properties

Molecular graphics is frequently used to map molecular properties onto the structure. The simplest approach is to place a glyph such as a dot or an arrow at every data point, with the size, orientation, and color of the glyph determined by a molecular property. Alternatively, surface or volume rendering may be used (see below). Both depend heavily on color coding, in which data values are mapped to a range of colors either continuously or discretely. The interpretation of the colored image is critically dependent on the choice of the mapping function.

One molecular property that is frequently displayed in this way is the electrostatic potential or gradient which is either mapped directly on the molecular surface or represented as an isopotential surface (cf. Lamotte-Brasseur et al., 1990; Vangelder and Wilhelms 1994; Hayd et al., 1995; Figure 4). Each of these representations conveys a different kind of information. Mapping the potential directly onto the surface of the molecule allows binding sites to be readily identified; for example a positively charged ligand is likely to bind to a region of the molecule with a large negative potential. This representation is also very useful when comparing the effects of mutations or changes in conformation on the local potential. Isopotential surfaces can, on the other hand, be used to identify channels through which a ligand may be funneled into the active site.
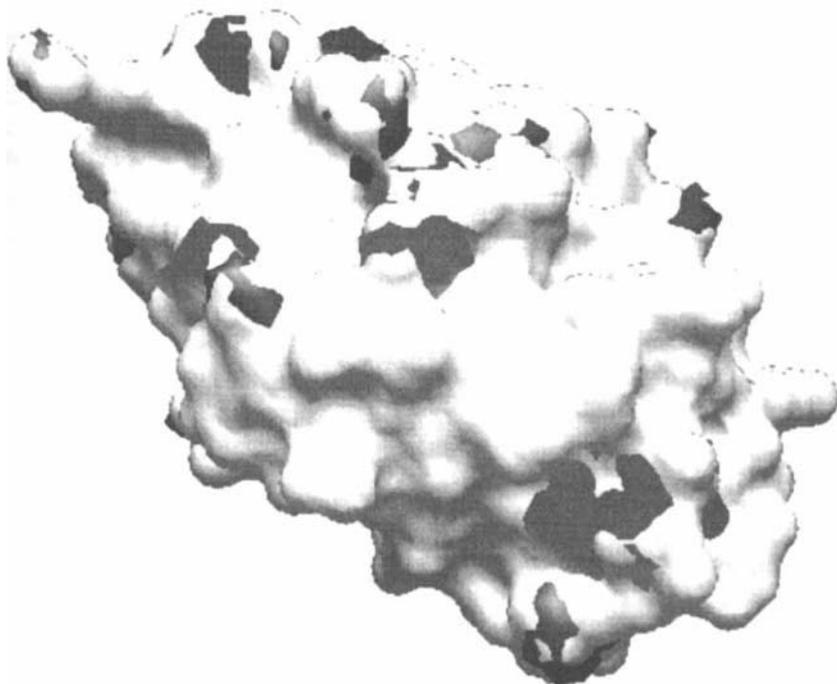
Figure 4. The electrostatic potential at the surface of lysozyme (PDB entry 2LYZA) produced with the program GRASP (Nicholls et al., 1991). Regions of positive electrostatic potential are shown as uniform gray surfaces rather then the usual blue. Regions of negative potential are shown as uniform black surfaces, rather than the usual red. Regions with varying shades of gray represent the molecular surface.

III. TECHNIQUES

A. Stereo Viewing

Each of the representations described above can be viewed in three dimensions using stereo views. Stereo views are generated by presenting the right and left eye with two different views, with one slightly rotated relative to the other. Hardware assisted and software based stereo viewers are available. In one implementation of

hardware stereo, the two images are drawn alternately on the same screen which the user views with special glasses that allow each eye to see only one image. Stereo views are indispensable when illustrating features such as side-chain motion, the relationship between structural features that are related in three dimensions, the opening and closing of an active site cleft, and the partial unfolding of a protein domain upon ligand binding.

B.   Surface and Volume Rendering

Surface rendering is a technique in which three-dimensional scalar data is converted to a geometrical representation and then rendered using conventional graphics methods. Isosurfaces and contours produce interpolated three-dimensional representations of the scalar data at fixed values of the scalar data. The surfaces can be colored according to the scalar value or with color codes that combine in one image two different properties, such as the electrostatic potential and the hydrophobicity (Figure 5). Surface rendering is widely used to visualize electrostatic potentials, electron densities and three-dimensional reconstructions of electron and light microscopic data. In volume rendering, the complete three-dimensional data set is displayed, without first generating a surface representation (reviewed in Koning et al., 1994). The main advantages of volume rendering are that it makes it possible to look at the variation in the data through space and to
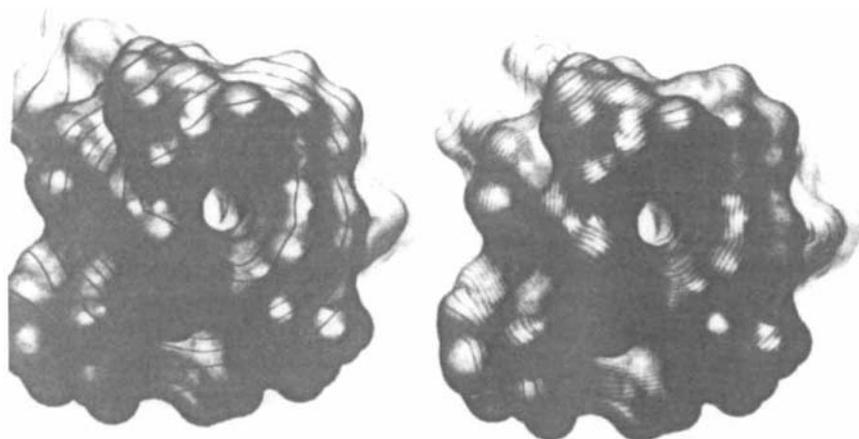
Figure 5.   An illustration of mapping a property to a molecular surface using textures. Isocontours of the hydrophobic potential are mapped onthe solvent-accessible surface of Gramicidin A. Using a different scaling for a texture map results in a different spacing of the isocontour lines (reproduced with permission from Teschner et al., 1994).

examine the whole molecule as well as the space around it by varying properties such as color and opacity.

C.   Texture Maps

By reducing geometric complexity, enhancing realism, and allowing several data sets to be presented at the same time, texture maps facilitate the presentation of large amounts of data in interactive environments. Texture mapping is a technique that applies an image to an object's surface as if the image were a sheet of cellophane. Textures exist in a parametric space, the texture space, that may be one-, two-, or three-dimensional; individual elements of the texture are called texels. Texture space can be thought of as an area in which information can be stored and linked to an object representation in three-dimensional space. An advantage of using textures is that the dimensionality of the texture space does not affect the object space. This independence makes it possible to show changes (such as a color code) immediately on the object by applying transformations in texture space instead of in the object space (Zachmann et al., 1995). Texture mapping has recently been used to display multichannel, three-dimensional structural information about molecular surfaces in real time (Teschner et al., 1994).

Volumetric information such as electrostatic potentials can also be sampled as a three-dimensional texture and then mapped on to the macromolecular surface. By taking the absolute value of the electric field and filtering with a delta function, one can pinpoint areas of highest local field gradient, often useful in identifying the binding site of a protein. This process can be used in combination with image processing techniques to screen large numbers of ligands of a molecule automatically.

D.   Animation

Animation techniques simulate continuous motion by rapidly displaying images in succession. The viewer is given the impression of watching the object in continuous motion. Even for moderately complex systems, computing the images may take several hours so that the individual frames must be either stored on disk or written to video tape. The first and simplest technique is flipbook animation in which the generated images are displayed one after the other. Its name refers to thumbing or flipping through a series of images. A second and more space and time efficient technique is keyframe animation in which only frames that correspond to changes in the characteristics of the motion are generated. Interpolation techniques are used to generate a set of images between two keyframes. The larger the interpolated set of images the smoother the conversion from one keyframe to the other will appear to the viewer.

E. Virtual Reality

Virtual reality is an emerging technology that attempts to immerse the user fully in an interactive computer generated environment. The participant in a virtual reality experience interacts with the system through a series of sensors and sophisticated output devices that give the impression that the participant is part of the synthetic environment. The Virtual Reality Modeling Language (VRML) (Bell et al., 1994) represents both an effort to develop a widely accepted specification for describing three-dimensional objects and a standard format for the storage and retrieval of three-dimensional data. A VRML scenario contains multi-participant virtual worlds hyperlinked through an appropriate transport mechanism and described using the VRML language. The chemistry extension to VRML allows the interactive visualization and distribution of three-dimensional structural information over the Web. The currently favored format for VRML 1.0 is a sub-set of the IRIS Inventor file format (Silicon Graphics Inc.).

IV. APPLICATIONS

A. Molecular Recognition and Drug Design

Because steric complementarity is the basis for the selectivity and specificity of enzymes, it has been used as a first test to discriminate among different modes of association of molecules. Optimal steric fit occurs when interacting atoms do not approach closer than their van der Waals radii and the free space between interacting atoms is minimized. Steric contacts that may restrict the passage of the ligand into the binding site must also be considered. Space filling models and molecular dot surface representations are frequently used to illustrate steric fit.

The chemical fit, which includes hydrogen bonding and electrostatic, hydrophobic, and hydrophilic complementarity, is generally the most important requirement for strong binding. Volumetric and isopotential electrostatic surfaces and color coded molecular surfaces are used to show chemical complementarity.

Drug design techniques include graphical model-building procedures and *de novo* search systems that simultaneously optimize the structural and chemical fit in a docking procedure (reviewed in Whittle and Blundell, 1994). Graphical approaches include visual examination of a binding site (Olson and Morris, 1993) in which interactive graphics is used to dock candidate ligands into the binding site manually. Potential ligands identified through either approach may be of widely varying sizes and have very little surface contact with the protein but still bind strongly through chemical interactions such as hydrophobic contacts and hydrogen bonds. Hence, early in the design process it is important to be able to display both the structural and chemical information in a manner that allows the visualization and evaluation of the complex.

De novo search techniques that optimize steric and chemical fit have been used to search for small molecules that fit a specific molecular interface (cf. Meng et al., 1994). Graphical techniques have been used to study structural complementarity in the trypsin-BPTI complex (Naray-Szabo, 1993) and in combination with *de novo* techniques to design a potent inhibitor of HIV-1 protease (DesJarlais and Dixon, 1994).

B. Molecular Surface Properties

The folding of a protein is accompanied by large changes in the solvent accessible area with different atom and residue types contributing differentially to the free energy change associated with the change of accessibility. Changes in solvent accessibilities labeled by magnitude and atom type and displayed on the molecular surface are frequently used to develop models of molecular recognition and folding (Fisher et al., 1990). Visualization of molecular surface changes is also used to study effects of mutations and ligand binding since even small changes in solvent accessibility often result in large effects on function.

In one recent application, the B-spline technique of calculating molecular surfaces was used to identify regions of several molecules that were at protein-protein interfaces and active sites. Surfaces at interfaces were found to be rough while active sites, although rough, were often surrounded by smooth regions (Colloc'h and Mornon, 1990). In a second application, graphical and statistical analysis of surface topology and sequence conservation of picornaviruses suggest that the canyons and surface depressions encircling the fivefold axes are the sites of receptor attachment in enteroviruses as well as human rhinovirus (Chapman and Rossman 1983).

Recently a technique for quantitatively visualizing molecular surface flexibility was used to analyze surface fluctuations of pancreatic trypsin inhibitor (PTI) in a molecular dynamics run; fluctuations between structures color coded onto the molecular surface using texture maps indicated that all exposed side-chains, with one exception, were flexible (Zachmann et al., 1995). The authors suggest that the high specificity of this inhibitor in binding to the active site of trypsin is achieved by close complementarity of the rigid regions of PTI and the protein, with the flexible regions interacting with the more mobile regions of the protein. These results underline the importance of the observation that both the flexibility and complementarity of ligands are critical in molecular recognition.

C. Electrostatic Potentials

Because ionizable residues frequently play a role in the functional mechanism of biological molecules, it is important to be able to visualize their charged states as a function of the pH, ionic strength, and other variables. Being able to calculate and visualize pH dependent electrostatic potentials is especially important for large

molecules, where it is difficult to access the information experimentally. Th electrostatic potential can be visualized using isopotential contours, isopotenti; surfaces, or rendered volumes. Examples are shown in Figures 6a and 6b.

Recently modeling studies using the Poisson-Boltzmann equation, molecula dynamics, and molecular graphics have been used to analyze the catalytic mecha nisms of several enzymes. Figure 6a shows that electrostatic interactions in sub tilisin are influenced by the shape of the boundary between the protein and solven

(a)

(b)

pH 2   pH 6   pH 11
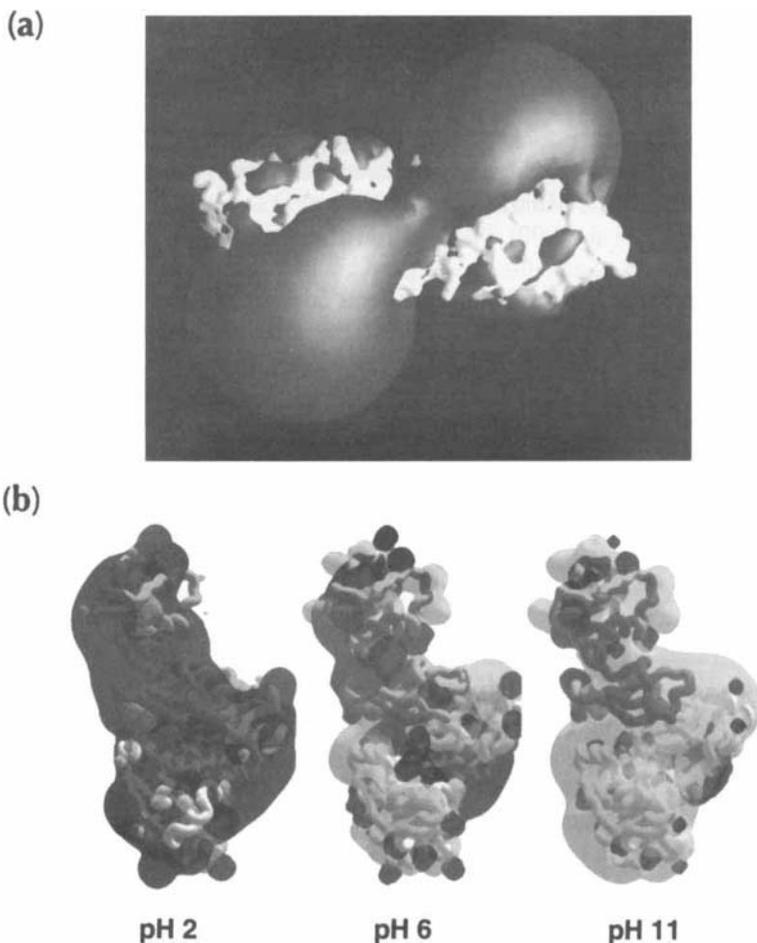
Figure 6. (a) Three dimensional potential surfaces ballooning from the two active sites of the acetylcholinesterase dimer (courtesy of Dr. Michael Gilson). (b) The pH-dependendent electrostatic potential around *Escherichia coli* aspartate transcarbamylase (PDB 6AT1). The very dark region on the molecule at 4 o'clock is a region of positive potential associated with the active site which binds negatively charged substrates. Note that it persists even at pH 11.

and that the region of increased electrostatic potential extends from the active site far into the solvent. Calculations for lysozyme showed that the magnitude of the electrostatic potential is strongly pH dependent (Oberoi and Allewell, 1993). The active site of aspartate transcarbamylase shows a persistent positive potential even at high pH where the protein is net negatively charged (Oberoi et al., 1996) illustrating how a negatively charged protein can bind a negatively charged ligand. Similarly, the electrostatic potentials at the active sites of two serine peptidases, α-chymotrypsin and subtilisin, were shown to have a large negatively charged volume (contoured at -10 kcal mol^{-1}) at the serine triad, with a very narrow channel in the immediate vicinity of the active site serine through which the ligand could enter the active site (Lamotte-Brasseur et al., 1990). Continuum methods have been used to propose an alternative entrance to the active site of acetylcholinesterase and a mutation to test this hypothesis (Gilson et al., 1994). The model improves agreement with kinetics studies; it is unlikely that this could have been discovered without the combined power of modeling and visualization.

D. Molecular Simulation Trajectories

The results of a molecular dynamics run are generally visualized using a combination of techniques (Ferrin et al., 1988; Miura et al., 1995). The simplest approach is to rapidly display wire-frame drawings of the succession of structures calculated from the molecular dynamics run to generate an animation of the dynamics of the molecule. Such animations are useful for identifying regions of the molecule that are undergoing rapid motion or that are relatively static, compared to the rest of the structure (Figure 7). Molecular dynamics simulations have been used to analyze the various stages of protein unfolding (reviewed in Karplus and Sali, 1995).

E. Crystallography

The molecular graphics packages used in three-dimensional structure determination provide a set of tools to display and interpret the electron density systematically. In the words of a crystallographer, an ideal package "would function very much like a drawing program in which it is possible to make and undo changes to the displayed image with ease, much like a three dimensional scratchpad" (Jones et al., 1991). An important requirement is that the user be able to switch between different representations of the data with ease. The package should also be able to maintain a database containing different bits of information about the molecule. While not all the requirements of an ideal system are present in any one single package today, there are several excellent alternatives.

Solving a crystal structure is a cyclic process in which the molecule is alternately fit to the electron density and then subject to energy minimization procedures. The packages available today have the ability to generate main-chain atoms from Cα

Figure 7. Molecular dynamics trajectory of lysozyme (PDB entry 2LYZA) at 5 ps intervals. The core of the protein is relatively rigid with very little movement in the backbone while the loop regions have moved up to 4 Å.

guide points using the fragment approach (Jones and Thirup, 1986), to select from a library of side chain rotamers; to carry out a real-space optimization of a rough model to the electron density, and to provide a residue-based index of fit between model and electron density map (Jones et al., 1991).

For example, in a typical application using the program O, the graphics display starts with a wire drawing of the molecule (the skeleton) superimposed on a simplified representation of the electron density which the user can rotate freely and view in stereo. The representation on the screen serves as a three-dimensional scratchpad that the user can use to modify the geometry of the skeleton. While providing a suitable structure to manipulate, this representation also allows an overview of the electron density map that is not possible with two-dimensional contour plots. At the outset, the skeletal atoms are classified as main-chain or side-chain; these categories are not necessarily correct, but are simply the program's best guess based on the connectivity of the skeleton. The user manipulates the skeleton to change the level of classification (color), the connectivity, and positions of the individual skeleton atoms. The program provides options for filtering and improving the skeleton automatically. The user can place pieces of three-dimensional text and labels as reminders in the map that can later aid in the task of deciding

where the sequence matches the density. Tools are provided to compare the real sequence with guesses that are made based on the shape of the density.

F. Electron Microscopy

Surface and volume rendering techniques are being used increasingly to visualize large scale assemblies such as cellular organelles, chromosomes, and large macromolecular assembles in their fully hydrated environment (Frank, 1995); for example several conformations of the nuclear pore complex have been analyzed using image reconstruction techniques (Pante and Aebi, 1994). Three-dimensional image reconstruction of the structures is done by combining several two-dimensional electron micrographic images of the structure to produce a three-dimensional volume of the structure. This volume is rendered using standard surface and volume rendering algorithms.

V. CHALLENGES

Our ability to visualize proteins depends on the availability and organization of the structural data and the tools we can use to look at it. There are substantial challenges in the organization, management, storage, and analysis of structural and sequence data. The absence of common formats for both storing and annotating coordinates makes comprehensive analyses difficult. Recent developments in this area include the Chemical Interchange Format (CIF) (Hall et al., 1991) that enables chemical information to be stored with structural information, and the chemical extension to VRML for distributing and visualizing molecular structures. The mmCIF format—based on the CIF format—is in the process of being adopted as the standard format for the PDB. Attempts to develop annotated databases of macromolecular structures represent the first step in developing databases that combine information from several different sources (Whittle and Blundell, 1994; Murzin et al., 1995).

Although the availability of protein and nucleic acid sequences and structures has been limiting in the past, several databases are now available online. Problems that still need to be solved include organizing molecular information in a way that is meaningful chemically, developing better metaphors for navigating through the information, and the development of relational databases.

While there has been significant progress in the development of software and hardware for graphics, the tools for retrieving and analyzing data can still be difficult to use. Typically one has to use a multitude of programs with incompatible user interfaces, file formats, and hardware requirements. The sheer volume of structural and sequence data being generated is also a major source of concern. This problem is most severe in molecular dynamics simulations, where both the techniques and tools for analysis are still in the early stages of development.

Standard techniques for displaying and rendering three-dimensional data have several shortcomings; for example each view has to be recalculated so it is difficult

to generate real time displays of complex systems. Methods need to be developed for displaying molecular properties along with structural data, analogous to those now used to map electrostatic potentials onto molecular surfaces. There is also little or no hardware support for rendering volumetric data on available graphics work-stations, so that volumetric rendering has to be done entirely in software. The design of software for the analysis of structural data is complicated by the fact that three-dimensional structural data is unstructured, so that there is no ordered progression from one data point to the next. Since computations are inherently sequential, developing computationally efficient algorithms for traversing and displaying the resulting data structures is challenging.

VI. FUTURE PROSPECTS

While it is difficult to predict the form the future will take, it is likely that graphics packages will make increasing use of comprehensive databases for storing and retrieving information. The integration of display and searching of three-dimensional data into the Web protocol (e.g., with kinemage and the chemistry extension to VRML) will almost certainly produce tools that will make it possible to carry out large scale analyses of structural, chemical, and sequence data from multiple sources. With the adoption of object-oriented techniques for software development, it should be possible to develop software that is reusable and can communicate with databases and other packages. Collaborative efforts to develop integrated and extensible tools for the storage and retrieval of sequence information are already underway (Prophet, Bolt, Beranek, and Newman, Cambridge).

The availability of powerful three dimensional browsers that link together databases maintained in different locations will foster collaborations between scientists in ways not even previously conceived. Application of concepts drawn from virtual reality to the display, analysis, visualization, annotation, and manipulation of molecular data will produce technologies that will enable abstract concepts to be comprehended more easily and therefore spark new ideas.

Advances in hardware capabilities and representations that reduce geometric complexity, such as texture mapping, will make it easier to visualize complex data. Cross disciplinary efforts such as the application of techniques used in computer-aided design and visualization for calculating molecular geometric properties will result in further improvements. Since visualization techniques make it possible to compress a large amount of information into one picture, techniques such as data-browsing and active data visualization will become increasingly important in the analysis of molecular data.

Active data visualization techniques make it easy to track information between linked windows into data. For example, in an evolving molecular dynamics run, each time one highlights a new area of interest in one of the windows, the equivalent position is simultaneously updated in each of the other display windows. These

techniques make it possible to identify correlations between different attributes of a macromolecule in both space and time. This makes it possible to view the data selectively and interactively in real time and can provide the basis for generating novel structures that do not themselves appear in the molecular dynamics run.

Data browsing techniques in which the user views the data at coarse resolution and then focuses into the areas of interest are particularly useful in viewing multidimensional data (Keim and Kriegel, 1994). For example, data selected from a library of protein motifs can be displayed as a two-dimensional array in which each display pixel on the screen represents one database item with the pixels arranged and colored to indicate their relevance to a query. When the user selects a pixel, it expands to a thumbnail view of structures that have similar motifs. By successively narrowing the focus of the view, the user can "drill down" into the data.

Because metaphors in the form of graphic images are so powerful in examining and understanding complex scientific concepts, computer graphics has already produced paradigm shifts in many areas of research. As the developments sketched out above come to fruition, further shifts are certain to occur.

VII.   RESOURCES

A.   Databases

While there are several databases that contain molecular structural information, the most comprehensive are:

- Protein Data Bank at Brookhaven (PDB) at the Brookhaven National Laboratories, USA (Bernstein et al., 1977)
- Nucleic Acid Database (NDB) at Rutgers University, USA
- Cambridge Crystallographic Database at the Cambridge Crystallographic Center, UK (Allen et al., 1991; Allen and Kennard, 1993)
- Drug Information System 3D Database at the NIH/NCI, USA.

Together these allow researchers easy access to more than 100,000 crystallographic or NMR structures of large and small molecules.

Databases that combine structural, chemical and other information that are being developed include the following:

- SCOP contains a description of the structural and evolutionary relationships between all proteins whose structure is known, including all entries in the PDB. It is available as a set of linked hypertext documents. The hypertext pages are linked to PDB entries, sequences, references, images, and interactive display systems. Unlike other databases, SCOP was created by manual

inspection and a set of automated methods. It provides a broad survey of all known protein folds, detailed information about the close relatives of any particular protein, and a framework for future research and classification (Murzin et al., 1995).

- 3D_ALI contains both protein structural and sequence information including structural superpositions among proteins with similar main-chain folds (Pascarella and Argos, 1992).
- FSSP (families of structurally similar proteins) lists structural alignments of proteins in the PDB based purely on the three-dimensional coordinates of the proteins and derived by automatic structure comparison programs (Holm et al., 1992). The significance of structural similarity is estimated statistically. Related databases include HSSP (Sander and Schneider, 1991) and DSSP (Kabsch and Sander, 1983).

A current list of widely available databases and molecular modeling resources is maintained at the NIH, USA and can be accessed through the Internet (URL http://www.nih.gov/molecular_modeling)

B.   Internet

The explosive growth of the Internet and the popularity of browsers such as Mosaic and Gopher have prompted the development of online databases and packages designed to allow scientists to query and retrieve information effortlessly. The implementation of URL exchange via the Common-Client Interface into molecular graphics packages will produce Web browsers that are capable of accessing information on the Internet. These browsers will make it possible for widely separated sites to exchange structural and chemical information. By exploiting the ability of Web documents to contain hyperlinks to other documents, the selection of an object in the three-dimensional browser can be used to invoke further actions, thus providing a feature-rich three-dimensional environment for exploring data. The client-server model of the Web also makes it possible to link databases for both retrieving and archiving data. NCI, NDB, and the PDB support both the retrieval and deposition of data using the Web.

VIII.   PROGRAMS

The programs listed below are widely used to construct, display, modify and analyze molecular structures. Some are commonly used specifically for X-ray and NMR structural analysis and have additional functions for structure solution. Other graphics packages are less specialized and are used to display proteins, nucleic acid, and small molecules in a variety of representations and color schemes.

- AVS is a general purpose visualization application software and development environment with a modular design. It accepts various forms of data and uses several visualization techniques for rendering objects. Modules for the display of molecules are available both from the vendor and through the Internet (AVS User Guide, 1994; http://www.avs.com).
- EyeChem is a modular visualization environment based on IRIS Explorer. The EyeChem suite of modules is designed for the collaborative visualization and manipulation of molecular data using the remote rendering capabilities of IRIS Explorer. Because the application is based on IRIS Explorer, new molecular formats can be rapidly and easily assimilated into EyeChem by developing new modules. With the recent addition of WWW and Multipurpose Internet Mail Extension (MIME) interfaces, the application can be used to access data remotely through the Internet (Casher et al., 1994).
- FRODO is one of the first graphics packages to be widely used (Jones, 1978). Developed in the late 1970s primarily for the X-ray crystallographic community, it is available in several "flavors" as different users have modified and enhanced the package over the years. Its principal author has recently developed another package, O, that is itself a sophisticated visualization and structure solution environment (Jones et al., 1991).
- GEMM (Generate, Edit, and Manipulate Molecules), is an interactive molecular graphics package developed at NIH. It is a tool for constructing, viewing, modifying, and manipulating a molecular structure in three dimensions. The molecular structure can be any collection of small and large molecules. The program is menu-driven and is specifically designed for the bench scientist (Lee, 1992).
- GRASP is a graphics program written for Silicon Graphics computers that is widely used by the structural biology community to visualize macromolecules. Its particular strengths are its facility with surfaces and with displaying electrostatic potentials both as rendered surfaces and mapped onto the molecular surface. The program also contains a scaled down version of the solution to the Poisson-Boltzmann equation (Nicholls et al., 1991).
- Insight is a comprehensive graphic molecular modeling program. In conjunction with the molecular dynamics/mechanics program DISCOVER, it can be used to build and manipulate molecules (Insight II Reference Guide, 1994).
- MAGE Kinemages, kinetic protein images, were introduced in the first issue of Protein Science (Richardson and Richardson, 1992). The kinemage format is proposed as one of the chemical MIME standards; specifically, chemical/x-kinemage. The entire image can be rotated in real time, parts of the display can be turned on or off, any point can be identified by picking it, and the transition between different structures can be animated. The kinemage format is a mid-level form that is not commonly used except within other programs.
- MolScript is a "batch" utility to generate publication quality graphics. It reads an input file that describes what to draw, and outputs a postscript file. The

input file specifies the coordinate file(s) to use, how to transform the coordi-
nates, what graphical objects to create, and various parameters such as colors,
shading, and line width (Kraulis, 1991).

- O is a multi-platform X-ray crystal structure solution package. It has an
  integrated database for the storage of molecular information, is driven by a
  macro command language and has a powerful user configurable interface
  (Jones et al., 1991).
- QMView allows the visualization of quantum mechanical data. It can display
  the chemical structure, quantum mechanically determined vibration modes,
  electronic properties and three-dimensional molecular orbitals (Baldridge
  and Greenberg, 1995).
- QUANTA is a general purpose molecular graphics pachage. It runs under the
  UNIX operating system and is available on several platforms (QUANTA
  User's Guide, 1994).
- RasMol is a freely available program that can be used to generate high
  quality images of molecules. The package can be used to view the
  molecule interactively as well as to generate animations using a scripting
  language. This program has the added advantage of being available on
  both microcomputers (Macintosh, personal computers) as well as high end
  machines that support the X windowing system. There are several en-
  hanced versions of this package including a Web aware version (Sayle,
  1994).
- Raster3D is a set of tools for generating raster images of proteins or other
  molecules for publication. The program renders spheres, triangles, and cyl-
  inders with specular highlighting, Phong shading, and shadowing. Ancillary
  programs process atomic coordinates from PDB files into rendering descrip-
  tions for pictures composed of ribbons, space-filling atoms, bonds, and ball
  and stick representations (Merit and Murphy, 1995).
- SCARECROW analyzes and displays the trajectories from several dynamics
  programs. Two-dimensional history plots of variables such as distances, bond
  angles, torsion angles, and energies variations during a dynamics simulation
  can be displayed (Laaksonen, 1992).
- SCULPT allows the interactive manipulation of atoms and secondary struc-
  ture, interactive docking of a ligand into a flexible receptor, and on-the-fly
  specification of movable and frozen side-chains and backbones. It supports
  amino and nucleic acids, cofactors, arbitrary small molecules, and electro-
  static and van der Waals interactions and visual feedback on van der Waals
  interactions (Surles, et al., 1994).
- SETOR allows the display of secondary, tertiary, and quaternary structures
  of molecules and use of a wide variety of rendering models. It provides the
  standard methods of displaying wire frame models of all or a subset of atoms
  in the molecule and can also be used to generate publication quality shaded
  models of solid surfaces (Evans, 1993).

- VMD is a molecular graphics program designed for ease of use, modifiability, and modularity. In addition to supporting several different display styles, it has an atom selection language in which multiple molecular representations can be applied to arbitrary groups of atoms simultaneously (Humphrey, et al., 1995).

ACKNOWLEDGMENTS

REFERENCES

Allen, F. H., & Kennard, O. (1993). 3D search and research using the Cambridge structural database. Chem. Design Automation News 8, 1; 31-37.

Allen, F. H., Davies, J. E., Galloy, J. J., Johnson, O., Kennard, O., Macrae, C. F., Mitchell, E. M., Mitchell, G. F., Smith, J. M., & Watson, D.G. (1991). The development of versions 3 and 4 of the Cambridge structural database system. J. Chem. Inf. Comp. Sci. 31, 187.

Anvir, D., Farin. D., & Pfeifer, P. (1984). Molecular fractal surfaces. Nature 308, 261-263.

AVS User Guide. (1994). Version 5.0, Advanced Visual Systems, Inc. Waltham.

Baldridge, K. K., & Greenberg, J. P. (1995). Qmview—a computational chemistry 3-dimensional visualization tool at the interface between molecules and mankind. J. Mol. Graphics 13, 63-66.

Bell, G., Parisi, A., & Pesce, M. (1994). The Virtual Reality Modeling Language, November 1994. See http://www.eit.com/vrml/vrmlspec.html.

Berners-Lee, T.J., Cailliau, R., Groff, J-F., & Pollermann, B. (1992). CERN, World-Wide Web: The Information Universe, in Electronic Networking: Research, Applications and Policy, Vol. 2 No 1, pp. 52-58, Meckler Publishing, Westport.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977). The Protein Data Bank: A computer-based archival file for macromolecular structures. J. Mol. Biol. 112, 535-542.

Casher, O., Green, S. M., & Repa, H. S. (1994). EyeChem 1.0: A modular chemistry toolkit for collaborative molecular visualization J. Mol. Graphics 12, 226-230.

Chapman, M. S., & Rossmann, M. G. (1993). Comparison of surface properties of picornaviruses: Strategies for hiding the receptor site from immune surveillance. Virology 195, 745-756

Colloc'h, N., & Mornon, J.-P. (1990). A new tool for the qualitative and quantitive analysis of protein surfces using B-spline and density of surface neighbourhood. J. Mol. Graphics 8, 133-140.

DesJarlais, R. L., & Dixon, J. S. (1994). A shape- and chemistry-based docking method and its use in the design of HIV-1 protease inhibitors. J. Comp-Aided. Mol. Design 8, 231-242.

Duncan, B. S., & Olson, A. J. (1993). Approximation and characterization of molecular surfaces. Biopolymers 33, 219-229.

Evans, S. V. (1993). Setor: Hardware lighted three dimensional solid representations of macromolecules. J. Mol. Graphics 11, 134-138.

Ferrin, T. E., Huang, C. C. Jarvis, L. E., & Langridge, R. (1988). The MIDAS display system. J. Mol. Graphics 6, 13-27.

Fisher, C. L., Tainer, J. A., Pique, M. E., & Getzoff, E. D. (1990). Visualization of molecular flexibility and its effect on electrostatic recognition. J. Mol. Graphics 8, 125-132.

Frank, J. (1995). Approaches to large-scale structures. Curr. Opin. Struct. Biol. 5, 194-201.

Gilson, M. K., Straatsma, T. P., McCammon, J. A., Ripoli, D. R., Faerman, C. H., Axelson, P. H., Silman, I., & Sussman, J. L. (1994). "Open back door" in a molecular dynamics simulation of acetylcholinesterase. Science 263, 1276-1278.

Goodsell, D. S., & Olson, A. J. (1993). Molecular illustration in black and white. J. Mol. Graphics 10, 253.

Gustchina, A., Sansom, C., Prevost, M., Richelle, J., Wodak, S. Y., Wlodawer, A., & Weber, I. T. (1994). Energy calculations and analysis of HIV-1 protease inhibitor crystal structures. Protein Eng. 7, 309-317.

Hall, R., Allen, F. H., & Brown, I. D. (1991). The crystallographic information file (CIF)—A new standard archive file for crystallography. Acta. Crystallogr. A47, 655

Harvey, S. C. (1989). Treatment of electrostatic effects in macromolecular modeling. Proteins 5, 78-92.

Hayd, H., Bergner, A., & Preuss, H. (1995). Game—A computer-graphics method for calculating and displaying the molecular electrostatic potential. J Mol. Graphics 13, 2-9.

Holm, L., Ouzounis, C., Sander, C., Tuparev, G., & Vriend, G. (1992). A database of protein structure families with common folding motifs. Protein Science 1, 1691-1698.

Humphrey, W., Dalke, A., Hamer, K., Leech, J., & Phillips, J. (1995). VMD Programmer's Guide Version 1.0. University of Illinois and Beckman Institute, Urbana.

Insight II (1994). User Manual, Version 2.3, Biosym Technologies, Burlington.

Insight II (1991). Reference Guide, Version 1.1.0 and Reference Guide 2.0.0, Biosym Technologies, San Diego.

Jones, T. A. (1978). A graphics model building and refinement system for macromolecules. J. Appl. Crystallogr. 11, 268-272.

Jones, T. A., & Thirup, S. (1986). Using known substructures in protein model building and crystallography. EMBO J. 5, 819-822.

Jones, T. A., Zou, J. Y., Cowan, S. W., & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and locatin of errors in these model. Acta Crystallogr. A47, 110-119.

Kabsch,W., & Sander,C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22, 2577-2637.

Karplus, M., & Sali, A. (1995). Theoretical studies of protein folding and unfolding. Cur. Opin. Struct. Biol. 5, 58-73.

Keim, D. A., & Kriegel, H. P. (1994). VisDB—Database exploration using multidimensional visualization. IEEE Comp. Graphics Appl. 14, 40-49.

Koning, A. H. J., Zuiderveld, K. J., & Viergever, M. A. (1994). Medical volume visualization on general-purpose parallel architectures. Med. Informatics 19, 283-293.

Kraulis, P. J. (1991). MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. J. Appl. Cryst. 24, 946-950.

Laaksonen L. (1992). A graphics program for the analysis and display of molecular dynamics trajectories. J. Mol. Graphics 10, 33-34.

Lamotte-Brasseur, J., Dive, G. Dehareng, D., & Ghuysen, J-M. (1990). Electrostatic potential maps at the quantum chemistry level of the active sites of the serine peptidases, a-chymotypsin and subtilisin. J. Theor. Biol. 145, 183-198.

Langridge, R. (1995). Three-dimensional molecular graphics. Science 267, 1890-1891.

Lee, B. K. & Richards, F. M. (1971). The interpretation of protein structures: Estimation of static accessibility. J. Mol. Biol. 55, 379-400.

Lesk, A. M., & Hardman, K. D. (1982). Computer generated schematic diagrams of protein structures. Science 216, 539-540.

Lesyng, B., & McCammon, J. A. (1993). Molecular modeling methods. Basic techniques and challenging problems. Pharmac. Ther. 60, 149-167.

Lewis, M., & Rees, D. C. (1985). Fractal surfaces of proteins. Science 230, 1163-1165.

Meng, E. C., Kuntz, I. D., Abraham, D. J., & Kellogg, G.E. (1994). Evaluating docked complexes with the HINT exponential function and empirical atom hyrophobicities. J. Comput-Aided Molec. Des. 8, 299-306.

Merit, E., & Murphy, M. E. P. (1995). Raster3D Version 2.0. A program for photorealistic molecular graphics. Acta. Crystallogr. D 50, 869-873.

Miura, R., Yamano, H., Yamauchi, R., Katagiri, M., Kubo, M. Vetrivel, R., & Miyamoto, A. D (1995). evelopment of RYUGA for 3-dimensional dynamic visualization of molecular-dynamics results. Catalysis Today 23, 409-416.

Murzin, A. G., Brenner, S. E., Hubbard, T., & Chothia, C. (1995). SCOP: A structural classification of proteins database for the investigation of sequences and structure. J. Mol. Biol. 247, 536-540.

Naray-Szabo, G. (1993). Analysis of molecular recognition: steric electrostatic and hydrophobic complementarity. J. Mol. Recog. 6, 205-210.

Nicholls, A., Sharp, K. A., & Honig, B. (1991). Protein folding and association: insights from the interfacial andthermodynamic properties of hydrocarbons. Proteins 11, 4, 282.

Oberoi, H., & Allewell, N. M. (1993). Multigrid study of the nonlinear Poisson-Boltzmann equation and calculation of titration curves. Biophy. J. 65, 48-55.

Oberoi, H., Trikha, J., Yuan, X., & Allewell, N. M. (1996). Identification and Analysis of Long Range Electrostatic Effects in Proteins by Computer Modeling: Aspartate Transcarbamylase Proteins Proteins: Structure, Function, and Genetics 25, 300-314.

Olson, A. J., & Morris, G. M. (1993). Seeing our way to drug design. Perspect. Drug. Discov. Des. 1, 1329-344.

Pante, N., & Aebi, U. (1994). Towards understanding the three-dimensional structure of the nuclear pore complex. Curr. Opin. Struct. Biol. 4, 187-196

Pascarella, S., & Argos, P. (1992). A data bank merging related protein structures and sequences. Protein Eng. 5, 121-137.

Peitsch, M. C., Stampf, D. R., Wells, T. N.C., & Sussman, J. L. (1995). The Swiss-3DImage collection and PDB-Browser on the World-Wide Web. TIBS 20, 82-84

QUANTA User's Guide (1994). Version 3.3, Molecular Simulations Inc. Waltham.

Richards, F. M. (1968). The matching of physical models to three dimensional electron density maps: A simple optical device. J. Mol. Biol. 37, 225-230.

Richards, F. M. (1977). Areas, volumes, packing and protein structure. Ann. Rev. Biophys. Bioeng. 6, 151-176.

Richardson, D. C., & Richardson, J. S. (1992). MAGE. Protein Science 1, 1-3.

Richardson, J. S. (1981). The anatomy and taxonomy of protein structure. Adv. Protein Chem. 34, 167-339.

Rzepa, H. S., Murray-Rust, P., & Whitaker, B. J. (1995). Internet Request for Comment (RFC).

Sander, C., & Schneider, R. (1991). Database of homology derived protein structures and the structural meaning of sequence alignment. Proteins 9, 56-68.

Satow, Y., Cohen, G. H., Padlan, E. A., Davis, D. R. (1987). Phosphocholine binding immunoglobulin MC/PC603. An x-ray diffraction study at 2.7 Angstroms. J. Mol. Biol. 190, 593-601.

Sayle, R. (1994). RasMol user manual Version 2.5 Glaxo Research & Development, Greenford, Middlesex.

Shah, A.V., Walters, W. P., Shah, R., & Dolata, D. P. (1994). Babel a tool for converting between molecular data formats. In: Computerized Chemical Data standards: databases, data interchange, and information systems ASTM STP 1214 (Lysakowski, R. & Gragg, C. E., Eds.), American Society for Testing and Materials, Philadelphia.

Surles, M. C., Richardson, J. S., Richardson, D. C., & Brooks, F.P. Jr. (1994). Sculpting proteins interactively: Continual energy minimization embedded in a graphical modeling system. Protein Sci. 3, 198-210.

Teschner, M., Henn, C., Vollhardt, H., Reiling, S., & Brickmann, J. (1994). Texture mapping: A new tool for molecular graphics. J. Mol. Graphics 12, 98-105.

Trikha, J., Theil, E. C. , & Allewell, N. M. (1995). High resolution crystal structure of amphibial red-cell
    L ferritin: Potential roles for structural plasticity and solvation in function. J. Mol. Biol. 248,
    949-967.
Vangelder, A., & Wilhelms, J. (1994). Topological considerations in isosurface generation. J. Mol.
    Graphics 13, 337-375.
Whittle, P. J., & Blundell, T. L. (1994). Protein structure based drug design. Annn. Rev. Biophy. Biomol.
    Struct. 23, 349-375
Zachmann, C. D., Kast, S. M., & Brickmann, J. (1995). Quantification and visualization of
    molecular-surface flexibility. J. Mol. Graphics 13, 89-97.
Zauhar, R. J. (1995). Smart—A solvent-accessible triangulated surface generator for molecular graphics
    and boundary-element applications. J. Comp.-Aided Molec. Design 9, 149-159.

ELECTROSTATICS AND HYDROGEN BONDING

Bertrand García-Moreno E.

I. INTRODUCTION

Interactions among and between polar and ionizable groups are generally classified as either electrostatic interactions or hydrogen bonds. The term electrostatic interactions usually implies interactions between ionizable groups; hydrogen bond refers to the interaction of a donor proton bearing partial positive charge with the electron density of an acceptor polar atom. Separation of these two types of noncovalent interactions is somewhat artificial as hydrogen bonds are fundamentally electrostatic. However, since the physical origins and the structural consequences of these two forces in proteins are significantly different, it is therefore convenient to distinguish between them when studying or interpreting their role in structure and function.

Electrostatic interactions and hydrogen bonding play important roles in many aspects of folding, stability, and activity of proteins. There has been debate in the past about the exact magnitude of the contribution by hydrogen bonding to folding vis-a-vis the strongly stabilizing contribution by the hydrophobic effect. Currently, hydrogen bonds are thought to play a preeminent role in determining specificity as well as stability of the native states of proteins. The magnitude of contributions by electrostatic interactions to stability of folded structures, on the other hand, are generally believed to be modest compared to those by hydrophobic, van der Waals, or hydrogen bonding interactions. Electrostatics mediates interactions between macromolecules and polar and ionic components of their environment, including water, protons, and salts, and they play a fundamental role in the physiological regulation of many proteins.

The molecular origins and physical character of electrostatic interactions in proteins are relatively well understood. Algorithms are even available for quantitation of electrostatic energies from X-ray elucidated structures based on principles of classical electrostatics and statistical mechanics. The scope of the discussion on electrostatics is restricted to physical and structural determinants of electrostatic forces and to manifestations of these forces in structure and function of proteins. In contrast, the molecular basis of the energetics of hydrogen bonds is not well understood at present. Much of what is known about the physical and structural character of hydrogen bonds has been determined from surveys of hydrogen bonds in X-ray elucidated proteins. Consequently the discussion on hydrogen bonds is focused on phenomenological analyses of the contributions by hydrogen bonds to protein folding.

II. PHYSICAL CHARACTER OF ELECTROSTATIC INTERACTIONS

The electrostatic properties of a protein are determined by the electrostatic potential (ψ_{el}). In its most general form the potential can be expressed as:

$$\psi el, q = \sum_{i=1}^{n} \frac{332\, Z_i}{A\, r_{iq}} \qquad (1)$$

where r_{iq} is the distance (Å) between charge i and point q, A represents an attenuation parameter to quantify the effects of the dielectric and ionic environment on ψel, and Z_i is the charge at site i. ψel describes the energy (kcal/mol) of a unit positive test charge located at point q.

To understand the molecular origin of electrostatic interactions in proteins, and to describe their consequences on structure and function, it is necessary to know the magnitude of the electrostatic potential, its sources, and the manner in which it determines chemical properties of proteins.

Three different classes of atoms constitute the covalent structure of a protein: nonpolar, polar, and charged or ionizable. All but the nonpolar atoms contribute directly to the electrostatic potential. The ionizable atoms are the atoms of weakly basic and acidic side-chains that can bear charge as a result of their interactions with hydrogen ions. Polar atoms do not normally react with hydrogen ions over the pH range where the peptide bond is stable and they do not bear formal charge equal to the valence of the atom. Instead they bind covalently to atoms of different electronegativity. The resulting permanent dipoles can be represented as partial or virtual charges localized at each end of the dipole. Most polar atoms in proteins establish complementary interactions with other polar atoms through the hydrogen bond, consequently the charge of most of these dipoles is neutralized locally.

Contributions by nonpolar atoms to the magnitude and properties of the electrostatic potential are indirect but extremely important. The dielectric properties of the interior of proteins are determined primarily by nonpolar atoms. The surface of proteins constitutes a boundary between two phases of different dielectric. The internal phase always has a lower dielectric. The exact value of the internal dielectric remains controversial but it is at least one order of magnitude lower than the dielectric of the solvent phase, which for water soluble proteins corresponds to the dielectric of water ($D_{H2O} = 78.5$ at 298 K). Understanding the relative contributions by the two dielectrics to the electrostatic potential is one of the central problems in electrostatics of proteins.

Charged atoms at the surface of proteins mediate the interactions of proteins with components of their environment such as water, ions, other macromolecules, and small molecular weight metabolites. For this reason the chemical composition of proteins often reflects the ionic character of their physiological milieu. Water soluble proteins, for example, usually have a higher content of ionizable groups than membrane-bound proteins. Some proteins have been tailored evolutionarily to achieve a specific balance of acidic and basic residues that can buffer pH in the range where the protein is most active. Hemoglobin and pepsin, for example, can buffer solutions in the neutral and acidic pH range, respectively; these are the pH

ranges optimal for their physiological function. The proportion of basic and acidic residues in a protein is also correlated with the ionic composition of its milieu. Extracellular proteins tend to be acidic, and it is tempting to conjecture that acidic proteins can take advantage of the high concentrations of extracellular Ca^{2+} and Na^+ by binding cations at clusters of acidic residues. The ensuing stabilization of the folded protein could minimize damage from unfortunate encounters with rampant extracellular proteases.

Charged groups mediate the interactions between proteins and salts. Ions in solution can respond without any hindrance to attractive or repulsive potentials emanating from a protein, therefore they can screen or neutralize the electrostatic potentials very effectively. In regions where the electrostatic potentials are weak the interactions between proteins and salts are correspondingly weak; ions hover in a diffuse cloud near the attracting charges without interacting chemically with the protein. Ions can become trapped or bound in wells when the attraction by the electrostatic potential exceeds the randomizing influence of thermal motion This capacity to recognize and bind specific ions has been harnessed by many proteins for structural and functional purposes.

A.   The Charged State of a Protein

One of the main objectives of the study of electrostatics in proteins is to elucidate microscopic, molecular mechanisms whereby protons and salts regulate structural and functional energetics. This commonly entails measurement and analysis of changes in the charged state of proteins during equilibrium reactions. Any equilibrium reaction that is accompanied by binding or release of protons can be regulated by pH. To understand the molecular mechanism of this proton linkage it is necessary: (1) to measure the number of protons that are bound or released during the reaction, (2) to identify the groups that bind or release protons, and (3) to understand why the affinity for protons is different in the two end states of the reaction (García-Moreno E., 1995; Matthew et al., 1985).

The charged state of a protein is determined primarily by the constituent weak acid and base moieties and by the concentration of hydrogen ions. Ions and permanent dipoles can also contribute to the charged state of a protein but to a lesser extent. The contribution by site-bound ions can even be proton linked. Permanent dipoles are electrically neutral and do not normally react with hydrogen ions; in a strict sense they do not really contribute to the charged state of a protein.

The number of protons bound or released during an equilibrium reaction (Δn_{H^+}) can be measured directly, as the difference in the number of protons bound to the different molecular species in equilibrium, or indirectly, with Wyman's linkage relationships (Wyman and Gill, 1990):

$$\frac{d \ln K_{eq}}{d \ln H^+} = \Delta n_{H^+} \tag{2}$$

The pH dependence of the charge of a protein is described by the proton titration curve, as shown in Figure 1. Titration curves reflect the collective titration of the constituent individual ionizable residues. The titration curves of individual residues, shown in Figure 2, represent equilibrium reactions such as $AH \leftrightarrow A^- + H^+$, that can be analyzed in terms of acid dissociation constants (pKa). The charge of a protein (Z) with n ionizable residues can be calculated at any pH from the pKa values of its component ionizable residues according to the following expression:

$$Z = \sum_{i=1}^{n} \alpha_i = \sum_{i=1}^{n} \frac{10^{pH-pK_i}}{1 + 10^{pH-pK_i}} \tag{3}$$

where α_i is the degree of dissociation of an individual proton binding site. $\alpha_i * z_i$ and $(1-\alpha_i) * z_i$ describe respectively the charged state of individual acidic or basic groups of valence z_i.

Titration curves can be measured by potentiometric methods for proteins in the folded and in the unfolded states. Individual site proton titration curves can also be measured experimentally with the methods of nuclear magnetic resonance (NMR). While decomposition of an overall titration curve into its component individual site titrations would be extremely useful for purposes of elucidating the molecular mechanism of proton-linked energetics of proteins, this would require knowledge of the pKa of all titratable residues. This is not possible at present for all but the smallest and simplest proteins.

Computational methods for structure-based calculation of pKa values offer an alternative approach for obtaining proton titration curves of proteins and for dissecting them into contributions by individual sites. These methods are also of great heuristic value. The physical models that are used to calculate pKa values from structure are also the models used to interpret the physical and structural origins of pKa values and titration curves measured experimentally.

B. Determinants of pKa Values

pKa values measured in model systems in solution (His, 6.6; N-term, 7.0; Lys, 10.5; Arg, 12.0; C-term, 3.5; Asp, 4.0; Glu, 4.5 (Matthew et al., 1985)) can be used with Equation 3 to compute hypothetical or ideal titration curves of proteins. Figure 1 shows the titration curve calculated in this way for sperm whale myoglobin. This curve is more similar to the titration curve measured in the denatured state than to the curve of the folded state. This is true for all proteins studied. The difference between the ideal titration curve and the curve of the folded protein reflects the effect of the macromolecular environment on the pKa values of titratable groups.

Charge is created or destroyed when a proton binds to a protein. The effects of the macromolecular environment on pKa values can therefore be described with the language of classical electrostatics. The free energy of proton binding at a site on a protein can be dissected into different contributions according to the following expression:
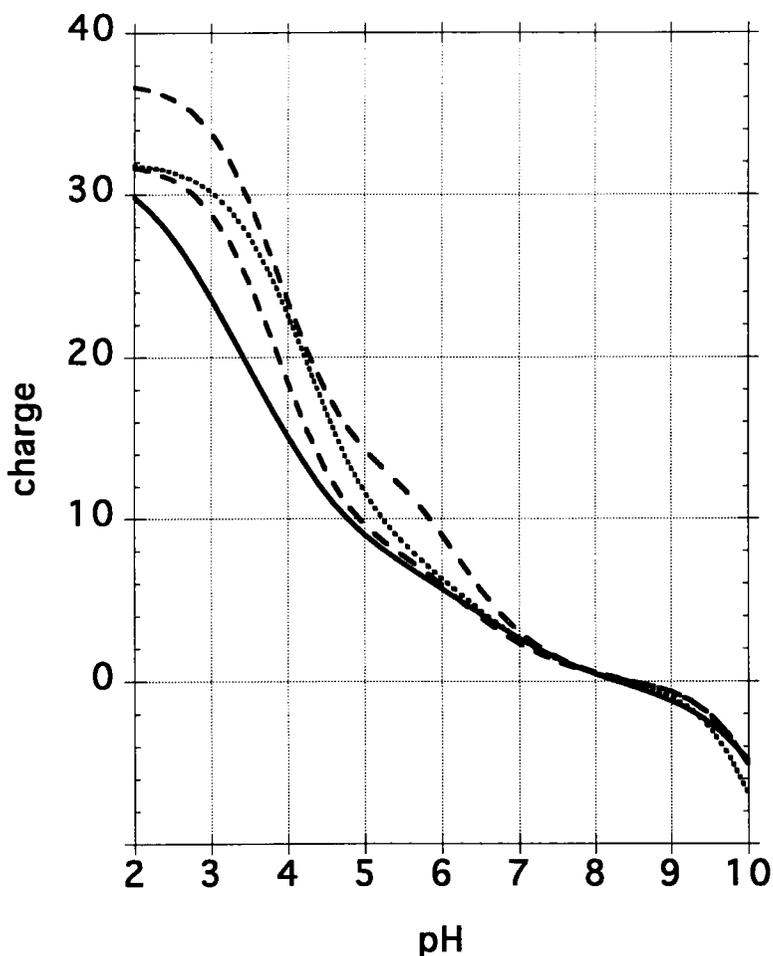
Figure 1. Proton titration curves of sperm whale myoglobin in the folded and unfolded states. These curves were calculated with Equation 3 with pKa values determined from the analysis of the titration curves measured experimentally. The titration curve in the native state (solid line) was determined experimentally in the pH range from 10 to 5. The segment from pH 5 to 2 was calculated with the SA-modified Tanford-Kirkwood algorithm as it is not accessible experimentally; under these conditions myoglobin is not in the native state. The titration curve of the denatured state (dashed line) was measured experimentally. Two different titration curves for the denatured state are shown; the lower one has been corrected so as not to include the charge contributions by the histidines that do not titrate in the folded holoprotein. The dotted curve refers to the ideal titration curve computed with Equation 3 and model system pKa values, as described in the text. All titrations were carried out under conditions of 100 mM KC1, at 298 K.
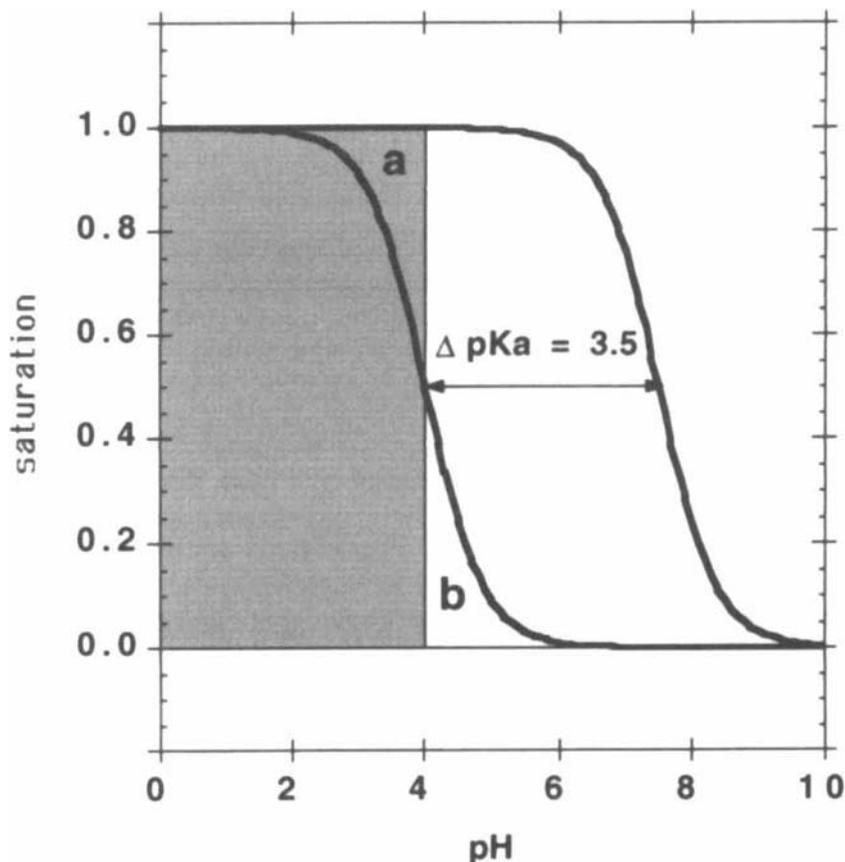
Figure 2. Titration of two residues with pKa of 4.0 and 7.4. The shaded area represents the area under the titration curve of pKa of 4.0 when the areas labeled a and b are identical. This area is equivalent to the pKa. The area between two titration curves corresponds to the difference in pKa between these two titrations.

$$\Delta G_i = \Delta G_{int,i} + \Delta G_{macro} = \Delta G_{int,i} + \Delta G_{coul,i} + \Delta G_{hydr,i} \qquad (4)$$

ΔG_{int} refers to the intrinsic free energy of proton binding described by the pKa values of model compounds. ΔG_{coul} describes the influence of the electrostatic potential emanating from all other charged atoms on the proton binding reaction at site i. ΔG_{hydr} quantitates the energetics of solvation of charged species.

Hydration of Charged Groups

The self energy (ΔG_{ii}) of an ion refers to the energy stored in its electrostatic field. It can be computed (kcal/mol) as the total work of charging a particle of cavity radius r_i:

$$\Delta G_{ii} = \int_0^{z_i} \frac{q \, dq}{r_i A} = 332 \frac{Z_i^2}{2 \, r_i A} \tag{5}$$

The parameter A in this expression is an attenuation parameter that accounts for dielectric effects by solvent and screening of the electrostatic potential counterions.

A charged group that can exist in two different dielectric environments will be stabilized preferentially in the environment with the higher dielectric by an amount equal to the difference in the self energy of the ion in the two environments. The contribution of hydration to the pKa of proton binding sites describes the difference in self energy of charged groups in the unfolded state, when they are presumably fully solvated, and in the folded state, where they are partially dehydrated by their abutment against a medium of low dielectric.

Coulombic Interactions

The work of charging a site against the electrostatic potential due to all other charges in the system (kcal/mol) can be calculated according to:

$$\Delta G_{ij} = \int_0^{z_i} \frac{Z_j dq}{r_{ij} A} = 332 \frac{Z_i Z_j}{r_{ij} A} \tag{6}$$

where r_{ij} refers to the distance between the interacting charges i and j. The charges Z_j that contribute to the electrostatic potential at charge i can include not only the charge of other ionizable residues but also the virtual charges of polar atoms and of bound water molecules and ions.

The coulombic contribution to the pKa values of titratable sites is generally calculated under the assumption that there are no interactions between titratable sites in the denatured state of proteins. In general this is a valid assumption, as demonstrated in some proteins by the agreement between the titration curve measured in the unfolded state and the titration curve computed with Equation 3 and model system pKa values.

Computation of pKa values

Equation 7 was obtained by combining and manipulating Equations 5 and 6. This expression is entirely generic, however, it is useful for interpreting physical origins of pKa values:

$$pK_i = pK_{int,i} - \frac{1}{2.303\ RT}\left(\frac{Z_i}{A_F ri} - \frac{Z_i}{A_U ri} + \sum_{i=j}^{n} \frac{Z_j}{A_F rij}\right) \tag{7}$$

Notice that only one coulombic term appears in this expression, indicating that the assumption that there are no coulombic interactions between groups in the unfolded, denatured state, is operative. A_F and A_U are parameters that describe the dielectric and screening effects in the folded and unfolded states, respectively. The difference of self energy terms in Equation 7 represents a hydration free energy analogous to that defined originally by Born (Bockris and Reddy, 1970).

Equation 7 shows that negative charges (i.e. $Z_j = -1$) raise the pKa of basic and acidic residues while positive ones lower the pKa values. Because A_U is always larger than A_F, the exposure of charges to the medium of low dielectric attendant with folding is always destabilizing. This is manifested by a decrease in the pKa values in basic residues and an increase in acidic residues.

There are other important physical determinants of pKa values that are not included in Equation 7. The effects of hydrogen bonding, for example, are not considered explicitly, although this effect can be approximated by including partial charges of polar atoms in the coulombic term. Electronic effects are also not considered explicitly. The negative charge in the side-chains of Arg, Glu, and Asp, for example, is delocalized and it is unlikely that the pKa of these groups reflects only the energetics of protons binding to a single atom. Instead they probably reflect an average of the energetics of the ensemble of electronic states accessible to this group. Similarly, the side-chain of histidines can exist in different tautomeric structures, thus the observed pKa of histidines can include contributions from multiple tautomeric species. It is also believed that interactions between charged and aromatic groups can affect pKa values due to the direct influence of the electron cloud of the aromatic side-chain on the potential at the proton binding site (Loewenthal et al., 1992).

C. Structure-Based Computation of Electrostatic Properties

Equation 7 delineates the main energetic factors that determine pKa values of titratable residues. Two separate problems need to be solved in order to use these types of models to calculate pKa values from X-ray elucidated structures. One is a problem in classical electrostatics that involves computation of electrostatic potentials in the macromolecular setting (i.e., evaluation of A parameters in Equation 7). The second is a problem in statistical thermodynamics that involves computation of the charged state of a system consisting of multiple, not identical, and interactive proton binding sites.

Calculation of the Electrostatic Potential

Many different approaches have been developed for the computation of electrostatic potentials from structure and they have been reviewed extensively (Matthew

et al., 1985; Sharp and Honig, 1990; references therein). The most useful ones for purposes of quantitating the effects of salt and pH on energetics of proteins avoid microscopic descriptions of the solvent and instead treat the solvent and the macromolecular phases in terms of macroscopic dielectric continua. The aim of these models is to calculate the potential by solving the Poisson-Boltzmann equation, which relates the electrostatic potential, the dielectric of the system, and the ionic strength of the solution (Bockris and Reddy, 1970). Meaningful evaluation of the operative dielectric is the central difficulty handled by these algorithms.

Tanford and Kirkwood developed a model-dependent solution of the linearized form of the Poisson-Boltzmann equation for the idealized case of spherical bodies (Matthew et al., 1985). This model is simply an extension of the original solution by Debye and Hückel. It has proven extremely useful for the prediction of proton titration curves of proteins. Values for the internal and external dielectric constants of the system, the ionic strength, and the approximate spherical size of the macromolecule need to be specified in this model. There is also a parameter that attempts to describe the distance of burial of charged atoms in the cavity of low dielectric. In practice, the most useful form of the Tanford-Kirkwood algorithm place the charges exactly at the interface between the two dielectrics. This limits the lower dielectric any interaction between two charges can assume to one that is the average of the internal and external dielectrics. In addition a parameter equal to the normalized accessible surface area is used to modulate the contribution of every charged atom to the electrostatic potential. Charged atoms that are fully accessible to solvent relative to a model peptide in extended configuration are considered not to contribute to the electrostatic potential as they are presumably fully neutralized by polarization of their environment. Conversely, buried groups that are sequestered from solvent are predicted to have the greatest impact on the magnitude of the electrostatic potential.

Considerably more rigorous numerical methods based on the method of finite differences have been developed recently to compute the electrostatic potential by solving the Poisson-Boltzmann equation (Yang et al., 1992; Oberoi and Allewell, 1993). These numerical methods avoid the spherical approximation of the original Tanford-Kirkwood model and take into consideration explicitly the topographical features of the dielectric boundary.

The electrostatic potentials predicted by the finite difference method are consistently higher in magnitude than those predicted with the Tanford-Kirkwood algorithm. In practice the computation of pKa values and titration curves of proteins with the modified Tanford-Kirkwood algorithm using a form of Equation 7 that does not include the self energy terms (i.e., hydration) gives consistently good agreement with experiment (Matthew et al., 1995). In fact, a calculation in which the potentials in the protein are computed with a constant dielectric of 55, corrected with a generic Debye-Hückel term to account for the effects of ionic strength, also gives a reasonably good approximation to the experimental titration curve if hydration effects are ignored. This highlights the important relationship between

hydration effects and proper handling of the dielectrics of the system, particularly the dielectric of the interior of the protein. Clearly it is desirable to deconvolute the pKa of a titratable site into separate contributions by coulombic and hydration terms in order to understand the molecular mechanisms of pH dependent processes. There is no question that for groups buried in the interior of a protein the value of the pKa is determined primarily by the penalty from dehydration. Models that do not consider hydration energies explicitly will fail to predict correctly the pKa values correctly in these cases. On the other hand, it appears that the pKa values of the majority of charged groups on the surface of a protein are not influenced much by desolvation, either because the loss of one or two waters of hydration is energetically inconsequential—hydration of charged species is cooperative—or maybe because the dynamic character of surface side-chains results in a dielectric effect higher than the one normally associated with the interior of the protein.

Calculation of the Charged State of a Protein

Proton binding is a highly cooperative process. The protonation state of an individual binding site is not independent of the protonation at the other binding sites. Therefore, the calculation of the pKa of a site should, in principle, require examination of a huge number of species corresponding to the different protonation states of the protein. Despite the staggering computing power available at present, the exact statistical treatment of this problem is not possible for all but the smallest of proteins. Several different strategies have been devised to overcome this computational problem. Tanford and Roxby used an iterative algorithm to compute the average charge of a titratable site (Roxby and Tanford, 1971). This mean field approximation is very accurate for all cases except for strongly interacting sites that titrate with similar pKa. These cases can, be treated with an exact statistical mechanical treatment (Bashford and Karplus, 1990). Alternatively, Monte Carlo methods have also been developed to compute the average protonation of a site from an equilibrium distribution of states generated by Monte Carlo sampling (Beroza et al., 1991). Long range interactions of surface charged groups are known to be weak, so the problem of coupled titrations is only significant for ion paired groups and for groups sequestered from solvent. Hybrid methods that apply the mean field approximation to all groups except those that are buried, clustered, or networked avoid the computational burden of the exact statistical treatment (Antosiewicz et al., 1994).

Limitations of Current Approaches

Most of the recent efforts in the development of computational algorithms for quantitation of electrostatic effects from structure have been focused on the techniques for solving the Poisson-Boltzmann equation and for computing the charged state of a protein. The validity of the methodologies that have been

developed has been established by direct comparison of titration properties meas-ured experimentally and titration behavior predicted from structure. However, most treatments developed thus far fail to consider systematically the dynamic nature of surface groups and the effects of site-specific binding of ions and waters.

The conformation of surface side-chains is not as rigid as that of residues buried in the core of a protein. Surface side-chains are usually in motion as evidenced by their large temperature factors measured crystallographically. Some side-chains sometimes even exhibit more than one distinguishable conformation and in extreme cases they are invisible in crystallographic experiments due to complete disorder. In addition, the conformations of some charged side-chains are influenced by intermolecular contacts with symmetry related molecules. Due to the uncertainties in the placement and conformation of surface side-chains the algorithms that are too sensitive to the geometry of the immediate environment of a charged group run the risk of translating and augmenting these uncertainties into incorrect predictions of electrostatic properties. The dynamic character of charged side-chains should be considered explicitly when detailed, molecular interpretation of predicted electrostatic behavior is attempted.

The Poisson-Boltzmann equation describes the screening of electrostatic poten-tials by clouds of counterions that are only loosely associated with proteins. Site-specific interactions between proteins and ions represents another mode of association that is not considered implicitly by Poisson-Boltzmann electrostatics (Matthew and Richards, 1982). Site-specific ion binding is ruled by the law of mass action. Site-binding presumably requires at least partial dehydration of the ligand and of ligating side-chains, therefore the binding characteristic of different ionic species are specific and related to their hydration properties. Binding of ions at specific sites in the proteins is relatively well understood for in the case of interactions between proteins and divalent cations (Ca^{2+}, Mg^{2+}) or transition metals. The weak interactions between proteins and monovalent anions (Cl^-) and cations (Na^+, K^+) are not well understood at present,although their physiological importance is recognized widely. There are only limited ways in which the energetics of these interactions can be studied by experimental or computational methods (Matthew and Richards, 1982; García-Moreno E., 1994). Because most of the site-bound ions are transparent in crystallographic experiments it is first necessary to identify ion binding sites from structure. For anions this can be done by identifying hot spots of positive potential in the electrostatic potentials, but for cations, which are stabilized by coordination in addition to simple electrostatic interactions, this is a more difficult problem that has not yet been solved satisfac-torily. Site-bound ions can be treated as titratable elements of the charge array of a protein. The occupancy of an ion binding site can be computed as a function of salt concentration and of pH with binding constants computed directly from the electrostatic potential under the assumption that binding is driven by electrostatics (García-Moreno E., 1994).

Similarly, it is evident from X-ray crystallographic and NMR experiments that waters can hydrogen bond tightly to polar and charged groups on the surfaces of

proteins (Jeffrey and Saenger, 1991). Algorithms that treat water as a dielectric continuum do not normally consider the energetic contributions by water binding. Handling site-bound waters will probably require development of more realistic and more accurate microscopic models in which water is represented at least partly as an ensemble of discrete molecules.

D. The Problem of the Dielectric Inside a Protein

All macroscopic models of electrostatic interactions in proteins use two different dielectric constants to describe the effects of polar and polarizable atoms on the energy of interaction between charged groups. The dielectric effects in the solvent phase are described in terms of the dielectric constant of water, 78.5 at 298 K. The interior of a protein is usually represented by a low dielectric constant with values in the range from 1 to 4. This range encompasses the dielectrics of many small organic molecules such as ethyl ester, benzene, oleic acid, n-pentane, and some amines. The low dielectric presumably reflects important contributions by electronic polarizability and the absence of a significant contribution by dipolar relaxation.

There is no question that the dielectric of the macromolecular substance is low. Several lines of experimental evidence support this. The presence of uncompensated charge buried inside a protein, for example, is known to be very severely destabilizing. Also, the magnitude of the interactions between charged atoms are stronger in the interior of proteins than at the surface; in fact many proteins have evolved to harness these strong electrostatic interactions to potentiate active sites to make and break covalent bonds. The questions that remain concern the value of the dielectric constant needed to model these effects. Use of a single constant value to treat as heterogeneous an environment as the inside of a protein is unlikely to be a realistic approximation.

The question surrounding the value of the internal dielectric of a protein is critical because the magnitude of the electrostatic effects is inversely proportional to the dielectric (Equation 3). The low dielectric inside a protein plays a much more important role than that of solvent in determining the magnitude of hydration energies. Experimental evidence supports the notion that the internal dielectric is much higher than previously thought: values as high as 12 and higher might be operative (Stites et al., 1991). Use of these high dielectrics results in accurate prediction of the titration behavior of surface groups, but they underestimate the electrostatic effects experienced by buried charged groups (Antosiewicz et al., 1994).

III. EXPERIMENTAL MANIFESTATIONS OF ELECTROSTATIC EFFECTS

It is generally accepted that all aspects of structure and function of proteins are influenced by electrostatics. The marked effects of pH and salts on physical and

physiological properties of proteins will be used to outline the contributions by electrostatics to conformational stability and functional energetics.

A.   pH Dependence of Conformational Stability

Most proteins are stable over a wide range of pH values, suggesting that in general the contributions by electrostatic interactions to the stability of proteins are not significant, or at least not under conditions of neutral pH. However, the conformational stability of many proteins is proton-linked and these proteins denature reversibly in acidic or basic conditions. The exact pH at which denaturation occurs varies from protein to protein and cannot be predicted from structure with any confidence at present. The pH of denaturation is the result of a complex balance of forces; denaturation occurs when stabilizing interactions are lost or destabilizing interactions are accrued that are not compensated for by other stabilizing contributions from hydrophobic, hydrogen bonding, and van der Waals interactions.

The conformational transition between folded and unfolded forms of proteins is proton linked when the pKa values of titratable residues are different in the two states. The difference in pKa values of equivalent groups reflects changes in their chemical environment and, as described by Equation 7, the most likely sources of pKa differences are loss of solvation and establishment of electrostatic contacts concomitant with folding. The standard free energy of proton binding contributed to stabilization by each individual group can be computed directly from the difference in pKa values in the two states ($\Delta G = $ -RT ln K). Notice in Figure 2 that the pKa of a symmetric titration curve corresponds to the area under the titration curve measured from pH = 0 (i.e., 1 M reference state) to pH = ∞. The standard free energy of titration of a single group can therefore be calculated by integration of the titration curve,

$$\Delta G_i^0 = -2.303\,RT \int_{pH=0}^{pH(\upsilon_i=0)} \upsilon_i d\,pH = -2.303\,RT\,pK_i \qquad (8)$$

The contribution by an individual charged group to the stability of the folded structure is proportional to the area between the titration curves measured in the folded and unfolded states (Figure 2). The free energy contributed by all ionizable groups to the stability of a protein can also be measured by integration. It is proportional to the area between proton titration curves of folded and unfolded forms of proteins such as those depicted in Figure 1. As can be surmised from the representative case of myoglobin illustrated in Figure 1, the contribution by the standard free energy of proton binding can be significant in some cases.

According to surveys of crystallographic structures and to the comparison of titration curves of folded and denatured proteins, the net contributions of electrostatic interactions are always stabilizing (Wada and Nakamura, 1981). This might

reflect the active participation of charged atoms during the folding process, which would result in all charges assuming spatial arrangements that maximize their favorable contacts with charges of the opposite polarity. It is also possible that protein sequences have been tuned during evolution to achieve this effect.

Electrostatic interactions between surface charged atoms are usually weak. On average the pKa of titratable residues measured by NMR rarely deviate by more than 0.50 pH units ($\Delta pH = 1$ is equivalent to 1.359 kcal/mol at 298 K) from model compound pKa values. Although the vast majority of titratable side chains are found in electrostatically favorable environments, the majority of them are actually not ion paired. Long range interactions between charged groups separated by distances greater than 5 Å are rarely significant under physiological conditions of ionic strength (Sali et al., 1991). The stabilizing contributions by surface ion pairs range anywhere from 0.0 to -5.0 kcal/mol (Anderson et al., 1990). The exact physical basis of the characteristic energetics of different ion pairs is poorly understood. The loss of solvation concomitant with ion pair formation is thought to counterbalance the stabilizing coulombic interaction (Hendsch and Tidor, 1994). Buried ion pairs represent an extreme case where strong stabilizing coulombic interactions are overcompensated by equally strong but destabilizing desolvation. The net contribution to stability by buried ion pairs is currently thought to be destabilizing.

Although experimental evidence argues convincingly against a major role for ion pairing in the stability of folded proteins, there are interesting cases of structural adaptations where ion pairs are adopted as a way of enhancing the conformational stability of a protein. Proteins of thermophilic organisms, for example, have evolved to withstand the rigors of their environment by enhancing the number of stabilizing ionic interactions in the form of ion pairs or cation binding sites (Jaenicke, 1981). This is a reasonable adaptation mechanism considering that the dielectric of water decreases with increasing temperature, therefore the strength of electrostatic interactions might increase with increasing temperatures. Protein engineers have mimicked this strategy for enhancing stability of recombinant proteins at high temperatures.

B. Other Proton-Linked Phenomena

Protons are probably the most widespread regulators of physiological functions of proteins. Many equilibrium reactions of proteins, such as conformational transitions between quaternary states and assembly of multimeric systems, are regulated physiologically by protons. Perhaps the best known example of a proton-linked transition between quaternary states is the Bohr effect of human hemoglobin. Protons bind preferentially to the T or deoxygenated state of hemoglobin, therefore the equilibrium between the oxygenated (R) and deoxygenated forms of hemoglobin can be regulated by pH. Physiologically this mechanism ensures the adequate supply of oxygen in exercised tissues, where protons are accumulated following an increase in the concentration of lactic acid and the

reaction of water with CO_2 produced by the tricarboxylic acid cycle. The Bohr effect is maximal near pH 7.4. At this pH two protons are bound preferentially to the T state over the R. The modest 3.2 kcal/mol of excess stability due to the binding of Bohr protons is enough to modulate the equilibrium reaction of hemoglobin with oxygen.

The molecular mechanism of the Bohr has been determined through a combination of experimental and computational studies (García-Moreno E, 1994, 1995). It arises due to changes in the pKa of histidines and of amino termini during the transition between the T and R states. While the effect is global in nature, arising from small differences in pKa of groups distributed throughout the molecule, one histidine in particular contributes half of the protons that are released upon binding of oxygen. The source of the pK shift in this histidine appears to be entirely due to electrostatic interactions with another charged group; the histidine is ion paired with a carboxylic acid in the T state and consequently has a high pKa value. The pKa of this histidine in the R state is depressed because the ion pair is not present, and in addition the histidine is found in a crevice, surrounded by other basic groups.

The Bohr effect is not unique to hemoglobin. Proton-linked transitions between quaternary structures are prevalent in multimeric systems. Protons also regulate physiologically the state of assembly of many multimeric systems. The picornaviruses exhibit one of the most exquisite examples of structural adaptation to harness protons as a regulatory signal (Rueckert, 1985). Viruses belonging to different genera in the picornavirus family exhibit grossly different response to changes in the pH of their environment. The rhinovirus, responsible for the common cold, is extremely sensitive to acid; the capsid of this virus is not stable in pH values other than neutral ones. The closely related poliovirus, which shares extensive chemical and structural homology with the rhinovirus and all other picornaviruses, is stable under extremely acidic conditions of pH. This structural adaptation is necessary for the viability of the poliovirus, which enters the body orally and has to survive the extremely acidic environment of the gastroenteric tract. The acid sensitivity of the rhinoviruses appears to be related to their highly contagious nature; it is likely that it plays a role in the breakdown of the capsid in the acidic vacuoles through which they gain access inside a cell. The capsids of polioviruses, which are not normally destabilized by acid, probably break down due to the difference in the concentration of free Cl^- in the intracellular environment (< 10 mM) relative to the extracellular space (180 mM) (Lee and Colter, 1979). Their sensitivity to anion concentration is maximal under the acidic conditions inside the vacuole. The detailed molecular mechanism of this proton and salt-linked electrostatic switch of picornaviruses is not known.

C.   Salt-Linked Phenomena

Macromolecular processes that are guided by electrostatic interactions are usually sensitive to the composition of the ionic environment and are therefore said

to be salt-linked. Determination of the molecular basis of salt effects on the energetics of proteins is not straightforward. The general analysis of the effects of salts on an equilibrium process through Wyman's linkage relationships is entirely analogous to that described earlier for proton-linked reactions (See Equation 2). The number of ions that are bound or released during an equilibrium reaction ($\Delta n_{a^{+/-}}$) can be quantitated from the slope of the equilibrium constant plotted against the concentration of the relevant ionic species:

$$\frac{d \ln Keq}{d \ln a^{+/-}} = \Delta n_{a^{+/-}} \tag{9}$$

However, the interpretation of $\Delta n_{a^{+/-}}$ in terms of ions that are bound or released stoichiometrically during a reaction is not entirely adequate. There are many different modes of association of ions with proteins, and the energetics described by Equation 9 accounts for all possible different modes (García-Moreno E., 1994).

In salt concentrations near physiological values salts interact with proteins primarily through the two mechanisms previously mentioned. One involves the delocalized, aspecific interactions between ions and proteins that are described implicitly by the Poisson-Boltzmann equation. The other entails site specific ion binding ruled by electrostatics and by the law of mass action. The attenuation of electrostatic potentials according to the Poisson-Boltzmann equation is independent of the nature of the salt studied and proportional to the ionic strength of the solution. The magnitude of ionic strength effects on pKa values can be gauged by observing the effect of salt on the proton titration curves of myoglobin shown in Figure 3. Increasing ionic strength shifts the titration curve of a folded protein toward the curve corresponding to the unfolded state, where presumably there are no interactions between charged groups. The effect of salt is to weaken repulsive and attractive electrostatic interactions and consequently to normalize all previously perturbed pKa values.

Due to the paucity of rigorous and systematic experimental studies of the effects of salts on proteins it is extremely difficult to make generalizations about the nature of ionic strength effects on the stability of proteins. Decreased stability due to an increase in salt concentration is usually interpreted as a signature of important contributions by ion pairs. Although the ionic strength of a solution is more effective in attenuating long range electrostatic interactions, attenuation of the less abundant but stronger short range interactions is more significant. In general the pH at which a protein denatures shifts toward the neutral pH range with increasing salt concentration.

The complexity of the mechanisms whereby simple salts can modulate electrostatic interactions in proteins, and the delicate balance between proton-linked and salt-linked phenomena, is illustrated by the effects of salts on proteins under extremely acidic conditions (Goto et al., 1990). In most proteins exposed to

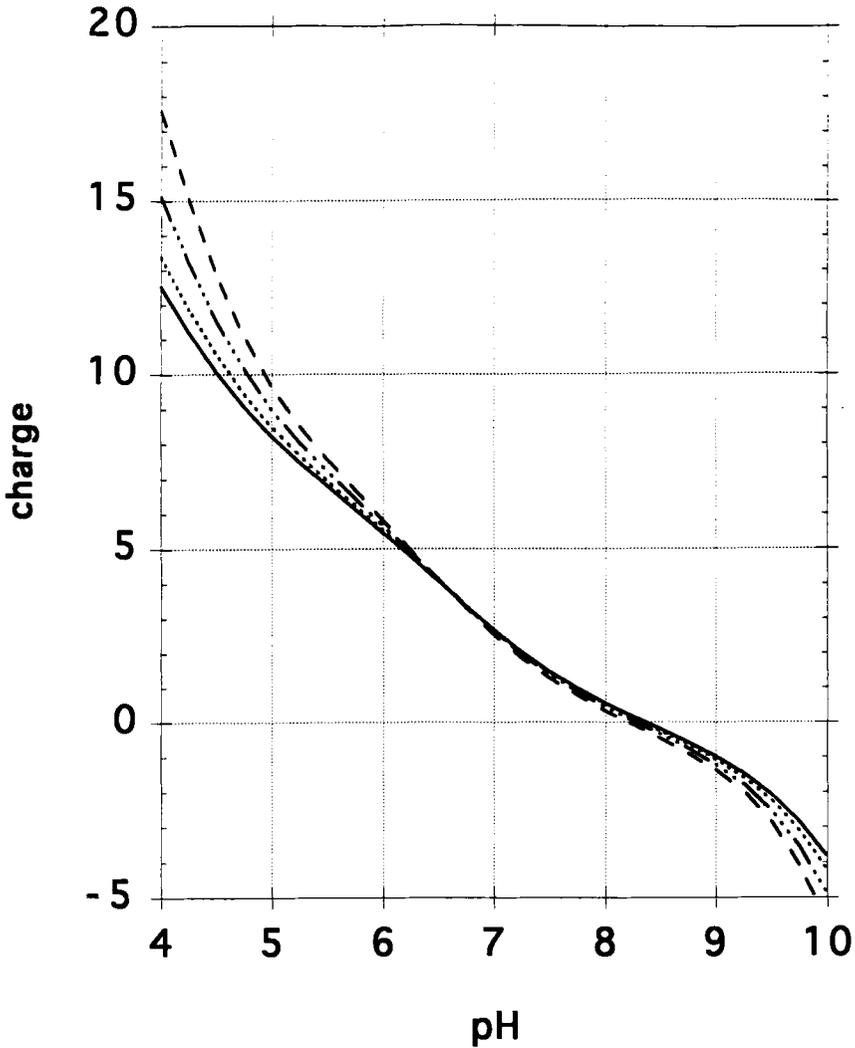*Figure 3.*   Proton titration curves of sperm whale myoglobin in 0.001 (solid line), 0.010 (dotted line), 0.100 (dotted-dashed line) and 1.0 (dashed line) M KC1. All curves were computed with the modified Tanford-Kirkwood algorithm at 298 K (Matthew et al., 1985).

extremely low pH (pH ≤ 2.0) all basic groups are positively charged and the acidic groups are protonated and uncharged. In some globular proteins the massive

build-up of uncompensated destabilizing electrostatic interactions at low pH can be overcome by the addition of salt. Proteins that existed in an extended, unfolded configuration can be partially refolded by the addition of salts at concentrations below 100 mM. The molecular mechanism of salt induced folding involves attenuation of repulsive interactions by ionic strength effects as well as concomitant stabilization of clusters of positively charged side-chains by site-specific anion binding.

Salts regulate physiologically important conformational transitions of proteins; macromolecular assemblies, in particular, are prone to regulation by their ionic environment. The R to T transition of hemoglobin, for example, is salt-linked. Anions preferentially stabilize the T state thereby ensuring the release of oxygen in the presence of elevated concentrations of the anionic effector diphosphoglycerate. Viral capsids are also exquisitely sensitive to the ionic composition of their environment. For example, capsids of some plant virus are stabilized in the extracellular environment by site specific interactions with monovalent and divalent anions (Durham et al., 1977). Once the virus enters the cell the cation binding sites are vacated in response to the low intracellular concentrations of Ca^{2+} and Na^+. The capsids fall apart under these conditions, thereby releasing the genetic material necessary to infect the cell. Capsid proteins operate as sensors and switches, capable of detecting changes in their ionic environment, ready to trigger major conformational transition when conditions for continuation of the viral life cycle are advantageous.

D. Consequences of Burying Charge in Hydrophobic Environments

According to Equation 7, removing charged groups from the solvent and burying them in the interior of a protein is an energetically unfavorable process. pKa shifts of buried groups are generally much greater that those of ion paired groups. Titratable side-chains that are buried are usually buried in the uncharged form, thereby potentiating proteins for denaturation by protons. In sperm whale myoglobin, for example, there are four buried histidines. These histidines are uncharged over the pH range where myoglobin is stable. The unfolding of myoglobin at pH values close to 5 is triggered by the protonation of some of these histidines (see Figure 1). Similarly, tyrosines in most proteins are buried in the uncharged state at neutral pH. They can become charged at pH values near 10 and are probably responsible for the denaturation of most proteins by base.

Burial of charged residues need not be intrinsically destabilizing. Common structural adaptations to stabilize a buried charge include burial of charged groups as ion pairs, surrounding the buried charge with a cage of polar atoms that can neutralize the charge locally, and burying waters bound to the charge atom. Some proteins harness the strong electrostatic potentials that result from burial of charge in a medium of low dielectric, for example, to perform chemistry at active sites or to modulate the redox potentials of metal centers.

The view that charged residues do not contribute greatly to the stabilization of the folded states of proteins diminishes the subtle but important way in which charged groups contribute to the folding process proper. During folding the charged side-chains must be prevented from becoming trapped in the low dielectric core of the nascent protein. Burial of charged groups in the low dielectric region during folding would be energetically prohibitive. The problem is solved by directing charged groups towards the surfaces of the protein to minimizes the energetic penalty from dehydration. Insofar as charged atoms are attracted towards the solvent phase they exert an influence on the folding process, not unlike that of the hydrophobic effect, that contributes to the definition of the inside and the outside of a protein.

E.   Role of Electrostatic Complementarity in Macromolecular Recognition

In the preceding discussion the physiological relevance of electrostatic interactions in proteins has been illustrated, not coincidentally, with examples of multimeric systems such as hemoglobin and icosahedral viruses. It is not surprising that electrostatic interactions play significant roles in the regulation of structure and function of macromolecular assemblies. The magnitude of contributions by electrostatic interactions to the stabilization of interfaces between macromolecules is comparable to the net stabilizing contributions by other noncovalent forces. Change in the charged state of a single residue can involve an energy comparable to the marginal energy of stabilization of an interface between two proteins. Electrostatic interactions across protein-protein interfaces therefore act as switches that poise assemblies to respond to subtle changes in the ionic composition of their milieu with conformational changes of physiological significance.

Interacting macromolecules often display structural features that have evolved to maximize long range electrostatic recognition of other macromolecules. Nucleic acid binding proteins, for example, are usually armed with a battery of basic residues in the face of the protein that recognizes and interacts with nucleic acids. The basic side-chains recognize the negatively charged potential from the phosphate backbone, and stabilize the incipient complex. Many protein-protein complexes are similarly stabilized by strong ionic interactions between groups of opposite polarity.

