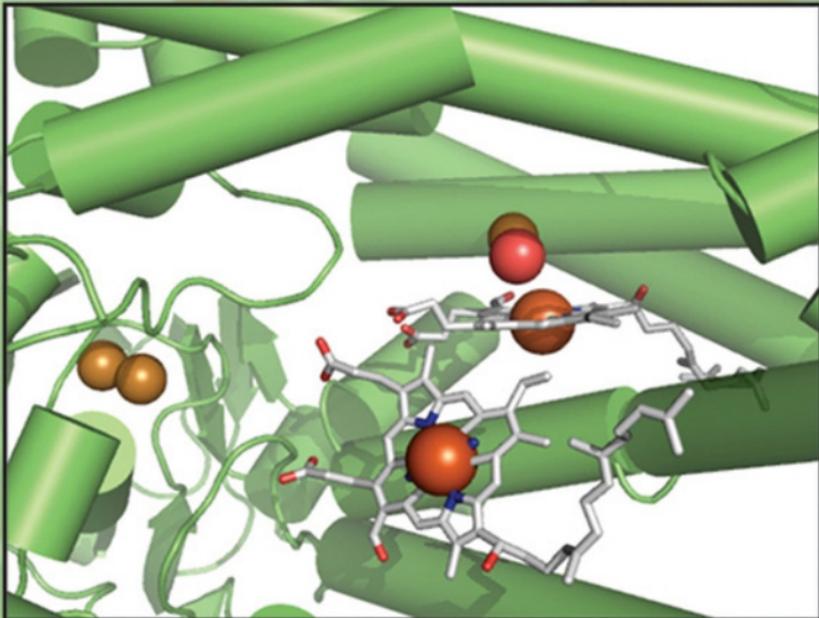


 WILEY

# BIOINORGANIC CHEMISTRY

A SHORT COURSE

**2**  
ND  
ITION



ROSETTE M. ROAT-MALONE

# **BIOINORGANIC CHEMISTRY**



---

THE WILEY BICENTENNIAL—KNOWLEDGE FOR GENERATIONS

---

Each generation has its unique needs and aspirations. When Charles Wiley first opened his small printing shop in lower Manhattan in 1807, it was a generation of boundless potential searching for an identity. And we were there, helping to define a new American literary tradition. Over half a century later, in the midst of the Second Industrial Revolution, it was a generation focused on building the future. Once again, we were there, supplying the critical scientific, technical, and engineering knowledge that helped frame the world. Throughout the 20th Century, and into the new millennium, nations began to reach out beyond their own borders and a new international community was born. Wiley was there, expanding its operations around the world to enable a global exchange of ideas, opinions, and know-how.

For 200 years, Wiley has been an integral part of each generation's journey, enabling the flow of information and understanding necessary to meet their needs and fulfill their aspirations. Today, bold new technologies are changing the way we live and learn. Wiley will be there, providing you the must-have knowledge you need to imagine new worlds, new possibilities, and new opportunities.

Generations come and go, but you can always count on Wiley to provide you the knowledge you need, when and where you need it!

Two handwritten signatures are shown side-by-side. The signature on the left is 'William J. Pesce' and the signature on the right is 'Peter Booth Wiley'.

WILLIAM J. PESCE  
PRESIDENT AND CHIEF EXECUTIVE OFFICER

PETER BOOTH WILEY  
CHAIRMAN OF THE BOARD

---

# **BIOINORGANIC CHEMISTRY**

---

## **A Short Course Second Edition**

**ROSETTE M. ROAT-MALONE**

Chemistry Department  
Washington College  
Chestertown, MD



**WILEY-INTERSCIENCE**

**A John Wiley & Sons, Inc., Publication**

Copyright © 2007 by John Wiley & Sons, Inc. All rights reserved

Published by John Wiley & Sons, Inc., Hoboken, New Jersey

Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at [www.copyright.com](http://www.copyright.com). Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

**Limit of Liability/Disclaimer of Warranty:** While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at [www.wiley.com](http://www.wiley.com).

Wiley Bicentennial logo: Richard J. Pacifico

***Library of Congress Cataloging-in-Publication Data:***

Bioinorganic chemistry : a short course / edited by Rosette M. Roat-Malone.—2nd ed.

p. ; cm.

Includes bibliographical references and index.

ISBN 978-0-471-76113-6 (pbk.)

1. Bioinorganic chemistry. I. Roat-Malone, Rosette M.

[DNLM: 1. Biochemistry. 2. Chemistry, Bioinorganic. QU 130 B614407 2008]

QP531.R63 2008

572'.51—dc22

2007019892

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

*To my young friends  
Allie, Andy, Anna, and Sebastian*



# CONTENTS

<b>Preface</b>	<b>xiii</b>
<b>Acknowledgments</b>	<b>xix</b>
<b>1 Inorganic Chemistry Essentials</b>	<b>1</b>
1.1 Introduction, 1	
1.2 Essential Chemical Elements, 1	
1.3 Metals in Biological Systems: A Survey, 3	
1.4 Inorganic Chemistry Basics, 6	
1.5 Biological Metal Ion Complexation, 8	
1.5.1 Thermodynamics, 8	
1.5.2 Kinetics, 9	
1.6 Electronic and Geometric Structures of Metals in Biological Systems, 13	
1.7 Bioorganometallic Chemistry, 19	
1.8 Electron Transfer, 22	
1.9 Conclusions, 26	
References, 27	
<b>2 Biochemistry Fundamentals</b>	<b>29</b>
2.1 Introduction, 29	
2.2 Proteins, 30	
2.2.1 Amino Acid Building Blocks, 30	
2.2.2 Protein Structure, 33	

2.2.3	Protein Sequencing and Proteomics, 39
2.2.4	Protein Function, Enzymes, and Enzyme Kinetics, 43
2.3	Nucleic Acids, 47
2.3.1	DNA and RNA Building Blocks, 47
2.3.2	DNA and RNA Molecular Structures, 47
2.3.3	Transmission of Genetic Information, 53
2.3.4	Genetic Mutations and Site-Directed Mutagenesis, 56
2.3.5	Genes and Cloning, 58
2.3.6	Genomics and the Human Genome, 61
2.4	Zinc-Finger Proteins, 63
2.4.1	Descriptive Examples, 67
2.5	Summary and Conclusions, 73
	References, 74

### **3 Instrumental Methods 76**

3.1	Introduction, 76
3.1.1	Analytical Instrument-Based Methods, 76
3.1.2	Spectroscopy, 77
3.2	X-Ray Absorption Spectroscopy (XAS) and Extended X-Ray Absorption Fine Structure (EXAFS), 78
3.2.1	Theoretical Aspects and Hardware, 78
3.2.2	Descriptive Examples, 81
3.3	X-Ray Crystallography, 83
3.3.1	Introduction, 83
3.3.2	Crystallization and Crystal Habits, 84
3.3.3	Theory and Hardware, 88
3.3.4	Descriptive Examples, 95
3.4	Nuclear Magnetic Resonance, 98
3.4.1	Theoretical Aspects, 98
3.4.2	Nuclear Screening and the Chemical Shift, 101
3.4.3	Spin-Spin Coupling, 104
3.4.4	Techniques of Spectral Integration and Spin-Spin Decoupling, 106
3.4.5	Nuclear Magnetic Relaxation, 107
3.4.6	The Nuclear Overhauser Effect (NOE), 108
3.4.7	Obtaining the NMR Spectrum, 110
3.4.8	Two-Dimensional (2D) NMR Spectroscopy, 111
3.4.9	Two-Dimensional Correlation Spectroscopy (COSY) and Total Correlation Spectroscopy (TOCSY), 112
3.4.10	Nuclear Overhauser Effect Spectroscopy (NOESY), 115
3.4.11	Multidimensional NMR, 116
3.4.12	Descriptive Examples, 117
3.5	Electron Paramagnetic Resonance, 122
3.5.1	Theory and Determination of g-Values, 122

3.5.2	Hyperfine and Superhyperfine Interactions,	127
3.5.3	Electron Nuclear Double Resonance (ENDOR) and Electron Spin-Echo Envelope Modulation (ESEEM),	129
3.5.4	Descriptive Examples,	129
3.6	Mössbauer Spectroscopy,	132
3.6.1	Theoretical Aspects,	132
3.6.2	Quadrupole Splitting and the Isomer Shift,	134
3.6.3	Magnetic Hyperfine Interactions,	136
3.6.4	Descriptive Examples,	137
3.7	Other Instrumental Methods,	139
3.7.1	Atomic Force Microscopy,	139
3.7.2	Fast and Time-Resolved Methods,	143
3.7.2.1	Stopped-Flow Kinetic Methods,	143
3.7.2.2	Flash Photolysis,	144
3.7.2.3	Time-Resolved Crystallography,	146
3.7.3	Mass Spectrometry,	148
3.8	Summary and Conclusions,	153
	References,	154

## 4 Computer Hardware, Software, and Computational Chemistry Methods

157

4.1	Introduction to Computer-Based Methods,	157
4.2	Computer Hardware,	157
4.3	Molecular Modeling and Molecular Mechanics,	160
4.3.1	Introduction to MM,	160
4.3.2	Molecular Modeling, Molecular Mechanics, and Molecular Dynamics,	161
4.3.3	Biomolecule Modeling,	166
4.3.4	A Molecular Modeling Descriptive Example,	167
4.4	Quantum Mechanics-Based Computational Methods,	170
4.4.1	Introduction,	170
4.4.2	Ab Initio Methods,	170
4.4.3	Density Function Theory,	171
4.4.4	Semiempirical Methods,	173
4.5	Computer Software for Chemistry,	174
4.5.1	Mathematical Software,	180
4.6	World Wide Web Online Resources,	181
4.6.1	Nomenclature and Visualization Resources,	181
4.6.2	Online Societies, Online Literature Searching, and Materials and Equipment Websites,	183
4.7	Summary and Conclusions,	185
	References,	185

<b>5 Group I and II Metals in Biological Systems: Homeostasis and Group I Biomolecules</b>	<b>189</b>
5.1 Introduction, 189	
5.2 Homeostasis of Metals (and Some Nonmetals), 192	
5.2.1 Phosphorus as Phosphate, 192	
5.2.2 Potassium, Sodium, and Chloride Ions, 193	
5.2.3 Calcium Homeostasis, 194	
5.3 Movement of Molecules and Ions Across Membranes, 195	
5.3.1 Passive Diffusion, 195	
5.3.2 Facilitated Diffusion, 197	
5.3.2.1 Gated Channels, 197	
5.3.3 Active Transport—Ion Pumps, 197	
5.4 Potassium-Dependent Molecules, 199	
5.4.1 Na <sup>+</sup> /K <sup>+</sup> ATPase: The Sodium Pump, 199	
5.4.2 Potassium (K <sup>+</sup> ) Ion Channels, 203	
5.4.2.1 Introduction, 203	
5.4.2.2 X-Ray Crystallographic Studies, 205	
5.5 Conclusions, 235	
References, 235	
<b>6 Group I and II Metals in Biological Systems: Group II</b>	<b>238</b>
6.1 Introduction, 238	
6.2 Magnesium and Catalytic RNA, 238	
6.2.1 Introduction, 238	
6.2.2 Analyzing the Role of the Metal Ion, 241	
6.2.3 The Group I Intron Ribozyme, 244	
6.2.4 The Hammerhead Ribozyme 261	
6.3 Calcium-Dependent Molecules, 301	
6.3.1 Introduction, 301	
6.3.2 Calmodulin, 302	
6.3.2.1 Introduction, 302	
6.3.2.2 Calmodulin Structure by X-Ray and NMR, 303	
6.3.2.3 Calmodulin Interactions with Drug Molecules, 308	
6.3.2.4 Calmodulin–Peptide Binding, 313	
6.3.2.5 Conclusions, 326	
6.4 Phosphoryl Transfer: P-Type ATPases, 327	
6.4.1 Introduction, 327	
6.4.2 Calcium P-Type ATPases, 327	
6.4.2.1 Ca <sup>2+</sup> -ATPase Protein SERCA1a and the Ca <sup>2+</sup> -ATPase Cycle, 329	
6.5 Conclusions, 337	
References, 338	

<b>7 Iron-Containing Proteins and Enzymes</b>	<b>343</b>
7.1 Introduction: Iron-Containing Proteins with Porphyrin Ligand Systems, 343	
7.2 Myoglobin and Hemoglobin, 343	
7.2.1 Myoglobin and Hemoglobin Basics, 345	
7.2.2 Structure of the Heme Prosthetic Group, 347	
7.2.3 Behavior of Dioxygen Bound to Metals, 348	
7.2.4 Structure of the Active Site in Myoglobin and Hemoglobin: Comparison to Model Compounds, 349	
7.2.5 Some Notes on Model Compounds, 352	
7.2.6 Iron-Containing Model Compounds, 353	
7.2.7 Binding of CO to Myoglobin, Hemoglobin, and Model Compounds, 356	
7.2.8 Conclusions, 359	
7.3 Introduction to Cytochromes, 359	
7.4 Cytochrome P450: A Monooxygenase, 361	
7.4.1 Introduction, 361	
7.4.2 Cytochrome P450: Structure and Function, 363	
7.4.3 Cytochrome P450: Mechanism of Activity, 365	
7.4.4 Analytical Methods: X-Ray Crystallography, 369	
7.4.5 Cytochrome P450 Model Compounds, 372	
7.4.5.1 Introduction, 372	
7.4.5.2 A Cytochrome P450 Model Compound: Structural, 372	
7.4.5.3 Cytochrome P450 Model Compounds: Functional, 374	
7.4.6 Cytochrome P450 Conclusions, 382	
7.5 Cytochrome b(6)f: A Green Plant Cytochrome, 382	
7.5.1 Introduction, 382	
7.5.2 Cytochrome b(6)f Metal Cofactor Specifics, 386	
7.6 Cytochrome bc <sub>1</sub> : A Bacterial Cytochrome, 388	
7.6.1 Introduction, 388	
7.6.2 Cytochrome bc <sub>1</sub> Structure, 389	
7.6.3 Cytochrome bc <sub>1</sub> Metal Cofactor Specifics, 391	
7.6.4 The Cytochrome bc <sub>1</sub> Q Cycle, 395	
7.6.5 Cytochrome bc <sub>1</sub> Inhibitors, 397	
7.6.6 Cytochrome bc <sub>1</sub> Conclusions, 408	
7.7 Cytochromes c, 408	
7.7.1 Introduction, 408	
7.7.2 Mitochondrial Cytochrome c (Yeast), 411	
7.7.3 Mitochondrial Cytochrome c (Horse), 416	
7.7.4 Cytochrome c Folding, Electron Transfer, and Cell Apoptosis, 422	
7.7.4.1 Cytochrome c Folding, 422	

7.7.4.2	Electron Transfer in Cytochrome c and Its Redox Partners,	424
7.7.4.3	Apoptosis,	427
7.7.5	Cytochrome c Conclusions,	429
7.8	Cytochrome c Oxidase,	429
7.8.1	Introduction,	429
7.8.2	Metal-Binding Sites in Cytochrome c Oxidase,	432
7.8.3	Dioxygen Binding, Proton Translocation, and Electron Transport,	434
7.8.4	Cytochrome c Oxidase Model Compounds and Associated Analytical Techniques,	440
7.8.5	Cytochrome c Oxidase Conclusions,	453
7.9	Non-Heme Iron-Containing Proteins,	454
7.9.1	Introduction,	454
7.9.2	Proteins with Iron–Sulfur Clusters,	454
7.9.2.1	The Enzyme Aconitase	455
7.9.3	Iron–Oxo Proteins,	458
7.9.3.1	Methane Monooxygenases	459
7.10	Conclusions,	465
	References,	466

# PREFACE

This second edition of *Bioinorganic Chemistry: A Short Course* adopts the same philosophy as the first—that is, chapters of introductory material followed by chapters featuring detailed discussions of specific bioinorganic chemistry topics. This approach foregoes any attempt to exhaustively survey the enormous range of bioinorganic topics that occupy the attention and research of theoreticians and experimentalists currently engaged in the field. In this second edition, introductory Chapters 1 and 2 cover inorganic chemistry essentials and biochemistry fundamentals for bioinorganic chemistry students whose background in these topics may be less than ideal. Chapter 3 (Instrumental Methods) concentrates on the physical and analytical methods used to describe the bioinorganic systems discussed in Chapters 5 through 7. Chapter 4 (Computer Hardware, Software, and Computational Chemistry Methods) describes some of the vast array of computer hardware, software, and drawing, visualization, computational, and modeling programs used by every researcher studying bioinorganic systems. Computational chemistry, for instance, allows researchers to predict molecular structures of known and theoretical compounds and to design and test new compounds on computers rather than at the laboratory bench. Chapter 5 (Group I and II Metals in Biological Systems: Homeostasis and Group I Biomolecules) discusses the vital roles of sodium and potassium ions in maintaining cellular integrity, and features the Nobel Prize-winning work of Roderick MacKinnon's research group on potassium ion channels. More structural work by the MacKinnon group confirming the selectivity of potassium ion channels for  $K^+$  over  $Na^+$  can be found in a recent *Science* magazine article (*Science* 2006, **314**, 1004–1007). Chapter 6 (Group I and II Metals in Biological Systems: Group II) describes the importance of

magnesium ions in catalytic RNA (ribozymes). Readers interested in the “RNA World hypothesis”, a theory connecting the origin of life with self-replicating ribozymes, will want to read the recent article by Michael Robertson and William Scott (*Science* 2007, **315**, 1549–1553). A background perspective on this article has been written by Gerald Joyce (*Science* 2007, **315**, 1507–1508). In addition, Chapter 6 discusses two calcium-containing biomolecules—calmodulin, a primary receptor for intracellular calcium ions and a switch in  $\text{Ca}^{2+}$ -dependent signaling pathways, and  $\text{Ca}^{2+}$ -ATPase, a major player in muscle contraction-relaxation cycles. Chapter 7 (Iron Containing Proteins and Enzymes) devotes much of its descriptive material to proteins and enzymes that contain their iron ions within a heme ligand system. This chapter extends the first edition’s discussion of myoglobin and hemoglobin, then reports on some members of the ubiquitous cytochrome family—cytochrome P450, a monooxygenase, cytochrome b(6)f, a green plant constituent, bacteria-based cytochrome bc<sub>1</sub>, members of the cytochrome c superfamily, and cytochrome c oxidase (CcO), the terminal electron transferring enzyme in the mitochondrial respiratory chain. An update reported recently by the Collman group (*Science* 2007, **315**, 1565–1568) connects the redox-active centers of cytochrome c oxidase—Fe<sub>a3</sub>, Cu<sub>B</sub>, and tyr244—to the rapid accumulation of four electrons. The four accumulated electrons are needed to reduce dioxygen, O<sub>2</sub>, to two oxide, O<sup>2-</sup>, ions while avoiding the production of partially reduced, tissue-damaging superoxide, O<sub>2</sub><sup>-•</sup>, or peroxide, O<sub>2</sub><sup>2-</sup>, ions. A shorter section in Chapter 7 discusses non-heme iron-containing proteins and enzymes, many of which, like aconitase, feature iron-sulfur clusters. Lastly, Chapter 7 reports on the enzyme methane monooxygenase (MMO), utilized by methanotrophic bacteria to oxidize methane to methanol with incorporation of one O<sub>2</sub> oxygen atom.

Many exciting bioinorganic topics are not covered in either the first or the present editions of *Bioinorganic Chemistry: A Short Course*. The new field of nanobioinorganic chemistry has become a prominent research area, especially in the medical field. Readers who wish to research this area might start with the review article: “Metal Nanoshells” in the *Annals of Biomedical Engineering* 2006, **34**(1), 15–22. In this article, Jennifer L. West and coworkers describe a new class of nanoparticles that have tunable optical properties. Chad Mirkin and coworkers describe oligonucleotide-modified gold nanoparticles that are being developed as intracellular gene regulation agents (*Science*, 2006, **312**, 1027–1030; *J. Am. Chem. Soc.* 2006, **128**(29), 9286–9287; *J. Am. Chem. Soc.* 2006, **128**(27), 8899–8903). These agents may eventually find applications in controlling the expression of specific proteins in cells for medical diagnostic and therapeutic purposes. The International Council on Nanotechnology (ICON) maintains a website at <http://icon.rice.edu/research.cfm> that includes links to other databases of interest, such as NIOSH (National Institute for Occupational Safety and Health) and the nanomedicine portal. ICON is particularly interested in informing researchers and nanotechnology users on environmental and safety issues related to this new, rapidly expanding field.

Readers interested in the connection between bioinorganic chemistry and catalysis might begin by reading an article entitled: “Better than Platinum? Fuels Cells energized by enzymes.” written by Marcetta Dahrensbourg, Michael Hall, and Jesse Tye (*Proc. Natl. Acad. Sci. U.S.A.* 2005, **102**(47), 16911–16912.) This article briefly describes the interest of bioinorganic chemists in the hydrogenase enzymes that biologically and reversibly accomplish proton reduction and dihydrogen oxidation. Since their discovery, hydrogenase enzymes, containing sulfur-bridged di-iron or nickel-iron active sites, have been presented as possible substitutes for expensive noble-metal based catalysts in the  $2\text{H}^+ + 2\text{e}^- \leftrightarrow \text{H}_2$  reaction. More recently, these researchers have published studies of synthetic di-iron(I) complexes as structural models of reduced Fe-Fe hydrogenase (*Inorg. Chem.* 2006, **45**(4), 1552–1559) and computational studies comparing computed gas-phase and experimental solution phase infrared spectra of Fe-Fe hydrogenase active site models (*J. Comput. Chem.* 2006, **27**(12), 1454–1462).

Readers with a more structural biology bent might be interested in the 2006 achievement of Jennifer A. Doudna’s group at the University of California, Berkeley in obtaining the first crystal structure of Dicer, an enzyme that initiates RNA interference (RNAi). This work, published in *Science* (2006, **311**, 195–198), helps confirm that two metal ions—in the X-ray crystallographic structure,  $\text{Er}^{3+}$  substitutes for the naturally occurring  $\text{Mn}^{2+}$  ions—participate in Dicer’s catalytic mechanism.

Intense research continues on the complex enzyme nitrogenase, described in the first edition’s Chapter 6. New X-ray crystallographic results for nitrogenase have led to the probable positioning of an atom, most plausibly nitrogen, as a central ligand in nitrogenase’s FeMo-cofactor (Rees, D. C., et al. *Science* 2002, **297**, 1696–1700). X-ray crystallographic data are deposited in the Protein Data Bank (PDB) at <http://www.rcsb.org/pdb> with the accession number 1M1N. (Note that the third character is the numeral one and not the letter “I”.) More recently, the Rees research group has structurally identified conformational changes in the nitrogenase complex during adenosine triphosphate (ATP) turnover (*Science* 2005, **309**, 1377–1380, PDB: 2AFH, 2AFI, 2AFK). Concurrent with structural studies, the Brian M. Hoffman group at Northwestern University has trapped  $\text{N}_2$ -derived intermediates bound to nitrogenase and synchronized the number of electrons arriving at the active site with possible nitrogenase  $\text{H}^{+}$ -,  $\text{H}^{\cdot}$ -, or  $\text{H}_2$ -containing intermediates. (Lukyanov, D., Barney, B. M., Dean, D. R., Seefeldt, L. C., Hoffman, B. M. *Proc. Natl. Acad. Sci. U.S.A.* 2007, **104**(5), 1451–1455; Barney, B. M., Lukyanov, D., Yang, T. C., Dean, D. R., Hoffman, B. M., Seefeldt, L. C. *Proc. Natl. Acad. Sci. U.S.A.* 2006, **103**(46), 17113–17118.) An excellent article with many references, available from the Royal Society at <http://www.journals.royalsoc.ac.uk> as a free download, reviews the structural basis of nitrogen fixation. (Rees, D. C., Tezcan, F. A., Haynes, C. A., Walton, M.Y., Andrade, S., Einsle, O., Howard, J.B. *Phil. Trans. R. Soc. A* 2005, **363**, 971–984.) In April 2007, a search of PubMed, [www.ncbi.nlm.nih.gov/pubmed](http://www.ncbi.nlm.nih.gov/pubmed), using the keyword nitrogenase and limiting the

search to the journal *Proceedings of the National Academy of Sciences U.S.A.*, and to the years 2005–2007, yielded thirteen pertinent articles, of which those published online more than one year ago are available as free downloads.

Researchers continue to extend their ability to study and analyze complex bioinorganic systems as new experimental and instrumental methods are developed and current ones are improved. For instance, protein structure determination in solution by nuclear magnetic resonance, NMR, received a boost in 2006 through a technique developed at Tokyo Metropolitan University. This technique, stereo-array isotope labeling, SAIL, will make it possible to routinely determine protein structures at least twice as large as those being determined using current NMR methods (Kainosho, M., Torizawa, T., Iwashita, Y., Terauchi, T., Ono, A. M., Guntert, P. *Nature* 2006, **440**, 52–57, PDB: 1X02). The solution structure of the  $\text{Ca}^{2+}$ -containing protein calmodulin described in the *Nature* article, as determined by the SAIL method, is compared to X-ray crystallographic structures in Section 6.3.2.2—see especially Figure 6.23.

In some cases, the increasing complexity of bioinorganic systems studied, and the increasing sophistication of the analytical methods used, has led to controversy over the interpretation of biomolecular structures and behaviors. In this text, variations in experimental results and their interpretations among different research groups are found in the discussions of potassium ion channels (Section 5.4.2), group I intron ribozymes (Section 6.2.3), and the hammerhead ribozyme (Section 6.2.4). This author has attempted to present material on all existent interpretations by different research groups working in good faith to solve thorny experimental problems. All researchers, including newcomers to these complicated subjects, should maintain an open mind, a continuing interest in and exploration of the problems, and a civil manner of discourse within the scientific literature.

Admission of errors can be part of this discourse, although, to my knowledge, these have not been called for in the research areas mentioned in the previous paragraph. Recently, however, retractions appeared in *Science* magazine concerning incorrect interpretations of X-ray crystallographic data gathered on the MsbA protein, an important member of a class of molecules that use energy from adenosine triphosphate, ATP, to transport molecules across cell membranes—the so-called ABC transporters. The erroneous structures arose not because of any fault in the data collection scheme or the protein crystals themselves, but because of a faulty data-analysis program used to massage the data into visualized molecular structures. The incorrectly visualized MsbA protein structures were featured in at least five journal articles now being retracted (Miller, G., News of the Week, *Science* 2006, **314**, 1856–1857; Chang, G., Roth, C. B., Reyes, C. L., Pornillos, O., Chen, Y-J., Chen, A. P. Letters, *Science* 2006, **314**, 1875; Miller C. Letters *Science* 2007, **315**, 459. No MsbA protein structures, faulty or otherwise, are discussed in this text. However, as will be said numerous times herein, the techniques of X-ray crystallography provide snapshots of biomolecules frozen into a solid crystalline lattice, not a

normal biomolecular physical state of being, and certainly not representative of every possible molecular conformation in the biological milieu. If errors in data interpretation are also introduced, one sees how incorrect biomolecule structure visualizations find their way into the literature. Confirmation of X-ray crystallographic structural results through experimental biochemistry and by the use of multiple analytical techniques—nuclear magnetic resonance (NMR), electron paramagnetic resonance (EPR), and Mössbauer spectroscopies to name a few—should always be sought by bioinorganic researchers.

Lastly, and importantly, researchers, academicians, and their students want to maintain ethical behaviors in their scientific endeavors. Although science practitioners have historically been self-policing in this regard, and continue to be so, science writers and thinkers now call for more consideration of ethical topics, especially for students in graduate and post-graduate years as well as for early-career scientists. Readers who wish more information on ethical issues may consult a recent article entitled: “A Code of Ethics for the Life Sciences” by Nancy Jones, an American Association for the Advancement of Science/National Institutes of Health (NIH) Science Policy Fellow and a faculty member at Wake Forest University School of Medicine. The article has been published in *Science and Engineering Ethics*, by Springer Netherlands, January 30, 2007, online at <http://www.springerlink.com>.

This text is appropriate for use in one-semester bioinorganic chemistry courses offered to fourth year undergraduate chemistry, biochemistry and biology majors or first year graduate students concentrating in inorganic and biochemical subject areas. After presentation of some introductory material in inorganic, biochemistry, and a review of selected instrumental and computer-based topics, I suggest choosing one to three bioinorganic chemistry topics from Chapters 5 through 7 for thorough discussion. Following that, students should be encouraged to choose their own bioinorganic topics for research and study. Their endeavors could lead to classroom presentations, laboratory experimentation, and submission of written term papers. Certainly, the subject area provides great opportunities for introducing the use of primary literature sources and the application of computer- and internet-based searching, visualization, and modeling techniques.

A website to accompany the second edition of *Bioinorganic Chemistry: A Short Course* can be found at <http://chemistry.washcoll.edu/roat/>. The website contains the book’s table of contents, a listing of online resources organized by chapter and subject area, additional figures organized by chapter section (best viewed while studying the section’s material), updated bibliographic references, study questions for each chapter, and communication links for questions, comments, and corrections submitted by instructors and students.

ROSETTE M. ROAT-MALONE  
*Washington College*



## **ACKNOWLEDGMENTS**

Many groups of people contributed to the creation and realization of this book. Thanks to former bioinorganic chemistry students, whose enthusiasm for the subject material inspired the book's manner of presentation. Professional colleagues at Washington College and other universities worldwide helped in many ways—as critical readers, and as advisers on important subject areas to be included. The book would not exist without the expert assistance of Wiley editors—Anita Lekhwani, Rebekah Amos, Danielle Lacourciere, Nancy Heimbaugh, and Dean Gonzalez. Lastly, I express heartfelt gratitude to family and friends for their patience during the many months of gestation.



---

# 1

---

## INORGANIC CHEMISTRY ESSENTIALS

### 1.1 INTRODUCTION

Bioinorganic chemistry involves the study of metal species in biological systems. As an introduction to the basic inorganic chemistry needed for understanding bioinorganic topics, this chapter will discuss the essential chemical elements, the occurrences and purposes of metal centers in biological species, the geometries of ligand fields surrounding these metal centers, and ionic states preferred by the metals. Important considerations include equilibria between metal centers and their ligands and a basic understanding of the kinetics of biological metal-ligand systems. The occurrence of organometallic complexes and clusters in metalloproteins will be discussed briefly, and an introduction to electron transfer in coordination complexes will be presented. Since the metal centers under consideration are found in a biochemical milieu, basic biochemical concepts, including a discussion of proteins and nucleic acids, are presented in Chapter 2.

### 1.2 ESSENTIAL CHEMICAL ELEMENTS

Chemical elements essential to life forms can be broken down into four major categories: (1) bulk elements ( $\text{H}/\text{H}^+$ , C, N,  $\text{O}^{2-}/\text{O}_2^\cdot/\text{O}_2^{2-}$ , P, S/ $\text{S}^{2-}$ ); (2) macro-minerals and ions ( $\text{Na}/\text{Na}^+$ ,  $\text{K}/\text{K}^+$ ,  $\text{Mg}/\text{Mg}^{2+}$ ,  $\text{Ca}/\text{Ca}^{2+}$ ,  $\text{Cl}^-$ ,  $\text{PO}_4^{3-}$ ,  $\text{SO}_4^{2-}$ ); (3) trace

elements ( $\text{Fe}/\text{Fe}^{\text{II}}/\text{Fe}^{\text{III}}/\text{Fe}^{\text{IV}}$ ,  $\text{Zn}/\text{Zn}^{\text{II}}$ ,  $\text{Cu}/\text{Cu}^{\text{I}}/\text{Cu}^{\text{II}}/\text{Cu}^{\text{III}}$ ); and (4) ultratrace elements, comprised of nonmetals ( $\text{F}/\text{F}^-$ ,  $\text{I}/\text{I}^-$ ,  $\text{Se}/\text{Se}^{2-}$ ,  $\text{Si}/\text{Si}^{\text{IV}}$ , As, B) and metals ( $\text{Mn}/\text{Mn}^{\text{II}}/\text{Mn}^{\text{III}}/\text{Mn}^{\text{IV}}$ ,  $\text{Mo}/\text{Mo}^{\text{IV}}/\text{Mo}^{\text{V}}/\text{Mo}^{\text{VI}}$ ,  $\text{Co}/\text{Co}^{\text{II}}/\text{Co}^{\text{III}}$ ,  $\text{Cr}/\text{Cr}^{\text{III}}/\text{Cr}^{\text{VI}}$ ,  $\text{V}/\text{V}^{\text{III}}/\text{V}^{\text{IV}}/\text{V}^{\text{V}}$ ,  $\text{Ni}^{\text{i}}/\text{Ni}^{\text{II}}/\text{Ni}^{\text{III}}$ ,  $\text{Cd}/\text{Cd}^{2+}$ ,  $\text{Sn}/\text{Sn}^{\text{II}}/\text{Sn}^{\text{IV}}$ ,  $\text{Pb}/\text{Pb}^{2+}$ ,  $\text{Li}/\text{Li}^+$ ). In the preceding classification, only the common biologically active ion oxidation states are indicated. (See references 3 and 21d for more information.) If no charge is shown, the element predominately bonds covalently with its partners in biological compounds, although elements such as carbon (C), sulfur (S), phosphorus (P), arsenic (As), boron (B), selenium (Se) have positive formal oxidation states in ions containing oxygen atoms; that is, S = +6 in the  $\text{SO}_4^{2-}$  ion or P = +5 in the  $\text{PO}_4^{3-}$  ion. The identities of essential elements are based on historical work and that done by Klaus Schwarz in the 1970s.<sup>1</sup> Other essential elements may be present in various biological species. Essentiality has been defined by certain criteria: (1) A physiological deficiency appears when the element is removed from the diet; (2) the deficiency is relieved by the addition of that element to the diet; and (3) a specific biological function is associated with the element.<sup>2</sup> Table 1.1 indicates the approximate percentages by weight of selected essential elements for an adult human.

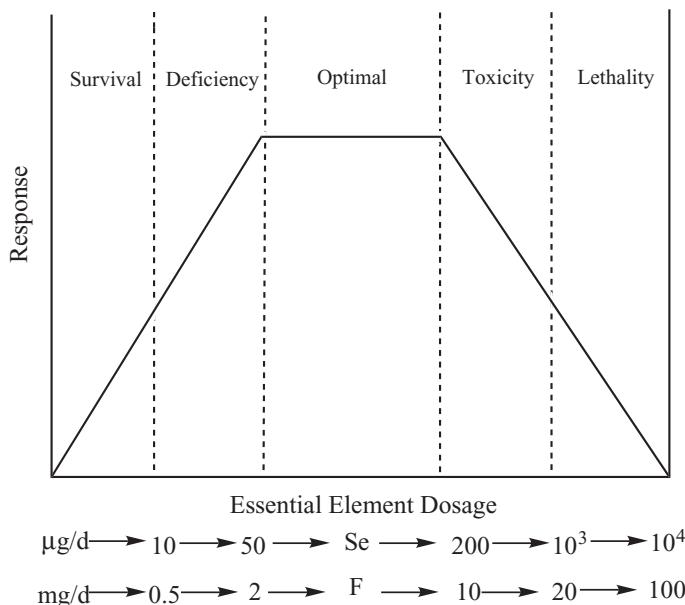
Every essential element follows a dose–response curve, shown in Figure 1.1, as adapted from reference 2. At lowest dosages the organism does not survive, whereas in deficiency regions the organism exists with less than optimal function. After the concentration plateau of the optimal dosage region, higher dosages cause toxic effects in the organism, eventually leading to lethality. Specific daily requirements of essential elements may range from microgram to gram quantities as shown for two representative elements in Figure 1.1.<sup>2</sup>

Considering the content of earth's contemporary waters and atmospheres, many questions arise as to the choice of essential elements at the time of life's origins 3.5 billion or more years ago. Certainly, sufficient quantities of the bulk elements were available in primordial oceans and at shorelines. However, the concentrations of essential trace metals in modern oceans may differ considerably from those found in prebiotic times. Iron's current approximate  $10^{-4}\text{ mM}$

**TABLE 1.1 Percentage Composition of Selected Elements in the Human Body**

Element	Percentage (by weight)	Element	Percentage (by weight)
Oxygen	53.6	Silicon, magnesium	0.04
Carbon	16.0	Iron, fluorine	0.005
Hydrogen	13.4	Zinc	0.003
Nitrogen	2.4	Copper, bromine	$2 \times 10^{-4}$
Sodium, potassium, sulfur	0.10	Selenium, manganese, arsenic, nickel	$2 \times 10^{-5}$
Chlorine	0.09	Lead, cobalt	$9 \times 10^{-6}$

*Source:* Adapted from reference 2.



**Figure 1.1** Dose-response curve for an essential element. (Used with permission from reference 2. Copyright 1985, Division of Chemical Education, Inc.)

concentration in seawater, for instance, may not reflect accurately its pre-life-forms availability. If one assumes a mostly reducing atmosphere contemporary with the beginnings of biological life, the availability of the more soluble iron(II) ion in primordial oceans must have been much higher. Thus, the essentiality of iron(II) at a concentration of 0.02 mM in the blood plasma heme (hemoglobin) and muscle tissue heme (myoglobin) may be explained. Besides the availability factor, many chemical and physical properties of elements and their ions are responsible for their inclusion in biological systems. These include: ionic charge, ionic radius, ligand preferences, preferred coordination geometries, spin pairings, systemic kinetic control, and the chemical reactivity of the ions in solution. These factors are discussed in detail by Frausto da Silva and Williams.<sup>3</sup>

### 1.3 METALS IN BIOLOGICAL SYSTEMS: A SURVEY

Metals in biological systems function in a number of different ways. Group 1 and 2 metals operate as structural elements or in the maintenance of charge and osmotic balance (Table 1.2). Transition metal ions that exist in single oxidation states, such as zinc(II), function as structural elements in superoxide dismutase and zinc fingers, or, as an example from main group +2 ions, as triggers for protein activity—that is, calcium ions in calmodulin or troponin C

**TABLE 1.2 Metals in Biological Systems: Charge Carriers**

Metal	Coordination Number, Geometry	Preferred Ligands	Functions and Examples
Sodium, Na <sup>+</sup>	6, octahedral	O-Ether, hydroxyl, carboxylate	Charge carrier, osmotic balance, nerve impulses
Potassium, K <sup>+</sup>	6–8, flexible	O-Ether, hydroxyl, carboxylate	Charge carrier, osmotic balance, nerve impulses

**TABLE 1.3 Metals in Biological Systems: Structural, Triggers**

Metal	Coordination Number, Geometry	Preferred Ligands	Functions and Examples
Magnesium, Mg <sup>2+</sup>	6, octahedral	O-Carboxylate, phosphate	Structure in hydrolases, isomerases, phosphate transfer, trigger reactions
Calcium, Ca <sup>2+</sup>	6–8, flexible	O-Carboxylate, carbonyl, phosphate	Structure, charge carrier, phosphate transfer, trigger reactions
Zinc, Zn <sup>2+</sup> ( $d^{10}$ )	4, tetrahedral	O-Carboxylate, carbonyl, <i>S</i> -thiolate <i>N</i> -imidazole	Structure in zinc fingers, gene regulation, anhydrases, dehydrogenases
Zinc, Zn <sup>2+</sup> ( $d^{10}$ )	5, square pyramid	O-Carboxylate, carbonyl, <i>N</i> -imidazole	Structure in hydrolases, peptidases
Manganese, Mn <sup>2+</sup> ( $d^5$ )	6, octahedral	O-Carboxylate, phosphate, <i>N</i> -imidazole	Structure in oxidases, photosynthesis
Manganese, Mn <sup>3+</sup> ( $d^4$ )	6, tetragonal	O-Carboxylate, phosphate, hydroxide	Structure in oxidases, photosynthesis

(Table 1.3). Transition metals that exist in multiple oxidation states serve as electron carriers—that is, iron ions in cytochromes or in the iron–sulfur clusters of the enzyme nitrogenase or copper ions in cytochrome c oxidase (Cu<sub>A</sub> site), azurin and plastocyanin (Table 1.4); as facilitators of oxygen transport—that is, iron ions in hemoglobin or copper ions in hemocyanin (Table 1.5); and as sites at which enzyme catalysis occurs—that is, copper ions in superoxide dismutase or cytochrome c oxidase (Cu<sub>B</sub> site), iron ions in heme a<sub>3</sub> of cytochrome c oxidase, or iron and molybdenum ions in nitrogenase (Table 1.6). Metal ions may serve multiple functions, depending on their oxidation state or location within the biological system so that the classifications in Tables 1.2–1.6 are necessarily incomplete, arbitrary, and/or overlapping.<sup>4,5</sup>

**TABLE 1.4 Metals in Biological Systems: Electron Transfer**

Metal	Coordination Number, Geometry	Preferred Ligands	Functions and Examples
Iron, Fe <sup>2+</sup> ( $d^6$ )	4, tetrahedral	<i>S</i> -Thiolate	Electron transfer, nitrogen fixation in nitrogenases
Iron, Fe <sup>2+</sup> ( $d^6$ )	6, octahedral	<i>O</i> -Carboxylate, alkoxide, oxide, phenolate	Electron transfer in oxidases
Iron, Fe <sup>3+</sup> ( $d^5$ )	4, tetrahedral	<i>S</i> -Thiolate	Electron transfer, nitrogen fixation in nitrogenases
Iron, Fe <sup>3+</sup> ( $d^5$ )	6, octahedral	<i>O</i> -Carboxylate, alkoxide, oxide, phenolate	Electron transfer in oxidases
Copper, Cu <sup>+</sup> ( $d^{10}$ ), Cu <sup>2+</sup> ( $d^9$ )	3, trigonal planar	<i>N</i> -Imidazole	Electron transfer in Type III heme–copper oxidases (Cu <sub>B</sub> in cytochrome c oxidase, for example)
Copper, Cu <sup>+</sup> ( $d^{10}$ ), Cu <sup>2+</sup> ( $d^9$ )	4, tetrahedral	<i>S</i> -Thiolate, thioether, <i>N</i> -imidazole	Electron transfer in Type I blue copper proteins and Type III heme– copper oxidases (Cu <sub>A</sub> in cytochrome c oxidase, for example)

**TABLE 1.5 Metals in Biological Systems: Dioxygen Transport**

Metal	Coordination Number, Geometry	Preferred Ligands	Functions and Examples
Copper, Cu <sup>2+</sup> ( $d^9$ )	5, square pyramid 6, tetragonal	<i>O</i> -Carboxylate, <i>N</i> -imidazole	Type II copper oxidases, hydroxylases Type III copper hydroxylases, dioxygen transport in hemocyanin
Iron, Fe <sup>2+</sup> ( $d^6$ )	6, octahedral	<i>N</i> -Imidazole, porphyrin	Dioxygen transport in hemoglobin and myoglobin

**TABLE 1.6** Metals in Biological Systems: Enzyme Catalysis

Metal	Coordination Number, Geometry	Preferred Ligands	Functions and Examples
Copper, Cu <sup>2+</sup> ( $d^9$ )	4, square planar	O-Carboxylate, <i>N</i> -imidazole	Type II copper in oxidases
Cobalt, Co <sup>2+</sup> ( $d^7$ )	4, tetrahedral	S-Thiolate, thioether, <i>N</i> -imidazole	Alkyl group transfer, oxidases
Cobalt, Co <sup>3+</sup> ( $d^6$ )	6, octahedral	O-Carboxylate, <i>N</i> -imidazole	Alkyl group transfer in vitamin B <sub>12</sub> (cyanocobalamin)
Cobalt, Co <sup>2+</sup> ( $d^7$ )	6, octahedral	O-Carboxylate <i>N</i> -imidazole	Alkyl group transfer in vitamin B <sub>12r</sub>
Cobalt, Co <sup>+</sup> ( $d^8$ )	6, octahedral, usually missing the 6th ligand	O-Carboxylate, <i>N</i> -imidazole	Alkyl group transfer in vitamin B <sub>12s</sub>
Nickel, Ni <sup>2+</sup> ( $d^8$ )	4, square planar	S-Thiolate, thioether, <i>N</i> -imidazole, polypyrrole	Hydrogenases, hydrolases
Nickel, Ni <sup>2+</sup> ( $d^8$ )	6, octahedral		Uncommon
Molybdenum, Mo <sup>4+</sup> ( $d^2$ ), Mo <sup>5+</sup> ( $d^1$ ), Mo <sup>6+</sup> ( $d^0$ )	6, octahedral	O-Oxide, carboxylate, phenolate, <i>S</i> -sulfide, thiolate	Nitrogen fixation in nitrogenases, oxo transfer in oxidases

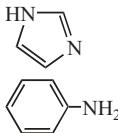
## 1.4 INORGANIC CHEMISTRY BASICS

Ligand preference and possible coordination geometries of the metal center are important bioinorganic principles. Metal ligand preference is closely related to the hard–soft acid–base nature of metals and their preferred ligands. These are listed in Table 1.7.<sup>6</sup>

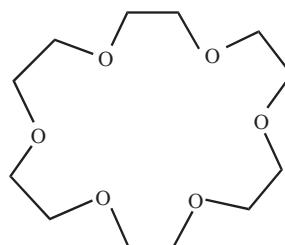
In general, hard metal cations form their most stable compounds with hard ligands, whereas soft metal cations form their most stable compounds with soft ligands. Hard cations can be thought of as small dense cores of positive charge, whereas hard ligands are usually the small highly electronegative elements or ligand atoms within a hard polyatomic ion—that is, oxygen ligands in  $(RO_2)PO_2^-$ , or  $CH_3CO_2^-$ . Crown ethers are hard ligands that have cavities suitable for encapsulating hard metal ions. The [18]-crown-6 ether shown in Figure 1.2 with its 2.6 to 3.2-Å hole provides a good fit for the potassium ion, which has a radius of 2.88 Å.<sup>6</sup>

It is possible to modify a hard nitrogen ligand toward an intermediate softness by increasing the polarizability of its substituents or the  $\pi$  electron cloud

**TABLE 1.7 Hard–Soft Acid–Base Classification of Metal Ions and Ligands**

Metals, Ions, Molecules				Ligands
<b>HARD</b>				<b>HARD</b>
$\text{H}^+$	$\text{Mg}^{2+}$	$\text{Al}^{3+}$	$\text{SO}_3$	Oxygen ligands in $\text{H}_2\text{O}$ , $\text{CO}_3^{2-}$ , $\text{NO}_3^-$ , $\text{PO}_4^{3-}$ , $\text{ROPO}_3^{2-}$ , $(\text{RO})_2\text{PO}_2^-$ , $\text{CH}_3\text{COO}^-$ , $\text{OH}^-$ , $\text{RO}^-$ , $\text{R}_2\text{O}$ , and crown ethers
$\text{Na}^+$	$\text{Ca}^{2+}$	$\text{Co}^{3+}$	$\text{CO}_2$	Nitrogen ligands in $\text{NH}_3$ , $\text{N}_2\text{H}_4$ , $\text{RNH}_2$ , or $\text{Cl}^-$
$\text{K}^+$	$\text{Mn}^{2+}$	$\text{Cr}^{3+}$		
	$\text{VO}^{2+}$	$\text{Ga}^{3+}$		
		$\text{Fe}^{3+}$		
		$\text{Tl}^{3+}$		
		$\text{Ln}^{3+}$		
		$\text{MoO}^{3+}$		
<b>INTERMEDIATE</b>				<b>INTERMEDIATE</b>
$\text{Fe}^{2+}$ , $\text{Ni}^{2+}$ , $\text{Zn}^{2+}$ , $\text{Co}^{2+}$ , $\text{Cu}^{2+}$ , $\text{Pb}^{2+}$ , $\text{Sn}^{2+}$ , $\text{Ru}^{2+}$ , $\text{Au}^{3+}$ , $\text{SO}_2$ , $\text{NO}^+$				$\text{Br}^-$ , $\text{SO}_3^{2-}$ , Nitrogen ligands in $\text{NO}_2^-$ , $\text{N}_3^-$ , $\text{N}_2$ ,
				 The image shows two chemical structures. The top structure is imidazole, which consists of a five-membered ring containing two nitrogen atoms (one double-bonded, one single-bonded) and three carbon atoms. The bottom structure is aniline, which consists of a benzene ring attached to an amino group (-NH <sub>2</sub> ).
<b>SOFT</b>				<b>SOFT</b>
$\text{Cu}^+$	$\text{Pt}^{2+}$	$\text{Pt}^{4+}$		Sulfur ligands in $\text{RSH}$ , $\text{RS}^-$ , $\text{R}_2\text{S}$ , $\text{R}_3\text{P}$ , $\text{RNC}$ , $\text{CN}^-$ , $\text{CO}$ , $\text{R}^-$ , $\text{H}^-$ , $\text{I}^-$ , $\text{S}_2\text{O}_3^{2-}$ , $(\text{RS})_2\text{PO}_2^-$ , $(\text{RO})_2\text{P(O)S}^-$
$\text{Au}^+$	$\text{Pb}^{2+}$			
$\text{Tl}^+$	$\text{Hg}^{2+}$			
$\text{Ag}^+$	$\text{Cd}^{2+}$			
$\text{Hg}_2^{2+}$	$\text{Pd}^{2+}$			

Source: Adapted from references 4 and 6.