Not all interactions between charged groups of interacting macromolecules are stabilizing. Apposition of charged groups of the same polarity across protein-protein interfaces is a motif encountered frequently. Such an arrangement can be stabilizing if ions can bind to these regions. The Cl^- binding properties of human hemoglobin, for example, arise from a clustering of eight basic side-chains in a region where β subunits contact one another (García-Moreno, 1995). Cation binding sites in viruses are usually located on the axes where multiple, symmetry-related subunits converge. Clustering of potentially repelling groups at interfaces between proteins can constitute a trigger that can sense the ionic composition of its environment and respond by stabilizing or destabilizing an assembly.

One of the most perplexing features of macromolecular interfaces is the widespread occurrence of charged groups removed from contact with solvent. Often charges at interfaces are buried as a pair with a charge of the opposite polarity. It is not uncommon, however, to find charges that are buried without any obvious mechanisms to compensate for the loss of solvation. There are complex networks and clusters of charged residues at most interfaces, and it is possible that stabilization of buried charges takes place through long range interactions enhanced by the low dielectric that results when water is squeezed from the interface. Alternatively buried ionizable groups are buried in the uncharged state, poising the interface for regulation by protons.

The energetics of charged groups at interfaces have not been studied rigorously by either experimental or computational methods. The manner and the extent to which clusters and networks of charges at interfaces contribute to the transduction of allosteric signals across interfaces is not yet clear, but there is no question that electrostatic interactions at interfaces are of paramount importance for the regulation of stability and function in macromolecular assemblies.

IV. HYDROGEN BONDING

A hydrogen bond is fundamentally an electrostatic interaction between two permanent dipoles. It is the interaction between a donor hydrogen bearing partial positive charge and the electron density of an acceptor atom. The average distance between electronegative atoms is approximately 2.8 Å. All the polar atoms of proteins can participate in hydrogen bonds. The > N - H of the peptide bond can donate a hydrogen to a single bond, while the backbone > C = O can accept two, one for each lone pair of electrons in the sp2 oxygen. All polar side-chains except Trp can form multiple hydrogen bonds. The variability in the geometry of hydrogen bonds observed in X-ray elucidated structures of small molecules and of macromolecules is enormous. These geometries have been catalogued previously and will not be discussed (Baker and Hubbard, 1984; Jeffrey and Saenger, 1991).

A. Distribution of Hydrogen Bonds in Proteins

Our understanding of the role of hydrogen bonds in defining the specific fold of proteins has been shaped primarily by extensive surveys of hydrogen bonds in X-ray elucidated structures of proteins (Baker and Hubbard, 1984; Stickle et al., 1992; McDonald and Thornton, 1994). Rose summarized recently some very distinctive patterns that have emerged from analysis of hydrogen bonds in X-ray elucidated structures (Stickle et al., 1992).

Most of the hydrogen bonds in proteins are local, which means that the donor and acceptor groups are proximal in sequence. Most of these bonds are between backbone atoms, and most (95%) are within elements of secondary structure

(helices, turns, parallel and antiparallel β sheets). Thus, it is not surprising that the total number of hydrogen bonds in a protein is a function of the helix and β sheet content.

Hydrogen bonds between side-chains and backbone atoms are clustered in the capping regions of helices. Helix caps are a structural motif where the first four > N - H donors and for the last four > C = O acceptors in a helix are hydrogen bonded to polar side-chains of residues flanking the helix. Most side-chain to acceptor hydrogen bonds are found at the amino terminal caps, and those between side-chain donor and main-chain acceptors are localized to the carboxyl terminal caps. The capping motif has been found to be widespread in proteins and it is believed to constitute a stereochemical code for stabilization of helices.

Hydrogen bonds between side-chains are also local except in the case of interactions between ionizable residues. Among the hydrogen bonds between side-chains there is a very high incidence of hydrogen bonded interactions between ion pairs.

Approximately 50% of all polar atoms are buried, but only 1.5% of all polar atoms in proteins do not form hydrogen bonds (McDonald and Thornton, 1994). Many donor or acceptor atoms actually participate in two or more hydrogen bonds, forming complex networks. The strength of individual bonds in these networks might be enhanced by cooperativity brought about by enhanced polarizability of permanent dipoles already involved in hydrogen bonding interactions.

B.   Role of Hydrogen Bonds in Folding

Pauling first identified the hydrogen bond as responsible for the structure of macromolecules (Pauling and Corey 1951; Pauling et al., 1951). It is clear that hydrogen bonds contribute significantly to the specificity of structures, but the exact magnitude of their contribution to the stability of proteins has been a debated issue. Pauling's original proposal that hydrogen bonds provide the force responsible for folding was eventually challenged by Kauzmann's proposal that the hydrophobic effect was the central driving force of protein folding. For many years the net contributions by hydrogen bonds to stability were thought to be neutral or even destabilizing. As the stability of a protein represents the energy difference between the unfolded and the native states, it was considered unlikely that the hydrogen bonding of buried polar groups would contribute to stability on the grounds that all of these groups are also hydrogen bonded with water in the unfolded state.

Not all hydrogen bonds are equivalent. Recent experimental measurements of the free energy of stabilization of hydrogen bonds made possible by protein engineering indicate that the strength of a bond can vary between -0.5 and -2.0 kcal/mol for an uncharged hydrogen bond, and can be as high as -5.0 kcal/mol for a charged bond (Fersht et al., 1985). Studies of the binding energy of amide-amide hydrogen bonds in water and in nonpolar solvents also revealed that the energy of a hydrogen bond in water can be as high as -3.0 kcal/mol in water and -5.0 kcal/mol

in tetrachloromethane (Doig and Williams, 1992). These measurements are in general agreement with calorimetric studies of entropies of transfer of various model compounds from the gaseous phase to water that have been interpreted as indicative that the net contribution by hydrogen bonding to the stability of proteins is of a magnitude comparable to that of the hydrophobic effect (Makhatadze and Privalov, 1995). In water the formation of hydrogen bonds is supposedly driven by the release of bound water molecules, whereas in nonpolar solvents there is an additional contribution from the electrostatic interaction between the dipoles.

The stabilizing contributions by hydrogen bonds to folding do not explain per se how buried polar groups are so successful in fulfilling their potential for hydrogen bonding. The prevalence of local hydrogen bonding interactions is thought to be consistent with a folding mechanism in which hydrogen bonds are established within elements of secondary structure before these coalesce to form the native tertiary structure (Stickle et al., 1992; Rose and Wolfenden, 1993; Srinivasan and Rose, 1995). The contributions by hydrogen bonds to the specificity of a folded structure is the result of the necessity to realize all possible hydrogen bonds.

REFERENCES

Anderson, D. E., Becktel, W. J., & Dahlquist, F. W. (1990). pH-Induced denaturation of proteins: A single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4) lysozyme. Biochemistry 29, 2403-2408.

Antosiewicz, J., McCammon, J. A., & Gilson, M. K. (1994). Prediction of pH-dependent properties of proteins. J. Mol. Biol. 238, 415-436.

Baker, E. N., & Hubbard, R. E. (1984). Hydrogen bonding in globular proteins. Prog. Biophys. Mol. Biol. 44, 97-179.

Bashford, D., & Karplus, M. (1990). pKa's of ionizable groups in proteins: Atomic detail from a continuum electrostatic model. Biochemistry 29, 10219-10225.

Beroza, P., Friedkin, D. R., Okamura, M. Y., & Feher, G. (1991). Protonation of interactions residues in a protein by a Monte Carlo methods: Application to lysozyme and the photosynthetic reaction center of rhodobacter sphaeroides. Proc. Natl. Acad. Sci. USA 88, 5804-5808.

Bockris, J. O., & Reddy, A. K. N. (1970). Modern Electrochemistry, Vol. I, Plenum Press, New York.

Doig, A. J., & Williams, D. H. (1992). Binding energy of an amide-amide hydrogen bond in aqueous and nonpolar solvent. J. Am. Chem. Soc. 114, 338-343.

Durham, A. C. H., Hendry, D. A., & Von Wechmar, M B. (1977). Does calcium ion binding control plant virus disassembly? Virology 77, 524-533.

Fersht, A. R., Shi, J.-P., Jack, K.-J., Lowe, D. M., Wilkinson, A. J., Blow, D. M., Brick, P., Carter, P., Waye, M. M. Y., & Winter, G. (1985). Hydrogen bonding and biological specificity analyzed by protein engineering. Nature 314, 235-238.

García-Moreno E., B. (1994). Estimating binding constants for site-specific interactions between monovalent ions and proteins. Methods Enzymol. 240, 645-667.

García-Moreno E., B. (1995). Probing the structural and physical basis of energetics linked to protons and salt. Methods Enzymol. 259, 512-538.

Goto, Y., Takahashi, N., & Fink, A. L. (1990). Mechanism of acid-induced folding in proteins. Biochemistry, 29, 3480-3488.

Hendsch, Z. S., & Tidor, B. (1994). Do salt bridges stabilize proteins? Protein Science 3, 211-226.

Jaenicke, R. (1981). Enzymes under extreme conditions. Ann. Rev. Biophys. Bioeng. 10, 1-67.

Jeffrey, G. A., & Saenger, W. (1991). In: Hydrogen Bonding in Biological Structures. Springer-Verlag, Berlin.

Lee, P. W. K., & Colter, J. S. (1979). Further characterization of mengo subviral particle: A new hypothesis for picornavirus assembly. Virology. 97, 266-274.

Loewenthal, R., Sancho, J., & Fersht, A. R. (1992). Histidine-aromatic interactions in barnase. J. Mol. Biol. 224, 759-770.

Makhatadze, G. I., & Privalov, P. L. (1995). Energetics of protein structure. Adv. Prot. Chem. 47, 307-425.

Matthew, J. B., Gurd, F. R. N., García-Moreno E., B. Flanagan, M. A., March, K. L., & Shire, S. J. (1985). pH-Dependent processes in proteins. CRC Crit. Rev. Bioch. 18, 91-197.

Matthew, J. B., & Richards, F. M. (1982). Anion binding in ribonuclease. Biochemistry 21, 4989-4999.

McDonald, I. K., & Thornton, J. M. (1994). Satisfying hydrogen bonding potentials in proteins. J. Mol. Biol. 238, 777-793.

Oberoi, H., & Allewell, N. M. (1993). Multigrid solution of the nonlinear Poisson-Boltzmann equation and calculation of titration curves. Biophys. J. 65, 48-55.

Pauling, L., & Corey, R. B. (1951). The structure of proteins: Two hydrogen bonded helical configurations of the polypeptide. Proc. Natl. Acad. Sci. USA 37, 729-740.

Pauling, L., Corey, R. B., & Branson, H. R. (1951). The pleated sheed, a new layer configuration of polypeptide chains. Proc. Natl. Acad. Sci. USA 37, 205-211.

Rose, G. D., & Wolfenden, R. (1993). Hydrogen bonding hydrophobicity packing and protein folding. Annul Rev. Biophys. Biomol. Struct. 22, 381-415.

Roxby, R. W., & Tanford, C. (1971). Hydrogen ion titration curve of lysozyme in 6 M guanidine hydrochloride. Biochemistry 10, 3349-3352.

Rueckert, R. R. (1985). Picornaviruses and their Replication. In: Virology (Fields, B. N., ed.) Raven Press, N. Y. 705-738

Sali, D., Bycroft, M., & Fersht, A. R. (1991). Surface electrostatic interactions contribute little to stability of barnase. J. Mol. Biol. 220, 779-788.

Sharp, K. A., & Honig, B. (1990). Electrostatic interactions in macromolecules. Annu. Rev. Biophys. Biophys. Chem. 19, 301-332.

Srinivasan, R., & Rose, G. D. (1995). LINUS: A hierarchic procedure to predict the fold of a protein. Proteins 22, 81-99.

Stickle, D. F., Presta, L. G., Dill, K. A., & Rose, G. D. (1992). Hydrogen bonding in globular proteins. J. Mol. Biol. 226, 1143-1159.

Stites, W. E., Gittis, A. G., Lattman, E. E., & Shortle, D. (1991). In a staphyolococcal nuclease mutant the side-chain of a lysine replacing valine 66 is fully buried in the hydrophobic core. J. Mol. Biol. 21, 7-14.

Wada, A., & Nakamura, H. (1981). Nature of the charge distribution in proteins. Nature, 293, 757-750.

Wyman, J., & Gill, S. J. (1990). In: Binding and Linkage, University Science Books, Mill Valley, CA.

Yang, A.-S., Gunner, M. R., Sampogna, R., Sharp, K., & Honig, B. (1992). On the calculation of pKa's in proteins. Proteins 15, 252-265.

PROTEIN STRUCTURE DETERMINATION BY NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY

Irina M. Russu

I. INTRODUCTION

In recent years, nuclear magnetic resonance (NMR) spectroscopy has emerged as a powerful method for determination of the three-dimensional (3D) structures of proteins in solution. Currently, it is possible to determine, using NMR, structures of proteins in the range of molecular weight from 15 to 25 kDa. The quality of these structures is often comparable to structures derived from X-ray crystallography at 2.0-2.5 Å resolution (Clore and Gronenborn, 1993). Although the structural information provided by NMR is comparable to that derived from X-ray crystallography, the two techniques are, in many ways, complementary. This complementarity results from the physical state of the sample and from the time scales of the two types of measurements (Wüthrich, 1989). X-ray crystallography requires single crystals of the protein for diffraction measurements as well as heavy atom derivatives to solve the phase problem. NMR measurements use proteins in solution state. Thus, for the NMR experiment to be successful, the protein of interest must be soluble to a concentration of at least 1 mM, and the solution must be stable for the duration of data collection i.e., as long as a few days. The time scale of processes amenable to study by NMR extends from picoseconds to seconds, and from seconds to days. Thus, in addition to structure, NMR methods can be used to measure thermodynamic, kinetic, and dynamic parameters, such as pK_a values of individual amino acid residues, hydrogen exchange rates, correlation times, and the amplitudes and rates of internal motions in the protein molecule.

The power of NMR spectroscopy for elucidation of structure and structure-function relationships in proteins became evident shortly after the discovery of the NMR phenomenon. In the 1960s and 1970s, the number and range of applications of NMR to proteins progressively increased, and these studies have provided important answers to biologically relevant questions (Jardetzky and Roberts, 1981). However, direct determination of protein structures in solution state by NMR awaited the introduction of superconducting high-field magnets and the development of one-dimensional (1D) and two-dimensional (2D) Fourier transform NMR spectroscopy (Aue et al., 1976; Ernst et al., 1987). The first complete protein structure determinations using NMR were reported in 1985 for the lac repressor headpiece (Kaptein et al., 1985) and for proteinase inhibitor IA (Williamson et al., 1985). Although of low resolution, these structures have set the stage for a rapid succession of important developments. The introduction of heteronuclear magnetic resonance methods and the extension of the NMR experiments to three and four dimensions allowed structure determinations for larger proteins. In parallel, the accuracy of the NMR-derived protein solution structures has improved due to the newer NMR experiments and advances in computational methods. This chapter presents an overview of the developments in protein structure determination by NMR spectroscopy in the last decade. Emphasis is placed on recent methodologies which promise to extend the size of proteins amenable to NMR investigations to 40 kDa, and beyond.

II. BASIC CONCEPTS

An NMR resonance is normally characterized by two parameters: the chemical shift, δ, which gives the position of the resonance in the spectrum (expressed in ppm) and the linewidth at half-height, $\Delta v_{1/2}$. The latter is related to the transverse relaxation time T_2 of the nucleus of interest by: $\Delta v_{1/2} = 1/\pi T_2$. Fine structures of NMR resonances are observed when the nucleus of interest is scalar coupled with one or more other nuclei. Scalar, or spin-spin coupling occurs through chemical bonds and is characterized by the scalar coupling constant ^nJ (where n stands for the number of chemical bonds separating the two interacting nuclei). Vicinal scalar couplings, ^3J, are of special importance in biological applications due to their dependence on the torsional angle of the three bonds. This dependence is generally expressed by the Karplus equation (Karplus, 1959, 1963):

$$^3J = A.\cos^2\Theta - B.\cos\Theta + C \tag{1}$$

where Θ is the torsion angle, and the coefficients A, B, and C can be calculated or deduced empirically for the torsion angle of interest. Most scalar couplings used in protein NMR spectroscopy involve ^1H, ^{13}C and ^{15}N nuclei. The ^1H-^1H homonuclear couplings are easiest to detect due to the high sensitivity of the proton and its high natural abundance. However, ^1H-^1H coupling constants in proteins are small (≤ 12 Hz). In contrast, many heteronuclear coupling constants involving ^{15}N and ^{13}C are much larger (Table 1) and, as shown below, can be successfully exploited to extend the size of proteins amenable to NMR studies.

Table 1. Selected Heteronuclear Scalar Coupling Constants (Hz) for Backbone Nuclei in Proteins

^1J [$^{13}C\alpha(i)$ - $^1H\alpha(i)$]	140
^1J [$^{13}C\alpha(i)$ - $^{13}CO(i)$]	55
^1J [$^{13}C\alpha(i)$ - $^{13}C\beta(i)$]	30-40
^1J [$^{13}C\alpha(i)$ - $^{15}N(i)$]	9-13
^1J [$^{15}N(i)$ - $^1HN(i)$]	90-100
^1J [$^{13}CO(i-1)$-$^{15}N(i)$]	15
^2J [$^{13}C\alpha(i-1)$-$^{15}N(i)$]	5-10
^2J [$^{13}CO(i)$ - $^1H\alpha(i)$]	4-7
^3J [$^{13}CO(i)$ - $^1H\beta(i)$]	8 (*trans* coupling)
^3J [$^{15}N(i)$ - $^1Hb(i)$]	5 (*trans* coupling)

One phenomenon of paramount importance in the determination of protein structure by NMR is the nuclear Overhauser effect (NOE). The effect can be described simply as an exchange of magnetization between spins which gives rise to variations in the intensities of the corresponding NMR resonances. The process underlying the NOE effect is cross-relaxation. The rate constant for magnetization exchange between two spins through cross-relaxation is directly related to the distance r between them, being proportional to $1/r^6$. The overall amplitude of the NOE depends also on the time allowed for the spins to exchange magnetization (i.e., mixing time). For short mixing times, the observed NOE is proportional to the cross-relaxation rate and is therefore directly related to $1/r^6$.

One of the problems in early NMR investigations of proteins was that of spectral resolution. Even for a protein of small size, the number of proton resonances is large (e.g., ~300 for a protein of 50 amino acids). In the absence of special spectroscopic effects such as paramagnetic or ring-current shifts, these proton resonances occur within a narrow frequency range (i.e., 10 to 15 ppm) and, thus, spectra suffer from severe resonance overlap.

A major breakthrough in NMR spectroscopy occurred with the introduction of 2D NMR methods (Ernst et al., 1987). These methods greatly increase spectral resolution by spreading out the correlations between spins in two frequency dimensions. The basic scheme of a 2D NMR experiment is shown in Figure 1. The experiment consists of: a preparation period, an evolution period of duration t_1, a mixing period, and a detection period of duration t_2. During the evolution period, the spins are labeled according to their chemical shifts. The mixing period yields the correlations between the spins that evolve during t_1 and those that are detected during t_2. The duration of the evolution period is incremented in successive experiments. Fourier transformation with respect to the time variables t_1 and t_2 produces a 2D spectrum that has on one axis, F1, the chemical shifts of the spins that evolved during t_1 and on the other, F2, the chemical shifts of the spins that were detected during t_2. Experimentally, spectral resolution in F2 is determined by digitization during the detection period. Spectral resolution in F1 is determined by the number of times the length of the evolution period is incremented.
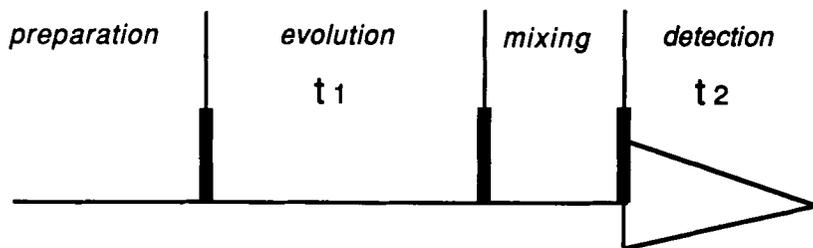

Figure 1.   Basic scheme of a 2D NMR experiment.

The presence of a cross-peak in the 2D spectrum indicates a correlation between the corresponding spins. Depending on the design of the experiment, the correlations observed can be scalar couplings or NOE effects. A scalar coupling between two nuclei generally indicates that they are separated by three chemical bonds at most. Thus, the two nuclei could belong to the same amino acid residue or, if located on the protein backbone, to two neighboring residues. An NOE effect between two spins indicates that the two nuclei are separated, in space, by a short distance, i.e., less than ~6 Å. A few examples of 2D NMR experiments are shown in Figure 2 and in Table 2. Scalar couplings between spins through one or several bonds are observed in COSY (Aue et al., 1976; Nagayama et al., 1980; Bax and Freeman, 1981) and TOCSY (or HOHAHA) experiments (Braunschweiler and Ernst, 1983; Bax and Davis, 1985). Other 2D NMR experiments that are based on scalar coupling correlations and have been used in protein studies are DQF-COSY (Piantini et al., 1982), MQF-COSY (Müller et al., 1986; Rance & Wright, 1986), SECSY (Nagayama et al., 1979), RELAY (Bolton and Bodenhausen, 1982; Eich et al., 1982; King and Wright, 1983; Wagner, 1983; Bax and Drobny, 1985) and DQ-Spectroscopy (Bax and Freeman, 1981; Braunschweiler et al., 1983; Mareci and Freeman, 1983). Through-space NOE connectivities are generally observed in NOESY experiments (Macura and Ernst, 1980). The pulse sequence for a NOESY experiment is shown in Figure 2. An alternative to the NOESY experiment, the ROESY experiment, is performed in the rotating frame (Bothner-By et al., 1984).

Over the last decade, proton 2D NMR experiments have been of great value for studies of small proteins of up to 100 amino acid residues (for a review, see Clore and Gronenborn, 1989). However, even the enhanced spectral resolution provided by 2D NMR proved to be insufficient for larger proteins. The number of protons increases roughly linearly with the size of the protein and, for proteins containing more than ~100 amino acids, 2D NMR spectra are affected by severe overlap of the cross-peaks. Furthermore, the linewidth of proton resonances increases approximately linearly with the protein size. Thus, for larger proteins, the linewidths are comparable to or larger than proton-proton scalar couplings. This fact compromises the observation by 2D NMR of proton coupling correlations which are essential for resonance assignments.

The problems associated with 2D NMR on larger proteins have been addressed by two distinct, yet complementary strategies: heteronuclear NMR and higher dimensionality of NMR spectra. Heteronuclear NMR uses scalar couplings between ^{1}H and heteronuclei such as ^{15}N or ^{13}C. Since many of these couplings greatly exceed ^{1}H-^{1}H couplings (Table 1), the problems created by broader proton resonances in larger proteins can be alleviated, at least in part. Heteronuclear scalar couplings can be used to edit ^{1}H-^{1}H NOESY experiments. An example of an edited NOESY experiment is shown in Figure 2 (Otting et al., 1986). The set of pulses following the mixing time τ_m act as a half-filter; they select for observation (in F2) only the protons that are directly attached to the heteronucleus X. The scalar couplings between protons and heteronuclei can also be used to correlate the
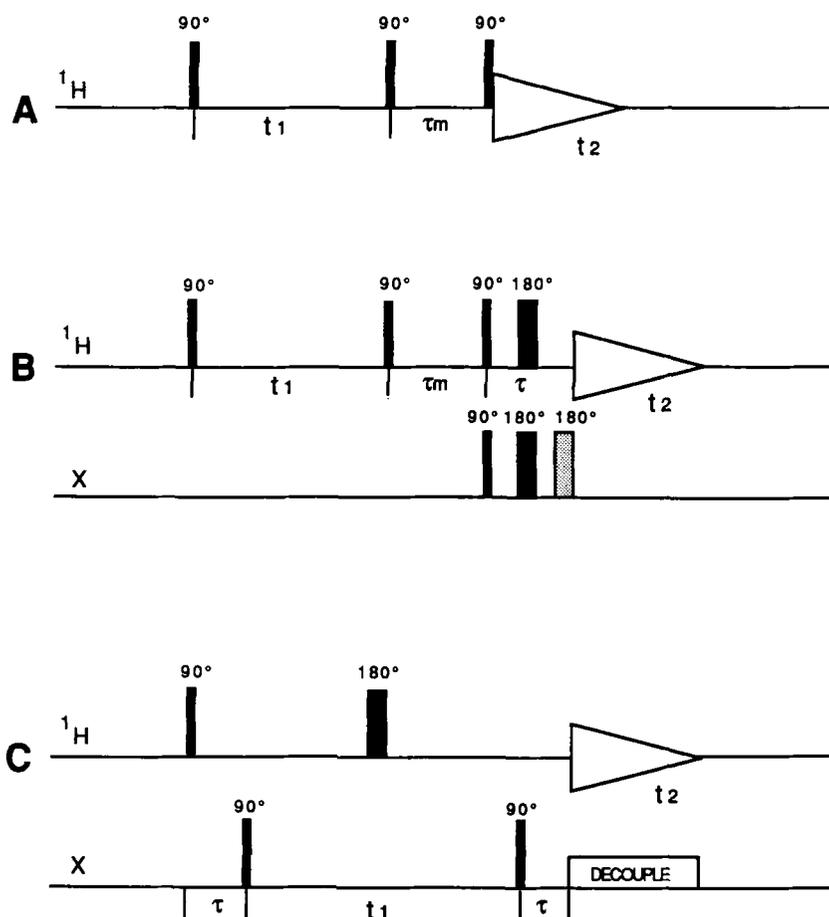
Figure 2. Examples of 2D NMR experiments. (A) ^{1}H-^{1}H NOESY; τ_m is the mixing time (Macura and Ernst, 1980); (B) ^{1}H-^{1}H NOESY with X-nucleus (F2) half-filter. The last 180° pulse on X (represented in a dotted pattern) is applied in every second transient. The delay τ is tuned to J[X-^{1}H]/2 (Otting et al., 1986). (C) One of the original schemes for the HMQC experiment. The delay τ is tuned to J[X-^{1}H]/2 (Bax et al., 1983).

frequency of a proton resonance with that of its directly attached heteronucleus. An example is the HMQC experiment shown in Figure 2 (Bax et al., 1983); resonances of the heteronucleus X appear in F1 and those of protons attached to X appear in F2.

The dimensionality of the NMR spectra can be extended to three and four dimensions to remove resonance overlap. 3D or four-dimensional (4D) NMR

Table 2. Selected 2D NMR Experiments

Abbreviated Name	Full Name
COSY	Correlation spectroscopy
DQF-COSY	Double-quantum filtered correlation spectroscopy
MQF-COSY	Multiple-quantum filtered correlation spectroscopy
E.COSY	Exclusive correlation spectroscopy
RELAY	Relayed coherence transfer spectroscopy
DQ-Spectroscopy	Double-quantum spectroscopy
MQ-Spectroscopy	Multiple-quantum spectroscopy
SECSY	Spin-echo correlated spectroscopy
TOCSY	Total correlated spectroscopy
or	or
HOHAHA	Homonuclear Hartman-Hahn spectroscopy
HMBC	Heteronuclear multiple-bond correlation spectroscopy
HMQC	Heteronuclear multiple quantum coherence spectroscopy
HSQC	Heteronuclear single quantum coherence spectroscopy
HSMQC	Heteronuclear single-multiple quantum coherence spectroscopy
NOESY	Nuclear Overhauser enhancement spectroscopy
ROESY	Rotating-frame nuclear Overhauser enhancement spectroscopy

experiments are normally constructed from blocks of 2D experiments (Griesinger et al., 1987, 1989; Oschkinat et al., 1988; Kay et al., 1990b). As shown in Figure 3, a 3D experiment is obtained by "condensing" two 2D experiments (labeled a and b in Figure 3). The detection period of the first experiment and the preparation period of the second are eliminated. In addition to the detection period $D(t_3)$, the new experiment has two evolution periods, $E(t_1)$ and $E(t_2)$, which are incremented independently. Fourier transformation with respect to the time variables t_1, t_2 and t_3 yields an NMR spectrum with three frequency domains, F1, F2, and F3. The addition of a third 2D experiment (labeled c in Figure 3) to the 3D pulse scheme results in a 4D experiment with detection period $D(t_4)$ and three evolution periods $E(t_1)$, $E(t_2)$ and $E(t_3)$. 3D and 4D experiments are generally named based on the parent pulse schemes. For example, 3D-TOCSY-NOESY indicates a 3D homonuclear experiment obtained from a 2D TOCSY and a 2D NOESY. In double- and triple-resonance experiments, the participating nuclei are often indicated. For example, 3D ^{1}H-^{15}N-NOESY-HMQC indicates a NOESY experiment in which ^{1}H-^{1}H NOEs are separated into the third dimension by the chemical shifts of ^{15}N nuclei. Similarly, 4D ^{1}H-^{13}C-HMQC-NOESY-^{1}H-^{15}N-HMQC indicates a NOESY experiment in which ^{1}H-^{1}H NOEs are separated into the third and fourth dimensions by ^{15}N and ^{13}C chemical shifts, respectively.

The resolving power of heteronuclear 3- and 4D experiments is illustrated by a simple example in Figure 4. In the ^{1}H-^{1}H 2D NOESY spectrum (plane to the left in the figure), three cross-peaks are assumed from three aliphatic protons (x, y, and

3D

Pa-Ea-(t1)-Ma-Da(t2)   Pb-Eb-(t1)-Mb-Db(t2)

Pa-Ea-(t1)-Ma-Eb(t2)-Mb-Db(t3)

4D

Pa-Ea-(t1)-Ma-Eb(t2)-Mb-Db(t3)   Pc-Ec-(t1)-Mc-Dc(t2)

Pa-Ea-(t1)-Ma-Eb(t2)-Mb-Ec(t3)-Mc-Dc(t4)
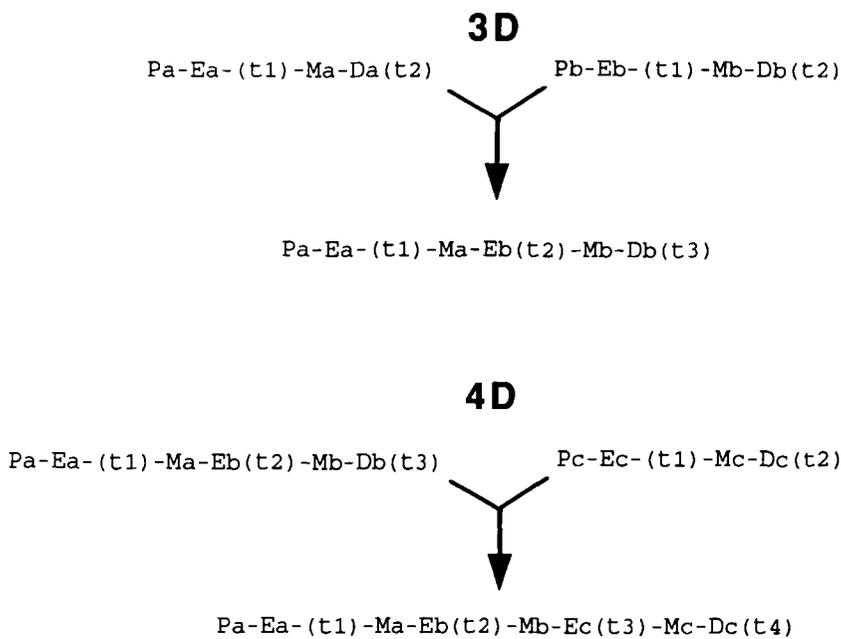
Figure 3.   Relationship between the basic schemes used in 2D, 3D, and 4D NMR. Abbreviations: P, preparation; E, evolution; M, mixing; and D, detection. The parent 2D NMR experiments are indicated by the letters a, b, and c.

z) in F1 to two ^{1}HN protons in F2. It is further assumed that the resonances of the two ^{1}HN protons occur at the same F2 frequency. In the 3D experiment, the ^{1}H-^{1}H NOEs are separated according to the ^{15}N chemical shift of the nitrogen in each NH group. In the example shown, one ^{1}HN proton is found to have NOE connectivities to aliphatic protons x and z, and the other to aliphatic proton y. For carbon-attached protons, further spectral resolution is achieved by extending the experiment into the fourth dimension. In the example shown in Figure 4, the plane of the 3D experiment containing the cross-peaks of protons x and z becomes a cube. The two protons are now edited by the chemical shift of the ^{13}C atom attached to each proton. The total number of cross-peaks in the 3D or the 4D heteronuclear NMR experiment is the same as that in the parent 2D experiments. However, the spectral resolution is greatly increased by observation of sub-sets of cross-peaks at separated frequencies in the third and fourth dimension.

In multi-dimensional NMR experiments, the nuclei that evolve during the evolution periods can be protons or other nuclei. In most instances, the nuclei detected during $D(t_3)$ or $D(t_4)$ in 3D or 4D experiments are protons. Proton detection

2D   **3D**   **4D**

Figure 4. A simple illustration of the resolving power of 3D and 4D heteronuclear NMR spectroscopy. The plane to the left repre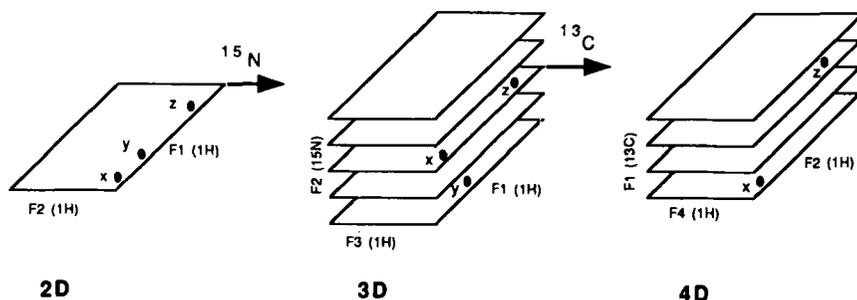sents a simplified ^{1}H-^{1}H 2D NOESY spectrum in which three NOE cross-peaks (filled circles) are assumed. The three aliphatic protons giving rise to these cross-peaks are arbitrarily labeled x, y, and z. In the ^{15}N-edited 3D experiment, the ^{1}H-^{1}H NOE cross-peaks are separated according to ^{15}N chemical shifts. In the ^{15}N-/^{13}C-edited 4D experiment, the plane of the 3D experiment containing the cross-peaks of protons x and z is further separated according to ^{13}C chemical shifts (modified after Clore and Gronenborn, 1993).

increases the sensitivity of observation as compared to an experiment that relies upon the observation of the heteronucleus. The gain in sensitivity is $n.(\gamma_H/\gamma_X)3/2$ where γ_H and γ_X are the gyromagnetic ratios of ^{1}H and heteronucleus X, respectively, and n is the number of protons attached to the X-nucleus. For example, for a methylene group (in which the X-nucleus is ^{13}C), the gain is 16, and for an amide group (in which the X-nucleus is ^{15}N), the gain is 31.

Heteronuclear NMR experiments require protein samples that are uniformly labeled with ^{15}N and/or ^{13}C, or samples in which only one or a few amino acid types are labeled (i.e., selective or semiselective labeling). Isotopic labeling of proteins has been made possible by the development of efficient protein-expressing systems. Uniform isotopic labeling can be achieved by growing the cells in minimal medium supplemented by ^{15}NH$_4$Cl (or K^{15}NO$_3$) and/or [^{13}C$_6$]-glucose as sole nitrogen and/or carbon sources. Selective labeling makes use of auxotrophic strains for the amino acids of interest. For detailed presentations of the procedures for isotopic labeling of proteins, the reader is refered to Stockman et al. (1988), Muchmore et al. (1989), and McIntosh and Dahlquist (1990).

III. RESONANCE ASSIGNMENTS

Assignment of resonances in NMR spectra to specific nuclei in the protein molecule represents an essential step for structure determination and for any studies of structure-function relationships. Most important are the assignments of proton

resonances, since proton-proton NOEs provide the distance information used in protein structure determination (section VB).

For smaller proteins, assignments can be obtained by homonuclear ^{1}H-^{1}H 2D experiments. This conventional strategy, developed by Wüthrich and co-workers (Wüthrich, 1986), involves the following steps: spin system identification, intra-residue connectivities to backbone ^{1}HN protons, and sequential assignments. In the first step, one seeks to identify the proton spin systems of individual amino acid side-chains. The underlying principle is that classes of amino acids give specific patterns in ^{1}H-^{1}H 2D correlation experiments. The patterns of some amino acids (namely Gly, Thr, Ile, Val, and Leu) are unique. They are illustrated in Figure 5
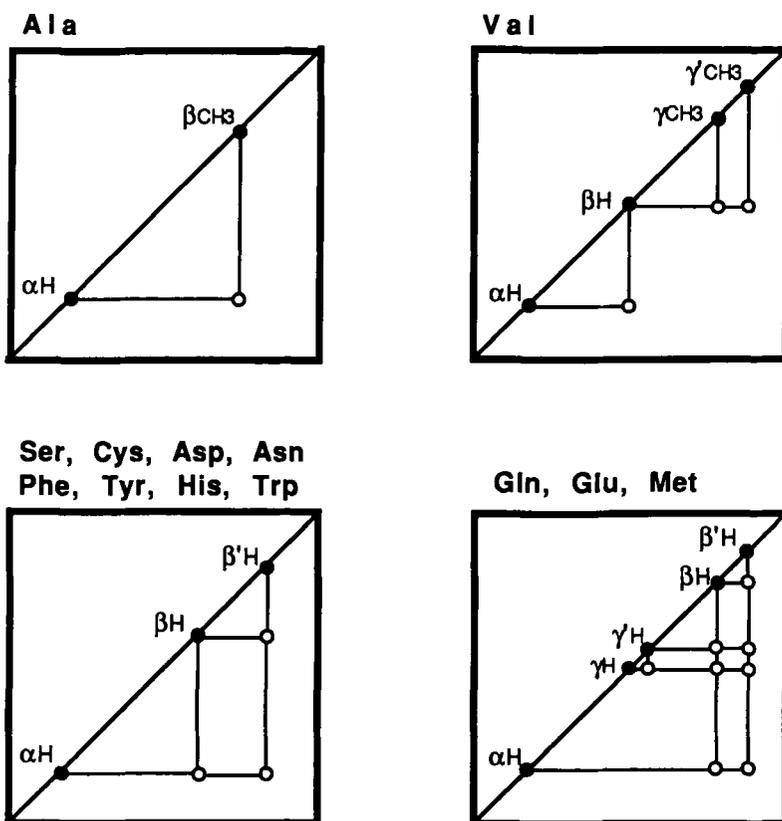

Figure 5. Patterns observed in 2D ^{1}H-^{1}H COSY spectra for selected amino acid side-chains. Closed circles: diagonal peaks. Open circles: COSY cross-peaks. The chemical shift scales are arbitrary.

for Ala and Val: Ala gives a single COSY cross-peak; Val has two γCH3-^{1}Hβ cross-peaks in addition to the ^{1}Hα-^{1}Hβ cross-peak. Amino acid residues such as Ser, Cys, Asp, Asn, Phe, Tyr, His, and Trp produce three COSY cross-peaks, whereas Gln, Glu, and Met have the characteristic pattern of 5-spin systems. In total, 10 different COSY connectivity patterns are observed for aliphatic protons, and four are observed for aromatic rings. For a complete presentation of these patterns, the reader is refered to Wüthrich (1986). The experiments for spin system identification are generally carried out in D_2O, after all labile protons have been exchanged with deuterium. For smaller proteins, COSY, DQF-COSY, and TOCSY experiments have been sufficient. As the size of the protein increases, severe overlap of cross-peaks in COSY spectra obscures the scalar connectivities for each side-chain. One solution to this problem is the integrated approach for spin system identification developed by Wright and co-workers (Chazin et al., 1988). In this method, side-chain magnetization is transferred to and from backbone protons (which are generally better resolved) using RELAY-COSY, DOUBLE-RELAY COSY, and TOCSY experiments. Further identification of amino acid types can be obtained by the analysis of NOESY cross-peaks between side chain and Cβ protons (Wagner, 1990).

The second step in the assignment strategy aims at establishing through-bond connectivities between backbone ^{1}HN and other protons within the same amino acid residue. ^{1}HN-^{1}Hα connectivities are observed in the "fingerprint" region of 2D correlation spectra (F1 $\simeq$ 1.5-6.0 ppm and F2 $\simeq$ 6.5-10.5 ppm) with each amino acid giving a single cross-peak. The exceptions are the Glycine residues which may give two cross-peaks, and N-terminal and Proline residues for which the cross-peaks are absent (Wüthrich, 1986). Due to the labile nature of the amide proton, the experiments must be carried out in H_2O using suitable water suppression techniques. For a detailed presentation of NMR techniques used to suppress the water resonance, the reader is refered to Gueron et al. (1991).

Sequential assignments are obtained by establishing connectivities between protons in successive amino acid residues using NOESY spectra in H_2O. The experiments rely upon NOE effects between ^{1}HN proton in one amino acid residue and ^{1}HN, ^{1}Hα, and ^{1}Hβ protons in the preceding residue (Figure 6). These connectivities are commonly abbreviated as d_{NN} for ^{1}HN(i)-^{1}HN(i+1), $d_{\alpha N}$ for ^{1}Hα(i)-^{1}HN(i+1), and $d_{\beta N}$ for ^{1}Hβ(i)-^{1}HN(i+1). The most useful ways to analyze these connectivities combine NOESY and COSY spectra. This procedure is illustrated in Figure 7 for $d_{\alpha N}$ connectivities involving a segment of six amino acid residues in hen egg white lysozyme (Redfield and Dobson, 1988). To obtain sequential assignments, one starts, for example, with the COSY cross-peak at 8.27 and 5.14 ppm which corresponds to the through-bond connectivity ^{1}HN-^{1}Hα in Thr43. Following the vertical and horizontal lines from this cross-peak into the NOESY spectrum, one finds the sequential connectivity between ^{1}Hα of Thr43 (5.14 ppm) and ^{1}HN of Asn44 (8.15 ppm). The latter proton also has a cross-peak in the COSY spectrum at F1=8.15 and F2=5.02 ppm which assigns the resonance

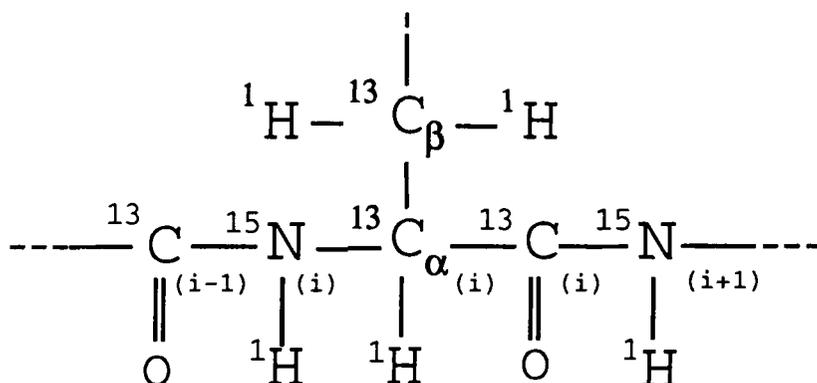$$^{1}H - ^{13}C_{\beta} - ^{1}H$$

$$----^{13}\underset{\underset{O}{\overset{\|}{(i-1)}}}{C} - ^{15}\underset{\underset{^{1}H}{|}{(i)}}{N} - ^{13}\underset{\underset{^{1}H}{|}}{C}_{\alpha}{(i)} \underset{\underset{O}{\overset{\|}{(i)}}}{C} - ^{15}\underset{\underset{^{1}H}{|}{(i+1)}}{N}----$$

Figure 6. Representation of the ^{1}H, ^{15}N, and ^{13}C nuclei along the protein backbone.

at 5.02 ppm to the $^{1}H\alpha$ of Asn44. This cycle of assignment is repeated for the subsequent amino acids.

Sequential assignments yield segments of the polypeptide chain for which the amino acid residue in each position is known, based on the identification of spin systems obtained in the first step of the assignment strategy. The peptide segments that are sufficiently long to be unique in the primary structure are then matched against the known amino acid sequence of the protein. The results are summarized in a diagram such as the one shown in Figure 10.

Proton resonance assignments have been obtained by homonuclear 2D NMR methods in proteins of molecular weights of up to ~15 kDa. A few representative examples are eglin c (Hyberts and Wagner, 1990), hen egg white lysozyme (Redfield and Dobson, 1988), ribonuclease A (Robertson et al., 1989), French bean plastocyanin (Chazin and Wright, 1988), and bacteriophage P22 Arc repressor (Breg et al., 1989). Homonuclear 2D NMR methods can be extended into the third dimension to alleviate spectral overlap. Several selective and nonselective homonuclear 3D experiments have been applied to resonance assignments in proteins, e.g., NOESY-HOHAHA (or NOESY-TOCSY; Oschkinat et al., 1988; Vuister et al., 1988) and HOHAHA-NOESY (Oschkinat et al., 1989). Although they yield enhanced spectral resolution, these methods have been of limited use due to their lower sensitivity. As explained in section II, the linewidths of proton resonances in larger proteins are comparable to or greater than proton-proton couplings. As a result, magnetization transfers through proton-proton couplings are inefficient, and the sensitivity of proton multidimensional experiments is decreased. In contrast, heteronuclear experiments have higher sensitivity since many heteronuclear coupling constants are larger than the linewidths of proton resonances. Heteronuclear NMR methods on isotopically-labeled proteins also have higher spectral resolution due to the larger chemical shift dispersion of ^{15}N and ^{13}C nuclei as compared to
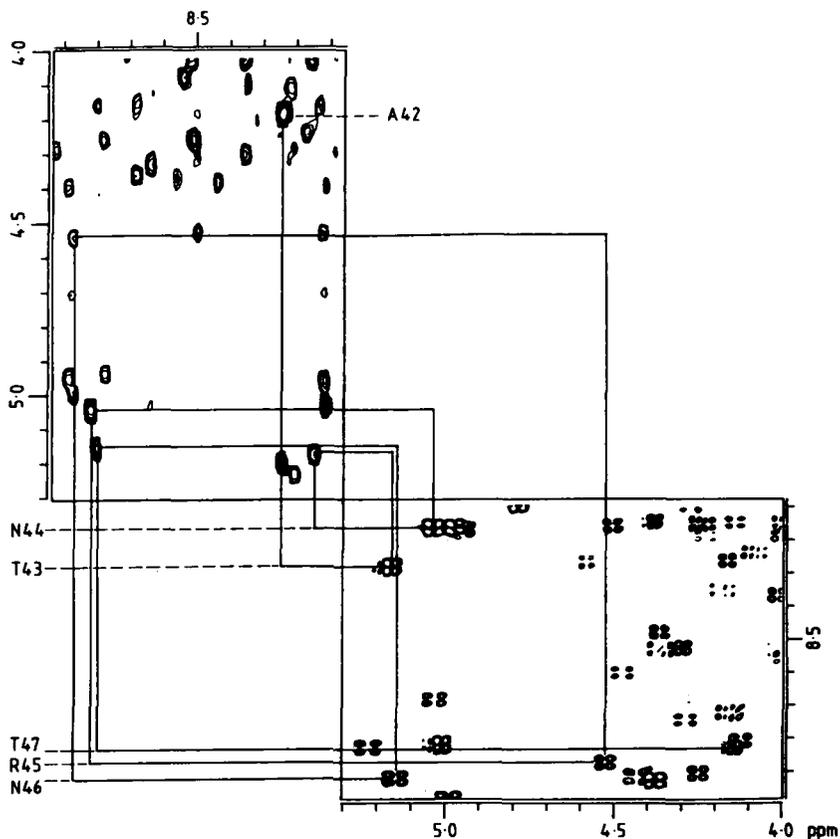
Figure 7. Selected regions of the ^{1}H-^{1}H NOESY (upper left) and ^{1}H-^{1}H COSY (lower right) spectra of hen egg white lysozyme. The sequential assignments of residues 42-47 via $d_{\alpha N}$ connectivities is shown (reproduced from Redfield and Dobson, 1988, with permission).

that of ^{1}H. The experiments provide, in addition to ^{1}H resonance assignments, assignments of heteronuclei such as ^{15}N and ^{13}C. Knowledge of these assignments is useful for obtaining additional conformational constraints based on coupling constants (section VA). Moreover, measurements of the relaxation rates of such heteronuclei provide important information on protein dynamics (Wagner et al., 1993).

In earlier investigations, isotopic labeling has been used to edit 1D and 2D spectra for resonances of protons directly attached to heteronuclei. These techniques have been pioneered by Redfield and co-workers (Griffey and Redfield,

1987). An example of the pulse sequences used is the edited 2D ^{1}H-^{1}H NOESY experiment shown in Figure 2. Isotopic labeling also allows direct observation of scalar couplings between protons and heteronuclei. The HMQC pulse sequence shown in Figure 2 is an example of these experiments. For uniformly ^{15}N-labeled proteins, ^{1}H-^{15}N HMQC spectra contain a ^{1}H-^{15}N cross-peak for every backbone amide group. These heteronuclear fingerprint spectra are significantly superior to ^{1}HN-^{1}Hα 2D correlation spectra due to the fact that ^{1}J(^{1}HN-^{15}N) scalar couplings are uniform (Table 1) and the spectral dispersion of ^{15}N chemical shifts is larger. The resolution can be further improved by using HSQC or HSMQC experiments (Bax et al., 1990b; McIntosh et al., 1990; Fairbrother et al., 1991). An example of the ^{1}H-^{15}N fingerprint spectrum of a protein is shown in Figure 8.

In earlier investigations, assignments of ^{1}H-^{15}N cross-peaks to individual residues have been obtained by comparing the spectrum of the uniformly-labeled protein to those of protein samples in which one or a few amino acid types were ^{15}N-labeled (McIntosh et al., 1990). One should note that this selective (or semi-selective) labeling also provides a direct way to identify resonances by amino acid type. Thus, the problem of analyzing scalar coupling patterns (Figure 5) in crowded ^{1}H-^{1}H correlation spectra is circumvented. 2D NMR methods on ^{15}N- or ^{13}C-labeled proteins have permitted resonance assignments for proteins of larger size such as bacteriophage T4 lysozyme (molecular weight 18.7 kDa; McIntosh et al., 1990) and staphylococcal nuclease (molecular weight 18 kDa; Torchia et al., 1989).

The first concerted 2D NMR strategy to assign ^{1}H, ^{13}C and ^{15}N resonances in proteins has been proposed by Markley and co-workers (Markley, 1989; Stockman et al., 1989). The first step in this strategy consists of measurements of one-bond ^{13}C-^{13}C correlations using homonuclear ^{13}C-^{13}C DQ-spectroscopy. One rationale for this choice is that the patterns of ^{13}C-^{13}C correlations are unique for all amino acid types (except Gln/Glu and Asp/Asn which are indistinguishable). Thus, amino acid types can be identified (Markley, 1989). Sequential connectivities are obtained by ^{13}C-^{15}N DQ-spectroscopy. ^{15}N and ^{13}C assignments are then transferred to protons, based on heteronuclear multiple-bond ^{1}H-^{15}N and one- and multiple-bond ^{1}H-^{13}C correlations. This strategy has been used for ^{1}H, ^{13}C, and ^{15}N resonance assignments in flavodoxin from *Anabaena* 7120 and cytochrome C-553 (Stockman et al., 1989).

The newest assignment strategy in protein NMR is based on 3D heteronuclear experiments on proteins uniformly labeled with ^{15}N and ^{13}C (Ikura et al., 1990; Kay et al., 1990a; Clore and Gronenborn, 1991a). The assignments usually start with the ^{1}H, ^{15}N, and ^{13}C nuclei along the polypeptide backbone (Figure 6). The goal is to correlate ^{1}HN, ^{15}N, ^{13}CO, ^{13}Cα and ^{1}Hα resonances of the same amino acid residue, and to sequentially relate them to the same resonances in the preceding and/or subsequent residue in the primary sequence. Some of the experiments employed in this assignment strategy are summarized in Table 3.

The HNCA experiment correlates ^{15}N and ^{1}HN resonances of a given amino acid residue with the ^{13}Cα resonances of the same and of the preceding residue
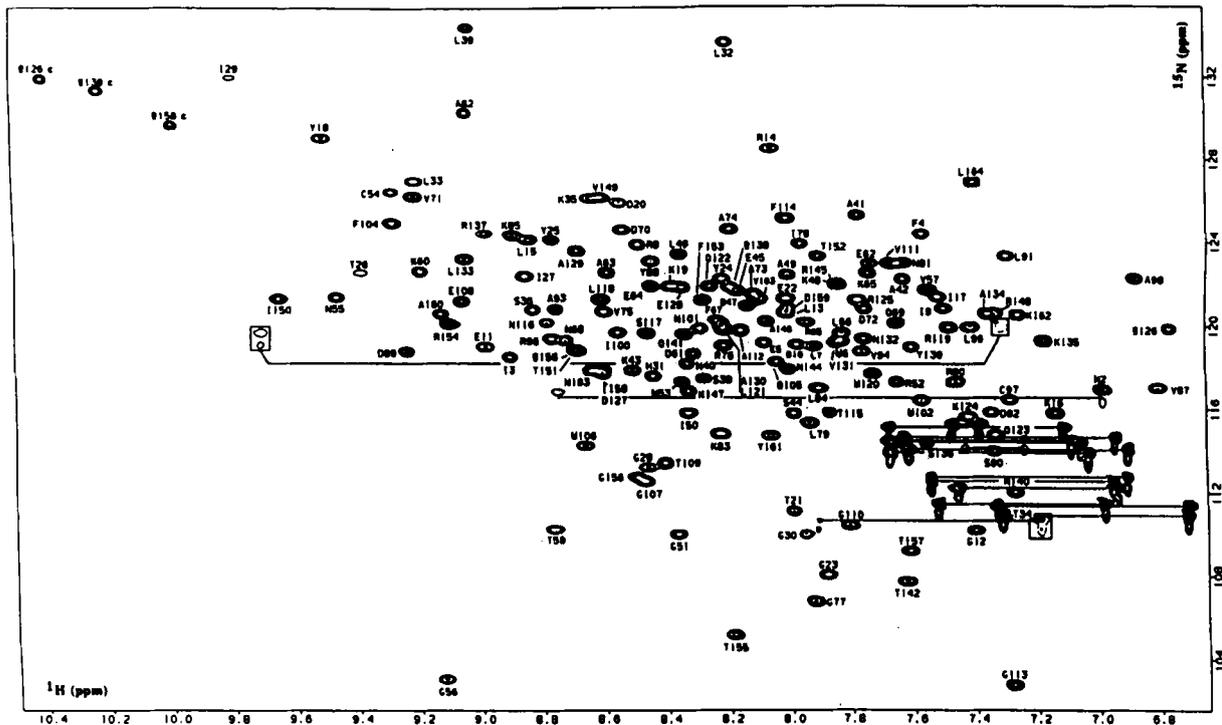
Figure 8. The amide region of the ^{1}H-^{15}N HSMQC spectrum of uniformly ^{15}N-labeled T4 lysozyme. The assignments of all 160 backbone amide and 3 tryptophan indole ^{15}Nε-^{1}H cross-peaks are indicated. The horizontal lines connect ^{15}N-^{1}H cross-peaks in amino groups of Asparagine and Glutamine side-chains (reproduced from McIntosh et al., 1990, with permission).

Table 3.   Selected 3D Heteronuclear NMR Experiments

Name	Correlation Observed	Reference[*]
HNCA	$^{13}C\alpha(i)$-$^{15}N(i)$-$^{1}HN(i)$ $^{13}C\alpha(i-1)$-$^{15}N(i)$-$^{1}HN(i)$	Kay et al., 1990a
HNCO	$^{13}CO(i-1)$-$^{15}N(i)$-$^{1}HN(i)$	Kay et al., 1990a
HCACO	$^{1}H\alpha(i)$-$^{13}C\alpha(i)$-$^{13}CO(i)$	Kay et al., 1990a
HN(CO)CA	$^{13}C\alpha(i-1)$-$^{15}N(i)$-$^{1}HN(i)$	Bax and Ikura, 1991
HN(CA)CO	$^{13}CO(i)$-$^{15}N(i)$-$^{1}HN(i)$ $^{13}CO(i-1)$-$^{15}N(i)$-$^{1}HN(i)$	Clubb et al., 1992
HCA(CO)N	$^{1}H\alpha(i)$-$^{13}C\alpha(i)$-$^{15}N(i+1)$	Kay et al., 1990a
H(CA)NH	$^{1}H\alpha(i)$-$^{15}N(i)$-$^{1}HN(i)$ $^{1}H\alpha(i-1)$-$^{15}N(i)$-$^{1}HN(i)$	Kay et al., 1991
HA(CACO)NH	$^{1}H\alpha(i-1)$-$^{15}N(i)$-$^{1}HN(i)$	Grzesiek and Bax, 1993
HN(CO)HB	$^{1}H\beta(i-1)$-$^{15}N(i)$-$^{1}HN(i)$ $^{1}H\alpha(i-1)$-$^{15}N(i)$-$^{1}HN(i)$	Grzesiek et al., 1992
HBHA(CO)NH	$^{1}H\beta(i-1)$/$^{1}H\alpha(i-1)$-$^{15}N(i)$-$^{1}HN(i)$	Grzesiek and Bax, 1993
HNHA	$^{1}H\alpha(i)$-$^{15}N(i)$-$^{1}HN(i)$	Vuister and Bax, 1993a
HNHB	$^{1}H\beta(i)$-$^{15}N(i)$-$^{1}HN(i)$	Archer et al., 1991
HCCH-COSY	$^{1}H(i)$-$^{13}C(i)$-$^{13}C(i\pm1)$-$^{1}H(i\pm1)$	Kay et al., 1990c
HCCH-TOCSY	$^{1}H(i)$-$^{13}C(i)$...$^{13}C(i\pm n)$-$^{1}H(i\pm n)$	Bax et al., 1990a

Note: * For constant-time versions of some of the experiments in the table the reader is refered to Bax and
                Grzesiek (1993).

(Kay et al., 1990a). The intraresidue correlations are based on the one-bond scalar couplings between $^{1}HN(i)$ and $^{15}N(i)$ ($^{1}J = 90\text{-}100$ Hz) and between $^{15}N(i)$ and $^{13}C\alpha$ (i) ($^{1}J = 9\text{-}13$ Hz) (Table 1). The interresidue correlations make use of the two-bond scalar coupling between $^{15}N(i)$ in one residue and $^{13}C\alpha(i-1)$ in the preceding residue ($^{2}J = 5\text{-}10$ Hz) (Table 1). The ^{15}N resonances appear in F1, the $^{13}C\alpha$ resonances appear in F2, and the ^{1}H resonances are directly detected and appear in F3.

The ^{15}N and ^{1}HN resonances of one amino acid residue can also be correlated with the ^{13}CO resonance of the preceding residue in the HNCO experiment (Kay et al., 1990a). Correlations between ^{1}HN and ^{15}N in a given amide group are established, as in the HNCA experiment, using the one-bond scalar couplings between ^{1}HN(i) and ^{15}N(i) (^{1}J = 90-100 Hz). The interresidue correlation relies on the one-bond scalar coupling between ^{15}N in one residue and ^{13}CO in the preceding residue (^{1}J = 15 Hz; Table 1)

For some of the experiments shown in Table 3, the transfer of magnetization occurs through "intermediate" nuclei that are not observed and are therefore shown in parentheses. For example, in the HN(CO)CA experiment (Bax and Ikura, 1991), the transfers are from ^{1}HN(i) to ^{15}N(i) (^{1}J = 90-100 Hz), from ^{15}N(i) to ^{13}CO (i-1) (^{1}J = 15 Hz), and from ^{13}CO(i-1) to ^{13}Cα(i-1) (^{1}J = 55 Hz). The observed chemical shifts are only those of ^{15}N(i) (F1), ^{13}Cα(i-1) (F2) and ^{1}HN(i) (F3). Thus, by using the ^{13}CO nucleus as an "intermediate," the experiment establishes correlations between the ^{1}HN and ^{15}N resonances of one amino acid residue and the ^{13}Cα resonance of the preceding residue.

The heteronuclear 3D correlation experiments summarized in Table 3 provide multiple, independent pathways to assign backbone ^{1}H, ^{15}N, and ^{13}C resonances in a protein. This increases the reliability of the assignments and removes many of the spectral ambiguities inherent to proteins of larger size. An illustration of the use of heteronuclear correlation experiments for assignments of protein backbone resonances is provided by the recent NMR study of the HU protein from *Bacillus stearothermophilus* (Vis et al., 1994). This study started with a conventional 2D ^{1}H-^{15}N HSQC experiment to observe intraresidue correlations between ^{1}HN(i) and ^{15}N(i). Out of the expected 85 ^{1}HN-^{15}N cross-peaks, 80 were resolved in this 2D spectrum. This spectral resolution was satisfactory for application of 3D heteronuclear methods to a protein sample uniformly labeled with ^{13}C and ^{15}N (Figure 9). The ^{1}HN resonances of each amino acid residue were correlated with the ^{13}Cα resonances of the preceding residues in an HN(CO)CA experiment (panel at top left in Figure 9). For example, the ^{1}HN resonance of Glu51 (F3 = 9.18 ppm) shows a cross-peak to the ^{13}Cα of the preceding residue, Phe50 (F2 = 54.7 ppm). This peak occurs in the plane of the 3D spectrum at F1 = 120.21 ppm which is, within spectral resolution, the same as the ^{15}N resonance for Glu51 (F1 = 119.9 ppm). At this F1 spectral position, the ^{15}N resonances of several other amino acids (i.e., Arg58, Glu67, Glu34, Leu85, Ala27, Phe50, and Ala84) are present. The number of these amino acid residues is small enough that their ^{1}Hα(i+1)- ^{13}Cα(i) cross-peaks are resolved from one another, and from that of Glu51-Phe50, in the F2 and F3 dimensions. The sequential and intraresidue connectivities between ^{1}HN and ^{13}Cα were further checked in an HNCA experiment (panel at top right in Figure 9). In the same plane corresponding to the ^{15}N frequency of Glu51, two cross-peaks are observed. One is the same as that in the HN(CO)CA spectrum (i.e., ^{1}HN of Glu51 to ^{13}Cα of Phe50); the other represents the intraresidue connectivity between ^{1}HN and ^{13}Cα in Glu51 (F3 = 9.18 ppm and F2 = 52.4 ppm).
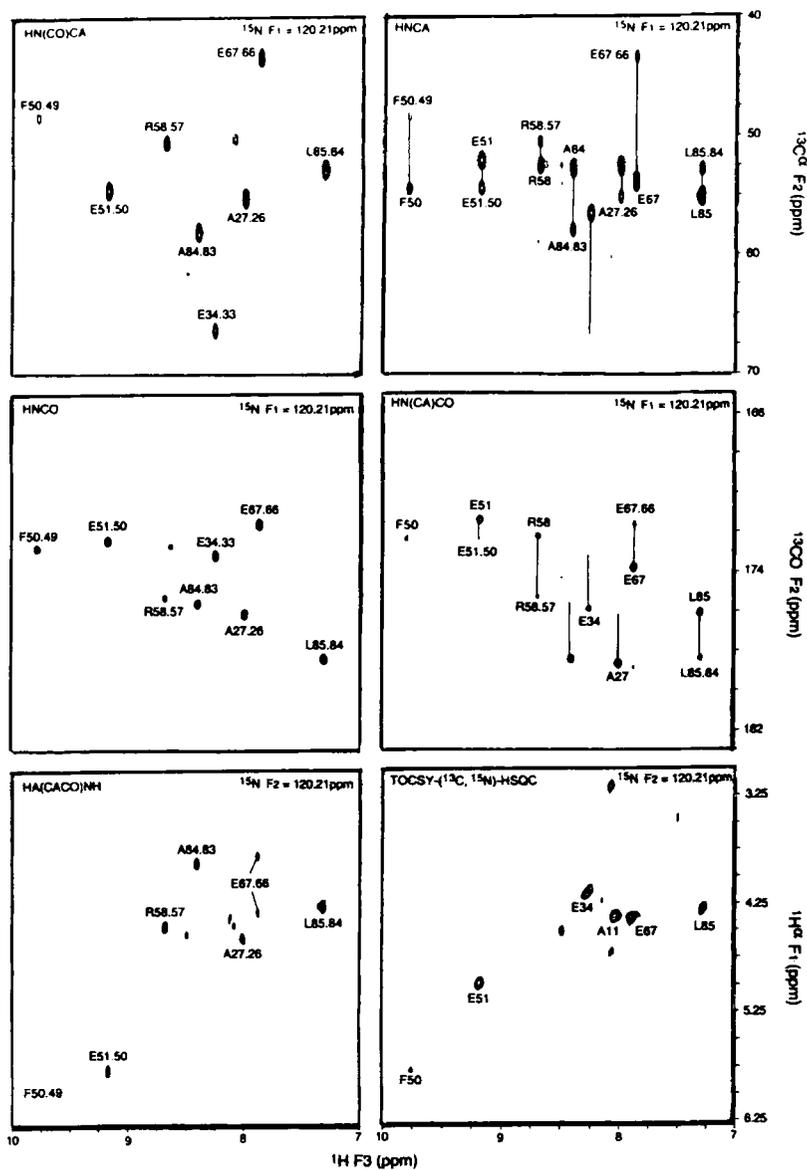
Figure 9. Selected $^{13}C\alpha/^{13}CO$(F2)-^1H(F3) planes of the HN(CO)CA, HNCA, HNCO, and HN(CA)CO spectra, and ^1H(F1)-^1H(F3) planes of the HA(CACO)NH and TOCSY-($^{13}C/^{15}N$)-HSQC spectra of the HU protein from *Bacillus stearothermophilus* (reproduced from Vis et al., 1994, with permission).

150

Sequential and intraresidue connectivities between ^{1}HN and ^{13}CO chemical shifts were established in a similar manner (central two panels in Figure 9). For example, in the HNCO spectrum, the ^{1}HN proton of Glu51 shows connectivity to ^{13}CO of Phe50 (F2 = 172.5 ppm). The same connectivity is present in the HN(CA)CO spectrum. This spectrum also contains the intraresidue connectivity between ^{1}HN and ^{13}CO of Glu51 (F3 = 9.18 ppm and F2 = 171.5 ppm).

The remaining backbone nuclei, ^{1}Hα, were assigned in TOCSY-(^{13}C,^{15}N)-HSQC and HA(CACO)NH experiments (lower two panels in Figure 9). The first experiment provided intraresidue connectivities between ^{1}HN and ^{1}Hα (e.g., cross-peak at F1 = 5.00 ppm and F3 = 9.18 ppm for Glu51). The second experiment provided sequential connectivities between ^{1}HN in one amino acid residue and ^{1}Hα in the preceding one (e.g., cross-peak at F1 = 5.83 ppm and F3 = 9.18 ppm for the connectivity between ^{1}Hα of Phe50 and ^{1}HN of Glu51).

The six 3D heteronuclear experiments described above for the HU protein involve connectivities to ^{1}HN protons. They are sufficient for complete assignments of all nuclei along the protein backbone. Independent confirmation of these assignments can be obtained from 3D experiments that correlate ^{1}Hα, ^{13}CO, ^{13}Cα, and ^{15}N nuclei only, without the participation of ^{1}HN protons. For example, the HCACO experiment gives intraresidue connectivities between ^{1}Hα, ^{13}CO and ^{13}Cα; similarly, the HCA(CO)N experiment gives sequential connectivities between ^{1}Hα and ^{13}Cα in one residue and ^{15}N in the subsequent one. These two experiments were used by Vis et al. (1994) as additional checks for backbone resonance assignments and for assigning ^{15}N chemical shifts in Proline residues.

Heteronuclear 3D experiments are also being used to delineate amino acid spin systems and to assign side chain resonances. Two of the most widely used methods are HCCH-COSY (Kay et al., 1990c) and HCCH-TOCSY (Bax et al., 1990a). These experiments rely on one-bond ^{1}H-^{13}C and ^{13}C-^{13}C couplings to transfer magnetization along the side-chain (Table 3). The transfer occurs in three steps: ^{1}H $\rightarrow$ ^{13}C $\rightarrow$ ^{13}C $\rightarrow$ ^{1}H. First, proton magnetization is transferred to directly-attached ^{13}C nuclei via the large ^{1}J[^{1}H-^{13}C] coupling (~ 140 Hz; Table 1). ^{13}C magnetization is then transferred to neighboring ^{13}C nuclei within a side-chain via the ^{1}J[^{13}C-^{13}C] coupling (33-45 Hz). Finally, ^{13}C magnetization is transferred back to protons via the ^{1}J[^{1}H-^{13}C] coupling. Since the coupling constants involved are large and independent of conformation, these methods are greatly superior to those involving homonuclear ^{1}H-^{1}H couplings (Fairbrother et al., 1992; Powers et al., 1992). Other methods for assignments of amino acid side chain resonances are described in section VA.

Heteronuclear-edited 3D experiments have been successfully used for resonance assignments in proteins of molecular weights up to 30 kDa. Examples of the largest proteins for which complete ^{1}H, ^{13}C, and ^{15}N backbone NMR assignments have been obtained are the 269-residue serine protease PB92 from *Bacillus alcalophilus* and the 269-residue protease subtilisin 309 from *Bacillus lentus* (Fogh et al., 1994; Remerowski et al., 1994). A repository of proteins with assignments of their NMR

resonances is being maintained by Markley and co-workers (Ulrich et al., 1989). With the assignments in hand, delineation of secondary structure and folding topology is possible. This aspect of protein structure determination is discussed in the next section.

IV. IDENTIFICATION OF REGULAR SECONDARY STRUCTURE ELEMENTS

Elements of secondary structure can be delineated from a quantitative analysis of NOEs and coupling constants involving protons on the polypeptide backbone. The NOEs of interest usually involve short ^{1}H-^{1}H distances, and their patterns follow the specific geometrical features of each secondary structure element. For example, an α-helix is distinguished by a stretch of strong sequential NOEs, d_{NN}, between ^{1}HN protons in successive amino acid residues. In an idealized α-helix, the distance between these two protons is 2.8 Å whereas, in β-sheets, the distance between the same protons is much larger (i.e., 4.2-4.3 Å; Wüthrich, 1986). Other NOE signatures of an α-helix result from the fact that, in this structure, amino acid residue i closely approaches residues (i+3) and (i+4). Hence, the following nonsequential NOEs are generally observed: $d_{\alpha N}$ (i,i+2), $d_{\alpha N}$(i,i+3), and $d_{\alpha N}$(i,i+4) (i.e., ^{1}Hα in residue i to ^{1}HN in residues (i+2), (i+3), and (i+4), respectively); d_{NN}(i,i+2) (i.e., ^{1}HN in residue i to ^{1}HN in residue (i+2)); and $d_{\alpha\beta}$ (i,i+3) (i.e., ^{1}Hα in residue i to ^{1}Hβ in residue (i+3)). These nonsequential medium range NOEs are absent in β-sheet structures, since the polypeptide chain is in an almost fully extended conformation. Instead, β-sheet structures are characterized by stretches of strong sequential $d_{\alpha N}$(i,i+1) NOEs. In addition, due to the close approach of neighboring strands, β-sheet structures also exhibit clusters of interstrand NOEs between ^{1}HN and ^{1}Hα protons in amino acid residues that are distant from each other in the primary structure. The patterns of short-distance proton NOEs on the protein backbone are commonly summarized along the amino acid sequence in diagrams such as that shown in Figure 10 for the first RNA binding domain of mammalian hnRNP A1 protein (Garrett et al., 1994).

Secondary structure elements in proteins can also be identified from scalar coupling constants of backbone protons. Generally, the intraresidue three-bond scalar coupling between ^{1}Hα and ^{1}HN protons, $^3J_{HN\alpha}$, is used. This coupling constant depends on the backbone torsion angle ϕ as in the Karplus equation (Equation 1). The equation has been parameterized for the torsion angle ϕ by correlating $^3J_{HN\alpha}$ values observed experimentally for basic pancreatic trypsin inhibitor (BPTI) with torsion angles ϕ derived from the crystal structure of the protein. The resulting dependence is expressed as (Pardi et al., 1984):

$$^3J_{HN\alpha} = 6.4 . \cos^2(\phi - 60°) - 1.4 . \cos(\phi - 60°) + 1.9 \qquad (2)$$

0   5   10   15   20   25   30   35   40   45   50   55   60   65   70   75   80   85   90

MSKSESPKEPEQLRKLFIGGLSFETTDESLRSHFEQWGTLTDCVVMRDPNTKRSRGFGFVTYATVEEVDAAMNARPHKVDGRVVEPKRAVSRR

NH Exchange

d$_{\alpha N(i,i+1)}$
d$_{NN(i,i+1)}$
d$_{\beta N(i,i+1)}$
d$_{\alpha N(i,i+2)}$
d$_{NN(i,i+2)}$
d$_{\alpha N(i,i+3)}$
d$_{\alpha \beta(i,i+3)}$
d$_{\alpha N(i,i+4)}$

$^3J_{HN\alpha}$ (Hz)   +8 / 0

$\Delta ^{13}C\alpha$ (ppm)   +4 / 0 / -4

$\Delta ^{13}C\beta$ (ppm)   +4 / 0 / -4

β1   α1   β2   β3   α2   β4
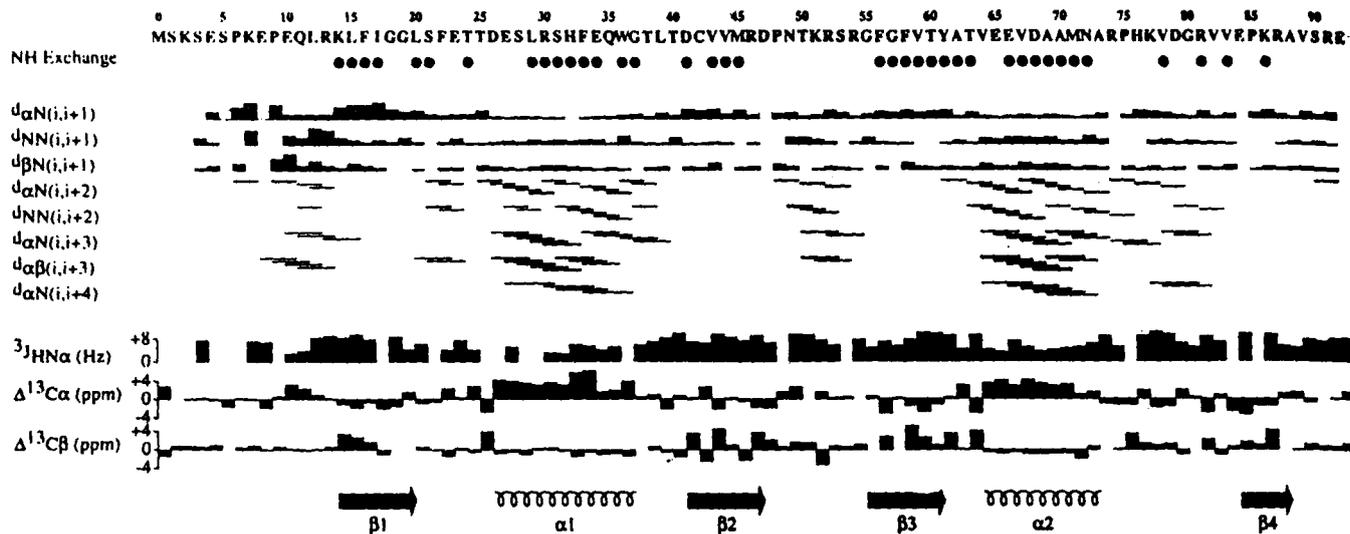
Figure 10. Summary of the sequential and medium-range NOEs involving ^1HN, $^1H\alpha$ and $^1H\beta$ protons, ^1HN exchange, $^3J_{HN\alpha}$, and $^{13}C\alpha$ and $^{13}C\beta$ secondary chemical shifts in the first RNA binding domain of mammalian hnRNP A1 protein. The thickness of the lines reflects the strength of the NOEs. ^1HN protons which are still present in the spectrum 10 hours after changing the sample into D_2O are indicated by closed circles. The elements of secondary structure deduced from the data are indicated at the bottom of the figure (reproduced from Garrett et al., 1994, with permission).

Accordingly, the following $^3J_{HN\alpha}$ coupling constants are predicted for various secondary structure elements: $^3J_{HN\alpha} = 3.9$ Hz for α-helix ($\phi = -57°$); $^3J_{HN\alpha} = 4.2$ Hz for 3_{10}-helix ($\phi = -60°$); $^3J_{HN\alpha} = 8.9$ Hz for antiparallel β-sheet ($\phi = -139°$); and $^3J_{HN\alpha} = 9.7$ Hz for parallel β-sheet ($\phi = -119°$) (Wüthrich, 1986). Thus, the $^3J_{HN\alpha}$ coupling constants are smaller for helices than for β-sheets. In practice, values of $^3J_{HN\alpha}$ less than 6 Hz are taken as indicators of helices, and values of $^3J_{HN\alpha}$ greater than 8 Hz as indicators of β-sheets (Figure 10).

The location of regular secondary structure elements in a protein can also be inferred from the rates of exchange of amide protons. Most amide protons in regular secondary structures are involved in hydrogen bonds (the exceptions are those in the first four residues in an α-helix, the first three residues in a 3_{10}-helix, and every second residue in the strands at the edges of a β-sheet). As a result of their participation in hydrogen bonds, these protons have slower rates of exchange with solvent protons. An example of this correlation between the location of secondary structure elements and proton exchange rates is shown in Figure 10.

Further confirmation for the location of secondary structure elements in the protein structure can be obtained from an analysis of chemical shifts. Recent comparative analyses of NMR and crystallographic databases have revealed that the deviations of NMR chemical shifts in proteins from their random coil values are strongly correlated with secondary structure elements (Spera and Bax, 1991; Wishart et al., 1991). For example, for $^{13}C\alpha$ nuclei, these secondary shifts are positive for residues located in α-helices ($+3.09\pm1.00$ ppm) and negative for those in β-sheets (-1.48 ± 1.23 ppm). The secondary shifts for $^{13}C\beta$ resonances exhibit the reverse pattern, with negative values for α-helices (-0.38 ± 0.85 ppm) and positive values for β-sheets ($+2.16\pm1.91$ ppm). These correlations are illustrated in Figure 10 for the secondary shifts of $^{13}C\alpha$ and $^{13}C\beta$ resonances of the first RNA binding domain of mammalian hnRNP A1 protein.

V.   PROTEIN STRUCTURE DETERMINATION

Determination of protein structures by NMR relies on torsion angles and interproton distances. Torsion angles are obtained from three-bond scalar coupling constants supplemented by intraresidue and sequential interresidue NOEs. Interproton distances are derived from through-space connectivities established by NOE effects. Torsion angles and interproton distances are then used as constraints in various algorithms that compute the protein 3D structure.

A.   Determination of Torsion Angle Constraints

The torsion angles of interest in protein NMR are the angles ϕ and ψ in the polypeptide backbone and the angles χ^1 and χ^2 in the amino acid side-chains. They are calculated from proton-proton and heteronuclear scalar coupling constants

using the Karplus equation (Equation 1). Since the value of the torsion angle derived from a given scalar coupling constant is not unique, the analysis is normally supplemented with intraresidue and sequential interresidue NOEs involving ^{1}HN, ^{1}Hα and ^{1}Hβ. Alternatively, for torsion angles ϕ and χ, values normally found in protein structures are selected (Wüthrich, 1986).

The backbone torsion angle ϕ is obtained from the three-bond intra-residue coupling constant $^3J_{HN\alpha}$ based on Equation 2. In earlier studies, $^3J_{HN\alpha}$ couplings were measured directly from the in-phase or anti-phase splitting of cross-peaks in the fingerprint region of 2D correlation spectra (Clore and Gronenborn, 1989). A limitation of these experiments is that, in larger proteins, the separation of multiplet components in ^{1}HN-^{1}Hα correlation cross-peaks is comparable to, or less than, resonance linewidths. As a result, the apparent splittings are significantly distorted and many smaller $^3J_{HN\alpha}$ couplings, such as those expected for α-helical regions of the protein, cannot be measured. This limitation has been overcome recently by ingenious uses of heteronuclear scalar coupling correlations. Some of the newly developed techniques remove part of the overlap in the correlation cross-peak by canceling multiplet components using E.COSY effects (Griesinger et al., 1986). The approach is most effective when one of the protons of interest, ^{1}HN or ^{1}Hα, has a large coupling to a third nucleus. The obvious choices for the third nucleus are ^{15}N for ^{1}HN (coupling constant 90-100 Hz, Table 1) and ^{13}Cα for ^{1}Hα (coupling constant 140 Hz, Table 1). A representative example of these pulse sequences is the ^{1}H-^{15}N-^{13}C triple-resonance scheme proposed by Wagner et al. (1991). In this experiment, the intraresidue ^{13}Cα-^{1}HN cross-peaks have E.COSY-like multiplet structures: only two components are present, and they are separated by $^3J_{HN\alpha}$ in F2 and by $^1J[^1$Hα-^{13}C$\alpha]$ in F1. Thus, the homonuclear coupling constant of interest, $^3J_{HN\alpha}$, can be measured independently of the linewidth. Alternative approaches use modifications of COSY experiments for example, the J-modulated (^{15}N,^{1}H) COSY pulse sequence proposed by Neri et al., 1990; others exploit the fact that the T_2 values of ^{15}N-^{1}H zero- and double-quantum coherences can be much longer than the corresponding one-quantum T_2 values of ^{1}HN protons. As a result, multiple-quantum linewidths are narrower, and more accurate $^3J_{HN\alpha}$ coupling constants can be extracted from pulse schemes built based on ^{1}H-^{15}N HMQC. For example, the 2D HMQC-J experiment (Kay and Bax, 1990) has allowed measurements of $^3J_{HN\alpha}$ couplings as small as 4 Hz for proteins as large as human thioredoxin (molecular weight 12 kDa) and staphylococcal nuclease (molecular weight 18 kDa) (Forman-Kay et al., 1990; Kay and Bax, 1990). Quantitative measurements of $^3J_{HN\alpha}$ coupling constants can also be made by the newly-developed HNHA experiment (Vuister and Bax, 1993a). This ^{15}N-separated ^{1}HN-^{1}Hα homonuclear J-correlation experiment correlates intraresidue ^{1}HN, ^{15}N, and ^{1}Hα resonances (Table 3). The ratio of the intensity of the diagonal cross-peak at [F1(^{15}N); F2(^{1}HN); F3(^{1}HN)] to the intensity of the cross-peak at [F1(^{15}N); F2(^{1}Hα); F3(^{1}HN)] also provides a direct measure of the $^3J_{HN\alpha}$ coupling constant. Coupling constants as low as 3 Hz have been obtained by this method in staphylococcal nuclease (Vuister and Bax, 1993a).

The backbone torsion angle ψ is obtained from the heteronuclear coupling constant $^3J_{N\alpha}$ between ^{15}N in one residue and $^1H\alpha$ in the preceding residue. This coupling constant is related to ψ by a Karplus-type relationship and has a maximum value of ~6 Hz for $\psi = -60°$. For $-120° < \psi < 0°$, $^3J_{N\alpha}$ is greater than 2 Hz, while for other values of ψ it is less than 2 Hz (Bystrov, 1976). Since ψ torsion angles range between -40 and -80° for α-helices and between 40 and 180° for β-sheets, larger coupling constants are expected for α-helices. $^3J_{N\alpha}$ coupling constants can be obtained from HMBC experiments (Clore et al., 1988). However, due to the small magnitude of $^3J_{N\alpha}$ and the strong dependence of the intensity of the $^{15}N(i)$-$^1H\alpha(i\text{-}1)$ cross-peaks on proton T_2, cross-peaks are normally observed only for α-helical regions of the protein.

Torsion angles χ^1 for amino acid side chains containing β-methylene groups have been derived from the three-bond intra-residue coupling constant $^3J\alpha\beta$ between $^1H\alpha$ and $^1H\beta$. The Karplus equation for this coupling constant has been parameterized using coupling constants measured for ferrichrome and torsion angles calculated based on the crystallographic structure of the peptide (DeMarco et al., 1978a):

$$^3J_{\alpha\beta}(\Theta) = 9.5.(\cos\Theta)^2 - 1.6.\cos\Theta + 1.8 \qquad (3)$$

where $\Theta = \chi^1 -120°$ for $^1H\beta2$ and $\Theta = \chi^1$ for $^1H\beta3$. The use of $^3J_{\alpha\beta}$ to derive the torsion angle χ^1 is intrinsically related to the stereospecific assignments of the β-methylene protons. $^3J_{\alpha\beta}$ coupling constants and multiplet structures of the corresponding cross-peaks in 2D correlation spectra have been used to differentiate $^1H\alpha$-$^1H\beta2$ and $^1H\alpha$-$^1H\beta3$ cross-peaks (Hyberts et al., 1987). To assign $^1H\beta2$ and $^1H\beta3$ protons stereospecifically, this analysis is supplemented by quantitative measurements of the intraresidue NOEs from ^1HN to $^1H\alpha$, $^1H\beta2$, and $^1H\beta3$ protons. The distances between these protons differ among various rotameric states around the $C\alpha$-$C\beta$ bond. Hence, for a range of χ^1 torsion angles, the relative intensities of the corresponding NOEs provide assignments for each β-methylene proton resonance. A similar method has been used for stereospecific assignments of amino acid side-chains containing pairs of methyl groups, e.g., Val and Leu (Zuiderweg et al., 1985; Hyberts et al., 1987). Current strategies employ systematic searches of databases created based on high-resolution crystallographic structures of proteins, or large sets of allowed conformers of short peptides with idealized geometry (Güntert et al., 1989; Nilges et al., 1990). This structural information is used to calculate intra-residue and sequential NOEs, and $^3J_{HN\alpha}$ and $^3J_{\alpha\beta}$ coupling constants (Equations 2 and 3). The experimental NOEs and coupling constants are matched with those in the databases to select correct stereospecific assignments and to determine allowed ranges for ϕ, ψ, and χ^1.

Determination of χ^1 and χ^2 torsion angles has greatly benefited from newer 2D and 3D double- and triple-resonance pulse sequences that measure heteronuclear coupling constants. For the torsion angle χ^1, the coupling constants $^3J[^1H\beta2\text{-}^{15}N]$

and $^3J[^1H\beta3-^{15}N]$ are of interest. They depend on the angle χ^1 as in the Karplus equation:

$$^3J[^1H\beta-^{15}N] = 4.4.(\cos\Theta)^2 - 1.2.\cos\Theta - 0.1 \tag{4}$$

where $\Theta = \chi^1 + 120°$ for $^1H\beta2$ and $\Theta = \chi^1 - 120°$ for $^1H\beta3$ (DeMarco et al., 1978b). These heteronuclear couplings can be measured from 2D or 3D H_NNH_{AB}-COSY experiments (Chary et al., 1991), or from the HNHB experiment (Archer et al., 1991). An additional HN(CO)HB experiment (Table 3) gives intraresidue $^3J[^1H\beta-^{13}CO]$ coupling constant (Grzesiek et al., 1992). $^3J[^1H\beta-^{15}N]$ and $^3J[^1H\beta-^{13}CO]$, in combination with $^3J[^1H\alpha-^1H\beta]$, determine the χ^1 angle and allow β-methylene protons to be unambiguously assigned. Torsion angles χ^1 and χ^2 can also be obtained by measuring scalar correlations involving carbon nuclei in methyl groups (Bax et al., 1994). The advantage in observing methyl groups is that the linewidth of their ^{13}C resonances is reduced, even in larger proteins, due to the rapid rotation of these groups about their threefold symmetry axis. Three-bond $^{13}C-^{13}C$ coupling constants can be measured in ^1H-detected [$^{13}C-^{13}C$] long-range correlation (LRCC) spectra (Bax et al, 1992). For example, scalar couplings between ^{13}C resonances of Cγ and Cδ methyl groups and $^{13}C\alpha$, ^{13}CO and $^{13}C\beta$ in Val, Ile, Leu, and Thr allow determination of χ^1 and χ^2 torsion angles in these amino acid residues. The χ^1 and χ^2 torsion angles in the same amino acid residues can also be obtained from the ^1H-detected [$^{13}C-^1H$] long-range correlation (LRCH) experiment which measures two- and three-bond correlations between methyl-^{13}C and ^1H nuclei (Vuister and Bax, 1993b; Vuister et al., 1993a). Of special interest for determination of χ^1 torsion angles are the three-bond couplings between methyl-^{13}C and backbone ^{15}N nuclei. These couplings can be measured in ^{13}C-$\{^{15}N\}$-Spin-Echo-Difference and ^{13}C-$\{^{13}CO\}$-Spin-Echo-Difference experiments (Grzesiek et al., 1993; Vuister et al., 1993b); they provide stereospecific assignments of methyl resonances in Val and the angle χ^1 in Val, Ile, and Thr residues.