**Figure 1.2** [18]-Crown-6 ether.

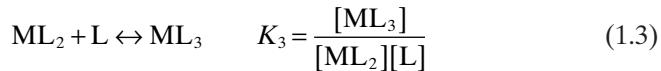
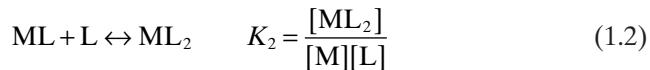
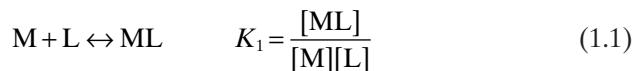
about it, an example being the imidazole nitrogen of the amino acid histidine. Increasing the softness of phosphate ion substituents can transform the hard oxygen ligand of  $(\text{RO})_2\text{PO}_2^-$  to a soft state in  $(\text{RS})_2\text{PO}_2^-$ . Soft cations and anions are those with highly polarizable, large electron clouds—that is,  $\text{Hg}^{2+}$ , sulfur ligands as sulfides or thiolates, and iodide ions. Also, note that metal ions can overlap into different categories. Lead as  $\text{Pb}^{2+}$ , for instance, appears in both the intermediate and soft categories. The  $\text{Fe}^{3+}$  ion, classified as a hard

cation, coordinates to histidine (imidazole) ligands in biological systems, whereas  $\text{Fe}^{2+}$ , classified as intermediate, can coordinate to sulfur ligands and the carbon atom of CO (see Section 7.2, for example, in which hemoglobin and myoglobin are discussed).

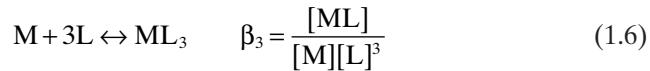
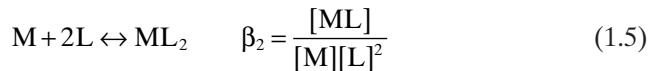
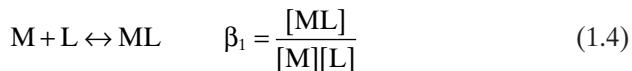
## 1.5 BIOLOGICAL METAL ION COMPLEXATION

### 1.5.1 Thermodynamics

The thermodynamic stability of metal ions is denoted by stepwise formation constants as shown in equations 1.1–1.3 (charges omitted for simplicity):



Alternately, they are indicated by overall stability constants as shown in equations 1.4–1.6:



The equation relating the stepwise and overall stability constants is indicated by equation 1.7:

$$\beta_n = K_1 K_2 \dots K_n \quad (1.7)$$

In biological systems, many factors affect metal–ligand complex formation. Hard–soft acid–base considerations have already been mentioned. Concentrations of the metal and ligand at the site of complexation are determined locally through concentration gradients, membrane permeability to metals and ligands, and other factors. Various competing equilibria—solubility products, complexation, and/or acid–base equilibrium constants—sometimes referred to as “metal ion speciation,” all affect complex formation. Ion size and charge, preferred metal coordination geometry, and ligand chelation effects all affect metal uptake. To better measure biological metal–ligand interactions, an

**TABLE 1.8**  $K_{ML}$  and  $K_{ML} \times [M]$  for Some Cations and Their Differentiating Ligands

	$K^+, Na^+$	$Ca^{2+}, Mg^{2+}$	$Zn^{2+}, Cu^{2+}$	Differentiating Ligand
$K^+, Na^+$				
$K_{ML}$	>10	<10 <sup>2</sup>	<10 <sup>6</sup>	<i>O</i> -Macrocycles such as crown ethers, cryptates and naturally occurring macrocyclic antibiotics such as nonactin and valinomycin
$K_{ML} \times [M]$	>1.0	<0.1	<0.1	
$Ca^{2+}, Mg^{2+}$				Oxygen donors such as di- or tricarboxylates
$K_{ML}$	1.0	<10 <sup>3</sup>	<10 <sup>6</sup>	
$K_{ML} \times [M]$	<0.1	>1.0	<0.1	
$Zn^{2+}, Cu^{2+}$				Nitrogen and sulfur ligands
$K_{ML}$	0.1	<10 <sup>2</sup>	>10 <sup>6</sup>	
$K_{ML} \times [M]$	<0.1	<0.1	>1.0	

Source: Adapted from reference 3.

“uptake factor” is defined as  $K_{ML} \times [M]$ , where  $K_{ML}$  is the stability constant  $K_1$  and  $[M]$  is the concentration of metal ion. Since naturally occurring aqueous systems have metal ion concentration varying roughly as

$K^+, Na^+$	$Ca^{2+}, Mg^{2+}$	$Zn^{2+}$	$Cu^{2+}$	$Fe^{2+}$
$10^{-1} M$	$\sim 10^{-3} M$	$<10^{-9} M$	$<10^{-12} M$	$\sim 10^{-17} M$

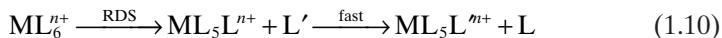
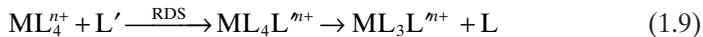
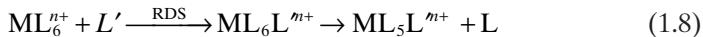
great selectivity for metal species is necessary to concentrate the necessary ions at sites where they are needed. Differentiating ligands are those preferred by the cation in question. A much more detailed discussion takes place in reference 3. Table 1.8 is adapted from this source.

### 1.5.2 Kinetics

As students learned in their introductory chemistry courses, rates of reaction are divided into several classes, depending on the number of reactants involved in rate determination. These are: (1) zero order—the reaction rate is independent of the concentration of that reactant; (2) first order—the reaction rate is dependent on the concentration of one reactant; (3) second order—the reaction rate is dependent on the concentration of two reactants; and (4) higher order—the reaction rate is dependent on more than two reactants. Higher-order reaction rates are very rare because the possibility of bringing more than two reactants together productively is very small. Bioinorganic kineticists, studying the reaction rates of complex enzymatic reactions, often simplify matters to isolate a reaction of interest and relate it to a proposed mechanism for the enzyme’s catalytic activity. For instance, in a pseudo-zero-order reaction—that is, one that would be first order under normal circumstances—the concentration of the enzyme may be held constant while a particular substrate’s concentration is varied but does not affect the reaction rate. This condition may apply when the enzyme is saturated with substrate over the range of

substrate concentration studied. In a pseudo-first-order reaction—that is, one that would normally be second order—the concentration of one reactant is held constant while the other is varied so that the reaction rate is directly proportional to the reactant whose concentration is varied. This is the most commonly used experimental technique used by enzyme kineticists.

In biological systems, as in all others, metal ions exist in an inner coordination sphere, an ordered array of ligands binding directly to the metal. Surrounding this is the outer coordination sphere consisting of other ligands, counterions, and solvent molecules. In stoichiometric mechanisms where one can distinguish an intermediate, substitution within the metals inner coordination sphere may take place through an associative (A),  $S_N2$  process as shown in equations 1.8 (for six-coordinate complexes) and 1.9 (for four-coordinate complexes) or a dissociative (D),  $S_N1$  mechanism as shown in equation 1.10 (RDS = rate determining step).



Associative mechanisms for metals in octahedral fields are difficult stereochemically (due to ligand crowding); therefore, they are rare for all but the largest metal ion centers. The associative mechanism is well known and preferred for four-coordinate square-planar complexes. Pure dissociative mechanisms are rare as well. When an intermediate cannot be detected by kinetic, stereochemical, or product distribution studies, the so-called interchange mechanisms (I) are invoked. Associative interchange ( $I_A$ ) mechanisms have rates dependent on the nature of the entering group, whereas dissociative interchange ( $I_D$ ) mechanisms do not.

The simplest reactions to study, those of coordination complexes with solvent, are used to classify metal ions as labile or inert. Factors affecting metal ion lability include size, charge, electron configuration, and coordination number. Solvents can be classified as to their size, polarity, and the nature of the donor atom. Using the water exchange reaction for the aqua ion  $[\text{M}(\text{H}_2\text{O})_n]^{m+}$ , metal ions are divided by Cotton, Wilkinson, and Gaus<sup>7</sup> into four classes:

**Class I.** Rate constants for water exchange exceed  $10^8 \text{ s}^{-1}$ , essentially diffusion controlled. These are classified as the labile species.

**Class II.** Rate constants for water exchange are in the range  $10^4$ – $10^8 \text{ s}^{-1}$ .

**Class III.** Rate constants for water exchange are in the range  $1$ – $10^4 \text{ s}^{-1}$ .

**Class IV.** Rate constants for water exchange are in the range  $10^{-3}$ – $10^{-6} \text{ s}^{-1}$ .

These ions are classified as inert.

Labile species are usually main group metal ions with the exception of  $\text{Cr}^{2+}$  (high-spin  $3d^4$ ) and  $\text{Cu}^{2+}$  ( $3d^9$ ) whose lability can be ascribed to Jahn–Teller

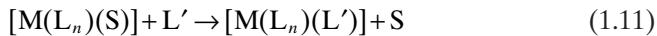
**TABLE 1.9 Rate Constants for Water Exchange in Metal Aqua Ions**

Class	Metal Ions	Rates $\log k$ ( $s^{-1}$ )
I	Group IA (1), Group IIA (2) except Be and Mg, Group IIB (12) except $Zn^{2+}$ ( $3d^{10}$ ), $Cr^{2+}$ ( $3d^4$ ), $Cu^{2+}$ ( $3d^9$ )	8–9
II	$Zn^{2+}$ ( $3d^{10}$ ) $Mn^{2+}$ ( $3d^5$ ) $Fe^{2+}$ ( $3d^6$ ) $Co^{2+}$ ( $3d^7$ ) $Ni^{2+}$ ( $3d^8$ ) $Mg^{2+}$	7.6 6.8 6.3 5.7 4.3 6.0
III	$Ga^{3+}$ $Be^{2+}$ $V^{2+}$ ( $3d^3$ ) $Al^{3+}$	3.0 2.0 2.0 $<0.1$
IV	$Cr^{3+}$ ( $3d^3$ ), $Co^{3+}$ ( $3d^6$ ), $Rh^{3+}$ ( $3d^6$ ), $Ir^{3+}$ ( $3d^6$ ), $Pt^{2+}$ ( $3d^8$ )	-3 to -6

Source: Adapted from references 7 and 8.

effects. Section 1.6 includes a formula for determining the number of  $d$  electrons in a transition metal ion, and Figures 1.4 and 1.7 show the placement of  $d$  electrons into nondegenerate (split)  $d$  orbitals in various ligand fields. Jahn-Teller effects arise (for the high-spin  $3d^4$  case) because the lone electron in the two destabilized, but degenerate,  $e_g$  orbitals causes further splitting of the  $e_g$  level with consequences for bond lengths between the metal ion and its ligands. Filling in the octahedral energy level diagram for the  $Cu^{2+}$  ( $3d^9$ ) case in Figure 1.4, readers should be able to show three electrons in the  $e_g$  level, again causing a loss of degeneracy in these orbitals. Transition metals of classes II and III are species with small ligand field stabilization energies, whereas the inert species have high ligand field stabilization energies (LFSE). Examples include  $Cr^{3+}$  ( $3d^3$ ) and  $Co^{3+}$  ( $3d^6$ ). Jahn-Teller effects and LFSE are further discussed in Section 1.6. Table 1.9 reports rate constant values for some aqueous solvent exchange reactions.<sup>8</sup>

Outer-sphere (OS) reaction rates and rate laws can be defined for solvolysis of a given complex. Complex formation is defined as the reverse reaction—that is, replacement of solvent (S) by another ligand (L'). Following the arguments of Tobe,<sup>9</sup> in aqueous solution the general rate law for complex formation (eliminating charge for simplicity),



takes the second-order form shown in equation 1.12:

$$-d \frac{[M(L_n)(S)]}{dt} = k'[M(L_n)(S)][L'] \quad (1.12)$$

The rate law frequently may be more complex and given as equation 1.13:

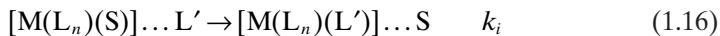
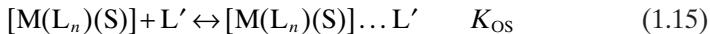
$$-d \frac{[M(L_n)(S)]}{dt} = \frac{k'K[M(L_n)(S)][L']}{(1+K[L'])} \quad (1.13)$$

Equation 1.13 reduces to the second-order rate law, shown in equation 1.12, when  $K[L'] \ll 1$  and to a first-order rate law, equation 1.14,

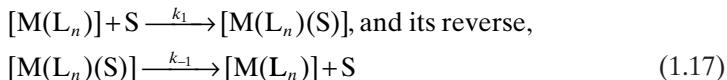
$$-d \frac{[M(L_n)(S)]}{dt} = k'[M(L_n)(S)] \quad (1.14)$$

when  $K[L'] \gg 1$ .

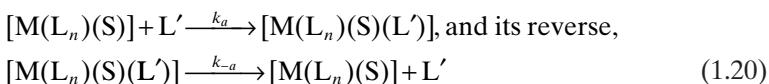
Interchange mechanisms ( $I_A$  or  $I_D$ ) in a preformed outer sphere (OS) complex will generate the following observed rate laws (which cannot distinguish  $I_A$  from  $I_D$ ) with the equilibrium constant =  $K_{OS}$  (equation 1.15) and  $k = k_i$  (equation 1.16).



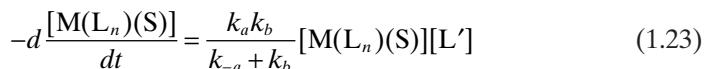
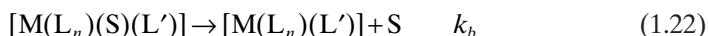
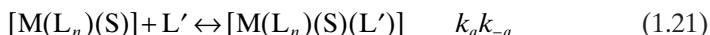
The dissociative (D or  $S_N1$ ) mechanism, for which the intermediate is long-lived enough to be detected, will yield equations 1.18 and 1.19 where  $k = k_1$  and  $K = k_2/(k_{-1}[S])$ . For the reaction:



The associative (A or  $S_N2$ ) will give the simple second-order rate law shown in equations 1.21 and 1.22 if the higher coordination number intermediate concentration remains small, resulting in the rate dependence shown in equation 1.23. For the reaction



we have



In all cases the key to assigning mechanism is the ability to detect and measure the equilibrium constant  $K$ . The equilibrium constant  $K_{OS}$  can be estimated through the Fuoss–Eigen equation<sup>10</sup> as shown in equation 1.24. Usually,  $K_{OS}$  is ignored in the case of  $L' = \text{solvent}$ .

$$K_{\text{os}} = \frac{4\pi N_A a^3}{3000} (e^{-V/kT}) \quad (1.24)$$

where  $a$  is the distance of closest approach of the oppositely charged ions ( $\sim 5 \text{ \AA}$ ),  $N_A$  is Avogadro's number, and  $V$  is the electrostatic potential at that distance (equation 1.25).

$$V = \frac{Z_1 Z_2 e^2}{4\pi \epsilon_0 \epsilon_R a} \quad (1.25)$$

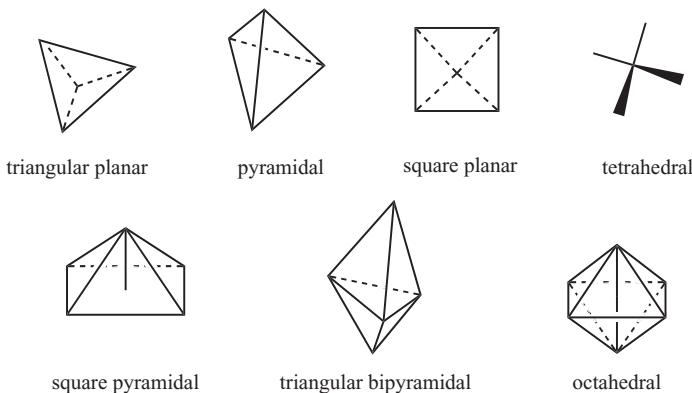
where

- $a$  = distance of closest approach of oppositely charged ions ( $\sim 5 \text{ \AA}$ )
- $N_A$  = Avogadro's number,  $6.022 \times 10^{23} \text{ mol}^{-1}$
- $V$  = electrostatic potential (dependent on distance between oppositely charged ions)
- $k$  = rate constant for a reaction
- $K$  = equilibrium constant for a reaction
- $Z_1 Z_2$  = absolute value of the charge on an ion
- $e$  = charge on the electron,  $4.8030 \times 10^{-10} \text{ esu}$  or  $1.6022 \times 10^{-19} \text{ Coulombs (C)}$
- $\epsilon_0 \epsilon_R$ :  $\epsilon_0$  = permittivity in a vacuum  $8.854 \times 10^{-12} (\text{C}^2/\text{Jm})$ ,  $\epsilon_R$  or  $\epsilon_r$  = dielectric constant = relative permittivity = 1 (for vacuum by definition, 80.4 for  $\text{H}_2\text{O}$  at  $20^\circ\text{C}$ ),  $\epsilon_0 \epsilon_R$  = actual permittivity

As the above discussion indicates, assigning mechanisms to simple anion reactions of transition metal complexes is not simple. The situation becomes even more difficult for a complex enzyme system containing a metal cofactor at an active site. Methods developed to study the kinetics of enzymatic reactions according to the Michaelis–Menten model will be discussed in Section 2.2.4. Since enzyme-catalyzed reactions are usually very fast, experimentalists have developed rapid kinetic techniques to study them. Techniques used by bioinorganic chemists to study reaction rates will be further detailed in Section 3.7.2.1 and 3.7.2.2.

## 1.6 ELECTRONIC AND GEOMETRIC STRUCTURES OF METALS IN BIOLOGICAL SYSTEMS

Tables 1.2–1.6 list some of the important geometries assumed by metal ions in biological systems. Common geometries adopted by transition metal ions that will be of most concern to readers of this text are illustrated in Figure 1.3. It is important to remember that in biological systems these geometries are usually distorted in both bond length and bond angle.



**Figure 1.3** Common transition metal coordination geometries.

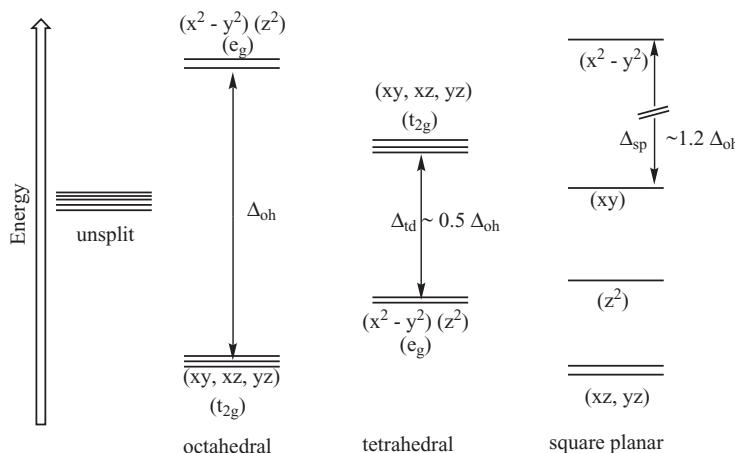
Transition metal ions play special roles in biological systems, with all elements from the first transition series except titanium (Ti) and scandium (Sc) occurring with great variety in thousands of diverse metalloproteins. Metal ions determine the geometry of enzymatic active sites, act as centers for enzyme reactivity, and act as biological oxidation–reduction facilitators. Molybdenum (Mo) appears to be the only transition element in the second transition series with a similar role. Vanadium (V), technetium (Tc), platinum (Pt), ruthenium (Ru), and gold (Au) compounds, as well as gadolinium (Gd) and other lanthanide complexes, are extremely important in medicinal chemistry. Tables 1.2–1.6 list the *d* electron configuration for transition metal ions common to biological systems. To find the number of *d* electrons for any transition metal ion, the following is a useful formula:

$$\begin{aligned} \text{Number of } d \text{ electrons} = & \text{ Atomic number for the element (Z)} \\ & - \text{oxidation state of the element's ion} - Z \text{ for the preceding} \\ & \text{noble-gas element.} \end{aligned}$$

$$\begin{array}{lll} \text{Examples:} & \text{Fe(II):} & 26 - 2 - 18 \text{ (argon)} = 6 \\ & \text{Mo(V):} & 42 - 5 - 36 \text{ (krypton)} = 1 \end{array}$$

Note that there are a number of different methods for indicating the oxidation state of a metal ion, especially transition metal ions that have variable oxidation states. As an example, the iron ion in its +2 oxidation state may be written as  $\text{Fe}^{2+}$ , Fe(II),  $\text{Fe}^{\text{II}}$ , or iron(II). In this text, the methods are used interchangeably.

As a consequence of their partially filled *d* orbitals, transition metals exhibit variable oxidation states and a rich variety of coordination geometries and ligand spheres. Although a free metal ion would exhibit degenerate *d* electron energy levels, ligand field theory describes the observed splitting of these *d* electrons for metal ions in various ligand environments. In all cases, the amount of stabilization or destabilization of *d* electron energy levels centers about the

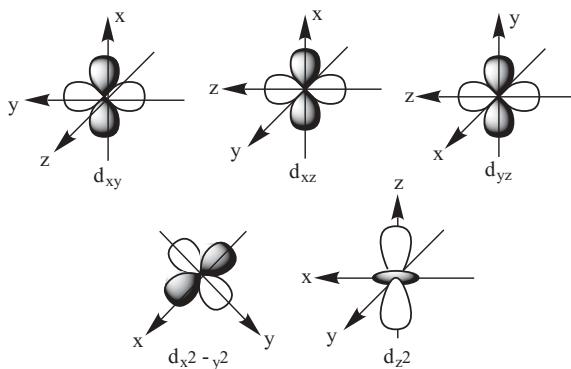


**Figure 1.4** Approximate energy levels for  $d$  electrons in octahedral, tetrahedral, and square-planar fields.

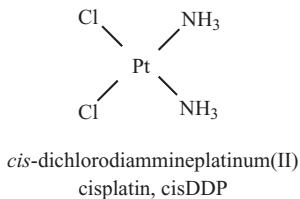
so-called barycenter of unsplit  $d$  electron energy levels. The most important splittings for bioinorganic applications are shown in Figure 1.4 for octahedral, tetrahedral, and square-planar ligand fields. The  $t_{2g}$  ( $d_{xy}$ ,  $d_{yz}$ , and  $d_{xz}$ ) and  $e_g$  ( $d_{x^2-y^2}$  and  $d_{z^2}$ ) energy level designations identify symmetry properties of the  $d$  orbitals and are often used to indicate the degenerate energy levels under discussion. (See LFSE discussion below). Generally, the energy gap between stabilized and destabilized  $d$ -electron energy levels for tetrahedral fields ( $\Delta_t$ ) is approximately one-half that for octahedral fields ( $\Delta_{oh}$ ), and that for square-planar fields is approximately  $1.2\Delta_{oh}$ . Many thermodynamic and kinetic properties of transition metal coordination complexes can be predicted by knowing the magnitude of  $\Delta$ . Measurement of ultraviolet and visible absorption spectra of transition metal complexes that arise from these quantum mechanically forbidden (but observed)  $d$ - $d$  transitions provide a measure of  $\Delta$ .

To describe the  $d$ -orbital splitting effect for the octahedral field, one should imagine ligand spheres of electron density approaching along the  $x$ ,  $y$ , and  $z$  axes where the  $d_{x^2-y^2}$  and  $d_{z^2}$  lobes of electron density point. Figure 1.5 illustrates representations of high-probability electron orbit surfaces for the five  $d$  orbitals.

For octahedral ( $O_h$ ) geometry the repelling effect of like charge approach of the ligand electrons toward regions of high  $d$  electron density along the  $x$ ,  $y$ , and  $z$  axes elevates the energy of the  $e_g$  ( $d_{x^2-y^2}$  and  $d_{z^2}$ ) orbitals while the  $t_{2g}$  ( $d_{xy}$ ,  $d_{yz}$ , and  $d_{xz}$ ) orbitals are proportionally lowered in energy. For the tetrahedral ( $T_d$ ) case ligands approach between the  $x$ ,  $y$ , and  $z$  axes, stabilizing  $d_{x^2-y^2}$  and  $d_{z^2}$  and destabilizing  $d_{xy}$ ,  $d_{yz}$ , and  $d_{xz}$  orbital energy levels. For the square-planar case, ligands will approach along the  $x$  and  $y$  axes. Distorted octahedral and tetrahedral geometries are quite common in biological systems. Octahedral geometries are found for iron ions in heme ligand systems to be discussed in Chapter 7—for instance, while copper ions occur in distorted



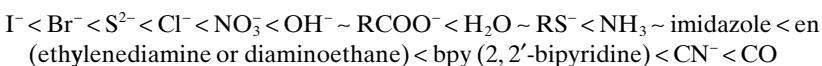
**Figure 1.5** Representations of the five  $d$  orbitals along  $x$ ,  $y$ , and  $z$  axes.



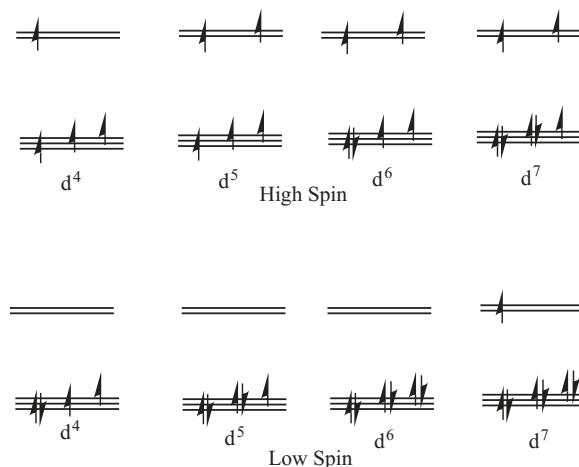
**Figure 1.6** The antitumor active platinum compound *cis*-dichlorodiammineplatinum(II).

pyramidal, tetrahedral, or even trigonal bipyramidal forms. Less commonly, square-planar geometries are found for  $d^8$  transition metal ions, especially for gold(III), iridium(I), palladium(II), and platinum(II) and for nickel(II) species in strong ligand fields. The platinum anticancer agent, *cis*-dichlorodiammineplatinum(II), shown in Figure 1.6, has a square-planar geometry all important for its utilization as an antitumor agent. While the other geometries shown in Figure 1.3 might be less common for metal ions in biological species, they do occur (also with distorted bond distances and angles) and will be described in discussions of the metal center in the specific protein or enzyme.

The strength of the ligand field at a metal center is strongly dependent on the character of the ligand's electronic field and leads to the classification of ligands according to a "spectrochemical series" arranged below in order from weak field (halides, sulfides, hydroxides) to strong field (cyanide and carbon monoxide):



Ligand field strength may determine coordination geometry. For example,  $\text{NiCl}_4^{2-}$  occurs as a tetrahedral complex (small splitting—small  $\Delta_{\text{td}}$ ), whereas



**Figure 1.7** High-spin and low-spin  $d$ -electron configurations for the octahedral field.

$\text{Ni}(\text{CN})_4^{2-}$  occurs in the square-planar geometry (large energy gap—large  $\Delta_{\text{sp}}$ ). In octahedral fields, ligand field strength can determine the magnetic properties of metal ions since for  $d^4$  through  $d^7$  electronic configurations both high-spin (maximum unpairing of electron spins) and low-spin (maximum pairing of electron spins) complexes are possible. Possible configurations are shown in Figure 1.7. In general, weak field ligands form high-spin (h.s.) complexes (small  $\Delta_{\text{oh}}$ ) and strong field ligands form low-spin (l.s.) complexes (large  $\Delta_{\text{oh}}$ ). Usually, tetrahedral complexes have high spin (small  $\Delta_{\text{td}}$ ) and will be paramagnetic. Square-planar complexes, usually found for metal ions having the electron configuration  $d^8$ , will be diamagnetic—all electrons paired—since a large energy gap occurs between the last filled orbital ( $d_{xy}$ ) and the  $d_{x^2-y^2}$  orbital (see Figure 1.4). Detection of paramagnetism (unpaired electrons) and diamagnetism (all electrons paired) in bioinorganic ligand fields can help determine coordination geometry at active enzymes sites. In the case of hemoglobin, for example, the  $d^6$  iron(II) center cycles between high-spin (paramagnetic) and low-spin (diamagnetic) configurations. The change is evident in electron paramagnetic resonance (EPR) spectroscopy in which a spectrum is determined only for paramagnetic species. See Section 3.5. Placement of  $d$  electrons also affects the placement of the iron center in or out of the plane of its porphyrin ligand in hemoglobin or myoglobin—high-spin systems require more room so that a h.s. Fe(II) ion will move out of the porphyrin ligand's planar coordination sphere. See Section 7.2 for further discussion with respect to this phenomenon in myoglobin and hemoglobin. In Type III copper enzymes, two  $d^9$  copper(II) centers become antiferromagnetically coupled resulting in a loss of the expected paramagnetism. See Section 7.8 for a discussion of binuclear copper centers in cytochrome c oxidase.

The sum of the  $d$  electron contributions to LFSE can be calculated with the formula shown in equation 1.26 for octahedral complexes.

$$\text{LFSE} = \frac{2}{5}(\# e^- \text{ in } t_{2g})\Delta_{\text{oh}} + \frac{3}{5}(\# e^- \text{ in } e_g)\Delta_{\text{oh}} \quad (1.26)$$

where  $\#e^-$  is the number of  $d$  electrons.

The 2/5 stabilization (negative energy values) and 3/5 destabilization (positive energy values) modifiers arise from the displacement of three  $d$  orbitals to lower energy versus two  $d$  orbitals to higher energy from the unsplit degenerate  $d$  orbital state before imposition of the ligand field. Splitting values for  $d$  orbital energy levels, based on  $\Delta_{\text{oh}} = 10$ , has been adapted from reference 7 and appears in Table 1.10.

The Jahn–Teller effect arises in cases where removal of degeneracy of a  $d$ -electron energy level is caused by partial occupation of a degenerate level. Two common examples are those of Cu(II),  $d^9$ , and high spin Cr(II),  $d^4$ , as shown in Figure 1.8. Electrons in the  $e_g$  level could be placed in either the  $d_{x^2-y^2}$  and  $d_{z^2}$  orbitals. Placing the odd electron in either orbital destroys the degeneracy of the  $e_g$  orbitals and usually has the effect of moving the ligands on one axis in or out. For Cu(II) complexes this effect is very common, resulting in longer bond lengths on what is usually taken as the complex's  $z$  axis. The effect is also seen for high-spin  $d^4$  Mn(III) and for low-spin  $d^7$  Co(II) and Ni(III) complexes.

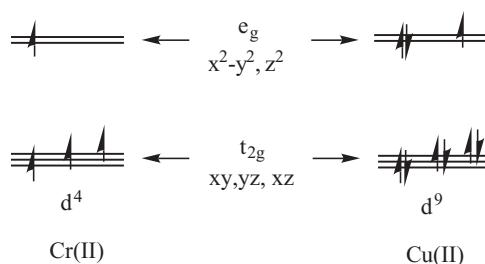
**TABLE 1.10** Splitting Values for d Orbitals in Common Geometries

C. N. <sup>a</sup>	Geometry	$d_{x^2-y^2}$	$d_{z^2}$	$d_{xy}$	$d_{xz}$	$d_{yz}$
4	Tetrahedral	-2.67	-2.67	1.78	1.78	1.78
4	Square planar <sup>b</sup>	12.28	-4.28	2.28	-5.14	-5.14
5	Square pyramidal <sup>c</sup>	9.14	0.86	-0.86	-4.57	-4.57
6	Octahedral	6.00	6.00	-4.00	-4.00	-4.00

<sup>a</sup>C. N. stands for coordination number.

<sup>b</sup>Bonds in xy plane.

<sup>c</sup>Pyramidal base in xy plane.



**Figure 1.8** Electron configurations for high-spin Cr(II) and Cu(II).

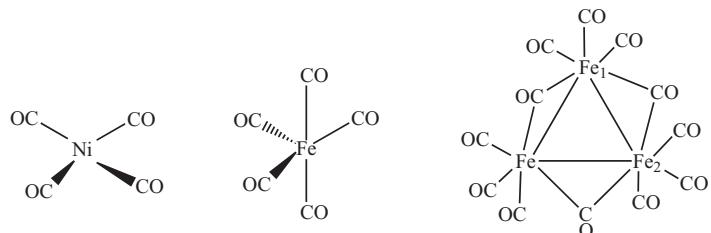
## 1.7 BIOORGANOMETALLIC CHEMISTRY

Organometallic complexes obey the so-called eighteen-electron (18-e) and sixteen-electron (16-e) rules as described below. These rules may also be applied to bioinorganic (bioorganometallic) systems. According to 16- or 18-e rules, the valence electrons of transition metals are considered to be filling the 4s, 3d, 4p or 5s, 4d, 5p shells. The most stabilized filled shell is determined to be eighteen electrons— $s^2, d^{10}, p^6$ , differing by the ten electrons of the filled d shell from main group element compounds stabilized by electron octets. Compounds or complexes fulfill the 18-e rule by addition of metal valence electrons and electron contributions from ligands. Metal valence shell electrons may be counted as if the metal, in its 0, +1, +2 oxidation states, combine with ligand electrons counted according to Table 1.11, which lists common ligands and their electron contributions. Table 1.11 counts electrons as if the ligands are neutral or are contributing electrons to the metal covalently. Another electron counting system classifies ligands according to an ionic contribution. Both these systems are more completely described in a website found at <http://www.ilpi.com/organomet/electroncount.html>.

Many stable coordination complexes can be counted as having sixteen electrons (16-e rule), especially those having square-planar geometry and those bonded to aromatic rings through their  $\pi$  electronic systems. Some of these complexes, belonging to a group of compounds called metallocenes, bind to DNA and have antitumor properties. Examples discussed in the literature usually contain two cyclopentadienyl ligands,  $\eta^5\text{-cp}$ , two chloride,  $\text{Cl}^-$ , ligands and the metals titanium, Ti, vanadium, V, molybdenum, Mo, or niobium, Nb. The complexes have the overall formula  $\text{Cp}_2\text{MCl}_2$  and are named, for example, as titanocene, vanadocene, or molybdocene complexes. The titanium antitumor complex, *bis*- $\eta^5$ -cyclopentadienyldichlorotitanium(II),  $\text{Cp}_2\text{TiCl}_2$ , is shown in Figure 1.9. This complex is the first metallocene, and the first non-platinum metal complex, to have undergone clinical trials as an anticancer agent.

**TABLE 1.11 Ligand Contributions to the 16- or 18-Electron Rule**

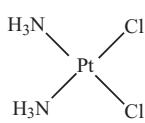
Ligand	Number of Electrons
Hydrogen $\text{H}^\bullet$ , chloride radical $\text{Cl}^\bullet$	1
Alkyl ( $\text{CH}_3, \text{CH}_3\text{CH}_2$ , etc) or acyl ( $\text{RC=O}$ ) groups	1
NO (bent)	1
Carbonyl groups ( $\text{RCOO}^-$ ), CO, $\text{CN}^-$ , RCN or RNC, ethers ( $\text{ROR}$ ), sulfides ( $\text{R}_2\text{S}$ ), ketones ( $\text{R}(\text{C=O})\text{R}$ )	2
Lewis bases $\text{Cl}^-$ , $\text{O}^{2-}$ , $\text{S}^{2-}$ , ammonia ( $\text{NH}_3$ ), amines ( $\text{NR}_3$ ), phosphines ( $\text{PR}_3$ )	2
Alkenes ( $\text{RCH=CH}_2$ )	2 per double bond
Nitrosyl group (NO) linear	3
Cyclopentadienyl, cp ( $\text{C}_5\text{H}_5$ )	5 per ring
Benzene ( $\text{C}_6\text{H}_6$ )	6 per ring ( $\pi$ donation)



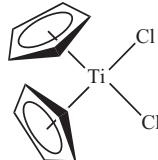
tetracarbonylnickel(0)  
 $\text{Ni}(0) = 10 \text{ e } (4s^23d^8)$   
 $4 \text{ CO} = 8 \text{ e}$   
Total = 18 e

pentacarbonyliron(0)  
 $\text{Fe}(0) = 8 \text{ e } (4s^23d^6)$   
 $5 \text{ CO} = 10 \text{ e}$   
Total = 18 e

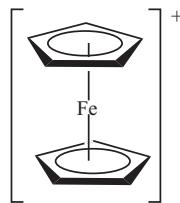
$\text{Fe}_1(0) = 8 \text{ e}$   
3 terminal CO = 6 e  
2 bridging CO = 2 e  
2 Fe-Fe bonds = 2 e  
Total = 18 e for each Fe



*cis*-dichlorodiammine-platinum(II)  
 $\text{Pt}(\text{II}) = 8 \text{ e}$   
 $2 \text{ NH}_3 = 4 \text{ e}$   
 $2 \text{ Cl}^- = 4 \text{ e}$   
Total = 16 e



bis- $\eta^5$ -cyclopentadienyl-dichlorotitanium(II)  
 $\text{Ti}(\text{II}) = 2 \text{ e}$   
 $2 \text{ cp} = 10 \text{ e}$   
 $2 \text{ Cl}^- = 4 \text{ e}$   
Total = 16 e



ferrocenium ion  
 $\text{Fe}(\text{I}) = 7 \text{ e}$   
 $2 \text{ cp} = 10 \text{ e}$   
+ charge = -1 e  
Total = 16 e

**Figure 1.9** Molecules obeying the 16-e and 18-e rules.

Structure–activity relationship studies have shown that the complex forms stable adducts with DNA.<sup>11</sup> Two unsubstituted cyclopentadienyl ligands appear to be important for activity because the compound  $(\text{MeCp})_2\text{TiCl}_2$  has been found to be biologically inactive. The corresponding molybdenum-containing metallocene,  $\text{Cp}_2\text{MoCl}_2$ , appears to have a different mode of cytotoxic (antitumor) activity because it does not form stable DNA adducts under the same conditions as the titanocene. Although the molybdocene readily forms soluble adducts with the amino acid cysteine and physiologically prevalent sulfur-containing compounds such as glutathione (see Figure 7.37), thiol derivatives of a molybdocene dichloride have been shown to lack antitumor activity. Researchers in the field have concluded that the biological chemistry of each of the metallocene dihalides is unique. The ongoing study of organometallic metallocenes as antitumor agents have been reported in various articles as referenced here.<sup>12</sup> The anticancer agent *cis*-platin, *cis*-dichlorodiammineplatinum(II), also obeys the 16-e rule.