B. Determination of Distance Constraints

The main source of distance constraints for protein structure determination are the $^1H-^1H$ NOE effects. A prerequisite in the use of NOEs for structure determination is the resolution and unambiguous identification of the proton pairs giving rise to the observed NOEs. Obviously, this identification relies upon assignments of resonances to specific protons in the backbone and in the side-chains. However, even when complete resonance assignments are available, resolution of individual NOE cross-peaks may be limited. To enhance resolution and unambiguously assign NOEs to specific proton pairs, 3D and 4D NMR experiments on proteins labeled with ^{15}N and/or ^{13}C are used. The experiments are constructed following the general scheme shown in Figure 3. Their experimental implementation relies upon the transfer of magnetization between protons and directly-bonded hetero-

nuclei by multiple quantum coherence (Bax et al., 1983). For example, in the
^{15}N-separated NOE (^{1}H-^{15}N NOESY-HMQC), incorporation of an HMQC scheme
in the evolution time t_2 selects for observation only NOE connectivities for amide
protons (Marion et al., 1989). All through-space connectivities of a given amide
proton appear in the plane of the 3D spectrum at the F2 frequency of the ^{15}N nucleus
in that amide group (as shown schematically in Figure 4). Extension to four
dimensions can be achieved by further editing of ^{13}C-bound protons. A 4D
^{13}C-^{15}N-edited NOE experiment is constructed from three 2D experiments: ^{1}H-^{13}C
HMQC, ^{1}H-^{1}H NOESY, and ^{1}H-^{15}N HMQC (Kay et al, 1990b). The ^{1}H-^{13}C
HMQC selects ^{13}C-bound protons for the F2 dimension; the ^{1}H-^{15}N HMQC selects
^{15}N-bound protons for the F4 dimension. The NOEs between these two classes of
protons develop during the central ^{1}H-^{1}H NOESY of the pulse sequence. The
chemical shifts of the ^{13}C nuclei bound to the originating protons are in F1, and
the chemical shifts of the ^{15}N nuclei bound to the destination protons are in F3. To
edit only the NOEs between ^{13}C-attached protons, a 4D ^{13}C-^{13}C-edited NOE
experiment can be used (Clore et al., 1991a). The experiment is conceptually similar
to the 4D ^{13}C-^{15}N-edited NOE experiment, except that the central ^{1}H-^{1}H NOESY
pulse scheme is now bracketed by two ^{1}H-^{13}C HMQC. In practice, the 4D
^{13}C-^{13}C-edited NOE experiment is significantly more demanding due to multiple
spurious magnetization transfer pathways that can result in observable signals. An
illustration of the results obtained by 4D ^{13}C-^{13}C-edited NOE for interleukin 1β is
given in Figure 11. In the ^{1}H-^{1}H NOESY spectrum (panel A in Figure 11), no
individual cross-peaks can be observed in the region between 1 and 2 ppm due to
the lack of resolution of the spectrum for a protein of the size of interleukin 1β (i.e.,
153 amino acids, molecular weight ~ 17.5 kDa). In contrast, in the plane of the 4D
^{13}C-^{13}C-edited NOE experiment at F1(^{13}C) = 44.3 ppm and F3(^{13}C) = 34.6 ppm,
the same spectral region contains only two cross-peaks, one of which originates
from the intraresidue NOE between ^{1}Hβ and ^{1}Hγ of Lys77 (box in panel B in Figure
11). The enhanced resolution in two other planes of the spectrum is illustrated in
panels C and D. For the detailed analysis of this spectrum, the reader is refered to
Clore et al. (1991a).

Interproton distances are quantitatively evaluated from the intensities of ^{1}H-^{1}H
NOE cross-peaks. The first fact to note is that, due to the strong dependence of the
NOE effect on the distance (namely, NOE α $1/r^6$), NOE intensities decrease very
rapidly as the distance between two protons increases. In a protein of average size
and at the magnetic fields currently available, NOE effects are observable for
protons separated by ~6 Å or less. Thus, NOE-derived interproton distances used
in structure determination are short-range. However, since their number is large and
they encompass both secondary and tertiary structure interactions, they are able to
define accurately the protein conformation.

In a NOESY experiment on a protein molecule, the intensities of the ^{1}H-^{1}H
cross-peaks can be represented as a matrix A in which the element A_{ij} corresponds
to the NOE between protons i and j. The dependence of the NOE cross-peak
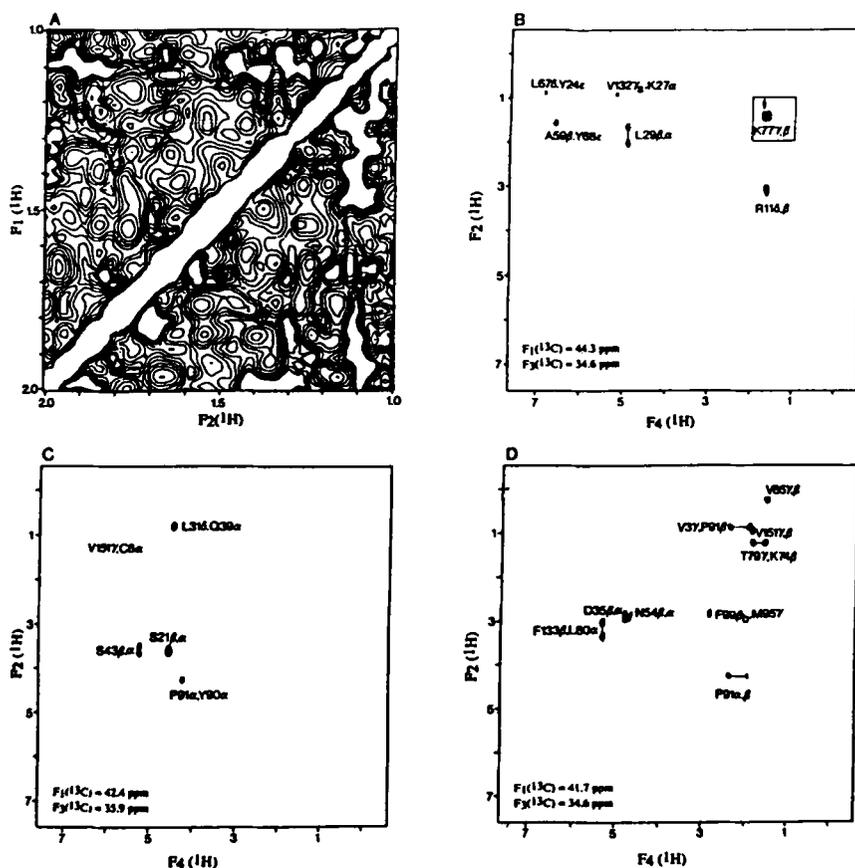
Figure 11. Comparison of the 2D and 4D NMR spectra of interleukin-1β. **(A)** Region between 1 and 2 ppm in the ^{1}H-^{1}H 2D NOESY spectrum. **(B)**-**(D)** Selected ^{1}H(F2)-^{1}H(F4) planes at several ^{13}C(F1) and ^{13}C(F3) frequencies of the 4D ^{13}C/^{13}C NOESY spectrum (reproduced from Clore and Gronenborn, 1993, with permission).

intensities on the mixing time of the experiment, τ_m, is then described by the following differential equation (Solomon, 1955):

$$dA(\tau_m)/d\tau_m = -R \cdot A(\tau_m) \tag{5}$$

with the solution:

$$A(\tau_m) = \exp(-R \cdot \tau_m) \tag{6}$$

in which R is the relaxation matrix for the protein of interest. The diagonal elements of the matrix, R_{ii}, are the self-relaxation rates of each proton while the off-diagonal elements, R_{ij}, are the cross-relaxation rates between protons i and j. The latter rates are of interest for protein structure determination since they are related to the distance between the two protons. For a pair of protons in which the interproton distance r_{ij} does not depend on time, the cross-relaxation rate is (Solomon, 1955):

$$R_{ij} = 0.1 \cdot (\gamma_H^4 \cdot \hbar^2/r_{ij}^6) \cdot [6J(2\omega_H) - J(0)] \qquad (7)$$

where γ_H is the proton gyromagnetic ratio, $\hbar$ is Planck's constant (divided by 2π), $J(\omega)$ is the spectral density function, and ω_H is the proton resonance frequency. The spectral density function $J(\omega)$ is dependent on the motions that modulate the dipole-dipole interaction between the two protons, e.g., overall rotation of the protein molecule and internal motions of side-chains.

Equations 5 to 7 show that, in general, derivation of interproton distances from NOE data involves two steps. First, the cross-relaxation rate R_{ij} must be evaluated from the matrix of NOE cross-peak intensities; second, the distance r_{ij} is calculated based on the cross-relaxation rate (Equation 7). Cross-relaxation rates can be obtained by using an approximation of $A(\tau_m)$ (Equation 6) which is valid for short mixing times:

$$A_{ij}(\tau_m) = - R_{ij} \cdot \tau_m \qquad (8)$$

In this approximation, only the initial slopes of the NOE build-up are measured, and the NOE intensities can be directly related to cross-relaxation rates. One limitation of this method is that NOESY experiments at short mixing times have low signal-to-noise ratios which lower the accuracy of the measurements and preclude observation of weaker cross-peaks. In addition, experimental precautions must be taken to insure that the measurements are made in the range where the NOE intensity depends linearly on the mixing time.

Calculation of interproton distances, r_{ij}, from cross-relaxation rates R_{ij} (Equation 7) requires some a priori knowledge of the molecular motions that affect the spectral density function $J(\omega)$. Although efforts are underway to incorporate these variables in protein structure determination (Brüschweiler and Case, 1994), most protocols assume that the spectral density functions are the same for all proton pairs and are determined by the overall correlation time of the protein molecule. In this "rigid-structure" model, interproton distances can be calculated based on Equation 7 by calibrating the intensity of the corresponding NOE cross-peak, Aij, against the cross-peak intensity, A_0, for a proton pair of known distance, r_0, as follows:

$$r_{ij} = r_0 \cdot (A_0/A_{ij})^{1/6} \qquad (9)$$

Given the theoretical complexities involved in deriving interproton distances from NOE data, the analysis described above is applied conservatively and, in most cases, distance constraints are qualitatively expressed as ranges between upper and lower bounds. The lower bound is the sum of the atomic core radii of the two protons (i.e., 1.8-2.0 Å depending on the atoms to which the protons are bound). The upper bounds are calibrated for each experiment based on distance ranges that are expected in a protein molecule. For example, the interresidue distances $^1H\alpha(i)$-$^1HN(i+1)$ usually range between 2.2 and 3.6 Å, while the intraresidue distances $^1H\alpha(i)$-$^1HN(i)$ range between 2.2 and 2.9 Å (Wüthrich, 1986). By comparing the NOE intensities for these pairs of protons to the intensities of the NOEs of interest, upper bounds for distances are obtained. Generally, the NOE database is classified into strong, medium, and weak NOEs. The upper distance bounds for these classes vary depending upon the experiment. A typical distribution would correspond to the following upper distance bounds: 2.7 Å for strong intensity, 3.3 Å for medium intensity, and 5.0 Å for weak intensity. For experiments of higher sensitivity, a fourth class is introduced, namely, very weak NOEs corresponding to distances up to 6.0 Å (Qin et al., 1994; Lodi et al., 1995). For NOEs involving ^1HN protons, the upper distance bounds for strong and medium intensities are often extended to 2.9 and 3.5 Å, respectively (Qin et al., 1994). Extensive NOE build-up data also allow closer upper bounds to be obtained, e.g., 2.4, 2.7, 3.0, 3.5, and 4.3 Å for backbone distances in French bean plastocyanin (Moore et al., 1991). Distances involving methyl and methylene protons which are not stereospecifically assigned are represented by replacing the corresponding group of atoms with a pseudoatom (Wüthrich et al., 1983). As expected, the use of pseudoatoms lowers the precision of the experimental structural constraints (Güntert et al., 1989). For backbone protons located in regular secondary structure elements, additional distance constraints are introduced in some stages of structure determination in order to reflect hydrogen-bonding. Typical values are 1.5-2.3 Å for the distance between the hydrogen and the acceptor atom, and 2.4-3.3 Å for the distance between the donor and the acceptor atom (Powers et al., 1993).

Interproton distance constraints have also been derived from 3D homonuclear NOESY-NOESY experiments. Analogous to Equation 6, the intensity of a cross-peak between protons i, j, and k in this 3D spectrum is expressed as:

$$A_{ijk}(\tau_m^{(1)}, \tau_m^{(2)}) = [\exp(-R \cdot \tau_m^{(2)})]_{ij} \cdot [\exp(-R \cdot t_m^{(1)})]_{jk} \qquad (10)$$

where $\tau_m^{(1)}$ and $\tau_m^{(2)}$ are the two mixing times (Figure 3). Various cross-peaks can be associated with various NOE transfers. For example, the cross-peak at F1=F(i), F2=F(j), F3=F(j) corresponds to a direct transfer from proton i to proton j; the cross-peak at F1=F(i), F2=F(j), F3=F(k) corresponds to a spin-diffusion transfer from proton i to proton j and from proton j to proton k; and the cross-peak at F1=F(i), F2=F(j), F3=F(i) corresponds to a back-transfer from proton i to proton j and back to proton i. Cross-peaks corresponding to spin-diffusion and back-transfers can be

used to quantitate spin-diffusion pathways and to correct the intensities of direct-transfer cross-peaks to pure cross-relaxation terms (Boelens et al., 1989; Kessler et al., 1991). Thus, more accurate distance constraints can be obtained. Recent applications of homonuclear 3D NOESY-NOESY experiments suggest that this approach could yield more accurate structures, especially for proteins of smaller size (Habazettl et al., 1992).

C.   Computation of Protein Three-Dimensional Structure

Protein structure determination is typically initiated by generating a set of initial structures that are consistent with stereochemical constraints imposed by the protein covalent structure and with experimentally-derived NMR constraints. At subsequent stages, these initial structures are refined and the quality of the final structure(s) is evaluated.

One method widely used to generate initial structural models based on NMR data is the metric matrix distance geometry. The method operates in distance space in which a macromolecule containing N atoms is described by N(N-1)/2 distances. The goal is to convert interatomic distances into the 3N atomic coordinates of the macromolecule in the Cartesian coordinate space (a procedure known as embedding). Rigorous mathematical arguments described elsewhere (Crippen and Havel, 1988) have established conditions in which a 3D structure exists that is consistent with a given distance matrix. The computational challenge when applying metric matrix distance geometry to protein structure determination is the fact that the number of interatomic distances provided by NMR is small. Although NMR-derived distances are supplemented with distance and planarity constraints dictated by the protein covalent structure, the fraction of known distances remains relatively small. Moreover, most distances are not uniquely known and are expressed as ranges between upper and lower bounds. In spite of these conceptual difficulties computationally efficient algorithms have been developed in the last decade to reduce the bounds ("bound-smoothing"), to select values of the input distances within the bounds, and to increase the efficiency of the metric matrix calculations by embedding subsets of atoms. For a detailed description of these and other computational methods, the reader is refered to the reviews by Clore and Gronenborn (1989), Havel (1991), Brünger and Nilges (1993), and Case and Wright (1993).

Initial structures obtained by metric matrix distance geometry are typically refined by restrained molecular dynamics methods. The basis of these methods is to integrate Newton's equations of motion using an empirical energy function. The energy function contains bonded interactions, nonbonded interactions (i.e., van der Waals, electrostatic, and hydrogen bonding), and effective potential terms that incorporate NMR-derived distance and angle constraints. The potential term for NMR-derived distances usually has the form of a square-well potential:

$$
\begin{aligned}
E_{NOE} &= K \cdot (r_{ij} - r_{ij}^{upper})^2 && \text{for } r_{ij} > r_{ij}^{upper} \\
&= K \cdot (r_{ij} - r_{ij\,lower})_2 && \text{for } r_{ij} < r_{ij}^{lower} \\
&= 0 && \text{for } r_{ij}^{lower} < r_{ij} < r_{ij}^{upper}
\end{aligned}
\qquad (11)
$$

where r_{ij}^{upper} and r_{ij}^{lower} are the upper and the lower bounds of the distance of interest, respectively, and K is the NOE force constant. The potential term for torsion angle constraints has a similar form.

Another method used to generate protein structures is a hybrid method that combines metric matrix distance geometry and dynamical simulated annealing (Clore and Gronenborn, 1989, 1993). Initial substructures are generated using metric matrix distance geometry for only subsets (approximately one-third) of the atoms. The structures have approximately the correct fold and are then used as starting structures for dynamical simulated annealing. In this algorithm, the target function maintains correct bond lengths and angles as well as planarity and chirality constraints. The NMR-derived constraints are introduced, as in restrained molecular dynamics, as square-well potential terms (Equation 11). In contrast to restrained molecular dynamics, all nonbonded interaction terms are replaced by simple van der Waals repulsion terms. This simplification increases the computational speed by a factor of at least two (Clore and Gronenborn, 1989). Alternative starting structures that have been used for simulated annealing are random structures with covalent geometry (Nilges et al., 1988a) and completely random arrays of atoms (Nilges et al., 1988b).

A new development in protein structure determination is the incorporation of ^{1}H and ^{13}C chemical shifts in computational algorithms (Case and Wright, 1993; Kuszewski et al., 1995a, b). This methodology and its impact on the accuracy and precision of protein structures have been recently tested for myoglobin (Ösapay et al., 1994) and human thioredoxin (Kuszewski et al., 1995a, b). Chemical shift restraints have been incorporated into structure refinement as pseudoenergy terms with a harmonic or square well potential of the form $E = k \cdot (\Delta\sigma)^2$, where k is a force constant and $\Delta\sigma$ is the difference between the observed and the expected chemical shifts. Expected ^{1}H chemical shifts were calculated based on high-resolution NMR or crystallographic structures of the protein using empirical models (Ösapay and Case, 1991; Williamson and Asakura, 1993). Expected [13]Cα and [13]Cβ chemical shifts were derived from an empirical correlation between [13]Cα and [13]Cβ chemical shifts and backbone torsion angles ϕ and ψ (Spera and Bax, 1991). In the case of myoglobin, incorporation of chemical shift constraints in structure refinement resulted in a narrower spread of calculated structures and a smaller deviation from the crystallographic structure. For thioredoxin, although the precision of the structure was not significantly affected, small atomic shifts in the coordinates were observed. Incorporation of chemical shifts into refinement procedures appears to be particularly useful for amino acid residues whose resonances are shifted by ring-current effects and for amino acids in which $^3J_{HN\alpha}$ coupling constants cannot be measured accurately.

A general feature of all methods for protein structure determination is that the refinement strategy is iterative. One typically starts with a conservative approach in which only NOE data that can be unambiguously interpreted are used. This yields a low resolution structure which can then be used to resolve original ambiguities in the identification of NOE cross-peaks, stereospecific assignments, and torsion angle restraints. The cycle of structure determination is repeated iteratively and, as the quality of the structures improves, more experimental constraints are incorporated.

The final NMR-derived structure is, in fact, an ensemble of structures calculated independently, based on the same experimental data but using different initial structures or conditions (Figure 12). The report on the structure normally includes the average root-mean-square (rms) deviation of the family of structures from their mean. Typical values of rms deviations for backbone atoms range from 0.4 to 1 Å. Higher rms deviations are obtained for the side-chains, depending on their location in the structure. Generally, NMR-derived structures are in good agreement with X-ray diffraction structures. For example, the NMR solution structures of BPTI differ from their mean by 0.4 Å, a value which is comparable to the average difference between the three crystallographic structures of the protein. Overall, the mean solution conformation is within 0.8 Å from the crystallographic one (Berndt et al., 1992). Similar agreements have been noted for interleukin 1β (Clore and Gronenborn, 1991b), reduced French bean plastocyanin (Moore et al, 1991), α-amylase inhibitor Tendamistat (Billeter et al., 1989), and the repressor from phage 434 (Neri et al., 1992).

The rms deviation of the ensemble of structures from their mean provides a direct measure of the precision of the NMR-derived structure. The quality of the structure is also evaluated by analyzing the number and the size of violations of structural constraints. These include deviations of distances and torsion angles from experimental constraints and energy terms associated with nonbonded interactions. Moreover, like in crystallographic analysis, deviations from idealized covalent geometries and from expected positions on the Ramachandran ϕ / ψ plots are considered. A low number of small violations is a good indicator that the structures have reached the global minimum of the target function that was searched for by the computational algorithm. Further evaluation of the quality of the structure is made by comparing the chemical shifts and NOE cross-peak intensities predicted by the structure to those observed experimentally (Case and Wright, 1993).

At the present time, active research efforts are being devoted to improving the methods for protein structure determination by NMR, and to developing new ones. As shown in this chapter, the constraints provided by NMR spectroscopy for interproton distances and torsion angles are imprecise. Hence, efforts are directed toward increasing the accuracy of interproton distance constraints by using the full solution of the relaxation equation (Equation 6) instead of the approximation in Equation 8 (Boelens et al., 1988; Borgias and James, 1988; Borgias et al., 1990). Other groups are exploring a completely different strategy based on the refinement

1st Generation
  ~ 7 restraints per residue
  rmsd   1.5Å for backbone atoms
          2.0Å for all atoms
  example: purothionin

2nd Generation
  ~ 10 restraints per residue
  rmsd: 0.9Å for backbone atoms
          1.2Å for all atoms
  example: BDS-I

3rd Generation
  ~ 13 restraints per residue
  rmsd: 0.7Å for backbone atoms
          0.9Å for all atoms
  example: BDS-I

4th Generation
  ~ 16 restraints per residue
  rmsd: 0.4Å for backbone atoms
          0.9Å for all atoms,
          ≤ 0.5Å for ordered side chains
  example: Interleukin-8

Figure 12. Illustration of the progressive improvement in the precision and accuracy of NMR protein structure determination with increasing number of experimental constraints (reproduced from Clore and Gronenborn, 1993, with permission).

of the structure directly against the observed NOE intensities (Yip and Case, 1991; Brüschweiler and Case, 1994). At the first sight, it is surprising that approximate constraints, such as those generated by NMR, can yield high-resolution protein structures. The explanation probably lies both in the nature as well as in the number of these constraints. Although approximate, the NMR-derived constraints are conformationally restrictive since many interproton vectors are highly correlated,

and short-distance constraints between amino acid residues far apart in the sequence greatly restrict the conformational states allowed to the polypeptide chain (Clore and Gronenborn, 1989). Moreover, a recent study suggests that it is the number of structural constraints that is the dominant factor in determining the precision and the accuracy of NMR-derived structures (Clore et al., 1993). This finding is qualitatively illustrated in Figure 12. Earlier protein structure determinations were based on 5-7 experimental constraints per amino acid residue. Their rms distributions were large and conformational details were obscured by the variations of individual structures around their mean. As the number of constraints per residue increased, the precision of structure determinations increased. In recent years, protein structures have been determined with an average number of constraints as high as 20 per residue and rms deviations for backbone atoms as low as 0.4 Å. The precision of NMR-derived structures continues to increase; for example, the recent determination of the solution structures of oxidized and reduced human thioredoxin was made on the basis of ~28 experimental restraints per residue to a rms deviation for backbone atoms of 0.18-0.19 Å (Qin et al., 1994). The following are several proteins for which high-resolution structures (rms deviations for backbone atoms < 1 Å) have been determined by NMR: the *Antennapedia* homeodomain from *Drosophila* (Billeter et al., 1990), *Escherichia coli* thioredoxin (Dyson et al., 1990), the α-amylase inhibitor Tendamistat (Billeter et al., 1989), interleukin 1β (Clore et al., 1991b), bovine pancreatic trypsin inhibitor (Berndt et al., 1992), reduced French bean plastocyanin (Moore et al., 1991), the DNA-binding domain of HIV-1 integrase (Lodi et al., 1995), the oligomerization domain of the tumor supressor p53 (Clore et al., 1994), the DNA-binding domain of the c-*myb* protooncogene product (Ogata et al., 1992), *xfin*-31 zinc-finger peptide (Lee et al., 1989, 1990), the double zinc finger of the human enhancer binding protein MBP-1 (Omichinski et al., 1992), interleukin-4 (Powers et al., 1993), and the DNA-binding domain of the 434 repressor (Neri et al., 1992). As the methodology is refined, NMR structures are also emerging for complexes of proteins with ligands, e.g., caldmodulin in complex with a target peptide (Ikura et al., 1992), and the DNA binding domain of transcription factor GATA-1 in complex with its cognate DNA site (Omichinski et al., 1993).

VI.   CONCLUDING REMARKS

In the last decade, protein structure determination by NMR spectroscopy has been revolutionized by the development of multi-dimensional heteronuclear NMR experiments. As the field continues to grow, new directions as well as new challenges are emerging. Many of the proteins of interest in biology are large. Moreover, the function of proteins often involves formation of complexes between proteins and ligands, substrates, nucleic acids, or receptors. Thus, at present, one challenge in protein NMR spectroscopy is to increase the size of molecular systems amenable

to NMR studies. Solving the structure of larger proteins or protein complexes using NMR presents both theoretical and practical problems. As the size of the protein increases, the transverse relaxation rates become faster and NMR resonances become broader. In multi-dimensional NMR experiments, the fast transverse relaxation during the experiment results in weaker signals during the detection period (Wagner, 1993). This problem can be circumvented, at least in part, by pulse sequences which use coherences that relax most slowly (Bax and Grzesiek, 1993; Wagner, 1993). Pulse field gradients are also being used to eliminate or select coherence transfer pathways, and thus decrease the time needed for data acquisition in multi-dimensional experiments. Higher field instruments (proton resonance frequency in the range from 750 MHz and 1 GHz) are becoming available, and they too are expected to increase the sensitivity of NMR measurements. Extending the size of proteins also places new demands on NMR data analysis. The large number of resonances and NMR structural constraints in proteins of larger size requires computer-aided analysis. Several software packages are already available for automatic peak picking and resonance assignments, and more are forthcoming.

At present, sustained efforts are also devoted to improving the quality of NMR-derived protein structures. As explained above, the number and the quality of NMR-derived structural constraints is continuously increasing, new computational algorithms are being developed, and the theoretical and computational basis for incorporation of dynamic information in structure determination is being tested and refined. As part of these efforts, NMR spectroscopy is also being coordinated with X-ray crystallography. The symbiotic relationship between NMR and X-ray crystallography began to emerge several years ago when the phase problem in the crystal structure determination for the α-amylase inhibitor Tendamistat was solved with a Patterson search using the NMR structure (Braun et al., 1989). Similarly, the solution of the crystal structure for the three zinc fingers from Zif268 in complex with DNA was facilitated by fitting the NMR structure of the *Xfin*-31 zinc finger into the electron density (Pavletich and Pabo, 1991; Lee et al., 1989). Recently, a new method has been developed in which NMR and X-ray crystallographic data are used simultaneously to refine a protein structure (Shaanan et al., 1992). This method is of special interest for the determination of structures of multisubunit proteins. In these systems, X-ray data could provide a low resolution structure of the whole protein, while NMR could focus on higher resolution structures of individual subunits.

The ultimate goal of NMR investigations of proteins and of their complexes with ligands is the understanding of biological function. Structural data are essential for this goal since they provide the physical foundation on which questions pertaining to function can be addressed. However, the protein function also depends on the mobility of structural domains and on internal motions of individual amino acid side-chains. Measurements of NMR relaxation rates on [15]N- and/or ^{13}C-labeled proteins are currently opening new ways to obtain this information (Wagner et al.,

1993). One can anticipate that, at a not-too-distant time in the future, the NMR description of a protein will include the 3D structure as well as the overall and local dynamics of the molecule. At that time, the potential of nuclear magnetic resonance spectroscopy for protein studies will have been fully utilized.

REFERENCES

Archer, S. J., Ikura, M., Torchia, D. A., & Bax, A. (1991). An alternative 3D NMR technique for correlating backbone ^{15}N with side chain Hβ resonances in larger proteins. J. Magn. Reson. 95, 636-641.

Aue, W. P., Bartholdi, E., & Ernst, R. R. (1976). Two-dimensional spectroscopy. Application to nuclear magnetic resonance. J. Chem. Phys. 64, 2229-2246.

Bax, A., & Freeman, R. (1981). Investigation of complex networks of spin-spin coupling by two-dimensional NMR. J. Magn. Reson. 44, 542-561.

Bax, A., & Drobny, G. (1985). Optimization of two-dimensional homonuclear relayed coherence transfer NMR spectroscopy. J. Magn. Reson. 61, 306-320.

Bax, A., & Davis, D. G. (1985). MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. J. Magn. Reson. 65, 355-360.

Bax, A., & Ikura, M. (1991). An efficient 3D NMR technique for correlating the proton and ^{15}N backbone amide resonances with the α-carbon of the preceding residue in uniformly ^{15}N/^{13}C enriched proteins. J. Biomol. Nucl. Magn. Reson. 1, 99-104.

Bax, A., & Grzesieck, S. (1993). Methodological advances in protein NMR. Acc. Chem. Res. 26, [13]1-138.

Bax, A., Griffey, R. H., & Hawkins, B. L. (1983). Correlation of proton and Nitrogen-15 chemical shifts by multiple quantum NMR. J. Magn. Reson. 55, 301-315.

Bax, A., Clore, G. M., & Gronenborn, A. M. (1990a). [1]H-^{1}H correlation via isotropic mixing of ^{13}C magnetization, a new three-dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C-enriched proteins. J. Magn. Reson. 88, 425-431.

Bax, A., Ikura, M., Kay, L. E., Torchia, D. A., & Tschudin, R. (1990b). Comparison of different modes of two-dimensional reverse-correlation NMR for the study of proteins. J. Magn. Reson. 86, 304-318.

Bax, A., Max, D., & Zax, D. (1992). Measurement of long-range [13]C- [13]C J couplings in a 20-kDa protein-peptide complex. J. Am. Chem. Soc. 114, 6923-6925.

Bax, A., Vuister, G. W., Grzesiek, S., Delaglio, F., Wang, A. C., Tschudin, R., & Zhu, G. (1994). Measurement of homo- and heteronuclear J couplings from quantitative J correlation. Methods Enzymol. 239, 79-105.

Berndt, K. D., Güntert, P., Orbons, L. P. M., & Wüthrich, K. (1992). Determination of a high-quality nuclear magnetic resonance solution structure of the bovine pancreatic trypsin inhibitor and comparison with three crystal structures. J. Mol. Biol. 227, 757-775.

Billeter, M., Kline, A. D., Braun, W., Huber, R., & Wüthrich, K. (1989). Comparison of the high-resolution structures of the α-amylase inhibitor Tendamistat determined by nuclear magnetic resonance in solution and by X-ray diffraction in single crystals. J. Mol. Biol. 206, 677-687.

Billeter, M., Qian, Y.-q, Otting, G., Müller, M., Gehring, W. J., & Wüthrich, K. (1990). Determination of the three-dimensional structure of the *Antennapedia* homeodomain from *Drosophila* in solution by ^{1}H nuclear magnetic resonance spectroscopy. J. Mol. Biol. 214, 183-197.

Boelens, R., Konig, T. M. G., & Kaptein, R. (1988). Determination of biomolecular structures from proton-proton NOEs using a relaxation matrix approach. J. Mol. Struc. 173, 299-311.

Boelens, R., Vuister, G. W., Konig, T. M. G., & Kaptein, R. (1989). Observation of spin diffusion in biomolecules by three-dimensional NOE-NOE spectroscopy. J. Am. Chem. Soc., 111, 8525-8526.

Bolton, P. H., & Bodenhausen, G. (1982). Relayed coherence transfer spectroscopy of heteronuclear systems: Detection of remote nuclei in NMR. Chem. Phys. Lett. 89, 139-144.

Borgias, B. A., & James, T. L. (1988). COMATOSE, a method for restraint refinement of macromolecular structure based on two-dimensional nuclear Overhauser spectra. J. Magn. Reson. 79, 493-512.

Borgias, B. A., Gochin, M., Kerwood, D. J., & James, T. L. (1990). Relaxation matrix analysis of 2D NMR data. Prog. Nucl. Magn. Reson. Spectrosc. 22, 83-100.

Bothner-By, A. A., Stephens, R. L., Lee, J.-m., Warren, C. D., & Jeanloz, R. W. (1984). Structure determination of a tetrasaccharide: Transient nuclear Overhauser effects in the rotating frame. J. Am. Chem. Soc. 106, 811-813.

Braun, W., Epp, O., Wüthrich, K., & Huber, R. (1989). Solution of the phase problem in the X-ray diffraction method for proteins with nuclear magnetic resonance solution structure as initial model. Patterson search and refinement for the α-amylase inhibitor Tendamistat. J. Mol. Biol. 206, 669-676.

Braunschweiler, L., Bodenhausen, G., & Ernst, R. R. (1983). Analysis of networks of coupled spins by multiple quantum N.M.R. Mol. Phys. 48, 535-560.

Braunschweiler, L., & Ernst, R. R. (1983). Coherence transfer by isotropic mixing: application to protein correlation spectroscopy. J. Magn. Reson. 53, 521-528.

Breg, J. N., Boelens, R., George, A. V. E., & Kaptein, R. (1989). Sequence-specific ^{1}H NMR assignment and secondary structure of the Arc repressor of bacteriophage P22, as determined by two-dimensional ^{1}H NMR spectroscopy. Biochemistry 28, 9826-9833.

Brünger, A. T., & Nilges, M. (1993). Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR spectroscopy. Quart. Rev. Biophys. 26, 49-125.

Brüschweiler, R., & Case, D. A. (1994). Characterization of biomolecular structure and dynamics by NMR cross relaxation. Prog. Nucl. Magn. Reson. Spectrosc. 26, 27-58.

Bystrov, V. F. (1976). Spin-spin coupling and the conformational states of peptide systems. Prog. Nucl. Magn. Reson. Spectrosc. 10, 41-81.

Case, D. A., & Wright, P. E. (1993). Determination of high- resolution NMR structures of proteins. In: NMR of Proteins (Clore, G. M., & Gronenborn, A. M., eds.), pp.53-91, CRC Press, Boca Raton.

Chary, K. V. R., Otting, G., & Wüthrich, K. (1991). Measurement of small heteronuclear [1]H-^{15}N coupling constants in ^{15}N-labeled proteins by 3D $H_N NH_{AB}$-COSY. J. Magn. Reson. 93, 218-224.

Chazin, W. J., & Wright, P. E. (1988). Complete assignments of the ^{1}H nuclear magnetic resonance spectrum of French bean plastocyanin. Sequential resonance assignments, secondary structure and global fold. J. Mol. Biol. 202, 623-636.

Chazin, W. J., Rance, M., & Wright, P. E. (1988). Complete assignments of the ^{1}H nuclear magnetic resonance spectrum of French bean plastocyanin. Application of an integrated approach to spin system identification in proteins. J. Mol. Biol. 202, 603-622.

Clore, G. M., & Gronenborn, A. M. (1989). Determination of three- dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy. Crit. Rev. Biochem. Mol. Biol. 24, 479-564.

Clore, G. M., & Gronenborn, A. M. (1991a). Applications of three- and four-dimensional heteronuclear NMR spectroscopy to protein structure determination. Prog. Nucl. Magn. Reson. Spectrosc. 23, 43- 92.

Clore, G. M., & Gronenborn, A. M. (1991b). Comparison of the solution nuclear magnetic resonance and X-ray crystal structures of human recombinant interleukin 1β. J. Mol. Biol. 221, 47-53.

Clore, G. M., & Gronenborn, A. M. (1993). Determination of structures of larger proteins in solution by three- and four- dimensional heteronuclear magnetic resonance spectroscopy. In: NMR of Proteins (Clore, G. M., & Gronenborn, A. M., eds.) pp.1-32, CRC Press, Boca Raton.

Clore, G. M., Bax, A., Wingfield, P., & Gronenborn, A. M. (1988). Long-range ^{15}N-^{1}H correlation as an aid to sequential proton resonance assignments of proteins. Application to the DNA- binding protein *ner* from phage Mu. FEBS Lett. 238, 17-21.

Clore, G. M., Kay, L. E., Bax, A., & Gronenborn, A. M. (1991a). Four-dimensional ^{13}C/^{13}C-edited nuclear Overhauser enhancement spectroscopy of a protein in solution: Application to interleukin 1β. Biochemistry 30, 12-18.

Clore, G. M., Wingfield, P. T., & Gronenborn, A. M. (1991b). High- resolution, three-dimensional structure of interleukin-1β in solution by three- and four-dimensional nuclear magnetic resonance spectroscopy. Biochemistry 30, 2315-2323.

Clore, G. M., Robien, M. A., & Gronenborn, A. M. (1993). Exploring the limits of precision and accuracy of protein structures determined by nuclear magnetic resonance spectroscopy. J. Mol. Biol. 231, 82-102.

Clore, G. M., Omichinski, J. G., Sakaguchi, K., Zambrano, N., Sakamoto, H., Appella, E., & Gronenborn, A. M. (1994). High- resolution structure of the oligomerization domain of p53 by multidimensional NMR. Science 265, 386-391.

Clubb, R. T., Thanabal, V., & Wagner, G. (1992). A constant-time three-dimensional triple resonance pulse scheme to correlate intraresidue ^{1}H^{N}, ^{15}N and ^{13}C chemical shifts in ^{15}N-^{13}C labeled proteins. J. Magn. Reson. 97, 213-217.

Crippen, G. M., & Havel, T. F. (1988). Distance Geometry and Molecular Conformation. Research. Studies Press, Taunton, UK.

DeMarco, A., Llinas, M., & Wüthrich, K. (1978a). Analysis of the 1H-NMR spectra of ferrichrome peptides. I. The non-amide protons. Biopolymers 17, 617-636.

DeMarco, A., Llinas, M., & Wüthrich, K. (1978b). Analysis of the ^{1}H-NMR spectra of ferrichrome peptides. II. The amide resonances. Biopolymers 17, 2727-2742.

Dyson, H. J., Gippert, G. P., Case, D. A., Holmgren, A., & Wright, P. E. (1990). Three-dimensional solution structure of the reduced form of *Escherichia coli* thioredoxin determined by nuclear magnetic resonance spectroscopy. Biochemistry 29, 4129-4136.

Eich, G., Bodenhausen, G., & Ernst, R. R. (1982). Exploring nuclear spin systems by relayed magnetization transfer. J. Am. Chem. Soc. 104, 3731-3732.

Ernst, R. R., Bodenhausen, G., & Wokaun, A. (1987). Principles of Nuclear Magnetic Resonance in One and Two Dimensions. Oxford University Press, New York.

Fairbrother, W. J., Cavanagh, J., Dyson, H. J., Palmer, A. G., III, Sutrina, S. L., Reizer, J., Saier, M. H. Jr., & Wright, P. E. (1991). Polypeptide backbone resonance assignments and secondary structure of *Bacillus subtilis* enzyme III^{glc} determined by two-dimensional and three-dimensional heteronuclear NMR spectroscopy. Biochemistry 30, 6896-6907.

Fairbrother, W. J., Palmer, A. G., III, Rance, M., Reizer, J., Saier, M. H. Jr., & Wright, P. E. (1992). Assignment of the aliphatic ^{1}H and ^{13}C resonances of the *Bacillus subtilis* glucose permease IIA domain using double- and triple-resonance heteronuclear three-dimensional NMR spectroscopy. Biochemistry 31, 4413-4425.

Fogh, R. H., Schipper, D., Boelens, R., & Kaptein, R. (1994). ^{1}H, ^{13}C and ^{15}N NMR backbone assignments of the 269-residue serine protease PB92 from *Bacillus alcalophilus*. J. Biomol. Nucl. Magn. Reson. 4, 123-128.

Forman-Kay, J. D., Gronenborn, A. M., Kay, L. E., Wingfield, P. T., & Clore, G. M. (1990). Studies of the solution conformation of human thioredoxin using heteronuclear ^{15}N-^{1}H nuclear magnetic resonance spectroscopy. Biochemistry 29, 1566-1572.

Garrett, D. S., Lodi, P. J., Shamoo, Y., Williams, K. R., Clore, G. M., & Gronenborn, A. M. (1994). Determination of the secondary structure and folding topology of an RNA binding domain of

mammalian hnRNP A1 protein using three-dimensional heteronuclear magnetic resonance spectroscopy. Biochemistry 33, 2852-2858.

Griesinger, G., Sorensen, O. W., & Ernst, R. R. (1986). Correlation of connected transitions by two-dimensional NMR spectroscopy. J. Chem. Phys. 85, 6837-6852.

Griesinger, C., Sorensen, O. W., & Ernst, R. R. (1987). A practical approach to three-dimensional NMR spectroscopy. J. Magn. Reson. 73, 574-579.

Griesinger, C., Sorensen, O. W., & Ernst, R. R. (1989). Three- dimensional Fourier spectroscopy. Application to high-resolution NMR. J. Magn. Reson. 84, 14-63.

Griffey, R. H., & Redfield, A. G. (1987). Proton-detected heteronuclear edited and correlated nuclear magnetic resonance and nuclear Overhauser effect in solution. Quart. Rev. Biophys. 19, 51-82.

Grzesiek, S., & Bax, A. (1993). Amino acid type determination in the sequential assignment process of uniformly $^{13}C/^{15}N$-enriched proteins. J. Biomol. Nucl. Magn. Reson. 3, 185-204.

Grzesiek, S., Ikura, M., Clore, G, M., Gronenborn, A. M., & Bax, A. (1992). A 3D triple-resonance NMR technique for qualitative measurement of carbonyl-Hβ J couplings in isotopically enriched proteins. J. Magn. Reson. 96, 215-221.

Grzesiek, S., Vuister, G. W., & Bax, A. (1993). A simple and sensitive experiment for measurement of J_{CC} couplings between backbone carbonyl and methyl carbons in isotopically enriched proteins. J. Biomol. Nucl. Magn. Reson. 3, 487-493.

Gueron, M., Plateau, P., & Decorps, M. (1991). Solvent signal supression in NMR. Prog. Nucl. Magn. Reson. Spectrosc. 23, 135-209.

Güntert, P., Braun, W., Billeter, M., & Wüthrich, K. (1989). Automated stereospecific ^1H NMR assignments and their impact on the precision of protein structure determinations in solution. J. Am. Chem. Soc. 111, 3997-4004.

Habazettl, J., Schleicher, M., Otlewski, J., & Holak, T. A. (1992). Homonuclear three-dimensional NOE-NOE nuclear magnetic resonance spectra for structure determination of proteins in solution. J. Mol. Biol. 228, 156-169.

Havel, T. F. (1991). An evaluation of computational strategies for use in the determination of protein structure from distance constraints obtained by nuclear magnetic resonance. Prog. Biophys. Mol. Biol. 56, 43-78.

Hyberts, S. G., & Wagner, G. (1990). Sequence-specific ^1H NMR assignments and secondary structure of eglin c. Biochemistry 29, 1465-1474.

Hyberts, S. G., Marki, W., & Wagner, G. (1987). Stereospecific assignments of side-chain protons and characterization of torsion angles in eglin c. Eur. J. Biochem. 164, 625-635.

Ikura, M., Kay, L. E., & Bax, A. (1990). A novel approach for sequential assignment of ^1H, ^{13}C, and ^{15}N spectra of larger proteins: Heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. Biochemistry 29, 4659-4667.

Ikura, M., Clore, G. M., Gronenborn, A. M., Zhu, G., Klee, C. B., & Bax, A. (1992). Solution structure of a calmodulin-target peptide complex by multidimensional NMR. Science 256, 632- 638.

Jardetzky, O., & Roberts, G. C. K. (1981). NMR in Molecular Biology. Academic Press, New York.

Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R., & van Gunsteren, W. F. (1985). A protein structure from nuclear magnetic resonance data. Lac repressor headpiece. J. Mol. Biol. 182, 179-182.

Karplus, M. (1959). Contact electron-spin coupling of nuclear magnetic moments. J. Chem. Phys. 30, 11-15.

Karplus, M. (1963). Vicinal proton coupling in nuclear magnetic resonance. J. Am. Chem. Soc. 85, 2870-2871.

Kay, L. E., & Bax, A. (1990). New methods for the measurement of NH-CαH coupling constants in ^{15}N-labeled proteins. J. Magn. Reson. 86, 110-126.

Kay, L. E., Ikura, M., Tschudin, R., & Bax, A. (1990a). Three- dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. J. Magn. Reson. 89, 496-514.

Kay, L. E., Clore, G. M., Bax, A., & Gronenborn, A. M. (1990b). Four-dimensional heteronuclear triple-resonance NMR spectroscopy of interleukin-1β in solution. Science 249, 411- 414.

Kay, L. E., Ikura, M., & Bax, A. (1990c). Proton-proton correlation via carbon-carbon couplings: A three-dimensional NMR approach for the assignment of aliphatic resonances in proteins labeled with carbon-13. J. Am. Chem. Soc. 112, 888- 889.

Kay, L. E., Ikura, M., & Bax, A. (1991). The design and optimization of complex NMR experiments. Application to a triple-resonance pulse scheme correlating Hα, NH, and ^{15}N chemical shifts in ^{15}N-^{13}C labeled proteins. J. Magn. Reson. 91, 84-92.

Kessler, H., Seip, S., & Saulitis, J. (1991). Determination of cross-relaxation rates by 3D NOE-NOE spectroscopy. J. Biomol. Nucl. Magn. Reson. 1, 83-92.

King, G., & Wright, P. E. (1983). Application of two-dimensional relayed coherence transfer experiments to ^{1}H NMR studies of macromolecules. J. Magn. Reson. 54, 328-332.

Kuszewski, J., Qin, J., Gronenborn, A. M., & Clore, G. M. (1995a). The impact of direct refinement against ^{13}Cα and ^{13}Cβ chemical shifts on protein structure determination by NMR. J. Mag. Reson. B 106, 92-96.

Kuszewski, J., Gronenborn, A. M., & Clore, G. M. (1995b). The impact of direct refinement against proton chemical shifts on protein structure determination by NMR. J. Mag. Reson. B 107, 293-297.

Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A., & Wright, P. E. (1989). Three-dimensional solution structure of a single zinc finger DNA-binding domain. Science 245, 635-637.

Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A., & Wright, P. E. (1990). Proton nuclear magnetic resonance assignments and solution structure of a synthetic zinc finger from *Xfin*. In: Structure and Methods, Vol. 2: DNA Protein Complexes and Proteins (Sarma, R. H. & Sarma, M. H., eds.), pp. 83-91. Adenine Press, Albany.

Lodi, P. J., Ernst, J. A., Kuszewski, J., Hickman, A. B., Engelman, A., Craigie, R., Clore, G. M., & Gronenborn, A. M. (1995). Solution structure of the DNA binding domain of HIV-1 integrase. Biochemistry 34, 9826-9833.

Macura, S., & Ernst, R. R. (1980). Elucidation of cross relaxation in liquids by two-dimensional N.M.R. spectroscopy. Mol. Phys. 41, 95-117.

Mareci, T. H., & Freeman, R. (1983). Mapping proton-proton coupling via double-quantun coherence. J. Magn. Reson. 51, 531-535.

Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., & Clore, G. M. (1989). Overcoming the overlap problem in the assignment of ^{1}H NMR spectra of larger proteins by use of three-dimensional heteronuclear ^{1}H-^{15}N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: Application to interleukin 1β. Biochemistry 28, 6150-6156.

Markley, J. L. (1989). Two-dimensional nuclear magnetic resonance spectroscopy of proteins: An overview. Methods Enzymol. 176, 12-64.

McIntosh, L. P., Wand, A. J., Lowry, D. F., Redfield, A. G., & Dahlquist, F. W. (1990). Assignment of the backbone ^{1}H and ^{15}N NMR resonances of bacteriophage T4 lysozyme. Biochemistry 29, 6341-6362.

McIntosh, L. P., & Dahlquist, F. W. (1990). Biosynthetic incorporation of ^{15}N and ^{13}C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. Quart. Rev. Biophys. 23, 1-38.

Moore, J. M., Lepre, C., Gippert, G. P., Chazin, W. J., Case, D. A., & Wright, P. E. (1991). High-resolution solution structure of reduced French bean plastocyanin and comparison with the crystal structure of poplar plastocyanin. J. Mol. Biol. 221, 533-555.

Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E., & Dahlquist, F. W. (1989). Expression and Nitrogen-15 labeling of proteins for proton and Nitrogen-15 nuclear magnetic resonance. Methods Enzymol. 177, 44-73.

Müller, N., Ernst, R. R., & Wüthrich, K. (1986). Multiple-quantum- filtered two-dimensional correlated NMR spectroscopy of proteins. J. Am. Chem. Soc. 108, 6482-6492.

Nagayama, K., Wüthrich, K., & Ernst, R. R. (1979). Two-dimensional spin echo correlated spectroscopy (SECSY) for ^{1}H NMR studies of biological macromolecules. Biochem. Biophys. Res. Commun. 90, 305-311.

Nagayama, K., Kumar, A., Wüthrich, K., & Ernst, R. R. (1980). Experimental techniques of two-dimensional correlated spectroscopy. J. Magn. Reson. 40, 321-334.

Neri, D., Otting, G., & Wüthrich, K. (1990). New nuclear magnetic resonance experiment for measurements of the vicinal coupling constants $^{3}J_{HN\alpha}$ in proteins. J. Am. Chem. Soc. 112, 3663-3665.

Neri, D., Billeter, M., & Wüthrich, K. (1992). Determination of the nuclear magnetic resonance solution structure of the DNA- binding domain (residues 1 to 69) of the 434 repressor and comparison with the X-ray crystal structure. J. Mol. Biol. 223, 743-767.

Nilges, M., Gronenborn, A. M., Brünger, A. T., & Clore, G. M. (1988a). Determination of three-dimensional structure of proteins by simulated annealing with interproton distance restraints: Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. Protein Eng. 2, 27-38.

Nilges, M., Clore, G. M., & Gronenborn, A. M. (1988b). Determination of three-dimensional structure of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms. FEBS Lett. 239, 129-136.

Nilges, M., Clore, G. M., & Gronenborn, A. M. (1990). ^{1}H-NMR stereospecific assignments by conformational data base searches. Biopolymers 29, 813-822.

Ogata, K., Hojo, H., Aimoto, S., Nakai, T., Nakamura, H., Sarai, A., Ishii, S., & Nishimura, Y. (1992). Solution structure of a DNA-binding unit of Myb: A helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core. Proc. Natl. Acad. Sci. USA 89, 6428-6432.

Omichinski, J. G., Clore, G. M., Robien, M., Sakaguchi, K., Appella, E., & Gronenborn, A. M. (1992). High-resolution solution structure of the double Cys_2His_2 zinc finger from the human enhancer binding protein MBP-1. Biochemistry 31, 3907- 3917.

Omichinski, J. G., Clore, G. M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S. J., & Gronenborn, A. M. (1993). NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. Science 261, 438-446.

Ösapay, K., & Case, D. A. (1991). A new analysis of proton chemical shifts in proteins. J. Am. Chem. Soc. 113, 9436-9444.

Ösapay, K., Theriault, Y., Wright, P. E., & Case, D. A. (1994). Solution structure of carbonmonoxy myoglobin determined from nuclear magnetic resonance distance and chemical shift constraints. J. Mol. Biol. 244, 183-197.

Oschkinat, H., Griesinger, C., Kraulis, P. J., Sorensen, O. W., Ernst, R. R., Gronenborn, A. M., & Clore, G. M. (1988). Three-dimensional NMR spectroscopy of a protein in solution. Nature (London) 332, 374-376.

Oschkinat, H., Cieslar, C., Gronenborn, A. M., & Clore, G. M. (1989). Three-dimensional homonuclear Hartmann-Hahn-nuclear Overhauser enhancement spectroscopy in H_2O and its application to proteins. J. Magn. Reson. 81, 212-216.

Otting, G., Senn, H., Wagner, G., & Wüthrich, K. (1986). Editing of 2D ^{1}H NMR spectra using X half-filters. Combined use with residue-selective ^{15}N labeling of proteins. J. Magn. Reson. 70, 500-505.

Pardi, A., Billeter, M., & Wüthrich, K. (1984). Calibration of the angular dependence of the amide proton-C^{α} proton coupling constants, $^{3}J_{HN\alpha}$, in a globular protein. Use of $^{3}J_{HN\alpha}$ for identification of helical secondary structure. J. Mol. Biol. 180, 741-751.

Pavletich, N. P., & Pabo, C. O. (1991). Zinc finger-DNA recognition: Crystal structure of a Zif268-DNA complex at 2.1 Å. Science 252, 809-817.

Piantini, U., Sorensen, O. W., & Ernst, R. R. (1982). Multiple quantum filters for elucidating NMR coupling networks. J. Am. Chem. Soc. 104, 6800-6801.

Powers, R., Garrett, D. S., March, C. J., Frieden, E. A., Gronenborn, A. M., & Clore, G. M. (1992). ^{1}H, ^{15}N, ^{13}C, and ^{13}CO assignments of human interleukin-4 using three-dimensional double- and triple-resonance heteronuclear magnetic resonance spectroscopy. Biochemistry 31, 4334-4346.

Powers, R., Garrett, D. S., March, C. J., Frieden, E. A., Gronenborn, A. M., & Clore, G. M. (1993). The high-resolution, three-dimensional solution structure of human interleukin-4 determined by multidimensional heteronuclear magnetic resonance spectroscopy. Biochemistry 32, 6744-6762.

Qin, J., Clore, G. M., & Gronenborn, A. M. (1994). The high-resolution three-dimensional solution structures of the oxidized and reduced states of human thioredoxin. Structure 2, 503-522.

Rance, M., & Wright, P. E. (1986). Analysis of ^{1}H NMR spectra of proteins using multiple-quantum coherence. J. Magn. Reson. 66, 372-378.

Redfield, C., & Dobson, C. M. (1988). Sequential ^{1}H NMR assignments and secondary structure of hen egg white lysozyme in solution. Biochemistry 27, 122-136.

Remerowski, M. L., Domke, T., Groenewegen, A., Pepermans, H. A. M., Hilberts, C. W., & van de Ven, F. J. M. (1994). ^{1}H, ^{13}C and ^{15}N NMR backbone assignments and secondary structure of the 269-residue protease subtilisin 309 from *Bacillus lentus*. J. Biomol. Nucl. Magn.Reson. 4, 257-278.

Robertson, A. D., Purisima, E. O., Eastman, M. A., & Scheraga, H. A. (1989). Proton NMR assignments and regular backbone structure of bovine pancreatic ribonuclease A in aqueous solution. Biochemistry 28, 5930-5938.

Shaanan, B., Gronenborn, A. M., Cohen, G. H., Gilliland, G. L., Veerapandian, B., Davies, D. R., & Clore, G. M. (1992). Combining experimental information from crystal and solution studies: Joint X-ray and NMR refinement. Nature (London) 257, 961-964.

Solomon, I. (1955). Relaxation processes in a system of two spins. Phys. Rev. 99, 559-565.

Spera, S., & Bax, A. (1991). Empirical correlation between protein backbone conformation and Cα and Cβ ^{13}C nuclear magnetic resonance chemical shifts. J. Am. Chem. Soc. 113, 5490-5492.

Stockman, B. J., Westler, W. M., Mooberry, E. S., & Markley, J. L. (1988). Flavodoxin from *Anabaena* 7120: Uniform nitrogen-15 enrichment and hydrogen-1, nitrogen-15, and phosphorus-31 NMR investigation of the flavin mononucleotide binding site in the reduced and oxidized states. Biochemistry 27, 136-142.

Stockman, B. J., Reily, M. D., Westler, W. M., Ulrich, E. L., & Markley, J. L. (1989). Concerted two-dimensional NMR approaches to hydrogen-1, carbon-13 and nitrogen-15 resonance assignments in proteins. Biochemistry 28, 230-236.

Torchia, D. A., Sparks, S. W., & Bax, A. (1989). Staphylococcal nuclease: Sequential assignments and solution structure. Biochemistry 28, 5509-5524.

Ulrich, E. L., Markley, J. L., & Kyogoku, Y. (1989). Creation of a nuclear magnetic resonance data repository and literature database. Protein Seq. Data Anal. 2, 23-37.

Vis, H., Boelens, R., Mariani, M., Stroop, R., Vorgias, C. E., Wilson, K. S., & Kaptein, R. (1994). ^{1}H, ^{13}C, and ^{15}N resonance assignments and secondary structure analysis of the HU protein from *Bacillus stearothermophilus* using two- and three-dimensional double- and triple-resonance heteronuclear magnetic resonance spectroscopy. Biochemistry 33, 14858-14870.

Vuister, G. W., Boelens, R., & Kaptein, R. (1988). Nonselective three-dimensional NMR spectroscopy. The 3D NOE-HOHAHA experiment. J. Magn. Reson. 80, 176-185.

Vuister, G. W., & Bax, A. (1993a). Quantitative J correlation: A new approach for measuring homonuclear three-bond J(H^{N}H$^\alpha$) coupling constants in ^{15}N-enriched proteins. J. Am. Chem. Soc. 115, 7772-7777.

Vuister, G. W., & Bax, A. (1993b). Measurement of two- and three-bond proton to methyl-carbon J couplings in proteins uniformly enriched with ^{13}C. J. Magn. Reson. B 102, 228-231.

Vuister, G. W., Yamazaki, T., Torchia, D. A., & Bax, A. (1993a). Measurement of two- and three-bond ^{13}C-^1H J couplings to the C_δ carbons of leucine residues in staphylococcal nuclease. J. Biomol. Nucl. Magn. Reson. 3, 297-306.

Vuister, G. W., Wang, A. C., & Bax, A. (1993b). Measurement of three-bond nitrogen-carbon J couplings in proteins uniformly enriched in ^{15}N and ^{13}C. J. Am. Chem. Soc. 115, 5334-5335.

Wagner, G. (1983). Two-dimensional relayed coherence transfer spectroscopy of a protein. J. Magn. Reson. 55, 151-156.

Wagner, G. (1990). NMR investigations of protein structure. Prog. Nucl. Magn. Reson. Spectrosc. 22, 101-139.

Wagner, G. (1993). Prospects for NMR of large proteins. J. Biomol. Nucl. Magn. Reson. 3, 375-385.

Wagner, G., Schmieder, P., & Thanabal, V. (1991). A new ^1H-^{15}N-^{13}C triple-resonance experiment for sequential assignments and measuring homonuclear H^α-H^N vicinal coupling constants in polypeptides. J. Magn. Reson. 93, 436-440.

Wagner, G., Hyberts, S., & Peng, J. W. (1993). Study of protein dynamics by NMR. In: NMR of Proteins (Clore, G. M., & Gronenborn, A. M., eds.), pp.220-257, CRC Press, Boca Raton.

Williamson, M. P., Havel, T.F., & Wüthrich, K. (1985). Solution conformation of proteinase inhibitor IA for bull seminal plasma by ^1H nuclear magnetic resonance and distance geometry. J. Mol. Biol. 182, 295-315.

Williamson, M. P., & Asakura, T. (1993). Empirical comparisons of models for chemical-shift calculation in proteins. J. Magn. Reson. B 101, 63-71.

Wishart, D. S., Sykes, B. D., & Richards, F. M. (1991). Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. J. Mol. Biol. 222, 311-333.

Wüthrich, K. (1986). In: NMR of Proteins and Nucleic Acids. Wiley & Sons, New York.

Wüthrich, K. (1989). Protein structure determination in solution by nuclear magnetic resonance spectroscopy. Science 243, 45- 50.

Wüthrich, K., Billeter, M., & Braun, W. (1983). Pseudo-structures for the 20 common amino acids for use in studies of protein conformation by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance. J. Mol. Biol. 169, 949-961.

Yip, P., & Case, D. A. (1991). A new method for refinement of macromolecular structures based on nuclear Overhauser effect spectra. J. Magn. Reson. 83, 643-648.

Zuiderweg, E. R. P., Boelens, R., & Kaptein, R. (1985). Stereospecific assignments of ^1H-NMR methyl lines and conformation of Valyl residues in the *lac* repressor headpiece. Biopolymers 24, 601-611.

This Page Intentionally Left Blank

MEMBRANE PROTEIN STRUCTURE

Michael P. McCarthy

I. INTRODUCTION

Membrane proteins are the class of proteins which interact directly with the intra-and extracellular membranes of cells. There are two subclasses of membrane proteins; intrinsic and extrinsic (Capaldi and Vanderkooi, 1972; Singer and

Nicholson, 1972). Intrinsic membrane proteins are inserted to some degree within the membrane lipid bilayer and are functionally defined as those proteins which can be removed from the lipid bilayer only through the use of detergents or harsh solvents. Extrinsic membrane proteins interact only with the surface of membranes, typically via ionic interactions with the charged head groups of membranous phospholipids or with intrinsic membrane proteins, and can be separated from the lipid bilayer by altering the pH or ionic composition of the medium. This review will only consider intrinsic membrane proteins (although, as will be discussed in greater detail below, the boundary between these two classifications of membrane proteins is becoming less distinct). Intrinsic membrane proteins are crucial to the energy-generating machinery of the cell, to communication between cells (including the actual physical linkages which unite cells into larger structures), to maintaining the basic structure of the cell, and to the transport of substances across both intra- and extracellular compartments. Some proteins are membrane-bound merely to localize their function near the surface of a membrane, while others perform an enzymatic activity which must take place within the interior of the lipid bilayer, such as generating or breaking-down lipid soluble compounds. The importance of membrane proteins to cellular metabolism, import and export, signaling and interaction with other cells and the outside environment, specialized enzymatic functions, and linking the cytoskeleton to the membrane skeletons of cells, explain why the structures of membrane proteins are a topic of such interest. Basically, the hope is that learning more about the structure of membrane proteins will allow us to understand better how these proteins are able to perform their unique functions.

A review of the structure of intrinsic membrane proteins is limited by the sparse number of high-resolution examples. As of this writing (Fall 1995), the structures of only four types of intrinsic membrane proteins have been determined at or near atomic-level resolution by X-ray diffraction techniques: the photosynthetic reaction centers of various species of photosynthetic bacteria (Deisenhofer et al., 1985; Allen et al., 1987), several related channel-forming porins of bacterial outer membranes (Weiss and Schulz, 1992; Cowan, 1993; Kreusch et al., 1994; Schirmer et al., 1995), prostaglandin H synthase (Picot et al., 1994), and the LH2 light-harvesting antenna complex of photosynthetic bacteria (McDermott et al., 1995), although the photosystem I structure, at a nominal resolution of 6 Å, is close (Krauss et al., 1993). However, while intrinsic membrane proteins have proved challenging subjects for X-ray crystallographic analyses, their propensity to form large, well-ordered two-dimensional arrays in lipid bilayers has been exploited to determine the high-resolution structure of bacteriorhodopsin (Henderson et al., 1990) and the plant light-harvesting chlorophyll *a/b* protein(LHC-II) complex II (Kühlbrandt et al., 1994) by electron microscopic image reconstruction techniques, bringing the total of known intrinsic membrane protein structures to six. Finally, many membrane proteins are multidomain structures, and in several cases the structures of important soluble domains of membrane proteins have been determined by X-ray crystal-

lographic techniques, providing valuable insights into at least part of the function of the membrane protein (Lange et al., 1994; Martinez et al., 1994).

In this review, the basic features of membrane proteins will be discussed in terms of classic protein structural hierarchy (primary, secondary, etc.), following a brief description of membrane protein topological models which are useful for comparing membrane protein structures at all levels. This will be followed by a limited descriptive analysis of the membrane proteins whose structures have been determined at adequate resolution, including some consideration of the techniques utilized. Earlier reviews of membrane protein structure include Eisenberg (1984), Unwin and Henderson (1984), Pattus (1990), McCarthy (1992), Reithmeier and Deber (1992), and White (1994).

II.   LEVELS OF MEMBRANE PROTEIN STRUCTURE

In this section, different aspects of membrane protein structure will be considered. First, the topology of membrane proteins, which refers to the pattern in which different parts of the protein are disposed relative to the membrane. Second, the primary structure of membrane proteins, with emphasis on unique features of membrane protein amino acid sequences and their known or predicted structure and topology. Third, the secondary structure of membrane proteins, particularly within the membrane-embedded regions, where the low dielectric of the bilayer interior, at least in theory, constrains the number of possible structures. Fourth, membrane protein tertiary structure, with emphasis on the idiosyncratic features of membrane protein tertiary structure. Fifth, and finally, the quaternary structure of membrane proteins, including both the complex oligomeric structure of many membrane proteins, and how changes in subunit interactions and state of oligomerization are important in signalling mechanisms.

A.   Topology

When discussing intrinsic membrane proteins, it is often useful to consider the manner in which they thread through the lipid bilayer. Knowledge of the number of membrane-spanning regions, the disposition of the amino- and carboxy-termini, and the positioning of important functional domains relative to the membrane is important for several reasons, including the classification membrane proteins within related types, the process of folding and assembly of membrane proteins, and the role membrane proteins play within the life of the cell. This information, particularly estimates of the number of membrane-spanning regions and determination of which portions of a membrane protein are extracellular, buried within the bilayer, or exposed to a particular cellular compartment, is often controversial due to the limitations of the determinative methods involved (Jennings, 1989). However, while the proposed topography of any one membrane protein may be somewhat or

entirely speculative, at the same time it is almost impossible to discuss that membrane protein without mentioning its predicted and/or partially-determined topography. In this review we will mention the predicted topology of several families of membrane proteins, but it should be kept in mind that in the absence of high resolution structural studies, or a large collection of convincing and consistent lower resolution information, these topographies are hypothetical models which remain to be proven or discarded.

The simplest classification of intrinsic membrane protein topologies is probably that of Blobel (Blobel, 1980). In this scheme, proteins which only dip into the bilayer but do not span it completely are termed monotopic, proteins which span the bilayer once and expose their amino- and carboxy-termini on opposite sides of the bilayer are termed bitopic, and proteins which span the bilayer multiple times are termed polytopic (Figure 1). A considerable number of small, "membrane-active" (e.g., channel-forming, fusogenic, and/or variably membrane-bound or soluble) peptides appear to be monotopic (Kaiser and Kezdy, 1987). However, large monotopic membrane proteins are relatively rare: cytochrome b_5 appears to be anchored to the membrane by a hairpin hydrophobic tail, at least under some circumstances, and pyruvate oxidase has a single amphipathic carboxy-terminal domain which may bind it to the membrane (Jennings, 1989; Picot et al., 1994). The most exciting example of a large, monotopic protein is prostaglandin H synthase, the structure of which has recently been determined to near atomic resolution by X-ray crystallographic techniques (Picot et al., 1994). As will be discussed in greater detail below, prostaglandin H synthase appears to be an intrinsic membrane protein, on the basis of biochemical studies, but its primary sequence is not characterized by the stretches of hydrophobic amino acids typical of intrinsic membrane proteins (Picot and Garavito, 1994). Instead, prostaglandin H synthase appears to be linked to the membrane by several amphipathic α-helices (Figure 9),
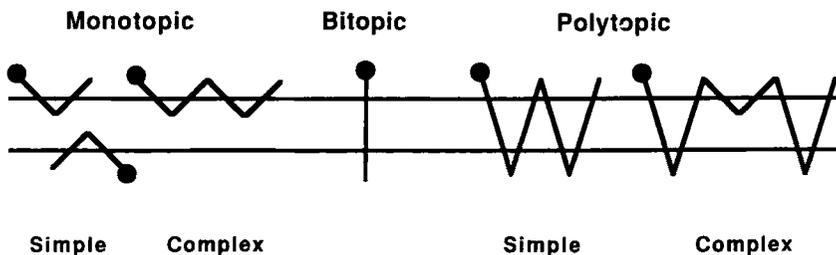
Figure 1. Classes of intrinsic membrane proteins, as defined by Blobel (1980). The lipid bilayer is indicated by the two horizontal lines. Monotopic proteins dip into, but do not span the bilayer, bitopic proteins span the bilayer only once, and polytopic proteins span the bilayer multiple times (n ≥ 2). In addition, complex versions of monotopic (with multiple membrane interactions) and polytopic (with mixed monotopic and polytopic features) proteins are indicated.

which presumably interact with both the hydrophobic interior of the bilayer as well as more polar groups at the lipid-solvent interface. To accomodate this new structural type, we have modified the original classification of Blobel (1980) to include a linked monotopic form, in which individual monotopic domains are connected by extramembranous regions of the protein (Figure 1).

Bitopic proteins are very common, and their functions are varied. In some cases, the single membrane spanning region appears merely to lock a functional domain to a particular face of a membrane. Examples include membrane-bound enzymes (guanylate cyclase, aminopeptidase, sucrase-isomaltase), antigen presentation structures (viral spike proteins, glycophorin, histocompatibility markers), and simple receptor/binding proteins (mannose 6-phosphate receptor, asialoglycoprotein receptor, transferrin receptor) (Popot and deVitry, 1990; Reithmeier and Deber, 1992). However, many bitopic proteins carry out more complex signaling functions, linking some signal on one side of a membrane to a biological response on the other. Examples include cell surface growth factor/tyrosine kinase receptors (e.g., epidermal growth factor (EGF) receptors, platelet-derived growth factor (PDGF) receptors, trk receptors, etc.) (Schlessinger and Ullrich, 1992) and integrins (e.g., (N-CAM, laminin, fibronectin, etc.) (Hynes, 1992). High resolution structures of bitopic proteins include the H subunit of the photosynthetic reaction center (Deisenhofer et al., 1984; Allen et al., 1987), and the α and β apoproteins of the LH2 complex (McDermott et al., 1995).

Polytopic membrane proteins contain two or more membrane-spanning segments. The number of membrane-spanning domains found in polytopic proteins varies; in "known" structures, the plant light-harvesting complex (LHC-II) monomer contains three (Kühlbrandt, 1994), the L and M subunits of the photosynthetic reaction center each contain five (Deisenhofer et al., 1984; Allen et al., 1987), bacteriorhodopsin contains seven (Henderson et al., 1990), and the bacterial porins contain 16-18 membrane-spanning regions (Cowan et al., 1992; Schirmer et al., 1995), while some membrane proteins, such as the sodium and calcium channel, are predicted to contain 24 or more membrane-spanning domains (Catterall, 1994). The polytopic membrane proteins are found in all types of membranes, and are primarily involved with various types of transport functions. In addition, some membrane proteins have been suggested to be composed of a mixed monotopic/polytopic structure (examples include the nicotinic acetylcholine receptor (Ratnam et al., 1986) and the family of voltage-gated ion channels (Guy and Conti, 1990; Lipkind and Fozzard, 1994)). Accordingly, this structural motif is included in Figure 1, also first suggested by Blobel, though in a different context (Blobel, 1980). As a generalization, the population of plasma membrane proteins appears evenly mixed between bitopic and polytopic proteins. Interestingly, however, some organelle membranes, in particular the inner membranes of mitochondria and chloroplasts, appear to contain a high percentage of very small bitopic proteins, many of which are little more than a single membrane-spanning region (Popot and de Vitry, 1990).

Additional classifications of membrane protein topology have also been proposed, largely in terms of the sidedness of the amino- and carboxy-termini of bitopic and polytopic proteins. The disposition of the termini relative to cytoplasmic or extracytoplasmic face of the lipid bilayer is important to the process of membrane protein insertion and translocation, reflecting in part the presence or absence of cleavable signal sequences. The most common classification, depicted in Figure 2, terms bitopic proteins with internal carboxy termini as Type I (probably the most populous form of eukaryotic membrane protein), bitopic proteins with internal amino termini as Type II, and polytopic membrane proteins as Type III. As indicated in the figure, the amino- and carboxy-termini of Type III proteins with an odd number of membrane crossings will be on opposite sides of the bilayer, while those with an even number of crossings will be on the same side of the bilayer.

B.   Primary Structure

Hydrophobicity

Like soluble proteins, the folded structures of membrane proteins are the product of the primary amino acid sequence of the protein, and the manner in which these component amino acids interact with each other and the surrounding milieu. The overall primary structures of membrane proteins differ from soluble proteins in several ways, however, in response to the demands of their peculiar environment. The first and most obvious difference, as noted early on (Capaldi and Vanderkooi, 1972; Singer and Nicholson, 1972), is that in general intrinsic membrane proteins are composed of a higher proportion of apolar amino acids than are soluble proteins.
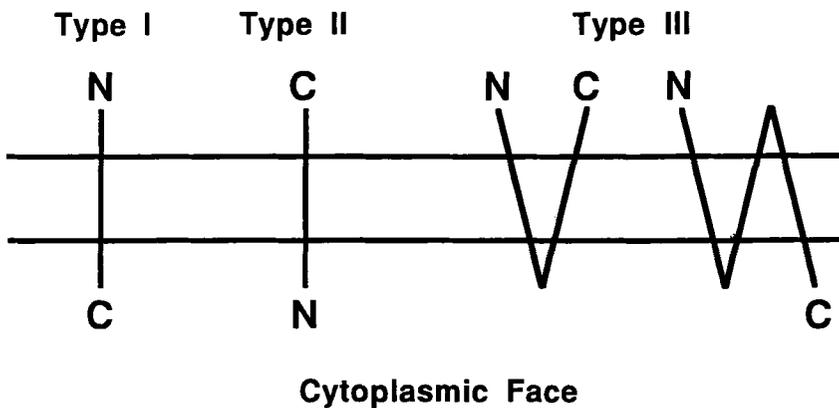

Figure 2.   Classification of membrane-spanning proteins into three types, derived from Singer, 1990. The number of membrane-spanning domains of Type III proteins can be greater than those indicated in the figure.

This is important for thermodynamic reasons, as amino acids with apolar side chains are more easily accomodated in the low dielectric interior of the lipid bilayer than amino acids with polar or charged side chains. The relative polarity of amino acids and their constituent side chains has been determined in a number of ways, primarily by directly measuring the free energy of transfer of some amino acids from a more polar to a less polar solvent or by comparison to the population of amino acids which are buried within the hydrophobic interior of soluble proteins, as well as more theoretical approaches (reviewed in Eisenberg, 1984; Engelman et al., 1986; Popot and Engelman, 1990; Reithmeier and Deber, 1992; White and Wimley, 1994). Summarizing between these different methods, isoleucine, valine, leucine, phenylalanine, cysteine, methionine and alanine are the most hydrophobic (in that order) and predicted most likely to be found in the internal membrane-spanning sequences of intrinsic membrane proteins, while unsurprisingly the charged amino acids are the least likely, with arginine by far the least preferred (Reithmeier and Deber, 1992). With some slight shuffling of the order, the same list of hydrophobic amino acids predominates the interior membrane-spanning sequences of most of the membrane proteins whose structures have been determined at sufficient resolution (cf., Rees et al., 1989; Reithmeier and Deber, 1992; Kühlbrandt, 1994; McDermott et al., 1995).

However, the generality that intrinsic membranes are less polar, on average, than soluble proteins does have exceptions, as demonstrated by two of the high resolution structures discussed in greater detail below. First, some members of the porin family of bacterial outer membrane proteins are not exceptionally hydrophobic in amino acid composition (Jähnig, 1990; Popot and de Vitry, 1990; Cowan, 1993). In particular, selective porins such as maltoporin (Schirmer et al., 1995), even though composed largely of membrane-spanning segments with relatively little extramembranous mass, are able to overcome the usual requirement for hydrophobic amino acids due to their novel construction in the form of a water-filled β-barrel (see below), in which only every other amino acid of the short membrane-spanning segments must face the hydrophobic interior of the bilayer. Second, for membrane proteins such prostaglandin H synthase or the H subunit of the photosynthetic reaction center, in which the membrane-spanning segments represent only a small proportion of the total mass of the protein, the contribution of the amino acids in the membrane-spanning regions is not sufficient to make the overall hydrophobicity of the protein noticeably different from that of a soluble protein. In the case of the monotopic type proteins, exemplified by prostaglandin H synthase (Garavito et al., 1994), it is not surprising that the limited degree of membrane-embedded structure has little impact on the overall hydrophobicity of the amino acid composition. However, the overall polarity of number of bitopic (such as many of the growth factor receptors (Schlessinger and Ulrich, 1992), integrins (Hynes, 1992), or the large form of the mannose 6-phosphate receptor) and even polytopic proteins (perhaps the extreme example is the ryanodine receptor (Takeshima et al., 1989)) is predominated by their extensive, more hydrophilic extramembranous regions.

An empirical demonstration of this fact is provided by the detergent Triton X-114, which will readily solubilize both intrinsic and extrinsic membrane proteins. However, the unique feature of this detergent is that upon raising the temperature to the detergent "cloud point," the solubilization mixture will separate into a detergent-poor (upper) phase, and a detergent-rich (lower) phase. It was anticipated that this feature could be used to differentiate between intrinsic membrane proteins and other proteins (with extrinsic membrane proteins and soluble proteins partitioning into the detergent-poor phase and intrinsic membrane into the detergent-rich phase) (Bordier, 1981), but it soon became apparent that some perfectly legitimate polytopic intrinsic membrane proteins also partitioned into the detergent-poor phase (Maher and Singer, 1985; Pryde, 1986). Therefore, while proteins that partition into the detergent-rich phase after solubilization in Triton X-114 are almost certainly lipoproteins or intrinsic membrane proteins (presumably highly hydrophic ones), other intrinsic membrane proteins (less notably hydrophobic) will partition with the soluble proteins in the detergent-poor phase, up to and including the sodium channel (Reber and Catterall, 1987), which is predicted to contain as many as 24 membrane-spanning segments.

Amphipathicity

Another important difference in the primary structure of membrane proteins (compared to soluble proteins) is that all membrane proteins must deal with multiple environments; aqueous solvent, the (potentially) highly charged surface of a membrane, and the low dielectric of the lipid bilayer hydrocarbon interior. These considerations often lead to an amphipathic segmentation of amino acids along the protein's primary sequence, as different linear stretches of the sequence (which often act as independent folding domains) have evolved to deal with their local surroundings. This segmentation is often most dramatic in membrane-active peptides which, while passably stable in aqueous solvent, are able to bind to and/or insert within lipid bilayers. Examples include peptides whose structures have been solved crystallographically, such as alamethicin (Fox and Richards, 1982; Hall et al., 1984) and melittin (Terwilliger et al., 1982; Bernheimer and Rudy, 1986), both of which are composed of ~20 hydrophobic amino acids followed by a shorter, polar tail (alamethcin, it should be noted, is composed of non-standard amino acids, including seven amino-isobutyrates and a C-terminal phenylalinol). In this instance, the segmental amphipathicity of these soluble peptides is important for membrane insertion and pore formation (Hall et al., 1984; Bernheimer and Rudy, 1986). The D helix of prostaglandin H synthase is also segmentally amphipathic (Garavito et al., 1994).

Not surprisingly, the membrane-spanning regions of bitopic and polytopic membrane proteins are frequently composed almost entirely of hydrophobic amino acids, with a smattering of polar amino acids interposed, presumably for reasons of folding or function. In particular, a hydrophobic stretch of amino acids spanning

18-24 amino acids, (long enough to form an α-helix 27-36 Å in length and span the apolar interior of a 35-45 Å wide eukayotic lipid bilayer), is frequently denoted a membrane-spanning structure even in the absence of any additional evidence. Amino acid sequence analysis programs (which typically analyze segmental distribution of apolar/polar residues in the form of hydrophobicity plots (Kyte and Doolittle, 1982; Argos, 1989; Fasman and Gilbert, 1990; Jähnig, 1990; Popot and de Vitry, 1990)), have made the exposure of such features straightforward. However, even though hydrophobicity analyses often generate more questions than they answer (i.e, where are the ends of the putative membrane-spanning regions, exactly how many are there, and how many polar and/or charged amino acids can be tolerated in a putative membrane-spanning region?), it is also true that hydrophobicity plots have successfully identified the membrane-spanning regions of bona fide intrinsic membrane proteins (Fasman and Gilbert, 1990; Jähnig, 1990; Popot and de Vitry, 1990). Hydrophobicity plots are also very useful for identifying evolutionarily-related membrane proteins, even in the case where divergence has almost completely erased primary sequence homology, such as between the most distantly-related members of the family of ligand-gated ion channels (Noda et al., 1983; Takeshima et al., 1989; Moriyoshi et al., 1991).

Distribution of Specific Amino Acids

Analysis of the amino acids which compose the interior membrane-spanning portions of intrinsic membrane proteins has identified certain "preferential" arrangements of amino acids in these regions. The aromatic amino acid tryptophan is found at the "flanks" of the transmembrane α-helices of the photosynthetic center (Deisenhofer et al., 1985), and is also observed to localize at the ends of predicted membrane-spanning α-helices in other membrane proteins, as are phenylalanine and tyrosine (Landolt-Marticorena et al., 1993). Interestingly, phenylalanine and tyrosine are also commonly found at the flanking ends of β-strand membrane-spanning sequences of several porins, where the presence of these residues clearly demarks the ends of the hydrophobic portion of the strand. Furthermore, these residues are differentially distributed on both faces of the protein, with tyrosines and phenylalanines both present on the periplasmic end of the molecule (where the phenylalanine side-chains all point inward to the membrane while the hydroxylated tyrosine side-chains all face out), but only tyrosine (and tryptophan) are found in the outer flank (Weiss and Schulz, 1992; Kreusch et al., 1994). Hydrophobic and hydrogen bonding interactions between aromatic amino acids, primarily tryptophan and tyrosine, also provide the major bonding forces between the (relatively) weakly-associated membrane-spanning α-helices of the α- and β-apoproteins of the LH2 light-harvesting complex of photosynthetic bacteria (McDermott et al., 1995). The N-terminal portion of a Type I membrane-spanning region is more likely to be enriched in isoleucine and valine, while the carboxy-terminal half is enriched in leucine (Landolt-Marticorena et al., 1993) (with phenylalanine, these four amino

acids comprise ~45-55% of the amino acids in interior membrane-spanning regions (Reithmeier and Deber, 1992)). The particular distribution of these amino acids across the hydrophobic membrane-spanning region may have to do with α-helix formation, or perhaps reflect a response by the protein to the local micro-environments of the lipid bilayer (White et al., 1994).

The placement of other hydrophobic amino acids in membrane-spanning regions is not as selective. Alanine and glycine are found throughout, perhaps with a slight predilection for the interior of multi-helix bundles (Reithmeier and Deber, 1992). Similarly, while serine and threonine residues do not appear to prefer the interior (protein-facing) or exterior (lipid-facing) surfaces of the membrane-spanning α-helices of the photosynthetic reaction center (Rees et al., 1989; Reithmeier and Deber, 1992), they are predicted to occupy the interior lining of the multi-helix bundles thought to form the membrane-spanning water-filled channels in the ligand-gated ion channel family of intrinsic membrane proteins (Stroud et al., 1990; Galzi and Changeux, 1994). Proline is a special case. As a hydrophobic amino acid, proline is not thermodynamically excluded from the interior of the lipid bilayer, but its capacity to form kinks into α-helices does influence its positioning. Accordingly, proline is rarely found in the middle of the (predicted) α-helices of bitopic proteins (Deber et al., 1986), particularly in those cases where the membrane-spanning segment functions largely to anchor the protein to the membrane. However, it has been noted that proline residues are frequently observed in the predicted membrane-spanning α-helices of polytopic signaling and transport proteins (Brandl and Deber, 1986). A kink introduced by a proline residue into an α-helix might help provide a binding pocket for transport substrates (this possibility is supported by experimental studies (Vogel, 1992)), but the suggested possibility the *cis/trans* isomerization of proline residues might be involved in the conformational changes which accompany transport and signaling (Brandl and Deber, 1986; Williams and Deber, 1991; Vogel, 1992) appears unlikely, given the relatively slow rate (seconds to minutes time scale) at which *cis/trans* proline isomerization occurs in model compounds (Brandts et al., 1975; Lin and Brandts, 1983). However, it is possible that proline isomerization is mechanistically involved in the characteristic desensitization of signaling proteins, which often occurs on the seconds to minutes time scale (McCarthy et al., 1986; Bormann et al., 1987; Hausdorff et al., 1990).

While we have focused above on the mostly apolar amino acids which comprise the bulk of the amino acids found in the membrane-spanning regions of intrinsic membrane proteins, it is also true that highly polar, and even potentially charged amino acids, do occur in membrane-spanning domains. The 11 membrane-spanning a-helices of the photosynthetic reaction center from *Rhodobacter sphaeroides* (Figure 6) contain roughly two asparagine and four glutamines, as well as five glutamates, eight histidines, and nine arginines (Yeates et al., 1987; Rees et al., 1989). It should be noted that both many of these residues are located at the "ends" of the membrane-spanning α-helices, (particularly glutamate, arginine, and aspartate) and that the exact boundaries of the membrane-embedded region of these

helices is inferred (as the membrane structure is lost during crystallization). Furthermore, other of the "charged" residues found deeply buried within the membrane-spanning regions of the photosynthetic reaction center, particularly the histidine residues, act as ligands for free iron or bacteriochlorophyl molecules (Yeates et al., 1987). This pattern, with polar and charged residues found at the expected boundaries of membrane-spanning regions, or buried in the interior for functional reasons, is also observed in other membrane-proteins whose structures are known in detail. For example, *Halobacterium halobium* bacteriorhodopsin contains four flanking glutamate residues, and five lysines, arginines, and aspartates postioned at various depths within the seven membrane-spanning α-helices. However, the majority of these charged residues, (including lysine 216, to which the retinal molecule is covalently attached (Bayley et al., 1981; Katre et al., 1981), are located within either the immediate retinal binding pocket or the proton-conducting channel (Henderson et al., 1990; Figure 4). Similarly, while the nonselective porins are thoroughly hydrophobic in their membrane-spanning domains (Cowan et al., 1992; Weiss and Schulz, 1992; Kreusch et al., 1994), the more selective variant maltoporin contains a number of paired polar and ionizable amino acids (probably too far apart to interact strongly) at the constricted central portion of its membrane-spanning channel (Schirmer et al., 1995), perhaps involved in stabilizing the structure. Most dramatically, two of the three membrane-spanning α-helices of the photosystem II-associated LHC-II appear to be linked by a pair of salt bridges between arginine and glutamate residues at the midpoint level of the bilayer (Figure 5), even though the existance of salt bridges in the interior of bilayers was long considered "impossible" on thermodynamic grounds. These residues are also involved in interactions with chlorophyl moities, as are some of the other five glutamate, four lysine and two arginine residues found within the membrane-spanning regions of the LHC-II (Kühlbrandt et al., 1994). Therefore, while it is clear that the membrane-spanning regions of intrinsic membrane proteins will be composed largely of hydrophobic amino acids, it is also clear that highly polar and charged amino acids are not precluded if their presence is justified in terms of protein function.

The amino acids that flank membrane-spanning domains on the outside of the lipid bilayer (e.g., at sites of helix initation or termination, linking regions, stop-transfer sequences, etc.) also show some characteristic specificities. It was first postulated by Blobel (1980) that there would be a signal at the ends of membrane-spanning regions, termed a stop-transfer sequence, to stop adjoining amino acids in the sequence from inserting into the bilayer. In practice, the stop-transfer sequence is not a specific amino acid sequence, but rather a short stretch of polar or charged amino acids which serve to fix the ends of the membrane-spanning sequence relative to the lipid bilayer. Interestingly, the particular charged amino acids found at the limits of membrane-spanning regions largely depends on the side of the bilayer in question. Positively-charged residues (mainly arginine and lysine) predominate in forming the stop-transfer signal on the inner (cis) side of mem-

brane-spanning sequences in both bitopic and polytopic membrane proteins, re-
gardless of their origin (i.e., the pattern is maintained for both prokaryotes and
eukaryotes, and for both cell surface membrane proteins, and internal membrane
proteins found in the endoplasmic reticulum, inner mitochondrial membrane, and
thylakoid membrane) (von Heijne and Gavel, 1988; von Heijne, 1992; Sipos et al.,
1993). The importance of the positively-charged amino acids at the inner boundary
of a membrane-spanning segment has been observed experimentally, as mutageni-
cally altering this pattern in the leader peptidase of *Escherichia coli* resulted in
re-orientation of the amino- and carboxy-termini of the protein (von Heijne, 1989).
Conversely, the outer (*trans*) side of membrane-spanning segments in the popula-
tion of intrinsic membrane proteins appears equally likely to contain either posi-
tively- or negatively-charged amino acids (von Heijne and Gavel, 1988), although
the outer ends of the membrane-spanning segments of the nonselective porins are
strongly negative in their composition (Weiss and Schulz, 1992; Cowan, 1993;
Kreusch et al., 1994). To a lesser extent, this is also true of the *trans* ends of the
membrane-spanning α-helices of the L and M subunits of the photosynthetic
reaction center (Reithmeier and Deber, 1992) and the LHC-II complex (Kühlbrandt,
1994). The underlying basis for the asymmetric distribution of charged residues on
the two sides of the lipid blayer is not entirely clear, although the importance of
electrostatic interactions has been stressed (Cramer et al., 1992). At the level of the
plasma membrane, at least, the preponderance of positively-charged residues at the
internal end of membrane-spanning regions may be related to the fact that the inner
membrane leaflet is more negatively-charged than the outer leaflet, and could
function to reduce repulsive interactions.

Finally, there are a few other peculiarities in the distribution of amino acids in
membrane proteins to consider. In addition to the charged residues found at the the
ends of membrane-spanning regions cited above, other amino acids also are
preferred. Proline is often found at the amino-terminal ends of α-helices, typically
as part of a turn leading into the helix, and appears to function as such for
membrane-spanning α-helices as well (amino-terminal prolines are found at the
ends of membrane-spanning helices in the LHC-II (Kühlbrandt, 1994), bacteriorho-
dopsin (Henderson et al., 1990), and the LH2 light-harvesting complex (McDer-
mott et al., 1995). Asparagine and serine also show a slight preference for the
amino-terminal end of Type-I membrane-spanning a-helices, probably functioning
as helix initiators (Landolt-Marticorena et al., 1993). Lastly, asparagine (in the
sequence Asn-X-Ser/Thr) is also the site of N-linked glycosylation, a common
feature of cell-surface glycoproteins.

C.   Secondary Structure

The secondary structural elements of membrane proteins are similar to soluble
proteins. The major difference is that, while the extra-membranous domains of
intrinsic membrane proteins, (of the few that are known), include the range of

structures expected from the study of soluble proteins, the membrane-spanning regions are much more limited in composition. Currently, the secondary structure of known membrane-spanning regions is limited to either α-helices or β-strands (which then anneal to form a β-barrel). Membrane-spanning random coil structures have not been observed (and may never be, at least not those which directly contact the lipid bilayer), nor have reverse turns, although the latter structure probably could be maintained in the interior of a lipid bilayer if there was a need for one. Similarly, 3_{10} helices, or other relatively rare hydrogen-bonded structures, may exist in the membrane-spanning domains of some membrane proteins, but they have not yet been observed. Quite simply, the low-dielectric environment of the lipid bilayer constrains the types of protein secondary structure which can be tolerated.