Many organometallic complexes are clusters involving multiple metals that feature metal–metal bonds. The electrons in Me–Me bonds are counted by contributing one electron to each metal connected. Bridging ligands contribute one-half of their electrons to each metal center. Some simple examples in Figure 1.9 illustrate application of the rules.

Iron–sulfur clusters, such as those found in the enzyme aconitase discussed in Section 7.9.2.1, cannot be treated using the 16-e or 18-e rules. Other frameworks exist to treat large metal clusters, and these have some utility in treating  $[Fe_xS_y]^{n+}$  clusters. One method treats the number of metal atoms and the metal–metal bonds in a cluster according to the following formula<sup>13</sup>:

$$\sum \text{Valence electrons} = \# \text{ of cluster Me atoms} \times 18 - (\text{no. of metal–metal bonds}) \times 2$$

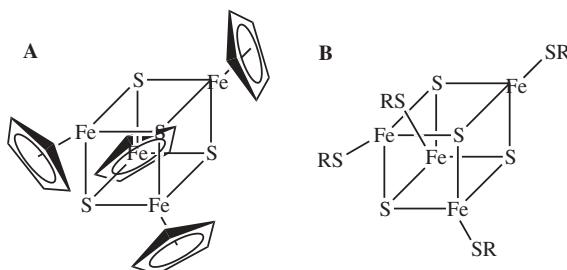
Applying this formula to the cubane  $[Fe(II)_4(\eta^5\text{-C}_5\text{H}_5)_4(\mu_3\text{-S})_4]$  shown in Figure 1.10A results in the following electron count:

$$\sum \text{Valence electrons} = 4 \times 18 - (6 \times 2) = 60 \text{ electrons}$$

the so-called “magic number” for four metal atoms in a cluster.

If one applies the same procedure to Figure 1.10B, an iron–sulfur cluster often used as a model for those in biological systems, the same magic number of 60 would be obtained. Cluster magic numbers would occur as: 48 e for a triangular clusters, 60 e for tetrahedral, 72 e for trigonal bipyramidal, 74 e for square pyramidal, 84 e for octahedral complexes like  $Zr_6I_{14}\text{C}$  or  $[\text{Mo}_6\text{Cl}_{14}]^{2-}$  or 86 e for octahedral complexes such as  $\text{Rh}_6(\text{CO})_{16}$  and  $[\text{Os}_6(\text{CO})_{18}]^{2-}$ , 90 e for trigonal prisms, and 120 e for cubic structures. Reference 13a contains a more complete discussion of cluster valence electron counting.

For biological systems such as ferredoxins, problems arise when counting electrons by the valence electron method. This system assumes six Fe–Fe bonds within the tetrahedral iron–sulfur clusters, but Fe–Fe bond distances within biological iron–sulfur clusters, as found by X-ray crystallography, often



**Figure 1.10** Cubanes (A)  $[Fe(II)_4(\eta^5\text{-C}_5\text{H}_5)_4(\mu_3\text{-S})_4]$  and (B)  $[Fe(II)_4(\text{SR})_4(\mu_3\text{-S})_4]^{4-}$ .

do not indicate Fe–Fe bonds. It is known that for the  $\text{Fe}_4\text{S}_4$  cubane found in biological systems, oxidations are accompanied by increasing distortion of the cubane frame. Nevertheless,  $^{57}\text{Fe}$  Mössbauer spectra indicate that the four iron atoms remain equivalent, suggesting delocalization within the Fe–S framework. Most biological iron–sulfur clusters deviate substantially from the electron-counting rules for iron–sulfur clusters discussed here.

In this text, iron–sulfur clusters are discussed because they appear in proteins and enzymes: (1) cytochrome b(6)f, Rieske [2Fe–2S] cluster (Section 7.5 and Figure 7.26); (2) cytochrome bc<sub>1</sub>, Rieske [2Fe–2S] cluster (Section 7.6 and Figure 7.30); and (3) aconitase, [4Fe–4S] cluster (Section 7.9.2.1, and Figure 7.50). The iron–sulfur protein (ISP) component of the cytochrome b(6)f and cytochrome bc<sub>1</sub> complexes, now called the “Rieske” ISP, was first discovered and isolated by John S. Rieske and co-workers in 1964 (in the cytochrome bc<sub>1</sub> complex). More information about the RISP is found in Section 7.5.1. Section 7.9.2 briefly discusses other proteins with iron–sulfur clusters—rubredoxins, ferredoxins, and the enzyme nitrogenase. The nitrogenase enzyme was the subject of Chapter 6 in the first edition of this text—see especially the first edition’s Section 6.3 for a discussion of iron–sulfur clusters. In this second edition, information on iron–sulfur clusters in nitrogenase is found in Section 3.6.4. See Table 3.2 and the descriptive examples discussed in Section 3.6.4.

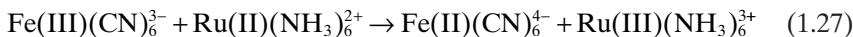
Many systems in organometallic chemistry involve the activation of hard-to-oxidize alkanes or other organic moieties. In this regard, several metalloenzymes discussed in this text are efficient alkane oxidizers. Cytochrome P450 (Section 7.4) enzymes are monooxygenases that insert one atom of a dioxygen molecule into a wide variety of organic substrates. The other oxygen atom of the dioxygen molecule is converted to water during the P450 catalytic cycle. Methane monooxygenase (Section 7.9.3.1) catalyzes the conversion of the hardest-of-all-to-oxidize alkanes, methane,  $\text{CH}_4$ , to  $\text{CH}_3\text{OH}$ . Biomimetic bioinorganic researchers model the active centers of these metalloenzymes as small molecules, to learn more about the metalloenzyme’s catalytic cycle, or to design efficient organometallic catalysts for industrial processes.

Researchers studying the metalloenzyme hydrogenase would like to design small compounds that mimic this enzyme’s ability to reversibly reduce protons to  $\text{H}_2$  and  $\text{H}_2$  to  $2\text{H}^+$ , using an active center that contains iron and nickel. Cobalamins (vitamin B<sub>12</sub> and its derivatives) contain an easily activated Co–C bond that has a number of biological functions, one of which is as a methyl transferase, 5-methyltetrahydrofolate-homocysteine methyltransferase (MTR). This enzyme converts homocysteine (an amino acid that has one more  $\text{CH}_2$  group in its alkyl side chain than cysteine; see Figure 2.2) to methionine as methylcobalamin is converted to cobalamin.

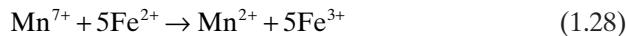
## 1.8 ELECTRON TRANSFER

Many reactions catalyzed by metalloenzymes involve electron transfer. On the simplest level, one can consider electron transfer reactions to be complemen-

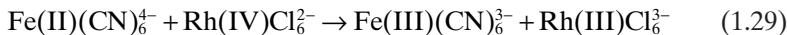
tary when there are equal numbers of oxidants and reductants and the metals transfer equal numbers of electrons as shown in equation 1.27:



Noncomplementary reactions, as shown in equation 1.28, involve unequal numbers of oxidants and reductants because the number of electrons gained or lost by each metal differs.<sup>6</sup> Noncomplementary reactions, especially for large biomolecules, must proceed through a number of bimolecular steps since the possibility of termolecular or higher-order collisions is very small.

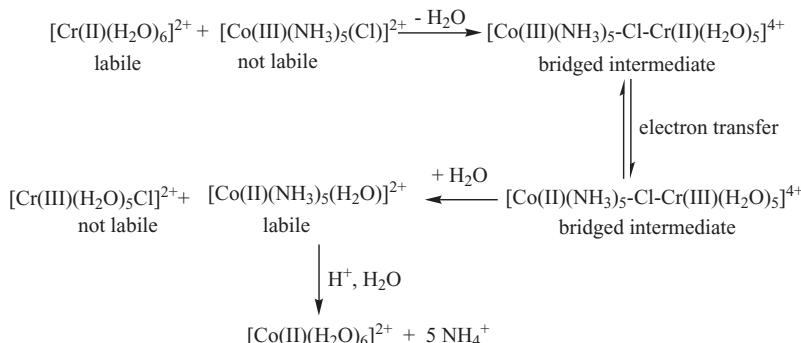


Two types of electron transfer mechanisms are defined for transition metal species. Outer-sphere electron transfer occurs when the outer, or solvent, coordination sphere of the metal centers is involved in transferring electrons. No reorganization of the inner coordination sphere of either reactant takes place during electron transfer. A reaction example is depicted in equation 1.29:

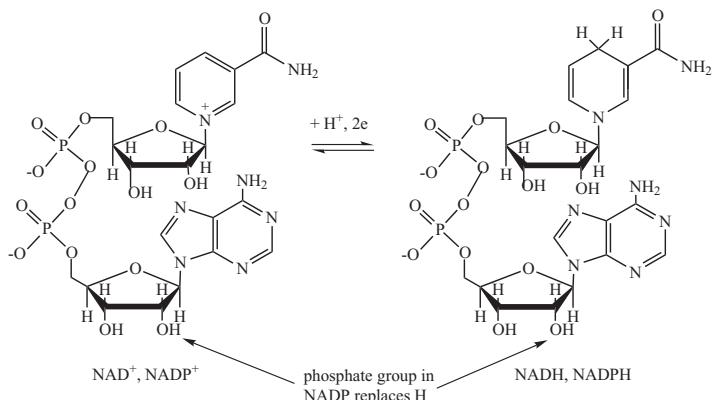


Inner-sphere electron transfers involve the inner coordination sphere of the metal complexes and usually take place through a bridging ligand. The classic example, typical of those studied and explained by H. Taube,<sup>14</sup> is illustrated by Figure 1.11's reaction sequence adapted from reference 7. In this reaction sequence, production of  $[\text{Cr(III)(H}_2\text{O})_5\text{Cl}]^{2+}$  implies that electron transfer through the bridged intermediate from Cr(II) to Co(III) and Cl<sup>-</sup> transfer from Co to Cr are mutually interdependent acts.

Harry B. Gray and Walther Ellis, writing in Chapter 6 of reference 15, describe three types of oxidation-reduction centers found in biological systems. The first of these, protein side chains, may undergo oxidation-reduction reactions such as the transformation of two cysteine residues to form the cystine dimer as shown in equation 1.30:



**Figure 1.11** An inner-sphere electron transfer reaction sequence. (Adapted from reference 7, p. 208. Copyright 1995, Wiley-VCH.)



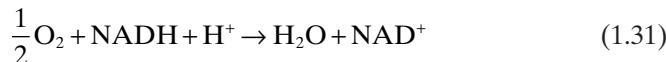
**Figure 1.12** Electron transfer cofactors NAD<sup>+</sup> or NADP<sup>+</sup>.



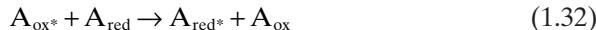
The second type of biological electron transfer involves a variety of small molecules both organic and inorganic. Examples of these are nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP) as two electron carriers and quinones and flavin mononucleotide (FMN) that may transfer one or two electrons. The structure of NAD and its reduced counterpart NADH are shown in Figure 1.12.

The third type of biological electron transfer involves metalloproteins themselves. These may be electron carriers (i.e., cytochromes) or proteins involved in the transport or activation of small molecules (i.e., cytochrome c oxidase). These so-called electron transferases have some or all of the following characteristics: (1) a suitable cofactor, such as NAD<sup>+</sup>/NADH or another protein, acting as an electron source or sink; (2) geometry that allows the cofactor close enough to the protein surface for the transfer of electrons; (3) a hydrophobic shell on the protein surface around or near the cofactor; and (4) architecture permitting changes in protein conformation to facilitate electron transfer. These last changes should be small.<sup>16</sup> Electron transferases that will be discussed in this text include the cytochromes and cytochrome c oxidase (Chapter 7). Cytochromes comprise several large classes of electron transfer metalloproteins widespread in nature. At least four cytochromes are involved in the mitochondrial electron transfer chain that reduces oxygen to water according to equation 1.31.

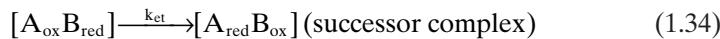
Other electron transferases include the rubredoxin and ferredoxin iron-sulfur proteins, so named because they contain iron-sulfur clusters of various sizes. Rubredoxins are found in anaerobic bacteria and contain iron ligated to four cysteine sulfurs. Ferredoxins are found in plant chloroplasts and mammalian tissue and contain spin-coupled [2Fe–2S] clusters. Further discussion of rubredoxin and ferredoxin proteins can be found in Chapters 6 and 7 of reference 15, and cytochromes will be extensively discussed in Chapter 7 of this text.



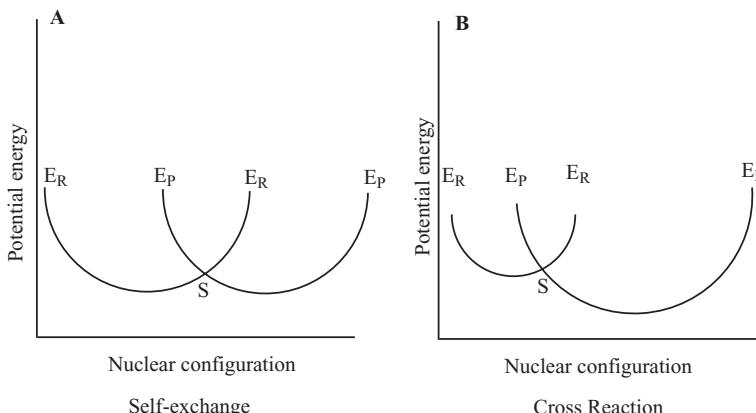
The simplest electron transfer reactions are outer sphere. The Franck–Condon principle states that during an electronic transition, electronic motion is so rapid that the metal nuclei, the metal ligands, and solvent molecules do not have time to move. In a self-exchange example,



the energies of donor and acceptor orbitals as well as bond lengths and bond angles remain the same during efficient electron transfer. Figure 1.13A illustrates this behavior using Marcus theory potential energy diagrams. In a cross reaction between two different species (as illustrated in Figure 1.13B), one can write the following set of equilibrium statements ( $K$ ) and rate equations ( $k_{\text{et}}$ ):



Electron transfer theory is further explained in a classic paper published by Rudolph A. Marcus using potential energy diagrams to describe electron transfer processes.<sup>17</sup> In the diagrams such as shown in Figure 1.13, electron donors and acceptors behave as collections of harmonic oscillators. The diagram expresses donor and acceptor in a single surface representing the precursor complex and one representing the successor complex. Point S represents the activated complex and  $E_R$  and  $E_P$  are the reactant and product surfaces, respectively.



**Figure 1.13** Potential energy diagrams describing electron transfer processes according to Marcus theory. (A) Self-exchange, (B) cross reaction.

It is beyond the scope of this text to continue the discussion of Marcus theory. Qualitatively the student should understand that electrons must find a path through the protein from the donor species to the acceptor. This may take place through bonds as outlined above or through electron tunneling events in which electrons travel through space between orbitals of the donor species to the acceptor species. Chapter 6 of reference 15 presents a clear explanation for further reading.

A four-day symposium held at the September 2006 meeting of the American Chemical Society honored Marcus' work on electron-transfer and reaction-rate theories and included talks by researchers continuing to use and update these theories in biological contexts. Recent biological system electron transfer experiments include oxygen binding and transport, photosynthesis, cellular respiration, and long-range electron transfer in proteins and DNA. One example comes from the laboratory of Alexei A. Stuchebrukhov, where a computational model is being developed to describe the proton-pumping mechanism of cytochrome c oxidase (to be discussed here in Section 7.8).<sup>18</sup> Essentially, the membrane-bound protein cytochrome c oxidase in mitochondria catalyzes the four electron reduction of O<sub>2</sub> to water using electrons and protons. The free energy generated by the reduction is used to "pump" protons to the outside of the mitochondrial membrane, generating an electrochemical gradient. The energy stored by this gradient is then used to synthesize adenosine triphosphate (ATP), the key energy transfer molecule involved in converting food into energy. The coupling of electron and proton transfer in cytochrome c oxidase is not well understood and is a subject of ongoing research in many bioinorganic laboratories.

An example of the need for extension of Marcus electron transfer theory was provided by Jacqueline K. Barton, one of Marcus' colleagues at the California Institute of Technology. Barton's group studies charge transport in DNA, attempting to elucidate fundamental mechanisms and kinetics of electron transport through DNA. A second research arm in the same laboratory studies how DNA becomes damaged in the genome and how this damage may be sensed and repaired. Experimental evidence indicates that DNA base excision repair proteins may use DNA charge transport to localize repair enzymes near the damage site. In both research areas, Barton describes the need for more theory to explain experimental results, particularly where ground-state DNA charge transport is monitored electrochemically.<sup>19</sup> Other biological applications of Marcus electron transfer and reaction rate theories described at the 2006 ACS symposium can be found in the *C&E News* article referenced here.<sup>20</sup>

## 1.9 CONCLUSIONS

The preceding brief review of inorganic chemistry has been oriented toward questions that will arise in subsequent discussion of bioinorganic systems. The

inorganic and bioinorganic chemistry texts referenced earlier in this chapter and here are good sources for answering the additional questions sure to arise in studying the behavior of metals in biological systems.<sup>21</sup> It is important to keep in mind that metal behavior in the biological milieu will be influenced greatly by the surroundings. Metal-ligand systems existing in thermodynamic equilibrium and slow to react to changing cellular or noncellular dynamics will not long endure. Therefore, most of the metalloenzyme systems to be described in following chapters contain metals in distorted and changeable ligand fields. These systems will continue to challenge the ingenuity of inorganic and bioinorganic chemists attempting to understand, modify, model, or design synthetic substitutes for them.

## REFERENCES

1. Schwarz, K. *Ged. Proc.*, 1974, **33**, 1748–1757.
2. Frieden, E. *J. Chem. Ed.*, 1985, **62**(11), 917–923.
3. Frausto Da Silva, J. R. R.; Williams, R. J. P. *The Biological Chemistry of the Elements: The Inorganic Chemistry of Life*, Clarendon Press, New York, 1991.
4. Lippard, S. J.; Berg, J. M. *Principles of Bioinorganic Chemistry*, University Science Books, Mill Valley, CA, 1994.
5. Hay, R. W. *Bio-Inorganic Chemistry*, Ellis Horwood Limited, Halsted Press, New York, 1984.
6. Cowan, J. A. *Inorganic Biochemistry, An Introduction*, 2nd ed., Wiley-VCH, New York, 1997.
7. Cotton, F. A.; Wilkinson, G.; Gaus, P. L. *Basic Inorganic Chemistry*, 3rd ed., John Wiley & Sons, New York, 1995, pp. 192–194.
8. (a) Eigen, M. *Pure Appl. Chem.* 1963, **6**, 105. (b) Bennetto, H. P.; Caldin, E. F. *J. Chem. Soc. A*, 1971, 2198.
9. Tobe, M. L. Substitution Reactions, in *Comprehensive Coordination Chemistry*, Wilkinson, G., ed., Pergamon Press, Oxford, 1987, pp. 281–329.
10. Shriver, D. F.; Atkins, P. W.; Langford, C. H. *Inorganic Chemistry*, Oxford University Press, Oxford, 1990, pp. 477–478.
11. Waern, J. B.; Harding, M. M. *Inorg. Chem.*, 2004, **43**(1), 206–213.
12. (a) Waern, J. B.; Harris, H. H.; Lai, B.; Cai, Z.; Harding, M. M.; Dillon, C. T. *J. Biol. Inorg. Chem.*, 2005, **10**(5), 443–452. (b) Waern, J. B.; Dillon, C. T.; Harding, M. M. *J. Med. Chem.*, 2005, **48**(6), 2093–2099. (c) Caruso, F.; Rossi, M. *Met. Ions Biol. Syst.*, 2004, **42**, 353–384.
13. (a) Wheeler, R. A.; Hoffmann, R. *J. Am. Chem. Soc.*, 1986, **108**, 6605–6610. (b) Mingos, D. M. P. *Acc. Chem. Res.*, 1984, **17**, 311–319. (b) Mingos, D. M. P., in *Comprehensive Organometallic Chemistry*, Wilkinson, G.; Stone, F. G. A.; Abel, E. W., eds., Pergamon, Oxford, 1981. (c) Elschenbroich, C. *Organometallics: A Concise Introduction*, VCH, New York, 1992.
14. Taube, H. *Electron Transfer Reactions of Complex Ions in Solution*, Academic Press, New York, 1970.

15. Gray, H. B.; Ellis, W. R., in Bertini, I.; Gray, H. B.; Lippard, S. J.; Valentine, J. S. *Bioinorganic Chemistry* University Science Books, Mill Valley, CA, 1994, pp. 315–363.
16. Adman, E. T. *Biochim. Biophys. Acta*, 1979, **549**, 107–144.
17. Marcus, R. A. *Annual Rev. Phys. Chem.*, 1964, **15**, 155–196.
18. (a) Quenneville, J.; Popovic, D. M.; Stuchebrukhov, A. A. *Biochim. Biophys. Acta*, 2006, **1757**(8), 1035–1046. (b) Popovic, D. M.; Stuchebrukhov, A. A. *FEBS Lett.*, 2004, **566**(1–3), 126–130. (c) Popovic, D. M.; Stuchebrukhov, A. A. *J. Am. Chem. Soc.*, 2004, **126**(6), 1858–1871.
19. (a) Yavin, E.; Stemp, E. D.; O’Shea, V. L.; David, S. S.; Barton, J. K. *Proc. Natl. Acad. Sci. USA*, 2006, **103**(10), 3610–3614. (b) Shao, F.; Augustyn, K.; Barton, J. K. *J. Am. Chem. Soc.* 2005, **127**(49), 17445–17452. (c) Yavin, E.; Boal, A. K.; Stemp, E. D.; Boon, E. M.; Livingston, A. L.; O’Shea, V. L.; David, S. S.; Barton, J. K. *Proc. Natl. Acad. Sci. USA*, 2005, **102**(10), 3546–3551. (d) Williams, T. T.; Dohno, C.; Stemp, E. D.; Barton, J. K. *J. Am. Chem. Soc.*, 2004, **126**(26), 8148–8158.
20. Ritter, S. K. *C&E News*, 2006, **84**(42), 45–48.
21. (a) Cotton, F. A.; Wilkinson, G.; Murillo, C. A.; Bochmann, M. *Advanced Inorganic Chemistry*, 6th ed., John Wiley & Sons, New York, 2006. (b) Atkins, P.; Overton, T.; Rourke, J.; Weller, M.; Armstrong, F. *Inorganic Chemistry*, 4th ed., Oxford University Press, Oxford, 2006. (c) Shriver, D.; Atkins, P. *Inorganic Chemistry*, 4th ed., W. H. Freeman, New York, 2006. (d) Fraústo da Silva, J. R. R. *The Biological Chemistry of the Elements*, Oxford University Press, Oxford, 2001. (e) Kraatz, H.-B.; Metzler-Nolte, N., eds. *Concepts and Models in Bioinorganic Chemistry*, John Wiley & Sons, New York, 2006. (f) *Biological Inorganic Chemistry: Structure and Reactivity*, University Science Books, New York, 2007.

---

# 2

---

## BIOCHEMISTRY FUNDAMENTALS

### 2.1 INTRODUCTION

Biochemistry concerns itself with the study of life at the molecular level. The subject area encompasses a huge body of information that grows and changes rapidly in the hands of thousands of capable teachers, researchers, and writers. An introduction to the biochemistry topics of proteins, protein kinetics, nucleic acids, and genetics will be presented in this chapter. In addition, a discussion of genomics and proteomics introduces topics of current relevance. The last section of the chapter discusses zinc-finger proteins, a topic that combines information introduced in this chapter on proteins, protein analysis, DNA, DNA binding, and cloning. Other important biochemical topic areas such as bioenergetics, saccharides and polysaccharides, lipids and lipoproteins, membrane compositions and dynamics, vitamins, hormones, biochemical pathways, and many others are not covered here. Students are referred to biochemistry texts<sup>1</sup> and, for the most up-to-date information, to the primary literature and to the internet. For instance, you might try the instructional “Biochemistry Online” website maintained by Dr. Henry Jakubowski at <http://employees.csbsju.edu/hjakubowski/classes/ch331/bcintro/default.html>. Wikibooks maintains a free biochemistry text at <http://en.wikibooks.org/wiki/Biochemistry>.

In introductory biochemistry, one becomes familiar with amino acids (aa) and how they combine (polymerize) to become peptides and proteins. Proteins fold into three-dimensional shapes and become enzymes, the catalysts of

biochemical reactions, or structural materials in muscles and tendons, or small molecule carriers such as the dioxygen-carrying metalloproteins myoglobin and hemoglobin. At least one-third of all proteins and enzymes contain metals and are known as metalloproteins and metalloenzymes. Some of these will be the bioinorganic chemistry topics treated in Chapters 5, 6, and 7 of this text. The recently coined term, proteomics, refers to the study of a large collection of proteins that occur and function together in a particular biochemical entity such as a cell or other organelle. Important interactions among proteins in a proteome distinguish the proteome's behavior as being more complex than that of a single protein studied individually.

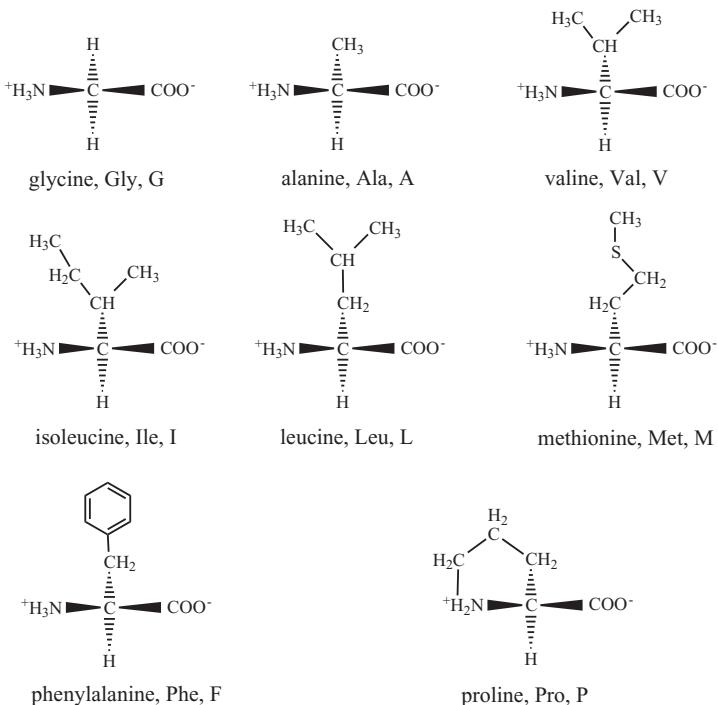
A second major area of biochemical importance concerns study of nucleotide polymerization to produce ribonucleic acids (RNA) and deoxyribonucleic acids (DNA). Genes, the basis for inherited characteristics, are contained in DNA double-helical sections incorporated into coiled and supercoiled DNA structures. Genomics, the study of the total genetic assemblage of any species, is now a well-known topic to all, especially with the announcement of the sequencing of the human genome in 2001. More information on this topic is given in Section 2.3.6.

The goal of this chapter is to provide a basic understanding of the structure of, and biochemical processes undertaken by, proteins and nucleic acids. This knowledge will be needed by students, teachers, and researchers wishing to understand and replicate the catalytic activity of metalloenzymes or to design compounds (many of them containing inorganic atoms) that will be able to detect, diagnose, and treat disease.

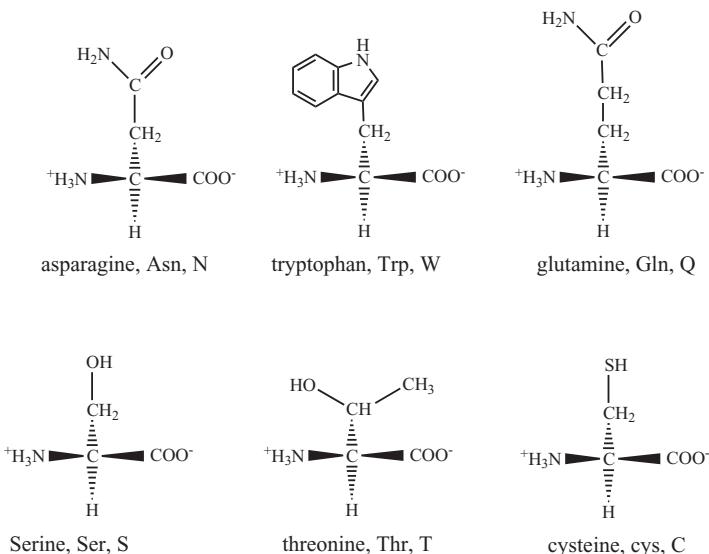
## 2.2 PROTEINS

### 2.2.1 Amino Acid Building Blocks

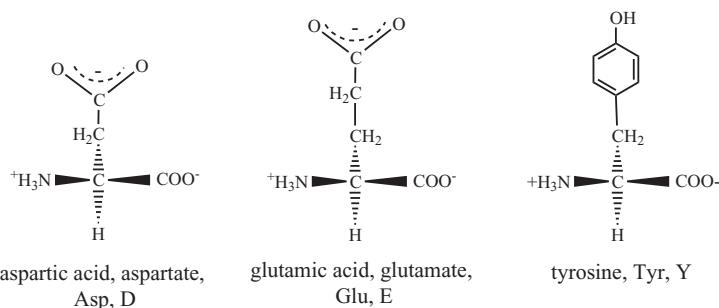
Polypeptides are formed through the polymerization of any combination of the 20 naturally occurring amino acids (aa). In humans, 10 of these amino acids are essential (cannot be synthesized by the body and must be ingested in the diet). The 10 essential amino acids (see Figure 2.1 for full names and structure) are: arg, his, ile, leu, lys, met, phe, thr, trp, and val. Relatively short amino acid chains, called polypeptides, have important hormonal (control) functions in biological species. Proteins are classified as polypeptide chains exceeding 50 amino acids in length, whereas enzyme molecules usually contain more than 100 amino acid residues. Amino acids contain a central carbon, called the  $\alpha$  carbon, to which four substituent groups are attached: (1) the amine group ( $-NH_2$ ); (2) a carboxylic acid group ( $-COOH$ ), a hydrogen atom ( $-H$ ), and a side chain ( $-R$ ) group unique to each amino acid. The structure of these basic building blocks are grouped in Figures 2.1–2.4 to indicate the neutral, polar, acidic, or basic characteristics of their R groups. The three- and one-letter common abbreviations for amino acids are shown in the figures. The



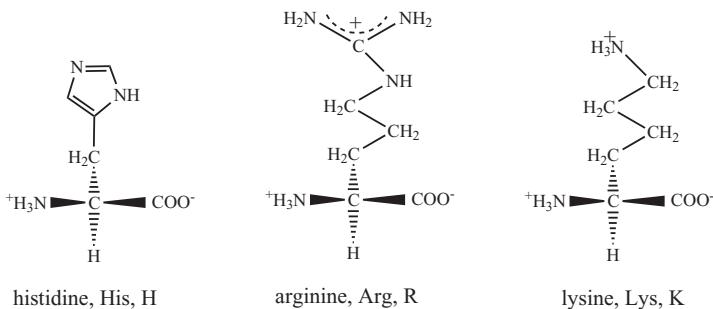
**Figure 2.1** Zwitterions of nonpolar hydrophobic amino acids at physiological pH.



**Figure 2.2** Zwitterions of polar neutral amino acids at physiological pH.



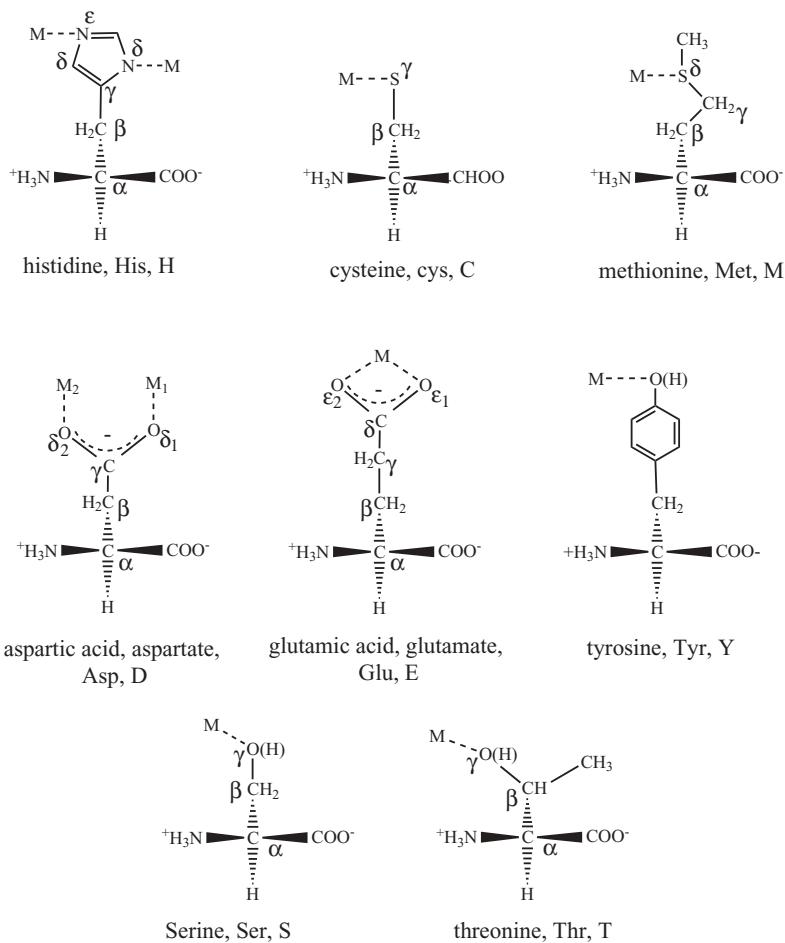
**Figure 2.3** Zwitterions of acidic amino acids at physiological pH.



**Figure 2.4** Zwitterions of basic amino acids at physiological pH.

amino acids are shown in their zwitterion form at pH 7 in which they have their COOH group ( $pK_a=2.35$ ) in the  $\text{COO}^-$  form and their NH<sub>2</sub> group ( $pK_a=9.69$ ) in the  $\text{NH}_3^+$  form.<sup>2</sup> Side-chain R groups will be protonated or deprotonated at physiological pH (~7.4) based on the pK values of the side chain. All amino acids found in proteins are called  $\alpha$ -amino acids because the amine group is bonded to the  $\alpha$  carbon. All amino acids found in proteins are L-stereoisomers with respect to the  $\alpha$  carbon (except glycine whose R group is H), although D-stereoisomers are found in bacterial cell walls and some peptide antibiotics.

In nomenclature for amino acids as ligands, the R group atoms are labeled with Greek letters starting with  $\beta$  for the first atom attached to the  $\alpha$  carbon, followed by  $\gamma$  and  $\delta$  and  $\epsilon$ . The common bioinorganic ligands from amino acid side chains have their atoms labeled in this manner in Figure 2.5. The labeling becomes important in identifying the ligand atom that may be the metal binding site in metalloproteins. In histidine, metals bind at either the  $\delta$  or  $\epsilon$  positions. In superoxide dismutase, one his ligand bridges two metal ions with Zn(II) coordinated at the  $\delta$  nitrogen and Cu(II) at the  $\epsilon$  nitrogen. For carboxylic acid side chains, metals may bind in monodentate mode (Asp-M), in bidentate chelate mode, or with the carboxylic acid oxygens bonding to two

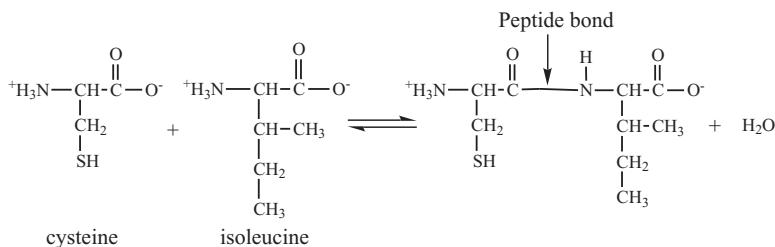


**Figure 2.5** Common metal ion bonding modes to amino acid side chains in proteins.

different metal ions ( $\text{Asp-M}_1\text{M}_2$ ). Some examples are shown in Figure 2.5, but it should be remembered that these modes are not all-inclusive. Metal ions exhibit a great variety of binding partners and bonding modes in biological species, not only to amino acid side- and main-chain atoms, but also to water, sugars, atoms of DNA and RNA, and other convenient atoms or ions in the physiological milieu.

### 2.2.2 Protein Structure

All proteins have at least three levels of structure: primary, secondary, and tertiary. Proteins with more than one polypeptide chain—hemoglobin and nitrogenase are examples—also possess quaternary structure. The primary,



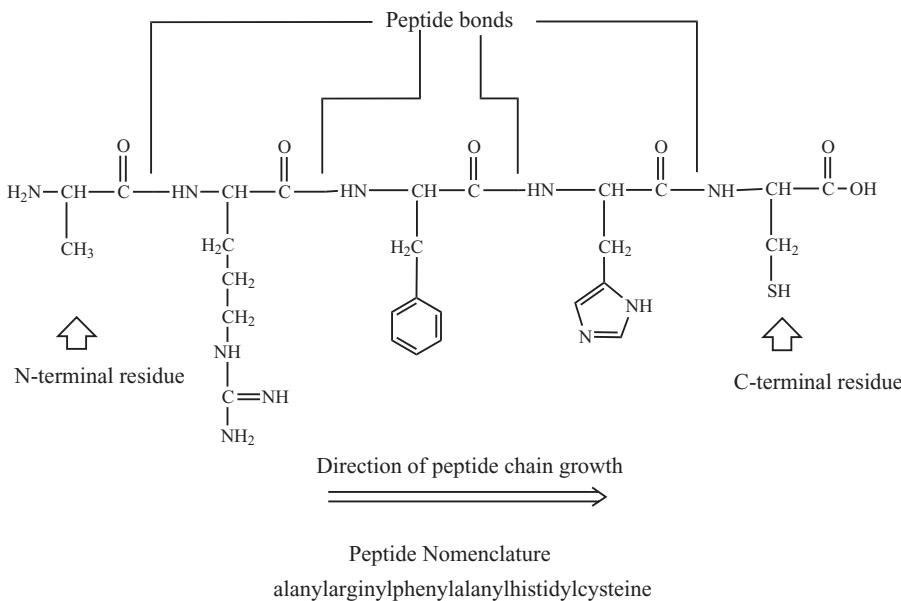
**Figure 2.6** Formation of a peptide bond.

secondary, tertiary, and quaternary structure of proteins controls their three-dimensional shape, which, in turn, affects their activity. The primary structure of proteins is formed by a condensation reaction resulting in a peptide bond as shown below in Figure 2.6. In this example, the zwitterions of cysteine and isoleucine combine to form the peptide bond via a condensation reaction between the  $\alpha$ -carboxyl group of one amino acid and the  $\alpha$ -amino group of another amino acid. A water molecule is eliminated as each successive peptide bond is formed. The peptide bond is rigidly planar and has a bond length intermediate between a single and a double bond. This rigid bond forms the backbone of the protein. Other single bonds in a polypeptide chain are flexible and can and do rotate.

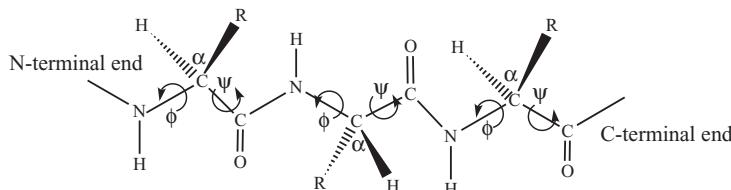
The N-terminal end of a polypeptide chain, with its free amino group, is conventionally known as the beginning of the chain, while the last amino acid, with its free carboxyl group, is the C-terminal end of the chain (see Figure 2.7).

A protein's secondary structure arises from the formation of intra-and intermolecular hydrogen bonds. All carboxyl group oxygens and amine hydrogens of a polypeptide participate in H-bonding. Protein secondary structure also derives from the fact that while all C–N bonds in peptides have some double-bond character and cannot rotate, rotation about the  $\text{C}_\alpha$ –N and  $\text{C}_\alpha$ –C bonds is possible. These rotations are affected by peptide R groups that have differing space and charge constraints. The illustration in Figure 2.8 indicates the  $\text{C}_\alpha$ –N bond angle  $\phi$  (phi) and  $\text{C}_\alpha$ –C bond angle  $\psi$  (psi) about which rotation is possible.

Two types of protein secondary structure are the  $\alpha$ -helix and  $\beta$ -pleated sheet, and their representations can be seen in Figure 2.9. The right-handed  $\alpha$ -helix occurs in globular proteins and is formed by intramolecular hydrogen bonds between the carboxyl group oxygen of one amino acid and the amine hydrogen of the fourth amino acid away from ( $i+4$  in Table 2.1). This helix completes one turn every 3.6 residues and rises approximately 5.4 Å with each turn. Other known types of protein helices are the  $\pi$ -helix and the 3–10 helix. The left-handed  $\alpha$ -helix is known but not found in protein structures. As can be seen in Table 2.1, these helical structures have differing  $\psi$  and  $\phi$  angles and different H-bonding patterns.<sup>3</sup> Many websites feature



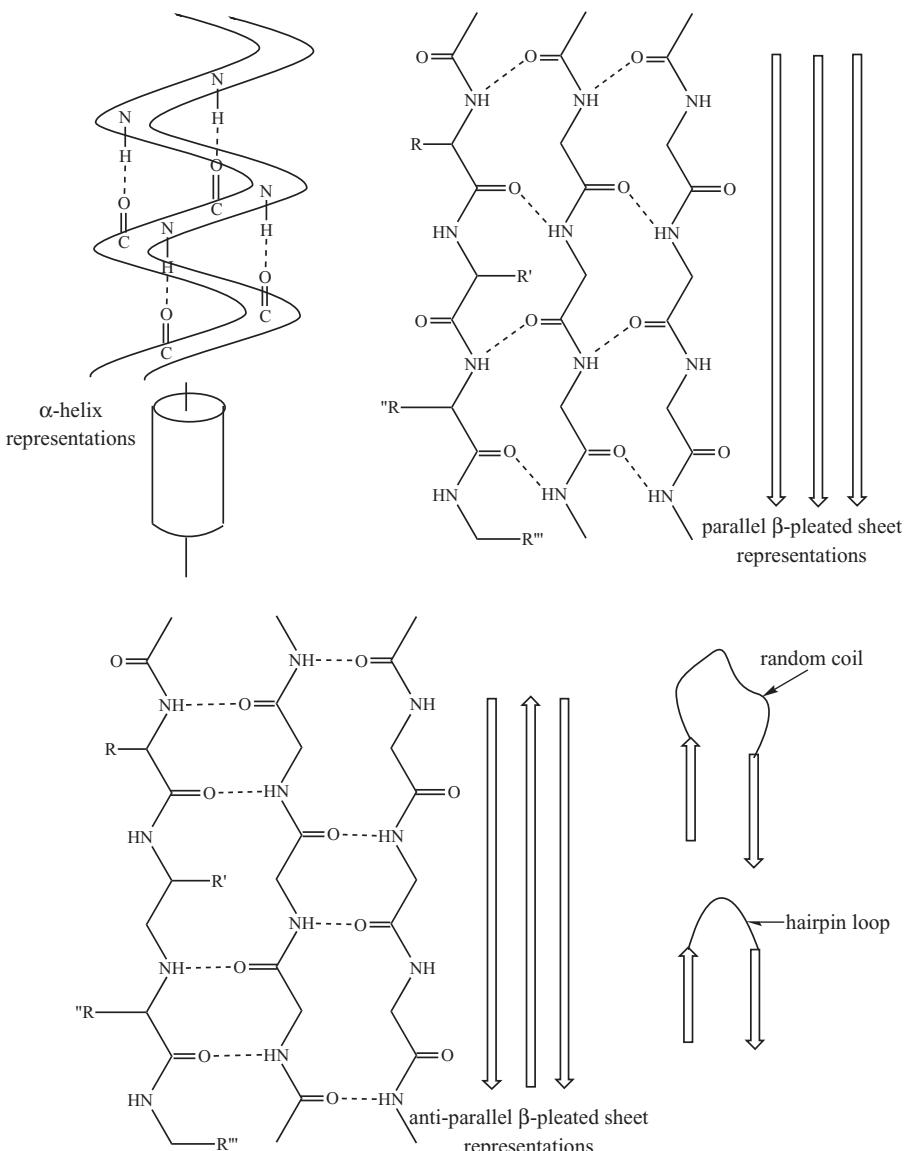
**Figure 2.7** A peptide of four amino acids.



**Figure 2.8** Illustration of  $C_\alpha$ -N bond angle  $\phi$  (phi) and  $C_\alpha$ -C bond angle  $\psi$  (psi) in a peptide.

information and graphic illustrations on proteins and peptides. Two examples are at <http://www.cem.msu.edu/~reusch/VirtTxtJml/protein2.htm#aacd7b> and [http://www.moleculesinmotion.com/protarch/page\\_sheet/menu.html](http://www.moleculesinmotion.com/protarch/page_sheet/menu.html) (requires Chime software; see Chapter 4).

The  $\beta$ -pleated sheet structure occurs in fibrous as well as globular proteins and is formed by intermolecular hydrogen bonds between a carboxyl group oxygen of one amino acid and an amine hydrogen of an adjacent polypeptide chain. Parallel  $\beta$ -pleated sheets form when the adjacent polypeptide chains are oriented in one direction (from N-terminal to C-terminal end or vice versa). Antiparallel  $\beta$ -pleated sheets form when the adjacent polypeptide chains are traveling in opposing directions (one chain N-terminal to C-terminal and the other C-terminal to N-terminal). Representations are shown in Figure 2.9. The secondary structure motifs of  $\alpha$ -helix and  $\beta$ -sheet are joined by unstructured areas called loops or coils, also shown in Figure 2.9.

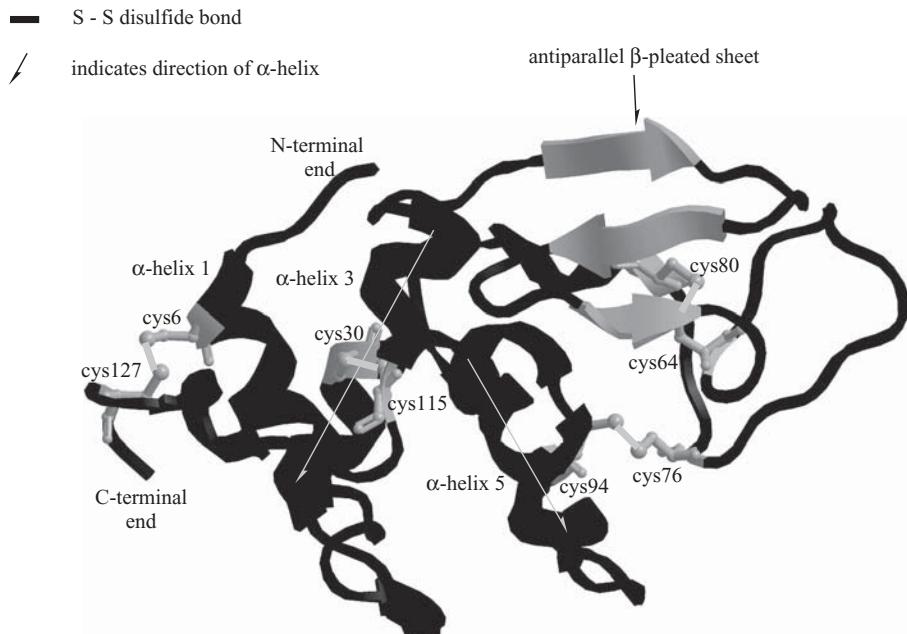


**Figure 2.9** Protein secondary structures (many R groups omitted for clarity). (Adapted with permission from Figure 1.19 of reference 15. Copyright 1997, Wiley-VCH, Inc.)

**TABLE 2.1 Characteristics of Protein Helical Structures**

Helix Type	Phi, $\phi$ (deg)	Psi, $\psi$ (deg)	H-Bond Pattern
Right handed α-helix	-57	-47	i+4
π-helix	-57	-70	i+5
3–10 helix	-49	-26	i+3

Lysozyme visualized from PDB: 2C8O



**Figure 2.10** Secondary and tertiary structure of the enzyme lysozyme, PDB: 2C8O. Visualized using Cambridge Soft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. ChemDraw Ultra, version 10.0. (Printed with permission of CambridgeSoft Corporation)

The tertiary structure of proteins arises from the interactions of the various R groups along the polypeptide chain. Some of the forces responsible for the tertiary structures include van der Waals forces, ionic bonds, hydrophobic bonds, and hydrogen bonds. Usually, hydrophilic R groups arrange themselves on the exterior of the tertiary structure and interact with the aqueous environment, whereas hydrophobic R groups usually orient themselves on the interior of the protein's tertiary structure and exclude water. Most proteins are defined as being amphipathic—that is, they contain both nonpolar and polar side chains and environments. The nonpolar groups or areas will align themselves with nonpolar or hydrophobic molecules, while the polar groups or areas will align themselves with polar or hydrophilic molecules.

One ionic bond that often helps establish tertiary structure is a disulfide bond between two cysteine side chain groups—for instance, in the enzyme lysozyme as shown in Figure 2.10. Lysozyme is not a metalloprotein, such as will be studied in this text, but it is a small enzyme and is illustrative of some secondary and tertiary structures found in the more complex molecules described in the following chapters. Lysozyme protects biological species from

bacterial infection by breaking the tough skin of carbohydrate chains, interlocked by short peptide strands, that form the protective cell walls of bacteria. The lysozyme action destroys the structural integrity of the bacterial cell, causing it to burst under its own internal pressure. The particular lysozyme modeled here is a synthetic construct that features a primary sequence of 129 amino acid residues. The X-ray crystallographic data visualized in Figure 2.10 was taken from the protein data bank (PDB) at <http://www.rcsb.org/pdb/home/home.do>. Searching the term “lysozyme” in this website yielded 991 deposited lysozyme structures in late 2006. The lysozyme structure chosen for visualization, with PDB accession number 2C8O, dates from 2006.<sup>4</sup> A possible confusion in entering letters versus numbers when using the PDB should be explained. The lysozyme PDB: 2C8O data are accessed with the last digit given as the letter “O”. If one enters the number 0 (zero) instead—2C80—a different data set for a different molecule is accessed. The visualization for PDB: 2C8O was carried out using CambridgeSoft’s Chem3D Ultra Version 10.0 software, saved as a .png file in Chem3D, then opened with ChemDraw Ultra Version 10.0 for additional notations. (See Chapter 4 for more information on CambridgeSoft software.)

Figure 2.10 visualizes the compact, globular, 129-amino-acid residue lysozyme protein that exhibits  $\alpha$ -helical,  $\beta$ -pleated sheet, random coil and connecting loop secondary structure. Seven  $\alpha$ -helices, some of them distorted and some short, can be identified: (1) residue arg5–arg14, contains cys6; (2) tyr20–gly22; (3) leu25–ser36, contains cys30; (4) cys80–leu84, contains cys80; (5) thr89–val99, contains cys94; (6) gly104–arg114; and (7) val120–ile124. Three antiparallel  $\beta$ -pleated sheet segments are also identified. The protein is knitted together into its tertiary, globular structure by four disulfide bonds between the eight cysteine residues found in lysozyme. The bonds are formed between (1) cys6 (in helix 1) near the protein’s N-terminal end and cys127 (in random coil) near the C-terminal end; (2) cys30 (in helix 3) and cys115 (in random coil connecting distorted helices 6 and 7); (3) cys76 (in a long loop connecting the final segment of the  $\beta$ -pleated sheet with helix 4) and cys94 (in helix 5); and (4) cys64 (also found in the long loop connecting the final segment of the  $\beta$ -pleated sheet with helix 4) and cys80 (the first residue in helix 4).<sup>5</sup>

Most proteins contain more than one polypeptide chain. The manner in which these chains associate determines quaternary structure. Binding involves the same types of noncovalent forces mentioned for tertiary structure: van der Waals forces, hydrophobic and hydrophilic attractions, and hydrogen bonding. However, the interactions are now interchain rather than intrachain (tertiary structure determination). The quaternary structure of hemoglobin (four almost identical subunits) will be discussed in Section 7.2.

Secondary structure, as well as tertiary and quaternary structure of proteins, is intimately dependent on the primary sequence of amino acids in the chain. In fact, the manner in which proteins fold into their ultimate structures in biological species is a subject of much research and continued uncertainty even

though much is known. Predicting protein folding is an enormous challenge. Most proteins contain dozens or hundreds of amino acids, so there is an astronomical number of ways in which these might be arranged into a compact, folded structure. While only a tiny fraction of the possible folds—perhaps 1000 to 10,000—are found in natural proteins, the challenge is to deduce the best fit of a particular protein sequence to one of these folds. This is called the protein-threading problem. Traditionally, the problem is tackled by assuming that each amino acid prefers to be surrounded by others of a specific kind, and then to look for the best compromise between the needs of all the amino acids. Success using this approach depends on how well we know what the amino acids prefer. Instead of trying to deduce this from physical and chemical principles, Jayanth Banavar of Pennsylvania State University and colleagues use a set of known protein structures to train a computer program to recognize the preferences of each amino acid. Once trained, the program, a neural network, can then predict unknown structures. These researchers have shown that the learning-based method is more successful than one based on *a priori* assumptions about amino acid preferences. The neural network correctly predicted the structures of 190 out of 213 test proteins, whereas the conventional approach identified only 137 structures correctly.<sup>5</sup> Some websites for more information include: <http://folding.stanford.edu/science.html> and <http://predictioncenter.org/> (the Protein Structure Prediction Center at the Genome Center at UCLA). Some recent review articles include those of reference 6.

The amino acid composition of a protein can be determined by cleaving all peptide bonds and identifying the constituent amino acids. The sequence of a given protein is determined by using various methods to cleave only selected peptide bonds and then assembling the information to deduce the amino acid sequence. Proteins with multiple subunits are usually broken down into individual subunits (denatured with heat or chemical reagents) before composition and sequencing analyses are carried out. Characterization of proteins on the basis of size and/or charge can be accomplished by a number of methods. These methods are described Section 2.2.3.

### 2.2.3 Protein Sequencing and Proteomics

Protein chemists have followed a basic strategy in sequencing proteins, as will be described in the following section. Denaturation of multi-subunit proteins through heat, changes in pH, or chemical reagents (urea, organic solvents, acids, or bases) first produces the constituent subunits. Disulfide bonds are broken by selective reduction or oxidation. Amino acid composition is determined by breaking the peptide bond through exhaustive enzymatic degradation or by hydrolysis by strong acid (6N HCl) or strong bases. Separation by chromatography is followed by identification and quantification of the individual amino acids by producing colored or fluorescent products, the measured intensity of which are proportional to the concentration of the amino acids.

Initial N-terminal and C-terminal sequence determination may be made using end-group analysis. Addition of the reagent fluorodinitrobenzene (FDNB), for instance, reacts with free amine to form a dinitrophenyl derivative used to identify the N-terminal amino acid. Carboxypeptidase will remove the C-terminal amino acid. Closed-end (circular) peptides or those with modified N-terminal or C-terminal amino or carboxyl groups will not be detected. The Edman degradation is a method for removing the N-terminal amino acid. The process may be repeated successively to sequence an entire peptide. Polypeptides may be broken down into fragments by cleaving at specific acids. Trypsin, for instance, will cleave residues with positively charged R groups (lys and arg) on their carboxyl side, and chymotrypsin will cleave residues with aromatic (phe, tyr, trp) or bulky aliphatic (ile, val) R groups. CNBr cleaves at the carboxyl side of the methionine peptide bond. End-group analysis, Edman degradations, and enzyme treatments may be repeated on different protein fragments to determine sub-sequences and create “overlaps.” Large peptide fragments are positioned relative to one another after a second or third treatment creates other fragments whose sequences extend across the initial cleavage points. Eventually the original polypeptide sequence is reconstructed.

Many other methods are used in the analysis of protein mixtures or purified proteins. Ultracentrifugation separates proteins according to size and can determine their molecular weight. Gel filtration (or size exclusion) chromatography separates proteins according to size. Elution time of the protein molecules through a column is related to molecular size, with the largest proteins eluting first and the smallest last. Ion exchange chromatography separates proteins based on their charge, whereas affinity chromatography separates proteins that bind specifically with certain chemical groups. Electrophoresis separates different proteins based on their net charge and based on mass if they have the same charge. Isoelectric focusing separates proteins based on their isoelectric points—the pH at which the individual amino acids exist as zwitterions. More information on these methods can be found in reference 7. The three-dimensional structure of a protein—its secondary, tertiary, and quaternary structure—has been determined for hundreds of proteins (and thousands of protein variations) using X-ray crystallography in the single-crystal solid state and nuclear magnetic resonance (NMR) spectroscopy in solution. These techniques are discussed in detail in Chapter 3. The amount and type of secondary structure in a protein may be determined by circular dichroism (CD) and magnetic circular dichroism (MCD) spectroscopy. A detailed description of CD and MCD spectroscopy applied to bioinorganic molecules is found in Chapter 5 of reference 8. Mass spectrometry (MS) can determine the structure of very small quantities of protein and has become a very popular analytical technique in proteomics, as discussed below.

The procedures above describe treatment of an individual protein or mixtures of protein subunits obtained by breaking down quaternary structures of individual proteins. More recently, researchers have been interested in numbers

of proteins having separate and distinct functions but that are grouped together in a cellular matrix. Proteomics, a term coined around 1995, attempts to study the interactions of the proteins as a group, focusing on their combined molecular behavior. Proteomics can be defined as the analysis of all the proteins expressed at a given time or in a given location rather than the study of one isolated protein as has been the scientific practice until recently. In describing the behavior of numbers of proteins functioning together in an organelle, the proteome becomes defined as the protein complement of the organism's genome. One can think of the human proteome as the universe of proteins encoded by gene sequences in the human genome. The genome, discussed further in the Section 2.3.3, is defined as all of the genes contained in a given organism.

Scientists believe that the molecular function of an isolated protein may be very different from the function of that protein in its complex cellular environment—that is, in the proteome. It is also thought that an understanding not only of the structure and function of the individual proteins, but also of their interactions with other proteins and other molecules, may lead to better disease diagnosis and treatment. Proteomics is a more complex topic than genomics because organisms may have well over an order of magnitude more proteins than genes. Because proteomics studies large, complex groups of proteins together at the same time, new techniques and technologies are being developed to assist in this effort. A few of these techniques and technologies are discussed briefly here. A good basic review of proteomics techniques and technologies can be found in the Chembytes ezine article written by Michael Dunn at the website <http://www.chemsoc.org/chembytes/ezine/1998/dunn.htm>.

The first requirement for proteome analysis, also known as protein profiling, is separating the complex protein mixture into its various protein components. Two-dimensional polyacrylamide gel electrophoresis (2-DE or 2-D gels) is still a common method used for separating sample proteins according to their different properties. The first dimension separation, called isoelectric focusing (IEF), separates proteins according to charge as the different proteins are focused at their respective isoelectric points ( $pI$ , the pH at which the protein's net charge equals zero). The second dimension separation, by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), separates the proteins by size—molecular weight,  $M_r$ . The orthogonal combination of the two separations carried out at right angles results in the proteins being distributed across the 2-D gel profile. Automated computer analysis algorithms, available commercially, are needed at this point for rigorous qualitative and quantitative analysis of the complex and partially overlapping patterns of proteins visualized by the 2-D separation process.

Following the separation process, the individual proteins must be identified and characterized. The 2-D separation has provided  $M_r$ ,  $pI$ , and relative abundance data but no information on protein identities or functions. Computer algorithms are available for matching the  $M_r$ ,  $pI$ , and relative abundance data

to known proteins in mixtures, but these are unlikely to provide unequivocal protein identification. One method used to identify individual proteins has been to excise a spot containing one protein from the 2-D gel and then digest the protein with a proteolytic enzyme to form peptide fragments which are then analyzed by mass spectrometry (MS). Peptide mass profiles obtained by MS analysis of given proteins can yield unique profiles leading to protein identification when compared directly to known protein MS peptide mass profile databases. One may also use Edman degradations—the same chemical method described above for successively releasing individual N-terminal residues of a protein—followed by MS of the degradation products to obtain peptide mass profiles of the protein. These methods are slow if one wishes to identify the hundreds of different proteins that may be separated on the 2-D gel.

The proteomics mass spectrometric and other analytical techniques must be capable of high-throughput sensitive screening of proteins separated on 2-D gels so that only those proteins that cannot be identified unequivocally or appear to be novel need further characterization by protein sequencing. High-pressure liquid chromatographic (HPLC) analyses assist here as some proteins have unique amino acid compositions that can be used for their identification when compared to databases of protein amino acid compositions. Gel electrophoresis, MS, and HPLC techniques are continually being refined and extended for more efficient proteome analyses. One final requirement for proteomic technology is that the generated data must be stored in databases that can be interrogated effectively in the laboratory and made available to other scientists worldwide through the internet or other mechanisms. Lists of known 2-D PAGE (two-dimensional polyacrylamide gel electrophoresis) protein databases are maintained at <http://www.expasy.ch/ch2d/2d-index.html>. The Swiss Proteomics Society maintains a page with links to publications and information in the proteomics area at <http://www.swissproteomicsociety.org/links.html>.

The proteomics research of a number of scientists was described in a *C&E News* report of the 2001 Pittcon meeting.<sup>9</sup> One group, that of Catherine Fenselau at the University of Maryland, studied a new method for proteolytic stable isotope labeling to provide quantitative and concurrent comparisons between individual proteins from two entire proteome pools.<sup>10</sup> Two serotypes of an adenovirus were selected for study. Two O<sup>18</sup> atoms were incorporated into the carboxyl termini of tryptic peptides (peptides digested using trypsin) during the proteolytic cleavage of all proteins in the first pool. Proteins in the second pool were also cleaved using trypsin digestion with the carboxyl termini of the resulting peptides containing two O<sup>16</sup> atoms (no labeling). The two peptide mixtures (<sup>18</sup>O mixture plus <sup>16</sup>O mixture) were pooled for fractionation and separation by HPLC, and the masses and isotope ratios of each peptide pair (differing by 4 Da) were measured by high-resolution mass spectrometry. Short sequences and/or accurate mass measurements combined with proteomics software tools allowed the peptides to be related to the precursor

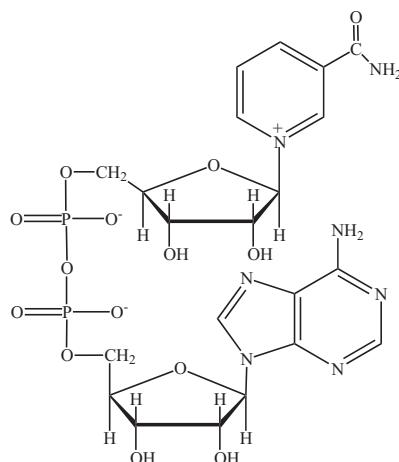
proteins from which they were derived. The authors stated that proteolytic O<sup>18</sup> labeling enabled a shotgun approach for proteomic studies with capability for quantification of the amount of a given protein in a complex mixture. The authors proposed this method as a useful tool for comparative proteomic studies of very complex protein mixtures. In other proteomic experiments, the Fenselau group studied *Bacilli* mixtures [*Bacillus cereus* strain T, *Bacillus globigii* strain 9372, wild-type *Bacillus subtilis* strain EMG 168, and *Bacillus thuringiensis* subs. Kurstakistrain HD-1 (ATCC 33679)],<sup>11a</sup> developed a method for quantitative determination of heme for forensic characterization of *Bacillus* spores,<sup>11b</sup> and developed another method for the characterization of enterobacteria in mixtures [*E. coli* ATCC 25404 (strain K12), *Enterobacter cloacae*, *Erwinia herbicola* (also known as *Pantoea agglomerans*), and *Salmonella typhimurium* (strain 98-9176)],<sup>11c</sup> all using the mass spectrometric method called MALDI-TOFS (matrix-assisted laser desorption/ionization time-of-flight mass spectrometry). This method is further described in Section 3.7.3.

Research in the Fenselau laboratory uses proteomics to study groups of cellular proteins that may be responsible for acquired drug resistance. Acquired drug resistance (ADR), causes chemotherapeutic agents to lose effectiveness over time. In one study, Fenselau and coworkers examined the abundance changes between proteins in (1) a drug susceptible MCF-7 breast cancer cell line and (2) a cell line (from the same parent line) exhibiting resistance to the anticancer drug mitoxantrone. The study identified 15 proteins with significant changes—including receptors, adhesion proteins, proteins involved in amino acid uptake or in glucose uptake.<sup>12a</sup> Alterations in the mitochondrial proteome of adriamycin resistant MCF-7 breast cancer cells were found in a 2D-gel based comparative proteomic analysis conducted in 2006 by the same group.<sup>12b</sup>

In addition to the stable isotope labeling (<sup>18</sup>O versus <sup>16</sup>O) of proteins for quantifiable proteomic analyses as described above, chemical approaches to the protein-labeling problem have developed in great variety. These so-called “affinity tags” can be used to label specific side chain groups such as sulfhydryl or amino groups, active sites for serine and cysteine hydrolases and many others. This active research area has been reviewed recently by A. Leitner and W. Lindner in a *Proteomics* article entitled: “Chemistry meets proteomics: The use of chemical tagging reactions for MS-based proteomics.”<sup>13</sup>

#### 2.2.4 Protein Function, Enzymes, and Enzyme Kinetics

Protein functions and interactions are infinitely varied in biological species—one of the major problems associated with complete classification of any proteome. Proteins may transport substances: myoglobin and hemoglobin (discussed in Chapter 7) transport oxygen, and carbon dioxide, in mammalian blood. Proteins called enzymes catalyze necessary biochemical reactions. The active site of an enzyme contains those amino acids that come in direct contact



**Figure 2.11** Cofactor nicotinamide adenine dinucleotide ( $\text{NAD}^+$ ).

with the substrate (substance acted upon) and bind it. Some enzymes require nonprotein molecules or groups, called *cofactors*, for their activity. Cofactors may be metal ions such as  $\text{Fe}^{2+,3+}$ ,  $\text{Cu}^{1+,2+}$ ,  $\text{Ni}^{1+,2+,3+}$ , or  $\text{Co}^{1+,2+,3+}$ , and so on. Cofactors may also be organic molecules such as nicotinamide adenine dinucleotide ( $\text{NAD}^+$ ), shown in Figure 2.11. These organic molecules are also called *coenzymes*. An enzyme without its cofactor, called an *apoenzyme*, usually will be inactive in its catalytic role or exhibit greatly reduced catalytic activity. An active enzyme with its cofactor is called a *holoenzyme*. Enzymes that require metal ions as cofactors are called *metalloenzymes*. Examples of metalloenzymes to be discussed in later chapters include metal cations from periodic chart Groups 1 and 2, which are extremely important in biological systems. These metal ions—sodium and potassium in  $\text{Na}^+/\text{K}^+$ -ATPases; magnesium in kinases, ATP pumps, and chlorophyll; calcium in P-type ATPases; and enzymes like troponin and calmodulin—will be discussed in Chapters 5 and 6. Enzymes that contain  $\text{Fe}^{2+/3+}$  and  $\text{Cu}^{2+/3+}$  ions—cytochromes, cytochrome c oxidase, aconitase—will be discussed in Chapter 7. Essential cofactors and coenzymes that must be ingested in the diet include water-soluble vitamins such as the cobalt-containing vitamin B<sub>12</sub>. *Isozymes* may be structurally different forms of the same enzyme or oligomeric proteins with differing sets of subunits. Isozymes may also be defined as multiple forms of an enzyme whose synthesis is controlled by more than one gene. Usually, isozymes catalyze similar types of reactions. *Allosteric* enzymes have their activity modulated by the binding of a second molecule, whereas nonallosteric enzymes do not. Proteins maintain biological structures; that is, the protein collagen constitutes part of fibrous connective tissues in skin, bone, tendon, cartilage, blood vessels, and teeth. Proteins may facilitate movement; that is, the proteins actin and myosin mediate muscle contraction. Proteins serve as storage compounds; that is, the

protein ferritin stores iron in the human body. *Immunoglobins* are proteins that react with and neutralize foreign compounds (antigens) in the body. The protein hormone insulin regulates glucose levels in the blood.

Many enzymes are named in classes by attaching “ase” to the type of reaction they catalyze. ATPase enzymes catalyze the hydrolysis of adenosine triphosphate (ATP) to adenosine diphosphate (ADP). Oxidoreductases catalyze oxidation-reduction reactions: Alcohol dehydrogenase catalyzes the oxidation of an alcohol to an aldehyde. Transferases catalyze the transfer of groups from one molecule to another. Hydrolases catalyze the breaking of covalent bonds using water; for example, peptidase hydrolyzes a peptide bond. Lyases either remove a group by splitting a bond and forming a double bond or add a group to a double bond to form a single bond; for example, decarboxylase removes a carboxyl group to form carbon dioxide. Isomerases catalyze internal atom rearrangements in a molecule. Ligases catalyze the formation of covalent bonds.

The general theory of enzyme kinetics is based on work of L. Michaelis and M. L. Menten, later extended by G. E. Briggs and J. B. S. Haldane.<sup>1a</sup> The basic reactions ( $E$ =enzyme,  $S$ =substrate,  $P$ =product) are shown in equation 2.1:



Assuming that the reactions are reversible and that a one-substrate enzyme-catalyzed reaction is being studied, one can derive the Michaelis-Menten rate:

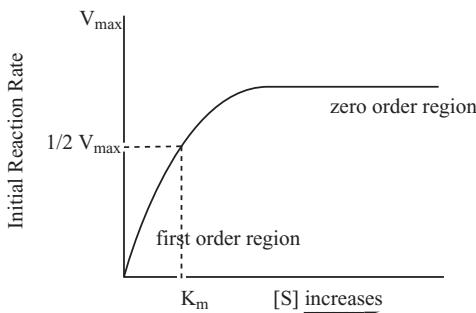
$$V = \frac{V_{\max}[S]}{[S] + K_m} \quad (2.2)$$

where  $V$  is the initial rate for first-order breakdown of the enzyme–substrate ( $[ES]$ ) complex into enzyme ( $E$ ) and product ( $P$ );  $V_{\max}$  is the maximum reaction rate for a given concentration of enzyme in the presence of saturating levels of substrate;  $[S]$  is the substrate concentration; and  $K_m$  is the Michaelis constant, the concentration of substrate required to achieve one-half the enzyme’s maximal velocity. Equation 2.2 applies to single-substrate reactions at a constant enzyme concentration. Only nonallosteric enzymes—those not dependent on binding of a second molecule in addition to the substrate—are treated by the Michaelis-Menten rate equation. When the substrate concentration is low, the rate of reaction ( $V$ ) increases in direct proportion to the substrate concentration (first-order kinetics). When the substrate concentration becomes high enough to saturate all available enzyme active sites, the reaction rate becomes constant (zero-order kinetics). The graphical representation of this behavior is shown in Figure 2.12.

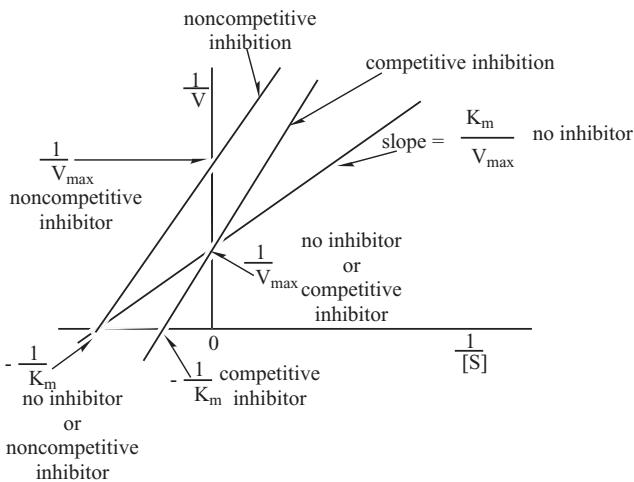
The Lineweaver-Burk plot uses the reciprocal of the Michaelis-Menten equation in the form of the equation of a straight line,  $y=mx+b$ , having the form shown in equation 2.3:

$$\frac{1}{V} = \left( \frac{K_m}{V_{\max}} \right) \frac{1}{[S]} + \frac{1}{V_{\max}} \quad (2.3)$$

Plotting  $1/V$  vs.  $1/[S]$ , one obtains a straight line having a slope of  $K_m/V_{\max}$  with a  $y$ -axis intercept of  $1/V_{\max}$  and an  $x$ -intercept of  $-1/K_m$  as shown in Figure 2.13. Lineweaver–Burk plots of enzyme activity in the presence of an inhibitor can distinguish the type of inhibitor. Competitive inhibitors have a molecular structure similar to that of the substrate and will alter  $K_m$  but not  $V_{\max}$  because they compete with the substrate for binding at the enzyme's active site but do not change the enzyme's affinity for substrate. Noncompetitive inhibitors bear no structural similarity to the substrate but bind the free enzyme or enzyme–substrate complex at a site other than the active site. They reduce the enzyme's



**Figure 2.12** Graphical representation of the Michaelis–Menten equation for nonallosteric enzymes.



**Figure 2.13** Lineweaver–Burk plot for no inhibitor, competitive inhibition, and noncompetitive inhibition.

affinity for substrate and decrease  $V_{\max}$  while not changing  $K_m$ . The lines indicating competitive and noncompetitive inhibition are included in Figure 2.13.

## 2.3 NUCLEIC ACIDS

### 2.3.1 DNA and RNA Building Blocks

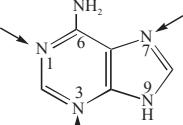
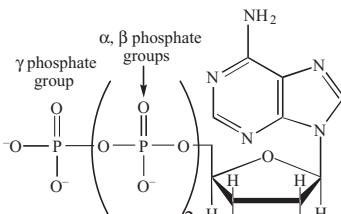
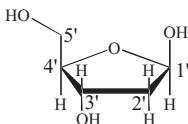
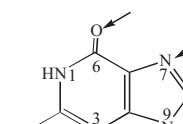
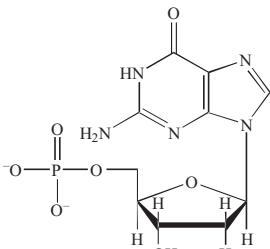
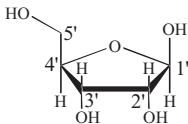
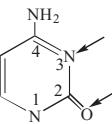
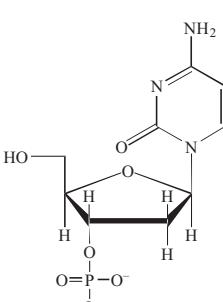
The sequence of amino acids in a given protein determines its structure. The amino acid sequence is controlled ultimately by sequences of genes that are made up of deoxyribonucleic acid (DNA). DNA is composed of repeating units of the nitrogenous bases adenine and guanine (purines) and thymine and cytosine (pyrimidines) linked to the five-carbon cyclic sugar  $\beta$ -D-2-deoxyribose. In turn, the sugars are connected by 3'-5' phosphodiester bonds forming the linkages in the DNA polymeric chain. In nucleic acid numbering, positions on the sugar ring carry a prime (' $\prime$ ') to distinguish them from the numbering of nitrogenous base positions. DNA chains form the well-recognized double helix through hydrogen bonding of adenine to thymine (A–T pairs) and guanine to cytosine (G–C pairs). More about the forms of the double helix is found in the next section. The complete nitrogenous base, sugar, and phosphate unit is called a *nucleotide*. Nitrogenous base plus sugar moiety are called *nucleosides*. Ribonucleic acids (RNA) resemble DNA in that nucleoside monophosphates are joined through phosphodiester bonds. RNAs differ in that the sugars are  $\beta$ -D-ribose units and the pyrimidine uracil is found in place of thymine. Molecular structures and nomenclature for nitrogenous bases, nucleosides, and nucleotides are delineated in Table 2.2.

Common metal binding sites on nucleobases are indicated by arrows in Table 2.2. Many of these are endocyclic (in the ring) nitrogen atoms that have more available lone-pair characteristics than exocyclic (exterior to the ring) amino groups whose lone pairs are delocalized into the ring by resonance. Experimental and theoretical studies indicate that the purine N7 sites are the best nucleophiles among all the possible metal binding sites.<sup>2</sup> They are exposed to solvent in the B DNA major groove and are not involved in Watson–Crick base pairing. The N7 sites of adenine and especially guanine are the preferred platinum ion coordination location for the platinum-containing anticancer drugs. Hard metal ions such as  $\text{Na}^+$ ,  $\text{K}^+$ , and  $\text{Mg}^{2+}$  prefer the hard negatively charged oxygen atoms of the phosphodiester groups, with the result that enzymes requiring ATP often require  $\text{Mg}^{2+}$  as well.<sup>2</sup> Various metal chelates of the  $\alpha$ ,  $\beta$ , and  $\gamma$  phosphate groups have been identified by X-ray crystallography and  $^{31}\text{P}$  NMR.

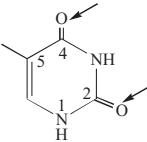
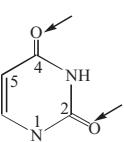
### 2.3.2 DNA and RNA Molecular Structures

The nitrogenous base pairs in DNA (A–T and C–G) link through the formation of hydrogen bonds within the double-stranded (ds) structure for DNA.

**TABLE 2.2** Nitrogenous Bases, Nucleosides, Nucleotides and Sugars Found in DNA and RNA

Base <sup>a</sup>	Nucleoside	Nucleotide Example	Sugar
 Adenine, A	Adenosine (RNA) or deoxyadenosine (DNA)	Adenosine triphosphate (ATP) 	$\beta$ -D-2- Deoxyribose <sup>b</sup> in DNA or $\beta$ - D-Ribose <sup>c</sup> in RNA 
 Guanine, G	Guanosine (RNA) or deoxyguanosine (DNA)	Deoxyguanosine 5'- monophosphate (5'-dGMP) 	$\beta$ -D-2- Deoxyribose <sup>b</sup> in DNA or $\beta$ - D-Ribose <sup>c</sup> in RNA 
 Cytosine, C	Cytidine (RNA) or deoxycytidine (DNA)	Deoxycytidine 3'- monophosphate (3'-dCMP) 	$\beta$ -D-2- Deoxyribose <sup>b</sup> in DNA or $\beta$ - D-Ribose <sup>c</sup> in RNA

**TABLE 2.2** *Continued*

Base <sup>a</sup>	Nucleoside	Nucleotide Example	Sugar
 Thymine, T 5-methyl uracil	Deoxythymidine (DNA)	Deoxythymidine 5'-diphosphate (5'-dTDP)	$\beta$ -D-2'-Deoxyribose <sup>b</sup> in DNA
 Uracil, U	Uridine	Uridine 3'-monophosphate (3'-UMP)	$\beta$ -D-Ribose <sup>c</sup> in RNA

<sup>a</sup>Arrows indicate common metal-binding sites

<sup>b</sup>H at 2' position

<sup>c</sup>OH at 2' position

One type of hydrogen bonding is shown in Figure 2.14. Watson–Crick pairing is shown; however, other hydrogen bonding conformations are found. Some of these are known as Hoogsteen pairs, reversed Hoogsteen pairs, and reversed Watson–Crick pairs. Note that in Watson–Crick base pairing A–T pairs form two hydrogen bonds while C–G pairs form three. RNA usually exists in the single-stranded (ss) form but may fold into secondary and tertiary structures through the formation of specific base pairings (A–U and C–G). Double-stranded RNA is also known.

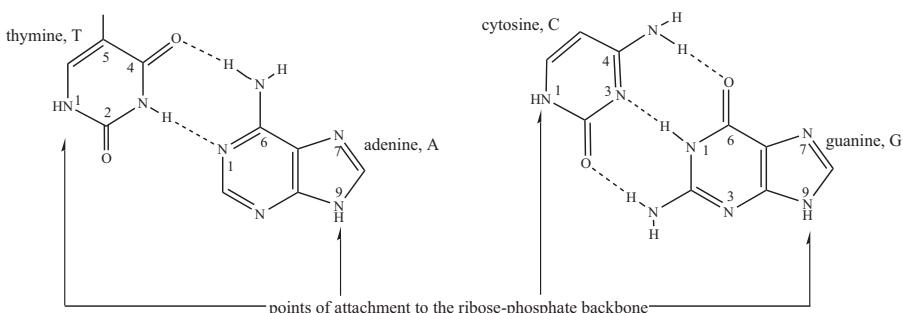
Double-stranded DNA exhibits complementarity in forming the double helix. The complementary sequences have opposite polarity; that is, the two chains run in opposite directions as in the following illustration:



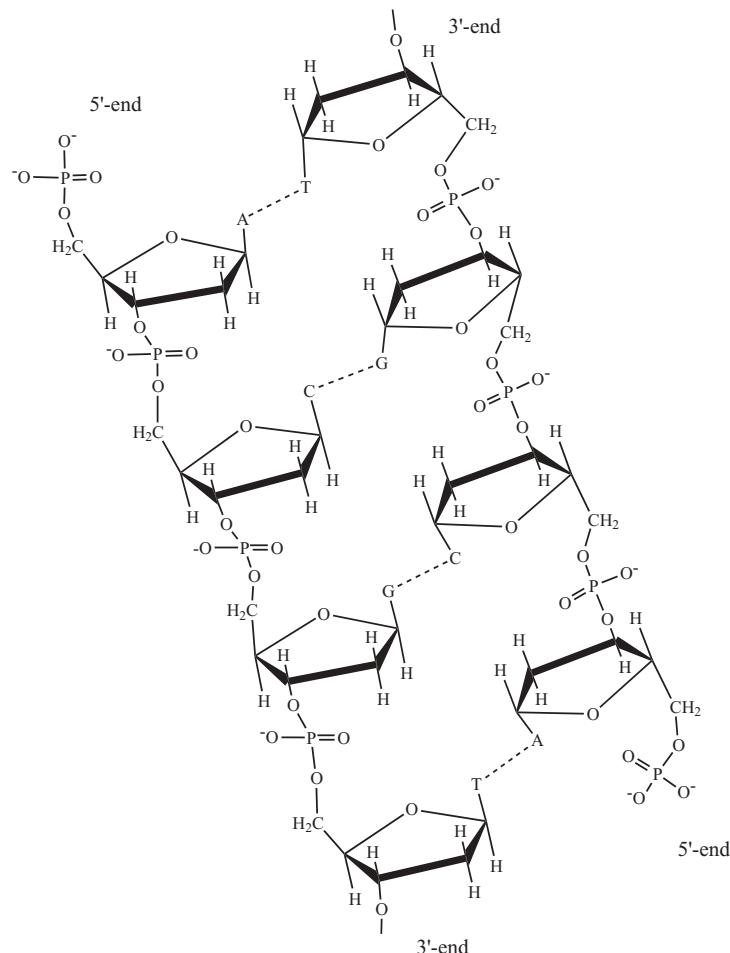
This is described as an antiparallel arrangement. This arrangement allows the two chains to fit together better than if they ran in the same direction (parallel arrangement). A sequence illustrating complementarity is shown in Figure 2.15.

In forming the double-helix polymeric DNA structure, the two sugar-phosphate backbones twist around the central stack of base pairs, generating a major and minor groove. Several conformations, known as DNA polymorphs, are possible. The classical form, B DNA, exhibits well-defined major and minor grooves; C<sub>2</sub>-endo deoxyribose sugar pucks (see Figure 2.16) and parallel stacked base pairs perpendicular to the helix axis. A and B DNA form right-handed helices, while Z DNA forms a left-handed helix. Usually, DNA adopts the B conformation; but in high salt or organic solvent conditions, the Z conformation may be found. Sequences containing alternating GC nucleotides favor the Z DNA conformation. Several different base-sequence-dependent DNA polymorphs may exist within a given genome. Double-stranded RNA or RNA-DNA hybrids normally form the A DNA conformer. Visualizations of all forms are found on the Jena image library of biological molecules site at [http://www.imb-jena.de/IMAGE\\_DNA\\_MODELS.html](http://www.imb-jena.de/IMAGE_DNA_MODELS.html). DNA conformations are dependent on a number of parameters, including the type of pucker found in the nonplanar ribose sugar ring, the syn or anti conformation of the nucleobase relative to the sugar moiety, and the orientation about the C<sub>4</sub>–C<sub>5</sub> bond. Figure 2.16, as adapted from references 14 and 15, illustrates some of the possibilities.<sup>14,15</sup>

The differing conformations of the ribose have been named with respect to that ring atom which puckers out of the plane given by the other ring atoms. The most prominent conformations are C<sub>2</sub>-*endo* in B-helices and C<sub>3</sub>-*endo* in A-helices. At room temperature, both conformers are in a dynamic equilibrium. Intermediates between C<sub>2</sub>-*endo* and C<sub>3</sub>-*endo* are found in several (time-



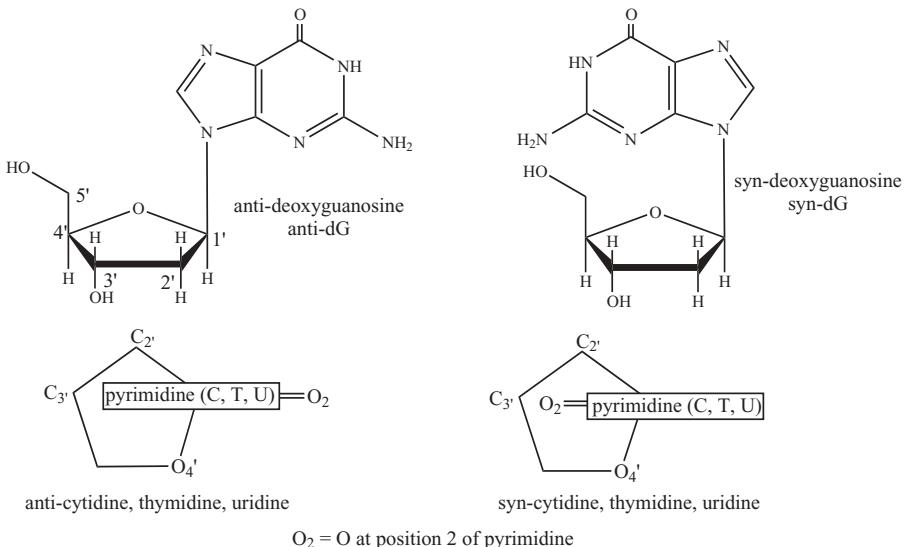
**Figure 2.14** Watson–Crick base pairing in DNA.



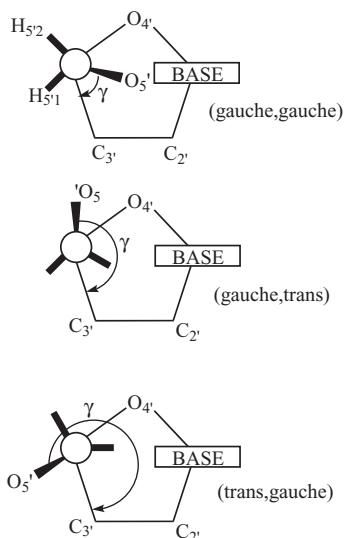
**Figure 2.15** Complimentary antiparallel double-stranded (ds) DNA.

averaged) structures obtained from X-ray crystallography or NMR. When changing the pucker phase angle from  $0^\circ$  to  $360^\circ$ , one steps through all possible conformations of the ribose ring. This pseudorotation cycle can be depicted as a conformational wheel such as that found in Figure 2.8 of reference 14. Figure 9.3 of reference 14 shows a view of  $C_{2'}\text{-endo}$  and  $C_{3'}\text{-endo}$  conformations observed in A- and B-type polynucleotides.