Helical Structures

Initially, it was generally considered that only helical protein structures would be maintainable in the interior of the lipid bilayer (Singer, 1971; Henderson 1975; Engelman and Steitz, 1981). The reasons are obvious; in a helix, all of the polar main-chain atoms of the peptide backbone are hydrogen-bonded, effectively removing the energetic penalty for positioning polar groups in a non-polar enviroment (assuming the side-chains of the amino acids are equally apolar). Even though we now know this restriction is not absolute, it is still the norm and, except in the specialized case of the porin family of bacterial outer membrane proteins, all of the membrane-spanning regions of the membrane proteins whose structures are known at high resolution are α-helical (including the α-helices of the monotopic membrane protein prostaglandin H synthase, which apparently only dip into the bilayer *(Picot et al., 1994)). As far as bitopic proteins are concerned, the membrane-spanning segment is still clearly limited to helical structures only (except perhaps in the* case of a multisubunit membrane protein complex, where a bitopic subunit could conceivably span the bilayer without ever contacting it). However, in polytopic membrane proteins, additional membrane-spanning structures are possible, including mixed α-helical, β-sheet forms.

However, while α-helical structures are still the most common membrane-spanning motif, the membrane-spanning α-helices of intrinsic membrane proteins appear to differ somewhat from their soluble counterparts in terms of helix folding and multihelical assembly (see next section). First, it is clear that the hydrophobic lipid environment often induces α-helical formation (Singer, 1971; Kaiser and Kezdy, 1987; Cserháti and Szögy, 1993). This is not surprising. The absence of aqueous solvent "forces" the polypeptide to form hydrogen bonds with itself. The feature is most clearly observed with membrane-active peptides, which often appear to be composed of random coil or β-structures in solution, but shift to α-helical structures upon insertion into the lipid bilayer. For example, monomeric melittin appears to be random coil (disordered) in solution, but becomes predominantly α-helical when inserted into membranes, on the basis of circular dichroic (Lauter-

wein et al., 1979; Vogel, 1981) and Raman spectroscopic studies (Levin et al., 1982; Vogel, 1987). This pattern is true for a number of other membrane-active peptides, including neuropeptide Y (McLean et al., 1990), β-endorphin (Zetta and Kaplein, 1984), and even isolated signal peptides (Gierasch, 1989). However, the tendency of the lipid environment to induce α-helical structure in peptides may not always be observed. The crystal structure of alamethicin (from crystals grown in 10% methanol/acetonitrile) described a pore composed of a ring of kinked α-helices (Fox and Richards, 1982). However, circular dichroism studies of alamethicin in lipids indicated much less α-helical structure, and a significant degree of β-structure in some (Jung et al., 1975; Cascio and Wallace, 1985; Rizzo et al., 1987), but not all (Vogel, 1987) studies. Ultimately, it is possible (perhaps even probable), that all these studies are correct, as it is clear that membrane-active peptides are very sensitive to their environments (the multiple proposed and supported conformations of gramacidin A are a good example (Wallace, 1986; Langs, 1988, 1989; Wallace and Ravikomar, 1988; Ketchem et al., 1993), or the various solved structures of leu-enkephalin (Camerman et al., 1983; Aubry et al., 1989; Milon et al., 1990)). In general, it is important to consider the solvent conditions (and peptide concentration) before deciding which conformation is the predominant one adopted by a membrane-active peptide when it is inserted into a lipid bilayer.

Another feature of α-helices to consider in terms of membrane proteins is bilateral amphipathicity. Amphipathic α-helices are characterized by an asymmetric distibution of polar/apolar amino acid side-chains, with one face of the helix predominantly apolar and the other face more or less polar. As demonstrated in Figure 3, amphipathic α-helices function in membrane proteins in two ways, both of which have been observed; they may lie along the surface of the bilayer with the polar face exposed to solvent or protein and the apolar face directed to the bilayer interior, or they may assemble into multihelical bundles to form a hydrophilic passage across a membrane. Examples of the former include the helix D of the LHC-II complex (Kühlbrandt, 1994; Figure 5), abbreviated helices which link membrane-spanning helices A to B and C to D in the photosynthetic reaction center M and L subunits (Deisenhofer et al., 1985; Allen et al., 1987; Figure 6), short helical segments attached to membrane-spanning helices in filamentous bacterio-phage coat proteins (Cross and Opella, 1994; Opella et al., 1994), membrane-active peptides such as magainin (Bechinger et al., 1993) and neuropeptide Y (McLean et al., 1990), and the essential (for membrane binding) helices A, B, and C of the complex monotopic protein prostaglandin H synthase (Garavito et al., 1994; Picot et al., 1994; Figure 9). Multiple, membrane-spanning amphipathic α-helices are rarer in solved structures, although quite common in predicted ones. The five helices which line the proton channel of bacteriorhodopsin are somewhat amphipathic in nature, particularly helices B, C, and G, although the charged and polar residues of bacteriorhodopsin are distributed among so many helices that no one helix is extraordinarily amphipathic (Henderson et al., 1990). A soluble fragment of the pore-forming domain of colicin A also contains multiple amphipathic α-helices,
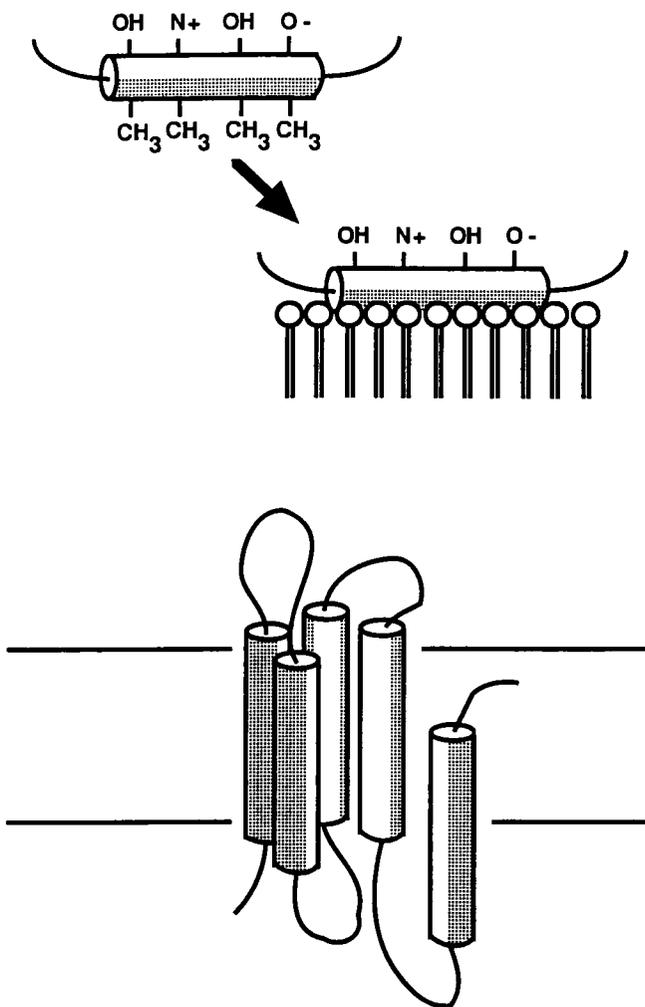
Figure 3. Modes of interaction of bilateral amphipathic α-helices with the lipid bilayer. Top, interaction of an amphipathic α-helix with the surface of a lipid bilayer. In this example, a single amphipathic α-helix, with polar and apolar faces as indicated, is demonstrated to sink into the surface of the bilayer, burying the apolar face of the α-helix within the hydrocarbon interior of the bilayer while the polar face remains exposed to solvent and the charged surface of the bilayer. Bottom, assembly of a group of amphipathic α-helices to form a membrane-spanning, hydrophilic channel. In this example, five amphipathic α-helices act as the staves of a helical "barrel," exposing the hydrophobic (shaded) faces to the interior of the lipid bilayer and generating a channel lined with more polar and charged residues (unshaded) which could serve to pass polar or charged atoms or molecules across the bilayer. The right front helix has been pulled back, to expose the interior of the passage.

191

on the basis of X-ray crystallographic studies (Parker et al., 1989), which have been speculated to form the cytotoxic colicin channel upon membrane insertion. Other, truly amphipathic helical membrane-spanning structures remain largely speculative, although there is good biochemical and electrophysiological data to suggest that (weakly) amphipathic α-helices form the channel lining in certain types of gated ion channels (Betz, 1990; Stroud et al., 1990; Galzi et al., 1991; Langs and Triggle, 1992; Galzi and Changeux, 1994).

β-Structures

β-strands, which assemble to form β-sheets, are the other form of membrane-spanning secondary structure. However, unlike individual α-helices, individual β-strands or even limited β-sheet structures are thermodamically very unstable in the hydrophobic interior of the lipid bilayer, due to the many unfulfilled hydrogen-bonding main-chain amide and carboxyl groups (Singer, 1971; Engelman and Steitz, 1981). To date, such structures have not been observed, and may never be observed. However, a very stable membrane-spanning structure (Jap and Wallian, 1990) is generated when β-sheets circularize to form a β-barrel, as exemplifed by the porin family of bacterial outer-membrane proteins (Weiss and Schulz, 1992; Cowan, 1993; Kreusch et al., 1994; Schirmer et al., 1995). By closing an extended β-sheet into a barrel, all of the polar main-chain amide and carboxyl groups become hydrogen bonded, removing the primary thermodynamic barrier to β-structure in the membrane interior. While the structure of the porins will be discussed in greater detail below (see Figure 8), some general aspects of membrane-spanning β-strand structure derived from study of the porins will be considered here. The individual β-strands which form the "staves" of the β-barrel can be much shorter than membrane-spanning α-helical sequences. Some of the membrane-spanning β-strands of the porins are as short as 6-9 residues long, although some strands are up to 18 residues long, forming a more extended β-structure. Furthermore, the strands are tilted 35-50° normal to the membrane (and also relative to the axis of the barrel). However, this is not a feature unique to membrane-spanning β-barrel structures, but is commonly found in the β-barrels of soluble proteins as well (Chou et al., 1990; Chothia and Murzin, 1994). Only antiparallel β-strands association is found in the porin structures. While a β-barrel formed of parallel β-strand is not denied on thermodynamic grounds, assembly of such a structure would be much more complex and therefore that much more unlikely. Finally, like α-helices, β-strands can also form amphipathic structure, with opposing polar and apolar faces, some evidence of which is found in the porin structures (Jap and Wallian, 1990; Cowan, 1993).

β-structure is associated with other membrane proteins as well. Gramacidin A, a channel-forming bacterial peptide noncoventionally composed of 15 alternating D and L amino acids, forms unusual β-helical structures (like a spiraled ribbon) which appear to differ in their state of assembly and orientation (parallel or anti

parallel dimers, or end to end dimers) as a function of solvent and method of study (Langs and Triggle, 1992; Ketchem et al., 1993). Spectral and electron microscopic studies have suggested appreciable β-structure in other, more traditional intrinsic membrane proteins as well, including myelin proteolipid, sarcoplasmic reticulum Ca^{2+} ATPase (Jennings, 1989), and the mitochondrial voltage-dependent anion carrier, which may be a mitochondrial homolog of the bacterial porins (Manella et al., 1986; Forte et al., 1987). Finally, the structures of two soluble toxins with the potential to oligomerize into membrane-spanning channels have been determined at atomic resolution by X-ray diffraction techniques. Both of the toxins, proaerolysin from *Aeromonas hydrophila* (Parker et al., 1994) and toxin γ from of the cobra *Naja nigricollis* (Bilwes et al., 1994), contain β-sheet domains which are speculated to form oligomeric pores (analogous to the porin β-barrel) when inserted into the lipid bilayer.

Mixed Membrane-Spanning Secondary Structures

While only all α-helical or all β-strand membrane-spanning structures have been observed in the high resolution structures known to date, mixed structures are not precluded. Basically, if the dimensions of the membrane-spanning domain are large enough (to shield polar protein structures from the protein-lipid interface by isolating them in the more adaptable protein interior), essentially all of the structures found in soluble proteins could also be found in membrane-spanning regions. However, actual physical evidence for such membrane-spanning structures is limited. Recent, moderate-resolution electron micrographic image reconstructions of the nicotinic acetylcholine receptors of *Torpedo californica*, based on two-dimensional arrays of the protein in native membrane fragments, suggest that the membrane-spanning regions of the receptor are composed of a mixed α-helical, β-sheet conformation. A ring of weakly amphipathic α-helices, predicted on the basis of mutagenesis and biochemical labeling experiments to form the lining of a water-filled channel (Stroud et al., 1990; Galzi et al., 1991), were observed in the interior of this multisubunit ligand-gated ion channel, but the remainder of the membrane-spanning structure observed was more consistent with an extended β-conformation (Unwin, 1993). If this conclusion, based on a 9 Å resolution structure, is confirmed at higher resolution, it will be the first example of a membrane-spanning region composed of a mixed secondary structure. As a note of caution, however, it should be kept in mind that the acetylcholine receptor has been difficult to characterize consistently in the past (for example, good biochemical evidence suggests that the one α-helix observed in the electron microscopic studies cited above is actually a β-strand as well (Akabas et al., 1992), while the number of membrane-spanning regions per subunit determined by immunological topographical studies has varied from four to as much as seven (Young et al., 1985; Ratnam et al., 1986; Dwyer, 1988; Jennings, 1989)). Other intrinsic membrane-protein complexes with multiple membrane-spanning segments have also proposed

to contain mixed secondary structure domains, the most interesting of which are probably the voltage-gated ion channels. These proteins were initially thought to be composed of four subunits (or domains) each composed of six membrane-spanning α-helices (including one novel helix containing a lysine or arginine residue at every third position, which most likely functions as the voltage sensor (Jan and Jan, 1989)), until mutagenesis studies exposed a short domain between the fifth and sixth predicted membrane-spanning α-helices which almost certainly forms a large part of the lining of the ion channel but which probably spans part or all of the bilayer in some non-helical conformation (Catterall, 1994).

Spectral Assessment of Secondary Structure

In the absence of high resolution structural information, estimates of the extent of protein secondary structure can be determined using spectrosopic techniques, primarily circular dichroism, Raman, and infrared spectroscopy, which can provide structural information about both the membrane-spanning and extramembranous domains of intrinsic membrane proteins, and also place limits upon proposed structures (Heyn, 1989; McCarthy, 1992; Vogel, 1992). Initially, the validity of these techniques in the study of membrane protein structure was questioned, as they depended to some extent upon conclusions based on the study of soluble proteins, and it was not clear that membrane proteins would follow the same rules. However, in most cases, the structures of the membrane proteins which have been solved at high resolution have confirmed the earlier spectral estimates of secondary structure. For example, polarized infrared spectroscopic analysis of the photoreaction center from *Rhodospeudomonas sphaeroides* correctly described the largely α-helical structure of the L and M subunits, the greater degree of β-structure in the H subunit, and even the slight tilting of the membrane-spanning α-helices as they cross the bilayer (Nabedryk et al., 1982, 1984). Similarly, the β-structure of the porins was established well in advance of the crystallographic analysis on the basis of multiple circular dichroic and infrared spectroscopic studies (Rosenbusch, 1974; Schindler and Rosenbusch, 1978; Kleffel et al., 1985; Vogel, 1986; Nabedryk et al., 1988), and the predominant α-helical membrane-spanning structure observed in the 6 Å structure of the photosystem I complex of green sulphur bacteria was first proposed on the basis of circular dichroism analysis (Krauss et al., 1993). However, the agreement between spectral and other measures of protein structure has not been universal. Perhaps the best example is from early studies of bacteriorhodopsin, where some circular dichroism and infrared spectra were interpreted to indicate a significant degree of β-structure (Cortijo et al., 1982; Jap et al., 1983; Glaeser and Jap, 1985; Lee et al., 1985; Downer et al., 1986), although most studies (Mao and Wallace, 1984; Nabedryk et al., 1985; Vogel and Gartner, 1987; Wallace and Teeters, 1987; Gibson and Cassim, 1989) predicted the all α-helical structure which was later confirmed by higher resolution electron microscopic image reconstruction analysis. Finally, in the absence of a high resolution structure, it is clear that

discrepant spectral studies can be difficult to evaluate, as evidenced by the trouble-some nicotinic acetylcholine receptor. The predictions of the types and extent of secondary structure on the basis of circular dichroic and infrared red spectroscopic studies have ranged from about 20% α-helix and 48% β-sheet to 43% α-helix and 20% β-sheet (Yager et al., 1984; Mielke et al., 1986; Fong and McNamee, 1987; Mielke and Wallace, 1988; Wu et al., 1990; Castresana et al., 1992; Naumann et al., 1993), and experience indicates that they can not all be correct.

D.   Tertiary Structure

In this section we will consider the association of membrane-spanning α-helices and β-strands into higher order structures, and the topographical domain structure of membrane proteins.

Helical Association

A feature by which membrane-spanning α-helices differ from their counterparts in soluble proteins is their "independence." Individual transmembrane α-helices are fairly stable structures, due largely to the hydrophobic bilayer they have managed to span, and each may be considered "a small autonomous folding domain" (Popot, 1993) which can then independently assemble to form higher order structures, a two-stage assembly model first proposed by Popot and Engelman (1990). The stability of membrane-spanning α-helices is supported by a number of lines of evidence; the existance of multiple bitopic proteins which are little more than a single membrane-spanning domain (Popot and deVitry, 1990), the propensity of the hydrophobic bilayer to induce α-helical structure (Singer 1971; Kaiser and Kezdy, 1987; Cserháti and Szögy, 1993), and the observation that truncated and/or substituted α-helical segments (of larger α-helical membrane-spanning complexes) assume the same conformation as they do in the intact protein (reviewed in Lemmon and Engelman, 1994). These stable α-helices then appear to assemble into higher order structures in a separate, discrete folding process, as suggested by the ability of preformed, smaller helical domains to combine to form functional, nativelike proteins. Examples include renaturation of functional bacteriorhodopsin by com-bining a fragment of bacteriorhodopsin composed of helices A and B, and a fragment composed of helices C-G (Popot et al., 1987) (other combinations of bacteriorhodopsin fragments also yield functional bacteriorhodopsin (Lemmon and Engelman, 1994)), as well as similar recombination experiments performed upon the structurally-related heptahelical receptors, the β-adrenergic receptor (Kobilka et al., 1988), and the muscarinic receptor (Maggio et al., 1993). While these studies are all restricted to the "seven membrane-spanning α-helix" family of polytopic proteins, the observation that assembly follows helix formation appears to be a general rule for α-helical membrane-spanning domains.

The interactions which occur between membrane-spanning α-helices and drive assembly, have recently been summarized as either specific or nonspecific (promis-

cuous) (Lemmon and Engelman, 1994). Hydrogen bond formation between membrane-spanning α-helices is rarely observed (Lemmon and Engelman, 1994), and it may be that ionic interactions (either at the ends of helices as in the LH2 complex (McDermott et al., 1995), or where helices cross as in the LHC-II complex (Kühlbrandt, 1994)) are more important for tight helical interactions. Other forces which may promote specific interactions include stereochemical close packing (knobs into holes) and the positioning of prosthetic groups (while prosthetic groups may not be important to the assembly of membrane-spanning domains in the majority of membrane-spanning proteins, the fact that five of the six membrane proteins whose structures are known at moderate-high resolution contain highly-ordered prosthetic groups may not be a coincidence). Conversely, van der Waals interactions, lipid exclusion, and the structure of the protein outside the bilayer in membrane proteins with significant extracellular domains, appear to underlie some of the more nonspecific interactions between membrane-spanning α-helices (Lemmon and Engelman, 1994). A prominent feature of α-helices in soluble proteins is the helix dipole, which supports the antiparallel association of helices to maximize attractive charge interactions (Hol et al., 1981). However, the importance of the helix dipole for maintaining interactions between membrane-spanning α-helices is not clear (Popot and Engelman, 1990). While it is true that the helical bundles composed of the membrane-spanning α-helices of the photosynthetic reaction center (Deisenhofer et al., 1984; Yeates et al., 1987) and bacteriorhodopsin (Engelman et al., 1980; Henderson et al., 1990) are predominantly anti-parallel, the helices of the α- and β-apoproteins of the LH2 complex (which interact largely at the ends of the helices) are parallel (McDermott et al., 1995), as are the most closely associated α-helices of the LHC-II complex (Kühlbrandt et al., 1994). It is possible that the accumulated counterions at the level of the phospholipid headgroups of the lipid bilayer may weaken the influence of the helix dipole in establishing helical assocations, and consequently weaken the tightness of helix interactions in membrane-spanning domains.

The types of helical bundles formed in the membrane-spanning regions of polytopic domains are only beginning to be understood. A common feature is that the helices typically are not perpendicular to the plane of the bilayer, but rather bisect the membrane at some angle. This was initially noted in the early electron microscopic reconstructions of bacteriorhodopsin (Henderson and Unwin, 1975), which demonstrated that helices 1-4 (membrane-spanning sequences A, E, F, and G on the basis of hydrophathy analysis (Engelman et al., 1980)) are tilted up to 20° from helices 1-3, which are essentially perpendicular to the bilayer (Figure 4). The 10 membrane-spanning α-helices of the L and M subunits of the photosynthetic center (Figure 6) are also tilted, crossing the bilayer at an average angle of 22-25° (Deisenhofer et al., 1985; Yeates et al., 1987), with the D helix tilted the most (35-38°). It is interesting to note that the D helix is 27-28 amino acids long, and the most hydrophobic of all the membrane-spanning helices in the photosynthetic reaction center, which leads to the question: is the D helix tilted so that it can bury

all of those apolar residues in the hydrophobic interior of the bilayer, or is it composed of a long stretch of apolar amino acids so that it can span the bilayer at a tilted angle? Probably the latter, as the D helices are involved in L and M subunit dimer formation, and also complexation with iron and bacteriochlorophyls (Deisenhofer et al., 1985; Yeates et al., 1987). In more recent studies, helices A and B of the LHC-II complex cross the bilayer tilted 32° normal to the membrane (Kühlbrandt et al., 1994), while the single α-helices of the α- and β-apoproteins of the LH2 complex are perpendicular, or tilted 15°, to the expected membrane normal (McDermott et al., 1995). Probably, α-helices span the bilayer at an angle for reasons related to the functions of the individual proteins, and not due to a more general mechanism underlying helical folding and assembly.

A special class of membrane-spanning α-helical bundle is represented by the seven membrane-spanning α-helix/heptahelical receptor/ G protein-coupled receptor family (of which over 400 members have been putatively identified by protein sequence analysis (Attwood and Findlay, 1994). The pattern of seven-associated transmembrane α-helices has been conserved from halobacteria to humans, although at the evolutionary extremes of the family little significant homology remains at the level of the amino acid sequence (Donelly and Findlay, 1994). Interestingly, recent electron micrographic studies of the structure of bovine rhodopsin suggest that the packing of the membrane-spanning helices in the seven helix bundle may also have diverged, as at similar resolution the moderate-resolution structure of rhodopsin (Schertler et al., 1993) appears condensed (more like a ring of helices), rather than the two rows of helices found in bacteriorhodopsin (Henderson et al., 1990). This has led to the suggestion of alternate arrangements of the seven helix bundle in eukaryotic heptahelical receptors, either in terms of helix packing or even linkage (Donelly and Findlay, 1994). If, as suggested above, membrane-spanning helices can function as autonomous folding domains (Popot, 1993), and if membrane-helical interactions are either specific or promiscuous (Lemmon and Engelman, 1994), it is possible that the membrane-spanning helical bundles of related membrane-spanning proteins may have different orientations, depending on which specific interactions are retained. At a more fundamental level, it may be that the membrane-spanning helical assemblies of some membrane proteins are structurally dynamic for signaling, allosteric, and transport functions, and are not restricted to a single "packing" arrangement (Lemmon and Engelman, 1994).

Finally, what is the size limit, if any, of the membrane-spanning helical bundles of polytopic intrinsic membrane proteins? The α- and β-apoproteins of the LH2 complex (McDermott et al., 1995) and the H subunit of the photosynthetic reaction center (Deisenhofer et al., 1985) are bitopic, single membrane-spanning proteins, while known polytopic proteins inclue the LHC-II complex, which contains three membrane-spanning a-helices (Kühlbrandt et al., 1994); the M and L subunits of the photosynthetic reaction center, which each contain five membrane-spanning α helices (Deisenhofer et al., 1985); bacteriorhodopsin, which contains seven mem-

brane-spanning α helices (Henderson et al., 1990); and the A and B subunits of thermophile photosystem I which may each contain eight membrane-spanning α-helices (Krauss et al., 1993). However, even larger helical bundles have been predicted, and to varying degrees confirmed. The many members of the family of non-ATP-dependent transporters (symporters and antiporters) are estimated to be composed of twelve membrane-spanning α-helices (Maloney, 1994), although all of the helices may not be essential for function, as some members of the family appear to have shed a pair of helices. Similarly, the larger members of the ATP-binding cassette (ABC) transporter family (Higgins, 1992), including such interesting proteins as the cystic fibrosis transmembrane conductance regulator and P-glycoprotein (the multi-drug resistance gene product) also appear to be composed of 12 membrane-spanning α-helices, although in this case the pattern of 12 helices is interrupted in the middle by a large, cytoplasmic ATP-binding domain. Fourteen member helical bundles are predicted for the family of mammalian ion exchange proteins (Reithmeier, 1994), and even larger helical bundles (14-17) are predicted for several NADH-reductase subunits (Popot and de Vitry, 1990). Finally, the α subunit of the sodium channel and the α_1 subunit of the calcium channel (Catterall, 1994) (very large proteins with molecular weights in the neighborhood of 300,000 daltons) are each likely to contain 24 membrane-spanning α-helices (and perhaps some other membrane-spanning structures as well). However, while even larger helical bundles may exist in some as yet uncharacterized membrane proteins, as the complexity of the structure increases, of course so does the difficulty of proper folding and assembly.

β-Barrel Formation

All of our knowledge of standard β-structure in membrane-spanning regions comes from the atomic resolution X-ray crystallographic analyses of the porin family of bacterial outer membrane proteins. Fortunately, some features of their construction and predictions for future structures can be made from this numerically limited data base. The porins are composed of large, antiparallel β-pleated sheets closed into a membrane-spanning β-barrel. Given the polarity expected of an unassembled β-barrel, it appears that the porins fold either prior to, or concurrently with, membrane insertion (Nikaido, 1992; Popot, 1993). Once assembled, however, this structure is exceptionally stable. Matrix porin (the product of the OmpF gene), can be completely delipidated, renatured following sodium dodecylsulphate denaturation (Rosenbusch, 1974), is pH stable from below 2 to above 12, effectively resists proteases and heat denaturation, and retains a significant degree of β-structure even in concentrated guanidium chloride (Schindler and Rosenbusch, 1984). Other members of the porin family appear equally robust (Cowan, 1993), clearly as a result of their sturdy β-barrel structure. Interestingly, the size of the β-barrel is not constant among the porin family. Both the original porin structure from photosynthetic bacteria (Weiss et al., 1991; Weiss and Schulz, 1992) and sub-

sequents structures of matrix porin and phosphorin from *E. coli* (Cowan et al., 1992) described β-barrels composed of 16 antiparallel β-strands. Recently, however, the β-barrel of maltoporin (another porin from *E. coli*), has been determined to be formed from 18 β-strands (Schirmer et al., 1995). Whereas this might appear a straightforward method to generate a larger pore, the maltoporin channel is more constricted at its narrowest point (Schirmer et al., 1995) than is matrix porin (Cowan et al., 1992), although in this instance the narrower pore diameter is determined largely by the positioning of side chains and the interposition of looping regions down into the pore and not the diameter of the β-barrel backbone. However, this observation also suggests that there is no one approved membrane-spanning β-barrel structure, and it is possible that both smaller (such as the eight strand β-barrels predicted for the OmpA and OrpF protein of *E. coli* and *Pseudomonas aeruginosa* (Cowan, 1993)) and larger membrane-spanning β-barrel structures will be described in the future.

Membrane Protein Domain Structure

While we mentioned above that individual membrane-spanning α-helices could be characterized as independent, stable folding domains, it is often useful to consider intrinsic membrane proteins simply in terms of their three topographical domains: external to the membrane (usually cytoplasmic or extracytoplasmic), or within the lipid bilayer. Frequently, this manner of classification reiterates the functional divisions of the membrane protein. In many cases, the extracellular domains or the cytoplasmic domains are functionally discrete structures, which can be cleaved from the native protein by proteolytic or chemical means or generated by directed mutagenesis, and yet retain their ability to maintain the appropriate stucture and perform their relevant enzymatic or binding activities (Jennings, 1989). Most of the membrane proteins whose structures are known at atomic resolution are largely embedded in the lipid bilayer and contain no discrete extramembranous domains, including the porins (Weiss et al., 1991; Cowan et al., 1992), bacteriorhodopsin (Henderson et al., 1990), and the LHC-II and LH2 complexes (Kühlbrandt, 1994; McDermott et al., 1995). However, the bitopic H subunit of the photosynthetic reaction center complex is largely composed of a single, cytoplasmic domain (which is anchored to membrane by a solitary membrane-spanning a-helix (Deisenhofer et al., 1985; Allen et al., 1987)), while the monotopic prostaglandin H synthase could be considered to be essentially an extracellular (or endoplasmic reticulum lumenal) domain that just happened to be bound to the membrane surface by a layer of amphipathic a-helices (Picot et al., 1994) (although, unlike more traditional unitary extramembranous domains, prostaglandin synthase can not be functionally released from the membrane by proteolytic cleavage (Picot et al., 1994)).

It is known from topological studies that many other membrane proteins are characterized by discrete extramembranous domains crucial to the function of the proteins. In the case of polytopic proteins, the extracellular domain of the ligand-

gated ion channels contains the agonist (and also many toxin and alkaloid binding sites) while the intracellular domain provides sites for phosphorylation and cytoskeletal attachment (McCarthy et al., 1986; Stroud et al., 1990), while the ATP cassette family of transporters are named for their conserved intracellular ATP-binding domain (Higgins, 1992). However, the topographic segmentation of functional domains is best characterized by bitopic proteins. Many bitopic membrane proteins function to localize (soluble) enzymatic activities (e.g., guanylate cyclase and aminopeptidase) or cell-surface antigens (e.g., histocampatibility markers or viral spike proteins) to the membrane surface (Popot and de Vitry, 1990; Reithmeier and Deber, 1992), while more sophisticated bitopic proteins function as cell surface growth factor/tyrosine kinase receptors (Schlessinger and Ulrich, 1992) and integrins (Hynes, 1992). While it has not proved possible, to date, to determine the entire structure of these bitopic proteins at high resolution (due to the general difficulties associated with crystallization of membrane proteins), in some cases the structures of the most interesting parts have been determined. For example, the (soluble, cytoplasmic) tyrosine kinase domain of the human insulin receptor was recently determined at high resolution using X-ray crystallographic techniques (Hubbard et al., 1994), which serves as a model for the tyrosine kinase domains of a vast number of related proteins. Further, combined with the knowledge of what the extracellular fibronectinlike domains of the insulin receptor must look like (based on the structure of human fibronectin), it is possible to recreate a fairly accurate model of the insulin receptor by assembling the known structures of its parts (McDonald et al., 1995). Other examples include the cytoplasmic domain of cytochrome b_5 (determined over 30 years ago), portions of the extracellular domain of the rat CD4 receptor (Lange et al., 1994), the lumenal domain of cytochrome f of the chloroplast cytochrome b_6f complex (Martinez et al., 1994), the HA_2 subunit of the hemagglutinin dimer of influenza virus (Weis et al., 1990; Baker and Agard, 1994), and the extracellular ligand binding domain of the aspartate receptor from *Salmonella* (Scott et al., 1993) (while the latter is not strictly speaking a bitopic protein, containing as it does two membrane-spanning segments, it is functionally equivalent to a bitopic protein). Basically, partial information is preferable to no information in any case and, in some cases, understanding the structure of the extramembranous domains of membrane proteins provides most of the information necessary to begin understanding how they function.

E.   Quarternary Structure

Like soluble proteins, individual membrane proteins associate to form two types of higher order structures: one, symmetric or pseudo-symmetric subunits assemble to form dimers, trimers, tetramers, etc.; or two, disparate subunits form larger, non-symmetric complexes. Both forms of membrane proteins quarternary structure have been described in high-resolution studies, although the former type is much more prevalent. At first glance, the pseudo-symmetric dimers formed by the bitopic

α- and β-apoproteins of the LH2 complex might appear to be the simplest example. However, the structure is much more complex than that, as the α-apoproteins form a nine-membered inner ring circled by a concentric nine-membered β-apoprotein outer ring, which serve to laminate 18 bacteriochlorophyll *a* molecules (McDermott et al., 1995; Figure 7). The pseudo-symmetric L and M subunits of the photosynthetic reaction center (Figure 6) form a more complex AV dimer (with the axis of symmetry perpendicular to the plane of the bilayer). The pentahelical membrane-spanning domain of each subunit overlaps the other, like one hand lying atop the other, slightly displaced, with the primary interactions mediated by the homologous D and E helices of each subunit (Deisenhofer et al., 1985; Allen et al., 1987). The monotopic protein prostaglandin H synthase also forms a stable dimer, even when solubilized in detergent (Smith and Marnett, 1994), but in this case the predominant interactions leading to dimer formation involve the large, extramembranous domains of the monomeric subunits (Picot et al., 1994). In some cases, membrane proteins may associate to form larger assemblies in the lipid bilayer for packing or stability reasons, and not as a direct role in the function of the protein. For example, while monomers of bacteriorhodopsin normally associate to form trimers in the membrane, proton pumping apparently can proceed through bacteriorhodopsin monomers (Dencher and Heyn, 1979), and trimerization may just be a result of the close-packing of bacteriorhodopsin in the enriched purple membrane. Porin monomers also form trimers in the bacterial outer membrane, but in this case the association is quite tight, and it has been suggested that a hydrophobic core at the center of this trimer may serve the same stabilizing fuction as the hydrophobic hearts of soluble proteins (Cowan et al., 1992). LHC-II complex monomers also form trimers, which are fairly stable and persist upon detergent solubilization, but it is not known whether trimer formation materially affects light capture and energy transfer by this antennary compound (Kühlbrandt, 1994).

A number of membrane proteins whose structures are not highly resolved are also formed of symmetric or pseudo-symmetric assemblies. The family of voltage-gated potassium and chloride ion channels are almost certainly tetramers formed from homologous subunits. The evidence that voltage-gated potassium and chloride ion channels are tetrameric is indirect but compelling; the conserved structure of the individual potassium and chloride channel subunits are present as four domains in the homologous voltage-gated calcium and sodium channels (Catterall, 1994), which unlike K^+ and Cl^- channels apparently do not require the ability to switch and swap subunits in order to expand function. The neuromuscular nicotinic acetylcholine receptor, which serves as a model for the ligand-gated family of ion channels, is an even more unusual oligomer; it is a pseudo-symmetric pentamer composed of four homologous subunits (Raftery et al., 1980). The other members of the ligand-gated ion channel family (the glutamate, glycine and 5 HT_3 receptors (Betz, 1990; Stroud et al., 1990)) have a simpler subunit structure, being pentamers composed of one or two subunit types, except for the γ-aminobutyric receptors, which are comprised of a wide variety of oligomers generated from a large number of subunits and subunit-subtypes (Macdonald and Olsen, 1994), rather like potassium channels.

Membrane proteins also form nonsymmetric oligomers composed from nonidentical or homologous subunits. Again, the photosynthetic reaction center provides a good example of the two types of nonsymmetric oligomerization found in membrane proteins: interactions with other intrinsic membrane protein subunits, and interactions with soluble protein subunits. First, the L and M subunits, which form a pseudo-symmetric dimer, also interact with their unrelated bitopic partner, the H subunit, which perhaps functions to align interactions between the reaction center and associated light-harvesting complex. Second, the extra-membranous surface of the L and M subunits, where they converge on the periplasmic side of the bilayer, serve to bind a soluble c-type cytochrome subunit in some, but not all, bacterial reaction centers (Deisenhofer and Michel, 1991a). A similar pattern of interactions (i.e., nonsymmetric association of intrinsic membrane protein subunits, as well as noncovalent linkage between membrane-embedded subunits and soluble subunits) was observed at lower resolution in photosystem I of photosynthetic bacteria (Krauss et al., 1993). The latter form of interaction is most interesting, as it provides a mechanism for the higher order assembly of cellular complexes, such as membrane skeletons, nuclear pore structures, integrin receptor complexes, etc. The structures of some soluble components of membrane-bound complexes have been determined. For example, the structure of the soluble F_1 subunit of bovine mitochondrial F_1F_0 ATPase has recently been solved at atomic resolution by X-ray crystallographic techniques (Abrahams et al., 1994). Examination of the structure of this subunit (itself a complex nonamer formed from five subunit types) provides clues as to how the soluble F_1 ATP-synthesizing subunit is functionally linked to the membrane-embedded F_0 proton channel (Abrahams et al., 1994; Barber, 1994).

The final form of membrane protein quarternary structure is the transient association of intrinsic membrane proteins, usually as a mechanism for signal transduction. The diffusion of membrane proteins is limited to two dimensions, (i.e., within the plane of the bilayer), increasing the likelihood of productive interactions, and therefore the rapid association of activated subunits has been utilized by the cell as a mechanism to transmit signals across membranes. The most studied example is probably the family of tyrosine-kinase/growth factor receptors, wherein binding of the growth factor ligand induces or stabilizes dimerization of the receptor subunits and consequent activation of the linked cytoplasmic tyrosine kinase domains (Schlessinger and Ullrich, 1992). However, other examples exist, including the proposed direct interactions between ligated muscarinic receptors and voltage-sensitive ion channels (Hille, 1992). The complex cell biology of these dynamic interactions is still the subject of considerable analysis, and undoubtedly additional examples will be uncovered.

III.   EXAMPLES OF STRUCTURES

The commonalities of membrane protein structure discussed above were largely determined utilizing three separate techniques; nuclear magnetic resonance

(NMR), electron microscopy, and X-ray crystallography. Some of the features of these methods relevant to the study of membrane proteins, as well as examples of protein structures determined using these methods, are described below.

A. Nuclear Magnetic Resonance

NMR, particularly multi-dimensional NMR, is a technique capable of determining the structure of proteins in solution at atomic resolution. However, its application to the study of membrane protein structure has been somewhat limited, due largely to the inevitable spectral line-broadening associated with slowly-reorienting membrane-bound proteins. (Structure determination by NMR, or X-ray crystallography for that matter, also requires milligram quantities of pure protein, but that is not a concern unique to the study of membrane proteins.) To get around the line-broadening problem, NMR studies of membrane protein structure (which have been limited to peptides and small proteins to date) are performed in two ways, ideally using isotopically-labeled proteins. First, solid-state NMR, in which sheets of membrane-bound proteins are aligned between two glass plates, and radio frequency irradiation and mechanical movements replace the protein tumbling motion which normally sharpens NMR spectra (Cross and Opella, 1994; Opella et al., 1994). Second, standard multidimensional NMR studies of membrane-proteins, with the bilayer replaced by organic solvents, selected detergents, or (rarely) small lipid vesicles (Henry and Sykes, 1994). The information derived from both approaches may be combined, with solid-state studies providing details of tertiary structure, solution studies describing the basic secondary structure, and both methods identifying side-chain and backbone dynamics (Opella et al., 1994).

As the techniques necessary for the study of membrane protein structure by NMR are still under development, only a few, model peptides have been well analyzed. The pore-forming peptide melittin was one of the first protein structures to be analyzed using the techniques of two-dimensional NMR, both in detergent micelles (Brown et al., 1982; Inagaki et al., 1989) and in methanol (Bazzo et al., 1988). In both media, melittin was observed to form a bent α-helix, and the angle of the bend was found to be more dynamic than that observed in the X-ray model (Terwilliger et al., 1982), which is not unexpected. Other membrane-active peptide structures have been determined by NMR as well, including neuropeptide Y (McLean et al., 1990), met and leu enkephalin (Behnam and Deber, 1984; Milon et al., 1990), and gramacidin A. The latter peptide is a special example, as the structures determined for this protein by X-ray crystallographic methods (Langs, 1988; Wallace and Ravikumar, 1988; Langs and Triggle, 1992) differ from those determined by NMR (Bystrov and Arseniev, 1988; Ketchem et al., 1993; Cross and Opella, 1994; Prosser et al., 1994). Both structural models have their merits and their advocates, but the NMR model of an end to end β-helical dimer appears to fit the solution and electrophysiological data better. The structures of intrinsic membrane proteins, and portions of larger ones, have also been studied by NMR. Perhaps

the most thoroughly studied examples are the fd and Pf1 filamentous bacteriophage coat protein, both of which appear to form a Γ-shaped structure composed of a membrane-spanning helix linked to an amphipathic helix which lies on the surface of the membrane (Shon et al., 1991; McDonnell et al., 1993). Studies of fragmented membrane proteins by NMR include the single membrane-spanning domain of glycophorin A and bacteriorhodospin, and in the latter case support the structural model for bacteriorhodopsin determined by electron micrographic techniques (Arseniev et al., 1988; Cross and Opella, 1994). While the study of membrane-protein structure by NMR is still limited, recent procedural advances, combined with the exquisite sensitivity of NMR to dynamic structures, suggest that in the future NMR will be one of the most important methods for the study of membrane protein structure.

B.   Electron Microscopy

Of all the techniques available for the study of membrane protein structure, electron microscopy is probably the most versatile. First, the conventions of electron microscopy allow membrane proteins to be studied in their native lipid membranes. In ideal circumstances, the electron microscopic analysis is performed upon membrane proteins quick frozen in glassy (non-crystalline) ice or embedded in glucose, preserving the membrane protein in its normal lipid environment and avoiding the necessity of staining (as sufficient visual contrast is present between the glassy ice or glucose and the protein). This approach requires fairly specialized equipment, but even using standard electron microscopic techniques, the protein can usually be conserved in its native lipids (albeit after staining, drying, or other forms of fixation), avoiding some of the concerns associated with solubilizing membrane protein in detergents and/or solvents. Second, the analysis of large macromolecular assemblies is not precluded by electron microscopy. In fact, huge complexes can been observed, as witnessed by the studies of the nucleopore complex, and important cellular formations, such as interactions between membrane and cytoskeletal proteins, or between integrins and their targets, which currently can not be visualized by other techniques can be visualized in the electron microsope. Third, in rare cases, it is possible to determine the three-dimensional structure of membrane proteins at atomic resolution in the electron microscope by reconstructing the averaged image of two-dimensional arrays of the membrane protein formed in the lipid bilayer. Membrane proteins are predisposed to the formation of two-dimensional crystalline arrays as a consequence of their restriction to the plane of the bilayer, and in some cases these arrays are large enough (i.e., contain sufficient repeating units) and sufficiently well-ordered (i.e., variation between the repeating units is minimized) that high-resolution structural studies can be performed. This combination of conditions is difficult to satsify, and to date the structure of only two membrane proteins has been solved to ~atomic resolution using primarily electron microscopic techniques. However, a considerable number of lower resolution studies have been performed, as summarized below, and while

they are limited in their detail, even so they provide a great deal of fascinating information about the tertiary and quaternary structure of important membrane proteins.

The two membrane proteins whose structures have been solved at atomic resolution are bacteriorhodopsin and the LHC-II complex. Bacteriorhodopsin is a 26,000 Da polytopic membrane protein naturally found in close-packed arrays in the purple membrane of the halophilic bacterium *Halobacterium halobium* (Khorana, 1988). The study of bacteriorhodopsin, in particular, is instructive of the history of membrane protein electron microscopy. Fifteen years elapsed between the first reported three-dimensional structure of bacteriorhodopsin, at a resolution of ~7.5 Å (Henderson and Unwin, 1975), intermediate level refinements (Tsygannik and Baldwin, 1987; Baldwin et al., 1988), and the current, near-atomic resolution structure (Henderson et al., 1990). To achieve this increase in resolution, a number of improvements were utilized. Fused membranes (to generate larger arrays) were adhered to aged sample-holding grids (for better order preservation), embedded in glucose (for contrast), and both tilted and untilted samples were photographed on three separate electron microscopes at near liquid nitrogen temperatures (to minimize sample movement and irradiation damage during imaging). The scanned images were then computationally corrected for local disorder, defocus, and spatial distortion. The result was a structure at a nominal resolution of 3.5 Å (parallel to the plane of the bilayer), but more poorly resolved perpendicular to the plane of the bilayer (Henderson et al., 1990). The poorer resolution perpendicular to the bilayer is inevitable, as it is not possible to collect images tilted greater than 60° relative to the electron beam (imaging tilted specimens is necessary, to collect three-dimensional structural information from two-dimensional arrays, except in the somewhat rare case of well-ordered helical assemblies (Brisson et al., 1985; Unwin, 1995)).

As shown in Figure 4, these technical improvements have allowed the connecting regions between the membrane-spanning α-helices of bacteriorhodopsin to be visualized, including a small connecting helix parallel to the bilayer between helices B and C and suggestions of helical connections between the other membrane-spanning α-helices on the cytoplasmic side of the protein. In addition, the side-chain positions of a number of amino acids important for retinal binding or proton transport were defined. The position of the prosthetic retinal group (which forms a Schiff base with lysine 216 in the middle of helix G) is also clearly established. The retinal is in the all *trans* conformation, extended through the interior of the helical bundle at a slight angle, in agreement with previous dichroism studies (Henderson et al., 1990). All of this information is particularly useful in this instance due to the enormous body of spectroscopic and mutational analyses of bacteriorhodopsin which has already been performed. One of the genuine values of a high-resolution structural determination is that it allows earlier biochemical and biophysical studies to be physically put into context. The 3.5 Å structural determination of bacteriorhodopsin has identified the retinal binding pocket and the

probable geometry of the proton channel, and, in combination with earlier functional data, has suggested an elegant model to explain the movement of protons through bacteriorhodopsin during its photocycle (Henderson et al., 1990). The electron microscopic structural refinement of bacteriorhodopsin is not complete, and the authors suggest future improvements in sample preparation, computational refinement, but mainly, electron microscope design, which they advocate will increase the resolution and ease of structural studies of bacteriorhodopsin and other membrane proteins.

The recent determination of the structure of the plant LHC-II complex at atomic resolution by electron microscopy was also the result of many years of labor. In its role as an antennary protein for photosystem II of green plants, the LHC-II is the most abundant light harvesting protein in the world (Kühlbrandt, 1994). Earlier low
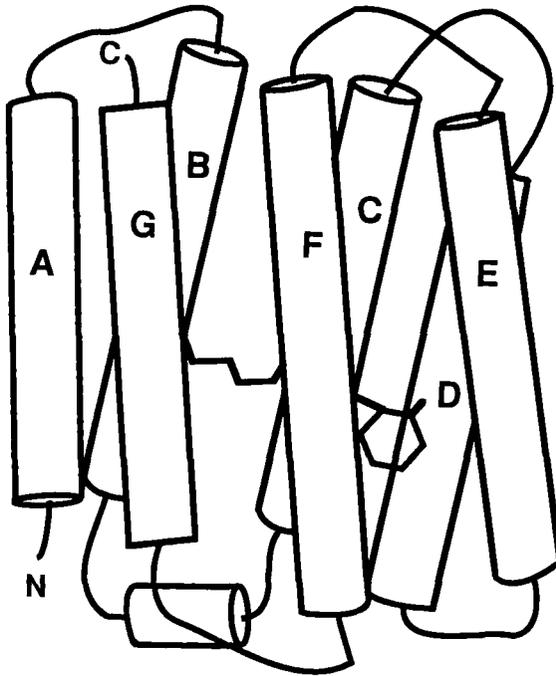
Figure 4.   Schematic representation of bacteriorhodopsin, modified from Figure14a of Henderson et al., 1990. The seven membrane-spanning α-helices of bacteriorhodopsin (as well as the short extra-cellular helix connecting helices B and C) are represented as cylinders, and the connecting regions are merely represented as simple loops. The locations of the extracellular N-terminus (at the A helix) and the cytoplasmic C terminus (at the G helix) are demonstrated, and the disposition of the retinal (attached to Lysine 216 on the G helix) as it slants through the interior of bacteriorhodopsin is also indicated.

resolution projection studies of two-dimensional crystals of the detergent-solub-lized LHC-II (Li and Hollingshead, 1982; Kühlbrandt, 1988; Kühlbrandt and Downing, 1989) were followed by three-dimensional reconstructions at increas-ingly higher resolution (Kühlbrandt, 1984; Kühlbrandt and Wang, 1991) until the most recent refinement, at a nominal resolution of 3.4 Å parallel to the plane of the bilayer (~ 4.9Å resolution perpendicular) (Kühlbrandt et al., 1994). Many of the same techniques utilized in the high-resolution structural determination of bacte-riorhodopsin were also applied here, primarily the use of an electron microscope chilled to 4.2 K to minimize electron beam damage and sample movement. An interesting technical difference is that the LHC-II two-dimensional arrays are actually thin crystals, formed from membranes solubilized in the non ionic deter-gent Triton X-100. However, the detergent micelle still contains native lipids, including one (digalactosyldiacylglycerol) which is necessary for crystal growth and another (phosphatidylglycerol) which is important for trimer formation (Kühl-brandt, 1994).

The protein structure of the 232 amino acid monomer of the LHC-II complex, as shown in Figure 5, is fairly simple (the monomers form trimers in detergent and presumably in the thylakoid membrane as well). The two tilted membrane-spanning α-helices A and B literally cross one another in the middle of the bilayer, where they are linked by salt bridges between arginine and glutamate residues. The C helix is somewhat isolated, acting almost like an independent single membrane-spanning α-helix, and the D helix forms an amphipathic structure which lies along the lumenal side of the bilayer. However, while the protein structure of the LHC-II may be minimal, the architecture of the light-gathering prosthetic groups is complicated. Each LHC-II monomer most likely complexes a minimum of seven chlorophyl *a* molecules, five chlorophyl *b* molecules, and two carotenoids. The chlorophyl molecules are largely positioned in the "open" spaces between the central paired A and B helices, and the flanking C and D helices. Calling upon earlier work, the arrangement of prosthetic groups suggests that photonic energy is first absorbed by the axial chlorophyl *b* molecules, rapidly transferred to the central chlorophyl *a* molecules, and transmitted to the photosystem II reaction center by the Förster mechanism. The two lutein carotenoids flank the central A/B helical supercoil, and function to prevent the formation of toxic oxygen radicals (Kühlbrandt, 1994; Kühlbrandt et al., 1994).

In addition to these high-resolution studies, a number of moderate and low resolution electron microscopic analyses of membrane proteins have been under-taken. While not as detailed as the two structures summarized above, these studies do illustrate both important structural features of membrane proteins and the special utility of electron microscopy in certain applications. In certain cases, the arrayed membrane proteins studied by electron microscopy retain some or all of the biological functions. For example, the light-harvesting proteins discussed above undergo many of the same spectral shifts as do their uncrystallized counterparts (Henderson et al., 1990; Kühlbrandt, 1994), and the aquaporins (a family of water
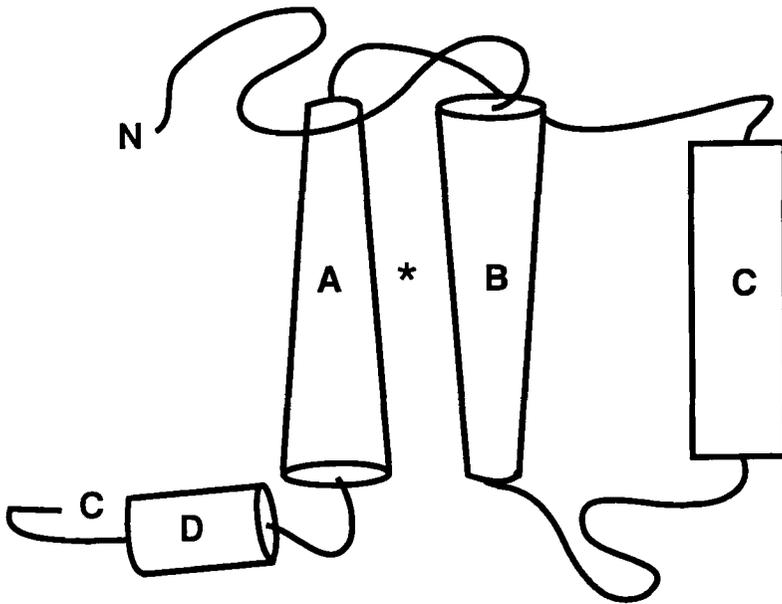
Figure 5. Schematic representation of a monomer of the light-harvesting chlorophyll a/b-protein (LHC-II) complex, modified from Figure 3a of Kühlbrandt et al., 1994. Indicated are the A and B helices which cross the bilayer at an angle of 30°, the somewhat removed C helix, and the ampipathic, surface D helix. The locations of the stromal N terminus and lumenal C terminus are indicated. The asterisk indicates the region of ionic interactions between glutamate and arginine residues on the A and B helices. Not shown are the approximately 12 a and b chlorophylls which interact primarily with the A, B, and C helices, and the luteins (which mirror the paths of the A and B helices and cross in the region indicated by the asterisk).

channels (Engel et al., 1994)) form two-dimensional arrays with characteristic water permeability (Walz et al., 1994). Electron microscopy is also perfectly suitable for the study of large structures. As a case in point, the nuclear pore complex is a massive structure of approximately 120 MDa, and basically too large to be analyzed by other high-resolution techniques. However, several recent three-dimensional electron microscopic studies of the nuclear pore have been performed, and while the ultimate resolution of these studies is limited, they do describe the overall tripartite structure of the pore complex and its dimensions, and suggest the presence of a controversial plug which in some, but not all, studies is observed to occlude the pore (Panté and Aebi, 1994). Much smaller, but still large by normal standards, the structure of the ryanodine receptor (~500,000 Da) has also been

determined in three dimensions at low resolution by electron microscopy, using a novel method based on the analysis of tilted and untilted single particles (Frank et al., 1988), and has been shown to comprise large tetragonal forms reminiscent of the major foot structure of the sarcoplasmic reticulum (Wagenknecht et al., 1989; Radermacher et al., 1992). Electron microscopy has also been used to study the membrane-bound form of soluble toxins, some of whose structures have been determined at atomic resolution by X-ray crystallization techniques. Cholera toxin forms pentameric pores in model membranes, as shown both by electron micrographic image reconstruction (Ribi et al., 1988) and by the relatively novel technique of atomic force microscopy (Yang et al., 1993), while the *Staphylococcus aureus* α-toxin (Ward and Leonard, 1992) and aerolysin (from *Aeromonas hydrophila* (Wilmsen et al., 1992)) appear to form hexameric and heptameric channels, respectively.

Unfortunately, the electron microscopic structural determination of many membrane proteins is still at an intermediate stage, similar to the penultimate analyses of bacteriorhosodpsin and the LHC-II complex. The resolution of electron microscopic structural studies of the nicotinic acetylcholine receptor has steadily increased (Brisson et al., 1985; Kubalek et al., 1987; Toyoshima and Unwin, 1988; Mitra et al., 1989), and now stands at 9 Å (Unwin, 1993). In this determination, the membrane-spanning domain of the acetylcholine receptor appears to be composed of both α-helices and β-strands, and changes in the packing of the channel-forming helices upon addition of the activating agonist acetylcholine have been observed (Unwin, 1995). Conversely, on the basis of electron microscopic analyses of two-dimensional crystals, the voltage-dependent anion channel of the mitochondrial outer membrane appears to form a β-barrel pore (Manella et al., 1986; Thomas et al., 1991), consistent with its functional homology with the bacterial porins. Band 3, the major anion exchanger of the erythrocyte, has been the subject of several electron microscopic studies (Dolder et al., 1993; Wang et al., 1993, 1994), and appears to have a pore-forming structure (a deep, canyonlike depression) quite distinct from the other families of ion channels (Reithmeier, 1993; Wang, 1994). The structures of a number of ATPases have also been studied by electron microscopy (Stokes et al., 1994), and in the case of the Ca^{2+} ATpases, are most consistent with 10 membrane-spanning α-helix models (Stokes and Nakamoto, 1994). Ultimately, as methods and machinery improve, many of the electron microscopic studies briefly referred to above may eventually join the ranks of bacteriorhodopsin and the LHC-II complex as high-resolution structures.

C.   X-ray Crystallography

While membrane proteins are notoriously refractory to crystallization (a necessary first step for X-ray diffraction analysis), more high-resolution membrane protein structures have been determined by this technique than any other, largely due to its impressive sensitivity once useful three-dimensional crystals are obtained. Crystallization of membrane proteins is difficult for several reasons. Membrane

proteins are difficult to purify in the large amounts required for crystallization trials, many membrane proteins are variably glycosylated and thus heterogeneous, and the habit of membrane proteins to form two-dimensional arrays often leads to thin crystals, or crystals disordered in the third dimension. However, the main complicating factor is the detergent. In order to produce three-dimensional crystals, it is first necessary to remove the membrane protein from the strictures of the bilayer by solubilization in non-denaturing detergent. Once a membrane protein is disposed in monomeric detergent micelles, it can be subjected to the same battery of crystallization methods as soluble proteins. However, the additional detail of determining the appropriate detergent type and concentration, as well as the method of solubilization (to which membrane proteins can be very sensitive), further exacerbates the already complicated process of optimizing crystallization conditions (pH, ionic strength, temperature, precipitant type and concentration, etc.). Detergent solubilization of membrane proteins prior to crystallization raises two more points. First is the general concern that solubilization in detergent may alter the conformation of the protein in gross or subtle ways, and that the structure determined by X-ray diffraction techniques will not reflect the true situation in vivo. For that reason, the detergents used are typically quite mild (mainly lauryl dimethylamine oxide (Michel, 1983; Allen and Feher, 1984; Papiz et al., 1989; Weiss and Schulz, 1991; Kreusch et al., 1994), β-octylglucoside (Allen and Feher, 1984; Garavito et al., 1984; Picot et al., 1994) and the polyoxyethylene detergent C_8E_5 (Weiss and Schulz, 1992), and in some cases it has been possible to visualize detergent molecules in the crystal, either directly (Deisenhofer and Michel, 1989; Weiss and Schulz, 1992; Kreusch et al., 1994) or by neutron diffraction analysis of deuterated detergent (Roth et al., 1989). Second, unless steps are taken to actively remove lipid (as has been the case with some photosynthetic reaction center and porin studies), the protein will still be exposed to native lipids in the form of mixed lipid/detergent micelles.

Despite crystallization problems, X-ray crystallographic studies have provided a great deal of structural information about membrane protein. The structures of many membrane-associated proteins have been solved at atomic resolution, including soluble pore-forming peptides and proteins (e.g., alamethecin (Fox and Richards, 1982), melittin (Terwilliger et al., 1982), gramacidin A (Langs, 1988; Wallace and Ravikumar, 1988), proaerolysin (Parker et al., 1994), the cytotoxic domain of colicin A (Parker et al., 1989), and perhaps toxin γ from cobra venom (Bilwes et al., 1994)), and the structures of cleaved soluble domains (e.g., the tyrosine kinase domain of the human insulin receptor (Hubbard et al., 1994), the intact lumenal domain of cytochrome f (Martinez et al., 1994), the entire extracellular region of the HA_2 subunit of influenza virus hemagglutinin (Weis et al., 1990; Baker and Agard, 1994), part of the extracellular domain of the rat CD4 receptor (Lange et al., 1994), and the ligand-binding domain of a bacterial aspartate receptor (Scott et al., 1993). However, the structures of only four intrinsic membrane proteins have been determined to atomic resolution by X-ray crystallographic methods, two of them within the last 18 months. These structures are: the photo-

synthetic reaction centers and the LH2 light-harvesting antennary complex of different species of photosynthetic bacteria, members of the porin family of bacterial outer membrane pore-forming proteins, and the mammalian prostaglandin H synthase.

The photosynthetic reaction center is a multisubunit photon-receptor complex localized to the thylakoid membranes of photosynthetic bacteria. This photosynthetic complex (from *Rhodospeudomonas viridis* and *R. sphaeroides*) was the first intrinsic membrane protein structure to be described at atomic resolution (2.65-3 Å), and has been extensively reviewed (Deisenhofer and Michel, 1989; Rees et al., 1989; Deisenhofer and Michel, 1991a,b; Okamura and Feher, 1992). The basic oligomer is formed from three subunits; the bitopic H subunit, and the polytopic L and M subunits. In some bacterial species, an extrinsic cytochrome subunit is associated with the periplasmic face of the intrinsic membrane portion of the complex. One of the most striking features of the photosynthetic reaction center is the helical bundle formed by the pseudo-symmetric dimer association of the L and M subunits, as shown in Figure 6 (the single membrane-spanning helix of the H subunit is also indicated). Most aspects of this helical domain have already been discussed above (in point of fact, many of the general characteristics attributed to membrane-spanning α-helices were extracted from detailed consideration of the photosynthetic reaction center structure, which for many years was the only available model).

A primary role of this helical bundle, which we have not discussed, is to bind and align a spatially-linked series of prosthetic groups which mediate the initial charge separation and consequent electron transfer that follow absorption of light energy. Each L or M subunit binds two bacteriochlorophylls, a bacteriopheophytin, and a quinone, and the pair of subunits share a free iron and a carotenoid between them (the specific types of chlorophyl pheophytin, quinone or carotenoid varies between species). Two shared bacteriochlorophylls (the spectrally-identified "special pair") form the origin of two electron transfer pathways which pass down the L and M subunits (Deisenhofer and Michel, 1989; Rees et al., 1989; Deisenhofer and Michel, 1991a,b; Okamura and Feher, 1992). A particular conundrum of the reaction center structure is that these two pathways appear almost identical in the crystal structure, but from spectral and kinetic studies it is known that transfer down one pathway is greatly preferred (Michel-Beyerle et al., 1988). A further concern was that the architecture of the prosthetic groups appeared to be significantly different between different species of photosynthetic bacteria. This concern was largely resolved by a recent, higher-resolution study of the discrepant species, performed using a different crystal form, in which the altered prosthetic groups positions were not observed (Ermler et al., 1994). As with bacteriorhodopsin and the LHC-II complex, the availability of a high resolution structure has allowed earlier models of electron transfer through the photosynthetic reaction center to be rigorously assessed (Okamura and Feher, 1992).

While the LH2 antennary complex of *R. acidophila* is chronologically the most recent membrane protein to be studied at recent high resolution by X-ray crystal-
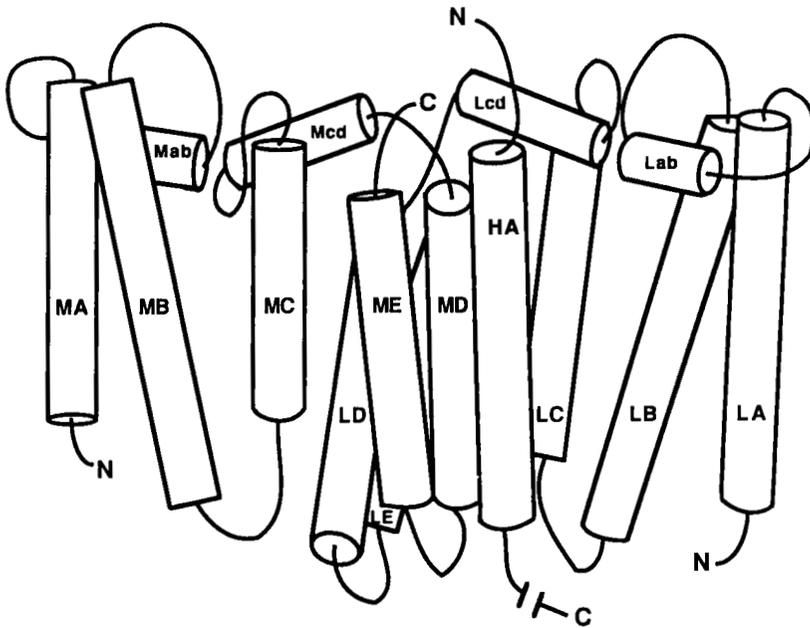
Figure 6. Schematic representation of the membrane-associated portion of the photosynthetic reaction center of *Rhodobacter sphaeroides*, modified from Figure 5 of Allen et al. 1987. The membrane-spanning α-helices of the L, M, and H subunits are represented as cylinders (as are the periplasmic membrane-surface helices which connect the A and B, and C and D helices of the L and M subunits). The helices are denoted according to subunit type (e.g., MA is helix A of subunit M, while Mab is the helix which connects membrane-spanning helices A and B of subunit M). The symmetry of the L and M subunits is apparent. Most of the H subunit is omitted, as are the cofactors involved in electron transport.

lography, it is appropriate to consider its structure in relation to the photosynthetic reaction center. The function of the LH2 complex is to capture and rapidly transfer light energy to the photosynthetic reaction center, and *in vivo* the two complexes are in close proximity in the bacterial membrane (McDermott et al., 1995). The LH2 complex structure, determined at 2.5Å resolution, illustrates several of the features common to membrane proteins discussed above. As depicted in Figure 7, the individual membrane-spanning α-helices of the bitopic α- and β-apoproteins are regular, apparently stable structures which further assemble (primarily via charge interactions at the ends of the helices and by complexing two prosthetic chlorophyll molecules) to form a simple dimer. Nine of these dimers then cyclize, and again the primary drving force appears to be complexation of additional
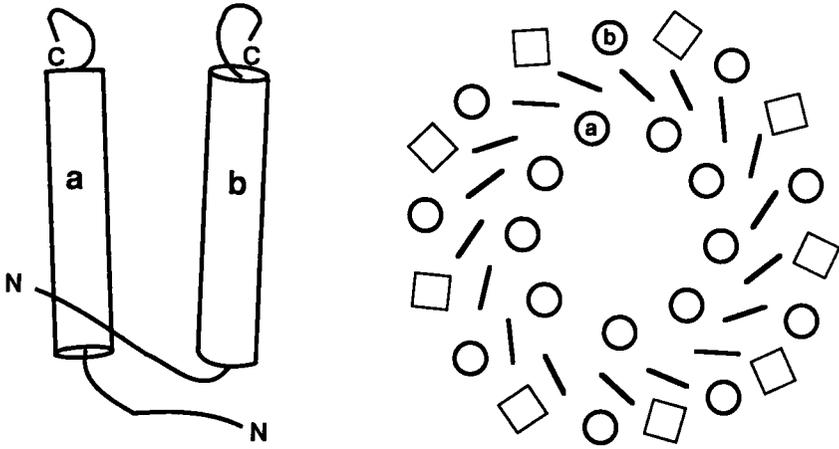
Figure 7. Schematic representation of the light-harvesting antenna (LH2) complex from *Rhodopseudomonas acidophila*, modified from Figures 3 and 4 of McDermott et al., 1995. On the left is depicted an a- and b-apoprotein dimer. The amino termini are on the cytoplasmic side of the bilayer, with the N-terminus of the a-apoprotein apparently buried ~9Å within the presumed bilayer region, where it coordinates with the magnesium of a bacteriochlorophyll. The carboxy termini are periplasmic. On the right, the manner in which the nine a- and b-apoprotein dimers assemble in the form of a nonameric ring is demonstrated (in this end-on view, the helical a- and b-apoproteins are represented as circles). Eighteen bacteriochlorophylls (depicted as dashes), oriented roughly perpendicular to the membrane normal, are sandwiched between the a- and b-aproproteins, while an additional nine bacteriochlorophylls (oriented parallel to the membrane surface and depicted as squares) are positioned between the b-apoproteins of the outer ring. The architecture of the bacteriochlorophylls presumably allows for the rapid lateral transfer of light energy to adjoining light-harvesting complexes and photosynthetic reaction centers.

chlorophyll, and perhaps carotenoid, molecules. The local environments of the chlorophyll molecules are different (as are the orientations of the planar faces of the chlorophylls), allowing the chlorophylls to absorb light across a broader spectrum. Furthermore, the ring of closely-associated chlorophyls allow light energy to be passed radially toward a waiting photosynthetic reaction or other antennary complex. Interestingly, the oligomeric structure of the LH2 complex (i.e., the number of dimers involved in forming the ring) appears to vary (Kühlbrandt, 1995), suggesting that the LH2 complex may not be limited to a single oligomeric form.

The porin family of outer membrane proteins also form ring-shaped structures, but differ greatly from the LH2 complex in their function and construction. Porins

form nonselective or partially selective trimeric pores in the outer membranes of prokayotes, and typically function to allow the passive import of nutrient molecules. The singular β-barrel structure of the porins was predicted well in advance of high resolution crystallopraphic and electron microscopic studies on the basis of circular drichroism, infrared, and Raman spectroscopic studies, which indicated that matrix porin (the most studied form) was at least 60-65% β-sheet (Schindler and Rosenbusch, 1978, 1984; Kleffel et al., 1985; Vogel and Jahnig, 1986; Nabedryk et al., 1988). While highly-ordered three-dimensional crystals of matrix porin from *E. coli* were available for over a decade (but unusable due to difficulties associated with generating isomorphic derivatives for phasing (Garavito and Rosenbusch, 1980)), the first porin structure solved at atomic resolution was of a nonspecific porin from the outer membrane of the photosynthetic bacteria *Rhodobacter capsulatus* (Weiss and Schulz, 1991; Weiss et al., 1992). As predicted on the basis of spectroscopic and electron microscopic work, the *R. capsulatus* porin monomers are composed of a 16-strand, antiparallel β-barrel (Figure 8), linked according to the nearest-neighbor rule which also applies for the α-helical bundles of bacteriorhodopsin and the photosynthetic reaction center. The strands of the β-barrel are highly tilted (30-60°) relative to the membrane normal, generating a squat, wide pore which is narrowed to approximately 7 × 9Å (the barrel is not circular, but rather an oval flattened on one face) by the interposition of an extracellular helical loop (L3) between strand number 5 and 6 (Weiss and Schulz, 1992; Cowan, 1993). However, the true dimensions of the pore, in terms of solute passage, are only about 3 × 6Å, as ordered water molecules observed in the structure are likely to constrict the pore even more (Weiss and Schulz, 1992). Furthermore, some of the β strands of *R. capsulatus* porin are estimated to be only six residues long, as the wall dimensions vary between 40 Å (on the outside, membrane-facing surface) and 20 Å (on the internal, trimer interface).

The structures of matrix porin (also called OmpF) and phosphoporin (also called PhoE) of *E. coli* are similar to that of *R. capsulatus* porin in general, but differ in several important details. All three porins are composed of 16 strand β-barrels. However, the *E. coli* β-barrels are taller and slimmer, as the β-strands are less tilted (35-50°) relative to the membrane normal, and longer (the shortest strands are 7-9 residues) (Cowan et al., 1992). However, matrix porin has the largest basic pore size (7 × 11Å) of all the porins, as the interposing L3 loop is backed more closely to the barrel wall. Except for L3, none of the connecting loops of the *E. coli* porins are ordered, while three of the loops (all extracellular) in *R. capsulatus* porin contain short a-helices (Weiss et al., 1991). The trimer interface of the porins is also interesting, and somewhat controversial. Early three-dimensional electron microsocopic studies of matrix porin suggested that the three pores observed on one face of the membrane merged to one pore on the other side (Engel et al., 1985), although this feature was not observed in higher resolution electron microscopic reconstructions performed upon the related phosphorin PhoE (Jap, 1989; Jap et al., 1991) or in atomic force microscopy topographs of matrix porin (Schabert and Engel, 1994).
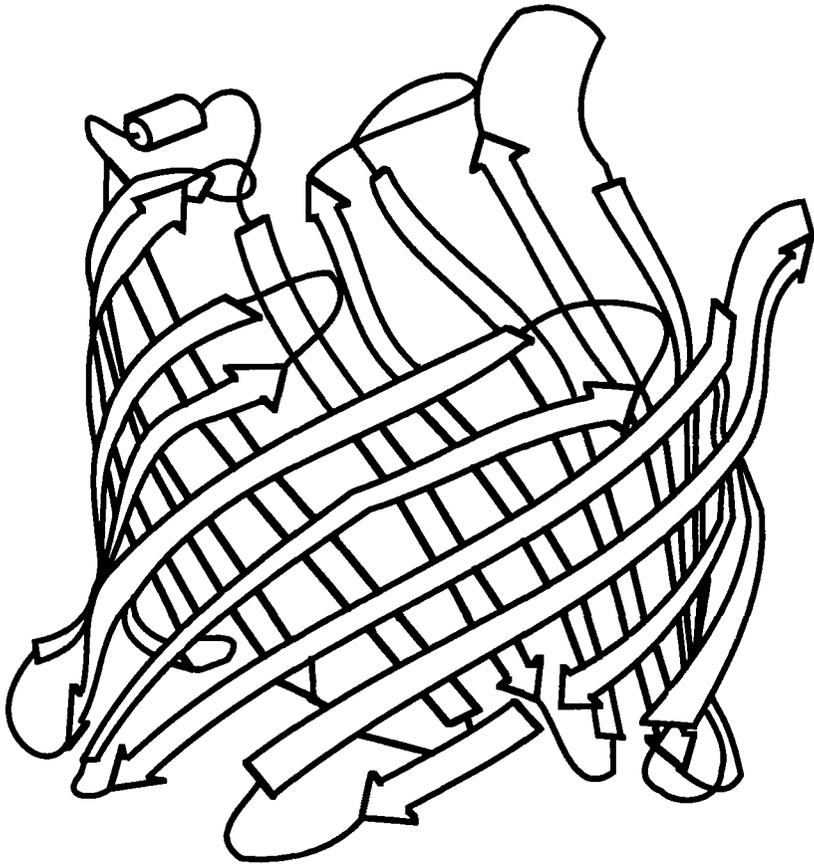
Figure 8. Schematic representation of a monomer of *Escherichia coli* matrix porin (OmpF), modified from Figure 2 of Cowan, 1993. The β-strands of the 16 member antiparallel β-barrel are represented as pointed ribbons. The front face of the model corresponds to the trimer interface region of the porin oligomer. The N and C termini are located in the interrupted strand at the front of the model. The L3 loop, which primarily determines the usable diameter of the porin channel, is not indicated, but a short helix connecting strands 10 and 11 is apparent on the extracellular surface of the monomer. The β strands are more tilted, and somewhat shorter, in the analogous view of the *Rhodobacter capsulatus* porin, while maltoporin is composed of an otherwise similar eighteen-stranded β-barrel.

While it had been suggested that the region of low density at the porin trimer interface might be occupied by "trapped" lipopolysaccharide (Jap, 1989), at least

for the *E. coli* porins it appears that the weak scattering density of the bulky, hydrophobic amino acid side-chains concentrated at the trimer interface, combined with the narrowness of the protein structure in this region, led to the appearance of a third "pore" at this position in some early electron microscopic studies (Cowan et al., 1992). Another photosynthetic bacteria porin structure, from *R. blastica*, has also been solved of late (Kreusch et al., 1994), and appears to be more similar to the *E.coli* porins (in terms of β-strand length, β-strand tilt-angle, and loop structure) than to *R. capsulatus* porin. Even more recently, the structure of the selective *E. coli* porin maltoporin (product of the lamB gene) has been solved at atomic resolution by X-ray diffraction techniques (Schirmer et al., 1995). This porin is composed of an 18-strand β-barrel, but is more constricted than other known porins due to the folding of three connecting loops (primarily the functionally-conserved L3) back into the pore. Maltoporin functions specifically in the uptake of maltose and other disaccharides, and a potential maltose transport pathway, composed of a series of hydrophobic aromatic side-chains which descend down the pore in a sinister spiral, has been proposed (Schirmer et al., 1995).

The last structure to consider is prostaglandin H synthase, which is unlike all of the other intrinsic membrane protein structures we have discussed. Prostaglandin H synthase (cyclooxygenase) is the target for the class of therapeutics known as non-steroidal anti-inflammatory drugs (more familiarly NSAIDs), which includes aspirin, acetaminophen, ibuprofen, etc. These drugs bind to the active site of prostaglandin H synthase and block the conversion of arachidonic acid (a lipid soluble substrate) into the vasoconstrictor and pro-inflammatory compound pro-staglandin H_2 (Garavito et al., 1994). Prostaglandin H synthase is an intrinsic membrane protein according to the standard operational definition (i.e., it remains associated with the membrane unless extracted by detergent or organic solvents and it contains no bound lipid (Picot et al., 1994)), but its primary structure gave no clues as to potential membrane-spanning regions. This conundrum was partially resolved by the determination of the structure of prostaglandin H synthase to 3.5 Å resolution by X-ray diffraction techniques (Picot et al., 1994), which indicated a surface layer of amphipathic α-helices, with the hydrophobic faces atypically turned outward from the interior of the protein. As shown in Figure 9, this structure most likely acts to bind prostaglandin H synthase to the cell surface or lumenal side of the endoplasmic reticulum membrane by embedding partially into the hydro-phobic interior of the bilayer in the manner of a monotopic protein. While this mechanism of membrane insertion of prostaglandin H synthase is still not explicitly proven (the location of the membrane can only be inferred in the crystal structure), it provides both a means of attaching prostaglandin H synthase to the membrane, and potential pathway from the catalytic site (in a conserved peroxidase fold catalytic domain) to the lipid bilayer. The structure (and function) of prostaglandin H synthase is similar to other enzymes, including arachidonic acid 15-lipooxy-genase (Boyington et al., 1993) and cytochrome P_{450} BM-3 (Ravichandran et al., 1993), which contain similar catalytic domains but are not intrinsic membrane
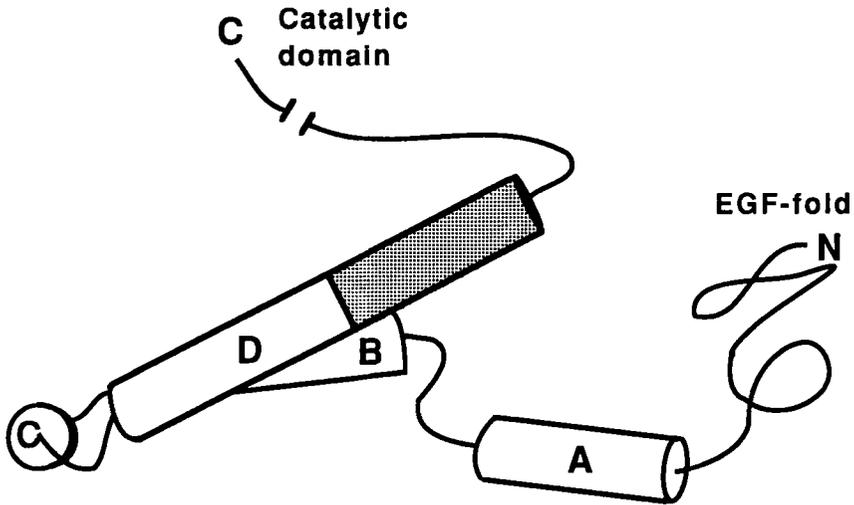
Figure 9. Schematic representation of the presumed membrane-binding domain of bovine prostaglandin H synthase, modified from Figure 1 of Garavito et al., 1994. Indicated are the positions of the bilateral amphipathic A, B, and C helices, which apparently dip into the hydrophobic interior of the bilayer and both anchor the protein to the membrane and provide a passage from the active site of the catalytic peroxidase domain for the apolar substrate. Also indicated is the the segmentally amphipathic D helix, which also functions to link the enzyme to the surface of the membrane, as well as the hydrophilic amino terminus with its well-characterized EGF-like fold. The catalytic domain, a complex helical bundle, is not depicted.

proteins. This suggests that, during evolution, members of this (and other) classes of soluble proteins could develop access to the interior of the lipid bilayer by appending a monotopic membrane-binding domain such as the one postulated for prostaglandin H synthase (Garavito et al., 1994).

IV. CONCLUSION

Initially, it was thought that membrane protein structures, specifically the membrane-spanning domain, would be restricted to simple helical structures composed almost entirely of hydrophobic amino acids. Therefore, it may just be happenstance that, of the six structures known at or near atomic resolution, one (the porins) is composed entirely of β-strands, one (prostaglandin H synthase) is a monotopic protein bound to the membrane by amphipathic surface helices, and several contain polar or even charged residues deep in the hydrophobic interior of the lipid bilayer

(including the salt bridges observed between helices in the LHC-II complex). Alternatively, it may be that membrane protein structure will be just as diverse and inventive as its soluble counterpart. The membrane protein structures outlined above were not "overnight successes" but rather the result of years (in some cases, decades) of concentrated labor. The near future promises the exposition of even more membrane protein structures at high resolution, hopefully accelerated by advances in the state of the art. Optimization of NMR methodology and data analysis should allow the dynamic structures of small to medium sized membrane proteins to be observed, and may prove particularly useful in the study of proteins which assume multiple conformations and resist crystallization. Electron micro- scopic studies are always ongoing, and the structures of band 3, the nicotinic acetylcholine receptor, the light-harvesting complex I, gap junction, and Ca^{2+} ATPase, may be the next to be determined at the atomic level, especially as improvements in electron microscope design and computational methods proceed. Finally, a number of variations are being applied to the study of membrane protein structure by X-ray diffraction techniques, including the use of cryo-protection measures (to prolong crystal "life" in the X-ray beam), utilizing synchrotron X-ray sources to obtain a more intense beam and reduce the amount of time needed to collect X-ray data, and expressing large amounts of membrane proteins in highly efficient heterologous expression systems to generate the required three-dimen- sional crystals. In combination, these parallel improvements should lead to the discovery of many novel, and intrinsically fascinating, membrane protein struc- tures.

REFERENCES

Abrahams, J. P., Leslie, A. G. W., Lutter, R., & Walker, J. E. (1994). Structure at 2.8 Å resolution of F_1-ATPase from bovine heart mitochondria. Nature 370, 621-628.

Akabas, M. H., Stauffer, D. A., Xu, M., & Karlin, A. (1992). Acetylcholine receptor channel structure probed in cysteine-substitution mutants. Science 258, 307-310.

Allen, J. P., & Feher, G. (1984). Crystallization of reaction center from *Rhodopseudomonas sphaeroides*: Preliminary characterization. Proc. Natl. Acad. Sci. USA 81, 4795-4799.

Allen, J. P., Feher, G., Yeates, T. O., Komiya, H., & Rees, D. C. (1987). Structure of the reaction center from *Rhodobacter sphaeroides* R-26: The protein subunits. Proc. Natl. Acad. Sci. USA 84, 6162-6166.

Argos, P. (1989). In: Protein Structure: A Practical Approach (Creighton, T. E., ed.), pp. 169-190, IRL Press, New York.

Arseniev, A. S., Maslennikov, I. V., Bystrov, V. F., Kozhich, A. T., Ivanov, V. T., & Ovichinnikov, Y. A. (1988). Two-dimensional 1H NMR study of bacteriorhodopsin-(34-65)-polypeptide conformation. FEBS Lett. 231, 81-88.

Attwood, T. K., & Findlay, J. B. C. (1994). Finger-printing G-protein-coupled receptors. Protein Eng. 7, 510-513.

Aubry, A., Birlirakis, N., Sakarellos-Daitsiotis, M., Sakarellos, C., & Marraud, M. (1989). A crystal molecular conformation of leucine-enkephalin related to the morphine molecule. Biopolymers 28, 27-40.

Baker, D., & Agard, D. A. (1994). Influenza hamagglutinin: Kinetic control of protein function. Structure 2, 907-910.

Baldwin, J. M., Henderson, R., Beckman, E., & Zemlin, F. (1988). Images of purple membrane at 2.8 Å resolution obtained by cryo-electron microscopy. J. Mol. Biol. 202, 585-591.

Barber, J. (1994). F1-ATPase structure: Getting our bearings for ATP synthesis. Structure 2, 889-890.

Bayley, H., Huang, K.-S., Radhakrishnan, R., Ross, A. H., Takagaki, Y., & Khorana, H. G. (1981). Site of attachment of retinal in bacteriorhodopsin. Proc. Natl. Acad. Sci. USA 78, 2225-2229.

Bazzo, R., Tappin, M. J., Pastore, A., Harvey, T. S., Carver, J. A., & Campbell, I. D. (1988). The Structure of melittin: A 1H-NMR study in methanol. Eur. J. Biochem. 173, 139-146.

Bechinger, B., Zasloff, M., & Opella, J. J. (1993). Structure and orientation of the antibiotic peptide magainin in membranes by solid-state nuclear magnetic resonance spectroscopy. Protein Sci. 2, 2077-2084.

Behnam, B. A., & Deber, C. M. (1984). Evidence for a folded conformation of methionine-and leucine-enkephalin in a membrane environment. J. Biol. Chem. 259, 14935-14940.

Bernheimer, A. W., & Rudy, B. (1986). Interactions between membranes and cytolytic peptides. Biochim. Biophys. Acta 864, 123-141.

Betz, H. (1990). Homology and analogy in transmembrane channel design: Lessons from synaptic membrame proteins. Biochemistry 29, 3591-3599.

Bilwes, A., Rees, B., Moras, D., Ménez, R., & Ménez, A. (1994). X-ray structure at 1.55 Å of Toxin γ, a cardiotoxin from I venom. J. Mol. Biol. 239, 122-136.

Blobel, G. (1980). Intracellular protein topogenesis. Proc. Natl. Acad. Sci. USA 77, 1496-1500.

Bordier, C. (1981). Phase separation of integral membrane proteins in Triton X-114 solution. J. Biol. Chem. 256, 1604-1607.

Bormann, J., Hamill, O. P., & Sakmann, B. (1987). Mechanism of anion permeation through channels gated by glycine and γ-aminobutyric acid in mouse cultured spinal neurons. J. Physiol. 385, 243-286.

Boyington, J. C., Gaffney, B. J., & Amzel, L. M. (1993). The three-dimensional structure of an arachidonic acid 15-lipoxygenase. Science 260, 1482-1486.

Brandl, C., & Deber, C. M. (1986). Hypothesis about the function of membrane-buried proline residues in transport proteins. Proc. Natl. Acad. Sci. USA 83, 917-921.

Brandts, J. F., Halvorson, H. R., & Brennan, M. (1975). Consideration of the possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. Biochemistry 14, 4953-4963.

Brisson, A., & Unwin, P. N. T. (1985). Quaternary structure of the acetylcholine receptor. Nature 315, 474-477.

Brown, L. R., Braun, W., Kumar, A., & Wuthrich, K. (1982). High resolution nuclear magnetic resonance studies of the conformation and orientation of melittin bound to a lipid-water interface. Biophys. J. 37, 319-328.

Bystrov, V. F., & Arseniev, A. S. (1988). Diversity of the gramacidin A spatial structure in solution: Two-dimensional 1H NMR study in solution H. Tetrahedron 44, 925-940.

Camerman, A., Mastrapaolo, D., Karle, I., Karle, J., & Camerman, N. (1983). Crystal structure of leucine-enkephalin. Nature 306, 447-450.

Capaldi, R. A., & Vanderkooi, G. (1972). The low polarity of many membrane proteins. Proc. Natl. Acad. Sci. USA 69, 930-932.

Cascio, M., & Wallace, B. A. (1985). The effects of environment on the conformation of the alamethicin membrane channel. Ann. N.Y. Acad. Sci. 435, 527-529.

Castresana, J., Fernandez-Ballester, G., Fernandez, A. M., Laynez, J. L., Arrondo, J.-L. R., Ferragut, J. A., & Gonzalez-Ros, J. M. (1992). Protein structural effects of agonist binding to the nicotinic acetylcholine receptor. FEBS Lett. 314, 171- 175.

Catterall, W. A. (1994). Molecular properties of a superfamily of plasma-membrane cation channels. Curr. Opin. Cell Biol. 6, 607-615.

Chothia, C., & Murzin, A. M. (1994). New folds for all-β Proteins. Structure 1, 217-222.

Chou, K. Y., Carlacci, L., & Maggiora, G. G. (1990). Conformational and geometrical properties of idealized β-barrels in proteins. J. Mol. Biol. 213, 315-326.

Cortijo, M., Alonso, A., Gomez-Fernandez, J. C., & Chapman, D. (1982) Intrinsic protein-lipid interactions: Infrared spectroscopic studies of gramacidin A, bacteriorhodopsin and Ca^{2+} ATPase in biomembranes and reconstituted systems. J. Mol. Biol. 157, 597-618.

Cowan, S. W., Schirmer, T., Rummell, G., Steiert, M., Ghosh, R., Pauptit, R. A., Jansomius, J. N., & Rosenbusch, J. P. (1992). Crystal structures explain functional properties of two E. coli Porins. Nature 358, 727-733.

Cowan, S. W. (1993). Bacterial porins: Lessons from three high-resolution structures. Curr. Opin. Struct. Biol. 3, 501-507.

Cramer, W. A., Engelman, D. M., von Heijne, G., & Rees, D. C. (1992). Forces involved in the assembly and stabilization of membrane proteins. FASEB J. 6, 3397-3402.

Cross, T. A., & Opella, S. J. (1994). Solid-state NMR structural studies of peptides and proteins in membranes. Curr. Opin. Struct. Biol. 4, 574-581.

Cserháti, T., & Szögyi, M. (1993). Interaction of phospholipids with proteins and peptides. New advances III. Int. J. Biochem. 25, 123-146.

Deber, C. M., Brandl, C. J., Deber, R. B., Hsu, L. C., & Young, X. K. (1986). Amino acid composition of the membrane and aqueous domains of integral membrane proteins. Arch. Biochem. Biophys. 251, 149-157.

Deisenhofer, J., Epp, O., Miki, K., Huber, R., & Michel, H. (1984). Structure of the protein subunits in the photosynthetic reaction center of Rhodopseudomonas viridis at 3 Å resolution. Nature 318, 618-624.

Deisenhofer, J., & Michel, H. (1989). The photosynthetic reaction center from the purple bacterium Rhodopseumonas viridis. Science 245, 1463-1473.

Deisenhofer, J., & Michel, H. (1991a). High-resolution structures of photosynthetic reaction centers. Ann. Rev. Biophys. Biophys. Chem. 20, 247-266.

Deisenhofer, J., & Michel, H. (1991b). Structures of bacterial photosynthetic reaction centers. Ann. Rev. Cell. Biol. 7, 1-23.

Dencher, N. A., & Heyn, P. M. (1979). Bacteriorhodopsin monomers pump protons. FEBS Lett. 108, 307-310.

Dolder, M., Walz, T., Hefti, A., & Engel, A. (1993). Human erythrocyte band 3: Solubilization and reconstitution in two-dimensional crystals. J. Mol. Biol. 231, 119-132.

Donelly, D., & Findlay, J. B. C. (1994). Seven-helix receptors: Structure and modelling. Curr. Opin. Struct. Biol. 4, 582-589.

Downer, N. W., Bruchman, T. J., & Hazzard, J. H. (1986) Infrared spectroscopic study of photoreceptor membrane and purple membrane. Protein secondary structure and hydrogen deuterium exchange. J. Biol. Chem 261, 3640-3647.

Dwyer, B. P. (1988). Evidence for the extramembranous location of the putative amphipathic helix of acetylcholine receptor. Biochemistry 27, 5586-5592.

Eisenberg, D. (1984). Three-dimensional structure of membrane and surface proteins. Ann. Rev. Biochem. 53, 595-623.

Engel, A., Massalski, A., Schindler, H., Dorset, D. L., & Rosenbusch, J. P. (1985). Porin channel triplets merge into single outlets in Escherichia coli outer membranes. Nature 317, 643-645.

Engel, A., Walz, T., & Agre, P. (1994). The aquaporin family of membrane water channels. Curr. Opin. Struct. Biol. 4, 545-553.

Engelman, D. M., Henderson, R., McLachlan, A. D., & Wallace, B. A. (1980). Path of the polypeptide in bacteriorhodopsin. Proc. Natl. Acad. Sci. USA 77, 2023-2027.

Engelman, D. M., & Steitz, T. A. (1981). The spontaneous insertion of proteins into and across membranes: The helical hairpin hypotheses. Cell 23, 411-422.

Engelman, D. M., Steitz, T. A., & Goldman, A. (1986). Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Ann. Rev. Biophys. Biophys. Chem. 15, 321-353.

Ermler, U., Fritzsch, G., Buchanan, S. K., & Michel, H. (1994). Structure of the photosynthetic reaction centre from *Rhodobacter sphaeroides* at 2.65 Å resolution: Cofactors and protein-cofactor interactions. Structure 2, 925-936.

Fasman, G., & Gilbert, W. A. (1990). The prediction of transmembrane protein sequences and their conformation: An evaluation. Trends Biochem. Sci. 15, 89-92.

Fong, T. M., & McNamee, M. G. (1987). Stabilization of acetylcholine receptor secondary structure by cholesterol and negatively charged phospholipids in membranes. Biochemistry 26, 3871-3880.

Forte, M., Guy, H. R., & Manella, C. A. (1987). Molecular genetics of the VDAC ion channel: Structural model and sequence analysis. J. Bioenerg. Biomembr. 29, 341-350.

Fox, R. O., Jr., & Richards, F. M. (1982). A voltage-gated ion channel model inferred from the crystal structure of alamethicin at 1.5-A resolution. Nature 300, 325-330.

Frank, J., Verschoor, A., Wagenknecht, T., Radermacher, M., Carazo, J.-M. (1988). A new non-crystallographic image-processing technique reveals the architecture of ribosomes. TIBS 13, 123-127.

Galzi, J.-L., Revah, F., Bessis, A., & Changeux, J.-P. (1991). Functional architecture of the nicotinic acetylcholine receptor: From electric organ to brain. Ann. Rev. Pharmacol. 31, 37-72.

Galzi, J.-L., & Changeux, J.-P. (1994). Neurotransmitter-gated ion channels as unconventional allosteric proteins. Curr. Opin. Struct. Biol. 4, 554-565.

Garavito, R. M., & Rosenbusch, J. P. (1980). Three-dimensional crustals of an integral membrane protein: An initial X-ray analysis. J. Cell Biol. 86, 327-329.

Garavito, R. M., Hinz, U., & Neuhaus, J.-M. (1984). The crystallization of outer membrane proteins from *Escherichia coli*. Studies of *lamB* and *ompA* gene products. J. Biol. Chem. 259, 4254-4257.

Garavito, R. M., Picot, D., & Loll, P. J. (1994). Prostaglandin H synthase. Curr. Opin. Struct. Biol. 4, 529-535.

Gibson, N., J., & Cassim, J. Y. (1989). Evidence for an all-type helical conformation for bacteriorhodopsin in the purple membrane. Biochemistry 28, 2134-2139.

Gierasch, L. M. (1989). Signal sequences. Biochemistry 28, 923-930.

Glaeser, R. M., & Jap, B. K. (1985) Absorption flattening in the circular dichroism spectra of small membrane fragments. Biochemistry 24, 6398-6401.

Guy, H. R., & Conti, F. (1990). Pursuing the structure and function of voltage-gated channels. Trends Neurol. Sci. 13, 201-206.

Hall, J. E., Vodyanoy, I., Balasubramanian, T. M., & Marshall, G. R. (1984). Alamethicin: A rich model for channel behavior. Biophys. J. 45, 233-247.

Hausdorff, W. P., Caron, M. G., & Lefkowitz, R. J. (1990). Turning off the signal: Desensitization of β-adrenergic receptor function. FASEB J. 4, 2881-2889.

Henderson, R. (1975). The structure of the purple membrane from *Halobacterium halobium*: Analysis of the X-ray diffraction pattern. J. Mol. Biol. 93, 123-138.

Henderson, R., & Unwin, P. N. T. (1975). Three-dimensional model of purple membrane obtained by electron microscopy. Nature 257, 28-32.

Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E., & Downing, K. H. (1990). Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. J. Mol. Biol. 213, 899-929.

Henry, G. D., & Sykes, B. D. (1994). Methods to study membrane protein structure in solution. Meth. Enzymol. 239, 515-535.

Heyn, M. P. (1989). Circular dichroism for determining secondary structure and state of aggregation of membrane proteins. Meth. Enzymol. 172, 575-584.

Higgins, C. F. (1992). ABC transporters: From microorganisms to man. Ann. Rev. Cell Biol. 8, 67-113.

Hille, B. (1992). Ionic Channels of Excitable Membranes, 2nd ed., pp. 183-186. Sinauer Associates, Inc., Sunderland MA.

Hol, W. G.-J., Halie, L. M., & Sander, C. (1981). Dipoles of the α-helix and β-sheet: Their role in protein folding. Nature 294, 532-536.

Hubbard, S. R., Eei, L., Ellis, L., & Hendrickson, W. A. (1994). Crystal structure of the tyrosine kinase domain of the human insulin receptor. Nature 372, 746-754.

Hynes, R. O. (1992). Integrins: Versatility, modulation, and signaling in cell adhesion. Cell 69, 11-25.

Inagaki, F., Shimada, I., Kawaguchi, K., Hirano, M., Terasawa, I., Ikura, T., & Go, N. (1989). Structure of melittin bound to perdeuterated dodecylphosphocholine micelles as studied by two-dimensional NMR and distance geometry calculations. Biochemistry 28, 5985-5991.

Jähnig, F. (1990). Structure predictions of membrane proteins are not that bad. Trends Biochem. Sci. 15, 93-95.

Jan, L. Y., & Jan, Y. N. (1989). Voltage-sensitive ion channels. Cell 56, 13-25.

Jap, B. K. (1989). Molecular design of phoE porin and its functional consequences. J. Mol. Biol. 205, 407-419.

Jap, B. K., & Wallian, P. J. (1990). Biophysics of the structure and function of porins. Quart. Rev. Biophys. 23, 367-403.

Jap, B. K., Wallan, P. J., & Gehring, K. (1991). Structural architecture of an outer membrane channel as determined by electron crystallography. Nature 350, 167-170.

Jap, B. K., Maestre, M. F., Hayward, S. B., & Glaeser, R. M. (1983) Peptide-chain secondary structure of bacteriorhodopsin. Biophys. J. 43, 81-89.

Jennings, M. L. (1989). Topography of membrane proteins. Ann. Rev. Biochem. 58, 999-1027.

Jung, G., Dubischar, N., & Leibfritz, D. (1975). Conformational changes of alamethicin induced by solvent and temperature: A ^{13}C-NMR and circular-dichroism study. Eur. J. Biochem. 54, 395-405

Kaiser, E. T., & Kezdy, F. J. (1987). Peptides with affinity for membranes. Ann. Rev. Biophys. Biophys. Chem. 16, 561-681.

Katre, N. V., Wolber, P. K., Stoecknius, W., & Stroud, R. M. (1981). Attachment sites(s) of retinal in bacteriorhodopsin. Proc. Natl. Acad. Sci. USA 78, 4068- 4072.

Ketchem, R. R., He, W., & Cross, T. A. (1993). High-resolution conformation of gramicidin A in a lipid bilayer by solid-state NMR. Science 261, 1457-1460.

Khorana, H. G. (1988). Bacteriorhodopsin, a membrane protein that uses light to translocate protons. J. Biol. Chem. 263, 7439-7332.

Kleffel, B., Garavito, R. M., Baumeister, W., & Rosenbusch, J. P. (1985). Secondary structure of a channel-forming protein: Porin from E. coli outer membranes. EMBO J. 4, 1589-1592.

Kobilka, B. K., Kobilka, T. S., Daniel, K., Regan, J. W., Caron, M. G., & Lefkowitz, R. J. (1988). Chimeric α_2-β_2-adrenergic receptors: Delineation of domains involved in effector coupling and ligand binding specificity. Science 240, 1310-1316.

Krauss, N., Hinrichs, W., Witt, I., Fromme, P., IPritzkow, W., Dauter, Z., Betzel, C., Wilson, K. S., Witt, H. T., & Saenger, W. (1993). Three-dimensional structure of system of photosynthesis at 6 Å resolution. Nature 361, 326-331.

Kreusch, A., Neubüser, A., Schiltz, E., Weckesser, J., & Schulz, G. E. (1994). Structure of the membrane channel porin from Rhodopseumonas blastica at 2.0 Å resolution. Protein Sci. 3, 58-73.

Kubalek, E., Ralston, S., Lindstrom, J., & Unwin, N. (1987). Location of subunits within the acetylcholine receptor by electron image analysis of tubular crystals from Torpedo marmorata. J. Cell Biol. 105, 9-18.

Kühlbrandt, W. (1984). Three-dimensional structure of the light-harvesting chlorophyll a/b-protein complex. Nature 307, 478-480.

Kühlbrandt, W. (1988). Structure of light-harvesting chlorophyll a/b protein complex from plant photosynthetic membranes at 7 Å resolution in projection. J. Mol. Biol. 202, 849-964.

Kühlbrandt, W. (1994). Structure and function of the plant light-harvesting complex, LHC-II. Curr. Opin. Struct. Biol. 4, 519-528.

Kühlbrandt, W. (1995). Many wheels make light work. Nature 374, 374-375.

Kühlbrandt, W., & Downing, K. H. (1989). Two-dimensional structure of plant light-harvesting complex at 37 Å resolution by electron crystallography. J. Mol. Biol. 207, 823-828.

Kühlbrandt, W., & Wang, D. N. (1991). Three-dimensional structure of plant light-harvesting complex determined by electron crystallography. Nature 350, 130-134.

Kühlbrandt, W., Wang, D. N., & Fujiyoshi, Y. (1994). Atomic model of the light-harvesting complex by electron crystallography. Nature 367, 614-621.

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydrophobic character of a protein. J. Mol. Biol. 157, 105-132.

Landolt-Marticorena, C., Williams, K. A., Deber, C. M., & Reithmeier, R. A. F. (1993). Non-random distribution of amino acids in the transmembrane segments of human type I single span membrane proteins. J. Mol. Biol. 229, 602-608.

Lange, G., Lewis, S. J., Murshodov, G. N., Dodson, G. G., Moody, P. C. E., Turkenburg, J. P., Barclay, A. N., & Brady, R. L. (1994). Crystal structure of an extracellular fragment of the rat CD4 receptor containing domains 3 and 4. Structure 2, 469-481.

Langs, D. A. (1988). Three-dimensional structure at 0.86Å of the uncomplexed form of the transmembrane ion channel peptide gramicidin A. Science 241, 188-191.

Langs, D. A. (1989). Structure of the ion channel peptide antibiotic gramacidin A. Biopolymers 28, 259-266.

Langs, D. A., & Triggle, D. J. (1992). In: The Structure of Biological Membranes (Yeagle, P. A., ed.), pp. 721-780. CRC Press, Boca Raton.

Lauterwein, J., Bosch, C., Brown, L. R., & Wuthrich, K. (1979). Physicochemical studies of the protein-lipid interactions in melittin-containing micelles. Biochim. Biophys. Acta 556, 244-264.

Lee, D. C., Hatward, J. A., Restall, C. J., & Chapmen, D. (1985). Second-derivative infrared spectroscopic studies of the secondary structure of bacteriorhodopsin and Ca^{2+} ATPase, Biochemistry 24. 4364-4373.

Lemmon, M. A., & Engelman, D. M. (1994). Specificity and promiscuity in membrane helix interactions. Quart. Rev. Biophys. 27, 157-218.

Levin, I. W., Lavialle, F., & Mollay, C. (1982). Comparative effects of melittin and its hydrophobic and hydrophilic fragments on bilayer organization by Raman spectroscopy. Biophys. J. 37, 339-349.

Li, J., & Hollingshead, C. (1982). Formation of crystalline arrays of chlorophyll a/b - light-harvesting protein by membrane reconstitution. Biophys. J. 37, 363-370.

Lin, L.-N., & Brandts, J. F. (1983). Determination of cis-trans proline isomerization by trypsin proteolysis. Application to a model pentapeptide and to oxidized ribonuclease A. Biochemistry 22, 553-559.

Lipkind, G. M., & Fozzard, H. A. (1994). A structural model of the tetrodotoxin and saxitoxin binding site of the Na^+ channel. Biophys. J. 66, 1-13.

Macdonald, R. L., & Olsen, R. W. (1994). GABA$_A$ receptor channels. Ann. Rev. Neurosci. 17, 569-602.

Maggio, R., Vogel, Z., & Weiss, J. (1993). Reconstitution of functional muscarinic receptors by co-expression of amino- and carboxy-terminal receptor fragments. FEBS Lett. 319, 195-200.

Maher, P. A., & Singer, A. J. (1985). Anomalous interaction of the acetylcholine receptor with the nonionic detergent Triton X-114. Proc. Natl. Acad. Sci. USA 82, 958-962.

Maloney, P. C. (1994). Bacterial transporters. Curr. Opin Cell Biol. 6, 571-582.

Manella, C. A., Ribeiro, A., & Frank, J. (1986). Structure of the channels in the outer mitochondrial membrane. Electron microscopic studies of the periodic arrays induced by phospholipase A2 treatment of neurospora membrane. Biophys. J. 49, 307-318.

Mao, D., & Wallace, B. A. (1984). Differential light scattering and absorption flattening optical effects are minimal in the circular dichroism spectra of small unilamellar vesicles. Biochemistry 23, 2667-2673.

Martinez, S. E., Huang, D., Szczerpaniak, A., Cramer, W. A., & Smith, J. L. (1994). Crystal structure of chloroplast cytochrome f reveals a novel cytochrome fold and unexpected ligation. Structure 2, 95-105.

McCarthy, M. P., Earnest, J. P., Young, E. F., Choe, S., & Stroud, R. M. (1986). The molecular neurobiology of the nicotinic acetylcholine receptor. Ann. Rev. Neurosci. 9, 383-413.

McCarthy, M. P. (1992). In: The Structure of Biological Membranes (Yeagle, P., ed.), pp. 395-447, CRC Press, Boca Raton.

McDermott, G., Prince, S. M., Freer, A. A., Hawthornthwaite-Lawless, A. M., Papiz, M. Z., Cogdell, R. J., & Isaacs, N. W. (1995). Crystal structure of an integral membrane light-harvesting complex from photosynthetic bacteria. Nature 374, 517-521.

McDonald, N. Q., Murray-Rust, J., & Blundell, T. L. (1995). The first structure of a receptor tyrosine kinase domain: A further step in understanding the molecular basis of insulin action. Structure 3, 1-6.

McDonnell, P. A., Shon, K., Kim, Y., & Opella, S. J. (1993). fd coat protein structure in membrane environments. J. Mol. Biol. 233, 447-463.

McLean, L. R., Buck, S. H., & Krstenansky, J. L. (1990). Examination of the role of the amphipathic α-helix in the interaction of neuropeptide Y and active cyclic analogues with cell membrane receptors and dimyristoylphosphatidylcholine. Biochemistry 29, 2016-2022.

Michel, H. (1983). Crystallization of membrane proteins. Trends Biochem. Sci. 8, 56-59.

Michel-Beyerle, M. E., Plate, M., Deisenhofer, J., Michel, H., Bixon, M., & Jortner, J. (1988). Unidirectionality of charge separation in reaction centers of photosynthetic bacteria. Biochim. Biophys. Acta 932, 52-70.

Mielke, D. L., Kaldany, R.-R., Karlin, A., & Wallace, B. A. (1986). Nicotinic acetylcholine receptor: Secondary structure determined by circular dichroism spectroscopy. Annals N.Y. Acad. Sci. 463, 392-395.

Mielke, D. L., & Wallace, B. A. (1988). Secondary structural analyses of the nicotinic acetylcholine receptor as a test of molecular models. J. Biol. Chem. 263, 3177-3182.

Milon, A., Miyazawa, T., & Higashijima, T. (1990). Transferred nuclear overhauser effect analyses of membrane-bound enkephalin analogues by ^{1}H nuclear magnetic resonance: Correlation between activities and membrane-bound conformations. Biochemistry 29, 65-75.

Mitra, A. K., McCarthy, M. P., & Stroud, R. M. (1989). Three-dimensional structure of nicotinic acetylcholine receptor and location of the major associated 43-kD cytoskeletal protein, determined at 22 Å by low dose electron microscopy and X-ray Diffraction to 12.5 Å. J. Cell Biol. 109, 755-774.

Moriyoshi, K., Masu, M., Ishii, T., Shigemoto, R., Mizuno, N., & Nakanishi, S. (1991). Molecular cloning and characterization of the rat NMDA receptor. Nature 354, 31-37.

Nabedryk, E., Tiede, D. M., Dutton, P. L., & Breton, J. (1982). Conformation and orientation of the protein in the bacterial photosynthetic reaction center. Biochim. Biophys. Acta 682, 273-280.

Nabedryk, E., Biaudet, P., Darr, S., Arntzen, C., & Breton, J. (1984). Conformation and orientation of chlorophyll-proteins in photosystem I by circular dichroism and polarized infrared spectroscopies. Biochim. Biophys. Acta 767, 640-647.

Nabedryk, E., Bardini, A. M., & Breton, J. (1985). Further characterization of protein secondary structures in purple membrane by circular dichroism and polarized infrared spectroscopies. Biophys. J. 48, 873-876.

Nabedryk, E., Garavito, R. M., & Breton, J. (1988). The orientation of β-sheets in porin. A polarized fourier transform infrared spectroscopic investigation. Biophys. J. 53, 671-676.

Naumann, D., Schultz, C., Görne-Tschelnokow, U., & Hucho, F. (1993). Secondary structure and temperature behavior of the acetylcholine receptor by fourier transform infrared spectroscopy. Biochemistry 32, 3162-3168.

Nikaido, H. (1992). Porins and specific channels of bacterial outer membranes. Mol. Microbiol. 6, 435-442.

Noda, M., Takahashi, H., Tanabe, T., Toyosato, M., Kikyotani, S., Furutani, Y., Hirose, T., Takashima, H., Inayami, S., Miyata, T., & Numa, S. (1983). Structural homology of *Torpedo californica* acetylcholine receptor subunits. Nature 302, 528-532.

Okamura, M. Y., & Feher, G. (1992). Proton transfer in reaction centers from photosynthetic bacteria. Ann. Rev. Biochem. 61, 861-896.

Opella, S. J., Kim, Y., & McDonell, P. (1994). Experimental nuclear magnetic resonance studies of membrane proteins. Meth. Enzymol. 239, 536-560.

Panté, N., & Aebi, U. (1994). Towards understanding the three-dimensional structure of the nuclear pore complex at the molecular level. Curr. Opin. Struct. Biol. 4, 187-196.

Papiz, M. Z., Hawthornthwaite, A. M., Cogdell, R. J., Woolley, K. J., Wightman, P. A., Ferguson, L. A., & Lindsay, J. G. (1989). Crystallization and characterization of two crystal forms of the b800-850 light-harvesting complex from the *Rhodopseumonas acidophila* strain 10050. J. Mol. Biol. 209, 833-835.

Parker, M. W., Buckley, J. T., Postma, J. P. M., Tucker, A. D., Leonard, K., Pattus, F. & Tsemoglou, D. (1994). Structure of the *Aeromonas* toxin proaerolysin in its water-soluble and membrane-channel states. Nature 367, 292-295.

Parker, M. W., Pattus, F., Tucker, A. D., & Tsernoglou, D. (1989). Structure of the membrane-pore-forming fragment of colicin A. Nature 337, 93-96.

Pattus, F. (1990). Membrane protein structure. Curr. Opin. Struct. Biol. 2, 681-685.

Picot, D., & Garavito, R. M. (1994). Prostaglandin H synthase: Implications for membrane structure. FEBS Lett. 346, 21-25.

Picot, D., Loll, P. J., & Garavito, R., Michael (1994). The X-ray crystal structure of the membrane protein prostaglandin H_2 synthase-1. Nature 367, 243-249.

Popot, J.-L. (1993). Integral Membrane Protein Structure: Transmembrane α-helices as autonomous folding domains. Curr. Opin. Struct. Biol. 3, 532-540.

Popot, J.-L., Gerchman, S. E., & Engelman, D. M. (1987). Refolding of bacteriorhodopsin in lipid bilayers: A thermodynamically controlled two-stage process. J. Molec. Biol. 210, 829-847.

Popot, J.-L., & Engelman, D. M. (1990). Membrane protein folding and oligomerization: The two-stage model. Biochemistry 29, 4031-4037.

Popot, J.-L., & de Vitry, C. (1990). On the microassembly of integral membrane proteins. Ann. Rev. Biophys. Biophys. Chem. 19, 369-403.

Prosser, R. S., Daleman, S. I., & Davis, J. H. (1994). The structure of an integral membrane peptide: A deuterium NMR study of gramacidin. Biophys. J. 66, 1415-1428.

Pryde, J. G. (1986). Triton X-114: A detergent that has come in from the cold. Trends Biochem. Sci. 11, 160-163.

Radermacher, M., Wagenknecht, T., Grassucci, R., Frank, J., Inui, M., Chadwick, C., & Fleischer, S. (1992). Cryo-EM of the native structure of the calcium release channel/ryanodine receptor from sarcoplasmic reticulum. Biophys. J. 61, 936-940.

Raftery, M. A., Hunkapiller, M. W., Strader, C. D., & Hood, L. (1980). Acetylcholine receptor: Complex of homologous subunits. Science 208, 1454-1457.

Ratnam, M., Nguyen, D. L., Riveir, J., Sargent, P. B., & Lindstrom, J. (1986). Transmembrane topography of nicotinic acetylcholine receptor: Immunochemical tests contradict theoretical predictions based on hydrophobicity profiles. Biochemistry 25, 2633-2643.

Ratnam, M., Sargent, P. B., Sarin, V., Fox, J. L., Nguyen, D. L., Rivier, J., Criado, M. & Lindstrom, J. (1986). Location of antigenic determinants on primary sequences of subunits of nicotinic acetylcholine receptor by peptide mapping. Biochemistry 25, 2621-2632.

Ravichandran, K. G., Boddupalli, S. S., Hasemann, C. A., Peterson, J. A., & Deisenhofer, J. (1993). Crystal structure of hemoprotein domain of P450MB-3, a prototype for microsomal P450s. Science 261, 731-736.

Reber, B. F. X., & Catterall, W. A. (1987). Hydrophobic properties of the β1 and β2 subunits of the rat brain sodium channel. J. Biol. Chem. 262, 11369-11374.

Rees, D. C., Komiya, H., Yeates, T. O., Allen, J. P., & Feher, G. (1989). The bacterial photosynthetic reaction center as a model for membrane proteins. Ann. Rev. Biochem. 58, 607-633.

Reithmeier, R. A. F. (1993). The erythrocyte anion transporter (band 3). Curr. Opin Struct. Biol. 3, 515-523.

Reithmeier, R. A. F. (1994). Mammalian exchangers and co-transporters. Curr. Opin Cell Biol. 6, 583-594.

Reithmeier, R. A. F., & Deber, C. M. (1992). In: The Structure of Biological Membranes (Yeagle, P., ed.), pp. 337-393, CRC Press, Boca Raton.

Ribi, H. O., Ludwig, D. S., Mercer, K. L., Schoolnik, G. K., & Kornberg, R. D. (1988). Three-dimensional structure of cholera toxin penetrating a lipid membrane. Science 239, 1272-1276.

Rizzo, V., Stankowski, S., & Schwarz, G. (1987). Alamethicin incorporation in lipid bilayers: A thermodynamic study. Biochemistry 26, 2751-2759.

Rosenbusch, J. P. (1974). Characterization of the major envelope protein from *Escherichia coli*. Regular arrangement on the peptidoglycan and unusual dodecyl sulfate binding. J. Biol. Chem. 249, 8019-8029.

Roth, M., Lewit-Bentley, A., Michel, H., Deisenhofer, J., Huber, R., & Oesterhelt, D. (1989). Detergent structure in crystals of a bacterial photosynthetic reaction centre. Nature 340, 659-662.

Schabert, F. A., & Engel, A. (1994). Reproducible acquisition of *Escherichia coli* porin surface topographs by atomic force microscopy. Biophys. J. 67, 2394-2403.

Schertler, G. F. X., Villa, C., & Henderson, R. (1993). Projection structure of rhodopsin. Nature 362, 770-772.

Schindler, H., & Rosenbusch, J. P. (1978). Matrix protein from *Escherichia coli* outer membranes forms voltage-controlled channels in lipid bilayers. Proc. Natl. Acad. Sci. USA 75, 3751-3755.

Schindler, M., & Rosenbusch, J. P. (1984). Structural transitions of porin, a transmembrane protein. FEBS Lett. 173, 85-89.

Schirmer, T., Keller, T. A., Wang, Y.-F., & Rosenbusch, J. P. (1995). Structural basis for sugar translocation through maltoporin channels at 3.1 Å resolution. Science 267, 512-514.

Schlessinger, J., & Ullrich, A. (1992). Growth factor signaling by receptor tyrosine kinases. Neuron 9, 383-391.

Scott, W. G., Milligan, D. L., Milburn, M. V., Privé, G. G., Yeh, J., Koshland, D. E., Jr., & Kim, S.-H. (1993). Refined structures of the ligand-binding domain of the aspartate receptor from *Salmonella typhimurium*. J. Mol. Biol. 232, 555-573.

Shon, K., Kim, Y., Colnage, L. A., & Opella, S. J. (1991). NMR studies of the structure and dynamics of membrane-bound bacteriophage Pf1 Coat protein. Science 252, 1303-1305.

Singer, S. J. (1971). In: Structure and Function of Biological Membranes (Rothfield, L. I., ed.), pp. 145-222. Academic Press, New York.

Singer, S. J. (1990). The structure and insertion of integral proteins in membranes. Ann. Rev. Cell Biol. 6, 247-296.

Singer, S. J., & Nicholson, G. L. (1972). The fluid mosaic model of the structure of cell membranes. Science 175, 720-731.

Sipos, L., & Von Heijne, G. (1993). Predicting the topology of eukaryotic membrane proteins. Eur. J. Biochem. 213, 1333-1340.

Smith, W. L., & Marnett, L. J. (1994). In: Metal Ions in Biological Systems (Sigal, H. & Sigel, A., eds.), pp. 163-199. Marcel Dekker, Inc., New York.

Stokes, D. L., & Nakamoto, R. K. (1994). Structure of P-type and F-type ion pumps. Curr. Opin. Struct. Biol. 4, 197-203.

Stokes, D. L., Taylor, W. R., & Green, N. M. (1994). Structure, transmembrane topology and helix packing of P-type ion pumps. FEBS Lett. 346, 32-38.

Stroud, R. M., McCarthy, M. P., & Shuster, M. (1990). Nicotinic acetylcholine receptor superfamily of ligand-gated ion channels. Biochemistry 29, 11009-11023.

Takeshima, H., Nishimura, S., Matsumoto, T., Ishida, H., Kangawa, K., Minamino, N., Hisayuki, M., Ueda, M., Hanaoka, M., Hirose, T., & Numa, S. (1989). Primary structure and expression from complementary DNA of skeletal muscle ryanodine receptor. Nature 339, 439-445.

Terwilliger, T. C., & Eisenberg, D. (1982). The structure of melittin. 1. Structure determination and partial refinement. J. Biol. Chem. 257, 6010-6015.

Terwilliger, T. C., Weissman, L., & Eisenberg, D. (1982). The structure of melittin in the form I crystals and its implication for melittin's lytic and surface activities. Biophys. J. 37, 353-361.

Thomas, L., Kocsis, E., Colombini, M., Erbe, E., Trus, B. L., & Steven, A. C. (1991). Surface topography and molecular stoichiometry of the mitochondrial channel, VDAC, in crystalline arrays. J. Struct. Biol. 106, 161-171.

Toyoshima, C., & Unwin, N. (1988). Ion channel of acetylcholine receptor reconstructed from images of postsynaptic membranes. Nature 336, 247-250.

Tsygannik, I. N., & Baldwin, J. M. (1987). Three-dimensional Structure of deoxycholate-treated purple membrane at 6 Å resolution and molecular averaging of three crystal forms of bacteriorhodopsin. Eur. Biophys. J. 14, 263-272.

Unwin, N. (1993). Nicotinic acetylcholine receptor at 9 Å resolution. J. Mol. Biol. 229, 1101-1124.

Unwin, N. (1995). Acetylcholine receptor channel imaged in the open state. Nature 373, 37-43.

Unwin, N., & Henderson, R. (1984). The structure of proteins in biological membranes. Sci. Amer. 250, 78-94.

Vogel, H. (1981). Incorporation of melittin into phosphatidylcholine bilayers: Study of binding and conformational changes. FEBS Letts. 134, 37-42.

Vogel, H. (1987). Comparison of the conformations and orientation of alamethicin and melittin in lipid membranes. Biochemistry 26, 4562-4572.

Vogel, H. (1992). Structure and dynamics of polypeptides and proteins in lipid membranes. Quart. Rev. Biophys. 25, 433-457.

Vogel, H., & Gartner, W. (1987). The secondary structure of bacteriorhodopsin determined by Raman and circular dichroism spectroscopy. J. Biol. Chem. 262, 11464-11469.

Vogel, H., & Jahnig, F. (1986). Models for the structure of outer-membrane proteins of *Escherichia coli* derived from Raman spectroscopy and prediction methods. J. Mol. Biol. 190, 191-199.

von Heijne, G. (1989). Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. Nature 341, 456-458.

von Heijne, G. (1992). Membrane protein structure prediction: Hydrophobicity analysis and the positive-inside rule. J. Mol. Biol. 225, 487-494.

von Heijne, G., & Gavel, Y. (1988). Topogenic signals in integral membrane proteins. Eur. J. Biochem. 174, 671-678.

Wagenknecht, T., Grassucci, R., Frank, J., Saito, A., Inui, M., & Fleischer, S. (1989). Three-dimensional architecture of the calcium channel/foot structure of sarcoplasmic reticulum. Nature 338, 167-179.

Wallace, B. A. (1986). Structure of gramicidin A. Biophys. J. 49, 295-306.

Wallace, B. A., & Ravikumar, K. (1988). The gramicidin pore: Crystal structure of a cesium complex. Science 241, 182-187.

Wallace, B. A., & Teeters, C. L. (1987). Differential absorption flattening optical effects are significant in the circular dichroism spectra of large membrane fragments. Biochemistry 26, 65-70.

Walz, T., Smith, B. L., Zeidel, M. L., Engel, A., & Agre, P. (1994). Biologically active two-dimensional crystals of aquaporin CHIP. J. Biol. Chem. 269, 1583-1586.

Wang, D. N. (1994). Band 3 protein: Structure, flexibility and function. FEBS Lett. 346, 26-31.

Wang, D. N., Kühlbrandt, W., Sarabia, V. E., & Reithmeier, R. A. F. (1993). Two-dimensional structure of the membrane domain of human band 3, the anion transport of the erythrocyte membrane. EMBO J. 12, 2233-2239.

Wang, D. N., Sarabia, V. E., Reithmeier, R. A. F., & Kühlbrandt, W. (1994). Three-dimensional map of the dimeric membrane domain of the human erythrocyte anion exchanger, band 3. EMBO J. 13, 3230-3235.

Ward, R. J., & Leonard, K. (1992). The *Staphylococcus aureus* α-toxin channel complex and the effect of Ca^{2+} ions on its interaction with lipid bilayers. J. Struct. Biol. 109, 129-141.

Weis, W. I., Cusack, S. C., Brown, J. H., Daniels, R. S., Skehel, J. J., & Wiley, D. C. (1990). The structure of a membrane fusion of the influenza virus haemagglutinin. EMBO J. 9, 17-24.

Weiss, M. S., Kreusch, A., Schiltz, E., Nestel, U., Welte, W., Weckesser, J., & Schulz, G. E. (1991). The structure of porin from *Rhodobacter capsulatus* at 1.8 Å resolution. FEBS Lett. 280, 379-382.

Weiss, M. S., & Schulz, G. E. (1992). Structure of porin refined at 1.8 Å resolution. J. Mol. Biol. 227, 493-509.

White, S. H. (1994). In: Membrane Protein Structure: Experimental Approaches. New York, Oxford University Press.

White, S. H., & Wimley, W. C. (1994). Peptides in lipid bilayers: Structural and thermodynamic basis for partitioning and folding. Curr. Opin. Struct. Biol. 4, 79-86.

Williams, R. W., & Deber, C. M. (1991). Proline in transmembrane helices: Structural dynamic role? Biochemistry 30, 8919-8923.

Wilmsen, H. U., Leonard, K. R., Tichelaar, W., Buckley, J. T., & Pattus, F. (1992). The aerolysin membrane channel is formed by heptamerization of the monomer. EMBO J. 11, 2457-2463.

Wu, C.-S., Sun, X. H., & Yang, J. T. (1990). Conformation of acetylcholine receptor in the presence of agonists and antagonists. J. Prot. Chem. 9, 119-126.

Yager, P., Chang, E. L., Williams, R. W., & Dalziel, A. W. (1984). The secondary structure of acetylcholine receptor reconsitituted in a single lipid component as determined by Raman spectroscopy. Biophys. J. 45, 26-28.

Yang, J., Yamm, L. K., Tillack, T. W., & Shao, Z. (1993). New approach for atomic force microscopy of membrane proteins. J. Mol. Biol. 229, 286-290.

Yeates, T. O., Komiya, H., Rees, D. C., Allen, J. P., & Feher, G. (1987). Structure of the reaction center from *Rhodobacter sphaeroides* R-26: Membrane-protein interactions. Proc. Natl. Acad. Sci. USA 84, 6438-6442.

Young, E. F., Ralston, E., Blake, J., Ramachandran, J., Hall, Z. W., & Stroud, R. M. (1985). Topological mapping of acetylcholine receptor: Evidence for a model with five transmembrane segments and a cytoplasmic COOH-terminal peptide. Proc. Natl. Acad. Sci. USA 82, 626-630.

Zetta, L., & Kaptein, R. (1984). Interaction of β-endorphin with sodium dodecyl sulfate in aqueous solution: ^{1}H-NMR investigation. Eur. J. Biochem. 145, 181-186.

STRUCTURAL FEATURES OF MEMBRANE PROTEINS

Tuomas Haltia

I. INTRODUCTION

Proteins associate with membranes in several ways: some bind electrostatically to the membrane surface, whereas others contain a covalently bound lipid that anchors an otherwise soluble protein hydrophobically to the membrane. For the purpose of this chapter, however, a membrane protein will be defined as a protein that penetrates into and, in most cases, traverses the hydrophobic core of the lipid bilayer. Owing to the presence of an exposed nonpolar surface that interacts with the membrane, integral membrane proteins are insoluble under aqueous conditions and they are normally "solubilized" by disrupting the biological membrane with a detergent. Thus the object of study in membrane protein research is actually a mixed micelle of protein, detergent and (frequently) some lipids. The choice of the right detergent is critical for the stability, purification and activity of a membrane protein. Apart from the inclusion of the detergent in all solutions, membrane proteins are typically purified using the same methods that are used for soluble proteins, although obtaining highly pure homogeneous preparations is often more difficult. In particular, the presence of a solubilizing detergent has complicated the crystallization of membrane proteins: only 13 membrane protein structures are currently known at an atomic resolution. However, recent development has been encouraging: in the years 1994-1995, seven new structures[1] have been reported, among them that of cytochrome c oxidase from bovine heart, the largest membrane protein structure (M_r = 205 kDa) solved so far (Tsukihara et al., 1995). At the same time another group published the structure of a homologous bacterial oxidase (Iwata et al., 1995). The two oxidase structures more than double the atomic resolution data on membrane proteins if estimated on the basis of the number of transmembrane helices in the database. In addition, the structures of two other respiratory complexes (cytochrome bc_1 complex and succinate dehydrogenase) can be expected in the near future (see, e.g., Lee et al., 1995). The high resolution structure of photosystem I is currently being solved (Krauss et al., 1993).

As the number of membrane protein structures in the database has increased, some of the rules that govern the folding and stability of this class of proteins have begun to emerge. This is important as the various genome sequencing projects continue to yield numerous sequences which are likely to encode previously unknown membrane proteins. For instance, the proportion of membrane protein sequences in the completely sequenced yeast chromosome VIII has been estimated to be as high as 20-30% (Rost et al., 1995). Thus knowledge-based structure

prediction methods are in a high demand. Moreover, membrane proteins are implicated in many diseases. To cite a few examples, cystic fibrosis, Alzheimer's, Batten's and Leber's diseases (Ashcroft and Röper, 1993; Hästbacka et al., 1994; Barinaga, 1995; Lerner et al., 1995; Savontaus, 1995) appear to involve mutations in genes encoding membrane proteins. In the newly discovered genes linked to the early onset form of Alzheimer's disease, several mutations appear in the putative transmembrane segments or close to their ends (Barinaga, 1995). The mechanism by which these affect the pathogenesis of the disease is not known.

The intrabilayer regions of membrane proteins fold and adopt their secondary, tertiary, and quaternary structures in an environment from which water is largely absent. Since the properties of water play a crucial role in determining the forces (electrostatic interactions, hydrogen bonding, and the hydrophobic effect) that govern the folding and stability of soluble proteins, the physical basis of membrane protein stability as compared with that of soluble proteins is of fundamental interest. How is a bundle of transmembrane helices stabilized? For example, with membrane proteins the role of the hydrophobic effect is to cause the partitioning of the nonpolar part of the protein into the lipid phase. It follows that inside the bilayer there is no hydrophobic effect that could drive the folding process in a way similar to that of soluble proteins. In the absence of competing water molecules the hydrogen-bonding potential of the peptide groups in a membrane-bound protein is satisfied intramolecularly by forming highly stable secondary structures (in most cases transmembrane α-helices). In this way the exposure of the polar peptide group to the lipid becomes minimized while the hydrophobic side-chains interact favorably with the hydrocarbons of the bilayer. In the final folding step, the helices interact with each other in a highly specific manner (in most cases, by packing optimally against each other) making up the tertiary fold of the membrane protein (Popot and Engelman, 1990; Lemmon and Engelman, 1994).

In this chapter, the characteristic features of membrane proteins and some aspects of their biochemistry are reviewed. A brief account of problems associated with crystallization of membrane proteins is given. As an example of a complex membrane protein structure, the newly determined structure of cytochrome *c* oxidase is dealt with in detail. A number of other reviews on membrane proteins have appeared recently (Jennings, 1989; Rees et al., 1989b, Michel, 1990; Popot and Engelman, 1990; Babcock and Wikström, 1992; Popot, 1993; Traxler et al., 1993; Lemmon and Engelman, 1994; Trumpower and Gennis, 1994; White, 1994; Haltia and Freire, 1995; Popot and Saraste, 1995).

II.   THE INTRAMEMBRANE SECONDARY STRUCTURE ELEMENTS

A.   The Transmembrane α-Helix

In an α-helical conformation, 20-25 nonpolar residues are needed to traverse the hydrophobic core of a biological membrane. Such stretches of non-polar amino

acids are routinely identified by hydropathy analysis (Kyte and Doolittle, 1982; Engelman et al., 1986; Michel et al., 1986): each residue is assigned a hydrophobicity value and the sequence is scanned using an algorithm that searches for regions whose hydrophobicity score is above a certain threshold. Typically, a sliding window average of 11 to 19 residues is used. Several hydrophobicity scales exist (Engelman et al., 1986; von Heijne, 1992). Comparison with known membrane protein structures (see below; Figure 4) shows that hydropathy analysis alone is able to identify the locations of transmembrane segments rather successfully (but is less reliable regarding the length of a transmembrane helix). However, false positives and negatives do occur (Rost et al., 1995). False positives are typically encountered when the protein in question has a large soluble domain with non-polar β-strands in its hydrophobic core. False negatives occur with multispanning membrane proteins because some rather amphipathic helices may be tolerated provided that they are inside a transmembrane helical bundle and are not exposed to the lipid. To overcome these problems, more sophisticated prediction methods have recently been developed (von Heijne, 1992; Jones et al., 1994b; Persson and Argos, 1994; Taylor et al., 1994; Rost et al., 1995). As additional constraints these algorithms have employed the positive-inside rule (see below; von Heijne, 1992) and analysis of conservation patterns, as well as sites for deletions and insertions among homologous sequences. The latter requires the generation of a multiple sequence alignment which is then used as the input for the prediction. The sequence alignment can be compiled by the reseacher or it can also be generated as part of the automated prediction protocol. Persson and Argos (1994) and Jones et al. (1994) divide the transmembrane helix into three zones (inside helix, middle helix, and outside helix) each with a preferential amino acid profile deduced from statistical analyses. Composition of the sequences flanking the transmembrane helix has been studied in order to make use of the positive-inside rule in predicting not only secondary structure but also topology (orientation) of the helices (von Heijne, 1992; Jones et al., 1994). Furthermore, attempts have been made to predict the relative positions of the included helices in bundles of helices (Baldwin, 1993; Donnelly et al., 1993; Taylor et al., 1994). As discussed by Rost et al. (1995) and Taylor et al. (1994), a prerequisite for accurate prediction of helix positions is the correct identification of the ends of the transmembrane helices. In section IV. A (see Table II) the predictive power of two such methods (Persson and Argos, 1994; Rost et al., 1995) is tested against the recently published high-resolution structure of a bacterial cytochrome c oxidase.

In addition to the 13 high resolution structures mentioned above, low or medium resolution structures of a number of membrane proteins have been determined (e.g., Cyrklaff et al., 1995; Haltia and Freire, 1995; Hebert et al., 1995; Unwin, 1995). For a much larger group of proteins, the topology of the polypeptide chain has been determined, typically using gene fusion experiments (Traxler et al., 1993) and a variety of biochemical methods (Jennings, 1989). Because of limited structural data, the use of the above prediction methods has attracted much attention in

studying the characteristics of (putative) membrane spanning segments. The main findings of these analyses are summarized below. In addition to the general non-polarity, transmembrane helices appear to have some specific features: (i) aromatic residues (Trp, Tyr, Phe) are common at the membrane-aqueous boundaries (Landolt-Marticorena et al., 1993; Jones et al., 1994); tryptophan residues appear especially at the outer boundary. This phenomenon is also observed in the 3D structures of bacteriorhodopsin, the bacterial reaction centers and porins (see Figure 8) and may represent a general structural element involved in protein-lipid-headgroup interaction (Cowan and Rosenbusch, 1994; Haltia and Freire, 1995). A number of interfacial tryptophan residues are present in the monotopic membrane anchor of prostaglandin synthase (Figure 1; Picot et al., 1994). In the structure of cytochrome *c* oxidase, numerous aromatic residues flank the aliphatic middle zone of the transmembrane helices (Iwata et al., personal communication; see Figure 7). Interestingly, model compound studies suggest that the indole ring partitions specifically to the headgroup region of the lipid bilayer which is clearly more polar ($\varepsilon = 10$-30) than the core of the bilayer ($\varepsilon = 2$) (Wimley and White, 1993). The zone rich in aromatic residues probably makes the transition of the polypeptide from a nonpolar, low dielectric environment to a solution with a very high dielectric constant ($\varepsilon = 80$) solution energetically feasible. (ii) Glycine and proline are common in transmembrane segments, particularly in those of multispanning proteins (Landolt-Marticorena et al., 1993; Jones et al., 1994a; cf. Figure 5). (iii) The transmembrane segments of multispanning proteins are less hydrophobic than those of single span proteins (Jones et al., 1994a). (iv) Residues that face the lipids are typically aliphatic and are less conserved than the residues that face the protein interior (Rees et al., 1989a,b; Donnelly et al., 1993; Jones et al., 1994a; Taylor et al., 1994; Samatey et al., 1995). Samatey et al. (1995) suggest that aromatic residues are enriched in the buried surfaces of the helices. The interiors of membrane and soluble proteins are of similar hydrophobicity (Rees et al., 1989a,b). (This conclusion is based on the analysis of the reaction center structure, and it remains to be tested using the new cytochrome oxidase structures.) (v) Charged residues, although rare, do occur in transmembrane segments. Functionally important charged residues are found in the helices of bacteriorhodopsin (Krebs and Khorana, 1993). Glutamate-arginine inter- and intrahelical ion-pairs are observed in the plant light-harvesting complex (Kuhlbrandt et al., 1994; Figure 9). Five potentially charged residues occur within the membrane spanning region of a bacterial cytochrome oxidase (Iwata et al., 1995). Intramembrane ion-pairs may also be present in lactose permease and the T-cell receptor as well as in voltage-gated ion channels (Haltia and Freire, 1995 and references cited therein). (vi) Lys and Arg residues are preferentially localized to the presumed loops which reside in the inner of the membrane, a phenomenon termed the positive inside rule (von Heijne, 1992; Landolt-Marticorena et al., 1993); the rule appears to be valid with extramembrane segments which are shorter than about 60 residues. The positive inside rule is probably the most important determinant of transmembrane helix orientation.
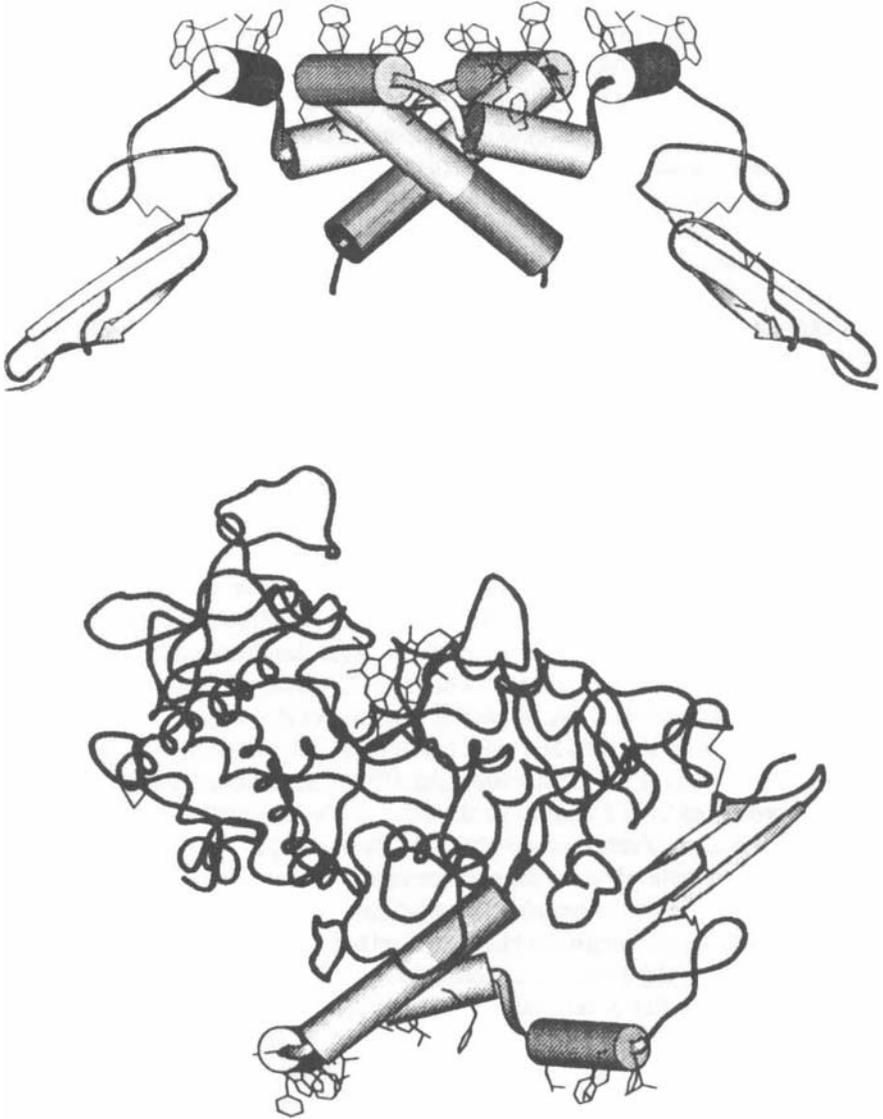
Figure 1. Prostaglandin H_2 synthase, located in the membrane of the endoplasmic reticulum, is a bifunctional enzyme that converts the fatty acid arachidonate into prostaglandin H_2. The enzyme is a membrane-bound homodimer. Each monomer has two active sites (the cyclooxygenase and peroxidase sites) located in the extramembrane domain of the protein. However, its substrate is present in the membrane. Accordingly,

While an isolated helix under aqueous conditions is only marginally stable, a transmembrane helix is a highly stable and rigid structure (Lemmon and Engelman, 1994; Haltia and Freire, 1995). Two factors bring about the high stability: the strength of the helical hydrogen bond inside the bilayer and the hydrophobic effect. The strength of intrabilayer hydrogen bonds (estimated to be 6 kcal/mol, a value much larger than under aqueous conditions; Engelman et al., 1986) is caused by the absence of water, which otherwise would compete for the formation of hydrogen bonds with the peptide carbonyls and amides, disfavoring the formation of the intramolecular, helix-stabilizing hydrogen bonds. The hydrophobic effect, on the other hand, drives the partition of the hydrophobic peptide into the bilayer where the interaction of the nonpolar side-chains with the lipid also stabilizes the helix. The free energy of stabilization of a 23-residue hydrophobic intrabilayer helix has been estimated to be about 70 kcal/mol (Lemmon and Engelman, 1994). For comparison, the free energies of stabilization of the folded conformations of many soluble proteins range from 5-15 kcal/mol. The extremely high stability of intrabilayer α-helices is manifested in the very low denaturation enthalpies of membrane proteins in calorimetric experiments (Haltia et al., 1994; Haltia and Freire, 1995), which suggest that transmembrane helices are so stable that they do not unfold thermally (see also Arrondo et al., 1994). In the case of the band 3 protein (the red cell anion transporter), which is a helical membrane protein, the calorimetric enthalpy appears to reflect mainly the strength of interhelix interactions (and not the unfolding of the transmembrane helices (Moriyama et al., 1992; Haltia and Freire, 1995)). Likewise, transmembrane helices in the band 3 protein and other proteins are often highly resistant toward chemical denaturants such as guadinine hydrochloride (Rizzolo and Tanford, 1978; Oikawa et al., 1985; Hill et al., 1988). Solvent denaturation of transmembrane helices tends to be gradual as a function of denaturant concentration, whereas soluble proteins usually display sharp cooperative transitions over a narrow concentration range of the perturbant (e.g. Rizzolo & Tanford, 1978; Oikawa et al., 1985).

Figure 1. (continued) the enzyme has a hydrophobic channel that leads from the membrane to the cyclooxygenase active site (for details, see Loll et al., 1995). In the upper part of the figure, the structure of a prostaglandin synthase monomer is shown. The four helices which mediate the binding of the otherwise soluble-like protein to the membrane are shown as cylinders. They run nearly parallel to the membrane plane, surrounding the mouth of the substrate channel and anchoring the enzyme to the membrane without traversing the bilayer. Because of this type of membrane-protein interaction, prostaglandin synthase has been termed a monotopic membrane protein (Picot et al., 1994). In the lower part of the figure, the membrane binding motif in the prostaglandin synthase dimer is shown in more detail. Note the many aromatic residues found in the membrane exposed faces of the short helices which mediate the binding of the dimer to the membrane. Figures were taken from Garavito et al. (1995) Protein Engineering 8 (Supplement), 41 by Permission of Oxford University Press.

The realization that an isolated transmembrane helix is stable as such has led to the concept of a transmembrane helix as an autonomous folding domain in membrane proteins (Popot and Engelman, 1990; Popot, 1993). (As discusssed by Lemmon and Engelman (1994), this is also an assumption inherent in the various transmembrane helix prediction methods.) In essence, the folding of a membrane protein is suggested to take place in two stages: first, the transmembrane helices form, and, in the second stage, they associate to make up the tertiary fold. What are the forces that bring the helices together? It appears that in many cases the critical interhelical interaction occurs inside the lipid phase rather than in the extramembrane domains. For instance, most loops in bacteriorhodopsin can be either cleaved, shortened, or made longer without interfering with the assembly of functional protein (Popot and Engelman, 1990; Gilles-Gonzalez et al., 1991; Teufel et al., 1993; Hansen et al., 1994; Haltia and Freire, 1995). Moreover, although there are examples of interhelical ion-pairs and hydrogen bonds, the interaction between helices can be very strong and specific in the absence of strongly polar residues. Mutagenesis studies with three different model systems (Arkin et al., 1994; Lemmon et al., 1994; Williams et al., 1995) suggest that close and specific stereochemical packing performs a crucial role in tight helix-helix association. Lemmon et al. (1994) have shown that dimerization of glycophorin A depends on a unique seven residue motif LIxxxGVxxGVxxT present in its single transmembrane domain. When introduced into a polyleucine helix, the seven residue motif brings about formation of sodium dodecyl sulfate-stable dimers. Even a conservative mutation of a residue in the motif is enough to cause disruption of the dimer. Modeling based on simulated annealing suggests that the critical residues are located on the interhelical surface in a right-handed supercoil. The role of the two glycines in the motif may be to allow a close approach of the backbones, thereby bringing about a favorable electrostatic interaction between the helices. The aliphatic residues are probably involved in optimizing the inter-helical packing interactions while the threonine may participate in interhelical hydrogen bonding (Lemmon et al., 1994).[2]

Another case in which glycines are important for helix-helix interaction is the oligomerization of the major histocombatibility complex subunits: the association of the single transmembrane segments of the α- and β-subunits of the complex appears to be mediated by glycine-rich sequences (Cosson and Bonifacino, 1992). In a helical wheel projection the glycines fall on the same face. More generally, the relatively high incidence of prolines and glycines in putative transmembrane helices (Jones et al., 1994a) may be related to the need for a slight perturbation of ideal helical geometry in order to promote interhelical association (or binding of cofactors between helices) within the membrane. The observation that the transmembrane segments of membrane proteins with a single transmembrane span are clearly more hydrophobic than those in multispanning proteins may have the same explanation. Conceivably, the extremely high stability of the transmembrane helix makes it energetically possible to tolerate some intramembrane polar residues and allows

the exposure of the polar backbone when necessary for the structure and/or function of the protein (Lemmon and Engelman, 1994).

B. Transmembrane β-Strands

Owing to the extended conformation of a β-strand, only 7-9 residues are needed to traverse the core of the bilayer as a β-strand. However, the hydrogen-bonding potential of the peptide bond cannot be satisfied locally as in a transmembrane α-helix. Instead, as shown by the high-resolution porin structures, it can be satisfied globally by a barrel of 16 or 18 antiparallel β-strands interconnected by hydrogen bonds (Cowan et al., 1992; Weiss and Schulz, 1992; Kreusch and Schulz, 1995; Schirmer et al., 1995). Intramembrane β-strands might be present in the acetyl-choline receptor (Unwin, 1993, 1995; Görne-Tschelnokow et al., 1994) and in a microsomal glutathione transferase (Hebert et al., 1995), both of which also contain transmembrane α-helices and may therefore be examples of more complex membrane protein folding patterns. On the basis of Fourier transform infrared spectroscopy (FTIR) spectra, a water channel (reviewed by Agre et al., 1995) consists of about equal amounts of α-helix and β-sheet (Haris et al., 1995; Van Hoek et al., 1993). However, in most models the protein is helical.

If α/β transmembrane elements really exist, recognition of structural motifs of this type from the primary structure remains a challenge (Popot, 1993; Schirmer and Cowan, 1993; Sun and Parthasarathy, 1994). Prediction of porin-type transmembrane β-strands from sequence is not straightforward either, as a transmembrane β-barrel can be produced by non-homologous sequences that do not differ much from those of soluble proteins. In transmembrane β-strands of porins only every second residue is consistently hydrophobic while the rest can be either hydrophilic or hydrophobic depending on whether they are exposed to the aqueous pore or to the protein interior. Based on the analysis of the 16-strand porins, Schirmer and Cowan (1993) developed a prediction method which correctly identified 16 out of the 18 strands of maltoporin (Schirmer et al., 1995). In addition to searching for sequences where every second residue is hydrophobic, their method made use of the observation that in porins aromatic residues cluster at the membrane-aqueous interface (Figure 8).

C. Monotopic Membrane Anchors

Neither transmembrane α-helices nor β-strands are seen in the three-dimensional (3D) structure of the integral membrane protein prostaglandin H$_2$ synthase. Instead, it contains four short amphipathic helices which run parallel to the membrane plane, exposing their hydrophobic face to the hydrophobic core of the bilayer (Picot et al., 1994; Picot and Garavito, 1994; see Figure 1). Most probably, prostaglandin synthase only penetrates into one leaflet of the bilayer and is therefore classified a monotopic membrane protein. Another example of this type of protein

may be mandelate dehydrogenase (Mitra et al., 1993). How common monotopic membrane proteins are is not clear; at the moment sequence-based prediction methods that would identify this mode of membrane-protein interaction do not exist. In fact, until the 3D structure of prostaglandin synthase became known, the enzyme was predicted to have one or two transmembrane helices (Picot and Garavito, 1994).

III. BUILDING A MODEL AND SOLVING THE STRUCTURE

A. Biochemistry, Spectroscopy, and Protein Engineering

The traditional way to construct a working model of a membrane protein involves, in addition to sequence-based modeling, biochemical experiments aiming at determining the membrane topography of the protein under investigation. In some cases proteolysis experiments and the use of lipophilic and hydrophilic labeling reagents, as well as antibodies (reviewed in Jennings, 1989), have resulted in fairly consistent two-dimensional (2D) folding models, although unambiguous experimental demonstration of the predicted topology of a polytopic membrane protein is a considerable task and results are sometimes difficult to interpret. For example, hydropathy analysis suggests that the band 3 protein has 14 transmembrane helices but the topology of the last six helices has been difficult to verify biochemically (Reithmeier, 1993). For some proteins, such as the nicotinic acetylcholine receptor very different models have been proposed (Fischbarg et al., 1993; Unwin, 1993; Görne-Tschelnokow et al., 1994; Akabas & Karlin, 1995; cf. Agre et al., 1995; Jap and Li, 1995; Mitra et al., 1995; and Walz et al., 1995 which discuss the emerging structure of a water channel).

Spectroscopic methods such as FTIR have been used to estimate the relative amounts of different types of secondary structure (Arrondo et al., 1993, 1994; Surewicz et al., 1993; Haltia et al., 1994; Echabe et al., 1995). Because typical transmembrane helices are much more stable than soluble helices (Lemmon and Engelman, 1994; Haltia and Freire, 1995), recording the FTIR spectra as a function of temperature has been used to assess the relative proportions of the intra- and extramembrane secondary structures (Haltia et al., 1994; Echabe et al., 1995; cf. Naumann et al., 1993). However, some transmembrane helices such as those in cytochrome oxidase subunit III (see below) are unstable, whereas many transmembrane helices have hydrophilic ends that may behave like soluble helices in thermal denaturation experiments, complicating an analysis based on differential stabilities (cf. Echabe et al., 1995 and Iwata et al., 1995). Because of the high stability of hydrogen bonds in the low-water environment, intrabilayer secondary structures are very resistant to hydrogen/deuterium (H/D) exchange, which can be quantified by FTIR spectroscopy (Earnest et al., 1990; Naumann et al., 1993; Görne-Tschelnokow et al., 1994; but see also Haris et al. (1995) who found rapid H/D exchange

in a water channel protein). Combination of proteolysis and FTIR spectroscopy has been used to demonstrate the likely presence of intrabilayer β-strands in the acetylcholine receptor (Görne-Tschelnokow et al., 1994). With a bacterial cytochrome oxidase α-helicity determined by infrared spectroscopy (Echabe et al., 1995) agrees quite well with the X-ray structure (Iwata et al., 1995). Circular dichroism (CD) spectroscopy has also been used to evaluate secondary structures of membrane proteins (Park et al., 1992); the latter work suggested that a transmembrane α-helix gives rise to a characteristic CD spectrum which is different from that of a soluble helix. In general, FTIR spectroscopy may be the most reliable spectroscopic method for quantitative analysis of secondary structures in membrane proteins. The accuracy of FTIR spectroscopy depends critically, however, on the method used to analyze the spectra (reviewed by Arrondo et al., 1993 and Surewicz et al., 1993).

The advent of molecular biology and protein engineering has opened new avenues for studying the structure and function of membrane proteins. Topologies of numerous membrane proteins have been determined with the gene fusion method in which the gene of a reporter enzyme is fused to the gene of the protein. Depending on the exact site of fusion along the sequence, the reporter enzyme is located either on the cytoplasmic or periplasmic side of the membrane, which is reflected in its activity (Traxler et al., 1993). Many membrane proteins are present in minute quantities or are known only as genes; even abundant proteins are often difficult to obtain in a highly purified and homogeneous form. Both these problems are now being addressed using modern expression methods (for a review see Grisshamer and Tate, 1995) and affinity tags which enable one to use affinity chromatography for purification of the protein (Kleymann et al., 1995; Kuusinen et al., 1995; Mitchell and Gennis, 1995; Wu et al., 1995). This often decreases the number of required purification steps and increases the yield. As discussed by Popot and Saraste (1995), a repertoire of novel techniques is available for manipulation of membrane proteins, especially the helical ones. Among these perhaps the most versatile and systematic is based on cysteine mutagenesis, i.e. introduction of cysteine residues for chemical modification with spin labels (Altenbach et al., 1994; Hubbell and Altenbach, 1994), fluorescent labels such as pyrenes (Jung et al., 1993, 1994) or other types of reagents (Sahin-Toth and Kaback, 1993a; Akabas and Karlin, 1995; Ghaim et al., 1995; Wu et al., 1995) in order to obtain structural information. The cysteine side-chain is moderately hydrophobic and can often be introduced into transmembrane helices without side effects. Studies with bacteriorhodopsin (Altenbach et al., 1990) and lactose permease (van Iwaarden et al., 1991; Sahin-Toth and Kaback, 1993a,b) suggest that it is often possible (i) to replace cysteines in the wild type protein with other residues (which is necessary only if these are reactive under the experimental conditions needed for the desired derivatization, cf. Ghaim et al. (1995)), and (ii) to systematically introduce single cysteine residues into consecutive positions in transmembrane segments and the loops connecting them without major changes in the biological activity and hence in the structure. Each mutant protein, having a cysteine residue in a different position, is

then derivatized, for instance with a nitroxide spin label (Hubbel and Altenbach, 1994). Derivatization of an internally located residue may require unfolding of the protein; for membrane proteins such as bacteriorhodopsin (Huang et al., 1981; Altenbach et al., 1990; Booth et al., 1995), porin (Eisele and Rosenbusch, 1990; Surrey et al., 1996) and the plant light-harvesting complex (Hobe et al., 1994) unfolding and refolding methods exist, but this is a potential methodological problem with many membrane proteins. In most cases studied so far, the spin label appears not to perturb the structure seriously. The electron paramagnetic resonance (EPR) signal of the spin probe depends on its mobility and on its frequency of collision with other paramagnets such as dioxygen and water-soluble transition metal compounds. A protein-facing probe will be relatively immobilized and experience a low oxygen concentration whereas a lipid-exposed probe will be conformationally flexible and, under aerobic conditions, will collide frequently with O_2. If the probe faces water, it will be mobile, experience an intermediate oxygen concentration, and its EPR signal will be sensitive to water-soluble relaxation agents. Consequently, the environment of a cysteine-linked probe can be infererred from its EPR spectrum. The technique is suitable for mapping helix-helix interactions and aqueous-lipid boundaries. Moreover, in a helical structure with one face exposed to the lipid and the other to the protein interior, the EPR signal of the probe should reflect the helical periodicity (i.e., every third or fourth residue in the sequence points in the same direction), providing a means to verify the secondary structure of a given segment. A β-strand can be demonstrated in an analogous manner. Spin labels have also been used to monitor conformational changes on a millisecond time scale during the photocycles of rhodopsin (Farabahbakhsh et al., 1993) and bacteriorhodopsin (Steinhoff et al., 1994). The spin label technique might also be suitable for studying events during protein folding (Hubbel and Altenbach, 1994; Kreimer et al., 1994).

The most recent application of cysteine derivatization is the use of a cysteine-linked backbone-cleavage reagent to map 3D relationships in a two-subunit cytochrome d quinol oxidase complex (Ghaim et al., 1995). The cleavage reagent which is a reductively activated iron-peroxo complex (Ghaim et al., 1995) hydrolyzes a suitably located peptide bond that is not farther than about 12 Å from the cysteine residue. The reaction appears not to be dependent on the sequence but rather on the 3D orientation of the peptide bond relative to the reagent; in the reported cases only one or two peptide bonds were cleaved, meaning that the identification of the cleavage site by N-terminal sequencing is straightforward. The advantage of cysteine-directed backbone cleavage is that, in addition to helping to establish the 2D topology of a membrane protein, it can provide information about subunit interfaces in a multisubunit complex. A related method that involves a cysteine-linked copper-oxo phenantroline compound has been used to study the arrangement of helices in lactose permease (Wu et al., 1995).

A further example of the versatility of sulfhydryl chemistry is disulfide bond formation between two nearby cysteine residues, which can be used to obtain structural information: if the cysteine residues introduced into two neighboring

helices reside on the interhelical surface, they will form disulfide crosslinks either spontaneously or after addition of oxidant (Lynch and Koshland, 1991; Pakula and Simon, 1992; Stoddard et al., 1992; Whitley et al., 1993). Thus the ability to form a crosslink can be utilized to define which amino acid residues are on the interhelical interface.

Many more ways to engineer membrane proteins have been developed (for a review, see Popot and Saraste, 1995). The engineering experiments often produce information which tests the correctness of the current working model of the protein under study. For example, one can insert epitopes or proteolytic cleavage sites into the extramembrane loops (Teufel et al., 1993; Sahin-Toth et al., 1995) or one can fuse genes for individual subunits of an oligomeric protein (Ma et al., 1993). Transmembrane helices can be joined to soluble proteins for which established purification protocols exist (Lemmon et al., 1992). Together biochemistry, protein engineering, and sequence-based structure prediction are able to produce a very rough structural model of a protein, and for a structural biologist these techniques are perhaps best seen as means for getting proper material for crystallization experiments. To determine a structure at an atomic level well-ordered 2D or 3D crystals are needed, and the scarcity of known membane protein structures reflects the extreme difficulty in producing good-quality membrane protein crystals. The major challenge in membrane protein research undoubtedly is: how to crystallize a membrane protein?

B.   Crystallization of Membrane Proteins

While the crystallization of soluble proteins has a relatively high success rate and hundreds of new structures are reported each year, obtaining membrane protein crystals continues to be a difficult problem. In the past 20 years 11 high-resolution membrane protein structures have been solved using X-ray crystallography. In addition, two structures have been determined by electron microscopy of 2D crystals. The conditions of these successful crystallizations are summarized in Table 1. Membrane proteins are thought to form crystal lattices in three different ways, as shown schematically in Figure 2. In general, the probability of getting diffracting crystals may be higher with 2D crystals and electron crystallography, in part because in this technique the detergent plays a less decisive role than in the formation of 3D crystals necessary for X-ray crystallography (see below; Kuhl-brandt, 1992). Moreover, 2D crystallization requires less material, meaning that with many proteins it is the only realistic possibility. Classical crystallography however often leads to a high resolution structure (Table 1) while with electron crystallography the progress appears to be more gradual. So far many low- or intermediate-resolution structures have been determined using 2D crystals and electron microscopy.

In the presence of a detergent at a concentration above its critical micellar concentration (CMC) membrane proteins can be handled much like soluble proteins

Table 1. Crystallization Conditions Which Have Yielded Crystals Used in High-Resolution Structure Determination of Membrane Proteins

Protein	Detergent	Precipitant and Buffer	Temperature During Crystallization	Resolution
3D crystals				
Cytochrome c oxidase (Tsukihara et al., 1995)	Decyl maltoside	PEG 4000; 40 mM phosphate pH 6.8	4 °C	2.6 Å
Cytochrome c oxidase from Paracoccus denitrificans plus the F$_v$-fragment (Iwata et al., 1995; Ostermeier et al., 1995)	Dodecyl maltoside	PEG 2000, 200-400 mM ammonium acetate pH 8.0	20 °C	2.8 Å
Bacterial light-harvesting complex (Papiz et al., 1989; McDermott et al., 1995)	1) Dimethyldecylamine oxide plus benzamidine 2) Octyl glucoside plus benzamidine	1) 2.4 M ammonium sulphate pH 9.5; 0.9 M phosphate pH 9.5 2) 2.2 M ammonium sulphate pH 9.5; 0.9 M phosphate pH 9.5	20 °C	2.5 Å
Maltoporin (Stauffer et al., 1990; Keller et al., 1994; Schirmer et al., 1995)	Decyl maltoside plus dodecyl-nonaoxyethylene	PEG 2000; 20 mM Hepes pH 7.0	Room temperature	3.0 Å
Prostaglandin synthase (Garavito and Picot, 1990; Picot et al., 1994)	Octyl glucoside	PEG 4000 and 100-200 mM NaCl; 20 mM phosphate pH 6.7	Not reported	2.8 -3.5 Å
Porin from Rhodopseudomonas blastica (Kreusch & Schulz, 1994)	Octyltetraoxyethylene	1) PEG 600 and 300 mM LiCl; 20 mM Tris-HCl pH 7.2 2) 200-400 mM ammonium phosphate pH 4.5	1) 20 °C 2) 20 °C	1) 1.8 Å 2) 2.3 Å
Porin from Rhodobacter capsulatus (Kreusch et al., 1991; Weiss and Schulz, 1992)	Octyltetraoxyethylene	PEG 600 and 300 mM LiCl; 20 mM Tris-HCl pH 7.2	20 °C	1.8 Å

Matrix porin from *E. coli* (Pauptit et al., 1991; Cowan et al., 1992) (and phosphoporin from *E. coli* (Cowan et al., 1992))	1) Octyltetraoxyethy lene and octyl-2-hydroksiethylsulfoxide 2) Decyldimethylamine oxide and octyl glucoside	1) PEG 2000 and 700 mM MgCl$_2$; 50 mM Tris-HCl pH 9.8 2) PEG 2000 and 250 mM MgCl$_2$;100 mM piperazine/glycylglycine pH 7.0	Not reported	2.7 Å
Photosystem I from *Synechococcus* (Witt et al., 1988; Krauss et al., 1993)	Dodecyl maltoside	100 - 6 mM MgCl$_2$ ("reverse salting in"); 5 mM MOPS pH 6.4	0-50 °C	4.0 Å
Reaction center from *Rhodobacter sphaeroides* (Buchanan et al., 1993; Ermler et al., 1994)	Lauryldimethylamine oxide and heptanetriol	1.0-1.5 M potassium phosphate pH 6.5-7.8	Five days at 14 °C and then 21 days at 18 °C	2.65 Å
Reaction center from *Rhodopseudomonas viridis* (Michel, 1982; Deisenhofer et al., 1995)	Lauryldimethylamine oxide and heptanetriol	2.2.-2.4 M ammonium sulphate		

2D crystals

Plant light-harvesting complex (Wang and Kuhlbrandt, 1991; Nussberger et al., 1993; Kuhlbrandt et al., 1994)	Triton X-100, nonyl glucoside and endogenous or added lipids	10 mM glycine pH 7.2 and 40% (v/v) glycerol	25 °C for 48 h followed by 30-60 min at 35 °C	3.4 Å
Bacteriorhodopsin (Henderson et al., 1990)	2D crystals occur naturally in the purple membrane of *Halobacterium salinarium*			3.5 Å

and the same crystallization protocols apply. Yet, the change from a protein/buffer/precipitant system to one containing an additional component, the detergent, makes the system significantly more complex (see reviews by Kuhlbrandt, 1988; Garavito and Picot, 1990; Michel, 1990; Reiss-Husson, 1992). The reason for this lies in the amphiphilic nature of both the detergent molecule and the protein: the nonpolar part of the detergent interacts with the lipid-exposed part of the protein,
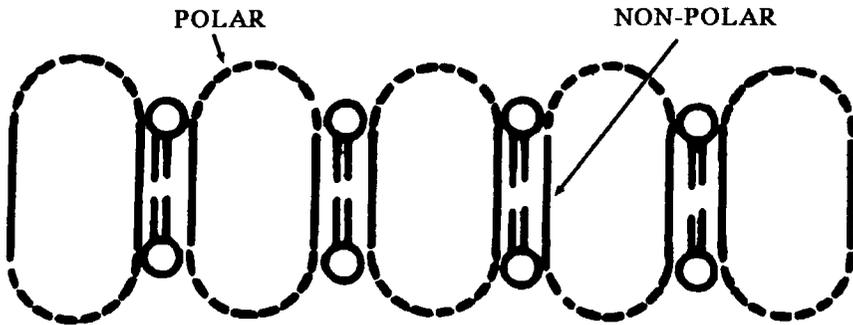
POLAR   **NON-POLAR**

(a) 2D CRYSTAL

Figure 2. Schematic models of three different types of membrane protein crystals, adapted from Michel (1983) and Kuhlbrandt (1992). The polar and nonpolar regions of the protein are marked. **A**. A 2D crystal in which the protein molecules interact laterally much like in a lipid bilayer. In most cases, 2D crystallization requires the presence of lipids which in the figure are represented by the spheres with two tails. Bacteriorhodopsin and the plant light-harvesting complex are examples of proteins which form highly ordered 2D crystals. Both proteins require specific lipids for the formation of the 2D crystal lattice. **B**. A type I crystal. A 3D lattice in which the protein molecules interact primarily via their nonpolar parts. These sort of crystals have been thought to be rare. However, two *E. coli* porins have been found to form well-ordered crystals of this type (Cowan et al., 1992). **C**. A type II crystal. Most membrane protein crystals reported so far belong to this class. All the protein-protein contacts in the lattice are between the polar parts of the protein molecules. Lattice formation is therefore crucially dependent on the availability of polar surface, which in hydrophobic membrane proteins can be limited. In the recently reported cytochrome *c* oxidase crystals (Ostermeier et al., 1995b; Figure 3), the size of the polar domain was increased by complexing the enzyme with an antibody fragment prior to crystallization. The detergent (the squares with one tail) binds as a monolayer to the nonpolar part of the protein. If the the detergent molecules are too bulky, they may not fit into the lattice. In particular, if the detergent headgroups do not interact favorably, crystallization is unlikely.
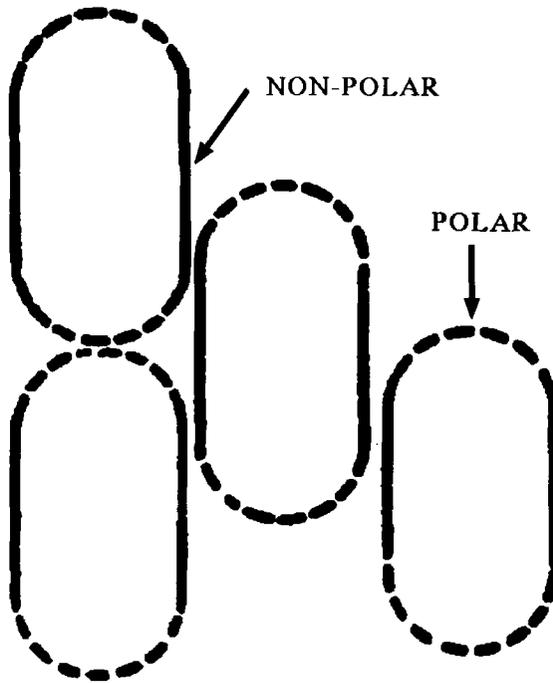
(b) TYPE I CRYSTAL
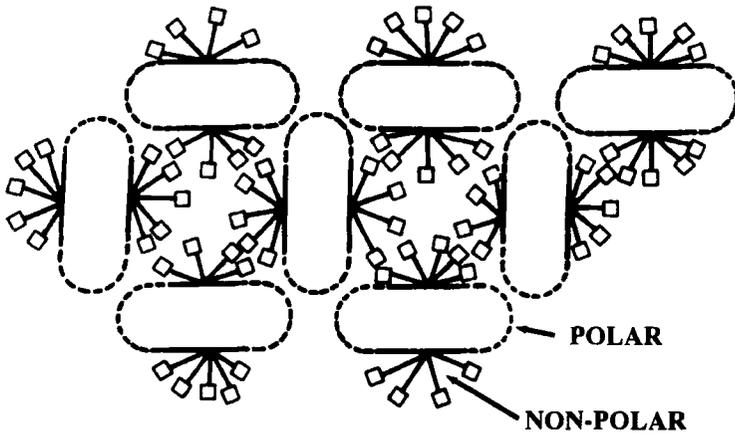
(c) TYPE 2 CRYSTAL

Figure 2 Continued.

replacing the lipid, whereas the polar part of the detergent mimics the lipid headgroups and is in contact with the extramembrane part of the protein. A typical protein-detergent complex may contain one-third of its mass in detergent (Le Maire et al., 1983; Moller and Le Maire, 1993), meaning that the behavior of the system is strongly influenced by the colloidal chemistry of the detergent (Zulauf, 1990).

The properties of a detergent (e.g., CMC, micelle size, phase behavior) depend both on its nonpolar part and on its headgroup. Usually detergents which have a small and charged head group together with a short nonpolar alkyl chain are denaturing for most proteins, whereas a mild detergent that is able to preserve weak protein-protein interactions typically has a large carbohydrate polar group joined to a long alkyl chain. The relative and absolute sizes of the polar and nonpolar parts of membrane proteins vary considerably and so do their sensitivities toward detergents. Of the proteins in the membrane protein structural database, the porins, both light-harvesting complexes and bacteriorhodopsin, have most of their mass inside the bilayer whereas prostaglandin synthase is mostly extramembranous. The reaction centers and cytochrome oxidases have sizeable peripheral and transmembrane domains. All the proteins in the database, perhaps excluding the two cytochrome oxidases, are rather robust and tolerate strong detergents. The choice of an optimal detergent remains an empirical task, complicated by the fact that a detergent that is optimal for stability (see De Grip, 1982) is not necessarily optimal for crystallization (see below). Most membrane proteins whose high-resolution structure has been solved were crystallized using lauryl- or decyldimethylamine oxide and octylglucoside (Table 1). These are relatively strong detergents and cannot be used if one wants to preserve the four subunit structure of cytochrome oxidase from *Paracoccus denitrificans*, for example. Octyltetraoxyethylene appears optimal for porin crystallization. It also seems that that many of the successful crystallizations have been carried out at a relatively high temperature.

The presence of an amphiphilic detergent influences the crystallization experiment at least in two ways: first, the detergent micelle has its own precipitant-induced and temperature-dependent phase behavior (Zulauf, 1990). If a detergent-rich phase forms, the protein also tends to partition into it with the consequence that the very high detergent concentration may denature the protein. However, precipitant concentrations approaching (but not crossing) the phase boundary are often favorable for crystal formation (Kuhlbrandt, 1988; Michel, 1990). Second, the species being crystallized is actually a mixed micelle of protein and detergent. Therefore the micelle size and/or the size of the headgroups influence the formation of the crystal lattice. Too large a micelle or too bulky a headgroup may prevent crystallization or result in poor crystal quality, in particular because protein-protein interactions within the polar, extramembrane domains are thought to be critical in the lattice formation in most cases (Kuhlbrandt, 1988; Michel, 1990; Ostermeier et al., 1995b) and a large micelle may sterically block the contacts between the polar domains (see Figure 2). A bulky headgroup may have the same effect because it interacts with the polar domain and reduces surface available for lattice contacts. On the

other hand, favorable detergent-detergent and detergent-solvent interactions probably play a subtle role in lattice energetics (Roth et al., 1989, 1991). In addition, bound detergent molecules are seen in the *Rhodopseudomonas viridis* reaction center (Deisenhofer et al., 1995) and porin (Weiss and Schulz, 1992; Kreusch and Schulz, 1994) structures, although the high crystallographic temperature factors suggest that the detergent-protein intereaction is not strong. A detergent molecule has also been found in the structure of a bacterial light-harvesting complex (McDermott et al., 1995) and see Note Added in Proof.

Protein-protein interaction in the nonpolar region may also take place. In the crystal lattice of *Escherichia. coli* porin, two porin trimers form a unit cell in which a major protein-protein interaction takes place between the hydrophobic transmembrane regions of the porin β-barrels (Cowan et al., 1992); the interaction surface appears to be rich in aromatic side-chains. Although contacts of this type have been thought to be rare in 3D crystals, they are essential for the formation of 2D crystals (Kuhlbrandt, 1992). A 2D crystal, shown schematically in Figure 2A, is essentially a monomolecular layer in which many protein molecules interact laterally through their intramembrane domains but the crystal is only one molecule thick in the third dimension, resembling the situation in the biological membrane (Kuhlbrandt, 1992). Formation of such crystals normally requires the presence of lipids, either endogenous as with the plant light-harvesting complex (Wang and Kuhlbrandt, 1991; Nussberger et al., 1993; Kuhlbrandt et al., 1994) and cytochrome oxidase (Frey et al., 1978; Frey and Murray, 1994) or added (Wang et al., 1994; Hebert et al., 1995; Jap and Li, 1995; Warne et al., 1995). However, lipids may not always be necessary for 2D crystallization (Cyrklaff et al., 1995).

Bacteriorhodopsin occurs naturally as highly ordered 2D crystals in the purple membrane of *Halobacterium salinarium* (Henderson et al., 1990) whereas halorhodopsin has been shown to arrange *in vivo* into a 2D lattice only when the protein is overexpressed (Havelka et al., 1993). Polar *Halobacterium* lipids are crucial to the formation of the 2D bacteriorhodopsin lattice, probably because the polar headgroups participate in the lattice formation (Popot et al., 1987; Sternberg et al., 1992). Lipids may also be important in 3D crystallization: two types of thylakoid lipids are needed to maintain the plant light-harvesting complex crystallization-competent (Nussberger et al., 1993). A specific lipid-binding site has been found in a bacterial cytochrome *c* oxidase (Iwata et al., 1995; see Figure 6 and Note Added in Proof) and in a porin (Kreusch and Schulz, 1994).

C.   Solving the Structure of Cytochrome *c* Oxidase from *Paracoccus denitrificans*

In 1995, unexpectedly, two groups announced that they had solved the structure of cytochrome *c* oxidase, the terminal enzyme of the respiratory chain of mitochondria and many bacteria (Iwata et al., 1995; Tsukihara et al., 1995). Iwata et al. (1995) reported the complete structure of a four-subunit enzyme from the bacterium *P.*

denitrificans (M_r = 127 kDa), whereas the paper of Tsukihara et al. (1995) describes the structures of the metal sites of bovine heart cytochrome oxidase (M_r = 205 kDa, composed of 13 polypeptides of which the three largest have homologues in the bacterial oxidase complex). The bovine enzyme crystals were obtained using a conventional trial-and-error method; the breakthrough finally came after years of effort. The approach employed by Michel and colleagues, although also labor-intensive, makes use of the tight and specific binding of an antibody to its epitope: the bacterial enzyme was co-crystallized with a monoclonal antibody specifically engineered for this purpose. For this reason, the method used with the bacterial enzyme could be a more general one and be applicable to crystallization of other membrane proteins as well. The enzyme from *Paracoccus* and the strategy used by Iwata et al. (1995) and Ostermeier et al. (1995b) will be discussed next.

Respiratory cytochrome *c* oxidases (EC 1.9.3.1) are enzymes that catalyze the reduction of dioxygen to water with concomitant oxidation of cytochrome *c*. Part of the free energy derived from this reaction is used to translocate protons across the membrane housing the oxidase (Babcock and Wikström, 1992; Wikström et al., 1994). Cytochrome *c* oxidase from *P. denitrificans* has four dissimilar integral membrane protein subunits. Polypeptides I and II contain the catalytically active metal sites: two A-type hemes, of which one is low and the other high spin, and two copper centers (Cu_A, a mixed-valence dinuclear center in subunit II, and Cu_B, a mononuclear center located in the proximity of the high spin heme in subunit I). Subunits III and IV are devoid of redox cofactors and their role in the enzyme remains to be understood. The four subunits contain a total of 22 transmembrane helices. The locations and sequences of the 21 helices in the three largest subunits are shown in Figures 4 and 5.

The protocol of Michel colleagues (Kleymann et al., 1995; Ostermeier et al., 1995b) involves (i) purification of the oxidase using conventional ion-exchange chromatography in the presesence of the mild non-ionic detergent dodecyl maltoside, (ii) raising monoclonal antibodies against the native enzyme complex, (iii) making a construct that encodes the variable portions of the antibody (the F_v fragment) and expressing it in *E. coli*, (iv) affinity purification of the antibody using an engineered streptavidin affinity tag, and (v) crystallization of the oxidase-F_v complex. The affinity-tagged antibody could also be utilized to purify the oxidase (Kleymann et al., 1995), but in the crystallization a conventionally purified preparation was used (Ostermeier et al., 1995b).

As discussed above, the critical protein-protein contacts in the crystal lattices of membrane proteins occur almost exclusively between the polar extramembrane parts of the protein molecules (see Figure 2). (The only exception so far are the two *E. coli* porins, as mentioned above (Cowan et al., 1992)). Consequently, 3D crystallization of proteins which are largely intramembranous may pose a problem, in particular when one has to use a detergent with a large polar headgroup and a large micelle. The rationale in using the F_v-fragment-oxidase complex is to increase the polar part of the oxidase and thus increase the likelihood of polar interactions

among the protein complexes. In addition, as the antibody recognizes a conformational epitope (i.e., the antibody binds to residues that are located near each other in the 3D structure but are not necessarily close in the amino acid sequence), it may stabilize a single conformation of the enzyme and thereby promote crystallization. One should also note that the particular F_v-fragment used shows a strong tendency to yield highly ordered crystals on its own (Ostermeier et al., 1995a), which may be an important factor in the successful outcome of the experiment (Ostermeier et al., 1995a). The crystals of the F_v-fragment diffract beyond 1.28 Å resolution.

The antibody against the oxidase complex recognizes an epitope in the peripheral part of subunit II (see Figures 3 and 6A). In addition, it binds to the cytoplasmic loops of subunit III and to another F_v in the neighboring unit cell. All the lattice contacts in the a, b plane of the oxidase- F_v -co-crystals (Figure 3) are between the F_v antibody molecules: none of the oxidase molecules touch another oxidase molecule. In conclusion, the crystallization is entirely mediated by the F_v-fragment. The arrangement of the molecules in the crystal lattice lend support to the general idea that polar interactions are decisive for the formation of most membrane protein crystals.

IV. THE STRUCTURE OF A FOUR-SUBUNIT CYTOCHROME *C* OXIDASE

A. Structure and Structure Prediction

The X-ray structure of the oxidase solved by Iwata et al. (1995) at 2.8 Å resolution (see Figure 6) shows that subunits I, III, and IV are largely helical with 12, 7, and 1 long transmembrane helices, respectively. Subunit II has only two transmembrane spans and most of the subunit constitutes a compact peripheral domain with antiparallel β-strands as the major secondary structure element. Subunits III and IV bind to subunit I mostly via lateral interactions that take place in the membrane plane. There are only four short helical segments which do not belong to the transmembrane helices: two helices in subunit I loops and two in the extramembrane domain of subunit II (see Figure 5). About 50% of the oxidase is in the α-helical conformation. The dinuclear Cu_A-site is located in a loop between two β-strands in subunit II, whereas all the other redox-active metal sites are in or near the helical region of subunit I. The low spin heme (heme *a*) is ligated by two histidines in subunit I helices II and X . The latter helix also ligates the high spin heme, known as heme a_3 ; the Cu_B-site is located 5.2 Å from Fe $_{a3}$ and is ligated by three histidines, one from helix VI and two from just beyond the end of helix VII. Spectroscopic evidence suggests that a water or hydroxyl oxygen is coordinated to Cu_B as well (Fann et al., 1995). The metal center ligands and their relation to the transmembrane helices are shown in Figure 5.