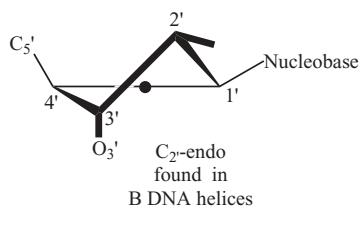
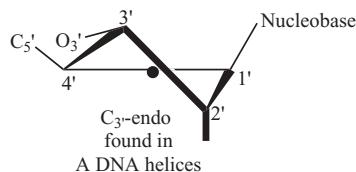
The distance between two subsequent base pairs along the helical axis is called helical rise ( $h$ ). The pitch ( $P$ ) is the length of the helix axis for one complete helix turn. The turn angle per nucleotide or twist angle ( $t$ ) is given by  $360^\circ/\text{number of nucleotides per turn}$ . Data describing some properties of



Definition of torsion angles about the C<sub>4</sub>-C<sub>5'</sub> bond  
looking in the direction C<sub>5'</sub> → C<sub>4'</sub>



Common sugar puckers



**Figure 2.16** Orientations found in DNA helices. [Adapted with permission from Figure 2.11 of reference 14 (copyright 1984, Springer-Verlag) and Figure 1.22 A, B, C of reference 15 (copyright 1997, Wiley-VCH).]

**TABLE 2.3 Some Properties of A DNA, B DNA, and Z DNA**

Property	A DNA	B DNA	Z DNA (the repeat unit is the dimer -G C-)
Helix handedness	Right	Right	Left
Sugar pucker	$C_{3'}\text{-endo}$	$C_{2'}\text{-endo}$	$C_{3'}\text{-endo}$
Number of nucleotides per pitch or base pairs per turn	11	10.4	12
Turn angle per nucleotide or twist angle ( $t, {}^\circ$ )	32.7	34.6	30
Helical rise ( $h, \text{\AA}$ )	2.56	3.38	3.7
Pitch ( $P, \text{\AA}$ )	24.6	34.0	45.6
Diameter ( $\text{\AA}$ )	25.5	23.7	18.4
Conformation of glycosidic bond	Anti	Anti	Anti at C Syn at G
Major (minor) groove width ( $\text{\AA}$ )	2.7(11.0)	11.7(5.7)	
Major (minor) groove depth ( $\text{\AA}$ )	13.5(2.8)	8.5(7.5)	

A, B, and Z DNA structure are found in Table 2.3 as adapted from Table 1.10 of reference 15 and the Jena image library at [http://www.imb-jena.de/IMAGE-DNA\\_MODELS.html](http://www.imb-jena.de/IMAGE-DNA_MODELS.html).

In the nucleus of eukaryotic cells, a compacted DNA is formed by winding in a shallow, left-handed superhelix around a group of eight histone proteins to form nucleosome core particles. The core particles are further organized into chromatin through interaction with additional histone proteins. Chromatin is anchored to a scaffold of proteins within the chromosomes. In order for replication, transcription, or translation to take place, supercoiled DNA must unravel and the DNA strands separate, at least temporarily. It is thought that the unraveling process, or some version of it, may be necessary before platinum antitumor agents such as *cis*-dichlorodiammineplatinum(II), cisDDP, can bind to DNA, their major target *in vitro* and *in vivo*. Much more about the platinum-containing drugs is found in Chapter 7 of the first edition of *Bioinorganic Chemistry: A Short Course*.

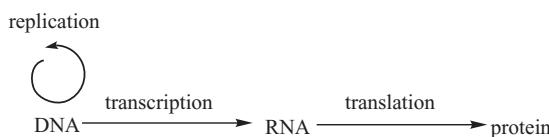
### 2.3.3 Transmission of Genetic Information

Three major components in the transmission of genetic information are deoxyribonucleic acids (DNA), ribonucleic acids (RNA), and proteins. The genetic code expressed through DNA ultimately determines which proteins a cell will produce. Coiled and supercoiled DNA molecules contain numerous sequences of nucleotides that may be transcribed as RNAs and translated to many different proteins. DNA molecules also contain long sequences of nucleotides not coding for protein and whose purpose is not completely understood. A gene is a specific sequence of DNA that encodes a sequence of messenger

RNA (mRNA) codons needed to synthesize a complete protein. Genes may be collected together in a chromosome, the DNA packaging unit in eukaryotes. The human genome is an ordered sequence of almost 3 billion adenine, guanine, cytosine, and thymine bases found on 46 chromosomes. Most cells transfer genetic information DNA → RNA, but retroviruses (such as the human immunodeficiency virus, HIV) transmit their information RNA → DNA. Cells have three major types of RNA. Messenger RNA (mRNA) is produced directly from a DNA template (the sequence of nitrogenous bases along the DNA chain) in a process called transcription. Transfer RNA (tRNA), found in the cytosol, transfers an amino acid to the ribosome for protein synthesis in the translation process. Ribosomal RNA (rRNA) is an essential component of the ribosome, the organelle in which particles of RNA and protein come together to synthesize proteins. The process of RNA interference, RNAi, prevents genes from being translated into proteins. This significant “gene-silencing” discovery, described experimentally in 1998 by biologists Andrew Fire and Craig Mello,<sup>16</sup> revolutionized genetics, transformed understanding of cellular behavior, and has spurred development of new medical treatments. Fire, of Stanford University School of Medicine, and Mello, of the University of Massachusetts Medical School, shared the 2006 Nobel Prize in Physiology or Medicine for their work.<sup>17</sup>

The processes of replication, transcription, and translation are illustrated schematically in Figure 2.17 and further explained in the following paragraphs.

When a cell divides, DNA duplicates itself in a multistep process called *replication*. DNA replication takes place in a semiconservative manner, meaning that the DNA double strand unwinds and each single strand acts as a template for a new complementary strand. Thus, each daughter DNA strand is half comprised of molecules from the old strand. DNA replication is neither a passive nor spontaneous process. Many enzymes are required to unwind the double helix and synthesize a new DNA strand. The enzyme topoisomerase, for instance, is responsible for the initiation of DNA unwinding. Nicking a single DNA strand releases the tension holding the helix in its coiled and supercoiled structure. Another enzyme, helicase, requires energy in the form of ATP to unwind the original double strand, held together by their A–T and G–C hydrogen bonds, finally forming the single-stranded DNA (ssDNA) to be replicated. DNA polymerase holoenzyme, a complex aggregate of several different protein subunits, then proceeds along the single DNA strand, bring-



**Figure 2.17** Schematic representation of replication, transcription, and translation.

ing free deoxynucleotide triphosphates (dNTPs) to hydrogen bond with their complementary nucleotide on the single strand. Energy stored in the incoming dNTP is used to covalently bind each deoxynucleotide to the growing second strand assisted by DNA polymerase III. Primase is an enzyme component of the protein aggregate called the primosome. It attaches a small RNA primer to the single-stranded DNA to act as a substitute 3'-OH for DNA polymerase to begin synthesis of the second strand. The RNA primer is eventually removed by RNase H, and the gap is filled in by DNA polymerase I. The enzyme ligase can catalyze the formation of a phosphodiester bond given adjacent but unattached 3'-OH and 5'-phosphate groups. Platinum-containing anticancer agents attach themselves preferentially to adjacent guanines in double-stranded DNA. The DNA platination causes kinking in the ds DNA, and this is believed to prevent DNA replication by the mechanism just discussed. There are many repair mechanisms that can remove the DNA platination lesion and other mechanisms for skipping over the platinated portion of dsDNA. These repair mechanisms are believed responsible for some of the acquired resistance that cancerous cells develop, eventually rendering the platinum drugs ineffective.

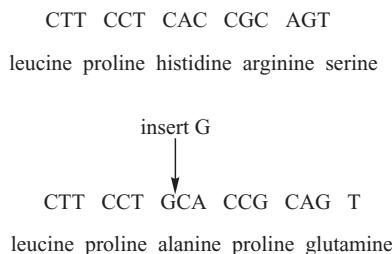
*Transcription* is the process that transfers sequence information from the gene regions of DNA to messenger RNA (mRNA) so that it can be carried to the ribosomes in the cytoplasm. The primary protein aggregate responsible for effecting this process is RNA polymerase holoenzyme. This enzyme aggregate directs the synthesis of mRNA on a DNA template. To begin the process, RNA polymerase must be able to recognize a particular DNA sequence at the beginning of genes called the promoter. The promoter is a unidirectional sequence on one strand of the DNA that tells RNA polymerase both where to start and in which direction (on which strand) to continue synthesis. The DNA strand from which RNA polymerase copies is called the antisense or template strand, and the other DNA strand the sense or coding strand. The RNA polymerase gathers ribonucleic nucleotide triphosphates (NTPs or rNTPs) proceeding to synthesize the single RNA strand in the 5' to 3' direction. Transcription terminates when the RNA polymerase reaches a stop signal on the gene.

In the *translation* process, proteins are synthesized in the cellular factory called the ribosome. The ribosome consists of structural RNA and about 80 different proteins. When inactive, the ribosome contains a large subunit and a small subunit. When the small subunit encounters mRNA, the process of translation to protein begins. Amino acids necessary for protein synthesis bind at two sites on the large subunit, these sites being close enough together to facilitate formation of a peptide bond. One site (A) accepts a new tRNA carrying an amino acid, while the second site (P) bears the tRNA attached to the growing protein chain. Each tRNA has a specific anticodon and acceptor site. The genetic code is composed of 64 triplet *codons* ( $4^3$ , 4 being the nucleobases A, U/T, C, G with the triplet code raising 4 to the third power) so that each of the 20 amino acids has multiple codons. Some examples are as follows: CAT/U

and CAC code for histidine (his); GGT/U, GGC, GGA, and GGG code for glycine (gly); T/UA<sub>A</sub>, T/UAG, and T/UGA are STOP codons that end the protein sequence. Each tRNA also has a specific charger protein, called an aminoacyl tRNA synthetase, that binds to a particular tRNA and attaches the correct amino acid to the acceptor site. The start signal for translation is the AU/TG codon, coding for methionine (met). Not every protein starts with met, so that often this amino acid must be removed during post-translational protein processing. This so-called post-translational modification (PTM) is one of a huge number of protein modifications that take place following protein synthesis. For example, newly synthesized proteins may have certain tyrosine residues phosphorylated or asparagine residues glycosylated. PTMs are important research targets in proteomics, because they have far-reaching implications for cellular behavior (and misbehavior in disease processes). The ribosomal large subunit binds to the mRNA and the ribosomal small subunit and elongation begins. Entering tRNAs with attached amino acids attach to the A site matching their codon to that on the mRNA. A peptide bond is formed with the amino acid attached to the tRNA at the P site followed by a shift of the new peptide to the P site freeing the A site for the next tRNA. When the ribosome reaches a stop codon, no aminoacyl tRNA binds to the empty A site, the ribosome breaks into its large and small subunits, and the newly synthesized protein and mRNA are released.

### 2.3.4 Genetic Mutations and Site-Directed Mutagenesis

Two major types of genetic mutation are possible: *chromosomal mutations* and *point mutations*. Chromosomal mutations can alter or shift large sections of chromosomes (collections of genes), leading to changes in the way their genes are expressed. In *translocations*, large DNA segments are interchanged between two different chromosomes, causing possible changes in gene expression. *Inversions* occur when a region of DNA flips its orientation with respect to the rest of the chromosome, also leading to changes in gene expression. Sometimes large sections of a chromosome may be deleted, leading to loss of important genes. Chromosome *nondisjunction* arises when one daughter cell receives more or less DNA than it should during cell division. Point mutations arise from single base-pair changes. A *nonsense mutation* creates a stop codon where none previously existed, possibly removing essential regions in the resulting protein—that is, U/TAA (stop) instead of U/TAC (tyr). A *missense mutation* changes the mRNA code, placing a different amino acid into the primary sequence and possibly changing the shape or the metal coordination sphere of the resulting protein. For example, AAC (asparagine, asn, N) could be substituted for CAC (histidine, his, N) by a missense point mutation. A *silent mutation* codes for the same amino acid in the protein primary sequence; that is, CAC and CAU/T both code for histidine. Within a gene, insertion of a small number of bases not divisible by three will result in a *frame shift mutation*,



**Figure 2.18** Frame shift mutation caused by insertion of one base, G.

resulting in an incorrect series of amino acids following the frame shift. An example is shown in Figure 2.18.

Organisms have elaborate proofreading and repair mechanisms that recognize false base-pairings and other types of DNA damage, and they repair these. Genetic mutations that occur in spite of proofreading and repair mechanisms cause genetic diseases. Recent advances in genetic manipulation have allowed the source of genetic diseases to be discovered and in some cases treated. Publication of the human genome sequence in 2001, as discussed below, constitutes a large step forward in the information available for scientists studying genetic disease and treatment.

Scientists use mutations to study proteins by introducing amino acid substitutions in crucial locations that determine protein structure or catalytic, binding, and regulatory functions. In order to carry out these mutations a process called *site-directed mutagenesis* is invoked, during which the coding gene must be isolated, cloned, sequenced, and then changed in a specific manner. These processes of DNA isolation and cloning, called recombinant DNA techniques, will be discussed further in the next section. Site-directed mutagenesis enables the scientist to change any residue in any protein to any other residue by changing the DNA code and then transcribing and translating the mutated DNA. The essential process involves extracting the plasmid containing the gene of interest from bacterial cells (usually *Escherichia coli* cells) and then using a series of enzymatic reactions to switch codons. The plasmid is then reintroduced to the *E. coli* cells for expression—transcription and translation—into the mutated protein. The techniques are described in more detail in the following section. The three-dimensional structure of the mutated protein is studied by many methods, one of the most popular being X-ray crystallography. One such study has been carried out on the copper-containing electron transfer metalloenzyme azurin. In studies of the electron transfer properties of this protein, it was found that the his35 residue (the histidine residue found 53 residues along from the N-terminal end of the protein) was part of a hydrophobic region or patch possibly involved in electron transfer between azurin and itself—electron self-exchange (ese)—or a redox partner

such as cytochrome  $c_{551}$ —electron transfer (et).<sup>18</sup> The authors created two mutant azurins, one in which a glutamine residue was substituted for his35 (his35gln or H35Q mutant) and the other in which a leucine residue was substituted for his35 (his35leu or H35L mutant). An important finding resulting from X-ray crystallographic structure of the H35Q mutant compared to the wt azurin structure was that the  $\beta$  barrel strand containing the mutation and the unmutated strand to its left were shifted with respect to the configuration found in the wt azurin. The cleft created in the changed H35Q mutant exposes the Q35 residue to solvent, a feature not found for the wt azurin in which his35 is shielded from solvent and buried within the protein structure. Figure 6 of reference 18 shows the differences in tertiary structure in a stereo diagram. (Stereo diagrams permit 3-D visualization of a molecule on paper or a computer screen. Helpful hints for stereoviewing are found in Section 4.6.1.) The technique for introducing point mutations using site-directed mutagenesis is discussed in the following section.

### 2.3.5 Genes and Cloning

Much of the information contained in this section and that following has been obtained from the websites <http://www.accessexcellence.org/RC/VL/GG/>, <http://www.emc.maricopa.edu/faculty/farabee/biobk/BioBookDNAMOLGEN.html>, and references 1a, 2, 7, and 15. In eukaryotes, species whose cells contain a nucleus and other intracellular structure, gene expression is more complex than in prokaryotes, simpler life forms whose cells have no intracellular compartments. In eukaryotes, DNA is segmented into exons (regions that encode for protein) and introns (regions that do not). When messenger RNA (mRNA) is initially transcribed, both exon and intron regions are included. Subsequently, the mRNA is spliced together at points so that only the exon regions will be translated into protein in the ribosome. DNA molecules are large and complex to work with in their native state. Usually, scientists wish to work with only the DNA exons, and particularly the exons that code for the specific protein(s) of interest. *Recombinant DNA* technology (also called *cloning* and *genetic engineering*) combines a number of different techniques that allow scientists to manipulate and replicate DNA as well as study its structure and function. Recombinant DNA is DNA that has been created artificially by techniques to be described in the following paragraphs. The recombinant DNA molecule must be replicated many times to become useful for laboratory research; this process is called *cloning*.

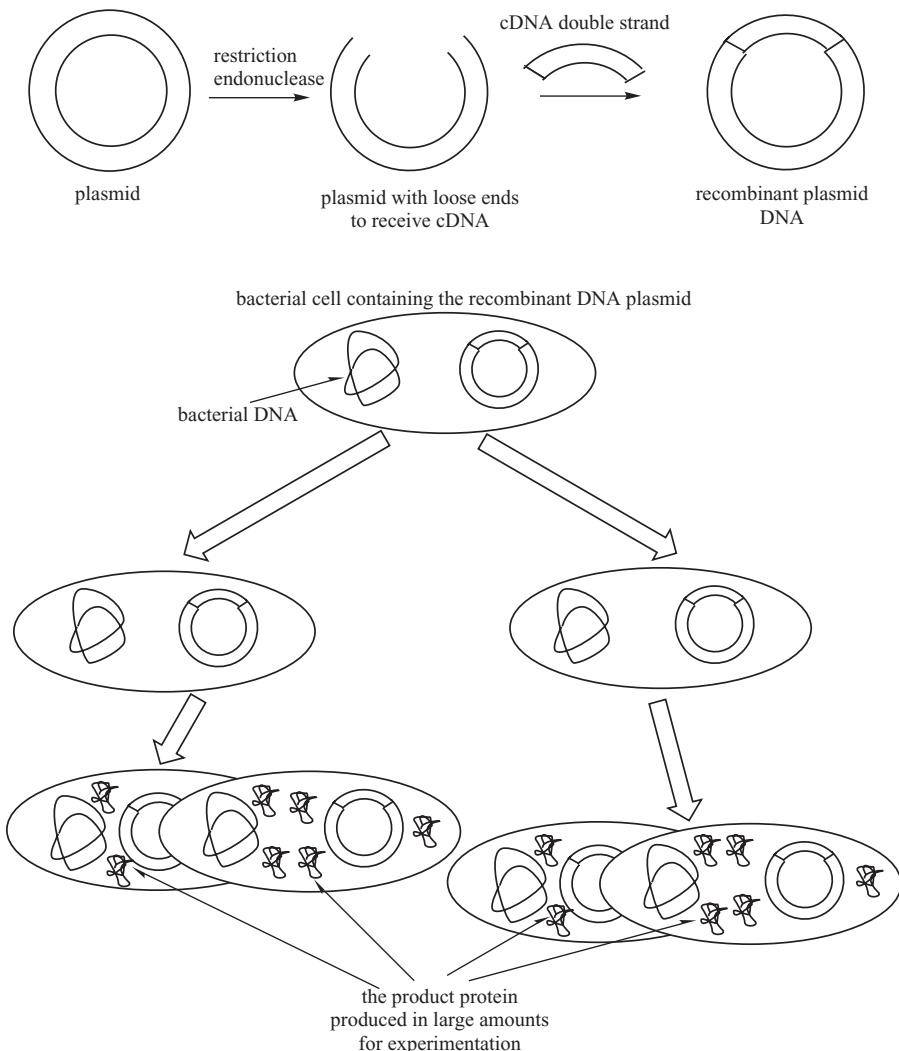
The first step in the cloning process is to break the large DNA molecule down into fragments. Sanger sequencing (also called the chain termination or dideoxy method) uses an enzymatic procedure to produce DNA chains of varying length in four different reactions, stopping DNA replication at positions occupied by one of the four bases and then determining the resulting fragment lengths. This method uses naturally occurring enzymes called *restriction endonucleases*. Restriction endonucleases are a class of enzymes, generally

isolated from bacteria, that recognize and cut specific sequences—restriction sites—in DNA. One restriction enzyme, BamHI, locates and cuts any occurrence of



The enzyme clips after the first G in each strand. Both strands contain the sequence GGATCC but in antiparallel orientation. This type of recognition site is called *palindromic*. Another sequencing method, Maxam–Gilbert (also called the chemical degradation method), uses chemicals to cleave DNA at specific bases, resulting in fragments of different lengths. The resulting DNA restriction fragments are separated by electrophoresis on an agarose gel in a method similar to that described for proteins in Section 2.2.3. Even fragments that differ in size by a single nucleotide can be resolved. The gel-separated fragments are converted to single-stranded DNA (ssDNA) by treatment with strong base. In the Southern blot technique, the gel is then covered with nitrocellulose paper and compressed with a heavy plate. The nitrocellulose paper preferentially takes up the ssDNA, producing a pattern of DNA identical to that on the gel. A DNA probe (either a ssDNA or RNA fragment with a complementary base sequence to the DNA of interest) then identifies the desired single strands of DNA through use of a  $^{32}\text{P}$  radioactively tagged nucleotide probe. For eukaryotic DNA, one usually begins with an mRNA template that will synthesize the desired DNA by treatment with reverse transcriptase, an enzyme that transcribes RNA to DNA, the reverse of the normal transcription process. The synthesized DNA is called cDNA because it carries a complementary base sequence to the mRNA template. In this cloning method, only the DNA exon portion that codes for the desired protein is transcribed, the intron portion that does not code for protein is not transcribed.

Once the desired cDNA or gene fragment has been obtained, it is produced (expressed) in large quantity through the cloning process. DNA cloning *in vivo* (in a living cell) can be carried out in a unicellular prokaryote such as the bacterium *Escherichia coli*, in a unicellular eukaryote such as yeast, or in mammalian cells grown in tissue culture. In any case, the recombinant DNA must be taken up by the cell in a form that can be replicated and expressed. This is accomplished by incorporating the cDNA into a vector, often a plasmid. *Plasmids* are small (a few thousand base pairs) circular DNA molecules that are found in bacteria separate from the bacterial chromosome. Usually they carry only one, or at most a few, genes. The same restriction enzyme that yielded the cDNA of interest can be used to cleave the plasmid, and then the enzyme DNA ligase is used to splice the cDNA of interest into the plasmid. The result is an edited or recombinant DNA molecule. When the recombinant DNA plasmid is inserted into the host, say *E. coli*, the *E. coli* cells will produce many copies of the DNA, which will in turn be transcribed and translated into the desired protein. The process, much simplified, is illustrated in Figure 2.19.



**Figure 2.19** Production of protein via a plasmid containing a cloned cDNA fragment. (Adapted with permission from Figure 3.8 of reference 2. Copyright 1994, University Science Books, Sausalito, CA.)

One of the in vitro (in the test tube) processes used to clone DNA is called the *polymerase chain reaction* (PCR). A vial in which PCR is to be carried out contains all the necessary components for DNA duplication: the piece of DNA to be cloned; large quantities of the four nucleotides, A, T, C, G; large quantities of a primer sequence, a short sequence of about 20 nucleotides synthesized by the primase enzyme; and DNA polymerase. To conduct the process, the vial is first heated to 90–95°C for 30 seconds to separate the two DNA chains in

the double helix. Next, the vial is cooled to 55°C so that the primers will bind (anneal) to the DNA strands. This process takes about 20 seconds. In the final step, the DNA polymerase (usually the Taq polymerase from the hot springs bacterium *Thermophilis aquaticus*, which works best around 75°C) starts making complete copies of the DNA template. The complete sequence of steps for the PCR process takes about two minutes, and a million or so copies of the desired DNA can be produced in one or two hours. The PCR method has found applications in clinical medicine, genetic disease diagnostics, forensic science, and evolutionary biology. For example, very small samples of DNA, perhaps from a crime scene or from an ancient mummy, can produce sufficient copies to carry out forensic tests such as DNA profiling.

### 2.3.6 Genomics and the Human Genome

Each cell in an organism contains its complete genome; but depending on the cell type, only the genes necessary for conducting the work of that cell type are expressed. The human genome sequence consists of an ordered listing of the adenine, cytosine, guanine, and thymine bases found on the 46 human chromosomes. Only about 1% of the genome sequence codes for proteins necessary for human life. Most of the rest of the genome consists of large repetitive noncoding regions whose function is not well understood. It is known, however, that critical clues to diseases such as cancer, diabetes, and osteoporosis lie in areas of the genome that do not code for protein. About one-fourth of the genome contains long, gene-free segments, whereas other regions contain much higher gene concentrations. The number of protein coding regions (genes) in the human genome, estimated at 30,000–40,000, is surprisingly small, given that the fruit fly has 13,000 genes and the thale cress plant has 26,000.<sup>19</sup>

Francis Collins, director of The National Human Genome Research Institute, and Ari Patrinos, head of the Office of Biological and Environmental Research at the Department of Energy's (DOE) Office of Science, managed and coordinated the sequencing of the human genome program, initiated in 1990. A substantially complete version of the 2.9 billion base-pair human genome sequence was published in 2001.<sup>20</sup> Reference 20 reported the results of an international collaboration to produce and make freely available a draft sequence of the human genome, presented an initial analysis of the draft sequence data, and described some of the insights that could be gleaned from the sequence. In addition to the draft sequence, pages 745–964 of this February 2001 issue of *Nature* contained accompanying discussions from many scientists on diverse topics in evolutionary, medical, and biological sciences that would benefit from this accomplishment.

The Human Genome Project used the shotgun sequencing method in which enzymes cut DNA into hundreds or thousands of random bits that were then sent to automated sequencing machines capable of handling DNA fragments up to 500 bases long. After sequencing, the fragments were pieced back

together to become part of the sequenced genome. The shotgun approach was applied to cloned DNA fragments that had already been mapped; that is, the fragment's location on the genome was already known.

The Human Genome Project delivered a complete human genome sequence available to scientists in a freely accessible database when it released the first reference sequence for the human genome in April 2003. The National Institutes of Health website address for current human genome sequence data and information is <http://www.ncbi.nlm.nih.gov/genome/guide/human/>, and the National Human Genome Research Institute's researcher resources website is at <http://www.genome.gov/>.

The Human Genome Project's goal was to produce high-quality, accurate, finished DNA sequences according to the following standards: (1) The DNA sequence is 99.99% accurate. (2) The sequence must be assembled; that is, the smaller lengths of sequenced DNA have been incorporated into much longer regions reflecting the original piece of genomic DNA. (3) The task must be affordable (the project funds technology development to reduce costs as much as possible). (4) The data must be accessible. To this end, verified DNA sequencing data are deposited in public databases on a daily basis (see the website <http://www.ncbi.nlm.nih.gov/genome/guide/human/>). In Section 2.3.5, the Sanger and Maxam–Gilbert methods for DNA cleavage followed by gel electrophoresis for DNA sequencing was described. These so-called first-generation gel-based sequencing technologies can be used to sequence small regions of interest in the human genome; however, these methods are too slow and too expensive for individual chromosomes, let alone a complete genome. The Human Genome Project, in carrying out their goal of affordability, has focused on the development of automated sequencing technology that can accurately sequence 100,000 or more bases per day at a cost of less than \$.50 per base. To achieve this goal, The National Human Genome Research Institute (NHGRI) has continued to support the development of technologies to dramatically reduce the cost of DNA sequencing. In 2005, NHGRI awarded close to \$25 million (to be spent over several years) to support new and emerging technologies, including multiple projects that are based on nanotechnology, specifically on nanopores. Modified protein nanopores, for instance, are being developed. These include (1) genetically engineered  $\alpha$ -hemolysin (a bacterial protein exotoxin capable of hemolysis) pores that would recognize specific bases based on restricted current flow through the pore and (2) chemically modified pores for base recognition using either natural or synthetic nucleobases acting as molecular brakes to slow DNA transit time.

Concurrently with work published by the Human Genome Project, a complete human genome sequence was reported by a consortium of 14 academic, nonprofit, and industrial research groups in work coordinated by Celera Genomics.<sup>21</sup> The following text is excerpted from the abstract of reference 21. In this work, a 2.91-billion base-pair (bp) consensus sequence of the euchromatic portion (the portion containing genes) of the human genome was generated by the whole-genome shotgun sequencing method. Two assembly

strategies—a whole-genome assembly and a regional chromosome assembly—were used, each combining sequence data from Celera and the publicly funded genome effort. Analysis of the genome sequence revealed 26,588 protein-encoding transcripts for which there was strong corroborating evidence and an additional approximately 12,000 computationally derived genes with mouse matches or other weak supporting evidence. Although gene-dense clusters are obvious, almost half the genes are dispersed in low G + C sequence separated by large tracts of apparently noncoding sequence. Only 1.1% of the genome is spanned by exons, whereas 24% is in introns, with 75% of the genome being intergenic DNA. DNA sequence comparisons between the consensus sequence and publicly funded genome data provided locations of 2.1 million single-nucleotide polymorphisms (SNPs). An SNP is a change in which a single base in the DNA under study differs from the usual base at that position. Many SNPs are normal variations in the genome, whereas others are responsible for diseases such as sickle cell anemia. A random pair of human haploid genomes differed at a rate of 1 bp per 1250 on average, but there was marked heterogeneity in the level of polymorphism across the genome. Less than 1% of all SNPs resulted in variation in proteins, but the task of determining which SNPs have functional consequences remains an open challenge.

An interesting short history of efforts to carry out the sequencing of the human genome—“Controversial from the Start” by reporter Leslie Roberts—was published in Science magazine in 2001.<sup>22</sup> Scientists continue to use the information generated by the human genome sequencing publications to understand how genes function, how genetic variations predispose the organism to disease, and how gene function can be used in disease detection, prevention and treatment regimens.

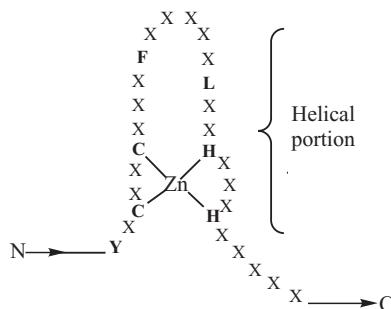
## 2.4 ZINC-FINGER PROTEINS

Zinc-finger proteins, discussed briefly here, provide examples of a bioinorganic topic intimately associated with biochemical knowledge of both proteins and nucleic acids. Sporting a well-recognized finger-like motif, these proteins are known to participate in one of the many and varied protein–DNA interactions that command the attention of bioinorganic researchers. It was known in the 1970s that zinc was crucial to DNA and RNA synthesis and to cell division. In the 1980s it was discovered that the African clawed toad *Xenopus*' transcription factor IIIA (TFIIIA) contained 2–3 mol zinc/mol of protein.<sup>23</sup> TFIIIA is a site-specific DNA-binding regulatory protein that activates the transcription of the 5S RNA gene into DNA. It was found that protein isolated from the 5S RNA complex, containing zinc, was bound to a 45-base-pair DNA sequence and that the protein protects the DNA from nuclease digestion.<sup>2</sup> It was soon discovered that the two cysteine (cys) and two histidine (his) residues per a 30-amino-acid unit of TFIIIA form a tetrahedral coordination complex with each of 7–11 zinc ions. These generate peptide domains, now called

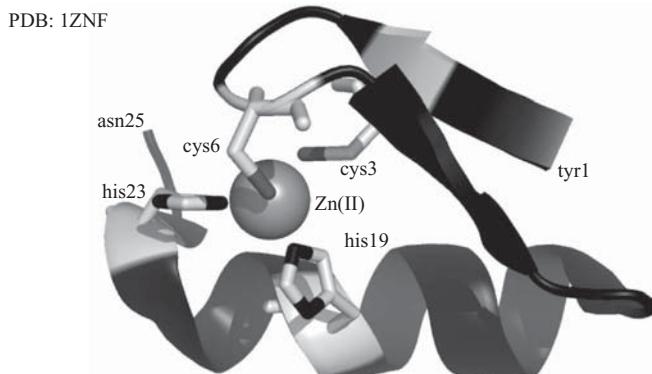
zinc-fingers, that interact with DNA.<sup>24</sup> Zinc ions are required for correct protein folding into its tertiary structure and required for specific DNA interactions, although the Zn(II) ions themselves do not interact directly with the DNA strands. Since its sequence has become available, researchers have attempted to find the most common protein domains encoded within the human genome. The cys<sub>2</sub>his<sub>2</sub> zinc-finger protein domain has been identified as one of the most common and recognizable; in 2004, more than 4000 such domains had been found in over 700 proteins.

Further research has led to a classification of zinc-finger proteins to include various criteria: (1) one or multiple repeat units of about 30 amino acids per finger; (2) both two cys and two his (cys<sub>2</sub>his<sub>2</sub>) with their spacing conserved (the same amino acids occurring in the same locations in many different biological species); (3) two aromatic residues, usually phenylalanine (F), and tyrosine (Y), and the hydrophobic aa leucine (L) conserved in the TFIIIA repeat units; (4) a three-dimensional structure of the sequence that resembles a finger (see Figure 2.20).<sup>25</sup> Comparisons with metalloproteins of known structure have allowed the development of a detailed three-dimensional model for these domains consisting of two antiparallel  $\beta$ -sheets followed by an  $\alpha$ -helix (see Figure 2.21). The proposed structure provides a basis for understanding the detailed roles of the conserved residues and allows construction of a model for the interaction of these proteins with nucleic acids in which the proteins wrap around the nucleic acids in the major groove.<sup>26</sup> Usually, zinc-finger proteins play a regulatory function and are found in nucleic acid polymerases and transcription factors. The role of zinc ion is not catalytic but structural, maintaining the structure of proteins that bind to DNA to activate and deactivate genes.

The consensus aa sequence for zinc-finger repeats, commonly called cys<sub>2</sub>his<sub>2</sub> type zinc-fingers, is often (phe, tyr)-X-**cys**-(X)<sub>2-5</sub>**-cys**-(X)<sub>3</sub>-phe-(X)<sub>5</sub>-leu or other hydrophobic residue-(X)<sub>2</sub>-**his**-(X)<sub>3-5</sub>**-his**-(X)<sub>5</sub>, where X is any amino acid.<sup>27</sup> The residues in bold type indicate amino acids ligands involved in zinc ion binding.



**Figure 2.20** Schematic representation of a zinc-finger domain from TFIIIA and related proteins. X represents any amino acid, and conserved amino acids are histidine (H), cysteine (C), tyrosine (Y), phenylalanine (F), and leucine (L).



**Figure 2.21** Zinc coordination sphere in a zinc-finger protein from the synthetic construct corresponding to *Xenopus* Xfin-31 domain from PDB: 1ZNF. Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft Corporation)

TFIIIA zinc-finger units and others with very similar aa sequences and DNA binding motifs—Zif268, TTK-2, YY1-3, 1MEY and others—have come to be known as “canonical fingers.” Many other known zinc-fingers adopt different aa sequences and DNA binding contacts and have become known as “non-standard fingers.” Figure 3 of reference 27 summarizes both types.

Zinc-finger aa sequences fold in the presence of zinc to form a compact  $\beta\beta\alpha$  domain—two antiparallel  $\beta$ -sheets followed by an  $\alpha$ -helix. (Figure 2.21 illustrates the three-dimensional structure.) The ending  $X_5$  sequence often has the amino acid sequence TGEKP forming a flexible linker between the multiple zinc fingers in a specific protein. The zinc ion coordinates tetrahedrally to two cysteines, lying either in one segment of the  $\beta$ -sheet or in the connector loop between the two segments of antiparallel  $\beta$ -sheet, and to two histidines in the C-terminal end of the  $\alpha$ -helical portion of the protein sequence. Zinc coordination establishes the protein folding pattern, and substituting a residue other than cys or his at one of the ligand positions usually results in a loss of function. In binding to DNA, zinc fingers have several kinds of contacts, the most important of these being hydrogen bonding contacts between aa residues on the  $\alpha$ -helix and bases primarily on one strand of ds DNA. This strand is called the *primary strand*. DNA conformational changes that occur on zinc finger binding include an enlarged major groove resulting from a combination of negative base pair displacement and unwinding of the DNA. Other kinds of interactions between zinc fingers and DNA include hydrophobic and phosphate contacts. Most phosphate contacts are made to the primary DNA strand, the most conserved of these is made through the N $\delta$  of the histidine at position 7 of the protein  $\alpha$ -helix (his7). This histidine is also a zinc ligand through its N $\epsilon$  atom, bringing the phosphate–zinc-finger interaction into close proximity to the DNA strand recognized by the finger.

The criteria described have broadened as more zinc-finger proteins have become known, so that the “zincfinger” designation now serves to describe any relatively short protein sequence that contains four or more cys and/or his residues and which is believed to interact with a nucleic acid binding domain.<sup>28</sup> Small differences in the primary, secondary, and tertiary structure of the proteins are common; for instance, the  $\alpha$ -helical region of the peptide for TFIIIA extends through both histidine ligands; however, the  $\alpha$ -helical region for the ADR1 peptide extends only to the first histidine ligand. In individual zinc fingers, the helical region varies from 5 to 11 amino acids long. Other zinc fingers exhibit 3–10 helices rather than the more common  $\alpha$ -helix. In all cases, the helical region is intimately involved in DNA interaction.

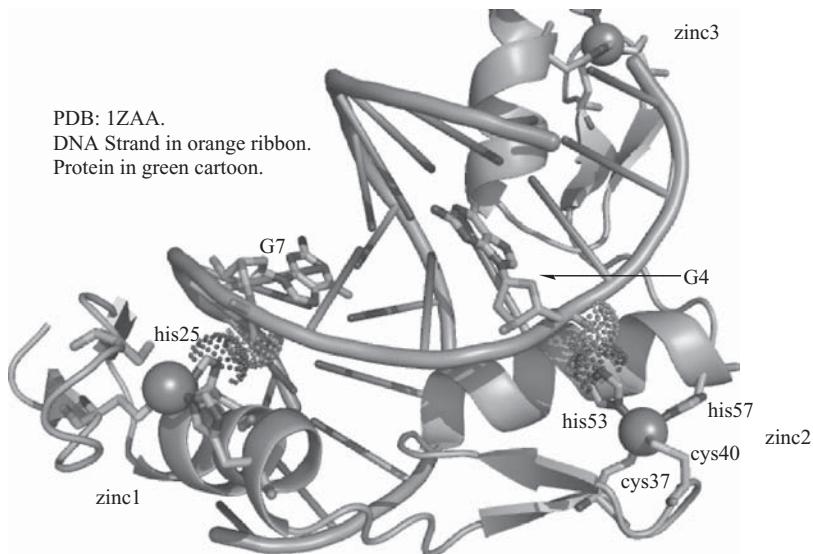
Other zinc-containing structural motifs are known, including the protein GAL4, a transcription factor required for galactose utilization in *S cerevisiae*.<sup>29</sup> The crystal structure of this protein bound to an oligonucleotide indicates that GAL4 is a protein dimer. Each monomer unit contains a binuclear zinc cluster with two zinc ions tetrahedrally coordinated by six cysteines. Two of the cysteines are bridging. Zinc-finger proteins have been found in many species including *Xenopus* TFIIIA as mentioned above, the yeast alcohol dehydrogenase regulatory gene ADR1,<sup>30</sup> the mouse protein ZIF268 extensively studied by many techniques including X-ray crystallography,<sup>31</sup> and the human oncogene GLI protein,<sup>32</sup> one of more than 700 zinc-finger proteins now known to be encoded within the human genome.

In summary, the zinc ions in zinc-finger proteins provide a structural center to direct folding of the protein. The metal ions influence the protein’s three-dimensional shape and define the shape or folding pattern of the peptide domain that interacts specifically with DNA. The proteins target specific sites on DNA, meaning that at least one element of the site recognition by DNA regulatory proteins appears to be recognition of complementary shapes. Zinc fingers change the three-dimensional structure of the B DNA to which they are bound, opening up the major groove. It is now believed that the zinc-finger domain represents a ubiquitous structural motif for eukaryotic DNA-binding proteins.<sup>33</sup> Zinc may be chosen for this purpose for at least two reasons: (1) its natural abundance and (2) the absence of redox activity associated with Zn(II) ions avoiding DNA damage, as might be the case with redox active metal centers such as Fe(II)/(III) or Cu(I)/(II). The number of zinc-finger domains in a single protein ranges from one to as many as 37. A minimum of three fingers seems to be needed for optimal DNA recognition and binding. Several representative zinc-finger domains are shown in Figure 7.4 of reference 2. Some discussion of structural characterization by NMR and X-ray crystallography of individual zinc-finger proteins and attempts to design zinc-finger proteins for binding to specific DNA sequences are discussed in the following section. References 34a and 34b refer readers to two recent reviews on this active research topic, while 34c and 34d refer to advances specific to therapeutic applications of zinc-finger–DNA interactions.<sup>34</sup>

### 2.4.1 Descriptive Examples

One of the first three-dimensional structures of a zinc-finger domain, determined by nuclear magnetic resonance (NMR) spectroscopy, was published in *Science* magazine in 1989.<sup>35</sup> The structure showed what is now known as a typical, compact zinc-finger domain with the tertiary structure featuring two antiparallel  $\beta$ -sheets and one  $\alpha$ -helical region—the  $\beta\beta\alpha$  motif. The structure visualizes a synthetic construct corresponding to the 31st zinc-finger motif from the *Xenopus* (frog) oocyte protein (Xfin-31). Figure 2.21 is a visualization of the Xfin-31 zinc-finger domain rendered from data deposited in the protein data bank (PDB: 1ZNF, found at website <http://www.rcsb.org/pdb>). The zinc ion, buried in the interior of the protein, is shown in silver space fill form, with its cys and his ligands in stick format. The ligands cys3 (in  $\beta$ -sheet 1) and cys6 (in the loop connecting the two antiparallel  $\beta$ -sheets) are shown above and in front of the Zn ion conforming to the zinc-finger pattern CX<sub>2</sub>C. The cysteines ligate to the zinc ion through their S<sub>γ</sub> atoms (shown as dark gray). The ligands his19 and his23, having their N<sub>ε</sub> atoms as zinc ion ligands (histidine nitrogens shown in black), are positioned in the  $\alpha$ -helical region below and to the right of the zinc ion in Figure 2.21, conforming to the zinc-finger pattern HX<sub>3</sub>H. Residues 1–10 of the construct form an antiparallel  $\beta$ -sheet domain with the  $\beta$ -sheets connected by a tight turn. A short loop leads to the  $\alpha$ -helical domain, comprised of residues 12–24. Hydrogen bonds conforming to those typical for antiparallel  $\beta$ -sheets: (1) tyr1 N–H $\cdots$ O–C phe10 and (2) phe10 N–H $\cdots$ O–C tyr1. The  $\alpha$ -helical N–H $\cdots$ O–C hydrogen bonds form between the residues: 16 $\rightarrow$ 12, 17 $\rightarrow$ 13, 18 $\rightarrow$ 14, 19 $\rightarrow$ 15, and 20 $\rightarrow$ 16. The helical motif then changes to a 3<sub>10</sub> pattern with N–H $\cdots$ O–C hydrogen bonds form between the residues: 21 $\rightarrow$ 18, 22 $\rightarrow$ 19, and 23 $\rightarrow$ 20. This early zinc-finger protein NMR solution structure has been followed by a great variety of both X-ray crystallographic and NMR solution studies—in late 2006, a search of the protein data bank (PDB) for “zincfinger protein” yielded 994 hits. Several of these will be discussed below.