Figure 3. Structure of the cytochrome c oxidase-F_v-fragment crystal lattice in the a, b plane of the lattice (Ostermeier et al., 1995b). Two unit cells comprising four oxidase and four antibody molecules each are shown. The antibody is colored black, subunit I is white, subunit II dark gray, subunit III light gray and the single helix of subunit IV is black. Note the central position of the F_v-molecule in the crystal lattice: not only does the antibody bind to its epitope in the peripheral domain of subunit II but it is also in contact with subunit III of another oxidase molecule and with another antibody fragment. Hence, the lattice is held together by the antibody. Courtesy of Drs. Iwata and Michel. Plotted with MOLSCRIPT (Kraulis, 1991).

250

The number and approximate locations of the transmembrane helices correspond to those predicted using a Kyte and Doolittle (1982) hydropathy analysis (see Saraste, 1990; cf. Engelman et al., 1986; Michel et al., 1986; Figure 4). However, the structure shows that the actual helices are longer than predicted, ranging from 19 to 38 residues in length (Iwata et al., 1995). One reason for this is that the helices are tilted relative to the membrane normal; the other is that some of them have hydrophilic ends beyond the hydrophobic core of the bilayer. As shown in Figure 5, the helices in subunits I and III commonly start and end with prolines and glycines. Aspartates and asparagines are also common as helix caps. Residues near the ends of the helices tend to be conserved (cf. Saraste, 1990). Five strongly polar residues, shown in bold and underlined in Figure 5, occur in the membrane-spanning region: a glutamate and a lysine in subunit I helices VI and VIII, respectively, a glutamate and an aspartate in helices III and VII of subunit III as well as a lysine in the single helix of subunit IV (not shown in Figure 5). The

(a) Subunit I

(b) Subunit II

(c) Subunit III
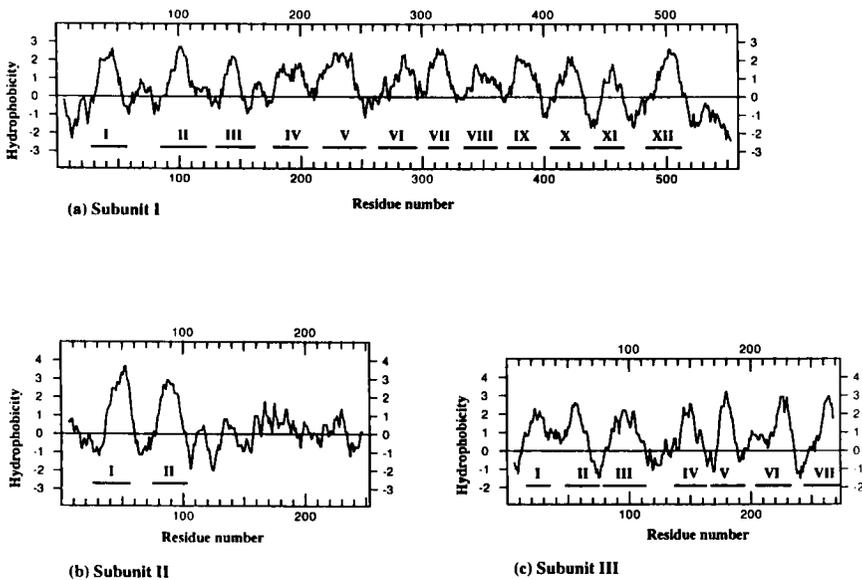
Figure 4. Kyte-Doolittle hydropathy analysis of the sequences of the three large subunits of cytochrome c oxidase from *Paracoccus denitrificans.* A window of 11 residues was used. Hydrophobicity increases upwards. The bars marked with Roman numerals beneath the hydropathy plots indicate the true locations of the transmembrane helices in the structure at 2.8Å resolution (Iwata et al., 1995). It can be concluded from the figure that the hydropathy analysis alone is able to predict the number and approximate locations of the 21 helices correctly (cf. Table 2). The sequences of subunit I, II, and III are from Raitio et al. (1990), Steinrucke et al. (1987), and Raitio et al. (1987), respectively.

exact functions of these residues are not known; those in subunits I and III are conserved in all aa_3-type cytochrome c oxidases. Mutagenesis of the subunit I glutamate and lysine shows that both residues influence the electron transfer activity (Fetter et al., 1995). Iwata et al. (1995) postulate that they are involved in delivering protons to the active site (see below). The conserved glutamate in subunit III helix III may form an ion pair with a conserved histidine in helix VI of subunit III (Iwata et al., 1995).

As discussed in Section II.A, porins, bacteriorhodopsin, as well as the reaction centers and prostaglandin synthase all have aromatic residues that cluster at the

A
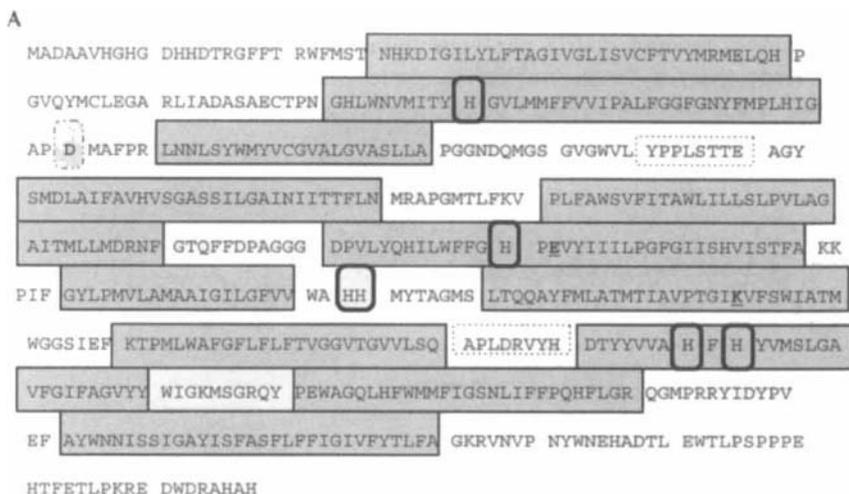
Figure 5. The locations of the transmembrane helices (shaded boxes) and extramembrane helices (dashed transparent boxes). The metal ligands are indicated by black frames. **A**. Subunit I. The ligand histidines are in helices II (heme a), VI (Cu_B) and X (hemes a and a_3). The two histidines near the end of helix VII are ligands to Cu_B. The aspartate residue (shaded box with broken line) between helices II and III is located at the mouth of a proton channel and is essential for proton translocation (Thomas et al., 1993). Note the presence of two strongly polar residues (shown in bold and underlined) in the hydrophobic region of the transmembrane helices: a glutamate in helix VI and a lysine in helix VIII. Both termini of subunit I are in the inside of the membrane (i.e., in the cytoplasm). **B**. Subunit II. The carbonyl oxygen of the glutamate (shaded box with dashed line) between the ligand cysteines is also a ligand to the Cu_A-center. The sequences of the β-strands in the peripheral domain are underlined and in bold. Both termini of subunit II are on the outer (periplasmic) side of the membrane. **C**. Subunit III. Note the long helices and the short loops, especially on the cytoplasmic side. Strongly polar residues in the hydrophobic region are in bold and underlined. The two residues in helices III and VI (shown in a larger font) are associated with Leber's disease in humans (Johns and Neufield, 1993). The N-terminus of subunit III is in the bacterial cytoplasm and the C-terminus in the periplasm.

B

Q DVLGDLPVIG KPVNGGMNFQPASSP LAHDQQWLDHFVLYIITAVTIFVCLLLLICIVRF N

RRANPVPARF TH NTPIEVIWTLVPVLILVAIGAFSLPILFRSQE MPNDPDLV **IKAIGH** QWY

WSYEY PNDG **VAFDA** LML EKEALADAG YSEDEYLLATDNP **VVV** PVGK **KVLVQVT**

ATDVI H **AWTI** PAFA **VKQDA** VPGRI **AQLW** FSVDQ **EGVYFGQ** C S E L C GIN

H AY M **PIVVKAV** SQEKYEAWLAGAKEE FAA

C

AHVKNHDYQI LPPS IWPFFGAIGAFVMLTGAVAWM KGITF FGLPVEG PWMFLIGLVGVLY

VMFGWWADVVNEGETG EH TPVVRIGLQYGFILFIMSEVMFFVAWFWAPIKNALY PMGPDS

PIKDGVWPPE GIVTFDPW HLPLINTLILLLSGVAVTWAHHAFVLE GD RKTTINGLIVAVI

LGVCFTGLQAYEYSHA AFGLAD TVYAGAFYMATGFHGAHVIIGTIFLFVCLIRLLK GQMT

QKQ HVGFEAAAWYWHFVDVVWLFLFVVIYIWGR

Figure 5 Continued.

assumed lipid-aqueous boundary. In Figure 7 the aromatic residues in the bacterial oxidase are highlighted. It is clear that in the oxidase the locations of aromatic residues follow the same interfacial pattern as in the other membrane proteins. Most of the tyrosine residues appear on the periplasmic lipid-aqueous boundary and are nearest to the aqueous phase. The middle part of the interface is rich in tryptophan residues whereas phenylalanine residues face the hydrophobic core of the membrane. It can be concluded that interfacial aromatic residues are a common feature among the membrane protein structures known at present.

Recently a number of novel prediction methods have been developed (e.g., the method of Persson and Argos (1994), based on multiple sequence alignments and statistical analysis of hydropathy predicted transmembrane segments, and the neural network method of Rost et al. (1995), which also utilizes a multiple sequence alignment). The new oxidase structure with its 22 transmembrane helices provides an excellent opportunity to test the predictive power of these methods (see Table 2). In general, both the method of Persson and Argos (1994) and Rost et al. (1995) are able to predict correctly the number of the transmembrane helices in subunits I and II. The algorithm of Persson and Argos identifies the ends and lengths of the helices in these two subunits slightly more accurately than does the method of Rost and co-workers: when each difference in sequence position relative to the structure counts one point, the Persson and Argos method gets 95 points while the neural

Table 2.   The Location of the Transmembrane Helices in Subunits I, II, and III of
                Cytochrome *c* Oxidase From *Paracoccus denitrificans*

| | | Residues (Length) | | |
| | | Structure | Predicted | Predicted (Persson |
Subunit		(Iwata et al., 1995)	(Rost et al., 1995)	and Argos, 1994)
Subunit I				
Helix	I	27-59 (33)	35-51 (17)	28-56 (29)
	II	84-121 (38)	89-120 (32)	92-120 (29)
	III	130-151 (22)	135-147 (13)	129-154 (26)
	IV	178-206 (29)	183-204 (22)	179-207 (29)
	V	218-251 (34)	219-244 (26)	219-247 (29)
	VI	263-298 (36)	273-294 (22)	266-294 (29)
	VII	304-322 (19)	307-327 (21)	301-327 (27)
	VIII	334-362 (29)	341-359 (19)	337-360 (24)
	IX	370-395 (26)	373-393 (21)	371-399 (29)
	X	404-430 (27)	407-432 (26)	407-435 (29)
	IX	441-468 (28)	448-462 (15)	449-477 (29)
	XII	483-513 (31)	489-511 (23)	482-505 (24)
Subunit II				
Helix	I	27-59 (33)	36-58 (23)	33-61 (29)
	II	74-105 (32)	78-102 (25)	74-101 (28)
Subunit III				
Helix	I	15-35 (21)	22-35 (14)	16-44 (29)
	II	48-76 (29)	48-65 (18)	not predicted
	III	79-114 (36)	86-113 (28)	85-113 (29
	IV	139-165 (27)	139-157 (19)	137-161 (25)
	V	168-196 (29)	173-191 (29)	167-195 (29)
	VI	203-236 (34)	202-235 (34)	208-236 (29)
	VII	244-273 (30)	249-272 (24)	246-273 (28)

network algorithm of Rost et al. collects 126 points. In 14 out of 42 cases, the
Persson and Argos algorithm predicts one of the helix ends to within one residue
of the correct location. The average lengths of transmembrane helices in subunits
I and II are 27.9 (Persson and Argos, 1994) and 21.8 (Rost et al., 1995) residues
while the average length according to the structure is 29.8. It should be noted that
the Persson and Argos method assumes that a transmembrane helix cannot be longer
than 29 residues. It is possible that without this restriction the method would have
yielded an even better estimate for the helix length. However, the Persson and Argos
algorithm fails to identify helix II of subunit III as a transmembrane segment, most
likely owing to inclusion of too distantly related sequences in the alignment used
as the input for the prediction. (Within the heme-copper oxidase family some
members lack the two first helices of subunit III (Castresana et al., 1994); in the
structure of Iwata et al. (1995) (Figure 6) these two helices form a separate

subdomain.) The problem can be eliminated by inspecting the automatically generated alignment and by deleting the shorter subunit III versions from the alignment. The manually corrected alignment is then used as the input. The network method of Rost et al. (1995) uses a different protocol to generate the sequence alignment which does not include the shorter subunit III homologues and appears therefore to avoid the problem. In general, it can be concluded that neither of the
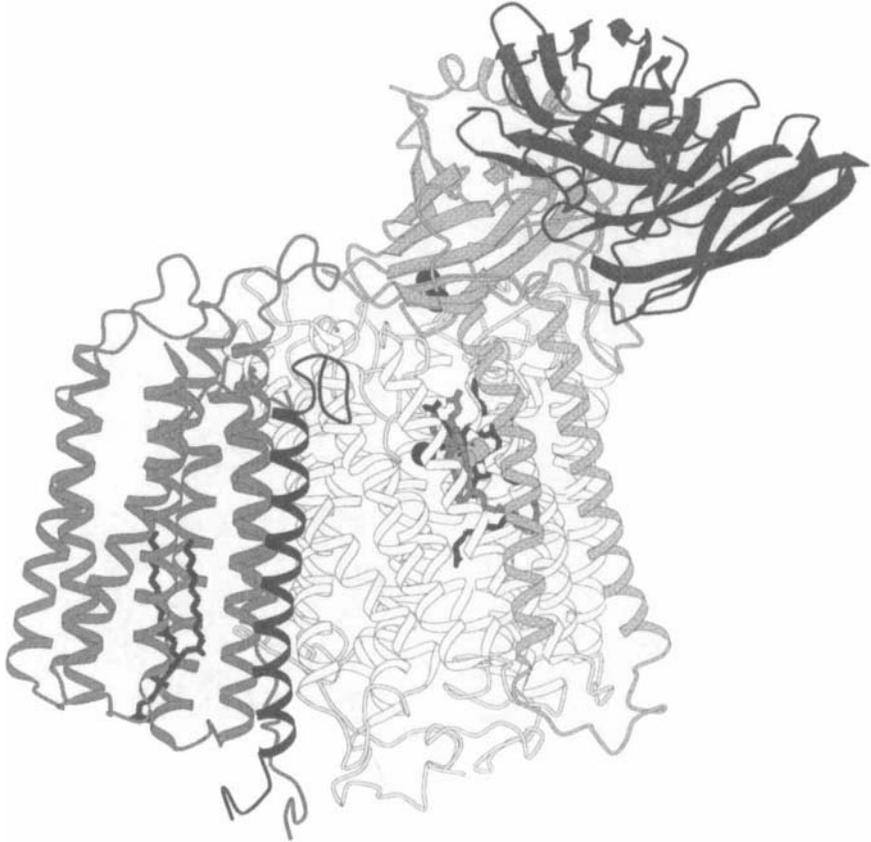
Figure 6A. The structure of cytochrome c oxidase-F$_v$-fragment complex as seen from the membrane plane. Subunit I is shown in white, subunit II in light gray, subunit III in dark gray, and subunit IV in black. The F$_v$-antibody which binds to subunit II is colored black in the right corner of the structure. The Cu$_A$-center in subunit II is marked with two black spheres, hemes a and a$_3$ are black and dark gray, respectively, whereas both heme irons are light gray. Cu$_B$ is shown as a black sphere to the left from heme a$_3$. The bound phospholipid molecule in subunit III is also indicated. Courtesy of Drs. Iwata and Michel. Plotted with MOLSCRIPT (Kraulis, 1991).

methods tested is able to predict the helix ends very reliably. For many hydrophobic sequences such as those of the three oxidase subunits, "old-fashioned" hydropathy analysis (Figure 4) performs almost equally well.

The function of the enzyme is to transfer electrons from cytochrome c to dioxygen which is reduced to water. The most intriguing aspect of the cytochrome oxidase activity is that this redox reaction is coupled to proton translocation across the bacterial cell membrane, i.e., cytochrome oxidase is a redox-linked proton pump (Wikström et al., 1994). The electrons from cytochrome c enter the enzyme at Cu_A, from which they are transferred via the low spin heme to the oxygen reduction site, made up by the high spin heme and Cu_B. The catalysis at the latter site involves two types of protons: first, the protons translocated across the membrane (vectorial protons) and, second, the protons used to make water. These latter protons (scalar
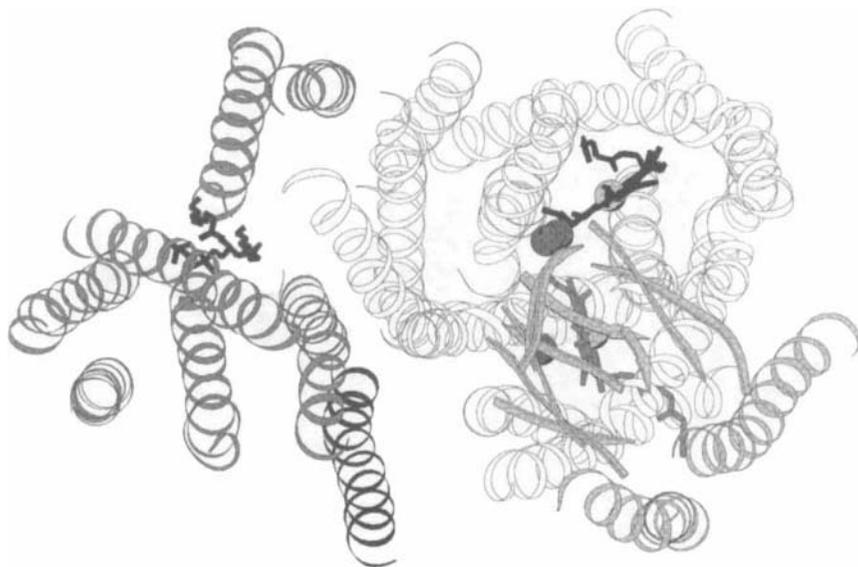
Figure 6B. The structure of cytochrome c oxidase from *P. denitrificans* (Iwata et al., 1995) as viewed above the membrane from the periplasmic side. Only the transmembrane helices and part of the peripheral Cu_A-binding domain composed of β-strands are displayed. Subunit I is shown in white, subunit II in light gray, subunit III in dark gray and subunit IV in black. Cu_A is represented by two spheres at one corner of subunit II above the approximate center of the three helical bundles in subunit I. Heme a is black and heme a_3 is gray; the gray sphere which represents Cu_B, although difficult to see, is located to the left from heme a_3 beneath the edge of subunit II. The phospholipid molecule between the helices II and VI in subunit III is shown in black. The transmembrane helices are identified in Figure 6C. Courtesy of Drs. Iwata and Michel. Plotted with MOLSCRIPT (Kraulis, 1991).

protons) are also taken up from the inside of the membrane. The structure shows that the active site is situated quite close to the outer (periplasmic) side of the membrane (see Figure 6D), suggesting that there must be rather extensive protein structures that control the access of protons from the inner aqueous phase to the heme a_3-Cu$_B$ center. Iwata et al. (1995) suggest that the transmembrane helices in subunit I form two proton channels that lead to the active site.

The 12 helices in subunit I are organized in an anticlockwise fashion, forming three bundles with a blocked pore in the middle of each (Figure 6). The three bundles comprise helices III, IV, V, and VI (lining pore A), helices VII, VIII, IX, and X (pore B) and helices XI, XII, I, and II (pore C). The last helix in each bundle (VI, X, and II) is located in the core of the whole structure and contains strictly conserved metal ligands (Figure 6C). Pores A and B lead from the cytoplasmic side to near the active center and are proposed to make up the proton channels, of which one might be specific for the vectorial (pore A) and the other for the scalar protons (pore B) (Iwata et al., 1995). On the basis of mutagenesis studies, several polar residues in pore A helices II and III and in the cytoplasmic loop between them may
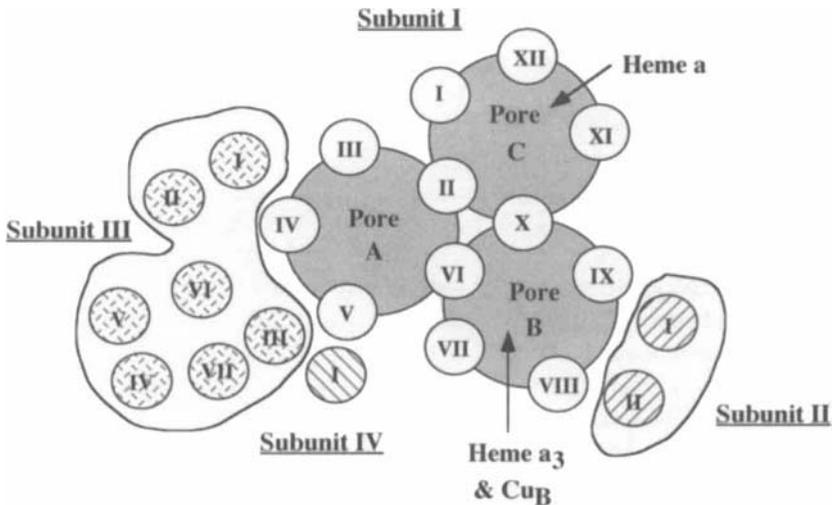
Figure 6C. A schematic model of the oxidase shown in Figure 6B. Note the arrangement of the helices in subunit I, which results in three pore-like channels. Pores A and B may be involved in controlling the access of both scalar and vectorial protons to the heme a_3-Cu$_B$ center which is the site for oxygen reduction (Iwata et al., 1995). The histidine ligands of the redox active metals in subunit I are provided by the centrally located helices II (heme a), VI (one Cu$_B$-ligand), and X (both heme a and a_3), and by VII (two Cu$_B$-ligands). Note also the unforeseen arrangement of the seven helices in subunit III.

participate in proton transfers (Thomas et al., 1993; Fetter et al., 1995; Garcia-Horsman et al., 1995). In particular, an aspartate in the loop (shown in Figure 5A) is essential for proton translocation (but not for electron transfer; Thomas et al., 1993). In the structure it appears to be located at the orifice of one of channels. On the other hand, some mutations in polar residues (e.g., the intramembrane lysine shown
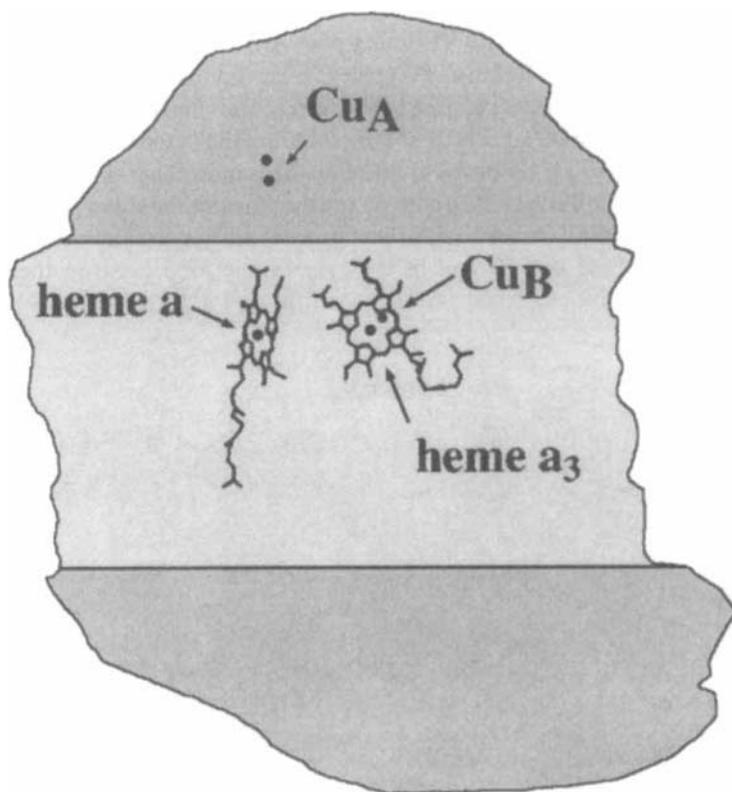
Figure 6D. A side view of cytochrome oxidase showing the locations of the redox-active metal centers relative to the membrane surface. The transmembrane part of the molecule (shown in lighter shading), defined by the transmembrane helices, is about 48 Å wide. All the metal centers are close to the outer side of the membrane. The heme planes are parallel to the membrane normal whereas the angle between the heme planes is 104° (bovine enzyme) or 108° (the bacterial enzyme). The distance from Cu_A to the iron of heme a is 19 Å ; the heme irons are within 14 Å from each other. The figure is adapted from Tsukihara et al. (1995) and is based on the structure of the bovine cytochrome oxidase. (Although the mammalian enzyme is larger than the bacterial one, the structures of the redox-active metal centers are very similar in both.)

in Figure 5A) in the amphipathic helix VIII in pore B result in inactive or almost inactive enzyme without affecting the structures of the metal centers (Fetter et al., 1995). Together with the 3D structure these results have been interpreted as suggestive of separate paths for the pumped and substrate protons. Further work is needed, however, to verify whether the conceptual distinction between the vectorial and scalar protons is realized physically. Moreover, it is not clear at this time how dioxygen finds its way to the active site or how the product, water, is expelled from the interior of the membrane.

The molecular events which couple oxygen reduction to proton translocation are not known, but many of the proposed mechanisms involve the ligands to the active site metal ions. Much recent attention has been focused on the role of the imidazoles that bind Cu_B (Haltia and Wikström, 1992; Morgan et al., 1994; Wikström et al., 1994; Iwata et al., 1995), especially the two adjacent histidines located near the end of the shortest transmembrane helix, helix VII: these two residues lie just outside the periplasmic end of the helix, i. e. at the borderline of the rigid intramembrane helical structure and a possibly more flexible extramembrane domain (cf. Figure 5). One of the histidine side-chains has no electron density in the crystals of Iwata et al. (1995), which may indicate that the histidine has multiple conformations. This observation is interesting because current models of proton translocation propose a mechanism that involves a redox-induced change in the geometry of the Cu_B-center, coupled to vectorial protonation and deprotonation of the ligand imidazoles.

B.   Subunit II

Subunit II associates with subunit I pore B tightly both via its two long transmembrane helices and the periplasmic soluble domain (Figure 6). The two coppers of the Cu_A-center bind at one edge of the soluble domain and are located above the middle of the three helical bundles of subunit I. The copper site is buried from the solvent. The fold of the periplasmic domain resembles the β-barrel structure of plastocyanin, a soluble type 1 copper protein. However, the Cu_A-site in subunit II is not mononuclear as in plastocyanin but a dinuclear copper center ligated by two cysteines, two histidines, a methionine and a peptide carbonyl oxygen, as indicated in Figure 5. The strongest copper ligands occupy symmetric positions around the metal ions: the two cysteines are located equatorially between the coppers bridging them whereas the histidines are on the opposite sides of the complex acting as terminal ligands. This basic symmetry probably results in very similar chemical environments for both coppers, consistent with the functioning of the dinuclear center as a one-electron acceptor and donor and with the formal valence of 1.5 for both coppers in the oxidized state.

The function of subunit II is to accept electrons from cytochrome *c* and transfer them further via Cu_A to the low spin heme and the oxygen reduction site. Accordingly, subunit II provides at least a part of the docking site for cytochrome *c*.

Although subunit II is tightly associated with subunit I, deletion of the transmembrane helices and separate expression of the extramembrane part yields a soluble protein with spectroscopic characteristics similar to those of Cu_A (Lappalainen et al., 1993). The Cu_A-site has also been created in a domain of a quinol oxidase which has a homologous subunit without the metal center (van der Oost et al., 1992; Kelly et al., 1993). The structure of this soluble fragment, solved by Willmanns and colleagues (1995), closely resembles that of subunit II as reported by Iwata et al. (1995).

C.   The Fold of Subunit III

Subunit III is a highly hydrophobic protein with seven long transmembrane helices (21-36 residues; see Figures 5 and 6), three of which are in side-to-side contact with the pore A helices of subunit I. Some of the extramembrane loops on the periplasmic side appear to interact with loops from the two other subunits while helix III of subunit III faces the single helix of subunit IV. None of these interactions are strong however: subunit III (along with subunit IV) can be removed readily either biochemically using a stronger detergent or genetically without major perturbation of subunits I and II (Haltia et al., 1989, 1994). Thus subunit III is bound weakly to the tight complex of subunits I and II, constituting a relatively independent but unstable domain (see below). Within the subunit, helices I and II form a subdomain, separated from the other helices by a large groove in which a phospholipid molecule binds (Figure 6A). The helices that contain the most conserved residues are III, VI, and VII; they occupy the most deeply buried positions in the five-helix bundle. Interestingly, calorimetric and spectroscopic experiments show that subunit III is thermally rather unstable when compared to many other membrane proteins (Haltia et al., 1994; Haltia and Freire, 1995; Echabe et al., 1995). Unlike the two catalytic subunits and many other membrane proteins, it is also very sensitive toward the denaturant guanidine hydrochloride (Hill et al., 1988; Haltia et al., unpublished): the calorimetric transition at 46-50 °C caused by denaturation of subunit III (Haltia et al., 1994) completely and specifically disappears at about 2 M guadinine hydrochloride. The solvent denaturation appears to be reversible, as removal of the denaturant restores the calorimetric transition (Haltia et al., unpublished). The low stability of subunit III probably results from the peculiar arrangement of the long and tilted helices, from the very short loops particularly on the cytoplasmic side and from the absence of stabilizing cofactors; together these features, along with the general hydrophobicity of the subunit, may cause the structure to be relatively open, perhaps allowing efficient hydration upon denaturation.

The function of subunit III in the oxidase complex is unknown. However, two point mutations in human oxidase subunit III have been found to be associated with Leber's disease (Johns and Neufield, 1993) (see also Note Added in Proof). The locations of these mutations (a Gly → Ser substitution in the polar N-terminal end

of helix III and an Ala → Thr substitution near the start of helix VI) are indicated in Figure 5C. Considering the low stability of subunit III discussed above, these mutations might interfere with the structural integrity of the subunit. Although the specific role of the subunit has not been elucidated, some results suggest that subunit III influences the electron transfer activity of the oxidase (Brown et al., 1993; Haltia et al., 1994).

V. NOTES ON OTHER MEMBRANE PROTEIN STRUCTURES

A. Porins

The function of porins is to form specific or nonspecific channels to facilitate the diffusion of small solutes across the outer membrane of gram-negative bacteria. Currently the high-resolution structures of five porins are known (Cowan et al., 1992; Weiss and Schulz, 1992; Kreusch and Schulz, 1994; Schirmer et al., 1995). Two of the *E. coli* porins, matrix porin and phosphoporin, are highly homologous to each other (63% sequence identity) (Cowan et al., 1992) but are not sequence-related to the mutually weakly homologous (24% identity) porins from *Rhodobacter capsulatus* (Weiss and Schulz, 1992) and *Rps. blastica* (Kreusch and Schulz, 1994). Maltoporin, which is specific for disaccharides (Schirmer et al., 1995), is not homologous to any of the above porins. Yet, all these porins have a similar overall fold that consists of a barrel of 16 (18 in maltoporin) tilted antiparallel β-strands with an aqueous pore in the middle. Moreover, all five porins are homotrimers in which the monomers associate so that the wall of an individual barrel interacts either with lipid or with a neighboring porin monomer. A porin monomer is shown in Figure 7A.

Porins exhibit several interesting features: despite being integral membrane proteins, the porins have hydrophilic primary structures similar to soluble proteins. No long hydrophobic stretches are present in the sequence. The porin trimer is highly stable and is denatured only at temperatures near 100 °C even when sodium dodecyl sulfate is present. Inspection of the structure suggests that a number of features are responsible for the extraordinary stability of porins (which almost certainly explains why so many high resolution structures are available for this class of membrane protein): First, the β-barrel, which includes more than half of the residues, is stabilized by extensive interstrand hydrogen bonding. Many of the hydrogen bonds are short (less than 3 Å). In addition, their strength is expected to be high because of the absence of water inside the lipid bilayer. Both aqueous-lipid borderlines of the barrel are rich in aromatic residues (see Figure 8); the width of the nonpolar surface of the β-barrel, including the interfacial aromatics, is 24-25 Å. Second, the strands are connected by loops, some of which have low mobilities. In particular, the long loop between strands 5 and 6 is rigid and highly structured; it folds back into the pore and makes extensive contact with the inner wall of the barrel and defines the pore size; its structure may also influence the shape of the
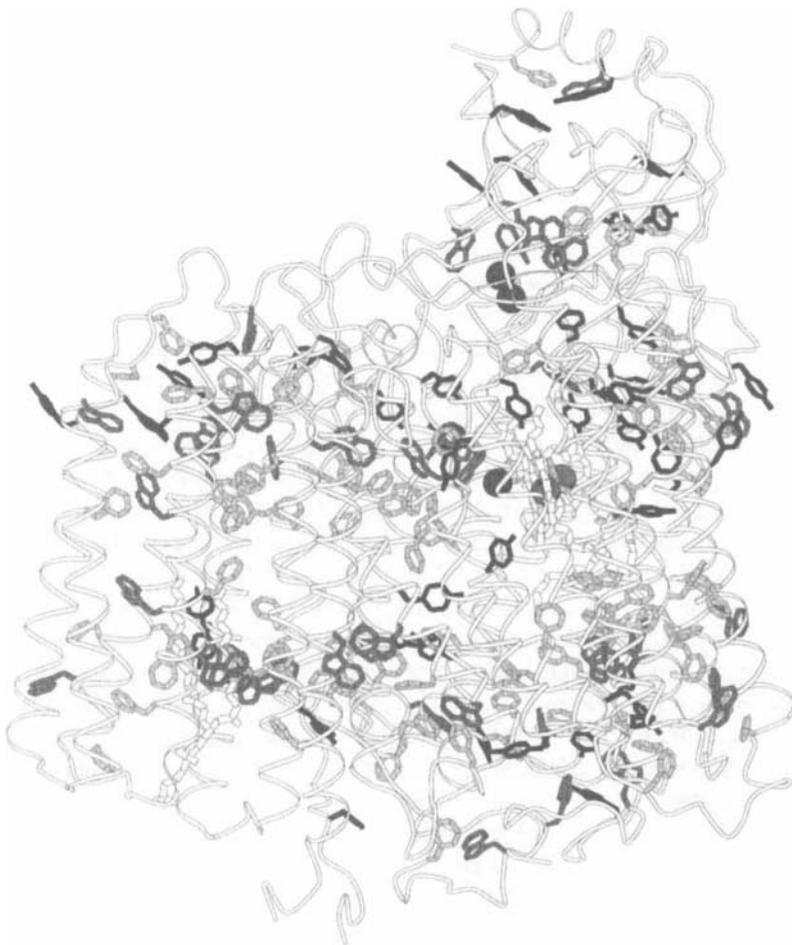
Figure 7. Distribution of aromatic residues in the bacterial cytochrome c oxidase. Tyrosine residues are colored black, tryptophans dark gray and phenylalanines light gray. The oxidase molecule is seen from the same angle as in Figure 6A. The aromatic residues cluster at both membrane-aqueous interfaces and and in the core of the peripheral domain of subunit II. Courtesy of Drs. Iwata and Michel. Plotted with MOLSCRIPT (Kraulis, 1991).
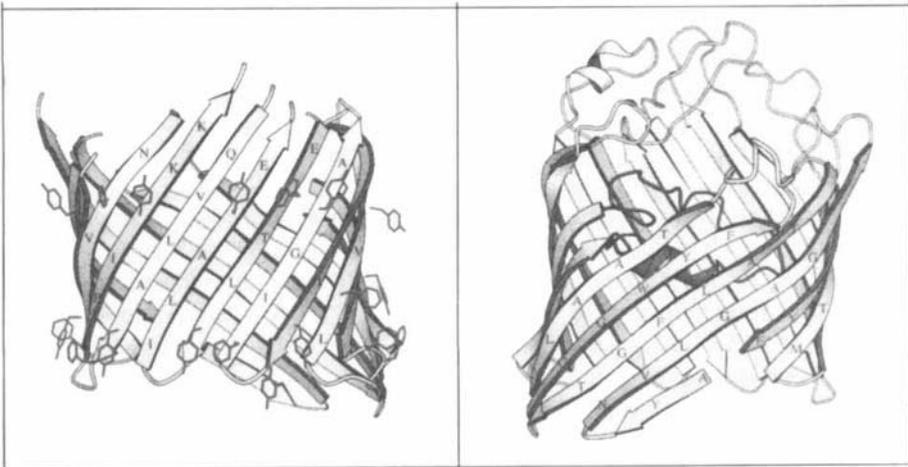
Figure 8. An *E. coli* porin monomer seen from the membrane plane (left panel) or from the center of a porin trimer (right panel). In the membrane three porin monomers associate to form a highly stable trimer. A porin monomer shown in the figure is composed of 16 transmembrane antiparallel β-strands joined by loops (shown only in the right panel) which are short on the periplasmic (bottom of the figure) and long and structured on the extracellular (top) side. One of the loops folds into the barrel channel and defines the permeability of the channel. As indicated by the letters in the strands in the left panel, the surface exposed to the lipid is hydrophobic, comprising mainly short aliphatic side-chains. The membrane spanning part of the β-barrel is flanked by aromatic (mostly tyrosine) residues. In the right panel the residues facing the viewer are those located on the trimer interface, which is tightly packed and resembles the hydrophobic core of soluble proteins. Large aromatic side-chains from one barrel wall interact favorably with small residues in the wall of the other monomers. This structural feature at least partially explains the extraordinary stability of porins. Reprinted with permission from Cowan and Rosenbusch, Science 264, 914-916. Copyright 1994 American Association for the Advancement of Science. Courtesy of Dr. Rosenbusch.

β-barrel. (The *E. coli* and *Rb. capsulatus* porins have a slightly different shape.) Third, the trimer interface (25-28% of all residues; four full strands) is tightly packed and well-ordered, resembling the hydrophobic core of soluble proteins. Inside the bilayer many aromatic residues participate in the intersubunit interactions whereas, at the level of the membrane surface and the external loops, the trimer is stabilized by intersubunit hydrogen bonds. Additional stabilizing factors include bound calcium ions (at the trimer interface and in the pore-constricting loop of *Rb. capsulatus* porin (Weiss and Schulz, 1992)) as well as interactions of the external side of the porin with the lipopolysaccharide (a polar lipid present in the outer leaflet of the outer membrane of gram-negative bacteria). A general conclusion derived from the analysis of porin structures is that a hydrophilic primary structure can fold

into a highly stable 3D membrane protein structure. While the lipid-exposed faces of porins are hydrophobic, the key players in the structural stability of porins are hydrogen-bonds, many of which are located in a water-poor environment, and stereochemically optimized van der Waals interactions between the monomers.

B. Bacterial Reaction Centers

The bacterial reaction centers were the first membrane proteins whose structures were solved at a high resolution in 1980s. Consequently, they have been a major focus in discussions of membrane protein structure (Rees et al., 1989a,b). In what follows the recently published refined structures at 2.3 Å (*Rhodopseudomonas viridis*) and 2.65Å (*Rhodobacter sphaeroides*) (Ermler et al., 1994; Deisenhofer et al., 1995) are discussed briefly.

The core of the photoreaction centers from *R. viridis* and *R. sphaeroides* comprises three subunits, designated L, M, and H. The photosynthetic cofactors are located in the L and M subunits, each of which has five transmembrane helices (named A, B, C, D, and E). The L and M polypeptides are homologous to each other (26% identity) and there is a local twofold symmetry in the structure. Deletions and insertions in these subunits have occured only outside the transmembrane segments. The H polypeptide has a single transmembrane helix, and most of its mass associates with the complex of L and M on the cytoplasmic side. A tetraheme cytochrome is bound to the periplasmic side of the *R. viridis* reaction center.

The length of the transmembrane helices in the *R. viridis* reaction center varies between 23 and 27 residues. Although the helix ends are polar, especially in helices B of both L and M subunits, there is always a 30 Å wide region that is highly nonpolar and faces the hydrophobic core of the membrane. The hydrophobic zone is flanked by aromatic residues, resembling the pattern of interfacial aromatic residues observed in porins. Glycines and prolines are found at or near the ends of the helices; these are often conserved (Michel et al., 1986). The C helices are curved owing to the presence of a proline in a position four residues after the kink, which leaves the carbonyl oxygens of the residues at the kink without a regular hydrogen bonding partner (Deisenhofer et al., 1995). However, a proline does not always cause a bend in a helix: although there is an additional proline in helix C of the L subunit, in this case the helix compensates locally for the missing hydrogen-bonds by forming a short 3_{10} helix. A similar arrangement occurs in the equivalent helix of the M polypeptide, despite the fact that this helix does not have an additional proline. In the single transmembrane helix of the H subunit a stretch of π-helix is observed.

Subunits L and M have a large interaction surface that involves numerous van der Waals interactions between the subunits and also pigment-protein interactions. The transmembrane domain of the L-M complex is the most highly ordered part of the structure, reflecting the rigid nature of transmembrane helices. The majority of the interhelical and intersubunit polar interactions occur outside the hydrophobic core of the membrane, mostly on the cytoplasmic side. The polar ends of helices

LB and MB are involved in many of these, as are structural water molecules located on the subunit interfaces. These water molecules typically form hydrogen bonds to more than one subunit. Inside the bilayer, there are eight interhelical hydrogen bonds within the L and M subunits and one intersubunit hydrogen bond. A bound iron ion which has ligands from the helices of both subunits also stabilizes the L-M complex. At least one water molecule is located in the membrane interior and participates in an intersubunit hydrogen bond network between L and M. Five waters are seen in the transmembrane part of the L-M complex.

C. Light-Harvesting Complexes

Another example of helical multisubunit membrane proteins is the nonameric light-harvesting complex from *Rhodopseudomonas acidophila* (McDermott et al., 1995) whose function is to capture light quanta and funnel them via another, related light-harvesting complex into a reaction center which is similar to those discussed above. The structure contains two concentric rings of nine plus nine short subunits (α-chains (53 residues) in the inner and β-chains (41 residues) in the outer ring) with one transmembrane helix in each chain (see Figure 7 of chapter 6 in this volume). In addition, the structure contains chlorophylls and carotenoids. The 27 chlorophylls are organized into two groups: 18 chlorophylls with histidine as the axial ligand are situated between the two helical rings so that they are alternatingly ligated by α-subunit and β-subunit transmembrane helices; the planes of the chlorin rings are perpendicular to the membrane plane. The second type of chlorophylls have their chlorin planes parallel to the membrane plane and alternate with the β-chains to form the outer ring of the nonamer. The axial ligand of the parallel chlorins are formyl oxygens of α-chain N-terminal formyl-methionines, which are at the beginning of an amphipathic 3_{10}-helix running approximately parallel to the membrane surface. The N-terminus penetrates some 9 Å into the membrane. Both the α- and β-subunit transmembrane helices are 26 residues long and their axes are nearly parallel to the membrane normal. A turn caused by a proline begins and ends the helix in the α- and β-chains, respectively. An intrahelix ion-pair is seen between Asp-17 and Arg-20 of the β-subunit.

A distinct feature of the bacterial light-harvesting complex structure is that there are no lateral contacts between the transmembrane helices; the subunits interact only through their short N-terminal domains which lie on the membrane surface. Inside the bilayer, the helices are bound together by the pigments. The light-harvesting complex of plants, although performing a function similar to that of the bacterial complex discussed above, presents a totally different structure (Kuhlbrandt et al., 1994) (Figure 9). The plant light-harvesting complex is a homotrimer. Each monomer has three relatively polar transmembrane helices, two of which are tilted 32° relative to the membrane normal and form a left-handed supercoil. The third, separate helix is less tilted and short (31 Å). In contrast, the supercoiled helices are long (43 Å and 51Å) and have hydrophilic ends; both begin with a
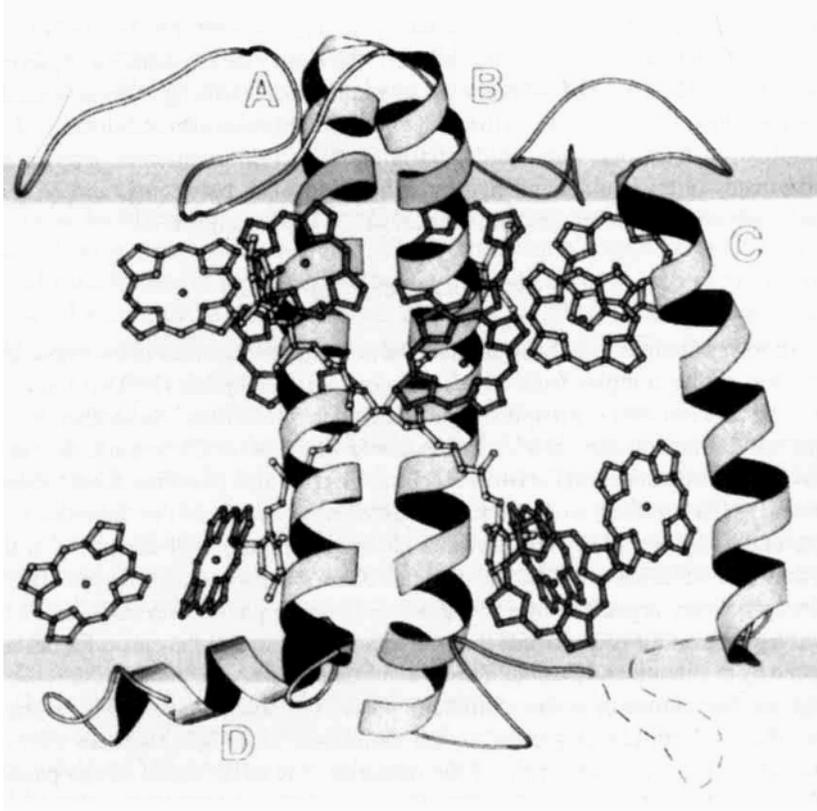
Figure 9. The structure of the plant light-harvesting complex as determined by electron crystallography of 2D crystals (Kuhlbrandt et al., 1994). The structure of a monomer seen from the membrane plane is shown; in reality the protein is a homotrimer both in the membrane and in detergent-solubilized state. The structure comprises three transmembrane helices (A, B, and C), a short interfacial amphipathic helix D, parts of the loops, 12 chlorophylls, and two carotenoids. The estimated location of the membrane surface is indicated by the two horizontal lines. The long and tilted helices A and B pack closely together to form a left-handed supercoil in which the distance between the helix centers is 8.5 Å. The interhelical contact involves about two turns of both helices in the middle of the bilayer. Interestingly, two intramembrane Glu-Arg ion pairs contribute to the helix-helix interaction. All helices including the short helix D provide ligands to the chlorophylls. Courtesy of Dr. Kuhl Brandt. Preprinted with permission from Nature 367, 614-621. Copyright 1994 Macmillan Magazines Limited.

proline and one ends with a glycine. The helices are homologous for most of their sequence and there is a twofold local symmetry in the structure. All three trans-membrane helices and a short surface helix provide ligands to at least 12 chloro-

phylls, arranged around the helices and ligated by histidines, glutamines, asparagines, glutamates, and by a peptide carbonyl; only eight ligands have been identified so far. Two membrane spanning carotenoid molecules bind to the supercoil. As with the bacterial light-harvesting complex, the pigments constitute a large fraction of the mass of the molecule.

The contact between the two supercoiled helices involves a seven residue sequence in each helix (Ser-<u>Arg</u>-Trp-Ala-Met-Leu-Gly-Ala (residues 69-76) in helix B and Gly-<u>Arg</u>-Trp-Leu-Ala-Met-Phe-Ser-Met (residues 184-191) in helix A). The helix axes are within 8.5 Å in the contact region, an indication of tight complementary packing. Interestingly, the underlined arginines in both sequences are involved in interhelical ion-pairs with glutamates which are ligands to two of the chlorophylls. The third helix has a similar intrahelix Arg-Glu pair that also participates in chlorophyll binding.

VI. CONCLUDING REMARKS

Membrane proteins are probably as old as is the living cell, and without a knowledge about their structure a highly important aspect of life remains poorly understood. The recent achievements in crystallization of membrane proteins may, however, represent a major breakthrough in the field. If the co-crystallization method of Michel and colleagues turns out to be a general one, many more membrane protein structures may soon be solved. Apart from using antibodies, the size and nature of the polar domain of a membrane protein (which appear to be critical for the crystallization process) can be manipulated by fusing the membrane protein with a suitable soluble protein. Advances may also take place in detergent chemistry, as only relatively few detergents have so far been tested by membrane protein researchers. In the near future, we may expect to know significantly more about such questions as how diverse are the folds found in membrane proteins? Do all the transmembrane parts really belong to either exclusively α-helical or exclusively β-sheet containing types? How are ions and water translocated specifically across membranes? What is the structural basis of energy-coupling in mitochondria or signal transduction in neurons?

VII. SUMMARY

Integral membrane proteins are soluble only in the presence of a detergent which binds to the nonpolar lipid-exposed part of the protein. The presence of detergent molecules has made crystallization of membrane proteins difficult: only 13 high-resolution structures are currently known. The published structures fall into three categories: α-helical membrane proteins (bacteriorhodopsin, two bacterial reaction centers, two cytochrome oxidases, and two light-harvesting complexes), β-barrel proteins (four 16-strand porins and one 18-strand porin), and a monotopic, non-

transmembrane membrane protein (prostaglandin synthase). Owing to the absence of water in the hydrophobic core of the lipid bilayer, the transmembrane secondary structure elements in both the α-helical and β-barrel proteins are generally much more stable than secondary structures in soluble proteins or in the extramembrane domains of membrane proteins. In both the bacterial reaction centers and porins the most highly ordered parts reside in the intrabilayer domain. In determining the 3D structure of a membrane protein the rate-limiting step is crystallization. Two basic methods have been used: the atomic resolution structures of bacteriorhodopsin and a plant light-harvesting complex have been solved using electron crystallography of 2D crystals while most structures have been determined with X-ray crystallography of 3D crystals. A recent method developed for the crystallization of cytochrome c oxidase makes use of an engineered monoclonal antibody fragment as a crystallization reagent. Earlier experience suggests that in most cases the formation of a membrane protein crystal lattice is critically dependent on polar interactions between the protein molecules. The rationale in making a protein-antibody -complex is to increase the polar domain of the oxidase, forcing the enzyme to co-crystallize with the help of a highly crystallizable antibody. The structure of cytochrome c oxidase from $P.$ $denitrificans$, solved using the co-crystallization method, comprises 22 transmembrane helices, two hemes, three copper atoms, and a phospholipid molecule in four integral membrane protein subunits. The structure of the oxidase as well as those of a few other membrane proteins are discussed. In addition, the new oxidase structure offers the opportunity to asses the accuracy in predicting the locations of transmembrane helices. The predictive power of two recently developed algorithms is tested against the X-ray structure. It is concluded that in the case of cytochrome oxidase the more sophisticated prediction methods do not perform significantly better than traditional hydropathy analysis.

ACKNOWLEDGMENTS

NOTE ADDED IN PROOF

By early 1997, several developments have taken place in the field. While the successful application of the antibody fragment technology to other membrane proteins has not yet been reported, a new crystallization method employing lipidic cubic phases has been used in the crystallization of bacteriorhodopsin (landau and Rosenbusch, Proc. Natl. Acad. Sci. USA 93, 14532-14535, 1996; de Kruijiff, Nature 386, 129-130, 1997. A 3D structure of bacteriohodopsin is being solved using the crystals grown in cubic phase lipids. It will be of interest to compare the

X-ray structure to the refined electron crystallographic structure of bacteriorhodopsin (Grigorieff et al., J. Mol. Biol. 259, 393-421, 1996).

Regarding cytochrome oxidase, the whole structure of bovine heart cytochrome *c* oxidase was published in 1996 (Tsukihara et al., Science 272, 1136-1144, 1996). In addition to the 13 dissimilar protein subunits and 28 transmembrane α-helices, the structure comprises five phosphatidyl ethanolamines and three phosphatidyl glycerols. Moreover, the enzyme has two bound cholate molecules, which is the detergent used during the purification of the enzyme. An analysis of the architecture of the oxidase structure has been published (Wallin et al., Prot. Sci. 6, 808-815, 1997); the structure appears to be consistent with many of the principles deduced earlier from the analysis of the bacterial reaction centers. A structural model of the *E. coli* cytochrome *bo* quinol oxidase, based on electron microscopy of 2D crystals, has also been published (Gohlke et al., EMBO J. 16, 1181-1188, 1997).

A number of mutations in the three largest subunits of cytochrome *c* oxidase appears to be linked to human disease: A five-residue deletion of helix III of subunit III has been reported in a patient with low oxidase acticity and myoglobinuria (Keightley et al., Nature Genetics 12, 410-416, 1996). A point mutation in the subunit III gene is associated with motochondrial encephalomyopathy, lactic acidosis and stroke-like episodes, a condition known as MELAS, (Manfredi et al., Neuromusc. Disord. 5, 391-398, 1995). Very recently, mutations or polymorphisms in cytochrome *c* oxidase genes for subunit I and II have been found in patients with late-onset Alzheimer's disease (Davis et al., Proc. Natl. Acad. Sci. USA 94, 4526-4531, 1997).

In late 1995, when the preceding chapter was written, the two presenilin loci that are linked to hereditary early-onset Alzheimer's disease had just been identified. Over the past two years, more than 30 point mutations have been found in the genes of these membrane proteins when screening early-onset Alzheimer's disease patients (Hardy, Trends Neurosci, 20, 154-159, 1997).

NOTES

1. The structures of two light-harvesting complexes, two chrochrome *c* oxidases, two porins and a prostaglandin synthase have been published in 1994-1995.

2. The structure of the dimeric transmembrane domain of glycophorin A was recently solved with nuclear magnetic resonance spectroscopy (MacKenzie et al., Science 276, 131-133, 1997). There are extensive backbone-backbone contacts which are possible because of the small size of the two glycines in the dimerization motif. In essence, the dimerization is based on optimal interhelical van der Waals interactions. There are no interhelical hydrogen bonds: instead, the threonine interacts with the preceding glycine in the same monomer.

REFERENCES

Agre, P., Brown, D., & Nielsen, S. (1995). Aquaporin water channels: Unanswered questions and unresolved controversies. Curr. Opin. Cell Biol. 7, 472-483.

Akabas, M. H., & Karlin, A. (1995). Identification of acetylcholine receptor channel-lining residues in the M1 segment of the α-subunit. Biochemistry 34, 12496-12500.

Altenbach, C., Greenhalgh, D. A., Khorana, H. G., & Hubbell, W. L. (1994). A collision gradient method to determine the immersion depth of nitroxides in lipid bilayers: Application to spin-labeled mutants of bacteriorhodopsin. Proc. Natl. Acad. Sci. USA 91, 1667-1671.

Altenbach, C., Marti, T., Khorana, H. G., & Hubbel, W. L. (1990). Transmembrane protein structure: Spin labeling of bacteriorhodopsin mutants. Science 248, 1088-1092.

Arkin, I. T., Adams, P. D., MacKenzie, K. R., Lemmon, M. A., Brunger, A. T., & Engelman, D. M. (1994). Structural organization of the pentameric transmembrane α-helices of phospholamban, a cardiac ion channel. EMBO J. 13, 4757-4764.

Arrondo, J. L. R., Castresana, J., Valpuesta, J. M., & Goni, F. M. (1994). Structure and thermal denaturation of crystalline and noncrystalline cytochrome oxidase as studied by infrared spectroscopy. Biochemistry 33, 11650-11655.

Arrondo, J. L. R., Muga, A., Castresana, J., & Goni, F. M. (1993). Quantitative studies of the structure of proteins in solution by Fourier-transform infrared spectroscopy. Prog. Biophys. Mol. Biol. 59, 23-56.

Ashcroft, F. M., & Röper, J. (1993). Transporters, channels and human disease. Curr. Opin. Cell Biol. 5, 677-683.

Babcock, G. T., & Wikström, M. (1992). Oxygen activation and the conservation of energy in cell respiration. Nature 356, 301-309.

Baldwin, J. M. (1993). The probable arrangement of the helices in G protein-coupled receptors. EMBO J. 12, 1693-1703.

Barinaga, M. (1995). Missing Alzheimer's gene found. Science 269, 917-918.

Booth, P. J., Flitsch, S. L., Stern, L. J., Greenhalgh, D. A., Kim, P. S., & Khorana, H. G. (1995). Intermediates in the folding of the membrane protein bacteriorhodopsin. Nature Struct. Biol.2, 139-143.

Brown, S., Moody, A. J., Mitchell, R., & Rich, P. (1993). Binuclear centre structure of terminal protonmotive oxidases. FEBS Lett. 316, 216-223.

Buchanan, S. K., Fritsch, G., Ermler, U., & Michel, H. (1993). New crystal form of the photosynthetic reaction centre from *Rhodobacter sphaeroides* of improved diffraction quality. J. Mol. Biol. 230, 1311-1314.

Castresana, J., Lubben, M., Saraste, M., & Higgins, D. G. (1994). Evolution of cytochrome oxidase, an enzyme older than atmospheric oxygen. EMBO J. 13, 2516-2525.

Cosson, B., & Bonifacino, J. S. (1992). Role of transmembrane domain interactions in the assembly of class II MHC molecules. Science 258, 659-662.

Cowan, S. W., & Rosenbusch, J. P. (1994). Folding pattern diversity of integral membrane proteins. Science 264, 914-916.

Cowan, S. W., Schirmer, T., Rummel, G., Steiert, M., Ghosh, R., Pauptit, R. A., Jansonius, J. N., & Rosenbusch, J. P. (1992). Crystal structures explain functional properties of two *E.coli* porins. Nature 358, 727-733.

Cyrklaff, M., Auer, M., Kuhlbrandt, W., & Scarborough, G. A. (1995). 2-D structure of the *Neurospora crassa* plasma membrane ATPase as determined by electron microscopy. EMBO J. 14, 1854-1857.

De Grip, W. J. (1982). Thermal stability of rhodopsin and opsin in some novel detergents. Methods Enzymol. 81, 256-265.

Deisenhofer, J., Epp, O., Sinning, I., & Michel, H. (1995). Crystallographic refinement at 2.3 Å resolution and refined model of the photosynthetic reaction centre from *Rhodopseudomonas viridis*. J. Mol. Biol. 246, 429-457.

Donnelly, D., Overington, J. P., Ruffle, S. V., Nugent, J. H. A., & Blundell, T. L. (1993). Modeling a-helical transmembrane domains: The calculation and use of substitution tables for lipid facing residues. Prot. Sci. 2, 55-70.

Dunten, R. L., Sahin-Toth, M., & Kaback, H. R. (1993). Cysteine scanning mutagenesis of putative helix XI in the lactose permease of *Escherichia coli*. Biochemistry 32, 12644-12650.

Earnest, T. N., Herzfeld, J., & Rothschild, K. J. (1990). Polarized Fourier transform infrared spectroscopy of bacteriorhodopsin. Transmembrane alpha helices are resistant to hydrogen/deuterium exchange. Biophys. J. 58, 1539-1546.

Echabe, I., Haltia, T., Freire, E., Goni, F. M., & Arrondo, J. L. R. (1995). Subunit III of cytochrome *c* oxidase influences the conformation of subunits I and II: An infrared study. Biochemistry 34, *13565-13569*.

Eisele, J.-L., & Rosenbusch, J.P. (1990). *In vitro* folding and oligomerization of a membrane protein. J. Biol. Chem. 265, 10217-10220.

Engelman, D. M., Steiz, T. A., & Goldman, A. (1986). Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Ann. Rev. Biophys. Biophys. Chem. 15, 321-353.

Ermler, U., Fritzsch, G., Buchanan, S. K., & Michel, H. (1994). Structure of the photosynthetic reaction centre from *Rhodobacter sphaeroides* at 2.65 Å resolution: Cofactors and protein-cofactor interactions. Structure 2, 925-936.

Fann, Y. C., Ahmed, I., Blackburn, N. J., Boswell, J. S., Verhovskaya, M. L., Hoffman, B. M., & Wikström. M. (1995). Structure of Cu_B in the binuclear heme-copper center of the cytochrome aa_3-type quinol oxidase from *Bacillus subtilis*: An ENDOR and EXAFS study. Biochemistry 34, 10245-10255.

Farahbakhsh, Z. T., Hideg, K., & Hubbell, W. L. (1993). Photoactivated conformational changes in rhodopsin: A Time-resolved spin label study. Science 262, 1416-1419.

Fetter, J. R., Qian, J., Shapleigh, J., Thomas, J. W., Garcia-Horsman, A., Schmidt, E., Hosler, J., Babcock, G. T., Gennis, R. B., & Ferguson-Miller, S. (1995). Possible proton relay pathways in cytochrome *c* oxidase. Proc. Natl. Acad. Sci. USA 92, 1604-1608.

Fischbarg, J., Cheung, M., Czegledy, F., Li, J., Iserovich, P., Kuang, K., Hubbard, J., Garner, M., Rosen, O. M., Golde, D. W., & Vera, J. C. (1993). Evidence that facilitative glucose transportes may fold as β-Barrels. Proc. Natl. Acad. Sci. USA 90, 11658-11662.

Frey, T. G., Chan, S. H. P., & Schatz, G. (1978). Structure and orientation of cytochrome c oxidase in crystalline membranes: Studies by electron microscopy and by labeling with subunit specific antibodies. J. Biol. Chem. 253, 4389-4395.

Frey, T. G., & Murray, J. M. (1994). Electron microscopy of cytochrome *c* oxidase crystals. Monomer-dimer relationship and cytochrome *c* binding site. J. Mol. Biol. 237, 275-297.

Garavito, R. M., & Picot, D. (1990). The art of crystallizing membrane proteins. Methods: A Companion to Methods in Enzymology 1, 57-69.

Garavito, R. M., Picot, D., & Loll, P. J. (1995). The structure of the integral membrane enzyme prostaglandin H_2 synthase-1 at 3.1 Å resolution. Prot. Eng. 8 (suppl), 41.

Garcia-Horsman, J. A., Puustinen, A., Gennis, R. B., & Wikström, M. (1995). Proton transfer in cytochrome bo_3 ubiquinol oxidase of *Escherichia coli*: Second-site mutations in subunit I that restore proton pumping in the mutant Asp135 → Asn. Biochemistry 34, 4428-4433.

Gillez-Gonzalez, M. A., Engelman, D. M., & Khorana, H. G. (1991). Structure-function studies of bacteriorhodopsin XV. Effects of deletions in loops B-C and E-F on bacteriorhodopsin chromophore and structure. J. Biol. Chem. 266, 8545-8550.

Görne-Tschelnokow, U., Strecker, A., Kaduk, C., Naumann, D., & Hucho, F. (1994). The transmembrane domains of the nicotinic acetylcholine receptor contain α-helical and β *structures. EMBO J. 13, 338-341.*

Ghaim, J. B., Greiner, D. P., Meares, C. F., & Gennis, R. B. (1995). Proximity mapping the surface of membrane protein using an artificial protease: Demonstration that the quinone-binding domain of subunit I is near the N-terminal region of subunit II of cytochrome *bd*. Biochemistry 34, 11311-11315.

Grisshammer, R., & Tate, C. G. (1995). Overexpression of integral membrane proteins for structural studies. Q. Rev. Biophys. 28, 315-422.

Haltia, T., & Freire, E. (1995). Forces and factors that contribute to the structural stability of membrane proteins. Biochim. Biophys. Acta 1228, 1-27.

Haltia, T., & Wikström, M. (1992). Cytochrome oxidase: Notes on structure and mechanism. In: Molecular Mechanisms in Bioenergetics (Ernster, L., Ed.), pp. 217-239, Elsevier, Amsterdam.

Haltia, T., Finel, M., Harms, N., Nakari, T., Raitio, M., Wikström, M., & Saraste, M. (1989). Deletion of the gene for subunit III leads to defective assembly of bacterial cytochrome oxidase. EMBO J. 8, 3571-3579.

Haltia, T., Semo, N., Arrondo, J. L. R., Goni, F. M., & Freire, E. (1994). Structural and thermodynamic stability of cytochrome c oxidase from *Paracoccus denitrificans*. Biochemistry 33, 9731-9740.

Hansen, O. K., Pompejus, M., & Fritz, H.-J. (1994). Two-chain bacteriorhodopsin synthesized by *Schizosaccharomyces pombe*. Biol. Chem. Hoppe-Seyler 375, 715-719.

Haris, P. I., Chapman, D., & Benga, G. (1995). A Fourier-transform infrared spectroscopic investigation of the hydrogen-deuterium exchange and secondary structure of the 28-kDa channel-forming integral membrane protein. Eur. J. Biochem. 233, 659-664.

Hästbacka, J., de la Chapelle, A., Mahtani, M.M., Clines, G., Reeve-Daly, M.P., Daly, M., Hamilton B.A., Kusumi, K., Trivedi, B., Weaver, A., Coloma, A., Lovett, M., Buckler, A., Kaitila, I., & Lander, E. (1994). The diastrophic dysplasia gene encodes a novel sulfate transporter: Positional cloning by fine-structure linkage disequilibrium mapping. Cell 78, 1073-1087.

Havelka, W. A., Henderson, R., Heymann, J. A. W., & Oesterhelt, D. (1993). Projection structure of halorhodopsin from *Halobacterium halobium* at 6 Å resolution obtained by electron cryo-microscopy. J. Mol. Biol. 234, 837-846.

Hebert, H., Schmidt-Krey, I., & Morgenstern, R. (1995). The projection structure of microsomal glutathione transferase. EMBO J. 14, 3864-3869.

Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E., & Downing, K. H. (1990). Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. J. Mol. Biol. 213, 899-929.

Hill, B. C., Cook, K., & Robinson, N. C. (1988). Subunit dissociation and protein unfolding in the bovine heart cytochrome oxidase complex induced by guanidine hydrochloride. Biochemistry 27, 4741-4747.

Hobe, S., Prytulla, S., Kuhlbrandt, W., & Paulsen, H. (1994). Trimerization and crystallization of reconstituted light-harvesting chlorophyll *a/b* complex. EMBO J. 13, 3423-3429.

Hubbell, W. L., & Altenbach, C. (1994). Investigation of structure and dynamics in membrane proteins using site-directed spin labeling. Curr. Op. Struct. Biol. 4, 566-573.

Huang, K.-S., Bayley, H., Liao, M.-J., London, E., & Khorana, H. G. (1981). Refolding of an integral membrane protein. J. Biol. Chem. 256, 3802-3809.

Iwata, S., Ostermeier, C., Ludwig, B., & Michel, H. (1995). Structure at 2.8 Å resolution of cytochrome c oxidase from *Paracoccus denitrificans*. Nature 376, 660-669.

Jap, B. K., & Li, H. (1995). Structure of the osmo-regulated H_2O-channel, AQP-CHIP, in projection at 3.5 Å resolution. J. Mol. Biol. 251, 413-420.

Jennings, M. L. (1989). Topography of membrane proteins. Annu. Rev. Biochem. 58, 999-1027.

Johns, D. R., & Neufeld, M. J. (1993). Cytochrome c oxidase mutations in Leber hereditary optic neuropathy. Biochem. Biophys. Res. Commun. 196, 810-815.

Jones, D. T., Taylor, W. R., & Thornton, J. M. (1994a). A mutation data matrix for transmembrane proteins. FEBS Lett. 339, 269-275.

Jones, D. T., Taylor, W. R., & Thornton, J. M. (1994b). A model recognition approach to the prediction of all-helical membrane protein structure and topology. Biochemistry 33, 3038-3049.

Jung, K., Jung, H., & Kaback, H. R. (1994). Dynamics of lactose permease of *Escherichia coli* determined by site-directed fluorescence labeling. Biochemistry 33, 3980-3985.

Jung, K., Jung, H., Wu, J., Prive, G., & Kaback, H. R. (1993). Use of site-directed fluorescence labeling to study proximity relationships in the lactose permease of *Escherichia coli*. Biochemistry 32, 12273-12278.

Keller, T. A., Ferenci, T., Prilipov, A., & Rosenbusch, J. P. (1994). Crystallization of monodisperse maltoporin from wildtype and mutant strains of various *Enterobacteriaceae*. Biochem. Biophys. Res. Commun. 199, 767-771.

Kleymann, G., Ostermeier, C., Ludwig, B., Skerra, A., & Michel, H. (1995). Engineered F_v fragments as a tool for the one-step purification of integral multisubunit membrane proteins. Bio/Technology 13, 155-160.

Kraulis, P. J. (1991). MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallog. 24, 946-950.

Krauss, N., Hinrichs, W., Witt, I., Fromme, P., Pritzkow, W., Dauter, Z., Betzel, C., Wilson, K. S., Witt, H. T., & Saenger, W. (1993). Three-dimensional structure of system I of photosynthesis at 6 Å resolution. Nature 361, 326-331.

Krebs, M., & Khorana, H.G. (1993). Mechanism of ligh-dependent proton translocation by bacteriorhodopsin. J. Bact. 175, 1555-1560.

Kreimer, D. I., Szosenfogel, R., Goldfarb, D., Silman, I., & Weiner, L. (1994). Two-state transition between molten globule and unfolded states of acetylcholinesterase as monitored by electron paramagnetic resonance spectroscopy. Proc. Natl. Acad. Sci. USA 91, 12145-12149.

Kreusch, A., & Schulz, G. E. (1994). Refined structure of the porin from *Rhodopseudomonas blastica*. Comparison with the porin from *Rhodobacter capsulatus*. J. Mol. Biol. 243, 891-905.

Kreusch, A., Weiss, M. S., Welte, W., Weckesser, J., & Schulz, G. E. (1991). Crystals of an integral membrane protein diffracting to 1.8 Å resolution. J. Mol. Biol. 217, 9-10.

Kuhlbrandt, W. (1988). Three-dimensional crystallization of membrane proteins. Q. Rev. Biophys. 21, 429-477.

Kuhlbrandt, W. (1992). Two-dimensional crystallization of membrane proteins. Q. Rev. Biophys. 25, 1-49.

Kuhlbrandt, W., Wang, D. N., & Fujiyoshi, Y. (1994). Atomic model of plant light-harvesting complex by electron crystallography. Nature 367, 614-621.

Kuusinen, A., Arvola, M., Oker-Blom, C., & Keinänen, K. (1995). Purification of recombinant GluR-D glutamate receptor produced in Sf21 insect cells. Eur. J. Biochem. 233, 720-726.

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105-132.

Landolt-Marticorena, C., Williams, K. A., Deber, C. M., & Reithmeier, R. A. F. (1993). Non-random distribution of amino acids in the transmembrane segments of human type I single span membrane proteins. J. Mol. Biol. 229, 602-608.

Lee, J. W., Chan, M., Law, T. V., Kwon, H. J., & Jap, B. K. (1995). Preliminary cryocrystallographic study of the mitochondrial cytochrome bc_1 complex: Improved crystallization and flash-cooling of a large membrane protein. J. Mol. Biol. 252, 15-19.

Le Maire, M., Kwee, S., Andersen J. P., & Moller, J. V. (1983). Mode of interaction of polyoxyethyleneglycol detergents with membrane proteins. Eur. J. Biochem. 129, 525-532.

Lemmon, M. A., Treutlein, H. R., Adams, P. D., Brunger, A. T., & Engelman, D. M. (1994). A dimerization motif for transmembrane α-helices. Nature Struct. Biol. 1, 157-163.

Lemmon, M. A., Flanagan, J. M., Hunt, J. F., Adair, R. B., Bormann, B. J., Dempsey, C. E., & Engelman, D. M. (1992). Glycophorin a dimerization is driven by specific interactions between transmembrane α-helices. J. Biol. Chem. 267, 7683-7689.

Lemmon, M. A., & Engelman, D. M. (1994). Specificity and promiscuity in membrane helix interactions. Q. Rev. Biophys. 27, 157-218.

Lerner, T. J., and the International Batten Disease Consortium. (1995). Isolation of a novel gene underlying Batten disease, CLN3. Cell 82, 949-959.

Loll, P. J., Picot, D., & Garavito, R. M. (1995). The structural basis of aspirin activity inferred from the crystal structure of inactivated prostaglandin H_2 synthase. Nature Struct. Biol. 2, 637-643.

Lynch, B. A., & Koshland, D. E. (1991). Disulfide cross-linking studies of the transmembrane regions of the aspartate sensory receptor of *Escherichia coli*. Proc. Natl. Acad. USA 88, 10402-10406.

Ma, J., Lemieux, L., & Gennis, R. B. (1993). Genetic fusion of subunits I, II, and III of the cytochrome *bo* ubiquinol oxidase from *Escherichia coli* results in a fully assembled and active enzyme. Biochemistry 32, 7692-7697.

McDermott, G., Prince, S. M., Freer, A. A., Hawthornthwaite-Lawless, A. M., Papiz, M. Z., Codgell, R. J., & Isaacs, N. W. (1995). Crystal structure of an integral membrane light-harvesting complex from photosynthetic bacteria. Nature 374, 517-521.

Michel, H. (1982). Three-dimensional crystals of a membrane protein complex. The photosynthetic reaction centre from *Rhodopseudomonas viridis*. J. Mol. Biol. 158, 567-572.

Michel, H. (1983). Crystallization of membrane proteins. Trends Biochem. Sci. 8, 56-59.

Michel, H. (1990). General and practical aspects of membrane protein crystallization. In: Crystallization of Membrane Proteins (Michel, H., ed.), pp. 73-88, CRC Press, Boca Raton, FL.

Michel, H., Weyer, K. A., Gruenberg, H., Dunger, I., Oesterhelt, D., & Lottspeich, F. (1986). The 'light' and 'medium' subunits of the photosynthetic reaction centre from *Rhodopseudomonas viridis*: Isolation of the genes, nucleotide and amino acid sequence. EMBO J. 5, 1149-1158.

Mitchell, D. M., & Gennis, R. B. (1995). Rapid purification of wildtype and mutant cytochrome c oxidases from Rhodobacter sphaeroides by Ni^{2+}-NTA affinity chromatography. FEBS Lett. 368, 148-150.

Mitra, A. K., van Hoek, A. N., Wiener, M. C., Verkman, A. S., & Yeager, M. (1995). The CHIP28 water channel visualized in ice by electron crystallography. Nature Struct. Biol. 2, 726-729.

Mitra, B., Gerlt, J. A., Babbit, P. C., Koo, C. W., Kenyon, G. L., Joseph, D., & Petsko, G. A. (1993). A novel structural basis for membrane association of a protein: Construction of a chimeric soluble mutant of (S)-mandelate dehydrogenase from *Pseudomonas putida*. Biochemistry 32, 12959-12967.

Moller, J. V., & Le Maire, M. (1993). Detergent binding as a measure of hydrophobic surface area of membrane proteins. J. Biol. Chem. 268, 18659-18672.

Morgan, J. E., Verkhovsky, M. I., & Wikström, M. (1994). The histidine cycle: A new model for proton translocation in the respiratory heme-copper oxidases. J. Bioenerg. Biomembr. 26, 599-608.

Moriyama, R., Ideguchi, H., Lombardo, C. R., van Dort, H. M., & Low, P.S. (1992). Structural and functional characterization of band 3 from Southeast Asian ovalocytes. J. Biol. Chem. 267, 25792-25797.

Naumann, D., Schulz, C., Görne-Tschelnokow, U., & Hucho, F. (1993). Secondary structure and temperature behavior of the acetylcholine receptor by Fourier transform infrared spectroscopy. Biochemistry 32, 3162-3168.

Nussberger, S., Dörr, K., Wang, D. N., & Kuhlbrandt, W. (1993). Lipid-protein interactions in crystals of plant light-harvesting complex. J. Mol. Biol. 234, 347-356.

Oikawa, K., Lieberman, D. M., & Rethmeier, R. A. F. (1985). Conformation and stability of the anion transport protein of human erythrocyte membranes. Biochemistry 24, 2843-2848.

Ostermeier, C., Essen, L.-O., & Michel, H. (1995a). Crystals of an antibody F_v fragment against an integral membrane protein diffracting to 1.28 Å resolution. Proteins 21, 74-77.

Ostermeier, C., Iwata, S., Ludwig, B., & Michel, H. (1995b). F_v fragment-mediated crystallization of the membrane protein bacterial cytochrome *c* oxidase. Nature Struct. Biol. 2, 842-846.

Pakula, A. A., & Simon, M. I. (1992). Determination of transmembrane protein structure by disulfide cross-linking: The *Escherichia coli* tar receptor. Proc. Natl. Acad. Sci. USA 89, 4144-4148.

Papiz, M. Z., Hawthornthwaite, A. M., Cogdell, R. J., Woolley, K. J., Wightman, P. A., Ferguson, L. A., & Lindsay, J. G. (1989). Crystallization and characterization of two crystal forms of the B800-850 light-harvesting complex from *Rhodopseudomonas acidophila* strain 10050. J. Mol. Biol. 209, 833-835.

Pauptit, R. A., Zhang, H., Rummel, G., Schirmer, T., Jansonius, J., & Rosenbusch, J. P. (1991). Trigonal crystals of porin from *Escherichia coli*. J. Mol. Biol. 218, 505-507.

Park, K., Perczel, A., & Fasman, G. D. (1992). Differentiation between transmembrane helices and peripheral helices by the deconvolution of circular dichroism spectra of membrane proteins. Prot. Sci. 1, 1032-1049.

Persson, B., & Argos, P. (1994). Prediction of transmembrane segments in proteins utilising multiple sequence alignments. J. Mol. Biol. 237, 182-192.

Picot, D., & Garavito, R. M. (1994). Prostaglandin H synthase: Implications for membrane structure. FEBS Lett. 346, 21-25.

Picot, D., Loll, P. J., & Garavito, R. M. (1994). The X-ray crystal structure of the membrane protein prostaglandin H_2 synthase-1. Nature 367, 243-249.

Popot, J.-L. (1993). Integral membrane protein structure: Transmembrane α-helices as autonomous folding domains. Curr. Op. Struct. Biol. 3, 532-540.

Popot, J.-L., & Engelman, D. M. (1990). Membrane protein folding and oligomerization: The two-stage model. Biochemistry 29, 4031-4037.

Popot, J.-L., Gerchman, S. E., & Engelman, D. M. (1987). Refolding of bacteriorhodopsin in lipid bilayers: A thermodynamically controlled two-stage process. J. Mol. Biol. 198, 655-676.

Popot, J.-L., & Saraste, M. (1995). Engineering membrane proteins. Curr. Opin. Biotech. 6, 394-402.

Raitio, M., Jalli, T., & Saraste, M. (1987). Isolation and analysis of the genes for cytochrome *c* oxidase in *Paracoccus denitrificans*. EMBO J. 6, 2825-2833.

Raitio, M., Pispa, J. M., Metso, T., & Saraste, M. (1990). Are there isoenzymes of cytochrome *c* oxidase in *Paracoccus denitrificans*? FEBS Lett. 261, 431-435.

Rees, D. C., DeAntonio, L., & Eisenberg, D. (1989a). Hydrophobic organization of membrane proteins. Science 245, 510-513.

Rees, D. C., Komiya, H., Yeates, T. O., Allen J. P., & Feher, G. (1989b). The bacterial photosynthetic reaction center as a model for membane proteins. Annu. Rev. Biochem. 58, 607-633.

Reiss-Husson, F. (1992). Crystallization of membrane proteins. In: Crystallization of Nucleic Acids and Proteins. A Practical Approach. (Ducruix, A., & Giege, R., eds.),. pp. 175-193, IRL Press, Oxford.

Reithmeier, R. A. F. (1993). The erythrocyte anion transporter (Band 3). Curr. Opin. Struct. Biol. 3, 515-523.

Rizzolo, L. J., & Tanford, C. (1978). Denaturation of the tryptic fragments of the calcium(II) adenosine triphosphatase from sarcoplasmic reticulum by guanidium hydrochloride. Biochemistry 17, 4044-4048.

Rost, B., Casadio, R., Fariselli, P., & Sander, C. (1995). Transmembrane helices predicted at 95% accuracy. Prot. Sci. 4, 521-533.

Roth, M., Arnoux, B., Ducruix, A., & Reiss-Husson, F. (1991). Structure of the detergent phase and protein-detergent interactions in crystals of the wild-type (strain Y). *Rhodobacter sphaeroides* photochemical reaction center. Biochemistry 30, 9403-9413.

Roth, M., Lewitt-Bentley, A., Michel, H., Deisenhofer, J., Huber, R., & Oesterhelt, D. (1989). Detergent structure in crystals of a bacterial photosynthetic reaction centre. Nature 340, 659-662.

Sahin-Toth, M., & Kaback, H. R. (1993a). Cysteine scanning mutagenesis of putative transmembrane helices IX and X in the lactose permease of *Escherichia coli*. Prot. Sci. 2, 1024-1033.

Sahin-Toth, M., & Kaback, H. R. (1993b). Properties of interacting aspartic acid and lysine residues in the lactose permease of *Escherichia coli*. Biochemistry 32, 10027-10035.

Sahin-Toth, M., Dunten, R., & Kaback, H. R. (1995). Design of a membrane protein for site-specific proteolysis: Properties of engineered factor Xa protease sites in the lactose permease of *Escherichia coli*. Biochemistry 34, 1107-1112.

Saraste, M. (1990). Structural features of cytochrome oxidase. Q. Rev. Biophys. 23, 331-366.

Samatey, F. A., Xu, C., & Popot, J.-L. (1995). On the distribution of amino acid residues in transmembrane α-helix bundles. Proc. Natl. Acad. Sci. USA 92, 4577-4581.

Savontaus, M.-L. (1995). mtDNA mutations in Leber's hereditary optic neuropathy. Biochim. Biophys. Acta 1271, 261-263.

Schirmer, T., & Cowan, S. W. (1993). Prediction of membrane-spanning β-strands and its application to maltoporin. Prot. Sci. 2, 1361-1363.

Schirmer, T., Keller, T. A., Wang, Y.-F., & Rosenbusch, J. P. (1995). Structural basis for sugar translocation through maltoporin channels at 3.1 Å resolution. Science 267, 512-514.

Stauffer, K. A., Page, M. G. P., Hardmeyer, A., Keller, T. A., & Pauptit, R. A. (1990). Crystallization and preliminary X-ray characterization of maltoporin from *Escherichia coli*. J. Mol. Biol. 211, 297-299.

Steinhoff, H.-J., Mollaaghababa, R., Altenbach, C., Hideg, K., Krebs, M., Khorana, H. G., & Hubbell, W. L. (1994). Time-resolved detection of structural changes during the photocycle of spin-labeled bacteriorhodopsin. Science 266, 105-107.

Steinrucke, P., Steffens, G. C., Panskus, G., Buse, G., & Ludwig, B. (1987). Subunit II of cytochrome *c* oxidase from *Paracoccus denitrificans*. DNA sequence, gene expression and the protein. Eur. J. Biochem. 167, 431-439.

Sternberg, B., L'Hostis, C., Whiteway, C. A., & Watts, A. (1992). The essential role of specific *Halobacterium halobium* polar lipids in 2D array formation of bacteriorhodopsin. Biochim. Biophys. Acta 1108, 21-30.

Stoddard, B. L., Bui, J. D., & Koshland, D. E. (1992). Structure and dynamics of transmembrane signaling by the *Escherichia coli* aspartate receptor. Biochemistry 31, 11978-11983.

Sun, S., & Parthasarathy, R. (1994). Protein sequence and structure relationship ARMA spectral analysis: Application to membrane proteins. Biophys. J. 66, 2092-2106.

Surewicz, W. K., Mantsch, H. H., & Chapman, D. (1993). Determination of protein secondary structure by Fourier transform infrared Spectroscopy: A critical assesment. Biochemistry 32, 389-394.

Surrey, T., Schmid, A., & Jähnig, F. (1996). Folding and membrane insertion of the trimeric β-barrel protein ompF. Biochemistry 35, 2283-2288.

Taylor, W. R., Jones, D. T., & Green, N. M. (1994). A method for α-helical integral membrane protein fold prediction. Proteins 18, 281-294.

Teufel, M., Pompejus, M., Humbel, B., Friedrich, K., & Fritz, H.-J. (1993). Properties of bacteriorhodopsin derivatives conctructed by insertion of an exogeneous epitope into extramembrane loops. EMBO J. 12, 3399-3408.

Thomas, J. W., Puustinen, A., Alben, J. O., Gennis, R. B., & Wikström, M. (1993). Substitution of asparagine-135 in subunit I of the cytochrome *bo* ubiquinol oxidase of *Escherichia coli* eliminates proton-pumping activity. Biochemistry 32, 10923-10928.

Traxler, B., Boyd, D., & Beckwith, J. (1993). The topological analysis of integral membrane proteins. J. Membr. Biol. 132, 1-11.

Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R., & Yoshikawa, S. (1995). Structures of metal sites of oxidized bovine heart cytochrome *c* oxidase at 2.8 Å. Science 269, 1069-1074; The whole structure of the 13-subunit oxidized cytochrome *c* oxidase at 2.8Å. Science 272, 1136-1144 (1996).

Trumpower, B. L., & Gennis, R. B. (1994). Energy transduction by cytochrome complexes in mitochondrial and bacterial respiration: The enzymology of coupling electron transfer reactions to transmembrane proton translocation. Annu. Rev. Biochem. 63, 675-716.

Unwin, N. (1993). Nicotinic acetylcholine receptor at 9 Å resolution. J. Mol. Biol. 229, 1101-1124.

Unwin, N. (1995). Acetylcholine receptor channel imaged in the open state. Nature 373, 37-43.

van der Oost, J., Lappalainen, P., Musacchio, A., Lemieux, L., Rumbley, J., Gennis, R. B., Aasa, R., Pascher, T., Malmström, B. G., & Saraste, M. (1992). Restoration of a lost metal-binding site: Construction of two different copper sites into a subunit of the *E. coli* cytochrome *o* quinol oxidase complex. EMBO J. 11, 3209-3217.

Van Hoek, A. N., Wiener, M., Bicknese, S., Miercke, L., Biwersi, J., & Verkman, A. S. (1993). Secondary structure analysis of purified functional CHIP28 water channels by CD and FTIR spectroscopy. Biochemistry 32, 11847-11856.

van Iwaarden, P. R., Pastore, J. C., Konings, W. N., & Kaback, H. R. (1991). Construction of a functional lactose permease devoid of cysteine residues. Biochemistry 30, 9595-9600.

von Heijne, G. (1992). Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. J. Mol. Biol. 225, 487-494.

Walz, T., Typke, D., Smith, B. L., Agre, P., & Engel, A. (1995). Projection map of aquaporin-1 determined by electron crystallography. Nature Struct. Biol. 2, 730-732.

Wang, D. N., & Kuhlbrandt, W. (1991). High-resolution electron crystallography of light-harvesting chlorophyll *a/b*-protein complex in three different media. J. Mol. Biol. 217, 691-699.

Wang, D. N., Sarabia, V. E., Reithmeier, R. A. F., & Kuhlbrandt, W. (1994). Three-dimensional map of the dimeric membrane domain of the human erythrocyte anion exchanger, band 3. EMBO J. 13, 3230-3235.

Warne, A., Wang, D. N., & Saraste, M. (1995). Purification and two-dimensional crystallization of bacterial cytochrome oxidases. Eur. J. Biochem 234, 443-451.

Weiss, M. S., & Schulz, G. E. (1992). Structure of porin refined at 1.8 Å resolution. J. Mol. Biol. 227, 493-509.

Whitley, P., Nilsson, L., & von Heijne, G. (1993). Three-dimensional model for the membrane domain of *Escherichia coli* leader peptidase based on disulfide mapping. Biochemistry 32, 8534-8539.

White, S. (ed.) (1994). In: Membrane Protein Structure: Experimental Approaches. Oxford Univ. Press, New York.

Wikström, M., Bogachev, A., Finel, M., Morgan, J. E., Puustinen, A., Raitio, M., Verkovskaya, M., & Verkhovsky, M. I. (1994). Mechanism of proton translocation by the respiratory oxidases. The histidine cycle. Biochim. Biophys. Acta 1187, 106-111.

Williams, K. A., Glibowicka, M., Li, Z., Li, H., Khan, A., Chen, Y. M. Y., Wang, J., Marvin, D. A., & Deber, C. M. (1995). Packing of coat protein amphipathic and transmembrane helices in filamentous bacteriophage M13: Role of small residues in protein oligomerization. J. Mol. Biol. 252, 6-14.

Willmanns, M., Lappalainen, P., Kelly, M., Sauer-Eriksson, E., & Saraste, M. (1995). Crystal structure of the membrane-exposed domain from a respiratory quinol oxidase complex with an engineered dinuclear copper site. Proc. Natl. Acad. Sci. USA 92, 11955-11959.

Wimley, W. C., & White, S. H. (1993). Membrane partitioning: Distinguishing bilayer effects from the hydrophobic effect. Biochemistry 32, 6307-6312.

Witt, I., Witt, H. T., Di Fiore, D., Rögner, M., Hinrichs, W., Saenger, W., Granzin, J., Betzel, Ch., & Dauter, Z. (1988). X-ray characterization of single crystals of the reaction center I of water splitting photosynthesis. Ber. Bunsenges. Phys. Chem. 92, 1503-1506.

Wu, J., Perrin, D. M., Sigman, D. S., & Kaback, H. R. (1995). Helix packing of lactose permease in *Escherichia coli* studied by site-directed chemical cleavage. Proc. Natl. Acad. Sci. USA 92, 9186-9190.