The X-ray crystallographic structure from the mouse protein Zif268 and a consensus DNA-binding site has been determined at 2.1-Å resolution, as reported by the authors of reference 31 (deposited as PDB: 1ZAA; see Figure 2.22). In this complex, the zinc fingers bind in the major groove of B-DNA and wrap part way around the double helix. Each zinc-finger domain consists of two antiparallel  $\beta$ -sheets containing two cys ligands (one cysteine in the loop connecting the two antiparallel  $\beta$ -sheets) and an  $\alpha$ -helix containing two histidine ligands. In Figure 2.22, the polypeptide chain is shown in cartoon form: The Zn(II) ions are shown in sphere format, while the histidines and the cysteines ligands are shown in stick format. The DNA double-helix segment is shown in ribbon form with two interacting guanines (G4 and G7) shown in stick format. The interatomic distance between Zn(1) and Zn(2) is 26.6 Å, and the interatomic distance between Zn(2) and Zn(3) is 27.4 Å. Each finger has a similar relation to the DNA and makes its primary contacts in a three-base



**Figure 2.22** Zinc-finger–DNA interactions for PDB: 1ZAA as described in reference 31 and the text. Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft Corporation.) (See color plate)

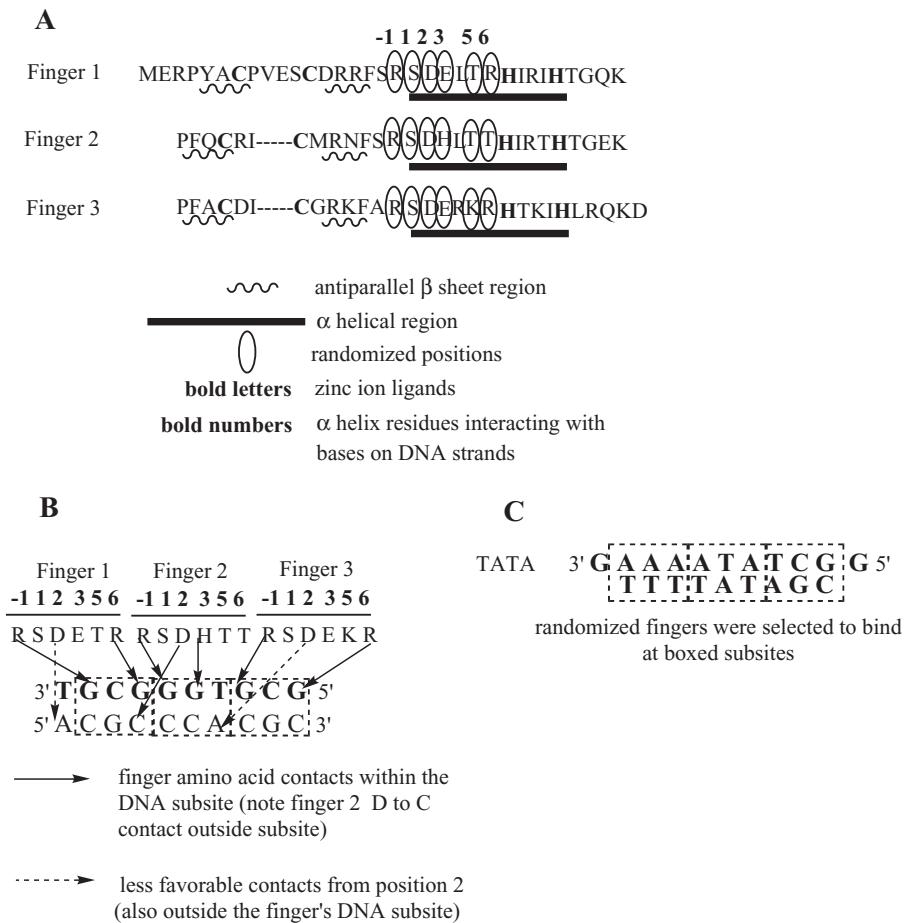
pair subsite. Each of the zinc fingers uses arginine, histidine, and aspartic acid amino acid residues from the N-terminal portion of an  $\alpha$ -helix to make contact the bases in the major DNA groove. Most of the zinc-finger–DNA contacts are made with the guanine-rich strand of the DNA. The  $N_{\epsilon 2}$  nitrogen atoms of his ligands coordinate to the Zn(II) ion; for example, the his53  $N_{\epsilon 2}$ –Zn $^{2+}$  bond distance is 2.00 Å. From the same histidine ligands, the  $N_{\delta 1}$  nitrogen atom hydrogen bonds to phosphodiester oxygens of DNA; for example, the his53  $N_{\delta 1}$ –O<sub>2P</sub> H-bond distance is 2.66 Å. In Figure 2.22, two  $N_{\delta 1}$ –O<sub>2P</sub> hydrogen bonds are indicated by dot surfaces.

Earlier experiments had already shown that zinc-finger protein design could test DNA binding specificity. For instance, Desjarlais and Berg in 1993 had designed three zinc-finger proteins with different DNA binding specificities.<sup>36</sup> The design strategy combined a consensus zinc-finger framework sequence with previously characterized recognition regions such that the specificity of each protein was predictable. The first protein consisted of three identical zinc fingers, each of which was expected to recognize the subsite GCG. The designed protein binds specifically to the sequence 5'-GCG-GCG-GCG-3' with a dissociation constant of approximately 11 μM. The second protein had three zinc fingers with different predicted preferred subsites and would bind to the predicted recognition site 5'-GGG-GCG-GCT-3' with a dissociation constant of 2 nM. A permuted version of the second protein was also constructed and shown to preferentially recognize the corresponding permuted site 5'-GGG-

GCT-GCG-3' over the nonpermuted site. The results indicated that observations on the specificity of zinc fingers can be extended to generalized zinc-finger structures and realized the use of zinc fingers for the design of site-specific DNA-binding proteins.<sup>36</sup>

The Pabo group has described a method for selecting DNA-binding proteins that recognize desired DNA sequences.<sup>37</sup> The research began with amino acid sequences and secondary structure of the three Zif268 zinc fingers, Zif1, Zif2, and Zif3. These are shown schematically in Figure 2.23A as adapted from reference 37. Figure 2.23B shows how the residues in the  $\alpha$ -helical regions of fingers Zif1, Zif2, and Zif3 interact with 3 base-pair subsites in the bound DNA. The bold numbers above the helical residue positions denote the residue within the helix, with -1 denoting a residue outside the N-terminal end of the helix. Note that many contacts are from arginine (R) residues to guanines on the same (primary) DNA strand. Other contacts from position 2 of the  $\alpha$ -helical portion often contact with bases on the complementary strand at positions not within the 3-base-pair subsite. In this research the amino acid residues in  $\alpha$ -helical portions of Zif268's Zif1, Zif2, and Zif3 zinc fingers known to make contact with DNA were randomized into a library. The proteins of the library then were subjected to multiple cycles of selection and amplification by genetic engineering techniques called *phage display*. Phage display has proven to be a powerful enabling technology in genomics and drug development. It allows the directed evolution of proteins engineered for specific properties and selectivity and provides an approach for the engineering of human antibodies, as well as protein ligands, and for such diverse applications as arrays, separations, and drug development. The use of phage display in screening for novel high-affinity ligands and their receptors has been useful in functional genomics and proteomics. Display methods promise to have benefit in the development of therapeutics targeting many different disorders, including cancer, AIDS, autoimmune disorders, and other diseases. In the Pabo group's work the phage display-selected proteins were successively substituted for Zif1, Zif2, and Zif3 in binding to different DNA strands such as the so-called TATA box sequence 3'-GAAAATATCGG-5' shown in Figure 2.23C. The protein sequences were deemed successful if they would bind tightly to the selected DNA sequences—that is, bind with nanomolar dissociation constants and discriminate at greater than 20,000-fold in binding to nonspecific DNA sequences. The authors believe that the protocol described could be adapted to finding zinc fingers capable of binding to many different DNA- and RNA-binding domains and also that their sequential selection strategy could be applicable to the designing of zinc-finger proteins to be used in gene therapy. Reference 34b discusses some recent developments in this research area.

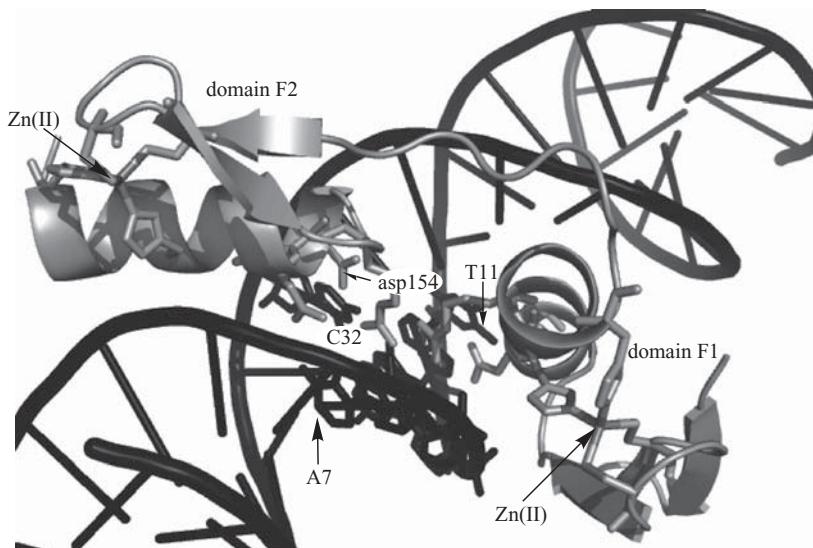
Various other research groups have been working since the early 1990s toward rule formation for zinc-finger proteins and their interactions with DNA. Researchers believe that there are key amino acid positions on the zinc-finger protein that interact with base or phosphate positions on DNA in similar ways for different zinc-finger–DNA systems. These positions may form



**Figure 2.23** (A) Amino acid sequence and secondary structure of Zif1, Zif2, and Zif3 zinc fingers of Zif268. (B) Base contacts of the -1, 1, 2, 3, 5, 6  $\alpha$ -helical amino acid residues of Zif1, Zif2, and Zif3. (C) DNA sequence of the TATA box strand for which zinc fingers were designed. (Figures adapted with permission from Figure 1 of reference 37. Copyright 1997, American Association for the Advancement of Science.)

a recognition code for zinc-finger binding to DNA. One example of such a system—in this case the so-called Tramtrack (TTK) transcriptional regulator from the *Drosophila* development gene *fushi-tarazu*—is shown in Figure 2.24.<sup>38</sup> The TTK structural data are deposited as 2DRP in the protein data bank (PDB).

Two identical zinc-finger–protein–DNA double strand interactions are found in the 2DRP crystal’s unit cell. A close-up view of one of these protein–DNA interactions is visualized in Figure 2.24. A description of the protein–



**Figure 2.24** Amino acid—double-stranded DNA contacts for zinc-finger protein PDB: 2DRP as described in reference 38 and the text. Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft Corporation.)

DNA interaction follows as adapted from reference 38. The 2DRP X-ray crystallographic structure was solved to 2.8-Å resolution. Two zinc-finger motifs (F1 and F2) form independent DNA-binding modules, each positioned with the N terminus of the protein’s  $\alpha$ -helix pointing into the major groove and making base-specific contacts. Most of the protein–DNA contacts are made on the primary strand of the 18-base-pair dsDNA, although one contact in F2—namely, asp154 (2) to C32—connects to the complementary DNA strand. The bold number shown in parentheses represents the position of the residue on the  $\alpha$ -helix. The aa residue, asp154, and the cytosine base, C32, are labeled in Figure 2.24. In Figure 2.24, the DNA strands are shown in black cartoon form and the protein domain is shown in gray. The specific nucleobases and amino acid residues that contact each other are shown in stick format—nucleobases in black and amino acids in gray. Notice that all of the contacting nucleobases are found on one DNA strand—namely, A7, G8, G9, A10, and T11—except for one, C32, which is found on the opposing strand. The three amino acids of zinc-finger domain F1 making contact with the primary strand of DNA are: ser124 (2) to T11, asn125 (3) to A10, arg128 (6) to G9. The bold numbers in parentheses indicate the position of the aa residue along the DNA-contacting  $\alpha$ -helix. (See Figures 2.23A and 2.23B, where the notations are shown for a different zinc-finger domain.) The two amino acids of F2 making contact with the primary DNA strand are: arg152 (−1) to G8 and asn155 (3) to A7. In F1, the Zn(II) ion (shown in stick form) is coordinated by the  $N_{\varepsilon_2}$

**TABLE 2.4 Selected Bond Lengths for the Tramtrack Zinc-Finger Protein**

<i>Coordinate-Covalent Bonds</i>	
Cysteine	Histidine
$S_{\gamma}-Zn(II)(\text{\AA})$	$N_{\epsilon 2}-Zn(II)(\text{\AA})$
cys143 2.26	his159 2.03
cys146 2.28	his164 1.95
<i>Hydrogen Bonds in the F1 Zinc-Finger Domain</i>	
Bonding Atoms	Bond Length ( $\text{\AA}$ )
ser124 (H of OH group, $O_{\gamma}$ ) with T11 (O at position 4, $O_4$ )	2.89
asn125 $O_{\delta 1}$ with A10 (H of $NH_2$ at position 6, $N_6$ )	2.97
arg128 (H of $NH_2$ side chain, $N_{H1}$ ) with G9 (O at position 6, $O_6$ )	2.73
<i>Hydrogen Bonds in the F2 Zinc-Finger Domain</i>	
Bonding Atoms	Bond Length ( $\text{\AA}$ )
arg152 (H of $NH_2$ side chain, $N_{H2}$ ) with G8 (O at position 6, $O_6$ )	2.80
asp154 $O_{\delta 2}$ with C32 (H of $NH_2$ at position 4, $N_4$ )	2.98
asn155 $O_{\delta 1}$ with A10 (H of $NH_2$ at position 6, $N_6$ )	3.17

atoms of two histidines from the  $\alpha$ -helical region—his129 and his134—and the  $S_{\gamma}$  atoms of two cysteines from the  $\beta$ -sheet and loop region—cys113 and cys116. In F2, the Zn(II) ion (shown in stick form) is coordinated by the  $N_{\epsilon 2}$  atoms of two histidines from an  $\alpha$ -helical region—his159 and his164—and the  $S_{\gamma}$  atoms of two cysteines from the  $\beta$ -sheet and loop region—cys143 and cys146. Bond and interaction distances are collected in Table 2.4.

The position of amino acids in sequence, with beginning and ending residue numbers shown, for F1 is—Y(111)XCX<sub>2</sub>**CX**<sub>3</sub>YX<sub>5</sub>FX<sub>2</sub>**HX**<sub>4</sub>HX<sub>6</sub>Y(141)—with the zinc ion ligand atoms shown in boldface. The amino acid sequence for F2, beginning with the right hand ending Y from the F1 sequence, is—Y(141)XC X<sub>2</sub>**CFX**<sub>2</sub>FX<sub>8</sub>**HX**<sub>4</sub>HX<sub>2</sub>(166)—. The  $\alpha$ -helical region of F1 runs from residues 123 to 134, encompassing all its DNA strand contacts and the histidine Zn(II) ligands. The  $\alpha$ -helical region of F2 runs from residues 153 to 164, encompassing all its DNA strand contacts and the histidine Zn(II) ligands.

Shi and Berg have reported on zinc-finger proteins that bind to DNA–RNA hybrids as well as those that bind to dsDNA.<sup>39</sup> Previous X-ray crystallographic structures such as that reported in references 31, 37, and 38 have indicated that zinc-finger proteins have more contacts with one strand of dsDNA than with the other. Also, the structures of the DNA strands in complexes with zinc-finger proteins have characteristics of both A- and B-form DNA. These findings led Shi and Berg to examine the binding of two different zinc-finger proteins (Sp1, a human transcription factor that contains three zinc fingers, and ZF-QQR) to dsDNA, to double-stranded RNA, and to DNA–RNA hybrids. RNA, of course, differs from DNA in the presence of the 2'-OH group

in RNA, which will in turn affect contacts between the zinc-finger proteins and the nucleic acids. While the binding of Sp1 to its preferred DNA–DNA and DNA–RNA segments was similar, ZF-QQR not only bound more tightly to its preferred DNA–DNA segments than Sp1 but bound DNA–RNA segments five times more strongly than the DNA–DNA segments. Both Sp1 and ZF-QQR proteins bound much less well to RNA–DNA hybrids. The authors found that interactions between the zinc-finger proteins and the DNA–RNA hybrids were dependent on which strand was RNA (DNA–RNA preferred over RNA–DNA by both Sp1 and ZF-QQR proteins) and were sequence-specific as well. It is also important to realize that although DNA–RNA hybrids are not well understood, it is known that they do adopt structures that are intermediate between A- and B-form DNA. The authors concluded that interactions with DNA–RNA hybrids should be considered with regard to the biological roles of zinc-finger proteins because these proteins could be designed to target specific DNA–RNA hybrid structures *in vivo*.

In conclusion, research on zinc-finger–DNA interactions has indicated that secondary structure of the zinc-finger proteins as well as their primary amino acid residue sequences are important in designing proteins to interact with specific DNAs. The points of DNA contact appear to be principally on one strand of dsDNA and, for many but certainly not all interactions, occur at guanine bases. Although early research suggested that the zinc-finger protein–DNA interactions would be readily characterized, copied, and applied to new systems through application of a “recognition code,” the situation now appears to be more complex than originally thought. Progress continues on research that will hopefully lead to advances in zinc-finger–DNA interactions along with the capability to use this knowledge to good effect in gene therapeutic and other medicinal applications. The latest review information on DNA recognition by cys<sub>2</sub>his<sub>2</sub> zinc-finger proteins at the time of this book was contained in the articles of reference 34.

## 2.5 SUMMARY AND CONCLUSIONS

The material introduced in this chapter was chosen to reflect the emphasis of information presented in the rest of this short text. In Chapter 3, more information on the X-ray crystallographic and NMR solution structural techniques applied to proteins and nucleic acids will be presented. Analysis of metalloproteins and metalloenzymes discussed in Chapters 5, 6, and 7 provides important information not only on structures but on functions of these bioinorganic systems. Knowledge of the protein composition and structural analyses of these metalloproteins and metalloenzymes will be assumed on the reader’s part from the background presented in this chapter and the next. The student is invited to refer to Chapter 2 materials often for greater understanding of the bioinorganic systems presented in the following chapters.

## REFERENCES

1. (a) Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry*, 4th ed., W. H. Freeman, New York, 2005. (b) Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry*, 6th ed., W. H. Freeman, New York, 2006. (c) Lehninger, A. L.; Nelson, D. L.; Cox, M. M. *Principles of Biochemistry*, 3rd ed., Worth Publishers, New York, 2000.
2. Lippard, Stephen J.; Berg, Jeremy, M. *Principles of Bioinorganic Chemistry*, University Science Books, Mill Valley, CA, 1994.
3. Creighton, T. E. *Proteins, Structures and Molecular Properties*, 2nd ed., W. H. Freeman, New York, 1992.
4. Verneude, X.; Lavault, B.; Ohana, J.; Nurizzo, D.; Joly, J.; Jacquemet, L.; Felisaz, F.; Cipriani, F.; Bourgeois, D. *Acta Crystallogr., Sect. D* 2006, **62**, 253.
5. Chang, I.; Cieplak, M.; Dima, R. I.; Maritan, A.; Banavar, J. R. *Proc. Natl. Acad. Sci. USA*, 2001, **98**(25), 14350–14355. (b) Cieplak, M.; Holter, N. S.; Maritan, A.; Banavar, J. R. *J. Chem. Phys.*, 2001, **114**(3), 1420–1423.
6. (a) Dunbrack, R. L., Jr. *Curr. Opin. Struct. Biol.* 2006, **16**(3), 374–384. (b) Bowie, J. U. *Nature*, 2005, **438**(7068), 581–589. (c) Thirumalai, D.; Hyeon, C. *Biochemistry*, 2005, **44**(13), 4957–4970. (d) Watters, A. L.; Baker, D. *Eur. J. Biochem.* 2004, **271**(9), 1615–1622.
7. Lyman, B. A. *Biochemistry*, Springhouse Corporation, Springhouse, PA, 1994.
8. Johnson, M. K. In Que, L., ed., *Physical Methods in Bioinorganic Chemistry: Spectroscopy and Magnetism*. University Science Books; Sausalito, CA, 2000, pp. 233–285.
9. Henry, C. M. *C&E News*, April 2, 2001, 47–49.
10. (a) Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. *Anal. Chem.* 2001, **73**, 2836–2842. Reynolds, K. J.; Yao, X.; Fenselau, C. J. *Proteome Res.*, 2002, **1**(1), 27–33. (c) Yao, X.; Afonso, C.; Fenselau, C. J. *Proteome Res.*, 2003, **2**(2), 147–152.
11. (a) Warsheid, B.; Jackson, K.; Sutton, C.; Fenselau, C. *Anal. Chem.*, 2003, **75**, 5608–5617. (b) Whiteaker, J. R.; Fenselau, C. C.; Fetterolf, D.; Steele, D.; Wilson, D. *Anal. Chem.*, 2004, **76**, 2836–2841. (c) Pribil, P.; Fenselau, C. *Anal. Chem.*, 2005, **77**, 6092–6095.
12. (a) Rahbar, A. M.; Fenselau, C. J. *Proteome Res.*, 2005, **4**(6), 2148–2153. (b) Strong, R.; Nakanishi, T.; Ross, D.; Fenselau, C. J. *Proteome Res.*, 2006, **5**(9), 2389–2395.
13. Leitner, A.; Lindner, W. *Proteomics*, 2006, **6**(20), 5418–5434.
14. Saenger, W. *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, 1984.
15. Cowan, J. A. *Inorganic Biochemistry, An Introduction*, 2nd ed., Wiley-VCH, New York, 1997.
16. Fire, A.; Mello, C. *Nature* 1998, **391**(6669), 806–811.
17. Couzin, J. *Science*, 2006, **314**, 34. (b) Arnaud, C. *C & E News*, 2006, **84**(41), 8.
18. Nar, H.; Messerschmidt, A.; Huber, R.; van de Kamp, M.; Canters, G. W. *J. Mol. Biol.*, 1991, **221**, 427–447.
19. Borman, S. *C&E News*, Feb. 12, 2001, 9.
20. International Human Genome Sequencing Consortium *Nature*, 2001, **409**, 860–921.

21. Venter, J. C., et al. *Science*, 2001, **291**, 1304–1351.
22. Roberts, L. *Science*, 2001, **291**, 1182–1188.
23. Hanas, J. S.; Hazuda, D. J.; Bogenhagen, D. R.; Wu, F. Y.-H.; Wu, C.-W. *J. Biol. Chem.*, 1983, **258**, 14120.
24. Klug, A.; Rhodes, D. *Trends Biochem. Sci.*, 1987, **12**, 464.
25. (a) Vallee, B. L.; Auld, D. S. *Acc. Chem. Res.*, 1993, **26**, 543–551. (b) Frankel, A. D.; Bredt, D. S.; Pabo, C. O. *Science*, 1998, **240**, 70.
26. Berg, J. M. *Acad. Sci. USA*, 1988, **85**, 99–102.
27. Wolfe, S. A.; Nekludova, L.; Pabo, C. O. *Annu. Rev. Biophys. Biomol. Struct.* 2000, **29**, 183–212.
28. Berg, J. M. *Curr. Opin. Struct. Biol.*, 1993, **3**, 11.
29. (a) Marmorstein, R.; Harrison, S. C. *Genes Dev.*, 1994, 2504–2512. (b) Marmorstein, R.; Carey, M.; Ptashne, M.; Harrison, S. C. *Nature*, 1992, **356**, 408.
30. Parraga, G.; Horvath, S. J.; Eisen, A.; Taylor, W. E.; Hood, L.; Young, E. T.; Klevit, R. E. *Science*, 1988, **241**, 1489.
31. Pavletich, N. P.; Pabo, C. O. *Science*, 1991, **252**, 809–817. (PDB: 1ZAA)
32. Pavletich, N. P.; Pabo, C. O. *Science*, 1993, **261**, 1701–1707. (PDB: 2GLI)
33. Bertini, I.; Gray, H. B.; Lippard, S. J.; Valentine, J. S. *Bioinorganic Chemistry*, University Science Books, Mill Valley, CA, 1994, pp. 491–493.
34. (a) Jantz, D.; Amann, B. T.; Gatto, G. J., Jr., Berg, J. M. *Chem. Rev.*, 2004, **104**, 789–799. (b) Dhanasekaran, M.; Negi, S.; Sugiura, Y. *Acc. Chem. Res.*, 2006, **39**, 45–52. (c) Klug, A. *FEBS Lett.*, 2005, **579**, 892–894. (d) Barbas, C. F., III. *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 11716–11721.
35. Lee, M. S.; Gippert, G. P.; Soman, K. V.; Case, D. A.; Wright, P. E. *Science* 1989, **245**, 635–637. (PDB: 1ZNF)
36. Desjarlais, J. R.; Berg, J. M. *Acad. Sci. USA*, 1993, **90**, 2256–2260.
37. Greisman, H. A.; Pabo, C. O. *Science*, 1997, **275**, 657–660.
38. Fairall, L.; Schwabe, J. W. R.; Chapman, L.; Finch, J. T.; Rhodes, D. *Nature*, 1993, **366**, 483–487. (PDB: 2DRP)
39. Shi, Y.; Berg, J. M. *Science*, 1995, **268**, 282–284.

---

# 3

---

## INSTRUMENTAL METHODS

### 3.1 INTRODUCTION

#### 3.1.1 Analytical Instrument-Based Methods

Lawrence Que, Jr., has stated in the preface to *Physical Methods in Bioinorganic Chemistry*: “By piecing together the various clues derived from the physical methods, bioinorganic chemists have been able to form a coherent picture of the metal binding site and to deduce the role of the metal ion in a number of biological processes.”<sup>1a</sup> Physical methods are used for analysis of systems under study by all chemists, whether or not they are trained analytical chemists. Therefore all researchers must become familiar with the capabilities and limitations of the analytical methods they use. In addition to instrumental and physical methods, chemists have additional investigational capabilities based on computer methodology. This chapter contains an introduction to some physical (instrumental), and Chapter 4 will address computer-based methods. The discussion is not intended as an exhaustive description or even listing of available methods but rather concentrates on those mentioned frequently in this text. The methods include X-ray absorption spectroscopy (XAS and EXAFS), X-ray crystallography, nuclear magnetic resonance, electron paramagnetic resonance (including ENDOR), mass spectrometry, and Mössbauer spectroscopies. The last method mentioned is especially relevant for characterization of bioinorganic iron-containing species. Other important methods used to analyze bioinorganic species such as ultraviolet-visible, infra-

red, resonance Raman, circular dichroism, magnetic circular dichroism spectroscopies, and many others are not discussed here. Students are referred to analytical and instrumental texts<sup>2</sup> for more information on all methods, discussed here or not. The volumes listed in reference 1 are especially helpful because they emphasize instrumental methods applied to the analysis of bioinorganic systems.

### 3.1.2 Spectroscopy

Students will be familiar with the absorption or emission of electromagnetic radiation as the basis for spectroscopic methods. Electromagnetic radiation itself is perceived as mutually perpendicular oscillating electric and magnetic fields. The total energy of the radiation, which has a number of components, is determined by the relationship shown in equation 3.1:

$$E_{\text{total}} = h\nu = \frac{hc}{\lambda} = E_{\text{translation}} + E_{\text{rotation}} + E_{\text{vibration}} + E_{\text{electron spin}} + E_{\text{nuclear spin}} + E_{\text{nuclear levels}} + \dots \quad (3.1)$$

where

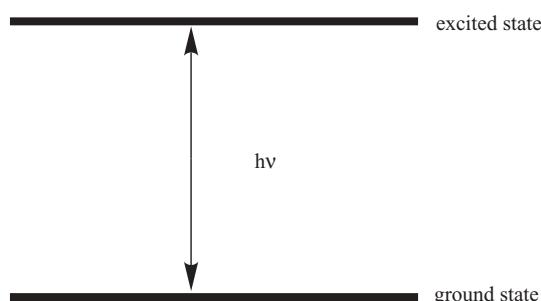
$h$ =Planck's constant= $6.626 \times 10^{-34}$  Js

$c$ =the speed of light= $3 \times 10^8$  m s<sup>-1</sup>

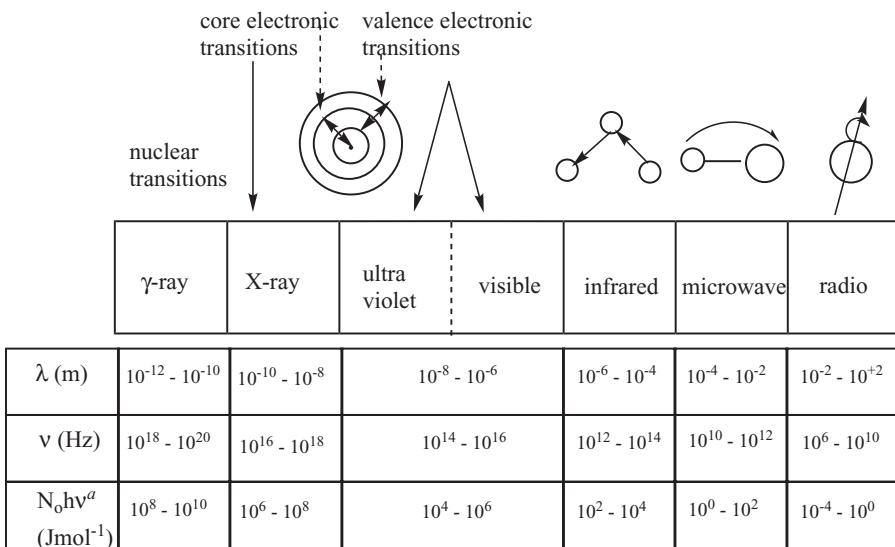
$\nu$ =the frequency of light in s<sup>-1</sup> (Hz)

$\lambda$ =the wavelength of light in meters (m)

Atomic and molecular energy levels represent specific quantum states (ground and excited) illustrated in Figure 3.1. Transitions between these states, which may be caused either by energy absorption or energy emission, are responsible for physical method observations. Some details are shown in Figure 3.2. Energy transition examples include:



**Figure 3.1** Energy transition (absorption or emission) between ground and excited states.



<sup>a</sup> $N_o$  = Avogadro's number =  $6.023 \times 10^{23}$

**Figure 3.2** Illustrated energy transitions for several useful regions of the electromagnetic spectrum. (Adapted with permission of John Wiley & Sons, Inc. from Figure 2.2 of reference 3. Copyright 1997, Wiley-VCH.)

1. Moving an electron from a ground state to an excited state ( $E_{\text{electronic}}$ ), leading to observations in the ultraviolet-visible spectroscopic region
2. Reorienting an electron in a magnetic field ( $E_{\text{electron spin}}$ ) as seen in electron paramagnetic resonance (EPR) spectroscopy
3. Reorienting a nuclear moment in a magnetic field ( $E_{\text{nuclear spin}}$ ) as observed in nuclear magnetic resonance (NMR) spectrometry
4. Detecting energy absorbed or emitted in deforming bonds between atoms in molecules ( $E_{\text{vibration}}$ ) as observed in infrared or resonance Raman spectroscopy

The entire spectrum of electromagnetic resonance, as shown in Figure 3.2 adapted from Cowan,<sup>3</sup> is available for various physical methods. High-energy, short-wavelength transitions occur at the left end of Figure 3.2, whereas low-energy, long-wavelength transitions occur at the right.

### 3.2 X-RAY ABSORPTION SPECTROSCOPY (XAS) AND EXTENDED X-RAY ABSORPTION FINE STRUCTURE (EXAFS)

#### 3.2.1 Theoretical Aspects and Hardware

X-ray diffraction from crystalline samples can result in a complete three-dimensional crystal structure of a molecule, but requires a single crystal suit-

able for proper diffraction (see Section 3.3). X-ray absorption spectroscopy (XAS) can yield limited molecular structural information on noncrystalline (amorphous) solid samples, frozen solutions, solutions, and gases. A tunable X-ray source is required because of the energy-dependent absorption coefficient of the material under study. Synchrotron radiation has provided a high-intensity tunable source of X-rays, and its use has been increasingly responsible for the popularity of this technique over the last three decades.<sup>4</sup> The following discussion summarizes the material found in reference 4. Another good source for basic information is the following: Matthew Newville at the Consortium for Advanced Radiation Sources, University of Chicago has posted a pdf file “Fundamentals of XAFS” at the website [http://cars9.uchicago.edu/xafs/xas\\_fun/xas\\_fundamentals.pdf](http://cars9.uchicago.edu/xafs/xas_fun/xas_fundamentals.pdf).

Figure 1 of reference 4 shows a typical X-ray absorption spectrum. A sharp rise, called the X-ray absorption edge, occurs at a well-defined X-ray photon energy, which is unique to the absorbing element. The absorption edge is due to electron dissociation from a core level of the absorbing atom to valence energy levels. Spectral features in the edge region, sometimes called the near edge, are related to the electronic structure of the absorbing atom and can often be used to identify the geometric arrangement of its ligands. For instance, one might be able to distinguish between octahedral and tetrahedral geometric arrangements about the absorbing atom through analysis of XAS edge spectral features. For the same absorbing atom, differences in absorption energy at the edge are related to its valence or oxidation state. Usually, higher oxidation states will result in absorption at higher energies. This analysis is often given the acronym XANES (X-ray absorption near-edge structure).

In the region above (to the right of) the edge, variation in the X-ray absorption coefficient known as extended X-ray absorption fine structure (EXAFS) may be analyzed to yield structural information about atoms in the ligand sphere (within 4- to 5-Å radius) of the absorbing atom. The EXAFS phenomenon arises through interaction of photoelectrons from the absorbing atom, symbolized by an “a,” with electron density of surrounding atoms. Scattering atoms or scatterers are symbolized by an “s.” Scattering atoms contribute damped sine waves of measurable frequency (related to distance between a and s), amplitude (related to coordination number about a and s atom types), and phase (related to s atom type). Analysis of EXAFS data can answer the question, How many of what type of atom are at what distance from the absorbing atom?

The XAS spectrometer is similar to a UV-visible system in that it consists of a source, a monochromator, and a detector. The most favorable XAS source, synchrotron radiation, is tunable to different wavelengths of desirable high intensity. A laboratory instrument for analysis of solids and concentrated solutions may use a rotating anode source (further described in Section 3.3). The monochromator for X-ray radiation usually consists of silicon single crystals. The crystals can be rotated so that the wavelength ( $\lambda$ ) of the X-rays produced depends of the angle of incidence ( $\theta$ ) with a Bragg lattice plane of

the crystal and the  $d$  spacing of the crystal according to Bragg's law as shown in equation 3.2:

$$n\lambda = 2d \sin \theta \quad (3.2)$$

Detectors for quantitative measurement of X-ray absorption spectra must measure the flux (photons  $s^{-1}$ ) of the X-ray beam. Ionization chambers consisting of X-ray transparent windows on each end of a chamber holding an inert gas work well as transmission detectors for concentrated samples. For transmission detectors,  $\ln(I_0/I)$  is proportional to the absorption coefficient of the absorbing atom,  $\mu$  ( $I_0$ =incident X-ray photon intensity,  $I$ =transmitted intensity), according to Beer's Law:

$$I = I_0 e^{-\mu t} \quad (3.3)$$

and the absorption coefficient varies according to the following:

$$\mu \approx \frac{pZ^4}{AE^3} \quad (3.4)$$

where

$\mu$ =absorption coefficient

$Z$ =atomic number

$A$ =atomic mass

$E$ =X-ray energy

Fluorescence excitation techniques provide a more sensitive detection system in which fluorescent X-ray photons (a fraction of the ionized absorbing atoms relax by emission of a fluorescent X-ray photon) are counted as the photon energy is scanned. The signal generated is proportional to the absorption coefficient,  $\mu$ , of the absorbing atom.

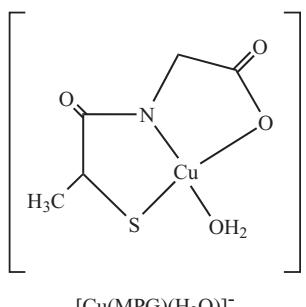
The data collected are subjected to Fourier transformation yielding a peak at the frequency of each sine wave component in the EXAFS. The sine wave frequencies are proportional to the absorber-scatterer (a-s) distance  $R_{as}$ . Each peak in the display represents a particular shell of atoms. To answer the question of how many of what kind of atom, one must do curve fitting. This requires a reliance on chemical intuition, experience, and adherence to reasonable chemical bond distances expected for the molecule under study. In practice, two methods are used to determine what the back-scattered EXAFS data for a given system should look like. The first, an empirical method, compares the unknown system to known models; the second, a theoretical method, calculates the expected behavior of the a-s pair. The empirical method depends on having information on a suitable model, whereas the theoretical method is dependent on having good wave function descriptions of both absorber and scatterer.

While extended X-ray absorption fine structure (EXAFS) in general does not give direct information about the geometry of ligand atoms about the metal center, it is possible to relate specific X-ray absorption edge behaviors to geometry for nickel(II) compounds, for instance. In the examples discussed by Scott in Figure 13 of reference 4, Ni(II) complexes with square-planar ( $D_4h$ ) symmetry give rise to a K edge spectra with a large characteristic pre-edge peak about 5 eV below the main edge. The peak is assigned to a  $1s \rightarrow 4p_z$  transition with simultaneous ligand  $\rightarrow$  metal charge transfer (LMCT) to a Ni(II) 3d orbital. Because neither octahedral nor tetrahedral Ni(II) complexes show this peak (but have their own characteristic spectral behavior), one can, in combination with information about ligands and bond lengths from EXAFS, assign metal complex geometry. Figure 13 of reference 4 also shows another typical behavior: Soft ligands result in lower edge energy. Information about the oxidation state of a metal ion can also be gathered from edge and pre-edge behavior. Cu(I) two- and three-coordinate complexes (tetrahedral Cu(I) complexes do not show the behavior) exhibit a pre-edge peak that disappears upon oxidation to Cu(II), allowing for quantitative determination of Cu(I) content in these systems.<sup>5</sup>

### 3.2.2 Descriptive Examples

Quantitative information about the first coordination sphere structure depends on analysis of EXAFS data. From analytical data or knowledge of common ligands in metalloenzymes (N, O, S, Se) one can decide which ligands are likely to be present in the coordination sphere. An example discussed by Scott<sup>4</sup> tests the hypothesis of a Cu(II)-S bond being present in the compounds shown in Figure 3.3.

Fourier transformation of Cu EXAFS data gathered on the Cu(MPG) complex reveals two separate peaks representing shells at distances of 1.9 and 2.3 Å. When tested for  $N_s$  (coordination number), metal-ligand distance ( $R_{as}$ ) and Debye-Waller parameter difference ( $\Delta\sigma_{as}^2$ ) followed by comparison to known model compounds, results show that the presence of both a Cu-(N,O)



**Figure 3.3** Cu(II) complex with potential chelating ligand mercaptopyrrolidine-2-carboxylate (MPG).

and Cu–S shell is necessary to obtain an adequate fit to the EXAFS data. Therefore it was concluded that a Cu–S bond is present in the compound.

Chapter 7 discusses the ubiquitous cytochrome proteins. Many of these proteins and their model compounds have been studied using EXAFS. For instance, the cytochrome P450 enzyme, one of a large group of enzymes that utilize atmospheric oxygen to functionalize biological molecules, exhibits a very short Fe–O bond at one point during its catalytic cycle. This has been assigned as a ferryl(oxo),  $\text{Fe}(\text{IV})=\text{O}$ , bond using EXAFS spectroscopy. EXAFS data indicated an  $\text{Fe}=\text{O}$  bond length of  $1.65 \pm 0.05 \text{ \AA}$ , similar to that found for the so-called “compound I” intermediates in many iron-containing enzymes such as hemoglobin, myoglobin, catalases, horseradish peroxidase, cytochrome c peroxidase, and chloroperoxidases.<sup>6</sup> Many researchers have sought to identify, characterize, and synthesize model compounds corresponding to the oxidizing species called compounds I and II that occur in iron-containing metalloenzymes. The compounds I and II intermediates are known to correspond to formal iron-oxidation states two and one above the resting iron(III) state. One sought-after model compound for cytochrome P450 would feature a  $\text{Fe}(\text{IV})=\text{O}$  state with a sulfur thiolate ligand attached directly opposite the oxo ligand. In this regard, Eckard Münck and Larry Que have recently published a report on the synthesis and characterization of  $[\text{Fe}^{\text{IV}}(\text{O})(\text{TMCS})]^+$ , where TMCS is a pendadentate ligand, the monoanion of 1-mercaptopethyl-4,8,11-trimethyl-1,4,8,11-tetraaza cyclotetradecane.<sup>7</sup> Figure 3 of reference 7 shows the EXAFS spectrum for the model compound discussed. The ligand provides a square pyramidal, non-heme,  $(\text{N}_4\text{SR})_{\text{apical}}$  ligand set that is similar to that in cytochrome P450 (see Figure 7.23). Extended X-ray absorption fine structure (EXAFS) indicates one oxygen atom at  $1.70(2) \text{ \AA}$ , three nitrogen/oxygen atoms at  $2.09(2) \text{ \AA}$ , one sulfur atom at  $2.33(2) \text{ \AA}$ , and four carbons at  $2.95(2) \text{ \AA}$ . The supposed  $\text{Fe}=\text{O}$  bond distance is slightly longer than that found for other characterized  $\text{Fe}=\text{O}$  species—usually found at  $1.65\text{--}1.67 \text{ \AA}$ . A detailed discussion of cytochrome P450 and model compounds synthesized to mimic its structure or function is found in Section 7.4.

Cytochrome c and ubiquinol oxidases are part of an enzyme superfamily coupling oxidation of ferrocyanochrome c (in eukaryotes) and ubiquinol (in prokaryotes) to the  $4 \text{ e}^-/4 \text{ H}^+$  reduction of molecular oxygen. At points in the cytochrome c oxidase catalytic cycle, dioxygen ( $\text{O}_2$ ), superoxide ( $\text{O}_2^-$ ) or peroxide ( $\text{O}_2^{2-}$ ) ions coordinate to Fe(II/III) and Cu(I/II) metal ions within the metalloenzyme. Kenneth Karlin, E. I. Solomon, and co-workers have studied cytochrome c oxidase biomimetic compounds—model compounds that mimic the structure or function of the parent protein or enzyme. In one example, the group studied the geometric and electronic structure of  $[(\text{F}_8\text{TPP})\text{Fe}^{\text{III}}-(\text{O}_2^{2-})-\text{Cu}^{\text{II}}(\text{TMPA})](\text{ClO}_4)$ . This complex is prepared by the reaction of a  $[(\text{TMPA})\text{Cu}(\text{I})(\text{RCN})]^+$  complex with a  $[(\text{F}_8\text{TPP})\text{Fe}(\text{II})]$  complex. The TMPA ligand is *tris*(2-pyridylmethyl)amine,  $\text{RCN} = \text{MeCN} = \text{acetonitrile}$ , and  $\text{F}_8\text{TPP} = \text{tetrakis}(2,6\text{-difluorophenyl})\text{porphyrinate}$  (−2 charge). The model complex was characterized using Cu and Fe K-edge EXAFS spectroscopy and

density functional theory. The data were fit with a Fe–Cu distance of ~3.72 Å. Very little change in bond distances or angles were found between the solid vs. solution EXAFS results.<sup>8</sup> The results are discussed in Section 7.8.4 (see also Figure 7.44).

### 3.3 X-RAY CRYSTALLOGRAPHY

#### 3.3.1 Introduction

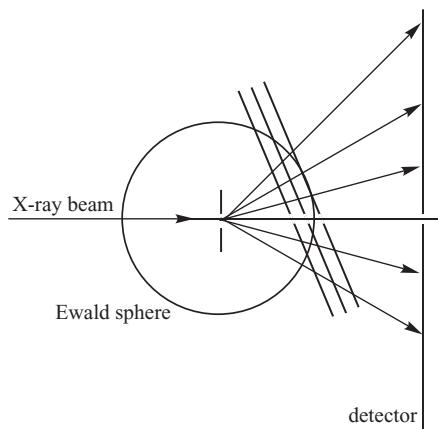
X-ray crystallographic molecular structures of proteins have been available since the 1960s and 1970s when pioneering work by Kendrew<sup>9</sup> and Perutz<sup>10</sup> produced X-ray diffraction structures of myoglobin and hemoglobin. These oxygen-carrying metalloproteins are discussed in Chapter 7. Since that time the introduction of sophisticated computer hardware and software has made the solution of protein structure in the solid state using X-ray crystallography more accurate and less time-consuming. The field continues to evolve as hardware and instrument design improvements are implemented and as crystallographers discover more powerful software algorithms for solving structures after the necessary data have been collected. At the time of this writing, 380+ X-ray crystallographic data sets were deposited in the Research Collaboratory for Structural Bioinformatics' Protein Data Bank (RCSB-PDB at <http://www.rcsb.org/pdb/>) for hemoglobin and hemoglobin as well as 245+ data sets for myoglobin and myoglobin mutant species. Nuclear magnetic resonance protein structure determination in solution provides a complementary structural technique that does not require the production of single crystals necessary for X-ray diffraction studies. However, at this time, NMR solution structures are limited to smaller proteins of molecular weights less than 30,000. In contrast, X-ray crystallography can produce structures of proteins of up to  $1 \times 10^6$  molecular weight. Recombinant DNA technology has aided the X-ray crystallographic study of proteins by allowing large amounts of a protein of interest to be produced through expression of its cloned gene in a microorganism. Site-directed mutagenesis of a selected protein's gene has allowed researchers to study three-dimensional structural changes brought about by amino acid replacement in the protein's primary amino acid sequence. These techniques are discussed in Sections 2.3.4 and 2.3.5. Much of the discussion in this section on X-ray crystallography has been taken from a recent text written by author and crystallographer Jan Drenth.<sup>11</sup> Readers are referred to the Department of Crystallography site at Würzburg University (<http://www.mineralogie.uni-wuerzburg.de/crystal/teaching/teaching.html>) for tutorial on X-ray diffraction methodology. The site includes interactive tutorials describing basic examples, reciprocal space, the crystallographic phase problem, and diffuse scattering and defect structures. Tutorials on convolution theorem, modification of a structure, solving a simple structure, anomalous scattering, and powder diffraction are also found on this site.

### 3.3.2 Crystallization and Crystal Habits

Protein structure determination using X-ray crystallography first requires the production of suitable crystals. Protein crystallization remains mainly a trial-and-error procedure complicated by impurities contaminating the selected protein, by the poorly understood process of crystal nucleation, and by other unknown factors. Protein purity requires not only that other compounds be absent but also that all molecules of the protein have the same surface properties (charge distribution) as the latter affects crystal packing of the molecules. After multistep purification procedures, the protein is dissolved in a suitable solvent (usually water–buffer mixtures for pH control) and the solution is brought to supersaturation so that nuclei for crystal growth appear. To avoid continued nucleation that will result in crystals too small for diffraction and to foster crystal growth, supersaturation must be decreased at this point. This is usually achieved by raising the temperature. The next step involves addition of a precipitant such as polyethyleneglycol (PEG), salt, or an organic solvent and perhaps adjustment of pH.

Crystallization techniques include liquid–liquid diffusion, several types of vapor diffusion, and dialysis through a membrane. Liquid–liquid diffusion of protein and precipitant solutions may conveniently take place in a melting point capillary tube using approximately 5 µl of each solution. The solution of higher density, which may be either the protein solution or the precipitant solution, is added to the capillary tube first using a syringe. The other solution is layered onto the top of the first forming a sharp boundary. Slow diffusion of the layers may produce suitable crystals, although it may be necessary to test many variations of solvent, solute concentration, and diffusing solvent before the desired crystal size and purity is attained. The most common vapor diffusion method is called the hanging drop method. Drops containing 3–10 µl of protein and precipitant solutions are placed on a siliconized (to prevent spreading of the drop) microscope glass cover slip. The slip is placed upside down over a depression in a tray. The depression is partially filled with precipitant solution. Diffusion of vapors from or to the precipitant solution may result in crystal formation given the same caveats as mentioned above for solvent diffusion. Many different membrane dialysis techniques are available. The advantage of this technique is that the protein solution is confined within a membrane allowing different precipitant solutions outside the membrane to be tried sequentially. Gilliland and Ladner<sup>12a</sup> have published a review of crystallization techniques, and Gilliland<sup>12b</sup> has assembled a database of macromolecule crystal and crystallization data. To be suitable for X-ray crystallography, a crystal is usually required to be  $\geq 0.1$  mm in its longest dimension.

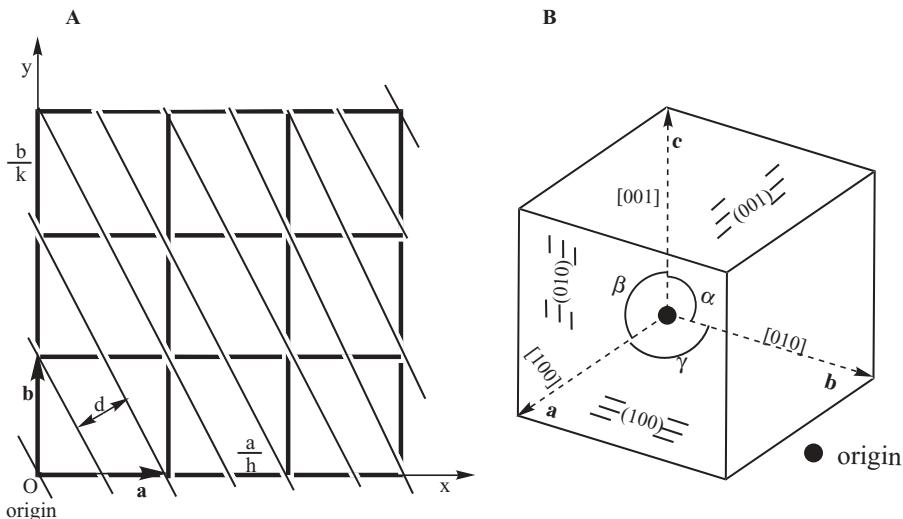
The flat faces of crystals reflect the regular packing of atoms, molecules, or ions within the crystalline structure, resulting in anisotropic (directional) behavior. Because of the regular arrangement of the crystal's atoms, molecules, or ions in planes with fixed angles and distances between the planes, single crystals will diffract X-rays. As an example, placing a crystal having a fourfold



**Figure 3.4** X-ray beam passing through the “Ewald Sphere” and diffracted by planes in a single crystal produces reflection spots. (Adapted with kind permission of Springer Science and Business Media from Figure 1.13 of reference 11. Copyright 1999, Springer-Verlag, New York.)

symmetry axis in an X-ray beam path and observing the X-ray pattern as the crystal oscillates over a small angle will produce reflection spots arranged in concentric circles. The concentric circle pattern of diffraction spots can be thought of as the intersection of a series of parallel planes with a sphere as shown in Figure 3.4. The planes represent the three-dimensional lattice that is not actually the crystal lattice but instead its reciprocal lattice. The unit distances in the lattice are reciprocally related to unit distances in the crystal, hence the name reciprocal lattice. The direction in which the X-ray beams are diffracted depends on two factors: the unit cell distances in the crystal and the X-ray wavelength. To be detected, the diffraction spots must be on, or pass through, the surface of the sphere represented by the circle in Figure 3.4. The radius of the sphere, called the sphere of reflection or the “Ewald sphere,” is reciprocal to the X-ray radiation wavelength—that is,  $1/\lambda$ . When the crystal is rotated, the reciprocal lattice rotates with it and different points within the lattice are brought to diffraction. The diffracted beams are called “reflections” because each of them can be regarded as a reflection of the primary X-ray beam against planes in the crystal.

A short discussion of crystalline habits and their important diffraction characteristics follows here; reference 11 contains much greater detail. Molecules of organic, organometallic, or inorganic materials, when precipitating from solution, attempt to reach their lowest free energy state. Frequently, this is accomplished by packing the molecules in a regular way in a crystalline habit. Regular packing of molecules in a crystal will define a unit cell through generation of three repeating vectors **a** (the *x* axis), **b**, (the *y* axis), and **c** (the *z* axis), with angles  $\alpha$ ,  $\beta$ , and  $\gamma$  between them. Planes can be constructed



**Figure 3.5** (A) Lattice planes in a two-dimensional lattice as taken from Figure 3.5 of reference 11 ( $h=2, k=1$ ). (B) Unit cell bounded by planes (100), (010), (001). Directions along **a**, **b**, and **c** are indicated by [100], [010], and [001]. (Adapted with kind permission of Springer Science and Business Media from Figure 3.6 of reference 11. Copyright 1999, Springer-Verlag, New York.)

through lattice points, and these repeat regularly throughout the crystal. Parallel planes are equidistant with distance  $d$  between them. Lattice planes cut the  $x$ ,  $y$ , and  $z$  axes into equal parts having whole numbers called indices. A set of lattice planes is determined by three indices  $h$ ,  $k$ , and  $l$  if the planes cut the  $x$  axis as  $a/h$ , the  $y$  axis as  $b/k$ , and the  $z$  axis as  $c/l$ . In Figure 3.5A, lattice planes are shown in a two-dimensional lattice. In this figure, taken from Figure 3.5 of reference 11,  $h=2$  and  $k=1$ . The lattice plane distance  $d$  is the projection of  $a/h$ ,  $b/k$ , and  $c/l$  on the line perpendicular to the corresponding lattice plane ( $h \ k \ l$ ). If a set of planes is parallel to an axis, that particular index is 0 (the plane intercepts the axis at infinity). Therefore the unit cell is bounded by planes (100), (010), and (001), with the parentheses indicating  $(h \ k \ l)$ . Line segments are given in brackets, that is, [100] is the line segment from the origin of the unit cell to the end of the  $a$  axis and [111] is the body diagonal from the origin to the opposite corner. These properties are illustrated in Figure 3.5B adapted from Figure 3.6 of reference 11.

The unit cell considered here is a primitive (P) unit cell; that is, each unit cell has one lattice point. Nonprimitive cells contain two or more lattice points per unit cell. If the unit cell is centered in the (010) planes, this cell becomes a B unit cell; for the (100) planes, an A cell; for the (001) planes, a C cell. Body-centered unit cells are designated I, and face-centered cells are called F. Regular packing of molecules into a crystal lattice often leads to symmetry

**TABLE 3.1** The Seven Crystal Systems

Crystal System	Conditions Imposed on Cell Geometry	Minimum Point Group Symmetry
Triclinic	None	1
Monoclinic	$\alpha=\gamma=90^\circ$ ( $b$ is the unique axis; for proteins this is a twofold axis or screw axis) or $\alpha=\beta=90^\circ$ ( $c$ is the unique axis; for proteins this is a twofold axis or screw axis)	2
Orthorhombic	$\alpha=\beta=\gamma=90^\circ$	222
Tetragonal	$a=b; \alpha=\beta=\gamma=90^\circ$	4
Trigonal	$a=b; \alpha=\beta=90^\circ; a=b; \gamma=120^\circ$ (hexagonal axes) or $a=b=c; \alpha=\beta=\gamma$ (rhombohedral axes)	3
Hexagonal	$a=b; \alpha=\beta=90^\circ; a=b; \gamma=120^\circ$	6
Cubic	$a=b=c; \alpha=\beta=\gamma=90^\circ$	23

Source: Adapted from reference 13.

relationships between the molecules. Common symmetry operations are two- or threefold screw (rotation) axes, mirror planes, inversion centers (centers of symmetry), and rotation followed by inversion. There are 230 different ways to combine allowed symmetry operations in a crystal leading to 230 *space groups*.<sup>13</sup> Not all of these are allowed for protein crystals because of amino acid asymmetry (only L-amino acids are found in naturally occurring proteins). Only those space groups without symmetry (triclinic) or with rotation or screw axes are allowed. However, mirror lines and inversion centers may occur in protein structures along an axis.

Seven crystal systems as described in Table 3.1 occur in the 32 point groups that can be assigned to protein crystals. For crystals with symmetry higher than triclinic, particles within the cell are repeated as a consequence of symmetry operations. The number of asymmetric units within the unit cell is related but not necessarily equal to the number of molecules in a unit cell, depending on how the molecules are related by symmetry operations. From the symmetry in the X-ray diffraction pattern and the systematic absence of specific reflections in the pattern, it is possible to deduce the space group to which the crystal belongs.

In summary, it is important to determine crystal quality, unit cell dimensions of the crystal (a larger crystal absorbs X rays more strongly, 0.3–0.5 mm is considered the optimal size), the crystal's space group, and how many protein molecules are in the unit cell and in one asymmetric unit. Actually, the great majority of crystals useable for X-ray crystallography are not ideal but contain lattice defects. This is true for protein crystals, which are also weak scatterers since the great majority of the component atoms are light atoms, C, N, and O.

One deduces the space group from the symmetry in the crystal's diffraction pattern and the systematic absence of specific reflections in that pattern. The crystal's cell dimensions are derived from the diffraction pattern for the crystal collected on X-ray film or measured with a diffractometer. An estimation of  $Z$  (the number of molecules per unit cell) can be carried out using a method called  $V_M$  proposed by Matthews.<sup>14</sup> For most protein crystals the ratio of the unit cell volume and the molecular weight is a value around  $2.15 \text{ \AA}^3/\text{Da}$ . Calculation of  $Z$  by this method must yield a number of molecules per unit cell that is in agreement with the decided-upon space group.

### 3.3.3 Theory and Hardware

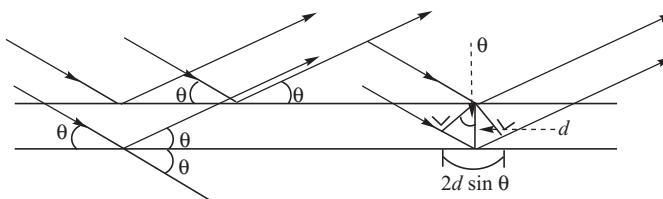
The mathematics necessary to understand the diffraction of X rays by a crystal will not be discussed in any detail here. Chapter 4 of reference 11 contains an excellent discussion. The arrangement of unit cells in a crystal in a periodic manner leads to the Laue diffraction conditions shown in equations 3.5, where vectors  $\mathbf{a}$ ,  $\mathbf{b}$ , and  $\mathbf{c}$  as well as lattice indices  $h$ ,  $k$ , and  $l$  have been defined in Figure 3.5 and  $\mathbf{S}$  is a vector quantity equal to the difference between the resultant vector  $\mathbf{s}$  after diffraction and the incident X-ray beam wave vector  $\mathbf{s}_0$ , so that  $\mathbf{S} = \mathbf{s} - \mathbf{s}_0$ .

$$\begin{aligned}\mathbf{a} \cdot \mathbf{S} &= h \\ \mathbf{b} \cdot \mathbf{S} &= k \\ \mathbf{c} \cdot \mathbf{S} &= l\end{aligned}\tag{3.5}$$

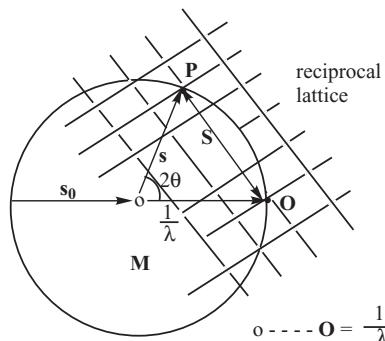
The same crystalline arrangement leads to the expression of Bragg's law applied to X-ray diffraction with incident X-ray beam of wavelength  $\lambda$  as shown in equation 3.6 and where the terms are defined as in Figures 3.5 and 3.6.

$$\lambda = 2d \sin \theta\tag{3.6}$$

As mentioned above, the formalism of the reciprocal lattice is convenient for constructing the directions of diffraction by a crystal. In Figure 3.4 the Ewald sphere was introduced. The radius of the Ewald sphere, also called the



**Figure 3.6** Two lattice planes separated by distance  $d$ . Incident and reflected X-ray beams make the angle  $\theta$  with the lattice planes. (Adapted with kind permission of Springer Science and Business Media from Figure 4.17 of reference 11. Copyright 1999, Springer-Verlag, New York.)



**Figure 3.7** The Ewald sphere used to construct the direction of the scattered beam. The sphere has radius  $1/\lambda$ . The origin of the reciprocal lattice is  $\mathbf{O}$ . The incident X-ray beam is labeled  $\mathbf{s}_0$  and the scattered beam is labeled  $\mathbf{s}$ . (Adapted with kind permission of Springer Science and Business Media from Figure 4.19 of reference 11. Copyright 1999, Springer-Verlag, New York.)

sphere of reflection, is reciprocal to the wavelength of X-ray radiation—that is,  $1/\lambda$ . The reciprocal lattice rotates exactly as the crystal. The direction of the beam diffracted from the crystal is parallel to  $\mathbf{MP}$  in Figure 3.7 and corresponds to the orientation of the reciprocal lattice. The reciprocal space vector  $\mathbf{S}(h\ k\ l) = \mathbf{OP}$  ( $h\ k\ l$ ) is perpendicular to the reflecting plane  $h\ k\ l$ , as defined for the vector  $\mathbf{S}$ . This leads to the fulfillment of Bragg's law as  $|\mathbf{S}(h\ k\ l)| = 2(\sin\theta)/\lambda = 1/d$ .

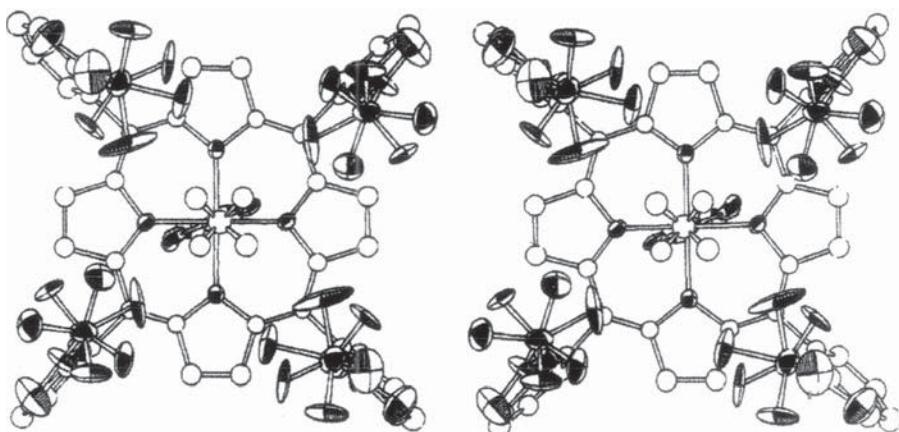
Molecules and atoms within a crystal are not in static positions but vibrate around an equilibrium position. Atoms around the periphery of a molecule will vibrate to a greater extent, whereas central atoms will have relatively fixed positions. The resultant weakening of X-ray beam intensity, especially at high scattering angles, is expressed as the *temperature factor*. In the simplest case, components of the atom's vibration are all in the same direction, called the isotropic case. The component perpendicular to the reflecting plane and thus along  $\mathbf{S}$  is equal for each  $(h\ k\ l)$ , and the temperature correction factor for isotropic atomic scattering is given by equation 3.7:

$$T_{\text{(iso)}} = \exp\left[-B \frac{\sin^2 \theta}{\lambda^2}\right] = \exp\left[-\frac{B}{4} \left(\frac{2 \sin \theta}{\lambda}\right)^2\right] = \exp\left[-\frac{B}{4} \left(\frac{1}{d}\right)^2\right] \quad (3.7)$$

Assuming isotropic and harmonic vibration the thermal parameter  $B$  becomes the quantity shown in equation 3.8, where  $\bar{u}^2$  is the mean square displacement of the atomic vibration.

$$B = 8\pi^2 \times \bar{u}^2 \quad (3.8)$$

For anisotropic vibration the temperature factor is more complex because  $\bar{u}^2$  now depends on the direction of  $\mathbf{S}$ . The anisotropic temperature factor is often



**Figure 3.8** Ellipsoids of vibration for a small molecule—an ORTEP view. Stereodigram of  $[(\text{Fe}(\text{O}_2)(\text{T}_{\text{piv}}\text{PP})(1\text{-MeIm})]$  looking down the  $\text{C}_2$  twofold axis of symmetry. (Reprinted with permission from Figure 6 of reference 16. Copyright 1978, American Chemical Society.)

displayed in the form of an ellipsoid of vibration, commonly so that the vibrating atom has a 50% chance of being within the ellipsoid.

This is illustrated in Figure 3.8 for a hemoglobin model compound, featuring one of a large group of so-called “picket fence porphyrin” ligand systems, synthesized in the James P. Collman laboratory in the late 1970s. More detail on hemoglobin model compounds will be found in Sections 7.2.4 through 7.2.7; here we will concentrate on the X-ray diffraction study of this important hemoglobin model. The researchers began with crystals of (2-methylimidazole)-*meso*-tetrakis( $\alpha,\alpha,\alpha,\alpha$ -*o*-pivalamidophenyl)porphyrinatoiron(II)-ethanol,  $[(\text{Fe}(\text{T}_{\text{piv}}\text{PP})(2\text{-MeIm}))\text{-EtOH}]$  (see Figures 7.9B and 7.10 for structures). They exposed the  $[(\text{Fe}(\text{T}_{\text{piv}}\text{PP})(2\text{-MeIm}))\text{-EtOH}]$  crystals to dioxygen, and then they determined the structure of the starting complex as well as the resulting dioxygen complex using X-ray diffraction.<sup>15</sup> The deoxy starting complex was found to have a  $\text{C}_2$  molecular symmetry with space group  $\text{C}_{2h}^6\text{-C2/c}$ ,  $Z=4$  (four molecules in the unit cell), and a unit cell with the following dimensions:  $\mathbf{a}=18.871(11)\text{\AA}$ ,  $\mathbf{b}=19.425(13)\text{\AA}$ ,  $\mathbf{c}=18.434(11)\text{\AA}$ ,  $\beta=91.48^\circ$ , and volume,  $V=6755.0\text{\AA}^3$ . The oxy complex was nearly isomorphous in space group  $\text{C}_{2h}^6\text{-C2/c}$ ,  $Z=4$ , and had very similar cell dimensions. The structures were refined by a full-matrix, least-squares methods, using 4176 and 5183 reflections for the deoxy and oxy complexes, respectively. Selected bond distances and angles for the oxy complex  $[(\text{Fe}(\text{O}_2)(\text{T}_{\text{piv}}\text{PP})(2\text{-MeIm}))\text{-EtOH}]$  are:  $\text{Fe}-\text{N}_{\text{porphyrin}}=1.997(4)$ ,  $1.995(4)\text{\AA}$ ;  $\text{Fe}-\text{N}_{\text{Im}}=2.107(4)\text{\AA}$ ;  $\text{Fe}-\text{O}=1.898(7)\text{\AA}$ ;  $\text{O}-\text{O}=1.22(2)\text{\AA}$ ;  $\text{Fe}-\text{O}-\text{O}$  angle =  $129(1)^\circ$ . The iron ion is displaced  $0.399\text{\AA}$  from the plane of the porphinato nitrogen atoms toward the imidazole ligand in the deoxy complex, but only  $0.086\text{\AA}$  in the oxy complex.

In the presence of ethanol, the deoxy complex was found to bind dioxygen reversibly, noncooperatively, and with lower affinity than when the sample was oxygenated after solvent removal. In the latter case, dioxygen uptake was found to be cooperative. The complex  $[(\text{Fe}(\text{T}_{\text{piv}}\text{PP})(2\text{-MeIm})]$  is a model for T (tense) state deoxy- and oxyhemoglobin (as discussed in Sections 7.2.2, 7.2.4, and 7.2.6; also see Figures 7.7, 7.9B, and 7.10), whereas its 1-methylimidazole analogue,  $[(\text{Fe}(\text{T}_{\text{piv}}\text{PP})(1\text{-MeIm})]$ , is a model for the R (relaxed) state. In comparing these two model compounds, the sterically active 2-methylimidazole substituent interacts with the porphyrin ring, whereas the 1-methylimidazole does not interact (or behaves in a similar manner to the normal hemoglobin ligand, histidine). If the X-ray crystallographic structures  $[(\text{Fe}(\text{O}_2)(\text{T}_{\text{piv}}\text{PP})(2\text{-MeIm})]$  and  $[(\text{Fe}(\text{O}_2)(\text{T}_{\text{piv}}\text{PP})(1\text{-MeIm})]$  are compared, the total distance of the Fe–O plus Fe–N<sub>fm</sub> bonds is  $>4.00 \text{ \AA}$  for  $[(\text{Fe}(\text{O}_2)(\text{T}_{\text{piv}}\text{PP})(2\text{-MeIm})]$  and  $3.82 \text{ \AA}$  for  $[(\text{Fe}(\text{O}_2)(\text{T}_{\text{piv}}\text{PP})(1\text{-MeIm})]$ .<sup>16</sup> Most of the difference ( $>0.15 \text{ \AA}$ ) is in the Fe–O bond length of  $[(\text{Fe}(\text{O}_2)(\text{T}_{\text{piv}}\text{PP})(2\text{-MeIm})]$ , leading the researchers to conclude that the sterically active 2-MeIm perturbs (lengthens) the Fe–O bond more than the Fe–N<sub>fm</sub> bond. This conclusion has significance for  $[(\text{Fe}(\text{O}_2)(\text{T}_{\text{piv}}\text{PP})(2\text{-MeIm})]$  as a model for hemoglobin's T state, which has a smaller O<sub>2</sub> affinity than the R state.

Figure 3.8 displays a so-called ORTEP view of  $[(\text{Fe}(\text{O}_2)(\text{T}_{\text{piv}}\text{PP})(1\text{-MeIm})]$ , drawn using 30% probability ellipsoids. Note that the outer atoms show greater uncertainty in position, as seen in their elliptical rather than spherical shape. The figure illustrates a stereodiagram, which will appear three-dimensional when the two halves of the figure are merged by the observer. Techniques for visualizing stereodiagrams are found in Section 4.6.1.

For proteins, X-ray structures usually are not determined at high enough resolution to use anisotropic temperature factors. Average values for  $B$  in protein structures range from as low as a few  $\text{\AA}^2$  for well-ordered structures to  $30 \text{ \AA}^2$  for structures involving flexible surface loops. Using equation 3.8, one can calculate the root mean square displacement  $\sqrt{\bar{u}^2}$  for a well-ordered protein structure at approximately  $0.25 \text{ \AA}$  (for  $B=5 \text{ \AA}^2$ ) and for a not-so-well-ordered structure at  $0.62 \text{ \AA}$  (for  $B=30 \text{ \AA}^2$ ). These seemingly small errors in atomic positions of C, N, and O atoms derive from the fact that the bond distances and angles for individual amino acids in small compounds are well known, and it is assumed that these do not change when the amino acids are incorporated into large protein molecules. In fact, the limited resolution of a protein X-ray diffraction pattern does **not** permit calculation of an electron density map at atomic resolution, although amino acid residues can be distinguished from differences in their side chains. Usually these are displayed in stereo diagrams such as seen in Figure 3.8.

In addition to the dynamic disorder caused by temperature-dependent vibration of atoms, protein crystals have static disorder due to the fact that molecules, or parts of molecules, do not occupy exactly the same position or do not have exactly the same orientation in the crystal unit cell. However, unless data are collected at different temperatures, one cannot distinguish between

dynamic and static disorder. Because of protein crystal disorder, the diffraction pattern fades away at some diffraction angle  $\theta_{\max}$ . The corresponding lattice distance  $d_{\min}$  is determined by Bragg's law as shown in equation 3.9:

$$d_{\min} = \frac{\lambda}{2 \sin \theta_{\max}} \quad (3.9)$$

The end result of an X-ray structural determination reports the electron density in the crystal. The fundamental equation for its calculation follows in equation 3.10:

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)| \exp[-2\pi i(hx + ky + lz) + i\alpha(hkl)] \quad (3.10)$$

The electron density at every position  $x, y, z$  in the unit cell,  $\rho(x y z)$ , is the *Fourier transform* of the structure factor  $\mathbf{F}(h k l)$ , which is in turn a function of the electron density distribution in the unit cell and the integrated intensity of the reflected beam, called  $I(h k l)$ . Values of  $I(h k l)$ , the integrated reflected beam intensity, are obtained from the diffraction pattern after the application of correction factors. The term  $|F(h k l)|$  is the structure factor amplitude of reflection  $(h k l)$  including the temperature factor,  $\alpha(h k l)$  is the phase angle, and  $x, y$ , and  $z$  are coordinates in the unit cell. The factor  $i$  is the mathematical imaginary term. The summation occurs over all the discrete directions in which diffraction by the crystal occurs. Sophisticated computer programs have been designed to calculate the electron density in the above equation except for the phase angles  $\alpha(h k l)$ , which cannot be derived in a straightforward manner from the diffraction pattern. The “phase angle” problem arises because the X-ray diffraction pattern registers only the intensity of the waves. To work back to a molecular structure, one needs to know the relative timing when each wave hits. This is the “phase data”—the position of wave crests and troughs relative to each other. Several methods have been developed to solve the phase angle problem as listed below.

1. The isomorphous replacement method requires attachment of heavy atoms to protein molecules in the crystal. In this method, atoms of high atomic number are attached to the protein, and the coordinates of these heavy atoms in the unit cell are determined. The X-ray diffraction pattern of both the native protein and its heavy atom derivative(s) are determined. Application of the so-called *Patterson* function determines the heavy atom coordinates. Following the refinement of heavy atom parameters, the calculation of protein phase angles proceeds. In the final step the electron density of the protein is calculated.
2. The multiple-wavelength anomalous diffraction method (MAD) relies upon sufficiently strong anomalously scattering atoms in the protein structure itself. In this method, diffraction data must be collected at a number of different wavelengths, usually requiring data collection with

synchrotron radiation. Anomalous scattering occurs when the electrons in atoms can no longer be considered free electrons, but rather electrons affected by the X rays themselves. The effect is negligible for light atoms C, N, and O but becomes useful for heavier atoms (S onwards). Anomalous scattering can be exploited for protein phase angle determination by the multiple wavelength anomalous dispersion (MAD) method. Determination of the absolute configuration of a protein structure is also possible.

3. The molecular replacement method assumes similarity of the unknown structure to a known one. This is the most rapid method but requires the availability of a homologous (similar) protein's structure. The method relies on the observation that proteins that are similar in their amino acid sequence (homologous) will have very similar folding of their polypeptide chains. This method also relies on the use of *Patterson* functions. As the number of protein structure determinations has been increasing rapidly, the molecular replacement method becomes extremely useful for determining protein phase angles.
4. The so-called direct methods rely on the principles that phase information is included in the intensities, that electron density is always positive, and that the crystal contains atoms that are or may be considered equal. Phase relations based on probability theory have been formulated and applied to clusters of reflections. Direct methods are standard techniques for determining phase angles in smaller molecules (<100 atoms). Recent advances in theory by H. A. Hauptman of the Hauptman-Woodward Medical Research Institute and Department of Structural Biology, School of Medicine and Biomedical Sciences, State University of New York at Buffalo have allowed the structures of small proteins to be solved using the so-called Shake and Bake statistical method. In essence, the Shake and Bake method executes phase angle determination in small protein molecules by applying it to the strongest reflections only. More information on this developing technique can be found in the reference 18 articles and on the website <http://www.psc.edu/science/Hauptman/Hauptman.html>.

A preliminary structural model of a protein is arrived at using one of the methods described above. Calculated structure factors,  $|F_{\text{calc}}|$ , based on the model generally are in poor agreement with the observed structure factors  $|F_{\text{obs}}|$ . The agreement is represented by an *R*-factor defined as found in equation 3.11, where  $k$  is a scale factor.

$$R = \frac{\sum_{hkl} |F_{\text{obs}}| - k |F_{\text{calc}}|}{\sum_{hkl} |F_{\text{obs}}|} \times 100 \quad (3.11)$$

Refinement takes place by adjusting the model to find closer agreement between the calculated and observed structure factors. For proteins the

refinements can yield *R*-factors in the range of 10–20%. An example taken from reference 11 is instructive. In a refinement of a papain crystal at 1.65-Å resolution, 25,000 independent X-ray reflections were measured. Parameters to be refined were the positional parameters (*x*, *y*, and *z*) and one isotropic temperature factor parameter (*B*) for each of the 2000 nonhydrogen atoms in the molecule. Four times 2000 yields 8000 parameters. With 25,000 measurements the ratio of observations to parameters is slightly more than 3, a poor (low) overdetermination. Incorporating bond length and angle data from the known dimensions of small molecules increases the number of “observations.” Another technique, called “solvent flattening,” which adjusts for disordered solvent molecules in the channels between protein molecules, is imposed. In all cases, because protein structures are large and complex, their refinement is a large project computationally requiring fast computers and application of fast Fourier transform methods. Application of refinement to the papain structure would probably result in an *R*-factor of about 16% (usually reported as a decimal, i.e.,  $R=0.16$ ). Application of refinement methods cannot completely rule out wrong interpretations of electron density maps and consequent entirely or partially incorrect protein structures. A number of checks for gross errors are available including stereochemistry checks for detailed analysis of protein geometry. One of these, PROCHECK,<sup>17a,b</sup> information at the website <http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html>, conducts a detailed analysis of all geometric aspects of proteins using data extracted from the Cambridge Crystallographic Data Centre (<http://www.ccdc.cam.ac.uk>) for bond lengths, angles, and planarity in peptide structures. For torsion angles, PROCHECK uses comparison data from high-resolution protein structures in the Protein Data Bank (PDB; <http://www.rcsb.org/pdb/home/home.do>).

The hardware necessary for collection of X-ray diffraction data includes an X-ray source and X-ray detector. Most commonly in research laboratories, the X-ray radiation is emitted from a copper source, Cu K $\alpha$  radiation, with a wavelength of 1.5418 Å. X-ray sources include sealed X-ray tubes and the more powerful, more usual system for small proteins, a tube with a rotating anode—that is, a rotating anode generator. Protein crystallography benefits from use of the rotating anode tube because the X rays emitted are of higher intensity. The rotating anode generator requires maintenance of high vacuum.

The most powerful X-ray source, a particle accelerator such as a synchrotron, also called a storage ring, requires the construction of a large and expensive facility. Extremely high-intensity, collimated synchrotron radiation is of value when collecting data from weakly diffracting crystals as well as proteins consisting of more than 150–200 amino acid residues. A synchrotron is essentially an accelerator that takes stationary charged particles, such as electrons, and drives them to velocities near the speed of light. Strong magnets arranged in particular arrays force electrons to travel around a circular storage ring, tangentially emitting electromagnetic radiation and, consequently, losing energy. This energy is emitted in the form of light and is known as synchrotron

radiation. Synchrotron radiation is tunable to wavelengths at or below 1 Å, an advantage for the multiple-wavelength anomalous diffraction (MAD) method for solving protein structures. A crystal may diffract shorter-wavelength radiation more weakly, but the crystal suffers less radiation damage because it absorbs less radiation at the shorter wavelength. A growing number of synchrotron radiation facilities exist that are available to researchers wishing to collect data and solve the structures of large proteins. A number of facilities in the United States are associated with universities or national laboratories. For instance, the Brookhaven National Laboratory (BNL) maintains multiple synchrotron beamlines as do the Stanford Synchrotron Radiation Laboratory (SSRL) and the Cornell High Energy Synchrotron Source. The European Synchrotron Radiation Facility (ESRF) serves researchers internationally. The Harima Institute's Riken SPring-8 Center has multiple beamlines serving Japanese researchers. The Protein Crystallography Research Resource at BNL (<http://www.px.nsls.bnl.gov/>) was running nine beamlines in late 2006, most of which were carrying out MAD-based protein crystallography. MAD is an acronym for multiple wavelength anomalous dispersion, as discussed previously.

The next-generation X-ray sources are called X-ray free-electron lasers (XFELs). XFELs promise X-ray beam lines that are vastly brighter, coherent (which synchrotron radiation is not), capable of delivering higher energy and shorter pulses than the synchrotron radiation used presently. Pulses as short as 100 femtoseconds could be capable of producing movies of chemical reactions—bond breaking and bond making—in progress. It is expected that XFELS could determine structures of proteins and other biological macromolecules from noncrystalline materials, a great advance over the present technology. The Linac Coherent Light Source (LCLS) at the Stanford University Linear Accelerator Center will be the world's first X-ray free electron laser (XFEL) facility when it becomes operational in 2009 (see the website <http://www-srsl.slac.stanford.edu/lcls/> for more information and updates).

X-ray detectors also come in several varieties: (1) single-photon counters which yield accurate results but require up to several weeks to acquire the 10,000–100,000 ( $10^4$ – $10^5$ ) reflections necessary to compile a complete data set for a protein crystal; (2) image plates that operate much like photographic film but are 10 times more sensitive; (3) area detectors, electronic devices that detect X-ray photons on a two-dimensional surface. Both fluorescent-type detectors, image plates and fast area detectors, are more sensitive at the shorter wavelengths of X-ray radiation from synchrotron sources.

### 3.3.4 Descriptive Examples

The following example of a calcium-ion-containing enzyme that hydrolyzes adenosine triphosphate (ATP) is taken from Section 6.4.2, where much more detail on the enzyme will be found. In general,  $\text{Ca}^{2+}$ -ATPases, especially those found in the sarco(endo)plasmic reticulum (SR), are P-type ATPases that play

a major role in muscle contraction–relaxation cycles and are responsible for transporting calcium into the lumen of the sarcoplasmic reticulum. The  $\text{Ca}^{2+}$ -ATPase from fast-twitch rabbit skeletal muscle sarcoplasmic reticulum (SERCA1a) is a large membrane protein of 110 kDa. The enzyme actively transports two  $\text{Ca}^{2+}$  ions per ATP hydrolyzed across the membrane from the cytoplasm to the lumen against a concentration gradient. At the same time, two to three protons are transported in the opposite direction. Activity of  $\text{Ca}^{2+}$ -ATPase is crucial to skeletal muscle function. In 2000, Japanese researchers determined the first atomic-scale structural model for any P-type ATP-dependent pump using crystals formed from the  $\text{Ca}^{2+}$ -ATPase SERCA1a protein.<sup>19</sup> The data are deposited in the Protein Data Bank (PDB) with the accession number 1SU4. Since then, various X-ray crystallographic structures corresponding to stages in the  $\text{Ca}^{2+}$ -ATPases catalytic cycle (see Figures 6.28–6.30) have been determined. The text that follows here is adapted from the Methods section of reference 19.

$\text{Ca}^{2+}$ -ATPase from rabbit hind leg muscle was purified, and crystals were grown by dialyzing the mixture of purified protein and phosphatidylcholine (a phospholipid) against a crystallization buffer. After one month, crystals of dimensions  $100 \times 500 \times 20 \mu\text{m}$  were obtained. The crystals belonged to the  $C_2$  space group with unit cell dimensions of  $a = 166 \text{ \AA}$ ,  $b = 64.4 \text{ \AA}$ ,  $c = 147.1 \text{ \AA}$ ,  $\beta = 98.0^\circ$  and contained one protein molecule in the asymmetric unit. The crystals used for crystallography were flash-frozen in cold nitrogen gas. Heavy atom derivatives were prepared by soaking native crystals in platinum, terbium, or lanthanum compounds or salts for one hour. After crystal screening with a laboratory X-ray source, the final data sets were collected at SPring-8 beamlines BL41XU ( $\lambda = 0.800 \text{ \AA}$ ) or BL44B2 ( $\lambda = 0.890$  and  $1.000 \text{ \AA}$ ) for native crystals and BL44B2 ( $\lambda = 0.890 \text{ \AA}$ ) for derivative crystals. Data were collected on eight native crystals and crystal derivatives obtained by soaking other native crystals in heavy metal compounds and salts. Table 1 of reference 19 contains a summary of the crystallographic analyses. Data were detected using either Rigaku R-AXIS IV imaging plate or Mar CCD 165 detectors. Charge coupled device (CCD) area detectors have faster readout times and better resolution than imaging plate detectors. Data taken with the Rigaku R-AXIS IV imaging plate detector were processed with Denzo and Scalepack (see reference 17d) software programs used to visualize and interpret X-ray crystallographic data. Data taken with the Mar CCD 165 detector were processed with MOSFLM and Scala, also software programs used to visualize and interpret X-ray crystallographic data. Merging of different data sets was carried out using Scalepack, subsequent processing using CCP4 program suites, available through the Collaborative Computational Project, Number 4. 1994 (The CCP4 Suite: Programs for Protein Crystallography, *Acta Crystallogr. D* **50**, 760–763). References to information on all these software programs are found in reference 19 and online. First, experimental phases were calculated using Miphare (part of the CCP4 program suites) to 2.8- $\text{\AA}$  resolution. The electron density map refined by density modification is shown in Figures 3a and 3b of

reference 19. A solvent-flattened map (solvent flattening techniques improve electron density maps) was calculated at 2.8-Å resolution and contoured at  $1.2\sigma$ , and a  $2|F_0| - |F_c|$  composite OMIT map was prepared with CNS. Its website at <http://cns.csby.yale.edu/v1.1/> states that the Crystallography & NMR System (CNS) is the result of an international collaborative effort among several research groups. The program has been designed to provide a flexible multilevel hierarchical approach for the most commonly used algorithms in macromolecular structure determination. For the  $\text{Ca}^{2+}$ -ATPase crystallographic study, the refined temperature factors were  $60.0 \text{ \AA}^2$  for protein atoms,  $48.7 \text{ \AA}^2$  for bound water molecules, and  $41.3 \text{ \AA}^2$  for  $\text{Ca}^{2+}$  ions. The assignment of secondary structure was carried out with the program DSSP. The DSSP program was designed by Wolfgang Kabsch and Chris Sander to standardize secondary structure assignment. The DSSP database is a database of secondary structure assignments (and much more) for all protein entries in the Protein Data Bank (PDB). Some refinement statistics taken from reference 19, Table 1 are the following: resolution range,  $15$ – $2.6 \text{ \AA}$ ; number of reflections; 48,373; number of protein atoms, 7673; number of water molecules, 276; number of calcium ions, 2. The calculated  $R_{\text{cryst}}$  factor (see equation 3.11) was 25.0%. Using a test set comprised of 10% of the data for refinement resulted in an  $R_{\text{free}}$  of 30.7%.

Mark Wilson and Axel Brunger published the 1.0-Å crystal structure of  $\text{Ca}^{2+}$ -bound calmodulin (PDB: 1EXR) in 2000.<sup>20</sup> Calmodulin (CaM) is a small acidic protein of 148 amino acid residues weighing approximately 17 kDa. CaM is present in all eukaryotic cells, serves as a primary receptor for intracellular  $\text{Ca}^{2+}$ , and is considered to be a switch in cellular processes controlled by  $\text{Ca}^{2+}$ -dependent signaling pathways. Calmodulin will be discussed in more detail in Section 6.3.2. Reference 20 presents a thorough X-ray crystallographic study of this small protein (about 1/10 the size of the  $\text{Ca}^{2+}$ -ATPase discussed previously in this section). The X-ray crystallographic data on the calmodulin crystals were collected at the Stanford Synchrotron Radiation Laboratory (SSRL). Essentially, Wilson and Brunger followed the same processes for collecting and analyzing data as described above for  $\text{Ca}^{2+}$ -ATPase. Their detailed explanation of the data collection and multiple-stage refinement of the calmodulin crystallographic model in reference 20 will be of interest to those readers wishing to learn more about these techniques.

It is obvious from the foregoing description that solving the structure of large proteins and other macromolecules requires technologically advanced hardware and software. The National Institutes of Health are supporting the Protein Structure Initiative, PSI (as discussed in Section 2.2.2). The mission of the PSI is to make the three-dimensional, atomic-level structures of most proteins easily available from knowledge of their amino acid sequences. To this end, NIH is supporting the development of tools to accelerate experimental structure prediction, whether by new technology in X-ray crystallography or nuclear magnetic resonance (NMR). NMR will be discussed in Section 3.4.

## 3.4 NUCLEAR MAGNETIC RESONANCE

### 3.4.1 Theoretical Aspects

While chemists are usually concerned with the behavior of electrons orbiting the nucleus, the instrumental method called nuclear magnetic resonance (NMR) is a technique based on the properties of the atom's nucleus. Interaction of the atom's electrons with the nucleus yields important structural and chemical information that can be gathered using NMR techniques. Nuclear magnetic resonance spectroscopy can provide information on structure, dynamics, kinetics, binding processes, electronic structure, and magnetic properties of bioinorganic molecules in solution. Basic information about the NMR technique as summarized here is taken from reference 21.