Zulauf, M. (1990). Detergent phenomena in membrane protein crystallization. In Crystallization of Membrane Proteins (Michel, H., ed.), pp. 53-72, CRC Press, Boca Raton, FL.

This Page Intentionally Left Blank

EXPERIMENTAL DISSECTION OF PROTEIN-PROTEIN INTERACTIONS IN SOLUTION

Michael L. Doyle and Preston Hensley

I. INTRODUCTION

The description of biological processes at the molecular level is a description
of events which are almost always governed by interactions that involve pro-
teins. Proteins interact with small molecules, nucleic acids, complex sugars,
lipids, and each other. Protein-protein interactions play crucial roles in a wide
variety of processes which define cell structure and function. This, and the fact
that advances in molecular biology have made an expanding assortment of
soluble proteins readily accessible, have heightened interest in quantitative
understanding of the structural and functional aspects of protein-protein inter-
actions. The structural aspects of protein-protein interactions are being defined
by high resolution approaches such as X-ray crystallography and nuclear
magnetic resonance (NMR) (Covell et al., 1994; Padlan, 1994; Clore and
Gronenborn, 1995; Jones and Thornton, 1995). With the steady advances being
made in molecular and structural biology, the need for accessible and rigorous
biophysical methods for determining the kinetics and thermodynamics of pro-
tein-protein interactions continues to grow. The purpose of the present discus-
sion is to describe the emerging tools and strategies for making these
measurements.

   During the last decade a number of biophysical approaches have emerged (or
re-emerged) as powerful tools for the characterization of protein interactions in
solution. These include isothermal titration calorimetry (ITC) (Wiseman et al.,
1989; Freire et al., 1990), analytical gel chromatography (AGC) (Ackers, 1970;
Valdes and Ackers, 1979), surface plasmon resonance (SPR) (O'Shannessy et al.,
1993; Malmqvist and Granzow, 1994; O'Shannessy, 1994), light scattering (Arak-
awa et al., 1994; Wen et al., 1996) and analytical ultracentrifugation (AUC) due to
the availability of modern instrumentation (Harding et al., 1992; Schuster and Laue,
1994). Spectroscopic approaches also continue to play an important role (Fairman
et al., 1995; Phizicky and Fields, 1995), though these approaches depend on specific
spectroscopic properties of proteins.

   In this chapter we present a practical guide for using several of these biophysical
methods. It is anticipated that these methods will play dominant roles as core tools
for studying the structural and functional molecular properties of protein-protein
interactions. In general these methods require relatively small amounts (milligram)
of protein material, and are universally applicable to the study of proteins. Key

features of these methods are that they are capable of elucidating molecular-level details of protein-protein interactions such as: (1) self-association state (i.e., quaternary structure) over a large range of protein concentration (since protein-protein interactions are concentration-dependent it is critical to know the self-association status of a protein at concentrations which are physiologically and/or mechanistically relevant), (2) molecular shape and conformational changes in solution, and (3) the quantitative thermodynamic and kinetic characteristics of protein binding interactions.

Although biophysical methods are capable of providing molecular-level details of protein-protein interactions, the increase in resolution to the molecular level is accompanied by an increased need to validate interpretation of the data. To facilitate this process we emphasize the synergistic benefits of using multiple independent methods. Most, if not all, methods have some degree of inherent systematic error. For example, does covalent modification of the protein alter its functional properties? Is the binding signal linear over the range studied? It is frequently the case that measurements of binding equilibrium or kinetic parameters are done on the edge of detectability for a given technique. A series of control experiments are then required to assess the accuracy of the physical parameters being determined, which may be laborious, and may or may not be sufficient to rule out systematic errors. Hence, one gains a great deal of validation when independent methods yield the same value for a specific parameter (e.g., an affinity constant determined by kinetic and equilibrium methods). Finally, an accurate description of the molecular mechanism of action for protein-protein interactions usually relies on mutually exclusive information provided by multiple methods. A common example is the use of one method to measure binding affinity and a different method to identify the self-association state of the reactants and products.

II. MEASUREMENT OF AFFINITIES AND STOICHIOMETRIES

A. Practical Issues: Determination of Molecular Mass and Concentration

The first step in an analysis of a macromolecular interaction is the determination of the molecular mass of the component species. This is now routinely done using one of several mass spectrometric approaches (Carr et al., 1991; Chait and Kent, 1992; Suzdak, 1994; Williams and Carr, 1995). Electrospray mass spectrometry (ESMS) is the method of choice for high resolution mass determination, though it is less forgiving in terms of solvent composition. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) is more forgiving in terms of solvent and more easily accommodates glycoproteins. The precision and accuracy of both methods is good, even for large proteins. For instance, the mass of a monoclonal antibody (mAb) against respiratory syncytial virus (RSV) was deter-

mined to be 149,265 ± 100 Da by ESMS and 149,519 ± 300 Da by MALDI-MS, compared to a predicted value, based on gene sequence and carbohydrate mass, of 149,154 Da (Roberts et al., 1995) (see Figure 1).

Mass determination is a critical measurement for at least two reasons. First, for non-covalently modified proteins, it allows the confirmation of the fidelity of gene expression. This should never be taken for granted. Second, for extensively glycosylated proteins it gives the total monomer molecular weight. This may be used to determine the mass due to carbohydrate and is required for the prediction of the partial specific volume term, $\bar{v}$, used in AUC (see below and Laue et al., 1992).

The accuracy in determining binding affinities and stoichiometries is limited to the accuracy with which the reactant concentrations are known. Several methods can be used to determine protein concentrations (see Stoscheck, 1990 for review). One of the easiest and most accurate methods for determining protein concentrations is from UV-visible absorbance measurements. Extinction coefficients at 280 nm can be calculated to 5% accuracy based on the amino acid composition of the protein (Gill and von Hippel, 1989; Pace et al., 1995). Absorbance at 205 or 210 nm can also yield protein concentrations to about 5% accuracy and does not require
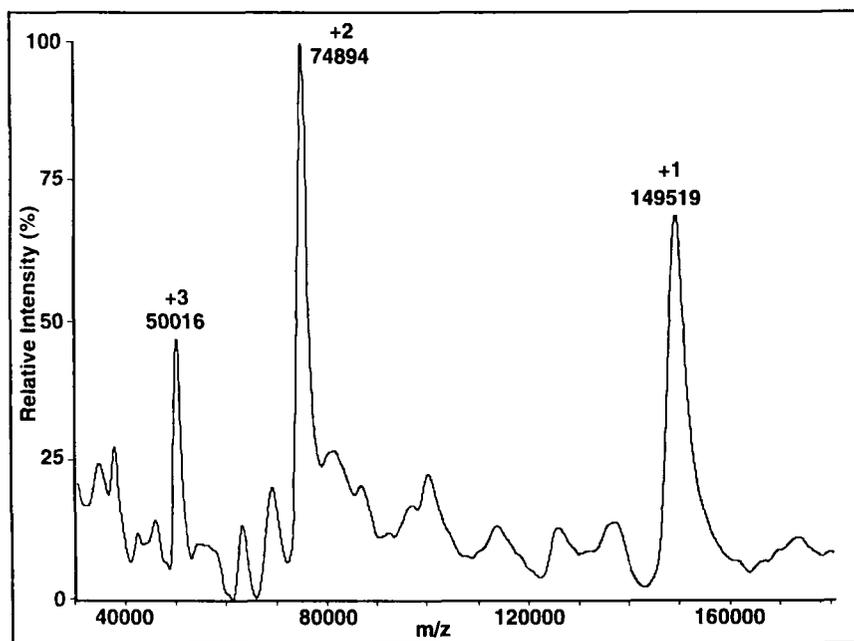
Figure 1.   MALDI mass spectrum of RSV mAb (unpublished observations; see Roberts et al., 1995).

knowledge of the amino acid composition (Stoscheck, 1990). Amino acid analysis is another accurate procedure, and is capable of yielding concentrations which are accurate to within a few percent (Strydom et al., 1992).

B. Analytical Ultracentrifugation

There are two general approaches to the characterization of protein-protein interactions in solution using AUC. Equilibrium sedimentation (SE) is thermodynamic and yields information about molecular mass and assembly state. Velocity sedimentation is hydrodynamic and yields information about molecular mass and shape, however, appropriate analysis of boundary shapes and sedimentation rate can also yield quantitative information about stoichiometries and equilibrium constants.

At equilibrium, the concentration distribution of a single, homogeneous species is exponential and is given by

$$c_r = c_i \exp\left(\frac{M_i(1-\bar{v}_i\rho)\omega^2(r^2 - r_m^2)}{2RT}\right) + \text{base} \tag{1}$$

where c_r and c_i are the concentrations of the protein at a radial position r and at a reference position, usually the meniscus, respectively. M_i is the protein molecular weight, $\bar{v}_i$ and ρ are the protein partial specific volume and solvent density, ω is the angular velocity, r is the distance in cm from the center of rotation, r_m is the radial position of the reference position in cm, R is the universal gas constant, T is the absolute temperature and base is a term describing non-sedimenting material. Equation 1 can be abbreviated as:

$$c_r = c_i \exp(\sigma_i) + \text{base} \tag{2}$$

$$\text{where} \qquad \sigma_i = \left(\frac{M_i(1-\bar{v}_i\rho)\omega^2(r^2 - r_m^2)}{2RT}\right)$$

Figure 2 shows typical SE data of the sort described by Equation 1. Here we have a single exponential protein distribution for a recombinant form of the first three short consensus repeat modules of human complement receptor Type-1 (sCR123) (Dodd et al., 1995).

This analysis yields a molecular mass of $21,629 \pm 255$ Da, compared to 21,696 Da predicted from the amino acid composition.

For a self-associating system, i.e. $A + A \leftrightarrow A_2$, Equation 2 will have two exponentials, one for the monomer and one for dimer, i.e.,
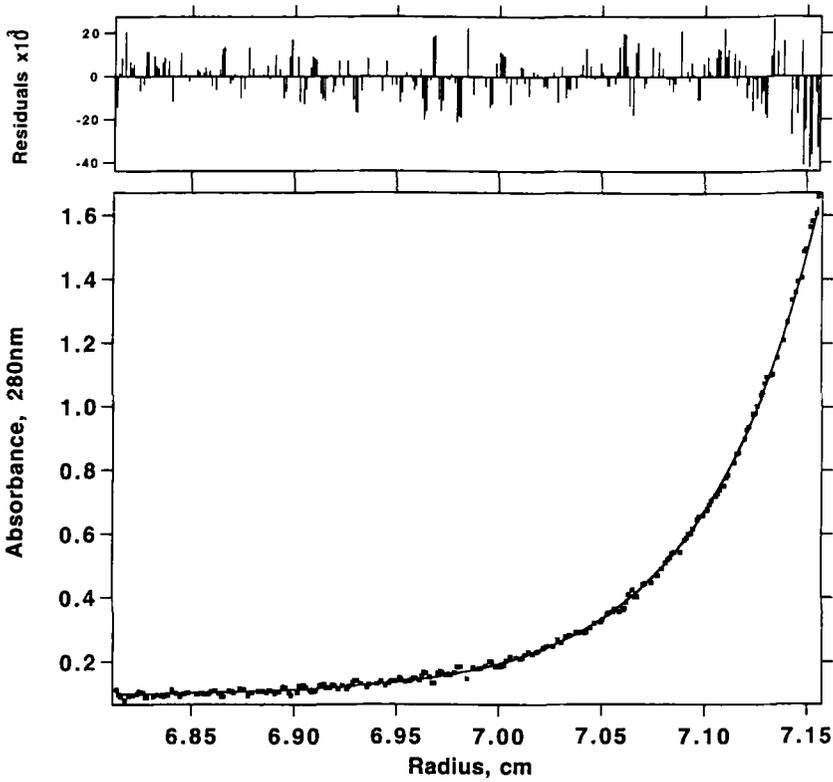
Figure 2. Equilibrium sedimentation for sCR123 at 30,000 rpm, from Dodd et al., 1995.

$$c_r = c_{mono} \exp (\sigma_{mono}) + c_{dimer} \exp (\sigma_{dimer}) + base \qquad (3)$$

From the definition of the equilibrium constant, $c_{dimer} = \dfrac{(c_{mono})^2}{K_{12}}$, Equation 3 may be re-written,

$$c_r = c_{mono} \exp(\sigma_{mono}) + \left(\frac{c_{mono}^2}{K_{12}} \right) \exp(\sigma_{dimer}) + base \qquad (4)$$

From Equation 4 it is clear that in fitting the sum of two exponentials, one for the monomer and one for the dimer, the equilibrium constant, K_{12}, can be fit for directly. For a heterologous association, i.e., A + B ↔ AB, Equation 4 becomes

$$c_r = c_A \exp(\sigma_A) + c_B \exp(\sigma_B) + \left(\frac{(c_A * c_B)}{K_{AB}}\right) \exp(\sigma_B) + base \qquad (5)$$

The thermodynamics of association may be studied from the temperature dependence of K_{12} or K_{AB}. The temperature range accessible to the centrifuge is 0-40 °C, which is limited. However, if ΔH and ΔC are significantly non-zero, a fairly complete description of the thermodynamics can be obtained with this approach (Yoo and Lewis, 1995).

As fitting of SE data is done in terms of sums of exponentials, the approach has limited numerical resolving power. Three exponentials, three species in chemical equilibrium, is about the upper limit if the stoichiometry is unknown or allowed to float, but 4-5 species and 3-4 equilibrium constants can be extracted if the model (stoichiometry) is known. Such analyses require global analysis of data carried out over several loading concentrations and speeds.

For a monomer $\leftrightarrow$ dimer $\leftrightarrow$ tetramer equilibrium, Equation 4 would have the form,

$$c_r = c_{mono} \exp(\sigma_{mono}) + \left(\frac{(c_{mono})^2}{K_{12}}\right) \exp(\sigma_{dimer}) + \left(\frac{(c_{mono})^4}{K_{12}K_{24}}\right) \exp$$

$$(\sigma_{tet}) + base \qquad (6)$$

From the form of Equation 6, it is clear that K_{12}, K_{24}, and c_{mono} will be highly correlated, however, this turns out not to be tragic. The values of these parameters can be recovered. A discussion of parameter correlation and parameter error determination has been presented (Brooks et al., 1994 a,b). Typical data for an RGD4 mutant form of the V_L dimer of REI is described by Equation 6 and is shown in Figure 3 (Chan et al., 1996).

A second general approach is available for the characterization of molecular mass and macromolecular interactions using AUC. The approach is hydrodynamic and is called velocity sedimentation (SV). Here proteins are sedimented at high rotor speed, where the molecule experiences net transport. As first shown by Svedberg (Equation 7), the rate of transport , $d \ln r_m/dt$, is proportional to the molecular mass, M, and to the translational diffusion coefficient, D (Svedberg and Pederson, 1959; Tinoco et al., 1985).

$$s = \left(\frac{1}{\omega^2}\right)\frac{d \ln r_m}{dt} = \frac{M(1 - \bar{v}\rho)D}{RT} \qquad (7)$$

Typical velocity sedimentation data for sCR123 is shown in Figure 4.
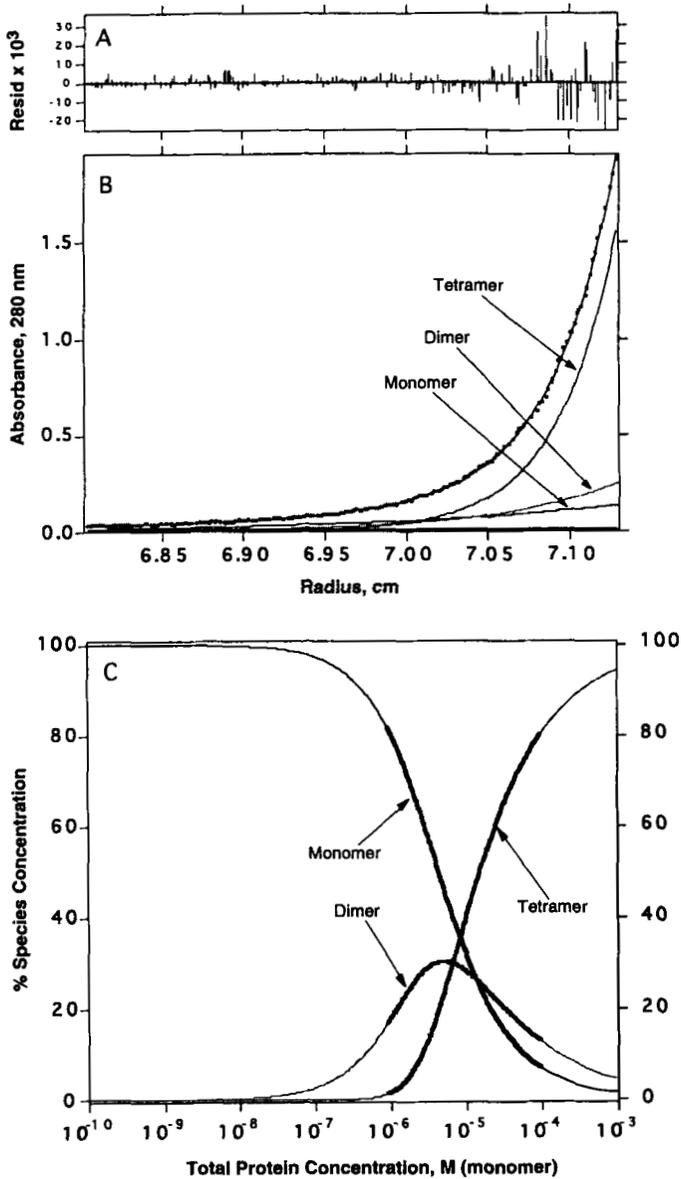
Figure 3. Equilibrium sedimentation of a monomer ↔ dimer ↔ tetramer equilibrium system. Panel **A** is the residuals (theoretical minus experimental data). Panel **B** is the primary data, acquired at 25,000 rpm, fitted to Equation 6. Panel **C** shows the relative concentrations of monomer, dimer, and tetramer, determined from an analysis of the data in Panel **B** according to equation 6. From Chan et al., 1996.
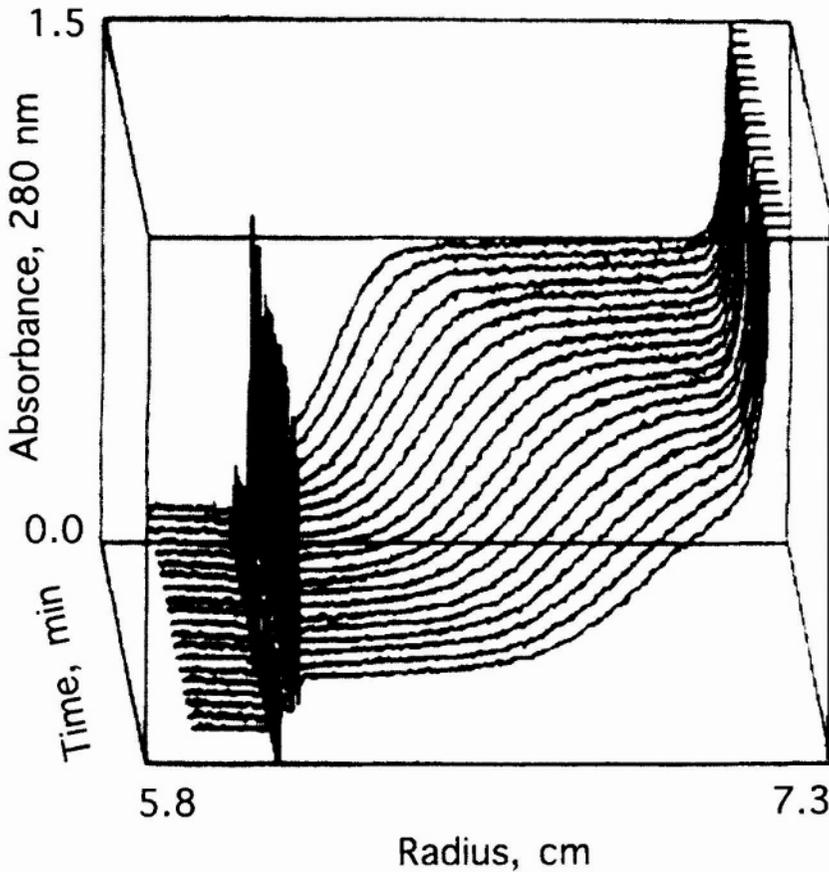
Figure 4. Sedimentation velocity data of sCR123 at 60,000 rpm. The x-axis is radial position (5.8 to 7.3 cm), the z-axis is absorbance (280 nm, 0 to 1.5 A) and the y-axis is time (scans taken at 2-minute intervals). From Dodd et al., 1995.

From Equation 7 we see that a plot of the natural log of the midpoint of the boundary, ln r_m, plotted versus $\omega^2 t$, defines the sedimentation coefficient, s (see Figure 5A). In this case the slope, when corrected to standard conditions, gives an $s_{20,w}$ of 2.05 S.

The translational diffusion constant, D, may be determined from the rate of spreading of the boundary. A practical method for doing this was described by Muramatsu and Minton (1988). Here a term, Z(t') is determined from the rate of boundary spreading. The time dependence of this term gives D, i.e.,

$$Z(t') = \left(\frac{\pi D}{4}\right)t' + \left(\frac{\pi D t_o}{4}\right)$$

(8)
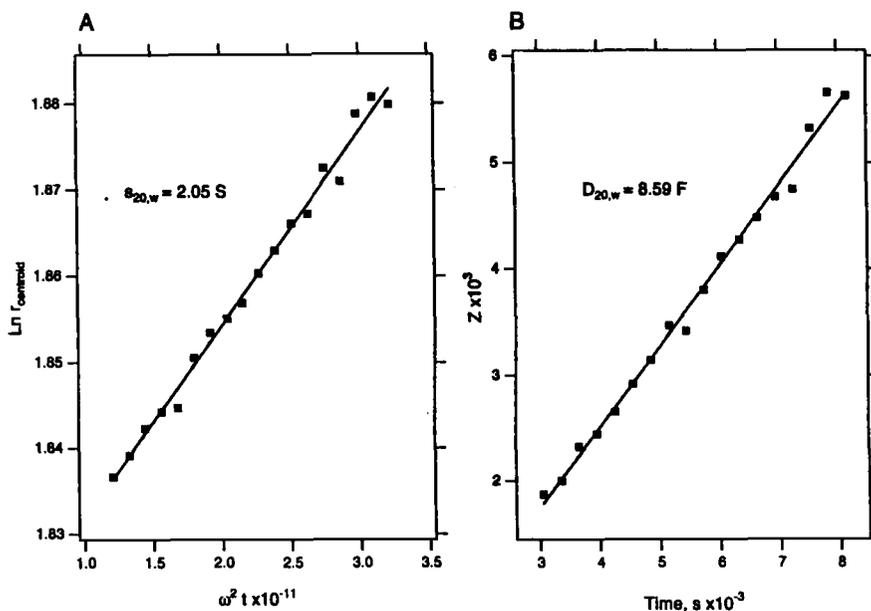
Figure 5. Analysis of the sedimentation velocity data for sCR123 shown in Figure 4. Panel **A**. determination of the sedimentation coefficient from a plot of ln r_m versus $\omega^2 t$. Panel **B**. determination of the translational diffusion coefficient from a plot of $Z(t')$ versus time, using the method of Muramatzu and Minton (1988). Data from Dodd et al., 1995.

Typical data, again for sCR123, are given in Figure 5B. These data yield a normalized diffusion constant of 8.59 F (Dodd et al., 1995). These two constants, s and D, in combination with Equation 7 can be used to determine the molecular mass. In this example, the molecular mass determined from s/D is 21,063 Da. This value compares to 21,696 Da, determined from sequence, and 21,629 Da, determined from equilibrium sedimentation (see Figure 2).

So, for very well behaved SV data, it is possible to determine molecular mass with precision and accuracy. But life is not always so kind. The signals may be smaller (and noisier), there may be more than one component, or species may be reversibly interacting over the period of the run (~1 hour). If any of these conditions holds, s and D, may not be determinable.

For these reasons, until recently, equilibrium approaches have been favored for the measurement of molecular mass or the characterization of protein-protein interactions in solution using sedimentation methods. But as mentioned above, equilibrium sedimentation is useful for systems of limited complexity (n $\leq$ 3 species) because the analysis is numerically ill-conditioned.

However, recently, Stafford has significantly added to the power of hydrodynamic (SV) method by noting that use of the time derivative (subtraction of boundaries) and signal averaging combined with the use of high precision Rayleigh optics, dramatically increases (by 10^2 to 10^3) the signal to noise, or the relative information content of the data (Stafford, 1992a,b, 1994a). Two other related SV approaches are also useful (van Holde and Weischet, 1978; Hansen, 1994; Hansen et al., 1994; Hansen and Wolffe, 1994; Philo, 1994; Schwarz and Hansen, 1994). The Stafford approach involves the computation of an apparent distribution function called g(s*). The resulting data yields a Gaussian peak for every macromolecular species. This makes visualization and characterization of complex mixtures (n > 3 species) in sedimentation experiments significantly easier. The maximum of the peak is the sedimentation coefficient and the area under each curve is proportional to concentration. Finally, the standard error, σ, of the Gaussian is related to the translational diffusion constant. Hence, from s, D, and the Svedberg equation (Equation 7), molecular mass may be determined with good precision in a single one hour sedimentation velocity run (Stafford, 1996; Stafford et al. 1996).

Typical data for a single species is shown in Figure 6B. This data is analogous to the data in Figure 4, except that this latter data was taken in the absorption mode and the data in Figure 6B was obtained using Rayleigh interference optics, with concentration measured in fringes. There are several things to note about this data *compared to that in Figure 4. First the concentration is much lower (~0.12 vs. ~1.0* A_{280}) and as a consequence, the systematic error is much greater. However, as the g(s*) approach uses the time derivative, closely spaced boundaries are subtracted, the systematic error is removed. Second, while systematic error is present, the data is inherently more precise. The result is that when the g(s*) transformation is complete, a smooth Gaussian is produced if the boundary is the result of a single sedimenting species (Stafford et al., 1996) (see Figure 6A).

As noted above, this Gaussian has two important properties. First, the peak gives the sedimentation coefficient. Second, the standard error of the Gaussian, σ, is related to the translational diffusion constant, D (Stafford, 1996). This is shown below:

$$D = \frac{(\sigma < \omega^2 t > r_m)^2}{2 < t >} \tag{9}$$

Hence, as mentioned above, from s, D, and the Svedberg equation, the molecular mass may be determined. Figure 6C,D shows the same result for two sedimenting species and Figure 6E,F shows this result for three sedimenting species.

It should be noted that this approach applies for monomers or for tightly associated complexes. Using titration calorimetry and surface plasmon resonance, it was shown in this case that the binding of the Fab to IL-5 had a K_d < 100 pM (Stafford et al., 1996). However, the g(s*) approach has been extended to weakly
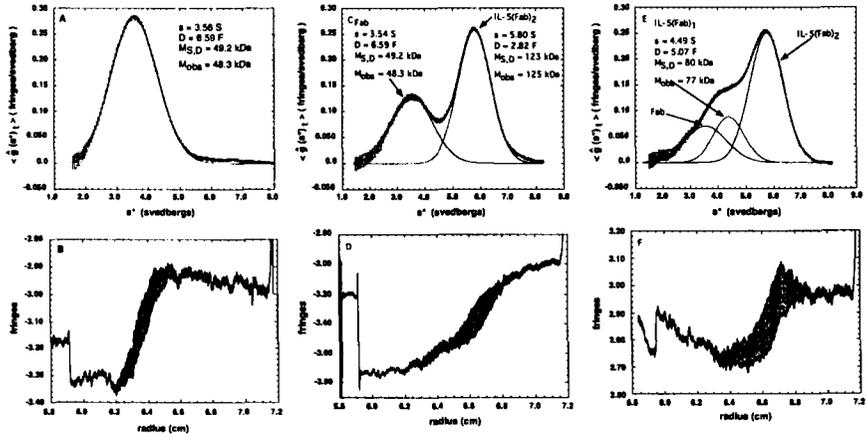
Figure 6.   g(s*) analysis of Fab binding to dimeric IL-5. Panels **B, D** and **F** are the primary sedimentation velocity data obtained using Rayleigh interference optics and expressed in fringe concentration units versus radius (cm). The data is analogous to data taken in the absorption mode shown in Figure 3. In each panel 40 boundaries are taken, 40 s apart. Note the extreme systematic error, especially in panels **B** and **F**. Panels **A, C** and **E** are the corresponding data transformed using the g(s*) time derivative and signal averaging approach of Stafford. In each case the values of s and D are determined with enough precision and accuracy to determine the molecular masses within a few percent. In these panels $M_{s,D}$ is the molecular mass calculated from s, D, and the Svedberg equation (Equation (7)), and M_{obs} is the molecular mass determined by MALDI. Note that for panel E, this is true even for complex, multi-species boundaries and even when the primary data is poor quality. From Stafford et al., 1996.

associating systems (Stafford, 1994b). There are two basic approaches to the analysis of interacting systems by sedimentation velocity. In the first, the velocity of the equivalent boundary position is measured to determine the weight average sedimentation coefficient, s_w. The concentration dependence of s_w can give information about the stoichiometry and equilibrium constants, as long as the sedimentation coefficient of each species can be estimated. Stafford has demonstrated that weight average sedimentation coefficients can be obtained from g(s*) and used to compute reliable values of the equilibrium constant (Stafford, 1994b), i.e.,

$$s_w = \frac{\int_0^\infty s * g(s*)ds *}{\int_0^\infty g(s*)ds *}$$

For example, for a monomer-dimer system, the weight fraction of monomer, f_m can be computed from,

$$f_m = \frac{s_2 - s_w}{s_2 - s_1}$$

where, s_1 and s_2 are the sedimentation coefficient of the monomer and dimer, respectively. The analysis of a sedimentation boundary in this way is directly analogous to the analysis of the trailing boundary from large zone analytical gel permeation experiments, discussed below.

In the second method, the boundary shape itself can be analyzed by comparison to simulated sedimentation boundaries for various stoichiometries to obtain estimates of the equilibrium constant. Curve fitting of simulated sedimentation velocity curves to experimental data has been reported and can be applied to interacting systems (Todd and Haschemeyer, 1981).

To summarize, two general sedimentation approaches are available to characterize protein-protein interactions, equilibrium and velocity. Equilibrium approaches are the most favored for describing reversibly associating systems with K_d's in the range of 10^{-3} to 10^{-8} M, as K_d is fit for directly. Using radiolabeled proteins and direct fractionation methods this range could be extended to 10^{-2} to 10^{-11} M (Attri and Minton, 1986, 1987; Chatelier and Minton, 1987; Darawshe et al., 1993; Darawshe and Minton, 1994). These approaches are thermodynamically rigorous, but are limited in terms of the number of chemical species which can be described. Velocity approaches now exist wherein the sedimentation and diffusion coefficients may be determined quickly and precisely, yielding molecular weights within experimental error to those determined by mass spectroscopy. These approaches are useful for high affinity complexes. Here molecular weight determination gives a precise measure of stoichiometry. This approach has also been successfully applied to analysis of interacting systems.

C.   Analytical Gel Chromatography

AGC is capable of measuring protein stoichiometries and self-association equilibria over a wide range of protein concentration (e.g., 100 μM (Turner et al., 1992) to pM (Beckett et al., 1991)). It also provides precise hydrodynamic information about both the monomer and oligomer forms of the protein. The theoretical and practical basis of this method has been extensively outlined by Ackers and colleagues (Ackers, 1970; Ackers, 1975; Valdes and Ackers, 1979; Turner, 1990). The method has recently been adapted for quantitative, reduced-scale measurements by HPLC (Nenortas and Beckett, 1994). AGC is particularly relevant to the present communication because it is a method which can be set up with tools which are familiar to laboratories that do protein purification. The principle underlying AGC is the molecular size-dependent partitioning of a protein between stationary

(inside the gel matrix) and mobile phases of a sizing column. Thus, a smaller protein will take longer to be transported through the column than a larger one, because it will spend more time in the stationary phase. In a "large zone" AGC experiment a sufficiently large volume of protein sample is applied to the column to ensure that the protein concentration remains constant. The concentration is therefore precisely known, which is required for a thermodynamically rigorous treatment of the data for self-associating proteins. The key experimental observable in large zone AGC is the elution volume V_e of the protein and is concentration dependent for self-associating proteins. The V_e for a given protein concentration is related to a weight-average partition coefficient, $\overline{\sigma}_W$, according to the following relation:

$$\overline{\sigma}_W = \frac{V_e - V_o}{V_i} \tag{10}$$

where V_o and V_i are the excluded (void) and internal volumes of the column. V_o may be determined experimentally as the elution volume of a large, totally excluded solute (such as Blue Dextran on a Sephadex G-100 matrix). V_i is determined as $V_i = V_t - V_o$ where V_t is the elution volume of a small solute which is totally nonexcluded (e.g., tryptophan or glycine). The partition coefficient represents the fraction of internal volume of the gel matrix which is penetrable by the protein, and contains hydrodynamic information about the size and shape of the protein (Ackers, 1970).

Figure 7A shows an example large zone AGC elution profile of lambda cI repressor at a concentration of 51 pM cI. The V_e is indicated by the position of the equivalent sharp boundary, which is determined experimentally as the centroid of the leading (and/or trailing) edge of the elution profile. Large zone experiments are conducted over a wide range of cI concentration. The elution volume of the protein at a specific concentration can then be compared to the elution volume of the pure monomer and dimer species in order to compute the fraction of protein as monomer. Figure 7B shows a dissociation curve for the monomer-dimer equilibrium of cI repressor series of elution profiles of cI repressor measured over a range of cI concentrations. Here the data are shown as the fraction of cI as monomer, f_M, which was calculated from $\overline{\sigma}_W$ according to:

$$f_M = \frac{\overline{\sigma}_W - \sigma_2}{\sigma_1 - \sigma_2} \tag{11}$$

where σ_1 and σ_2 are the partition coefficients of the pure monomer and dimer forms of cI, respectively. The data in Figure 7B were analyzed according to a monomer-dimer equilibrium model [i.e., $f_M = -1 + (1 + K_{12}C_T)^{1/2}/4K_{12}C_T$] which yielded an equilibrium dimerization constant $K_{12} = 1.8 \times 10^8$ M^{-1}. The best-fit values of the asymptotic parameters σ_1 and σ_2 can be correlated to the apparent molecular radius of the pure monomer and dimer species, respectively (Ackers, 1970).

Given these results, one may anticipate in favorable cases to be able to estimate dimerization constants as tight as 10^9 or 10^{10} M^{-1}. This is especially true for cases where the partition coefficient of the monomer species, σ_1, is known in advance and can therefore be constrained during regression analysis. An estimate of σ_1 may be determined in advance by characterizing a highly dissociated mutant form of the protein (Turner et al., 1992), or by characterizing the native form of the protein under high salt or temperature conditions where the monomer-dimer equilibrium
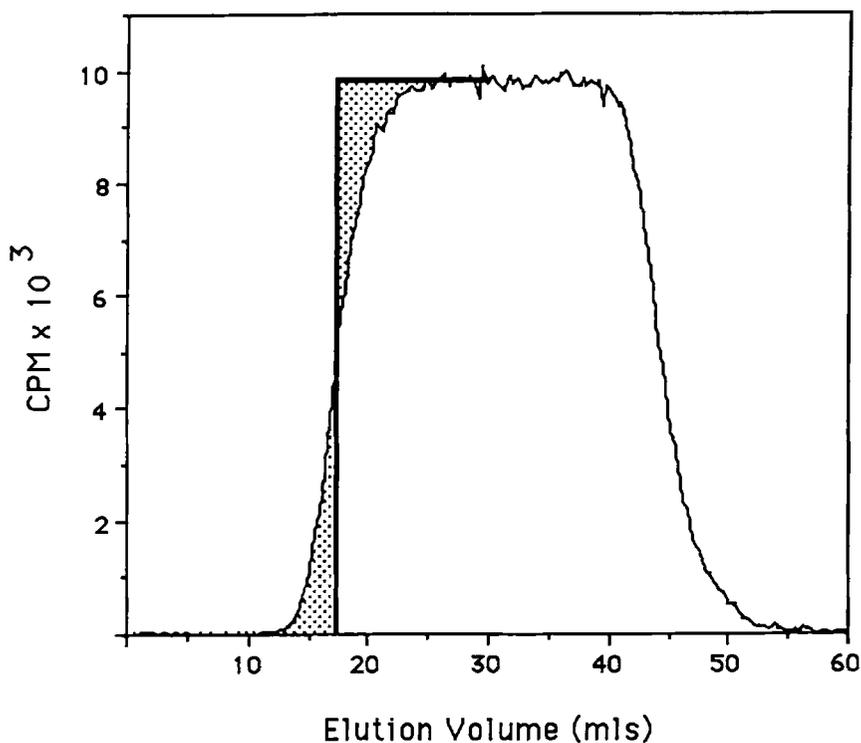
Elution Volume (mls)

Figure 7. AGC. **Upper panel.** large zone elution profile of lambda cI repressor on Sephadex G-100 column. Plateau concentration is 51 pM cI. The elution volume, Ve, of 51 pM cI is given by the equivalent sharp boundary of leading edge (vertical line) which was determined by numerical integration as the centroid (equal areas above and below the leading edge). **Lower panel.** Fraction of cI as monomer (Equation. 11) versus logarithm of cI molarity using Sephadex G-100 (open symbols) or BioGel P-100 (filled symbols) resin. Curve is based on a dimerization equilibrium constant of 1.8 x 10^8 M^{-1} determined by nonlinear least squares analysis. Monomer and dimer partition coefficients were determined from regression analysis as asymptotic parameters at infinite and zero cI concentration. Taken from Beckett et al., 1991.
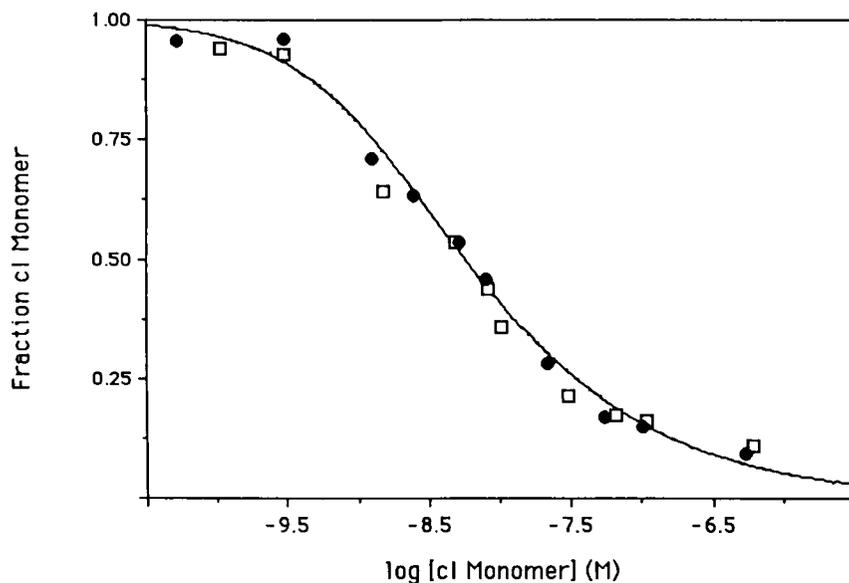
is sufficiently weakened. Although, it is possible that mutationally induced changes in hydrodynamics for a given protein could lead to a small, but detectable, shift in σ_1 with respect to the native monomer.

D.   Isothermal Titration Calorimetry

ITC (Figure 8) is a solution phase technique which measures binding reactions by virtue of binding enthalpy changes which accompany all binding reactions. Several reviews on ITC applications, instrumentation, and data analysis have been presented elsewhere (Biltonen and Langerman, 1979; Wiseman et al., 1989; Freire et al., 1990; Bundle and Sigurskjold, 1994; Bhatnagar and Gordon, 1995; Fischer and Singh, 1995). In the present communication we limit our discussion to the salient practical features of ITC and the simplest equations for data analysis.

ITC is often viewed as one of the most thermodynamically robust methods for investigating protein-protein interactions. There are several appealing features which give ITC this reputation. First, the reactants can be studied in their native chemical forms. Chemical modification or radiolabeling of proteins is not needed. Nor is it necessary to covalently attach proteins to surfaces. Second, reactants and products do not have to be physically separated after binding equilibrium has been obtained in order to quantitate binding. Thus, the amount of binding is determined

Figure 8. Simplified schematic of isothermal titration calorimeter, from Doyle et al., 1996. A solution containing one protein reactant (shown as half-circles) is titrated in the reaction cell with another protein (circles) by making sequential injections from a syringe. Binding is detected as heat uptake or release, by comparison to the reference cell which contains only solvent. Mixing of the proteins is done by continuous rotation of the syringe, which extends to the bottom of the reaction cell and has a stirring blade at the tip (not shown). A more detailed description can be found elsewhere (Wiseman et al., 1989).

at equilibrium, and the approach to equilibrium can usually be verified visually at each increment of the titration curve. Third, ITC measures equilibrium constants directly, in contrast to kinetic methods, which estimate equilibrium constants

Figure 8. Simplified schematic of isothermal titration calorimeter, from Doyle et al., 1996. A solution containing one protein reactant (shown as half-circles) is titrated in the reaction cell with another protein (circles) by making sequential injections from a syringe. Binding is detected as heat uptake or release, by comparison to the reference cell which contains only solvent. Mixing of the proteins is done by continuous rotation of the syringe, which extends to the bottom of the reaction cell and has a stirring blade at the tip (not shown). A more detailed description can be found elsewhere (Wiseman et al., 1989).

at equilibrium, and the approach to equilibrium can usually be verified visually at each increment of the titration curve. Third, ITC measures equilibrium constants directly, in contrast to kinetic methods, which estimate equilibrium constants

indirectly based on interpretation of on and off rate processes. Finally, ITC is rich in thermodynamic information. The intrinsic binding enthalpy for a protein-protein interaction can be deduced from the primary heat signal (Biltonen and Langerman, 1979; Doyle et al., 1995), and when combined with affinity, yields the binding entropy change. These thermodynamic parameters portray the molecular-level interactions which drive binding and provide a more detailed understanding about binding mechanism than is available from the binding affinity alone (Ladbury and Peters, 1994). Recent studies have shown the thermodynamic information available from ITC is a powerful measure of the amount of polar and apolar surface area which is buried during binding (Murphy and Freire, 1992; Freire, 1993; Murphy et al., 1995) and is also a measure of the amount of conformational change which is coupled to binding (Spolar and Record, 1994).

The main drawback of ITC is that high concentrations of reactants are needed due to limitations in instrumental sensitivity. For tight binding interactions ($K_d <$ 1 μM) the concentration of protein in the calorimeter syringe would typically be 25-100 μM. The concentration of protein in the calorimeter cell would be lower, but usually not much less than 1 μM. For weaker affinities the concentrations of both reactants would have to be even higher. Hence, the amounts of protein reagents needed for ITC are considerably higher than for other methods such as SPR or ELISA. Moreover, the higher concentrations also lead to a higher probability that self-association reactions may occur for the reactants and/or the products. Thus, it is especially important to assess the self-association status of the protein reactants and products at concentrations where ITC experiments are conducted. AUC is a convenient method for this purpose because it is both rapid and optimal for evaluating protein self-associations at concentrations which are relevant to ITC studies.

The experimental design of ITC studies has been outlined by Wiseman et al. (1989). It was shown that the concentration of protein in the cell is a key experimental variable, and that the product of this concentration times the equilibrium binding constant should be in the range of 1 to 1,000. Many protein-protein interactions have affinities which are in the range of nanomolar and below. The design strategy then becomes one of lowering the concentration of protein in the cell as much as possible without losing signal-to-noise altogether. For the most precise ITC instruments currently available, the signal-to-noise becomes poor when the concentration of protein in the cell is below micromolar. Thus the tightest affinity which can be measured directly is approximately a few nanomolar. Figure 9 shows ITC data for the titration of an antibody with its antigen, a soluble form of the T-cell receptor CD4.

The concentration of mAb in the reaction cell was 0.51 uM (1.02 uM binding sites) and the best-fit equilibrium binding constant was readily determined and equal to 1.0 x 10^8 M^{-1}. (The product of the protein concentration and binding constant is 110.) It is frequently the case that the affinity may be too tight to measure directly under a given set of solution conditions. Consequently, it may be necessary
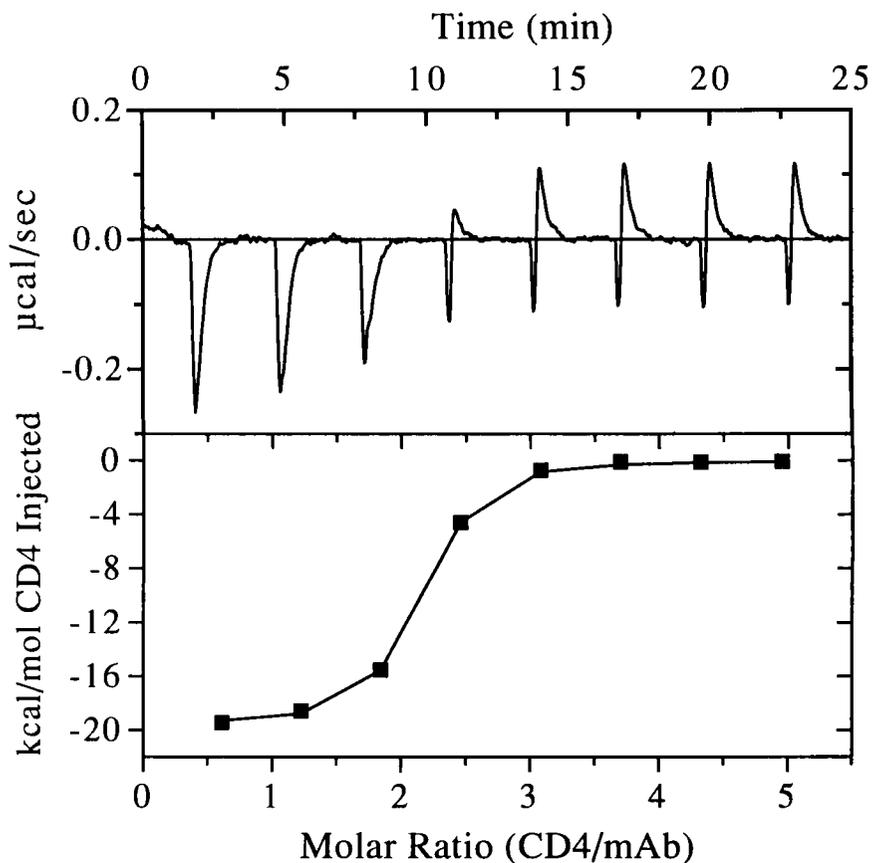
Figure 9. Illustrative ITC data: titration of an IgG antibody with a soluble single-site mutant form of the T-cell receptor CD4. Eight injections of CD4 were made into the reaction cell containing the mAb. **Top**: Power in ucal/second versus time. Downward spikes indicate exothermic heats of CD4 binding to the mAb. All available binding sites on the mAb are saturated by CD4 after 3-4 injections and the upward spikes are thus endothermic background heats in the absence of binding. **Bottom**: Integration of the spikes in the top part of the figure gives the heat per mole of CD4 which is injected and is plotted versus the molar ratio of CD4 which has been added per mole of mAb in the cell. Curve represents best-fit of the data to an independent and identical site binding model (Equations 12 and 14). The best-fit parameters of the model were: K = 1.0×10^8 M^{-1} (equilibrium binding constant), n = 1.88 (molar ratio of reactivity, or apparent stoichiometry), and ΔH^{app} = -19.6 kcal/mol CD4 (binding enthalpy change).

to alter solution conditions in order to weaken the interaction sufficiently to allow a direct measure of the affinity. The easiest solution condition to change is temperature, both experimentally and in terms of correcting for the temperature-dependence of the affinity, but it is also possible to change pH and still be able to correct for the pH-dependence of affinity. Both of these thermodynamic strategies are described below.

Equations for data analysis of ITC data relate changes in fractional ligand saturation and binding enthalpies to changes in reactant concentrations. The complexity of the equation depends on the binding model and can be adapted on a case-by-case basis (cf., Parody-Morreale et al., 1987; Wiseman et al., 1989; Freire et al., 1990; Wyman and Gill, 1990; Fischer and Singh, 1995). For the present purposes we focus on the simplest binding model wherein all sites bind independently and with identical affinity and enthalpy change. For a protein P1 which has n identical binding sites for a monomeric protein P2 the heat q_i absorbed or released at the ith titration step is given by:

$$q_i = n \Delta H^{app} [P1]_{total,i} V_{cell} (\overline{Y}_i - _i\overline{Y}_{-1}) \qquad (12)$$

Here ΔH^{app} is the apparent binding enthalpy change per mole of P2 bound, $[P1]_{total,i}$ is the total concentration of P1 in the cell after the ith injection, V_{cell} is the effective volume of the reaction cell which can be determined empirically (McKinnon et al., 1984), and $\overline{Y}_i$ and $\overline{Y}_{i-1}$ are the fractional saturations of P1 with P2 at the end of injections i and i-1, respectively. The term n is formally the number of binding sites on P1. However, in practice n is an adjustable parameter during least squares data analysis which is better described as the molar ratio of reactivity between P1 and P2 (or the apparent stoichiometry). Thus, experimental values of n are nonintegral, but nevertheless should be close to the true stoichiometry. Values which deviate significantly from the known stoichiometry suggest inaccuracy in determination of reactant concentrations or partial inactivation of the reactants.

ΔH^{app} in Equation 12 is comprised of the intrinsic binding enthalpy ΔH^{int} for P1-P2 binding but also includes heats of proton ionization ΔH^{ion} of the buffer which are proportional to uptake or release of any protons which may be coupled to P1-P2 binding. ΔH^{app} can be corrected to yield ΔH^{int} by measuring ΔH^{app} in buffers having different heats of ionization ΔH^{ion} (Biltonen and Langerman, 1979; Murphy et al., 1993; Doyle et al., 1995). The following relation holds for each buffer condition individually.

$$\Delta H^{app} = \Delta H^{int} + \Delta H^{ion} \Delta protons \qquad (13)$$

A plot of ΔH^{app} versus ΔH^{ion} (Figure 10) should be linear with a slope equal to Δprotons and an intercept equal to ΔH^{app}. Ideally, at least three buffers should be utilized to rule-out possible specific buffer effects. If one is only interested to know the intrinsic binding enthalpy, it is sometimes reasonable to carry out ITC experi-
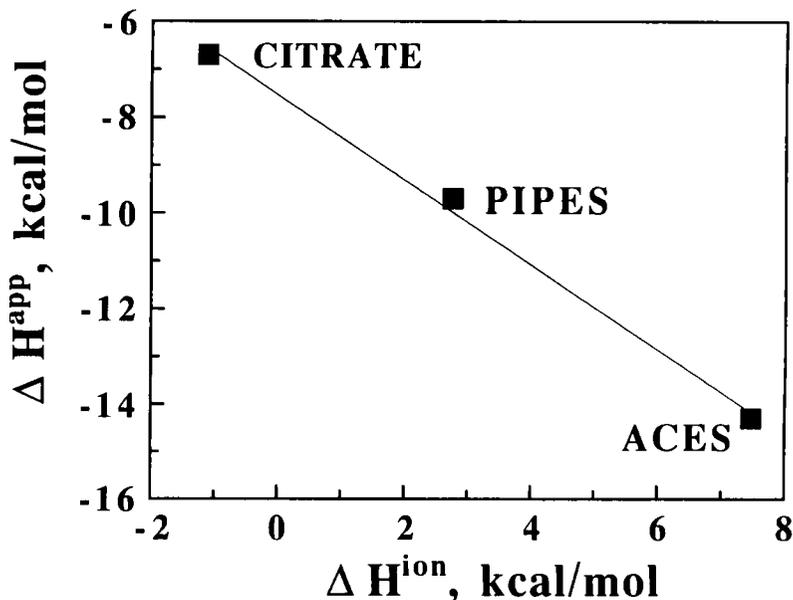
Figure 10. Apparent binding enthalpy changes for reaction of trypsin and ecotin as measured by ITC in buffers which have different ionization enthalpy changes. Ecotin is a 32 kDa homodimeric protein (McGrath et al., 1995). A linear relationship is evidence for the absence of specific buffer effects on the trypsin-ecotin binding reaction. Slope gives the number of protons linked to trypsin-ecotin binding (Δproton = -0.76 per trypsin), and intercept gives the intrinsic trypsin-ecotin binding enthalpy (ΔH^{int}= -10.1 kcal/mol trypsin) at 30 °C and pH 5.5.

ments in a buffer which has a ΔH^{ion} (e.g., phosphate (Christensen et al., 1976)) which is much lower than ΔH^{int} so that ΔH^{app} approximates ΔH^{int}. On the other hand, the number of protons coupled to a binding reaction is useful information and provides insight into the binding mechanism. The physical parameter Δproton can also be used to determine the pH dependence of binding affinity, as described below in the section on measuring tight binding affinities.

The fractional saturation function in equation 12, $\overline{Y}_{P1}$, for independent and identical binding of protein P2 to protein P1 is

$$\overline{Y}_{P1} = \frac{K[P2]_{free}}{1 + K[P2]_{free}} \tag{14}$$

where K is the equilibrium binding constant, and $[P2]_{free}$ is the concentration of free (unbound) P2 at equilibrium. Since most ITC experiments are conducted at

concentrations of P1 which are near the equilibrium dissociation constant, $[P2]_{free}$ must be calculated from knowledge of the total amount of P2 added at each step of the titration, $[P2]_{total}$, and the amount which is bound to P1. This is accomplished by solving for $[P2]_{free}$ as a quadratic in the following conservation of mass equation:

$$[P2]_{free} = [P2]_{total} - n\, [P1]_{total}\, \overline{Y}_{P1} \tag{15}$$

The total concentrations of P1 and P2 change upon each injection of a titration experiment. $[P1]_{total}$ decreases because a volume ΔV (the injection volume) is expelled from the effective reaction cell during each injection. $[P2]_{total}$ increases because it is being added to the cell with each injection, but a correction for the amount which gets expelled from the reaction cell during injection is also needed. The equations used to calculate changes in total concentrations of P1 and P2 have been reported in a couple of slightly different ways, depending on how the displaced-volume effects are treated (i.e., whether volume is displaced before mixing occurs or concomitantly) (Parody-Morreale et al., 1987; Bundle and Sigurskjold, 1994). The following equations are perhaps the most general because they allow for different injection volumes to be included within a single ITC experiment (adapted from Microcal, Inc. ITC Tutorial Guide, (Brandts, 1993)):

$$[P1]_{total} = [P1]_{total,o} \left(\frac{1 - (\Delta V / 2V_{cell})}{1 + (\Delta V / 2V_{cell})} \right) \tag{16a}$$

$$[P2]_{total} = [P2]_{total}^{syr}\, \Delta V / V_{cell}(1 - \Delta V / 2V_{cell}) \tag{16b}$$

These equations pertain to the change in concentration of P1 and P2 for an individual injection. Here $[P1]_{total,o}$ is the total concentration of P1 in the reaction cell before the injection, and $[P2]_{total}^{syr}$ is the total concentration of P2 in the syringe. Equations 16a and 16b assume that the concentrations of P1 and P2 in the volume of liquid which gets displaced from the reaction cell are equal to the average concentrations at the start and finish of the injection.

In addition to its primary usefulness for measuring heterologous protein-protein interactions, ITC can in favorable cases be used to measure dimerization equilibrium constants and thermodynamics of homologous protein-protein interactions. Based on the current sensitivity of ITC instrumentation, typical protein-protein interactions can be studied near micromolar concentrations or higher. One may expect to be able to measure dimerization constants down to about micromolar. The upper limit for experimentally accessible dimerization constants of proteins is probably near 100 uM due to the likelihood of solubility limits of proteins in the 10-100 mg/mL range. An application of ITC for studying the monomer-dimer equilibrium of interleukin-8 is given below (see Section III.D). Equations for the

analysis of ITC data for dimerizing molecules are given elsewhere (Murphy and Gill, 1989).

E. Kinetics Using Surface Plasmon Resonance

The technique of SPR is a method that has recently been commercialized by Pharmacia in their BIAcore instrument. It provides, in principle, a general mechanism for characterizing the kinetics of protein-protein interactions. With this instrument, a protein, the ligate is attached to a gold surface using specific or nonspecific chemical attachment approaches. Then, using a microfluidics system, a solution of analyte is flowed over the modified surface. As the analyte binds, the refractive index at the surface changes and this can be detected by the SPR sensor (Jonsson et al., 1991, 1993; Malmqvist, 1993; Malmqvist and Granzow, 1994). A schematic of the SPR detector is shown in Figure 11 (upper panel).

The lower panel in Figure 11 shows primary data from a typical BIAcore experiment. The first phase is a buffer wash and the signal stays constant. Subsequently, the analyte is flowed over the modified gold surface and it is bound to the ligate in a time dependent manner. As this occurs, the refractive index increases and the SPR detector responds with an increasing signal, proportional to binding. When the binding phase is complete, buffer is again flowed over the surface and the analyte dissociates with an exponential decay. Figures 12A and 12B show the association phases with various association constants and a fixed off constant. Similarly, Figures 12C and 12D show the dissociation phases with various dissociation constants. These data give a measure of the accessible ranges of rate constants obtainable from the SPR approach.

The chemistry of such a simple process is described by
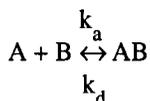
$$A + B \overset{k_a}{\underset{k_d}{\leftrightarrow}} AB$$

The rate of formation of product is given by

$$\frac{d[AB]}{dt} = k_a[A][B] - k_d[AB] = k_a[A]([B]_o - [AB]) - k_d[AB] \tag{17}$$

where k_a and k_d are the association and dissociation rate constants and B_o is the concentration of B at the start of the reaction. Since the response of the BIAcore is proportional to [AB], Equation 17 may be rewritten as

$$\frac{dR}{dt} = k_a C(R\max - R) - k_d R \tag{18}$$
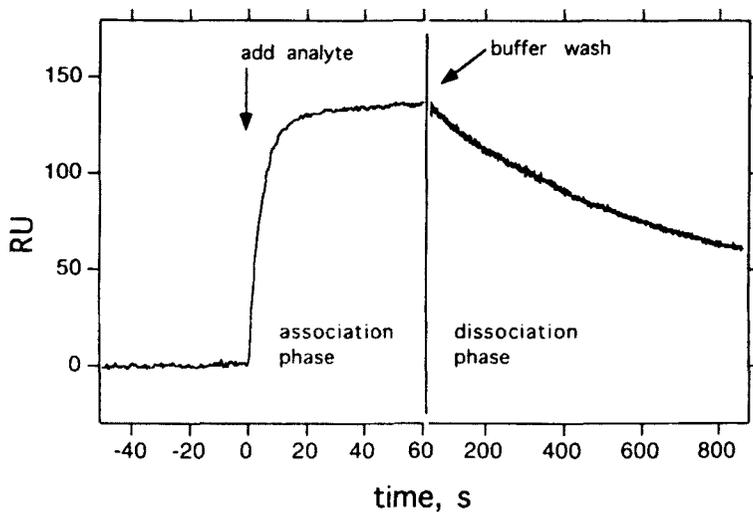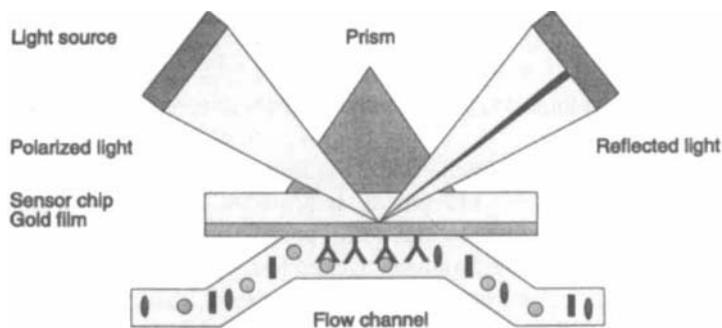
Figure 11. The use of surface plasmon resonance (SPR) to quantitate the kinetics of protein-protein interactions. **Upper panel**: a schematic of the SPR detection system. As molecules flowing past the sensor surface are bound by molecules attached to the sensor surface, the refractive index at the surface changes. This ultimately causes the angle of a minimum in the intensity of the reflected light to change. This change, measured in response units (RUs), allows the kinetics of protein-protein interactions to be visualized. From Malmqvist and Granzow, 1994. **Lower panel**: a diagram of the on- and off-progress curves, as visualized by SPR. As analyte binds to the surface, RUs increase. In this case, the two N-terminal domains of sCD4 and D1D2 are bound to a monoclonal antibody, CE9.1. Once the binding phase is complete, buffer is washed over the surface and the analyte dissociates and is washed away (Myszka, unpublished observations).
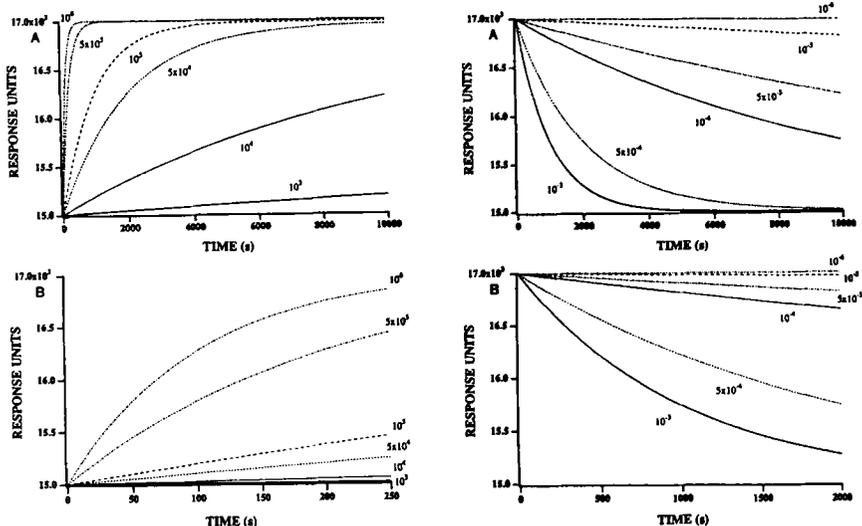
Figure 12. Simulated association and dissociation progress curves for a protein-protein interaction. Panels **A** and **B**. left: the effect of k_a on the association progress curve. Equation 13 was used to generate model data sets using the following parameter values: R_i = 15,000 RU, R_{max} = 2000 RU, C = 10^{-8} M, k_d = 10^{-5} s^{-1} and k_a varied from 10^3 to 10^6 M^{-1} s^{-1}. Panels **A** and **B**, right: the effect of k_d on the dissociation progress curve. Equation (14) was used to generate model data sets using the following parameter values: $R_t \rightarrow \infty$ = 15,000 RU, R_a = 2000 RU, k_d varied from 10^{-3} to 10^{-6} s^{-1}. From OShannessy et al., 1993.

where C is the constant concentration of ligate, Rmax is the capacity of the immobilized ligand surface, expressed in resonance units and (Rmax - R) is proportional to the number of unoccupied surface binding sites at time, t. Upon rearrangement of Equation 18, one obtains

$$\frac{dR}{dt} = k_a C\, R_{max} - (k_a C + k_d)R \qquad (19)$$

Historically, BIAcore data is analyzed as a plot of dR/dt versus R, yielding a slope, k_s which is defined as

$$k_s = k_a C + k_d \qquad (20)$$

Hence, a plot of k_s versus C gave a slope of k_a and an intercept of kd. However, a more straightforward and statistically rigorous approach is to analyze the data directly in terms of the integrated rate equation (O'Shannessy et al., 1993),

$$R_t = \left(\frac{C\,K_a R_{max}}{k_a\,C + k_d} \right)(1 - e^{-(k_a c + d_d)t}) + R_i \qquad (21)$$

where R_i is the signal at $t = 0$. For the dissociation phase, the equation is

$$R_t = R_a\, e^{-kat} + R\infty$$
$$(22)$$

where $R\infty$ is the response value at infinite time.

For simple models, this approach has proved very successful. However, for more complex systems, including terms for mass transport, conformational changes, heterogeneous attachment, etc., closed-form integrated rate equations, such as Equation 21, do not exist. Under these conditions, one has to employ numerical integration (Morton et al., 1995). Morton and colleagues have shown this to be a robust approach. However, these models are inherently more complex and these workers, and others, have shown in that addition to numerical integration, global analysis of data over wide concentrations ranges is needed to distinguish between models (Fisher et al., 1994a,b; Morelock et al., 1995; Morton et al., 1995).

If the binding of the analyte to the ligate surface comes to equilibrium (as opposed to stoichiometric association due to high affinity or [analyte] $>> K_d$), under the conditions of the SPR experiment, the plateau signal is a measure of the saturation of surface binding sites. This signal can then be measured as a function of analyte concentration and fit to an equilibrium binding function. From these data the equilibrium binding constant may be determined (see Figure 13). If data are analyzed globally, the value of the equilibrium binding constant (and its associated error) can be used to constrain implicitly the analysis of the kinetic data.

Similarly, the experiment can be done with fixed concentrations of the soluble ligate (or its analogues) in the presence of the analyte, in a competitive binding format. The concentration of the free analyte may then be quantified by a decrease in the plateau value. From an analysis of these data in terms of a competitive binding function, the true solution equilibrium binding constant of the analyte for the free ligate or for the competitive inhibitor may be determined (Morelock et al., 1995) (see Figure 14). Similar approaches have been developed by others (Karlsson, 1994; Ward et al., 1995).

Therefore, depending on the format of the SPR experiment, kinetic and/or equilibrium data for protein-protein (or in this case, protein-peptide) interactions may be obtained. That said, one should keep in mind that the SPR approach measures the interactions of macromolecules with ligands chemically bound to the surface of a detector. As this field matures, it is becoming evident that these interactions are more complicated than pure solution interactions. These complications include mass transfer effects, valency considerations, ligand density, fluid dynamics, inhomogeneity of ligand immobilization, rebinding during the dissocia-
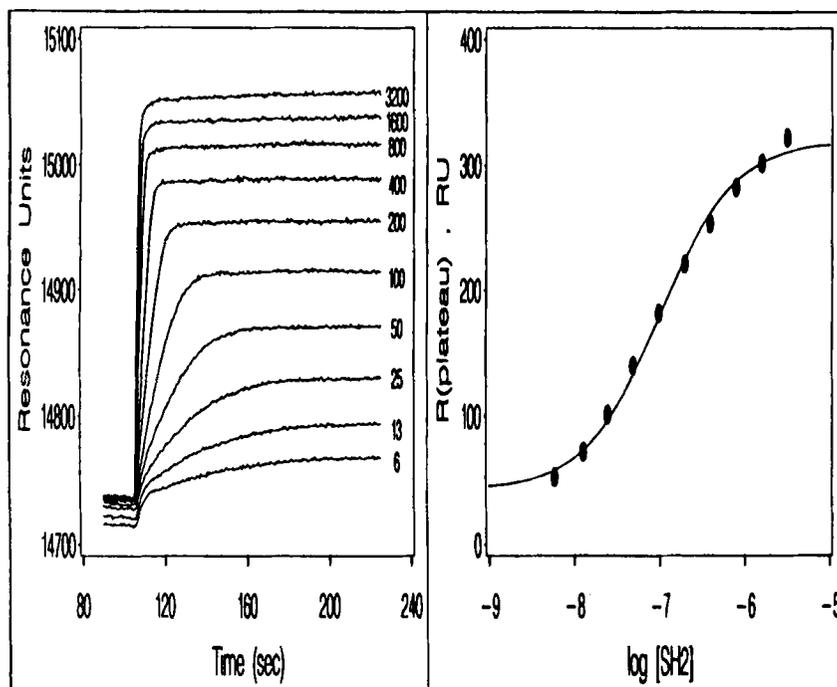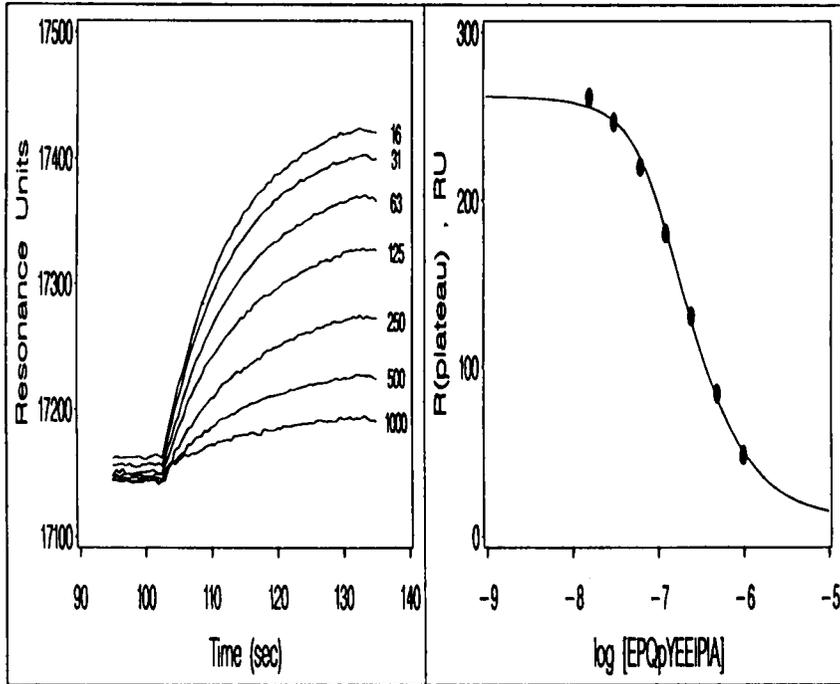
Figure 13. Determination of equilibrium dissociation constants for analyte binding to ligate on a surface from SPR kinetic data. Determination of the equilibrium dissociation constant for and SH2 domain analyte binding to an immobilized ligate. **Left panel** shows the progress curves at increasing concentrations of SH2 (nM). **Right panel** shows a plot of the plateau responses as a function of SH2 concentrations. These data can be analyzed by a Langmuir binding isotherm to determine the binding constant. From Morelock et al., 1995.

tion phase, crowding, and instrumental systematic errors (for a discussion see O'Shannessy and Winzor, 1996; Schuck, 1996). Careful design of experiments can minimize many of these problems.

F.   Other Methods

The kinetics and thermodynamics of protein associations have also been characterized by intrinsic and extrinsic spectroscopic approaches. Though much more dependent on the characteristics of the particular proteins or agents being studied, they often provide significant insight into protein-protein interactions (Aloj et al., 1973; Ingham et al., 1974, 1975; Formisano et al., 1977; Johnson et al., 1978; Lee

Figure 14. Determination of equilibrium binding constants for analyte binding to soluble ligate from SPR kinetic data. Determination of the equilibrium binding constant for peptides competing for the binding of an SH2 domain to an immobilized ligate. **Left panel** shows the progress curves at increasing concentrations of free ligate (nM). **Right panel** shows a plot of the plateau responses as a function of free ligate concentration. These data can be analyzed by a competitive binding function to determine the solution equilibrium constant. From Morelock et al., 1995.

et al., 1989a,b; Wensel and Stryer, 1990; Shapiro and Vallee, 1991; Otto-Bruc et al., 1993). These and other biophysical and biological approaches have recently been reviewed (Eisenstein and Schachman, 1989; Phizicky and Fields, 1995).

III. DEFINING THE MODEL

A.   Stoichiometry

Binding stoichiometry is often the first question addressed for a given protein-protein interaction. The stoichiometry defines the number of binding equilibrium or kinetic constants and thus places limits on the complexity of models which are

used during data analysis. The fractional saturation $\overline{Y}$ of a multimeric protein P1 with increasing concentrations of a monomeric protein P2 is given by the general expression:

$$\overline{Y}_{P1} = \frac{\displaystyle\sum_{i=1}^{n} i\, K_i\, [P2]^i}{i\left(1 + \displaystyle\sum_{i=1}^{n} K_i\, [P2]^i\right)} \tag{23}$$

Here the K_i's are overall equilibrium binding constants (Cantor and Schimmel, 1980; Wyman and Gill, 1990) for binding i P2 molecules onto P1, and n is the stoichiometry. For single-site binding mechanisms (n=1) or for cases where multiple sites behave independently and with identical affinity, Equation 23 simplifies to the familiar Langmuir isotherm given by Equation 14. We note that Equation 23 is "model-independent" and allows for heterogeneity of the binding sites on P1, and also allows for cooperative interactions to occur between sites. However, Equation 23 only applies to proteins which do not change self-association status during binding. For more complex equilibrium binding models (ligand-induced oligomerization, multivalent P1, multivalent P2, etc.) a thorough treatise on the binding partition function as it relates to binding models is given by Wyman and Gill (1990).

Stoichiometry is best determined by direct evaluation of the molecular mass in solution of the fully bound species (e.g., the fully bound species in Equation 23 is P1•P2$_n$). There are, nonetheless, several indirect methods for determining stoichiometry which generally yield only an apparent stoichiometry, or more specifically, a molar ratio of reactivity. ITC is one of the most precise methods for determining molar ratios of reactivity and readily measures this ratio with a precision of a few percent. That said, we have found AUC to be the method of choice for evaluating the true stoichiometry because it is capable of establishing the molecular mass of the complex, directly. A case study is given by work on interleukin-5 (IL-5), which is a covalent homodimeric cytokine that binds to membrane-bound and soluble forms of its α-receptor. Because of its homodimeric structure and the well-known theme of cytokine-induced receptor dimerization for initiation of signal transduction, it was expected that an IL-5 dimer might bind two receptor molecules. ITC experiments yielded a molar ratio of reactivity of only 0.9 IL-5 dimer per receptor (Johanson et al., 1995), which suggested a 1:1 stoichiometry but was not in itself proof. Questions remained as to whether the concentrations of IL-5 and its receptor were known accurately, and whether the IL-5 receptor was fully active. In contrast, equilibrium sedimentation studies (see Figure 15) provided solid evidence that the true stoichiometry was 1:1 based on evaluation of the molecular mass of the IL5-IL5R-α subunit complex directly (Johanson et al., 1995).

B.   Coupled Self-Association Reactions

Another critical step in defining the binding model for a protein-protein inter-action is to establish whether the binding reaction induces changes in self-associa-tion status of the reactants. Ligand-induced changes in self-association of proteins alters their functional properties. It causes inaccurate equilibrium and kinetic binding parameters and also produces apparent cooperativity in the binding curve (Doyle et al., 1986). It is possible to analyze binding curves with a model which accounts for ligand-induced self-association reactions and allows for the deconvo-lution of the equilibrium parameters for ligand binding to specific self-association
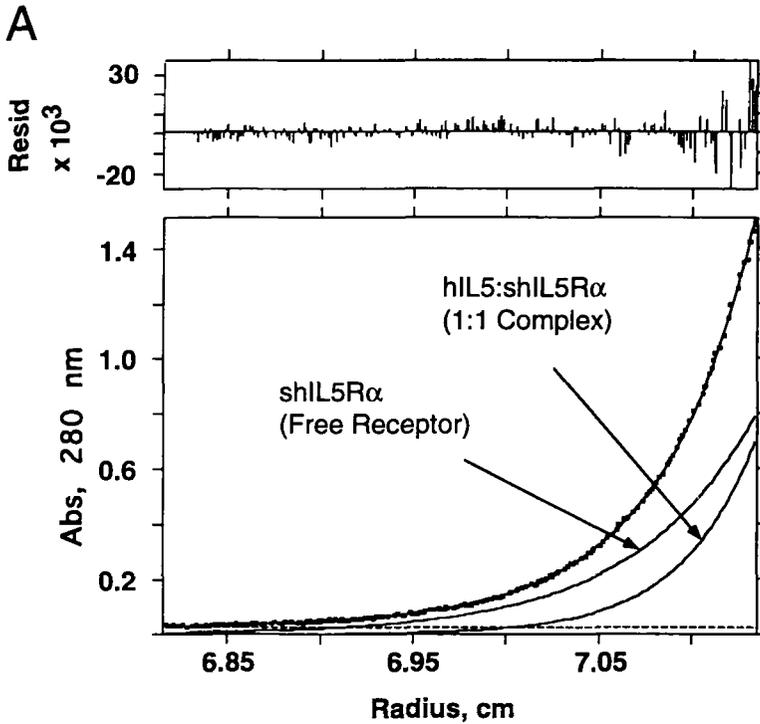
Figure 15.   Species analysis of SE data demonstrating 1:1 complex of hIL5 and shIL5R-α subunit. Panel **A**: SE data (20,000 rpm) for a mixture of hIL5 and shIL5R-α subunit with excess shIL5R-α subunit. The data were fit to model containing terms for shIL5R-α subunit and a 1:1 complex of hIL5 and shIL5R-α subunit, only. The fits did not converge when the 2:1 complex or free hIL5 were included. Panel **B**: the same experiment as in panel **A** (except at 18,000 rpm), except that the molar excess of shIL5R-α subunit has been increased. The fact that the same model fit two independent protein distributions, is strong support for the 1:1 complex. From Johanson et al., 1995.
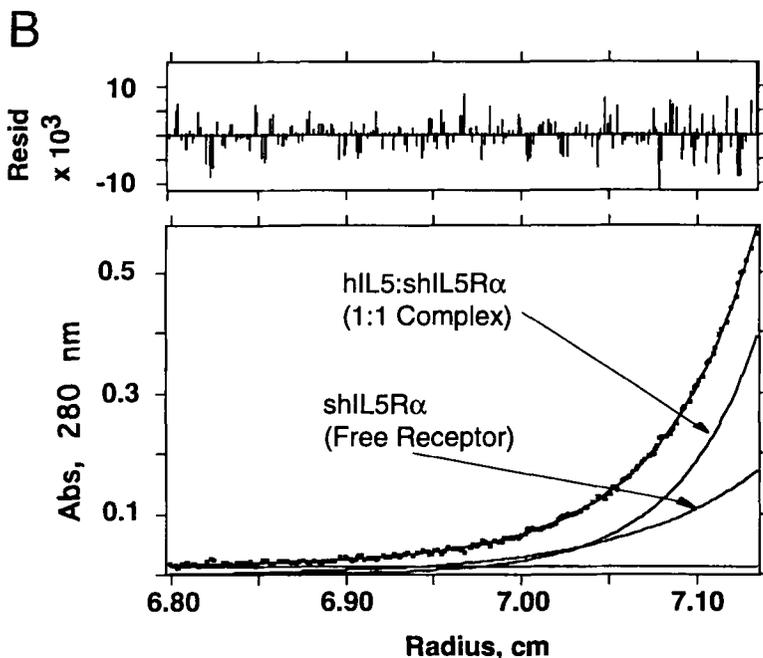
Figure 15. Continued.

forms as well as for the self-association equilibria themselves. However, in practice it is a considerable experimental challenge and may require independent measurements of self-association equilibria in the presence and absence of ligand, as well as high-precision ligand binding curves over a range of protein concentration (Ackers and Halvorson, 1974; Koblan and Ackers, 1991).

It is essential, therefore, to establish whether such coupled reactions are occurring during protein-protein interaction studies, either as a way to verify that these complicating effects are absent, or as a tool to assist in deconvoluting the energetics of ligand binding and self-association reactions. The method chosen for doing the self-association analyses depends on the concentration range that the binding studies are performed under. ITC binding studies, for example, are usually performed with reactant concentrations near 100 uM in the syringe and 1 μM in the reaction cell. Hence, analytical ultracentrifugation, using optical detection, is a convenient and rigorous method for defining the self-association status of the reactants and products in an ITC experiment. Experiments conducted at concentrations in the nanomolar range or lower would require a more sensitive method for evaluating self-association status, such as AGPC. Studies by AGPC have been conducted at concentrations in the picomolar range for radiolabeled lambda phage

cI repressor (Beckett et al., 1991). For this lower concentration range, again using radiolabeled proteins, AUC may also be used (Attri and Minton, 1986, 1987; Chatelier and Minton, 1987; Darawshe et al., 1993; Darawshe and Minton, 1994). Using radiolabels as tracers, one can use this same approach to characterize self-association at very high (> 100 mg/ml) concentrations.

C.   Coupled Thermal Unfolding Reactions

Another assumption when measuring protein-protein association kinetics and thermodynamics is that the binding reaction is taking place between natively folded structures of reactants and products. Ideally reactants and products should be analyzed for their thermal stabilities in order to define the maximum temperature at which binding studies can be conducted. This is an especially important concern when ITC is being done at elevated temperatures. For high affinity interactions, one often must go to significantly elevated temperatures to find conditions where the complex is measurably dissociated.

Differential scanning calorimetry (DSC) (Chowdry and Cole, 1989) is the method of choice for evaluating the precise thermodynamic basis for a proteins stability because the unfolding thermodynamics can be measured directly. However, spectroscopic methods (Eftink, 1995) can also be highly informative with regard to understanding the thermodynamic mechanism of protein folding, but also tend to be more convenient to implement and require less protein sample. In a practical sense, when studying specific protein-protein interactions, one is primarily interested in verifying that the protein reactants are safely away from temperatures where thermal unfolding may occur. We have found circular dichroism (CD) to be a rapid and reliable method for this purpose. Figure 16 shows the temperature dependence of the CD signal at 218 nm for IL-5. The data indicate that binding interactions with IL-5 could be conducted at temperatures below 60 °C without the adverse effects of coupled unfolding reactions of the IL-5. Similar experiments with the IL-5 receptor indicated that unfolding occurred at considerably lower temperatures, so that ITC experiments were limited to 35 °C and lower.

D.   The Use of Independent Methods

The strategy of using multiple methods is fundamental to characterizing protein-protein interactions. Most, if not all, methods have some degree of inherent systematic error. For example, does covalent modification of the protein alter its functional properties? Is the signal used for monitoring binding linear over the range studied? It is often the case that measurements of binding equilibrium or kinetic parameters are done on the edge of detectability for a given technique. A series of control experiments are then required to assess the accuracy of the physical parameters being determined, which may be laborious, and may or may not be sufficient to rule out systematic errors. Hence, one gains a great deal of reassurance
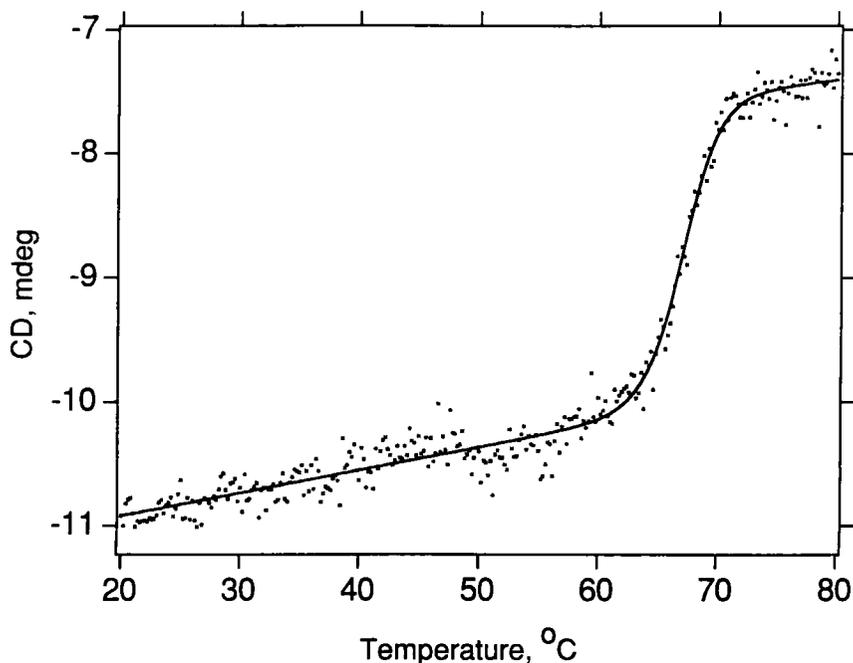
Figure 16. Thermal stability of IL-5 assessed by the temperature dependence of CD signal at 218 nm. Conditions were pH 7.4, 150 mM NaCl, 10 mM sodium phosphate, 0.27 mg/mL IL-5, and pathlength was 0.02 cm. Best-fit curve is for a two-state unfolding model (see Eftink, 1995 for a general discussion) with midpoint transition temperature of 67 °C and an unfolding enthalpy change of 144 kcal/mole. The equation used here was,

$$y_{obs} = \frac{(y_n^o + m_n T) - (y_u^o + m_u T)}{\left(1 + e^{\left(\frac{-\Delta H^o}{R}\left(\frac{1}{T} - \frac{1}{T_m}\right)\right)}\right)} + (y_u^o + m_u T)$$

where, y_{obs} is the observed CD, y^o_n and y_u^o are the CD values for the native and unfolded protein at zero °K, m_u and m_n are the slopes of the pre- and post transitional baselines and T_m is the temperature in °K at the midpoint of the transition. This analysis allows a useful estimate of the T_m and ΔH^o. In this analysis, ΔC was assumed to be zero. (Equation adapted from Hewlett Packard Application Note 12-5962-8098E, by Mathias Mücke and Franz-Xaver Schmidt, 1994.)

311

when multiple independent methods yield the same value for a specific parameter. A more obvious reason for using multiple methods is when the primary information gained from each method is mutually exclusive.

An illustrative example of the usefulness of the application of multiple methods to characterize a protein-protein interaction is the case of IL-8 dimerization. IL-8 is a chemokine involved in the inflammatory response and is an important agent in neutrophil activation (Herbert and Baker, 1993). The structure of IL-8 (Figure 17) was solved by NMR and x-ray crystallography and it was shown to be a dimer (Clore et al., 1989, 1990; Baldwin et al., 1991).

The dimeric IL-8 structure dominated thinking about the molecular basis for its mechanism of action for several years. As part of routine characterization of IL-8, SE experiments were performed to evaluate its quaternary structure. Instead of being a pure dimer, IL-8 was found to undergo a monomer-dimer equilibrium with a $K_d = 14$ μM at 25 °C (Figure 18).
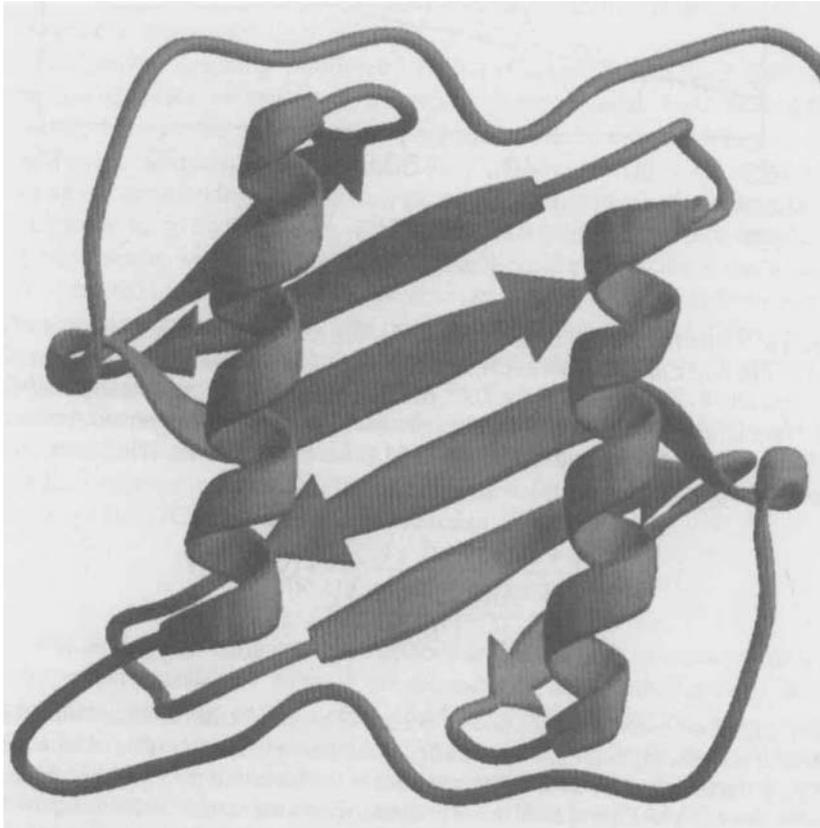
Figure 17. The dimeric structure of IL-8 as determined by NMR and X-ray crystallography. See Baldwin et al., 1991.
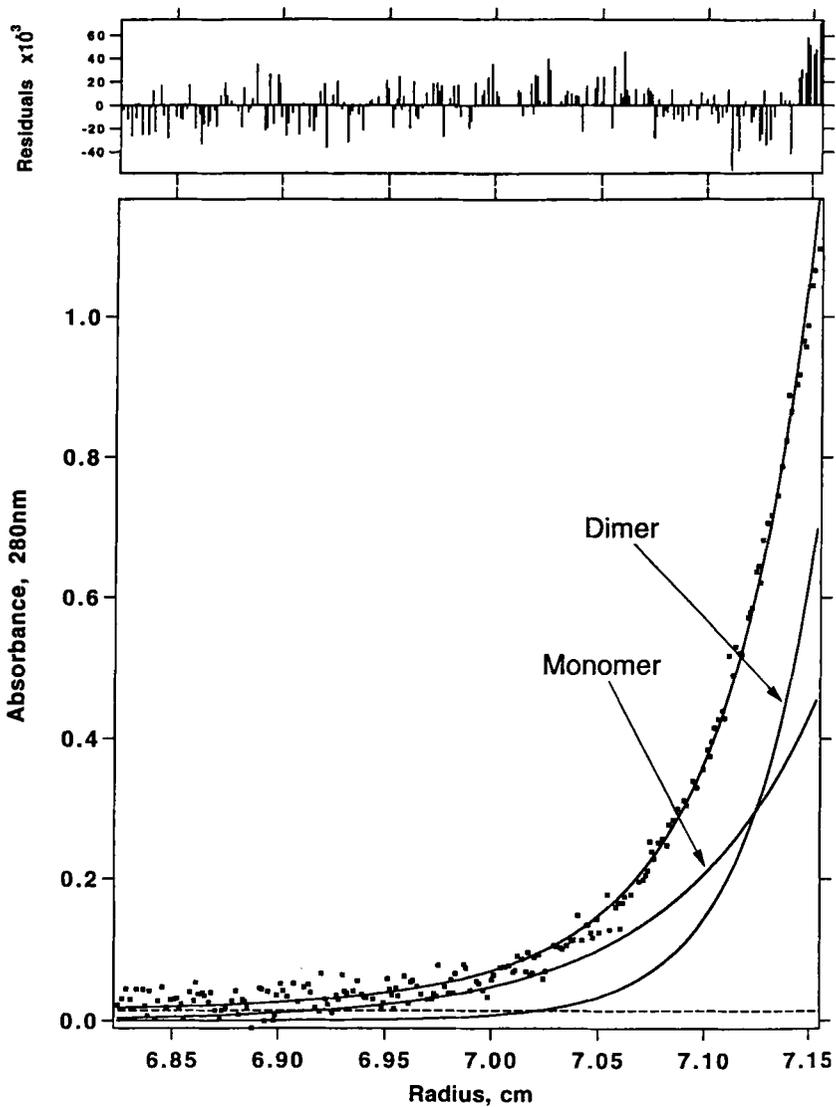
Figure 18. Fitting of IL-8 equilibrium sedimentation data to a monomer-dimer model. Panel **A**: residuals of the fit to the data. Panel **B**: fit of the primary data (45,000 rpm) to Equation 4. The total protein distribution is separated into components due to monomer and dimer. **Lower panel**: the fractions of monomer and dimer as a function of the total protein concentration as predicted from the determined value of K_d. As the physiological concentration of IL-8 is in the range of 10^{-8} M, the protein would be 100% monomer. From Burrows et al., 1994.
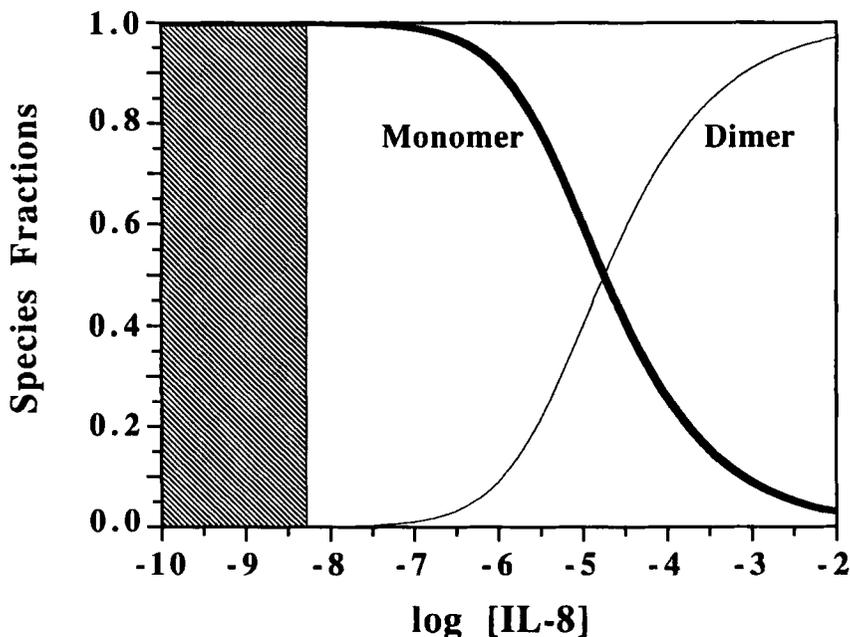
Figure 18. Continued.

These data demonstrated that the monomer was the form that existed at physiological concentrations (below 10^{-8} M; Figure 18) and implicated the monomer as the form which binds to the receptor. This was a paradigm changing observation, so it was necessary to carry out control experiments. Could it be that there was some artifact in the observation due to systematic error? This concern was amplified by the unusual observation that the dimerization K_d of IL-8 was largely insensitive to pH and salt concentration (Hensley and Doyle, unpublished observations).

But rather than focus exclusively on control experiments to assess possible systematic sources of error in the SE method, ITC experiments were carried out and proved to be an important means of verifying the SE results. Consecutive injections of concentrated IL-8 were made into the calorimeter which initially contained only buffer solution. Upon each injection, IL-8 was diluted from a concentration where it was 88% dimeric to a concentration where it was significantly dissociated to monomers. The observed heat at each injection was proportional to the fractional change in dimerization of IL-8 and the molar enthalpy change for dimerization. As the concentration of IL-8 increased with each injection, the fractional change in dimerization was reduced. The shape of the drop-off in heat per injection in Figure 19 is governed by the dimerization equilibrium constant

which was determined by nonlinear least squares analysis to be $K_d = 18 \, \mu M$ and was in excellent agreement with the 14 μM value determined by SE (Burrows et al., 1994).

In general the ITC data alone might not be able to distinguish a monomer-dimer equilibrium model from other self-association models. Hence, SE data played a key role in defining the model. The synergistic interplay between SE and ITC proved to be important for reliably establishing the model and verifying the results. Time resolved fluorescence anisotropy data were also consistent with a monomer-dimer equilibrium in the micromolar range (Burrows et al., 1994).

E.   Coupled Conformational Changes

Quantitation of the amount of conformational change which accompanies a protein-protein interaction is central to understanding the molecular mechanism of
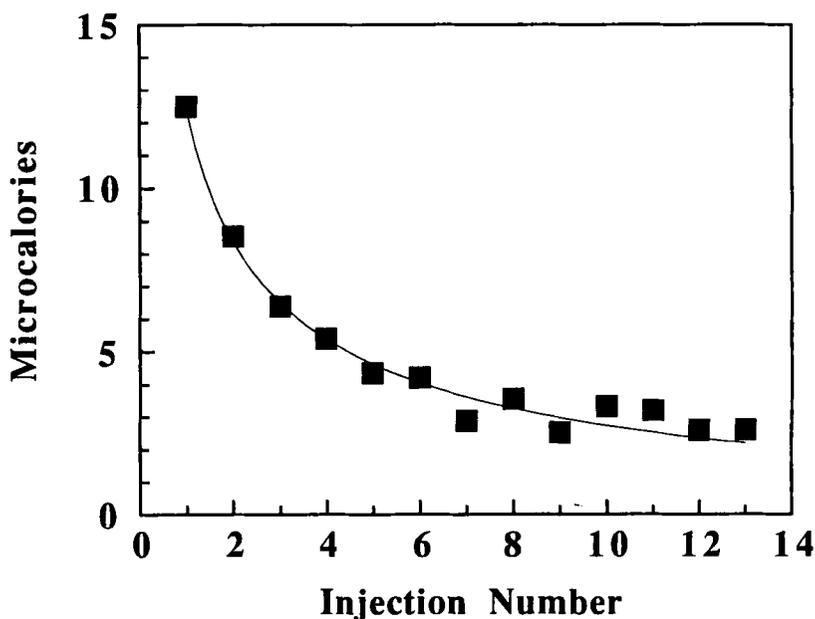
Figure 19.   The heats of dissociation of IL-8 dimer to monomers measured by ITC. **Upper panel**: the heat of IL-8 dimer dissociation as a function of injection number at 37 °C. Upon each injection of IL-8 into the calorimeter dissociation of the dimer to monomers was monitored by the accompanying uptake of heat. The concentration of IL-8 in the calorimeter increased with each injection and the extent of dissociation thus depends on concentration and K_d. Curve-fitting of the data gave $K_d = 18$ uM. **Lower panel**: the measured value of ΔH as a function of temperature. The slope of this line yields a ΔC of 185 ± 35 cal mol^{-1} K^{-1}. From Burrows et al., 1994.
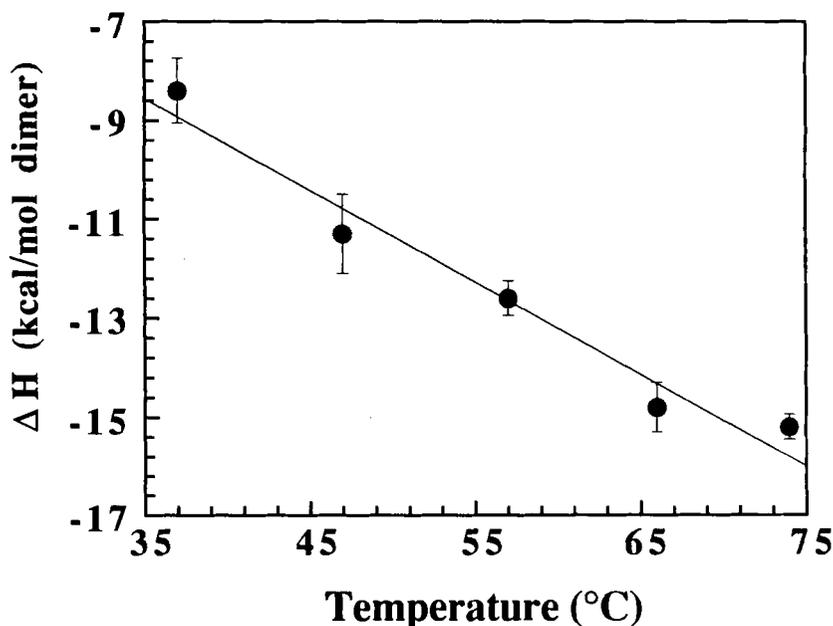
Figure 19. Continued.

action, but is difficult to assess without the 3-dimensional structures of the proteins individually and in complex. An emerging strategy for evaluating conformational changes in proteins comes from analyzing the thermodynamics of the interaction (Murphy and Freire, 1992; Spolar and Record, 1994). The quantitative correlation between binding thermodynamics and conformational change will continue to be refined as the thermodynamic and structural databases of protein-protein interactions expand. Nevertheless, there are presently a couple of methods available which are quite predictive. We describe briefly how these methods can be applied to protein-protein interactions. In the first example, conformational changes upon dimerization of IL-8 are implicated to be small. In the second example, evidence is given that considerable conformational rearrangement is associated with the binding interaction of IL-5 and its soluble α receptor.

The data in Figures 18 and 19 gave a free energy change for dimerization of IL-8 (from the familiar relationship $\Delta G = RT \ln K_d$) equal to -6.7 kcal/mol dimer at 37 °C. Further analysis of the binding thermodynamics was accomplished (Burrows et al., 1994) by measuring the dimerization enthalpy change as a function of temperature. The enthalpy showed a linear response with temperature (Figure 19), as expected for a simple protein-protein interaction over certain ranges of temperature. The slope of the enthalpy versus temperature plot in Figure 19 gives the dimerization heat capacity change (185 ± 35 cal mol^{-1} K^{-1}).

Hence a complete description of the dimerization thermodynamics for IL-8 was determined. Evaluation of conformational change upon dimerization was based on the approach of Murphy and Freire (1992) which is applied explicitly below in the case of IL-5 binding its receptor. Analysis of the dimerization thermodynamics of IL-8 indicated that the amount of surface area buried upon dimerization was 1,300 $Å^2$. Because this value was the same as the surface area buried at the dimer interface of the 3-dimensional NMR/crystal structures, it was concluded that conformational changes in the monomer upon dimerization were undetectably small. This was a useful observation, as it meant that the tertiary structure of the monomer within the dimer 3-dimensional structure could be taken as a good model for the monomer free in solution. Consequently, this information could be used to interpret mutagenesis studies and other work aimed at understanding its molecular mechanism of action (Herbert et al., 1991).

A counter example wherein binding thermodynamics implicated considerable coupled conformational change is given by the interaction between IL-5 and its soluble α receptor (Johanson et al., 1995). IL-5 is a disulfide cross-linked homodimeric cytokine (Figure 20) which plays a prominent role in maturation, proliferation and activation of eosinophils and is implicated in the etiology of asthma. It exerts its effects by binding to a cell surface receptor composed of an α- and a β-subunit. Most of the free energy of binding to the receptor is due to association with the α-subunit, but the presence of β-subunit is needed for signal transduction
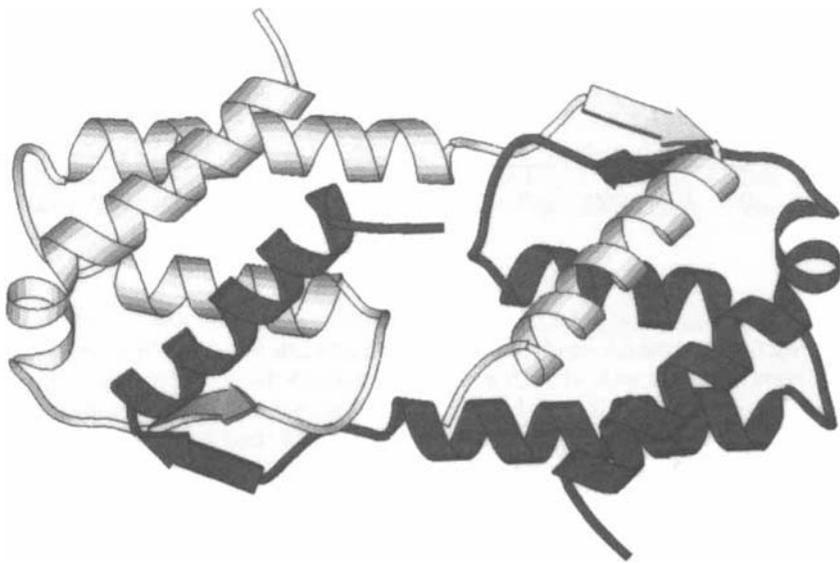
Figure 20. The three dimensional structure of IL-5 (see Milburn et al., 1993).
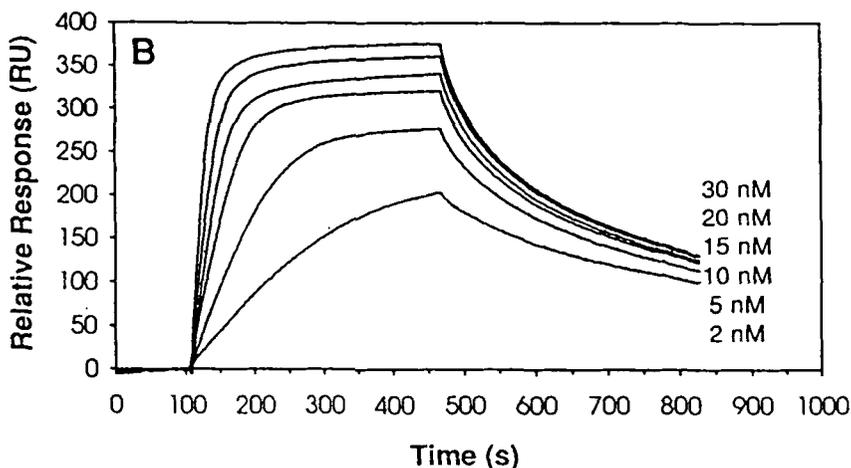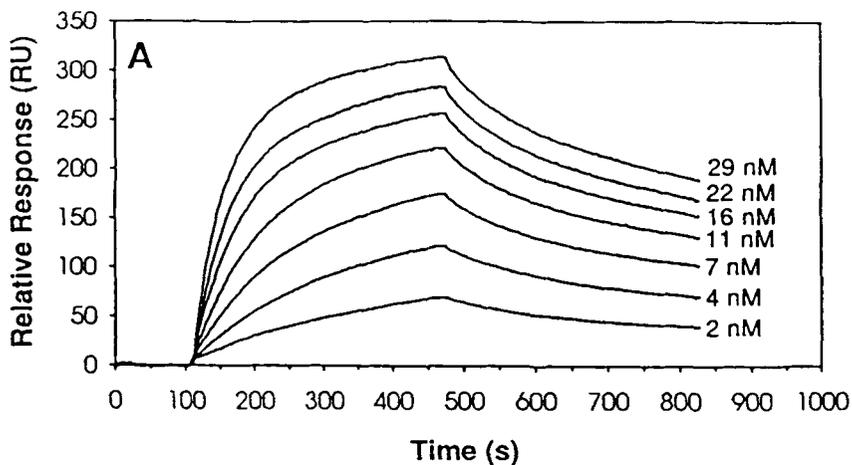
Figure 21. The time courses of the binding and dissociation of shIL5-Rα to / from hIL5. The numbers at the ends of each time course indicate the concentration of analyte (protein in solution). At 100 s the buffer is replaced by analyte. The increase in response shows the binding of analyte from solution to the immobilized ligand (the association phase). At 460 s the analyte injection ends and is replaced by buffer. The decay in the response represents the dissociation of the bound analyte (the dissociation phase) **Panel A**: overlays of the time courses for the binding and dissociation of shIL5-Rα to/from immobilized hIL5. **Panel B**: overlays of the time courses for the binding and dissociation of hIL5 to/from immobilized shIL5-Rα. From Morton et al., 1994.

(Tavernier et al., 1991). The molecular mechanism of IL-5-induced signal transduction is an area of current research interest both in terms of basic research and in drug discovery (Devos et al., 1995).

Figure 21 shows the association and dissociation kinetics for the interaction of IL-5 and its receptor, as monitored by SPR. Linear transformations of the association kinetics, according to Equations 19 and 20, are shown in Figure 22 and the deviations from linearity are indicative of heterogeneous rate processes. Rate constants determined from the linear portions of these data are summarized in Table 1. The association rate constant was estimated as $1 \times 10^6\,M^{-1}\,s^{-1}$ and the dissociation rate constant as $3 \times 10^{-3}\,s^{-1}$, yielding an overall K_d of 3×10^{-9} M. This kinetically derived dissociation constant agrees well with that determined by ITC.
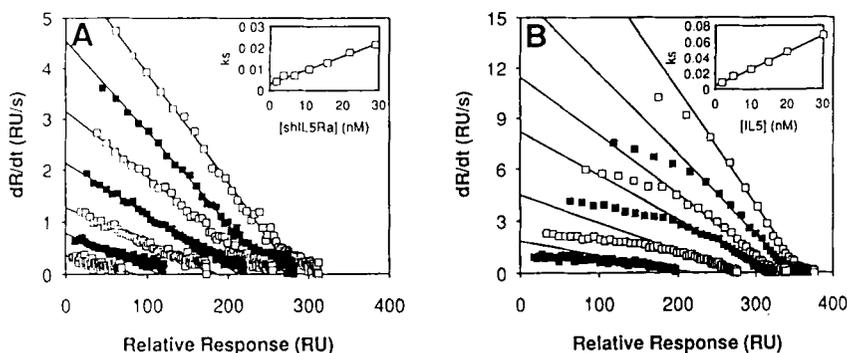
Figure 22. Analysis of the association phase of the data using a linear transform according to equation 12. The insets in the panels are replots of the slopes of the linearized data sets as a function of concentration according to equation 13. These replots are used to determine k_{on} and k_{off}. **Panel A**: overlays of the linearized time courses for the binding of shIL5-Rα to immobilized hIL5. **Panel B**: overlays of the linearized time courses for the binding of hIL5 to immobilized shIL5-Rα. From Morton et al., 1994.

Table 1. The Kinetics and Thermodynamics of Binding and Dissociation of shIL5-Rα to / From hIL5*

Ligand	Analyte	k_{on} $M^{-1}\,s^{-1} \times 10^{-6}$	k_{off} $s^{-1} \times 10^3$	$K_d{}^a$ $M \times 10^9$	$K_d{}^b$ $M \times 10^9$
hIL5	hIL5R-α	0.75 (0.45)[c]	2.8 (0.6)	4.8 (2.5)	3.1 (1.5)
hIL5R-α	hIL5	1.98 (0.28)	4.6 (0.50)	2.3 (0.2)	3.1 (1.5)

Notes: * From Morton et al., 1994.
[a] Calculated from k_{off} / k_{on}.
[b] Determined by titration calorimetry (Johanson et al., 1995).
[c] Values in parenthesis are standard errors.

One of several possible explanations for the heterogeneous kinetics seen with IL-5 and its receptor is a coupled conformational change (Morton et al., 1994). However, due to the complex nature of binding analytes to ligands which are attached to the SPR detector surface, there are other possible sources of heterogeneity that are unrelated to the native mechanism of the IL-5-receptor interaction (O'Shannessy and Winzor, 1996; Schuck, 1996). To further investigate whether conformational changes are coupled to the binding reaction of IL-5 and its receptor, it was therefore desirable to characterize the interaction from an independent biophysical method.

ITC has the potential for assessing coupled conformational changes based on analysis of the binding thermodynamics and, in particular, the binding entropy (Spolar and Record, 1994). However, as discussed above, for ITC to be useful in this regard, it must be independently shown that the only interaction that is being studied is the binding of IL-5 to its receptor, i.e., neither of these proteins self-associates in the concentration range of the ITC experiments. This was demonstrated by SE. Moreover, it was also shown (Figure 15) that the complex had a stoichiometry of 1:1 and that no further assembly of the complex occurred (Johanson et al., 1995).

ITC studies of the IL-5/receptor interaction yielded an equilibrium $K_d = 8 \pm 4$ nM and a molar ratio of reaction of 0.91 ± 0.10 IL-5 per receptor. These results are consistent with the SE results and together firmly establish the 1:1 stoichiometry. Moreover, they extend the characterization to define the affinity and thermodynamics of the interaction. Figure 23 shows ΔH measured as a function of temperature. The slope of this curve gives ΔC, the binding heat capacity change. These thermodynamic quantities are in turn valuable for assessing the magnitude of possible conformational changes coupled to binding.

Burial of apolar residues in protein-protein interactions has long been known to correlate to the heat capacity change of a given protein-protein interaction (Tanford, 1980; Privalov and Gill, 1988). Recent advances have refined correlations of buried surface area to heat capacity changes in proteins (Murphy and Freire, 1992; Spolar et al., 1992; Myers et al., 1995). Following the approach of Murphy and Freire (Murphy and Freire, 1992; Murphy et al., 1995) the experimental binding ΔC is related to changes in accessible surface area (ΔASA) of both apolar and polar residues as:

$$\Delta C = 0.45\ \Delta ASA_{apol} - 0.26\ \Delta ASA_{pol} \qquad (24)$$

Here ΔASA_{pol} is related to the experimentally deduced enthalpy change of the binding reaction at 100 °C (ΔH^{100}), where the contribution of apolar groups is believed to be negligibly small. The empirical relationship reported by Murphy and Friere (1992) is:

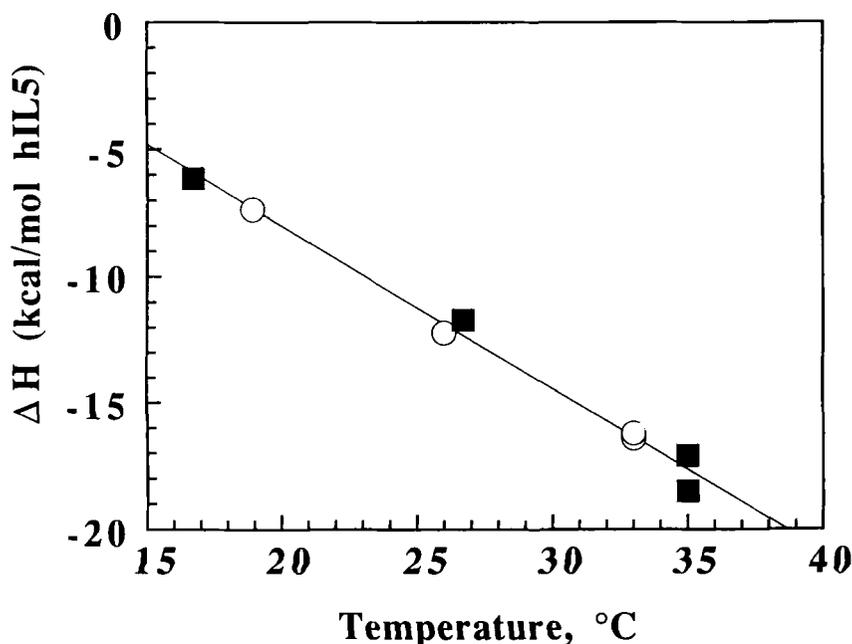$$\Delta ASA_{pol} = \Delta H^{100}/35 \qquad (25)$$

Figure 23. IL-5 binding enthalpy change of soluble IL5R-α versus temperature, as measured by ITC . From Johanson et al., 1995. Circles are for soluble IL-5Rα, squares are for IL-5Rα-Fc chimeric molecule. Slope of line equals the heat capacity change of -650 cal/degree/mol.

and ΔH^{100} is calculated from (Privalov and Gill, 1988):

$$\Delta H^{100} = \Delta H^{T_0} + \Delta C[100 - T_0]$$ (26)

ΔH^{T_0} is the binding enthalpy change measured at any reference temperature T_0. The amount of apolar and polar surface area buried when human IL-5 (hIL5) binds to shIL5R-α is calculated as 2,440 Å^2 and 1,720 Å^2, respectively, from equations 24 and 25. The total amount of buried surface is therefore estimated as 4,160 Å^2. The proportion of polar and apolar residues buried during the hIL5-shIL5R-α interaction are calculated from binding thermodynamics to be 41% and 59%, respectively, and are similar to the distribution of 55% apolar, 25% polar, and 20% charged reported by Janin and Chothia (1990) for typical protein-protein interfaces.

The very large value of 4160 Å^2 for the total amount of surface area buried during the binding of IL-5 to the shIL5R-α implies either an unusually large interfacial surface between the two proteins or structure changes coupled to the

binding reaction. If coupled structure changes are absent, the predicted interfacial area of the complex would be 2,080 Å^2, which is larger, by a factor of two, than that of typical protein-protein interfaces from the structural database reported for 15 crystallographic complexes (Janin and Chothia, 1990). The latter ranged from 625 to 975 Å^2. The largest side of the IL-5 dimer (Figure 20) exposes approximately 1,900 Å^2. Hence, unless that entire face is involved directly in receptor interaction, the surface burial analysis suggests that some type of conformational change occurs upon hIL5-receptor α chain interaction. Based on the likely rigidity of the IL-5 dimer expected from its structure (two four-helix bundles with exchanged helices D between the two bundle domains) large conformational changes in IL-5 seem unlikely (see Figure 18). Rather, the α-subunit receptor seems like a more plausible candidate for conformational change upon complex formation.

Further evidence for the possibility that conformational changes are coupled to the IL5-receptor binding reaction was provided from analysis of the binding entropy (Spolar and Record, 1994). Spolar and Record have demonstrated a strong empirical correlation between binding thermodynamics and coupled structure changes based on a comparative analysis of thermodynamic and structure databases for a series of protein-protein interacting systems where both complexed and free protein structures and corresponding binding thermo-dynamics are known. A signature for conformational changes coupled to bind-ing was deduced from analysis of the entropy of association for two macromolecules. The entropy of association is composed of several molecular sources and can be partitioned into entropy terms due to the hydrophobic effect (HE), rotational and translational degrees of freedom (rt) and other entropic effects, as (Spolar and Record, 1994):

$$\Delta S_{assoc} = \Delta S_{HE} + \Delta S_{rt} + \Delta S_{other} \tag{27}$$

Utilizing the general equation for calculating the temperature dependence of the entropy (Privalov and Gill, 1988)

$$\Delta S^{T} = \Delta S^{T_0} + \Delta C_p \ln\left(\frac{T}{T_0}\right) \tag{28}$$

where ΔS^{T_0} is the entropy at any reference temperature T_0, a characteristic temperature T_s can be computed where the total entropy of association ΔS_{assoc} is zero. At T_s Equation 27 becomes:

$$\Delta S_{HE} = \Delta S_{rt} + \Delta S_{other} \tag{29}$$

The entropic contribution due to changes in exposure of hydrophobic groups at T_s can be approximated by the following relation (Spolar and Record, 1994):

$$\Delta S_{HE} = 1.35 \Delta C_p \ln\left(\frac{T_s}{386}\right)$$ (30)

The rotational-translational entropy change upon association was estimated empirically as $\Delta S_{rt} = -50$ eu for typical macromolecular association processes. Hence, from these latter two terms ΔS_{other} may be determined. This term determines the extent of a potential conformational change. Spolar and Record report an empirical correlation between ΔS_{other} and the number of residues which fold upon binding, R^{th}, (i.e., $R^{th} = \Delta S_{other}/-5.6eu$).

For the case of IL-5 binding its α receptor a large unfavorable entropy change accompanies binding. Because it was shown that both IL-5 and its receptor did not undergo coupled self-association reactions (Figure 15), and also were thermally stable in the conditions of the ITC experiments (Figure 16), it is most likely that the unfavorable entropy change is due to a coupled conformational change. Following the analytical methods of Spolar and Record described above it was estimated that IL-5-receptor binding is accompanied by folding of 31 amino acid residues. Hence, from two independent interpretations of the binding thermodynamics, evidence for a conformational change being coupled to binding emerges.
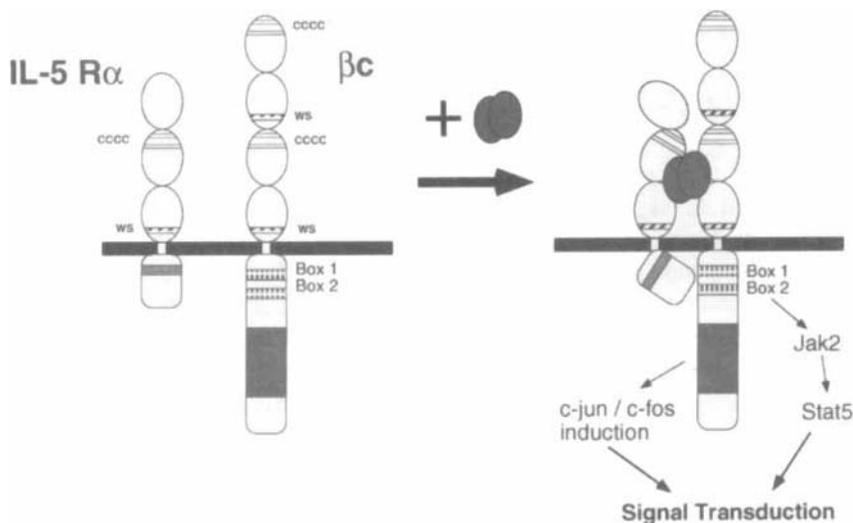
Figure 24. Model for IL-5 binding coupled conformational changes in the IL-5 receptor—implications for signal transduction. This model suggests how an IL5-binding-promoted structural transition in the α-subunit of the hIL5R may be coupled to signal transduction employing the β-subunit.

The conformational change may be in either IL-5, its receptor, or both. Arguments presented above suggest the majority of conformational change is the receptor (Figure 24).

IV.   STRATEGIES FOR MEASURING TIGHT BINDING

A.   Temperature Perturbation

Equilibrium affinity constants depend on the thermodynamic intensive variables of the experimental conditions, such as temperature and concentrations (chemical potentials) of solution components (pH, salt, etc.) (Wyman and Gill, 1990). Consequently, when faced with a protein-protein interaction which is too tight to measure under a specific set of solution conditions, a key strategy for determining the affinity is to change solution conditions until the affinity is weak enough to measure reliably. Correction of the measured affinity from the "weak affinity conditions" to the "tight affinity conditions" is often possible with exact thermodynamic principles.

One of the most sensitive thermodynamic variables for modulating binding affinity of protein-protein interactions is temperature. The temperature-dependence of an equilibrium constant, for example a binding equilibrium constant K^T, is given by (Wyman and Gill, 1990)

$$K^T = K^\circ * e\left(\frac{-\Delta H^\circ}{R}\left(\frac{1}{T} - \frac{1}{T_O}\right)\right) * e\left(\frac{\Delta C T_O}{R}\left(\frac{1}{T} - \frac{1}{T_O}\right)\right) * \left(\frac{T}{T_O}\right)^{\Delta C/R} \qquad (31)$$

where K° and ΔH° are the equilibrium constant and enthalpy change for the reaction at reference temperature T_0, ΔC is the heat capacity change (assumed to be constant), and R is the gas constant (1.987 cal/mol/degree). Thus, the binding affinity will decrease with increasing temperature for binding reactions which have an exothermic enthalpy change and the reverse for endothermic equilibrium reactions.

For protein-protein interactions, the binding enthalpies are usually exothermic. A simple strategy for measuring tight binding equilibrium constants is to perform ITC experiments as a function of increasing temperature until the equilibrium can be readily measured. Such was the case for the interaction of a soluble form of the T cell receptor CD4 with an anti-CD4 antibody (Doyle et al., 1996). Initial attempts were made to measure the affinity at 35 °C (Figure 25), but failed because the affinity was too tight. It could only be established that the affinity was sub-nanomolar. Since the binding enthalpy change was exothermic, ITC experiments were carried out at increasing temperatures until the affinity could be measured directly (at 45 °C K_d = 5 nM). The 45 °C affinity value was then corrected to a 25 °C value

with Equation 31 using the binding enthalpy and heat capacity changes measured experimentally (Figure 25).

An important assumption in the temperature-perturbation method is that there are no temperature-induced unfolding reactions. Whether or not this is valid can be tested in several ways. First, the thermal stabilities of both reactants can be evaluated
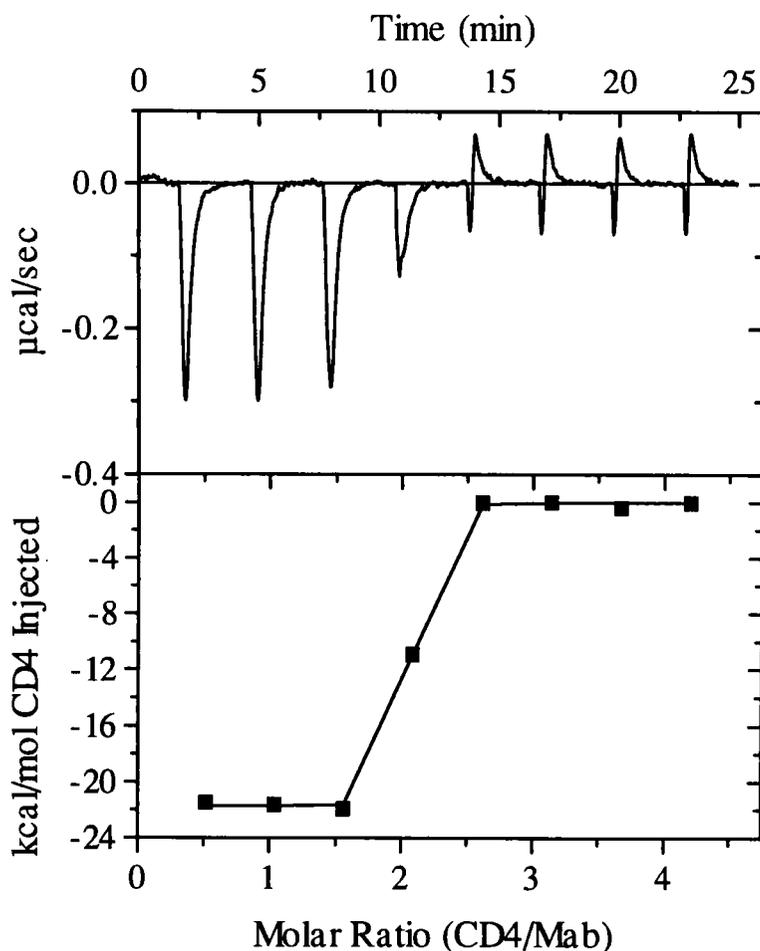

Figure 25. ITC data for sCD4-mAb binding reaction. Data were measured at 35 °C where the interaction was found to be too tight to measure an equilibrium binding affinity. Only a minimum affinity value of about 1 nanomolar could be established. ITC data for interactions which are too tight to measure have only one point in the transition region centered about the molar ratio equivalence point (for comparison to weaker binding data see Figure 9). See Figure 9 for general description of ITC data.

by either DSC or by monitoring the CD spectral signal versus temperature. In the case of CD4 and the mAb, the T_m values determined by DSC were 60 and 70 °C, respectively. Analysis of the DSC data demonstrated that there was no appreciable unfolding at temperatures up to about 45 °C, where the affinity was measured. Second, the binding stoichiometry and enthalpy change should be measured over the range of temperature which covers the temperature at which the affinity is determined and the temperature range where Equation 31 is to be applied. The stoichiometry should remain unchanged over the temperature range, and for a single equilibrium protein-protein binding reaction, a linear dependence of the binding enthalpy with temperature is expected (at least over the moderate temperature range from about 20-60 °C (Gomez et al., 1995)). Figure 25 shows a linear relationship for the CD4-antibody reaction and suggests that the binding reaction is not coupled to temperature-induced unfolding or self-association side reactions. A counter example has been reported for the interaction of ribonuclease S-protein with S-peptide (Thomson et al., 1994), where the binding reaction was coupled to folding of S-protein.
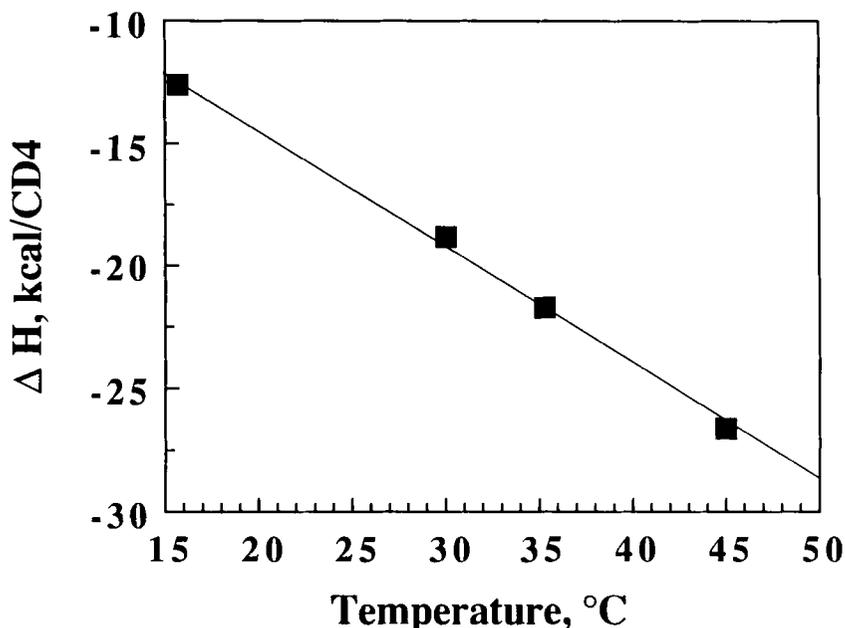
Figure 26. CD4 binding enthalpy change of anti-CD4 mAb versus temperature, as measured by ITC. Slope of line equals the heat capacity change of -480 cal/degree/mol CD4. From Doyle et al., 1996. Linearity of data suggests the absence of coupled side reactions such as temperature-induced global unfolding or self-association.

B.   pH Perturbation

One of the great virtues of thermodynamic state functions (including binding free energy changes) is that they are pathway independent. That is, they depend only on initial and final thermodynamic states (Nash, 1962). As such, experimental thermodynamic cycles may be constructed to enable indirect measurement of a change in a specific state function quantity which is difficult or impossible to measure directly.

Figure 27 depicts an example of a thermodynamic cycle relating binding free energies and protonation free energies of a macromolecule. Because proteins have ionizable groups which are often energetically linked to a protein-protein reaction, there is frequently a considerable pH dependence to the affinity of protein-protein interactions. Consequently, while the affinity may be too tight to measure at a
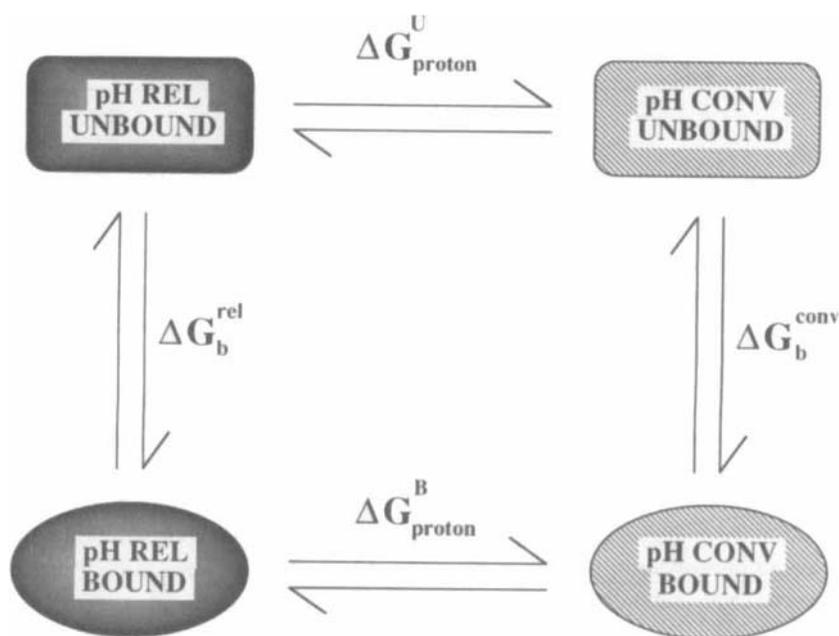
Figure 27.   Thermodynamic cycle linking ligand binding free energy changes (vertical equilibria) and protonation free energy changes (horizontal). pH conv refers to a pH condition where the ligand binding affinity can be conveniently measured, whereas pH rel is some relevant pH condition where the affinity may be too tight to measure directly. Hence, the relevant binding free energy change ΔG_b^{rel} can be determined by conservation around the thermodynamic cycle after experimental determination of the other three sides of the cycle. From Doyle et al., 1995.
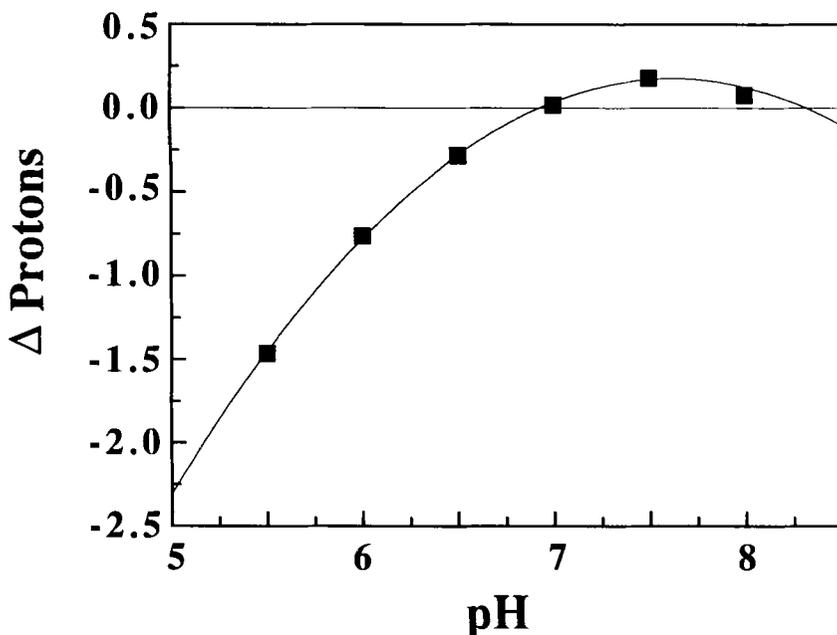
Figure 28. Experimental measures of protons coupled to CD4 binding reaction with anti-CD4 mAb. Protons taken up or released upon CD4-mAb binding were measured by titration calorimetry. Curve is an arbitrary function to allow for integration of the data. From Doyle et al., 1995.

specific, relevant pH, it might be easily measured at some other pH. Correction of the binding affinity from the convenient pH to the relevant pH can then be accomplished if the protonation free energies are known for the individual components. For the example given in Figure 27 this amounts to algebraic solution of the following conservation of energy equation, where three of four terms must be measured experimentally

$$\Delta G_b^{rel} + \Delta G^B_{proton} = \Delta G_b^{conv} + \Delta G^U_{proton} \tag{32}$$

Here ΔG_b^{rel} and ΔG_b^{conv} are ligand binding free energy changes in relevant and convenient pH conditions, and ΔG^B_{proton} and ΔG^U_{proton} are protonation free energy changes of the ligand bound and unbound macromolecule, respectively. The theoretical underpinnings from which the correction stems comes from work on the pH dependence for oxygen binding and subunit assembly of hemoglobin (Chu et al., 1984).

ITC can be used both to measure the affinity at a convenient pH and also to provide the necessary information on the protonation free energies (specifically, the difference in protonation free energies between bound and unbound forms (Doyle et al., 1995)). Figure 28 shows the number of protons which are thermodynamically coupled to binding CD4 to the anti-CD4 mAb. The coupled-protons in Figure 28 were measured by procedures which are described above (Figure 10, Equation 13). Integration of the data in Figure 28 gives the pH-dependence of binding affinity (Doyle et al., 1995). When combined with an independent measure of the affinity at any pH (i.e., $\Delta G_b{}^{conv}$ in Equation 32) within the range of data in Figure 28, the absolute affinity is determined over the entire pH range (Figure 29).

C. Differential Scanning Calorimetry

Any discussion of methods for measuring tight binding affinities would be incomplete without mention of DSC (For a review of DSC in biotechnology see
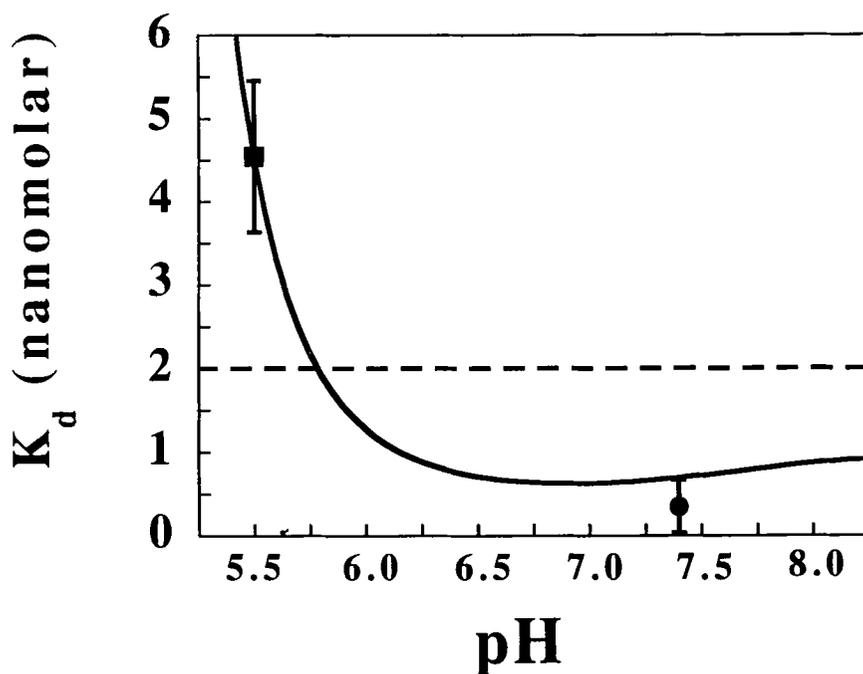
Figure 29. pH dependence of CD4 binding equilibrium with anti-CD4 mAb. Curve was determined by integration of the data in Figure 27. Square symbol is an experimentally determined equilibrium dissociation constant for the binding reaction measured by ITC. Circle was estimated by temperature dependence ITC studies. From Doyle et al., 1995.

Chowdry and Cole, 1989.) DSC has been demonstrated to be a useful method for determining binding affinities of relatively weak binding interactions such as cytodine-2′-monophosphate binding to ribonuclease A (Brandts and Lin, 1990; Straume and Freire, 1992). However, in principle DSC is also capable of estimating binding constants many orders of magnitude larger than most other methods can (Brandts and Lin, 1990).

Figure 30 shows simulated DSC curves for two proteins which were assumed to bind each other with various affinity constants. It can be seen that the thermal stabilities of the proteins are enhanced as the affinity constant increases, and that a wide range of affinities are measurable in the experimentally accessible temperature range presented. Brandts and Lin have analyzed DSC data from the literature on several protein-ligand or protein-protein interactions and have shown the DSC estimates of the binding affinities to compare favorably with literature values. However, for some cases the DSC estimate did not agree with the literature value for reasons unknown. Presently the quantitative reliability of the DSC method for measuring tight binding affinities is under investigation.
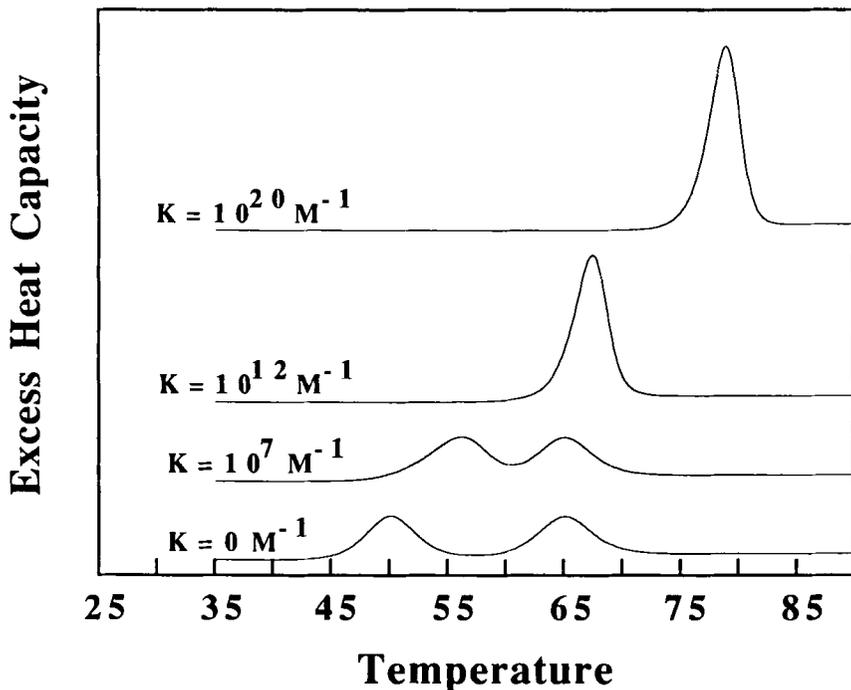
Figure 30. Simulated DSC curves for thermal unfolding of two proteins which are assumed to interact with various binding equilibrium constants as labeled. Redrawn from Brandts and Lin, 1990. It can be seen that extremely tight binding affinities are in principle measurable by DSC.

V. CONCLUDING REMARKS

The above discussion presents a practical guide for using several key biophysical methods to elucidate the molecular-level structural and functional properties of protein-protein interactions. We have seen that quantitative and self-consistent descriptions of molecular recognition events are possible using these approaches. We have also pointed out the synergistic relationship that exists when using multiple independent methods. Establishing the molecular-level functional properties for a given protein interaction (e.g., intrinsic affinity) usually requires the use of multiple methods not only to obtain mutually exclusive information, but also to serve as a powerful control strategy to assure accurate interpretation of the data. As more systems are characterized with multiple biophysical approaches, it is hoped that not only will we bring more clarity to our understanding of basic biological processes but that, in turn, these studies will help us to be more critical of and improve our analytical methods.

ACKNOWLEDGMENTS

REFERENCES

Ackers, G. K. (1970). Analytical gel chromatography of proteins. Adv. Protein Chem. 24, 343-446.

Ackers, G. K. (1975). In: Molecular Sieve Methods of Analysis. Academic Press, Inc. New York.

Ackers, G. K., & Halvorson, H. R. (1974). The linkage between oxygenation and subunit dissociation in human hemoglobin. Proc. Nat. Acad. Sci. USA 71, 4312-4316.

Aloj, S. M., Edelhoch, H., Ingham, K. C., Morgan, F. J., Canfield, R. E., & Ross, G. T.(1973). The rates of dissociation and reassociation of the subunits of human chorionic gonadotropin. Arch. Biochem. Biophys. 159, 497-504.

Arakawa, T., Wen, J., & Philo, J. S. (1994). Stoichiometry of heparin binding to basic fibroblast growth factor. Arch. Biochem. Biophys. 308, 267-273.

Attri, A. K., & Minton, A. P.(1986). Technique and apparatus for automated fractionation of the contents of small centrifuge tubes: Application to analytical ultracentrifugation. Analyt. Biochem. 152, 319-328.

Attri, A. K., & Minton, A. P.(1987). Simultaneous determination of the individual concentration gradients of two solute species in a centrifuged mixture: application to analytical ultracentrifugation. Analyt. Biochem. 162, 409-419.

Baldwin, E. T., Weber, I. T., St. Charles, R., Xuan, J. C., Appella, E., Yamada, M., Matsushima, K., Edwards, B. F., Clore, G. M., Gronenborn, A. M., & Wlodawer, A. (1991). Crystal structure of interleukin 8: Symbiosis of NMR and crystallography. Proc. Nat. Acad. Sci. USA 88, 502-506.

Beckett, D., Koblan, K. S., & Ackers G. K. (1991). Quantitative study of protein association at picomolar concentrations: The lambda phage cI repressor. Analyt. Biochem. 196, 69-75.

Bhatnagar, R. S., & Gordon, J. 1. (1995). Thermodynamic studies of myristoyl-CoA: Protein N-myristoyltransferase using isothermal titration calorimetry. Meth. Enzymol. 250, 467-486.

Biltonen, R. L., & Langerman, N. (1979). Microcalorimetry for biological chemistry: Experimental design, data analysis, and interpretation. Meth. Enzymol. 61, 287-318.

Brandts, J. F. (1993). In: ITC Data Analysis in Origin: Tutorial Guide. MicroCal, Amherst.

Brandts, J. F., & Lin, L. N. (1990). Study of strong to ultratight protein interactions using differential scanning calorimetry. [Review]. Biochemistry 29, 6927-6940.

Brooks, I., Watts, D. G., Soneson, K. K., & Hensley. P. (1994a). Determining confidence intervals for parameters derived from analysis of equilibrium analytical ultracentrifugation data. Meth. Enzymol. 240, 459-478.

Brooks, I., Wetzel, R., Chan, W., Lee, G., Watts, D. G., Soneson K. K., & Hensley, P. (1994b). Association of REI immunoglobulin light chain VL domains: The functional linearity of parameters in equilibrium analytical ultracentrifuge models for self-associating systems. In:Modern Analytical Ultracentrifugation Schuster, T. M., & Laue, T. M., eds.), pp.15-36, Birkhäuser, Boston.

Bundle, D. R., & Sigurskjold, B. W. (1994). Determination of accurate thermodynamics of binding by titration microcalorimetry. Meth. Enzymol. 247, 288-305.

Burrows, S. D., Doyle, M. L., Murphy, K. P., Franklin, S. G., White, J. R., Brooks, I., McNulty, D. E., Scott, M. O., Knutson, J. R., Porter, D., Young P. R., & Hensley P. (1994). Determination of the monomer-dimer equilibrium of interleukin-8 reveals it is a monomer at physiological concentrations. Biochemistry 33, 12741-12745.

Cantor, C. R., & Schimmel, P. R. (1980). In: Biophysical Chemistry Part III: The Behavior of Biological Macromolecules. W. H. Freeman, San Francisco.

Carr, S. A., Hemling, M. E., Bean, M. F., & Roberts, G. D. (1991). Integration of mass spectrometry in analytical biotechnology. [Review]. Analyt. Chem. 63, 2802-2824.

Chait, B. T., & Kent, S. B. H. (1992). Weighing naked proteins: Practical, high accuracy mass measurements of peptides and proteins. Science 257, 1885-1894.

Chan, W., Helms, L. R., Brooks, I., Lee, G., Ngola, S., Maleef, B., Hensley, P., & Wetzel, R. (1996). Mutations and inclusion body formation in the periplasmic expression of the immunoglobulin VL domain of REI. Protein Folding Design (in press).

Chatelier, R. C., & Minton, A. P. (1987). Sedimentation equilibrium in macromolecular solutions of arbitrary concentration. II. Two protein components. Biopolymers 26, 1097-1113.

Chowdry, B. Z., & Cole, S. C. (1989). Differential scanning calorimetry: Applications in biotechnology. Trends Biotech. 7, 11-18.

Christensen, J. J., Hansen, L. D., & Izatt, R. M. (1976). In: Handbook of Ionization Heats. Wiley, New York.

Chu, A. H., Turner, B. W., & Ackers, G. K. (1984). Effects of protons on the oxygenation-linked subunit assembly in human hemoglobin. Biochemistry 23, 604-617.

Clore, G. M., Appella, E., Yamada, M., Matsushima, K., & Gronenborn, A. M. (1989). Determination of the secondary structure of interleukin-8 by nuclear magnetic resonance spectroscopy. J. Biological Chem. 264, 18907-18911.

Clore, G. M., Appella, E., Yamada, M., Matsushima, K., & Gronenborn, A. M. (1990). Three-dimensional structure of interleukin 8 in solution. Biochemistry 29, 1689-1696.

Clore, G. M., & Gronenborn, A. M. (1995). Three-dimensional structures of alpha and beta chemokines. [Review]. FASEB J. 9, 57-62.

Covell, D. G., Smythers, G. W., Gronenborn, A. M., & Clore, G. M. (1994). Analysis of hydrophobicity in the alpha and beta chemokine families and its relevance to dimerization. Protein Science 3, 2064-2072.

Darawshe, S., & Minton, A. P. (1994). Quantitative characterization of macromolecular associations in solution via real-time and postcentrifugation measurements of sedimentation equilibrium: A comparison. Analy. Biochem. 220, 1-4.

Darawshe, S., Rivas, G., & Minton, A. P. (1993). Rapid and accurate microfractionation of the contents of small centrifuge tubes: Application in the measurement of molecular weight of proteins via sedimentation equilibrium. Analy. Biochem. 209, 130-135.

Devos, R., Plaetinck, G., Cornelis, S., Guisez, Y., Van, d. H. J., & Tavernier, J. (1995). Interleukin-5 and its receptor: A drug target for eosinophilia associated with chronic allergic disease. [Review]. J. Leukocyte Biology 57, 813-819.

Dodd, I., Mossakowska, D. E., Camilleri, P., Haran, M., Hensley, P., Lawlor, E. J., McBay, D. L., Pindar, W., & Smith, R. A. G. S. (1995). Overexpression in Escherichia coli, folding, purification and characterization of the first three short consensus repeat modules of human complement receptor-Type-1. Protein Expression Purification (in press).

Doyle, M. L., Brigham-Burke, M., Brooks, I. S., Blackburn, M. N., Smith, T. M., Sokoloski, T., Newman, R., Reff, M., Sweet, R. M., Truneh, A., Hensley, P., & OShannessy, D. J. (1997). The thermodynamics and kinetics of CD4 binding to IgG1 and IgG4 isotypes of a monoclonal antibody, in preparation.

Doyle, M. L., Gill, S. J., & Cusanovich, M. A. (1986). Ligand-controlled dissociation of chromatium vinosum cytochrome c'. Biochemistry 25, 2509-2516.

Doyle, M. L., Louie, G., Dal Monte, P. R., & Sokoloski, T. D. (1995). Tight binding affinities determined from thermodynamic linkage to protons by titration calorimetry. Methods in Enzymology 259, 183-194.

Doyle, M. L., Myszka, D. G., & Chaiken, I. (1996b). Molecular interaction analysis in ligand design using mass transport, kinetic, and thermodynamic methods. J. Molecular Recognition In press.

Eftink, M. R. (1995). Use of multiple spectroscopic methods to monitor equilibrium unfolding of proteins. Methods Enzymology 259, 487-512.

Eisenstein, E., & Schachman, H. K. (1989). Determining the roles of subunits in protein function. In:Protein Function: A Practical Approach., (Creighton, T. E., eds.), pp. 135-176, IRL Press, New York.

Fairman, R., Chao, H. G., Mueller, L., Lavoie, T. B., Shen, L. Y., Novotny, J., & Matsueda, G. R. (1995). Characterization of a new four-chain coiled-coil—influence of chain length on stability. Protein Science 4, 1457-1469.

Fischer, H. F., & Singh, N. (1995). Calorimetric methods for interpreting protein-ligand interactions. Methods Enzymology 259, 194-221.

Fisher, R. J., Fivash, M., Casas, F. J., Bladen, S., & Mcnitt, K. L. (1994a). Real-time BIAcore measurements of Escherichia coli single-stranded DNA binding (SSB) protein to polydeoxythymidylic acid reveal single-state kinetics with steric cooperativity. Methods (Orlando) 6, 121-133.

Fisher, R. J., Fivash, M., Casas, F. J., Erickson, J. W., Kondoh, A., Bladen, S. V., Fisher, C., Watson, D. K., & Papas, T. (1994b). Real-time DNA binding measurements of the ETS1 recombinant oncoproteins reveal significant kinetic differences between the p42 and p51 isoforms. Protein Science 3, 257-266.

Formisano, S., Johnson, M. L., & Edelhoch, H. (1977). Thermodynamics of the self-association of glucagon. Proc. Nat. Acad. Sci. USA 74, 3340-3344.

Freire, E. (1993). Structural thermodynamics: Prediction of protein stability and protein binding affinities. Arch. Biochem. Biophys. 303, 181-184.

Freire, E., Mayorga, O. L., & Straume, M. (1990). Isothermal titration calorimetry. Analyt. Chem. 62, 950A-959A.

Gill, S. C., & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. Anal. Biochem. 182, 319-326.

Gomez, J., Hilser, V. J., Xie, D., & Freire, E. (1995). The heat capacity of proteins. Proteins: Structure, Function Genetics 22, 404-412.

Hansen, J. C. (1994). Analysis of structural changes in steroid receptor proteins by partitioning. Methods Enzymology 228, 276-86.

Hansen, J. C., Lebowitz, J., & Demeler, B. (1994). Analytical ultracentrifugation of complex macromolecular systems. [Review]. Biochemistry 33, 13155-13163.

Hansen, J. C., & Wolffe, A. P. (1994). A role for histones H2A/H2B in chromatin folding and transcriptional repression. Proc. Nat. Acad. Sci. the USA 91, 2339-2343.

Harding, S. E., Rowe, A. J., & Horton, J. C., eds. (1992). In: Analytical Ultracentrifugation in Biochemistry and Polymer Science. The Royal Society of Chemistry, Cambridge.

Herbert, C. A., & Baker, J. B. (1993). Interleukin-8: A review. Cancer Investigation 11, 743-750.

Herbert, C. A., Vitangcol, R. V., & Baker, J. B. (1991). Scanning mutagensis of interleukin-8 identifies a cluster of residues required for receptor binding. J. Biol. Chem. 266, 18989-18994.

Ingham, K. C., Aloj, S. M., & Edelhoch, H. (1974). Rates of dissociation and recombination of the subunits of bovine thyrotropin. Arch. Biochem. Biophys 163, 589-599.

Ingham, K. C., Saroff, H. A., & Edelhoch, H. (1975). Ligand-induced self-association of human chorionic gonadotropin. Positive cooperativity in the binding of 8-anilino-1-naphthalenesulfonate. Biochemistry 14, 4751-4758.

Janin, J., & Chothia, C. (1990). The structure of protein-protein recognition sites. [Review]. J. Biol. Chem. 265, 16027-16030.

Johanson, K., Appelbaum, E., Doyle, M., Hensley, P., Zhao, B., Abdel, M. S., Young, P., Cook, R., Carr, S., Matico, R., McDonnell, P., Morton, T., Bennett, D., Sokoloski, T., McNulty, D., Rosenberg, M., & Chaiken, I. (1995). Binding interactions of human interleukin 5 with its receptor alpha subunit. Large scale production, structural, and functional studies of Drosophila-expressed recombinant proteins. J. of Biol. Chem. 270, 9459-9471.

Johnson, M. L., Formisano, S., & Edelhoch, H. (1978). Self-association of glucagon as measured by the optical properties of rhodamine 6G. J. Biol Chem 253, 1353-1356.

Jones, S., & Thornton, J. M. (1995). Protein-protein interactions: a review of protein dimer structures. [Review]. Prog. in Biophys. Mol. Biol. 63, 31-65.

Jonsson, U., Fagerstam, L., Ivarsson, B., Johnsson, B., Karlsson, R., Lundh, K., Lofas, S., Persson, B., Roos, H., & Ronnberg, I. (1991). Real-time biospecific interaction analysis using surface plasmon resonance and a sensor chip technology. Biotechniques 11, 620-627.

Jonsson, U., Fagerstam, L., Lofas, S., Stenberg, E., Karlsson, R., Frostell, A., Markey, F., & Schindler, F. (1993). Introducing a biosensor based technology for real-time biospecific interaction analysis. Ann. Biol. Clin. 51, 19-26.

Karlsson, R. (1994). Real-time competitive kinetic analysis of interactions between low-molecular-weight ligands in solution and surface-immobilized receptors. Anal. Biochem. 221, 142-151.

Koblan, K. S., & Ackers, G. K. (1991). Cooperative protein-DNA interactions: Effects of KCl on lambda cI binding to OR. Biochemistry 30 ,7822-7827.

Ladbury, J., & Peters, R. (1994). Turning up the heat on rational drug design: using enthalpy to probe binding interactions may put companies on the drug development fast track. Bio/Technology 12, 1083-1085.

Laue, T. M., Shah, B. D., Ridgeway, T. M., & Pelletier, S. L. (1992). Computer-aided interpretation of analytical sedimentation data for proteins. In:Analytical Ultracentrifugation in Biochemistry and Polymer Science (Harding, Science S. E., Rowe, A. J., & Horton, J. C. eds.), pp. 90-125. The Royal Society of Chemistry, Cambridge.

Lee, F. S., Auld, D. S., & Vallee, B. L. (1989a). Tryptophan fluorescence as a probe of placental ribonuclease inhibitor binding to angiogenin. Biochemistry 28, 219-224.

Lee, F. S., Shapiro, R., & Vallee, B. L. (1989b). Tight binding inhibition of angiogenin and ribonuclease by placental ribonuclease inhibitor. Biochemistry 28, 225-230.

Malmqvist, M. (1993). Surface plasmon resonance for detection and measurement of antibody-antigen affinity and kinetics. [Review]. Curr. Opin. Immunol 5, 282-286.

Malmqvist, M., & Granzow, R. (1994). Biomolecular interaction analysis. Methods (Orlando) 6, 95-98.

McGrath, M. E., Gillmor, S. A., & Fletterick, R. J. (1995). Ecotin: Lessons on survival in a protease-filled world. Protein Science 4, 141-148.

McKinnon, I. R., Fall, L., Parody-Morreale, A., & Gill, S. J. (1984). A twin titration microcalorimeter for the study of biochemical reactions. Analy. Biochem. 139, 134-139.

Milburn, M. V., Hassell, A. M., Lambert, M. H., Jordan, S. R., Proudfoot, A. E., Graber, P., & Wells, T. N. (1993). A novel dimer configuration revealed by the crystal structure at 2.4 A resolution of human interleukin-5. Nature 363, 172-176.

Morelock, M. M., Ingraham, R. H., Betageri, R., & Jakes, S. (1995). Determination of receptor-ligand kinetic and equilibrium binding constants using surface plasmon resonance: Application to the lck SH2 domain and phosphotyrosyl peptides. J. Med. Chem. 38, 1309-1318.

Morton, T. A., Bennett, D. B., Appelbaum, E. R., Cusimano, D. M,. Johanson, K. O., Matico, R. E., Young, P. R., Doyle, M., & Chaiken, I. M. (1994). Analysis of the interaction between human interleukin-5 and the soluble domain of its receptor using a surface plasmon resonance biosensor. J. Mol Recognition 7, 47-55.

Morton, T. A., Myszka, D. G., & Chaiken, I. M. (1995). Interpreting complex binding kinetics for optical biosensors: A comparison of analysis by linearization, the integrated rate equation and numerical integration. Anal. Biochem. 227, 176-185.

Muramatsu, N., & Minton, A. P. (1988). An automated method for rapid determination of diffusion coefficients via measurements of boundary spreading. Anal. Biochem. 168, 345-351.

Murphy, K. P., & Freire, E. (1992). Thermodynamics of structural stability and cooperative folding behavior in proteins. [Review]. Advances Protein Chemistry 43, 313-61.

Murphy, K. P., Freire, E., & Paterson, Y. (1995). Configurational effects in antibody-antigen interactions studied by microcalorimetry. Proteins: Structure, Function, Genetics 21, 83-90.

Murphy, K. P., & Gill, S. J. (1989). Thermodynamics of dissolution of solid cyclic dipeptides containing hydrophobic side groups. J. Chem. Thermodynamics 21, 903-913.

Murphy, K. P., Xie, D., Garcia, K. C., Amzel L. M., & Freire, E. (1993). Structural energetics of peptide recognition: Angiotensin II/antibody binding. Proteins: Structure, Function, Genetics 15, 113-120.

Myers, J. K., Pace, C. N., & Scholtz, J. M. (1995). Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. Protein Science 4, 2138-2148.

Nash, L. K. (1962). In: Elements of Chemical Thermodynamics. Addison-Wesley Publishing Company, Reading.

Nenortas, E., & Beckett, D. (1994). Reduced-scale large-zone analytical gel filtration chromatography for measurement of protein association equilibria. Anal. Biochem. 222, 366-373.

O'Shannessy, D. J. (1994). Determination of kinetic rate and equilibrium binding constants for macromolecular interactions: A critique of the surface plasmon resonance literature. Curr. Opin. Biotechnology 5, 65-71.

O'Shannessy, D. J., Brigham-Burke, M., Soneson, K. K., Hensley, P., & Brooks, I. (1993). Determination of rate and equilibrium binding constants for macromolecular interactions using surface plasmon resonance: use of nonlinear least squares analysis methods. Anal. Biochem. 212, 457-468.

O'Shannessy, D. J., & Winzor, D. J. (1996). Interpretation of Deviations from Pseudo-First Order Kinetic Behavior in the Characterization of Ligand Binding by Biosensor Technology. (in press).

Otto-Bruc, A., Anthony, B., Buong, T. M., Chardin, P., & Chabre, M. (1993). Interaction between the retinal cyclic GMP Phosphodiesterase Inhibitor and Transducin. Kinetics and affinity studies. Biochemistry 32, 8636-8645.

Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., & Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. Protein Science 4, 2411-2423.

Padlan, E. A. (1994). Anatomy of the antibody molecule. [Review]. Molecular Immunology 31, 169-217.

Parody-Morreale, A., Robert, C. H., Bishop, G. A., & Gill, S. J. (1987). Calorimetric studies of oxygen and carbon monoxide binding to human hemoglobin: Sequential binding heats for oxygen. J. Biol. Chem. 262, 10994-10999.

Philo, J. S. (1994). Measuring sedimentation, diffusion, and molecular weights of small molecules by direct fitting of sedimentation velocity concentration profiles. In:Modern Analytical Ultracentrifugation (Schuster, T. M., & Laue, T. M. eds.), pp. 156-170. Birkhäuser, Boston.

Phizicky, E. M., & Fields, S. (1995). Protein-protein interactions: methods for detection and analysis. [Review]. Microbiological Reviews 59, 94-123.

Privalov, P. L., & Gill, S. J. (1988). Stability of protein structure and hydrophobic interaction. Advances Protein Chemistry 39, 191-234.

Roberts, G. D., Johnson, W. P., S?., B., Anumula, K. R., & Carr, S. A. (1995). An integrated strategy of structural characterization of the protein and carbohydrate components of monoclonal antibodies: Application to anti-respiratory syncytial virus MAb. Anal. Chem. 67, 3613-3625.

Schuck, P. (1996). Kinetics of ligand binding to receptor immobilized in a polymer matrix, as detected with an evanescent wave biosensor. I. A computer simulation of the influence of mass transport. Biophysical J. (in press).

Schuster, T. M., & Laue, T. M. (1994). In: Modern Analytical Ultracentrifugation. Birkhäuser, Boston.

Schwarz, P. M., & Hansen, J. C. (1994). Formation and stability of higher order chromatin structures. Contributions of the histone octamer. J. Biol. Chem. 269, 16284-16289.

Shapiro, R., & Vallee, B. L. (1991). Interaction of human placental ribonuclease with placental ribonuclease inhibitor. Biochemistry 30, 2246-2255.

Spolar, R. S., Livingstone, J. R., & Record, M. J. (1992). Use of liquid hydrocarbon and amide transfer data to estimate contributions to thermodynamic functions of protein folding from the removal of nonpolar and polar surface from water. Biochemistry 31, 3947-3955.

Spolar, R. S., & Record, M. T. J. (1994). Coupling of local folding to site-specific binding of proteins to DNA. Science 263, 777-784.

Stafford, W. F., III (1992a). Boundary analysis in sedimentation transport experiments: a procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile. Anal. Biochem. 203, 295-301.

Stafford, W. F., III (1992b). Methods of obtaining sedimentation coefficient distributions. In:Analytical Ultracentrifugation in Biochemistry and Polymer Science (Harding, S. E., Rowe, A. J., & Horton, J. C., eds.), pp. 359-393. The Royal Society of Chemistry, Cambridge. UK.

Stafford, W. F., III (1994a). Boundary analysis in sedimentation velocity experiments. Methods Enzymology 240, 478-501.

Stafford, W. F., III (1994b). Sedimentation boundary analysis of interacting systems: Use of the apparent sedimentation coefficient distribution function. In: Modern Analytical Ultracentrifugation (Schuster, T. M., & Laue, T. M., eds.), pp.119-137. Birkhäuser, Boston.

Stafford, W. F., III (1996). Determination of the translational diffusion coefficient from the standard error of a Gaussian fit to g(s*) transformed sedimentation velocity data. Protein Science (in press).

Stafford, W. F., III, Silverman, C. C., McNulty, D. E., Doyle, M. L., Myszka, D. G., Porter, T. G., & Hensley, P. (1996). The Binding of an Fab to dimeric Interleukin-5: A solution interaction analysis using the time derivative method to interpret sedimentation velocity data. Protein Science (in press).

Stoscheck, C. M. (1990). Quantitation of protein. Methods Enz. 182, 50-68.

Straume, M., & Freire, E. (1992). Two-dimensional differential scanning calorimetry: Simultaneous resolution of intrinsic protein structural energetics and ligand binding interactions by global linkage analysis. Anal. Biochem. 203, 259-268.

Strydom, D. J., Tarr, G. E., Pan, Y. E., & Paxton, R. J. (1992). Collaborative Trial Analyses of ABRF-91AAA. Academic Press, San Diego.

Suzdak, G. (1994). The emergence of mass spectrometry in biochemical research. Proc. Nat. Acad. Sci. USA 91, 11290-11297.

Svedberg, T., & Pederson, K. O. (1959). In: The Ultracentrifuge. Johnson Reprint Corporation, New York.

Tanford, C. (1980). The Hydrophobic Effect, Second Edition. Wiley-Interscience, New York,.

Tavernier, J., Devos, R., Cornelis, S., Tuypens, T., Van, d. H. J., Fiers, W., & Plaetinck, G. (1991). A human high affinity interleukin-5 receptor (IL5R) is composed of an IL5-specific alpha chain and a beta chain shared with the receptor for GM-CSF. Cell 66, 1175-1184.

Thomson, J., Ratnaparkhi, G. S., Varadarajan, R., Sturtevant, J. M., & Richards, F. M. (1994). Thermodynamic and structural consequences of changing a sulfur atom to a methylene group in the M13Nle mutation in ribonuclease-S. Biochemistry 33, 8587-8593.

Tinoco, J., I., Sauer, K., & Wang, J. C. (1985). In: Physical Chemistry, Principles and Applications in Biological Sciences, 2nd Ed. Prentice Hall, Englewood Cliffs.

Todd, G. P., & Haschemeyer, R. H. (1981). General solution to the inverse problem of the differential equation of the ultracentrifuge. Proc. Nat. Acad. Sci. USA 78, 6739-6743.

Turner, G. J. (1990). In: Using Mutant and Chemically-Modified Hemoglobins to Probe the Molecular Mechanism of Cooperativity. Johns Hopkins University, Baltimore.

Turner, G. J., Galacteros, F., Doyle, M. L., Hedlund, B., Pettigrew, D. W., Turner, B. W., Smith, F. R., Moo, P. W., Rucknagel, D. L., & Ackers, G. K. (1992). Mutagenic dissection of hemoglobin cooperativity: Effects of amino acid alteration on subunit assembly of oxy and deoxy tetramers. [Review]. Proteins 14, 333-350.

Valdes, R., & Ackers, G. K. (1979). Study of protein subunit association equilibria by elution gel chromatography. Meth. Enzymol 61, 125-141.

van Holde, K. E., & Weischet, W. O. (1978). Boundary analysis of sedimentation velocity experiments with monodisperse and paucidisperse solutes. Biopolymers 17, 1387-1403.

Ward, L. D., Howlett, G. J., Hammacher, A., Weinstock, J., Yasukawa, K., Simpson, R. J., & Winzor, D. J. (1995). Use of a biosensor with surface plasmon resonance detection for the determination of binding constants: measurement of interleukin-6 binding to the soluble interleukin-6 receptor. Biochemistry 34, 2901-2907.

Wen, J., Arakawa, T., Talvenheimo, J., Welcher, A. A., Horaan, T., Tseng, J., Nicholson, M., & Phila, J. (1996). A light scattering/size exclusion chromatography method for studying the stoichiometry of a protein-protein complex. In:Techniques in Protein Chemistry (Crabb, J. W., eds.), Academic Press, San Diego. (in press).

Wensel, T. G., & Stryer, L. (1990). Activation mechanism of retinal rod cyclic gmp phosphodiesterase probed by fluorescein-labeled inhibitory subunit. Biochemistry 29, 2155-2161.

Williams, K. R., & Carr, S. A. (1995). Protein analysis by integrated sample preparations, chemistry and mass spectrometry. In:Molecular Biology and Biotechnology: A Comprehensive Desk Reference (Meyers, R. A., eds.), pp. 731-737. VCH Publishers, New York.

Wiseman, T., Williston, S., Brandts, J. F., & Lin, L. N. (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. Anal. Biochem. 179, 131-137.

Wyman, J., & Gill, S. J. (1990). In: Binding and Linkage: Functional Chemistry of Biological Macromolecules. University Science Books, Mill Valley.

Yoo, S. H., & Lewis, M. S. (1995). Thermodynamic study of the pH-dependent interaction of chromogranin A with an intraluminal loop peptide of the inositol 1,4,5-trisphosphate receptor. Biochemistry 34, 632-638.

Advances in
Molecular and Cell Biology

Edited by **E. Edward Bittar**, *Department of
Physiology, University of Wisconsin, Madison*

Cell biology is a rapidly developing discipline which brings to-
gether many of the branches of the biological sciences that
were separate in the past. The interrelations between cell
structure and function at the molecular and subcellular level
are the central connecting theme of this series. "As the twen-
tieth century nears its close," writes Kenneth Miller, "the devel-
opment of an enormous range of tools and techniques, some
physical, some chemical, some biological, has changed the
situation forever. Cell biology today crosses the boundary,
links the molecule with the organelle, associates the cellular
response with the larger organism." Most of the contributors
are cell and molecular biologists who are currently engaged in
research and have made important contributions to our
present understanding of cell function.

Volume 7, Biology of the Cancer Cell
1993, 268 pp. $109.50
ISBN 1-55938-624-X

Edited by **Gloria H. Heppner,** *Director,
Breast Cancer Biology Program,
Michigan Cancer Center*

CONTENTS: Preface, *Gloria H. Heppner.* Cellular Genetic Al-
terations: Models of Breast and Colon Cancer, *Sandra R.
Wolman and D. Visscher.* Altered Expression of Transforming
Growth Factor α and Transforming Growth Factor-Autocrine
Loops in Cancer Cells, *M.G. Brattain, K.M. Mulder, S.P. Wu,
G. Howell, L. Sun, J.K.V. Wilson, and B.L. Ziober.* Altered
Signal Transduction in Carcinogenesis, *Catherine A. O'Brian,
Nancy E. Ward, and Constantin G. Ioannides.* The Signifi-
cance of the Extracellular Matrix in Mammary Epithelial Car-
cinogenesis, *Calvin D. Roskelley, Ole W. Petersen, and Mina
J. Bissell.* Epithelial-Stromal Cell Interactions and Breast
Cancer, *Sandra Z. Haslam, Laura J. Counterman, and
Katherine A. Nummy.* The Tissue Matrix and the Regulation
of Gene Expression in Cancer Cells, *Kenneth J. Pienta, Brian
C. Murphy, Robert H. Getzenberg, and Donald S. Coffey.*
Tumor Cell Interactions in Cancer Growth and Expression of
the Malignant Phenotype, *Fred R. Miller and Bonnie E. Miller.*
Effect of Class I MHC Gene Products on the Immunobiological
Properties of BL6 Melanoma, *Misoon Kim and Elieser Gorelik.*
The Role of Angiogenesis in Tumor Progression and Metasta-
sis, *Janusz W. Rak, Erick J. Hegmann, and Robert S. Kerbel.*
Subject Index.

J
A
I

P
R
E
S
S

J

A

I

P

R

E

S

S

J

A

I

P

R

E

S

S

Volume 10, Molecular Processes of Photosynthesis
1994, 437 pp. $109.50
ISBN 1-55938-710-6

Edited by **J. Barber,** *Imperial College of Science,*
Technology and Medicine, London

REVIEWS: "Guest editor Jim Barber has done first-rate work
in compiling this updated volume from first-rate review
papers, written by front-line researchers in the various spe-
cialisms. . . . this new book is a welcome and justified expan-
sion of photosynthesis literature."

— *Acta Botanica Neerlandica*

"The current book fills an important gap as it provides, in one
reasonably-sized volume, highlights of the cellular and
molecular process of photosynthesis. . . . I recommend it
whole-heartedly and urge the libraries of universities and
institutes to purchase it for the benefit of their advanced
undergraduate, graduate and postgraduate students in biol-
ogy."

— *Photochemistry and Photobiology*

CONTENTS: Preface, *J. Barber.* Organization and Dynamics
of Thylakoid Membranes, *B. Andersson and J. Barber.* Anten-
na Pigment-Protein Complexes of Higher Plants and Purple
Bacteria, *J.P. Thornber, R. Cogdell, P. Chitnis, D.T. Morish-
ige, G.F. Peter, S. Gomez, S. Anandan, S. Preiss, B. Welty
Dreyfuss, A. Lee, T. Takeuchi, and C. Kerfield.* Adaptive Vari-
ations in Phycobilisome Structure, *A.N. Glazer.* Photoprotec-
tion and Photoinhibitory Damage, *W.S. Chow.* Molecular
Genetic Manipulation and Characterization of Mutant Photo-
synthetic Reaction Centres from Purple Non-Sulfur Bacteria,
E. Takahashi and C.A. Wraight. Proton-Translocating
NAD(P)-H+-Transhydrogenase and NADH Dehydrogenase in
Photosynthetic Membranes, *J.B. Jackson and A. McEwan.*
Structural Elements Involved in the Assembly and Mechanims
of Action of Rubisco, *S. Gutteridge and T. Lundqvist.* The
Ferredoxin/Thioredoxin System: Update on its Role in the
Regulation of Oxygenic Photosynthesis, *R. Buchanan.* Iden-
tification, Cellular Localization and Participation of Chaper-
onins in Protein Folding, *A.A. Gatenby, P. Viitanen, V. Speth,
and R. Grimm.* Translocation of Proteins Across Chloroplast
Membranes, *B. Bruce and K. Keegstra.* Subject Index.

Volume 11, Enzymology *In Vivo*
1995, 271 pp. $109.50
ISBN 1-55938-844-7

Edited by **Kevin M. Brindle,** *Department of
Biochemistry, University of Cambridge*

REVIEW: "The contributions included in this volume provide
critical overviews and recent developments on the methods

and strategies which should be followed in order to study enzymes in their natural environment, the living cell."

<div align="right">— Anticancer Research</div>

CONTENTS: Preface, *Kevin M. Brindle.* Metabolic Channeling in Organized Enzyme Systems: Experiments and Models, *Pedro Mendes, Douglas B. Kell, and G. Rickey Welch.* Metabolic Control Analysis in Theory and Practice, *Athel Cornish-Bowden.* Experimental Approaches to Studying Enzymes *In Vivo:* The Application of Nuclear Magnetic Resonance Methods to Genetically Manipulated Organisms, *Simon-Peter Williams, Alexandra M. Fulton, and Kevin M. Brindle.* Glycolysis *In Vivo:* Fluorescence Microscopy as a Tool for Studying Enzyme Organization in Living Cells, *Len Pagliaro.* The Cooperative Behavior of Krebs Tricarboxylic Acid Cycle Enzymes, *Paul A. Srere, Craig R. Malloy, A. Dean Sherry, and Balazs Sumegi.* NMR Studies of Erythrocyte Metabolism, *Hilary A. Berthon and Philip W. Kuchel.* Studies of Physiological Control of ATP Synthesis, *K.F. LaNoue and C. Doumen.* Hepatic High Energy Phosphate Metabolism in Transgenic Livers Expressing Creatine Kinase as Revealed by 31 P NMR, *Alan P. Koretsky, Kenneth R. Miller, and Jessica M. Halow.* Subject Index.

Volume 12, Cytoskeleton
1995, 227 pp. $109.50
ISBN 1-55938-845-5

Edited by **John Hesketh,**
Rowett Research Institute, Aberdeen

CONTENTS: Preface, *John Hesketh.* Aspects of the Muscle Cytoskeleton, *John Trinick.* The Sequestration of mRNA in the Cytoskeleton and Subcellular Structures, *Ian F. Pryme, Arild Johannessen, and Anni Vdeler.* Roles of the Cytoskeleton and Membrane-Cytoskeleton in Generating and Maintaining the Structural and Functional Organization of Polarized Epithelial Cells, *Kenneth A. Beck and W. James Nelson.* Development of the Axonal and Dendritic Cytoskeleton, *Beat M. Riederer.* Regulation of Adherens Junction Protein Levels: Its Role in Cell Motility and Tumorigenicity, *Avri Ben-Ze'ev, Jose Luis Rodriguez Fernandez, Ursula Gluck, Daniela Salomon, Margot Zoller, and Benjamin Geiger.* The Kinesin Superfamily: Variations of a Theme, *Janet L. Cyr, Ann O. Sperry, and Scott T. Brady.* Genetic Dissection of Drosophila Cytoskeletal Functions, *Eric Fyrberg. Keratin Diseases, E. Birgitte Lane.* Subject Index.

Volume 13, Cell Cycle
1995, 211 pp. $109.50
ISBN 1-55938-949-4

Edited by **Michael Whitaker,** *University of Newcastle, Newcastle on Tyne*

J

A

I

P

R

E

S

S

J

A

I

P

R

E

S

S

CONTENTS: Preface, *Michael N.Whitaker.* Centrosomes and the Cell Cycle, *Greenfield Sluder.* Regulation of the Centrosome Function During Mitosis, *Brigitte Buendia and Eric Karsenti.* The Essential Role of Calcium During Mitosis, *Robert M. Tombes and Gary G. Borisy.* Calcium and Calmodulin Regulation of the Nuclear Division Cycle of Aspergillus Nidulans, *Kung Ping Lu and Anthony R. Means.* Cell Cycle Control by Protein Phosphatase Genes, *Mitsuhiro Yanagida.* The cdc25 Phosphatase: Biochemistry and Regulation in the Eukaryotic Cell Cycle, *Ingrid Hoffmann, Paul R. Clarke, and Giulio Draetta.* Control of Nuclear Lamina Assembly/Disassembly by Phosphorylation, *Matthias Peter and Erich A. Nigg.* Dissection of the Cell Cycle Using Cell-Free Extracts from *Xenopus laevis, C.C. Ford and H. Lindsay.* Subject Index.

Edited by **Colin R. Jefcoate,**
University of Wisconsin, Madison

CONTENTS: Preface, *Colin R. Jefcoate.* P450 Cytochromes and Hormonal Regulation: An Overview, *Colin R. Jefcoate.* Electron Transfer Proteins of Cytochrome P450 Systems, *Israel Hanukoglu.* The Molecular Structure of P450S: The Conserved and the Variable Elements, *Sandra E. Graham-Lorence and Julian A. Peterson.* Diverse Molecular Mechanisms Regulate the Expression of Steroid Hydroxylase Genes Required for Production of Ligands for Nuclear Receptors, *Michael R. Waterman and Diane S. Keeney.* Cytochrome P450scc and Regulation of Cholesterol Conversion to Steroid Hormones, *Colin R. Jefcoate.* The Regulation of the Formation of Glucocorticoids and Mineralocorticoids *in Vivo, Jean-Guy LeHoux, Hugues Bernard, Lyne Ducharme, Andrée Lefebvre, Dennis Shapcott, André Tremblay, and Steve Véronneau.* Physiology and Molecular Biology of P450c21 and P450c17, *Bon-chu Chung.* The Aromatase Reaction, *Evan R. Simpson, Mala S. Mahendroo, Michael W. Kilgore, Gary D. Means, Sedar E. Bulun, Margaret M. Hinshelwood, and Carole R. Mendelson.* Regulation of Calcium Metabolism by the Vitamin D Hydroxylases, *H. James Armbrecht, Rama K. Nemani, and N. Wongsurawat.* The Regulation of Cholesterol Conversion to Bile Acids, *John Y.L. Chiang and Z. Reno Vlahcevic.* Eicosanoid Metabolism and Bioactivation by Microsomal Cytochrome P450, *Jorge H. Capdevila, Darryl Zeldin, Armando Karara, and John R. Falck.* Steroid Hormones and Other Physiologic Regulators of Liver Cytochromes P450: Metabolic Reactions and Regulatory Pathways, *David J. Waxman.* Index.