Nuclei of natural isotopes (atoms of chemical elements differing in the number of neutrons in their nuclei) may possess angular momentum or spin and therefore magnetic moments. One defines spin by the following equation:

$$\text{spin} = \hbar[I(I+1)]^{1/2} \quad (3.12)$$

where  $\hbar = h/2\pi$  and  $I = \text{nuclear angular momentum} = 0, 1/2, 1, 3/2, \dots$

The nuclear angular momentum,  $I$ , is quantized with magnitude  $\hbar m$ , where  $m = I, I-1, I-2, \dots, -I$ , yielding  $2I+1$  equally spaced spin states with angular quantum numbers  $I$ . Additionally, a nucleus with spin has a magnetic moment,  $\mu$ . Components of  $\mu$  having different spin states  $\mu m/I$  yield  $2I+1$  components for  $\mu$ . Application of an external magnetic field splits spin states into different potential energy states, accounting for the origin of the nuclear magnetic resonance phenomenon.

The angular momentum and magnetic moment of nuclei act as parallel or antiparallel vectors, and the ratio between these is known as the magnetogyric ratio,  $\gamma$ . The magnetogyric ratio is given in terms of equation 3.13:

$$\gamma = \frac{2\pi \mu}{h I} = \frac{\mu}{I\hbar} \quad (3.13)$$

For  $I > 1/2$ , nuclei also possess an electric quadrupole moment,  $Q$ . Quadrupolar nuclei exchange energy with electric fields in the rest of the molecule in which they are located causing profound effects on NMR spectra. Table 1.1 of reference 21 lists the nuclear properties of some elements. One property of nuclei, receptivity, or natural signal strength, depends on the intrinsic sensitivity of the nucleus weighted by its natural abundance. Usually the receptivity of a nucleus is high if the magnetic moment,  $\mu$ , is high. Usually, the  $^{13}\text{C}$  nucleus, which has a fairly weak signal and natural abundance of 1.108%, is arbitrarily given a receptivity of 1.00 with the receptivity of all other nuclei calculated relative to it. The resonant frequency in a particular magnetic field (2.348T in reference 21, Table 1.1) is proportional to the magnetogyric ratio and will vary slightly according to the chemical and electronic environment in which the

nucleus finds itself. This variation results in the “chemical shift effect” (to be explained further below), one of the most useful results of nuclear magnetic resonance spectroscopy.

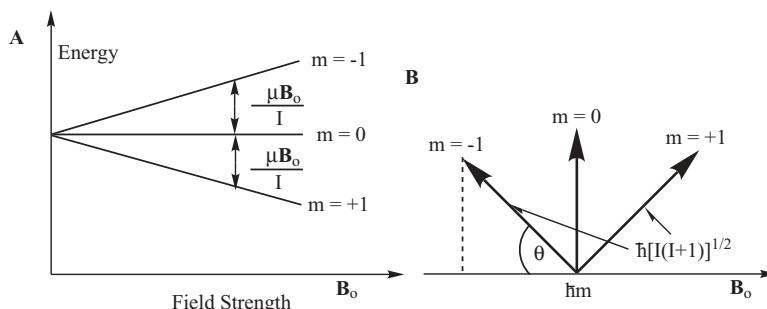
Nuclei possessing even numbers of both protons and neutrons have angular momentums equal to 0 ( $I=0$ ) and are magnetically inactive. Two major isotopes in organic and biological species,  $^{12}\text{C}$  and  $^{16}\text{O}$ , belong to this group, a fact that simplifies the NMR spectra of molecules containing them. Odd numbers of either (or both) protons or neutrons lead to nonzero spins, with the actual number of spin states being dependent on the number of possible arrangements. Because  $s$  bonding electrons have charge density within the nucleus and will perturb nuclear spin states, ultimately the NMR technique leads to information about molecular arrangements surrounding the nucleus under study. This is the information being sought in the NMR experiment.

Placing the nucleus with  $I=1$  and a positive magnetogyric ratio in a particular magnetic field  $\mathbf{B}_0$  (aligned along a  $z$  axis) leads to  $2I+1$  arrangements at differing potential energies split by  $(\mu\mu/\mathbf{I}) \mathbf{B}_0$ , as shown in Figure 3.9A.

Alignment of vectors relative to  $\mathbf{B}_0$  for each value of  $m$  results in vectors of length  $\hbar[I(I+1)]^{1/2}$  and a  $z$  component  $\hbar[I(I+1)]^{1/2}$ , where  $\cos\theta=m[I(I+1)]^{1/2}$  as shown Figure 3.9B. Differing energy states lead to interaction with electromagnetic radiation of the correct frequency according to  $\Delta E=h\nu$ . For nuclear magnetic resonance observations, equation 3.14 must hold because the selection rule for energy transitions between adjacent states ( $\Delta m=\pm 1$ ) operates and in terms of the magnetogyric ratio, one finds equation 3.15.

$$h\nu = \mu \frac{\mathbf{B}_0}{\mathbf{I}} \quad (3.14)$$

$$\gamma = \frac{2\pi \mu}{h} \frac{\mathbf{I}}{\mathbf{B}_0} \quad (3.15)$$



**Figure 3.9** (A) Nuclear spin energy for a nucleus having  $I=1$  (e.g.,  $^{14}\text{N}$ ) plotted as a function of  $\mathbf{B}_0$ . (B) Nuclear vectors relative to  $\mathbf{B}_0$  with vector length  $\hbar[I(I+1)]^{1/2}$  and  $z$  component  $\hbar m$  so that  $\cos\theta=m[I(I+1)]^{1/2}$ . (Adapted with permission of Nelson Thornes Ltd. from Figures 1a and 1b of reference 21.)

Substitution and rearrangement lead to equation 3.16:

$$v = \frac{\gamma}{2\pi} \mathbf{B}_0 \quad (3.16)$$

Values of the magnetogyric ratio,  $\gamma$ , are such that frequencies lie within the radio range (frequencies,  $v = < 1000$  MHz; see Figure 3.2).

The low-energy, low-frequency range for NMR transitions corresponds to a small change in energy,  $\Delta E$ . This has implications for the population of excited states, the Boltzmann distribution. For a spin-1/2 nuclei with  $\Delta E = \mu \mathbf{B}_0/I$  and  $I = 1/2$ , equation 3.17 applies. Since  $N^+ \approx N^-$ , one can write equation 3.18.

$$\frac{N^+}{N^-} = \exp\left(-\frac{\Delta E}{kT}\right) = \exp\left(-\frac{2\mu\mathbf{B}_0}{kT}\right) \quad (3.17)$$

$$\frac{N^+}{N^-} = 1 - \left(\frac{2\mu\mathbf{B}_0}{kT}\right) \quad (3.18)$$

For the proton ( $I = 1/2$ ,  $\Delta E = 2\mu\mathbf{B}_0$ ) in a magnetic field of 9.39 T (tesla)—resonance frequency of 400 MHz—and at a temperature of 300 K, the quantity  $2\mu\mathbf{B}_0/kT$  has a value of about  $6 \times 10^{-5}$ . In words the excess population in the lower energy state is extremely small, approximately one nucleus in 300,000. As will be seen in the following, the NMR experiment monitors the *relaxation* of molecules from an excited state to a ground state. This is in contrast to the EPR technique and many other instrumental techniques that detect *absorption* of energy in a species moving from a ground state to an excited state.

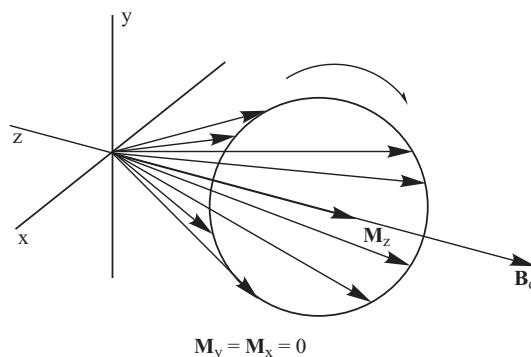
If one defines the length of the vector as in Figure 3.9B as  $[I(I+1)]^{1/2}$ , the angle  $\theta$  is given by equation 3.19:

$$\cos \theta = \frac{m}{[I(I+1)]^{1/2}} \quad (3.19)$$

It is possible to translate these, in three dimensions, as the motion of a magnet of moment  $\mu$  in an applied magnetic field. The magnet axis becomes inclined to the field, precesses around it, and is defined as the Lamour precession. The half-apex angle of the cone equals  $\theta$  and the angular velocity is  $\gamma\mathbf{B}_0$ , so that one finds equation 3.16 now rewritten as equation 3.20:

$$v = \left(\frac{\gamma}{2\pi}\right) \mathbf{B}_0 \quad (3.20)$$

Equation 3.20 defines the nuclear resonant frequency. The precession cone of excess low-energy nuclei is shown in Figure 3.10, with the magnetic axis indicated as the  $z$  axis leading to a net magnetization  $\mathbf{M}_z$  along the  $z$  axis. The  $x$ - and  $y$ -axis components average to zero, that is,  $\mathbf{M}_x = \mathbf{M}_y = 0$ .



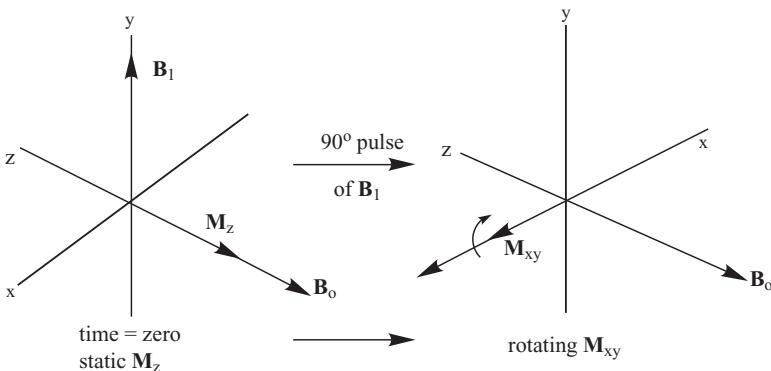
**Figure 3.10** Freely precessing nuclei in a magnetic field. The figure represents low-energy nuclei in a sample arising from different atoms (drawn with the same origin). (Adapted with permission of Nelson Thornes Ltd. from Figure 1.2b of reference 21.)

To produce an observable magnetic effect, the NMR technique resorts to an additional perturbation for detection of a resonance. The additional perturbation results from the application of a sinusoidally oscillating magnetic field with frequency  $= \gamma B_0 / 2\pi$  along the  $y$  axis of Figure 3.10. The perturbing field is generated by passing a radio-frequency (RF) alternating current through a coil wrapped around the sample space. The field generates a vector (called  $\mathbf{B}_1$ ) around which the nuclei precess. The precession yields a resultant magnetization vector  $\mathbf{M}_{xy}$  and, consequently, a net absorption of energy by the nuclei. When  $\mathbf{B}_1$  is cut off, the precession stops; however, the signal output due to the loss of the energy previously absorbed by the nuclei is observable for between 10 ms and 10 s. The point at which maximum current has been induced in the coil during  $\mathbf{B}_1$ 's application is known as a  $90^\circ$  pulse. If one observes only the resultant nuclear magnetization, this laboratory frame of reference is known as the rotating frame. A simplified version of this concept is shown in Figure 3.11. In this scheme the resultant nuclear magnetization  $\mathbf{M}_{xy}$  is illustrated with  $\mathbf{B}_1$ 's position frozen. The fact that all low-energy nuclei act in concert and at a single frequency leads to the name “resonance spectroscopy.”

A nuclear magnetic resonance spectrometer measures the frequency of “nuclear resonance” with sufficient accuracy to provide useful information. The spectrometer features a strong stable magnet containing a gap within which the sample is placed. The sample is surrounded by a transmitter-receiver coil. The usual modern system contains a superconducting cryomagnet for generation of magnetic field strengths between 200 and 750 to 1000 MHz.

### 3.4.2 Nuclear Screening and the Chemical Shift

If all the nuclei being detected in an NMR experiment (all protons in an organic ligand molecule, for instance) resonated at the same frequency, chemists would not be very interested because little information about structure,



**Figure 3.11** Rotating Frame when  $\mathbf{B}_1$  is applied in the form of a  $90^\circ$  pulse. (Adapted with permission of Nelson Thornes Ltd. from Figure 1.6 of reference 21.)

and so on, would be gained. However, when a magnetic field is applied during an NMR experiment, electrons surrounding nuclei in the molecules under study set up a secondary magnetic field. The secondary field opposes the main field, reducing the nuclear frequency. The magnitude of the frequency change is proportional to  $\mathbf{B}_0$ . This is important in that there will be larger separations between resonant frequencies at higher magnetic field strengths, allowing one to detect finer differences between the different protons in any liquid sample. The effect of electrons surrounding the nucleus on the nucleus in the applied magnetic field is termed screening (or shielding). Taking equation 3.20 and introducing the screening constant,  $\sigma$ , one finds equation 3.21:

$$\nu = \left( \frac{\gamma}{2\pi} \right) \mathbf{B}_0 (1 - \sigma) \quad (3.21)$$

The screening constant,  $\sigma$ , is dimensionless and usually recorded in parts per million (ppm). Contributors to  $\sigma$ , opposite in sign, are  $\sigma_d$  (the diamagnetic term) and  $\sigma_p$  (the paramagnetic term). The diamagnetic term depends upon the density of circulating electrons—the number of electrons surrounding the nucleus of interest. The paramagnetic effect in this context does *not* imply the presence of unpaired electrons (to be discussed below) but is substantial, and dominates, for heavier atoms with many electrons in outer orbitals involved in chemical bonding. Several factors affect  $\sigma_p$ :

1. The inverse of the energy separation,  $\Delta E$ , between ground and excited electronic states of the molecule. This means that there will be a correlation between NMR spectra and absorption in the visible and ultraviolet spectral regions.
2. The relative electron density in  $p$  orbitals involved in bonding.
3. The value of  $\langle 1/r^3 \rangle$ , the average inverse cube distance from the nucleus to the electronic orbitals involved.

The paramagnetic screening constant becomes disproportionately larger for heavier elements; thus while  $^1\text{H}$ , the proton, exhibits screening for its compounds within a range of 20 ppm, thallium ( $^{205}\text{Tl}$ ) compound screening constants range over 5500 ppm. Changes in screening of each nucleus do not increase continuously with atomic number but are periodic, following the value of  $\langle 1/r^3 \rangle$ , increasing along each period and then falling markedly at the beginning of the next. Screening constants change in complex manners dependent upon a number of factors including charge density near the nucleus ( $^{14}\text{N}$  nucleus is 25 ppm more shielded in  $\text{NH}_3$  than in  $\text{NH}_4^+$ ), the influence of neighboring  $\pi$  systems, and oxidation states or coordination number of the nucleus being observed ( $^{31}\text{P}$  screening increases in the series  $\text{PCl}_3 < \text{PCl}_4^+ < \text{PCl}_5 < \text{PCl}_6^-$ ). Usually, screening increases for substituted main group elements as the electronegativity of the substituent increases. The nephelauxetic effect (expansion of the electron cloud and increasing electron delocalization in legend–metal bonding) changes the screening effect down the halogen group, thus, while the difference between  $\text{AlCl}_4^-$  and  $\text{AlBr}_4^-$  is 22 ppm, that between  $\text{AlBr}_4^-$  and  $\text{AlI}_4^-$  is 47 ppm.

Anisotropic magnets may be formed in chemical bonds within a molecule so that nuclei in the vicinity may be screened or descreened. Anisotropic behavior would be found in the vicinity of a carbonyl bond, for instance. The benzene ring exhibits ring current anisotropy, leading to large descreening (downfield shifts) of benzene protons. Molecules containing electric dipoles perturb molecular orbitals and therefore the screening of a nuclei. The closer the nucleus is to the bond generating the electric field, the more they are descreened. In 1-chloropropane the descreening shifts, compared to  $\text{CH}_4$ , are  $\alpha\text{-CH}_2$  3.24 ppm,  $\beta\text{-CH}_2$  1.58 ppm, and  $\text{CH}_3$  0.83 ppm.

The electron ( $s=1/2$ ) has a very large magnetic moment that affects the NMR spectrum of any molecule possessing a paramagnetic center. If paramagnetic transition metal ions are present in the molecule, large effects are observed. The screening constants cover a much larger range than is normal for the nucleus under study because the unpaired electrons apparently can delocalize throughout the molecule and appear at or “contact” nuclei. The large resonance frequency shifts that result are called contact shifts. For proton spectra these shifts may have magnitudes of several hundred parts per million (ppm). The behavior has great utility in simplifying complex NMR spectra as illustrated here for some lanthanide elements. Octahedral complexes of lanthanide ions such as europium (Eu), dysprosium (Dy), praseodymium (Pr), and ytterbium (Yb), complexed with organic ligands to render the metals soluble in organic solvents normally used for NMR samples, are added to the system being studied. Because the lanthanides may assume higher coordination numbers, the so-called shift reagent may react with suitable donor sites (such as O or N) within the target molecule. The interaction produces a pseudocontact shift caused by the anisotropic magnetic moment of the shift reagent (similar to the neighbor anisotropy effect described above). The shift may not only move the resonance of protons to different locations in the spectrum but

affects each proton differently, causing larger separation of the resonances. Both effects are useful in revealing more information about the structure of the molecule. The effect of paramagnetic transition metals on NMR spectra have been used to great advantage in the analysis of bioinorganic systems.

Normally, one speaks of “chemical shifts” rather than “screening” when discussing NMR spectra. For two nuclei in different environments with screening constants  $\sigma_1$  and  $\sigma_2$ , from equation 3.21, the corresponding nuclear frequencies will result in equation 3.22. Applying equation 3.22 to  $\sigma_2$ , subtracting, eliminating  $\mathbf{B}_0$ , and setting  $\sigma_1 \ll 1$ , one finds equation 3.23:

$$\nu_1 = \left( \frac{\gamma}{2\pi} \right) \mathbf{B}_0 (1 - \sigma_1) \quad (3.22)$$

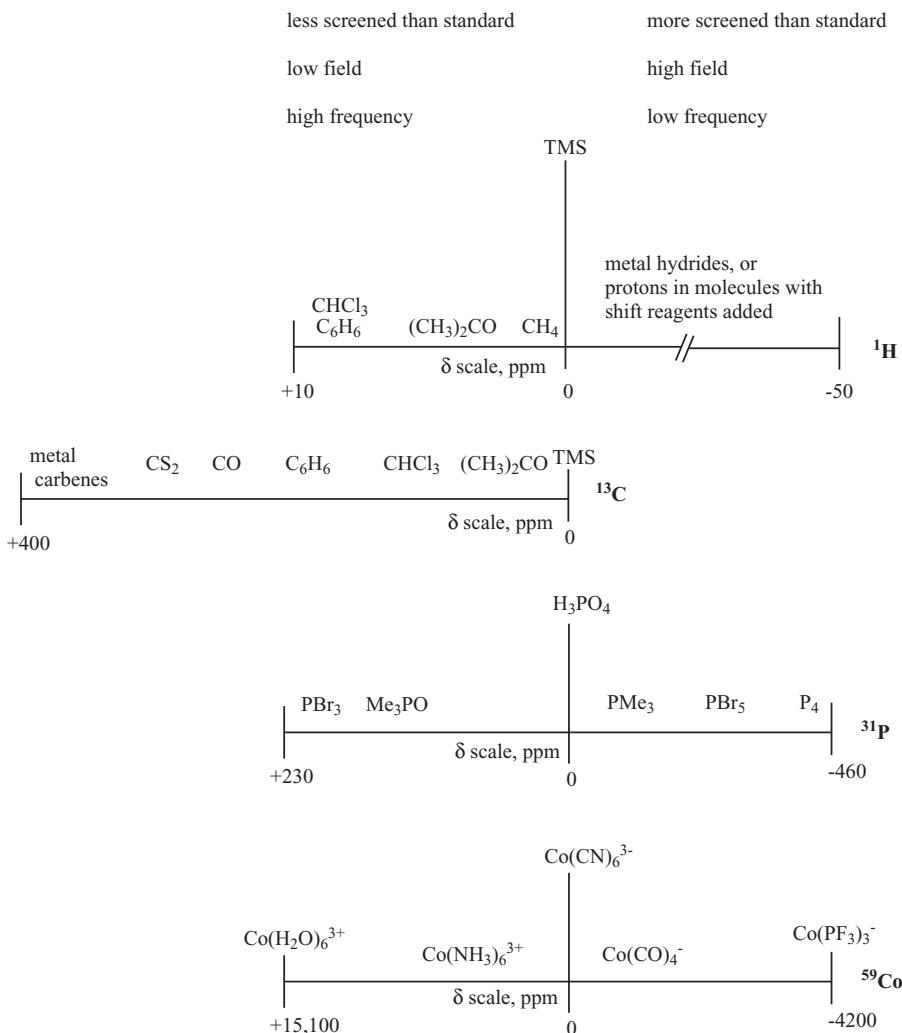
$$\left( \frac{\nu_1 - \nu_2}{\nu_1} \right) = (\sigma_2 - \sigma_1) \quad (3.23)$$

The quantity  $\sigma_2 - \sigma_1$  is given the symbol  $\delta$  and is called the chemical shift. Its value is expressed in parts per million (ppm). It is normal to establish a chemical shift scale for a given nucleus choosing a standard and defining its chemical shift arbitrarily at zero ppm. Tetramethylsilane,  $(\text{CH}_3)_4\text{Si}$ , or TMS, is used as the standard for the common nuclei,  $^1\text{H}$  and  $^{13}\text{C}$ . In recording spectra using TMS, it is normal to depict the descreened (deshielded) nuclei to the left of the standard. One calls the deshielded region “low field” or “high frequency.” A useful diagram detailing the terminology adapted from Figure 2.12 of reference 21 is shown in Figure 3.12.

### 3.4.3 Spin–Spin Coupling

A nucleus under study by nuclear magnetic resonance techniques is affected by other nuclei in the same molecule. This phenomenon is known as *spin–spin coupling*. The effect arises (in adjacent nuclei) from the two electrons joining the nuclei in a covalent bond. Suppose the energy of states in which the electrons in the bond have opposing spins is lower than the state in which the electron spins are parallel. Then the  $\Delta E$  between the two states (in this case a negative number) is called the coupling constant,  $J$ , expressed in frequency units, Hz. Internuclear spin–spin coupling constants may be either positive or negative and depend on a number of factors:

1. The number and bond order of bonds intervening between the nuclei as well as the bond angles. Usually the interaction is observed only through one to four bonds, and the effect is attenuated (the  $J$  value becomes smaller) as the number of intervening bonds increases.
2. The magnetic moments of the two interacting nuclei. These are directly proportional to the product of the magnetogyric ratios ( $\gamma_A \gamma_B$ ) of the interacting nuclei.



**Figure 3.12** Chemical shift scales and chemical shifts of some compounds. (Adapted with permission of Nelson Thornes Ltd. from Figure 2.12 of reference 21.)

3. The valence *s* electron density at the nucleus. This is affected by the *s* character of the bonding orbitals between the interacting nuclei.

Nuclei coupling to each other through spin–spin interactions may have very similar or very different chemical shifts. The difference or similarity will affect the appearance of the resonances associated with the coupled nuclei. Nuclei separated by small chemical shifts are denoted with the letters A, B, and C while sets of nuclei separated by large chemical shifts are designated A, M, and X.

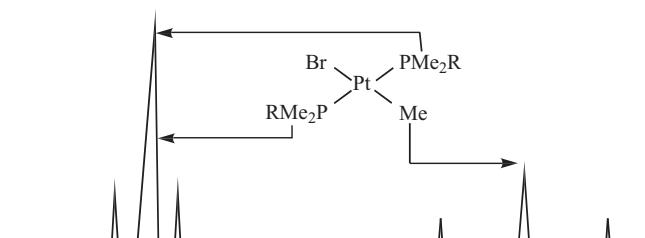
The number of nuclei in each letter category is indicated by a subscript. Using the proton as an example,  $\text{CH}_3\text{CH}_2\text{Cl}$  (chloroethane) is an example of an  $\text{A}_3\text{X}_2$  system while  $\text{CH}_2\text{CHCl}$  (vinyl chloride) is an example of an ABX system. When chemical shifts differences are large, coupling between protons on adjacent atoms will follow the simple  $n+1$  multiplicity rule for the number of peaks in a multiplet (the general rule is  $2nI+1$ , where  $I$  is the nuclear spin). This is named a first-order pattern. The ABX system is almost first-order, but  $\text{A}_a\text{M}_b\text{X}_x$  or  $\text{A}_a\text{B}_b\text{C}_c$  systems exhibit complex spin-coupling multiplet patterns.

An example of spin–spin coupling between the  $^{195}\text{Pt}$  nucleus ( $I=1/2$ , abundance = 33.8%) and the proton ( $^1\text{H}$ ,  $I=1/2$ , abundance = 99.985%) is shown schematically in Figure 3.13 for the complex *trans*-MeBrPt(PMe<sub>2</sub>R)<sub>2</sub> (where R is a 2,4-dimethoxyphenyl group).<sup>21</sup> The two major methyl proton resonances are indicated and connected to the responsible peaks. The height of the central downfield methyl resonance indicates that it corresponds to the protons of four methyl groups attached to phosphorus, and thence to the magnetically inactive platinum nucleus (66.2% abundant). The two smaller satellite peaks at one-quarter intensity on either side of the major downfield peak originate from the same methyl protons coupled to the magnetically active  $^{195}\text{Pt}$  nucleus. The longer coupling path from  $^{195}\text{Pt}$  through  $^{31}\text{P}$  to  $^1\text{H}$  results in a weaker, smaller coupling constant (a so-called  $^3J$  coupling) when compared to the upfield pattern for methyl protons of the methyl group directly attached to platinum. The 1:4:1 pattern for the upfield peak again indicates that the coupling corresponds to the 33.8% abundant platinum nucleus ( $^{195}\text{Pt}$ ). The upfield resonance corresponds to the protons of the methyl group directly attached to the platinum atom, and thus the satellite peaks exhibit an appreciably stronger coupling and consequently a larger  $J$  value.

A more complete discussion of spin–spin coupling may be found in Chapter 3 of reference 21 and many instrumental chemistry texts.<sup>2</sup>

### 3.4.4 Techniques of Spectral Integration and Spin–Spin Decoupling

Two other important NMR concepts and related NMR techniques are introduced in Figure 3.13. The first of these is that the size of the peak is proportional to the number of protons resonating at a given frequency. One can



**Figure 3.13**  $^1\text{H}$  spectrum of the complex *trans*-MeBrPt(PMe<sub>2</sub>R)<sub>2</sub>. (Adapted with permission of Nelson Thornes Ltd. from Figure 3.13 of reference 21.)

integrate the area under the different peaks in a spectrum containing a number of proton resonances and determine how many protons resonate at each given frequency. Frequently, this can lead to confirmation of a proposed molecular structure. The second concept might have occurred to the reader when considering the phosphorus nucleus in the compound.  $^{31}\text{P}$  nuclei are 100% abundant and should couple to both the magnetically active platinum and hydrogen nuclei. This fact should complicate the spin–spin couplings in Figure 3.13. This has not taken place because the effect of the  $^{31}\text{P}$  spin-1/2 nucleus has been removed by a technique called double irradiation or spin-decoupling. In this technique an NMR sample is irradiated at a resonant frequency for one particular nucleus (in this case the  $^{31}\text{P}$  nucleus). This irradiation causes the orientation of the nucleus to become indeterminate, and the resonance of adjacent nuclei will not show splitting due to spin–spin coupling with the irradiated nucleus. Thus the spectrum is simplified. This technique is applied in many different ways in NMR spectroscopy. For instance, heteronuclear spin-decoupling is essential for  $^{13}\text{C}$  NMR spectroscopy. The large number of  $^{13}\text{C}$ – $^1\text{H}$  couplings in an organic molecule or ligand would make the  $^{13}\text{C}$  spectrum extremely complex if the spins were not decoupled. However, because many protons in different electronic environments in organic molecules or ligands result in many resonances, a range of frequencies must be used for spin-decoupling. This has the favorable result of increasing the intensity of  $^{13}\text{C}$  nuclei (1.108% abundant) and the unfavorable result of not allowing integration of peaks to count numbers of carbon atoms of specific types. Spin–spin coupling is discussed further in the section on the nuclear Overhauser effect (NOE), Section 3.4.6.

### 3.4.5 Nuclear Magnetic Relaxation

If one perturbs a physical system from equilibrium and then removes the perturbing influence, the system will return to its original equilibrium condition. This does not happen instantaneously but occurs over some time, according to the equation

$$(n - n_e)_t = (n - n_e)_0 \exp\left(\frac{-t}{T}\right) \quad (3.24)$$

where  $(n - n_e)_t$  is the displacement from equilibrium,  $n_e$ , at time  $t$ ,  $(n - n_e)_0$  is the displacement from equilibrium,  $n_e$ , at time zero, and  $T$  is the relaxation time.

Two types of relaxation processes are known with possibly different relaxation times. These are known as  $T_1$  and  $T_2$ . Looking at Figure 3.11, one can imagine a  $180^\circ$  pulse that inverts the magnetization. Following the end of the pulse, relaxation processes begin to return magnetization to its initial state. This process is called  $T_1$  or longitudinal relaxation because it takes place in the direction of  $\mathbf{B}_0$ . If one uses a  $90^\circ$   $\mathbf{B}_1$  pulse, the magnetization is moved to the  $xy$  plane ( $\mathbf{M}_{xy}$ ) as in the rotating frame figure, Figure 3.11. This transverse magnetization rotates at the nuclear Larmor frequency; and because some

nuclear spins are faster and some are slower, the  $xy$  magnetization starts to fan out and lose coherence, eventually resulting in  $\mathbf{M}_{xy}=0$ . The characteristic time for this process is called transverse relaxation or  $T_2$ . The equations governing the behavior of transverse ( $\mathbf{M}_{xy}$ ) or longitudinal ( $\mathbf{M}_z$ ) magnetization and their return to equilibrium following a  $\mathbf{B}_1$  pulse are given by equations 3.25 (in which  $\mathbf{M}_z$  increases from zero to its equilibrium value) and 3.26 (in which the transverse magnetization falls from its maximum (equal to  $\mathbf{M}_z$ ) to zero when sufficient time has elapsed).

$$(\mathbf{M}_z)_t = (\mathbf{M}_z)_\infty \left[ 1 - \exp\left(\frac{-t}{T_1}\right) \right] \quad (3.25)$$

$$(\mathbf{M}_{xy})_t = (\mathbf{M}_{xy})_0 \exp\left(\frac{-t}{T_2}\right) \quad (3.26)$$

Methods for measuring  $T_1$  and  $T_2$  are discussed in Chapter 5 of reference 21. Suffice it to say here that understanding the method for measuring  $T_2$  (the Carr–Purcell pulse sequence or spin-echo method) becomes important for discussing two-dimensional NMR spectra. When spin–spin coupling is present, a modulation of spin echoes is produced, and it is this fact that is important in 2-D NMR spectroscopy. Nuclear relaxation rates and mechanisms become important when discussing the effect of paramagnetic metal centers on NMR spectroscopy.

### 3.4.6 The Nuclear Overhauser Effect (NOE)

The double resonance experiment can be used to simplify a spectrum as discussed in Section 3.4.4, or to probe correlations between different nuclei. Two types of double resonance experiments are described. In the homonuclear double resonance experiment the nuclei irradiated are the same isotope as those observed: Shorthand notation for this is, for example,  $^1\text{H}\{^1\text{H}\}$ . In heteronuclear double resonance, the nucleus irradiated may differ from that observed: Observing  $^{13}\text{C}$  while irradiating  $^1\text{H}$  has the notation  $^{13}\text{C}\{^1\text{H}\}$ .

As stated previously, for a spin=1/2 nuclei so that  $I=1/2$  and  $\Delta E=\mu\mathbf{B}_0/I$ , one obtains equation 3.27. Rewriting this in terms of the magnetogyric ratio, where  $\gamma=2\pi/\hbar(\mu/I)$ , yields equation 3.28.

$$\frac{N_+^I}{N_-^I} = \exp\left(\frac{-\Delta E}{kT}\right) = \exp\left(\frac{-2\mu\mathbf{B}_0}{kT}\right) \quad (3.27)$$

$$\frac{N_+^I}{N_-^I} = \exp\left(\frac{-\gamma_I\hbar\mathbf{B}_0}{kT}\right) \quad (3.28)$$

The number of nuclei in the upper energy state,  $N_+^I$ , is less than that in the lower energy state,  $N_-^I$ , and the probabilities of upwards and downwards transitions are different. The spin transitions are caused by the spins  $S$  of a nucleus, and the influence of these occurs directly through space. The transi-

tions and their relaxations may be between the same or different type of nucleus, but in either case they are chemically shifted from the spin  $I$ . At equilibrium, some relative spin  $I$  populations exist and the fractional difference in populations between the two energy states can be written (remembering that  $e^x$  is approximated by  $1+x$ ) as

$$\frac{(N_-^I - N_+^I)}{N_-^I} = \frac{\gamma_I \hbar \mathbf{B}_0}{kT} \quad (3.29)$$

A similar expression for the difference of the two spin  $S$  populations can be written as

$$\frac{(N_-^S - N_+^S)}{N_-^S} = \frac{\gamma_I \hbar \mathbf{B}_0}{kT} \quad (3.30)$$

Now if  $S$  is strongly irradiated, then it is saturated and  $S$  is no longer at its Boltzmann equilibrium. Therefore it cannot maintain the Boltzmann equilibrium of spins  $I$  and the intensity of the  $I$  signal is changed. Equalizing  $S$  populations produces a proportional change in  $I$  populations such that equation 3.31 can be written in which  $\eta_{IS}$  is called the nuclear Overhauser enhancement (NOE) factor.

$$\begin{aligned} \frac{\text{Intensity } I \text{ with } S \text{ irradiated}}{\text{Normal intensity } I} &= \frac{\frac{\gamma_I \hbar \mathbf{B}_0}{kT} + \frac{\phi \gamma_S \hbar \mathbf{B}_0}{kT}}{\frac{\gamma_I \hbar \mathbf{B}_0}{kT}} \\ &= 1 + \frac{\phi \gamma_S}{\gamma_I} \\ &= 1 + \eta_{IS} \end{aligned} \quad (3.31)$$

Two factors contribute to  $\eta_{IS}$ . One is the ratio of the magnetogryric ratios of the two different spins, and the other depends on relaxation mechanisms. Provided that the relaxation mechanism is purely dipole-dipole,  $\phi$  has the value 1/2. If other relaxation mechanisms affect spin  $I$ , then  $\phi$  may approach zero. Assuming that the dipolar mechanism is operational (no quadrupolar nuclei with  $I > 1/2$  are present),  $\eta$  has the value  $\gamma_S/2\gamma_I$  and is regarded as  $\eta_{max}$ . In the homonuclear case we have  $\eta_{max} = 1/2$ . Usually one chooses nuclei where  $\gamma_S > \gamma_I$  to ensure that the NOE is significant. For observation of  $^{13}\text{C}$  for instance, if the protons in the molecule are double irradiated, the ratio is 1.99 and  $1 + \eta_{max}$  equals approximately 3. To repeat a statement made above, proton broad-band irradiation enhances the intensity of the  $^{13}\text{C}$  nucleus, which otherwise has very low receptivity.

Practically speaking, NOEs occur only between atoms that are quite close together in space (within 4–5 Å) because the effect occurs with a radius dependence proportional to  $1/r^6$ . Experimentally, one speaks of strong (atoms closer

than 2.5 Å), medium (atoms 2.5–3.5 Å apart), and weak (atoms 3.5–5 Å apart) NOEs. Methods developed by the Wüthrich group use NOEs to determine the amino acid sequence in any small protein; that is, nearly all resonances can be assigned to a specific aa residue. (See Section 3.4.10.)

### 3.4.7 Obtaining the NMR Spectrum

Three parameters affect an NMR spectrum: the chemical shift, coupling, and nuclear relaxation. These must be accounted for when obtaining the NMR spectrum from the spectrometer's output. Obtaining the NMR spectral plot from the output (the free induction decay, FID) of a modern NMR spectrometer involves the analysis of the mathematical relationship between the time ( $t$ ) and frequency ( $\omega$ ) domains, known as the Fourier relationship:

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \exp(-i\omega t) dt \quad (3.32)$$

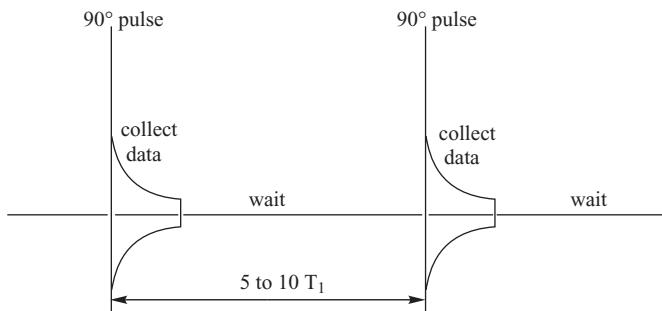
which is also written as

$$F(\omega) = \int_{-\infty}^{\infty} f(t) [\cos(\omega t) - i \sin(\omega t)] dt \quad (3.33)$$

The Fourier transform (FT) relates the function of time to one of frequency—that is, the time and frequency domains. The output of the NMR spectrometer is a sinusoidal wave that decays with time, varies as a function of time, and is therefore in the time domain. Its initial intensity is proportional to  $M_z$  and therefore to the number of nuclei giving the signal. Its frequency is a measure of the chemical shift, and its rate of decay is related to  $T_2$ . Fourier transformation of the FID gives a function whose intensity varies as a function of frequency and is therefore in the frequency domain.

When one applies the perturbing field  $\mathbf{B}_1$ , the nuclei in the sample precess as discussed above. The point at which maximum current has been induced during  $\mathbf{B}_1$ 's application is known as a 90° pulse. Short  $\mathbf{B}_1$  pulses—less than 100 µs—ensure that all nuclei in a sample, whatever their chemical shift, are swung around  $\mathbf{B}_1$  by an appropriate angle. Long  $\mathbf{B}_1$  pulses choose nuclei of a particular chemical shift to precess around  $\mathbf{B}_1$  without affecting other nuclei in the sample. The ideal Fourier transform experiment, allowing spins to relax to equilibrium before successive pulses are applied, is illustrated in Figure 3.14.

Assuming that the spectrometer is stable, a series of isolated 90° pulses will each give an identical nuclear response, and these can be added together in computer memory to yield a strong total response. Actually, most pulsing sequences omit the waiting time and have pulse sequences in the 40° to 30° range, depending on the nuclear isotope observed, the chemical shift range, the relaxation time  $T_1$ , and the computer memory size. Different nuclei in different parts of the molecule may have different relaxation times, so that pulse



**Figure 3.14** The ideal Fourier transform experiment. (Adapted with permission of Nelson Thornes Ltd. from Figure 5.12 of reference 21.)

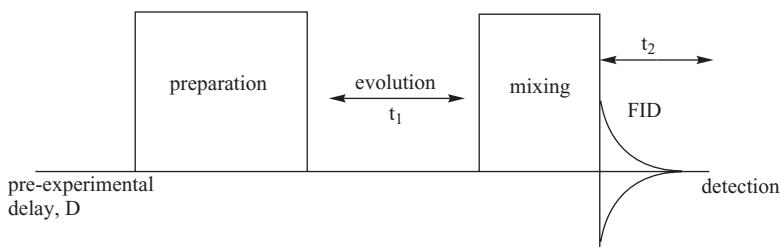
lengths for each are slightly different. Some distortion of signal intensity results from this factor; thus if precise quantitative data is required, the pulse sequence of Figure 3.14 is necessary. The intensity of a FID is proportional to the number of nuclei contributing to the signal. When transformed to the so-called absorption spectrum, integration of the area under the peaks relates to the number of nuclei resonating at a given frequency.

Each FID signal is accompanied by noise; however, the noise is incoherent—sometimes positive, sometimes negative—so that it increases more slowly than the desired nuclear signal. A series of  $N$  FIDs has a signal-to-noise ratio  $\sqrt{N}$  times better than a single FID, allowing spectroscopists to obtain useful chemical information from otherwise unreceptive nuclei or from dilute solution samples having few of the nuclei of interest.

The output of the NMR spectrometer must be transformed from an analog electrical signal into digital information that can be stored in the computer's dedicated computer. The minicomputers used in NMR spectroscopy have memory used for data accumulation, programs for manipulating the data, and storage devices to store large collections of data for future or additional manipulation into useful spectral results.

### 3.4.8 Two-Dimensional (2D) NMR Spectroscopy

Every NMR experiment must have a preparation sequence (inducing the nuclei to resonate) and detection capability (finding out what happened). Two-dimensional NMR spectroscopy adds two more domains between preparation and detection. These are an indirect evolution time,  $t_1$ , and a mixing sequence (see Figure 3.15). The two dimensions of two-dimensional NMR spectroscopy are those of time. In one time domain, FIDs containing frequency and intensity information about the observed nuclei is collected. The second time dimension refers to the time that elapses between some perturbation of the system and the onset of data collection in the time domain. The second time period is varied, and a series of FID responses are collected for each of the variations.



**Figure 3.15** Conceptual diagram of a two-dimensional (2D) NMR experiment. (Adapted with permission of Nelson Thornes Ltd. from Figure 8.1 of reference 21.)

Following this, each FID is Fourier-transformed to produce perturbed spectra. These are stacked, and the data sets at each frequency point are transformed once again. The results give data that are displayed on two frequency axes, with the peaks represented by contours on a relief map. The two-dimensional plot contains connectivities between the different nuclei. The connectivities or correlations may involve spin–spin coupling of nuclei through covalent bonds or show spatial proximity through nuclear Overhauser (NOE) effects. A simple conceptual diagram of the 2D experiment is shown in Figure 3.15.

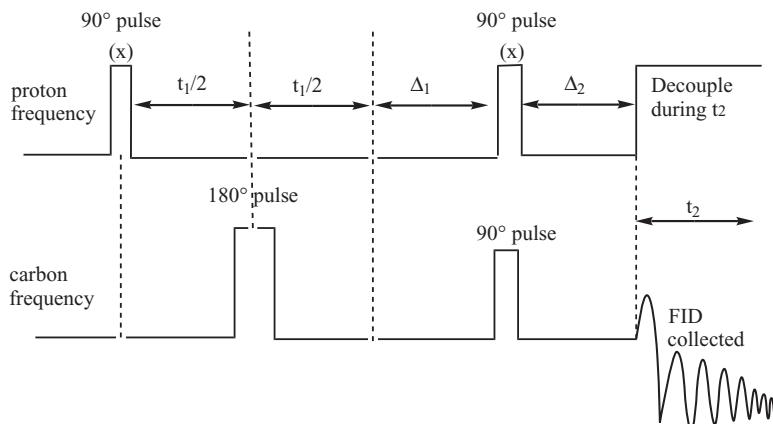
### 3.4.9 Two-Dimensional Correlation Spectroscopy (COSY) and Total Correlation Spectroscopy (TOCSY)

Three two-dimensional NMR methods have been commonly used in determining the three-dimensional structure of proteins, although there are many variations that we will not discuss here. The website at [http://www.genomicglossaries.com/content/nmr\\_email.asp](http://www.genomicglossaries.com/content/nmr_email.asp) contains a glossary and definitions of some other NMR methods. Two methods, COSY and TOCSY, transfer magnetization through the chemical bonds between adjacent protons. These techniques are discussed in this section. The third method, nuclear Overhauser effect spectroscopy (NOESY), is discussed in Section 3.4.10. COSY or correlation spectroscopy is a technique with many variations, all of them depending upon the existence of spin–spin coupling between nuclei. The coupling provides responses relating to the chemical shift positions of coupled nuclei. Its equivalent in one-dimensional spectroscopy would be a series of double resonance experiments at each multiplet in the spectrum. One can perform homonuclear (probing couplings between the same nuclei, i.e.,  $^1\text{H}$ – $^1\text{H}$ ) or heteronuclear (probing connectivities between different nuclei, i.e.,  $^1\text{H}$ – $^{13}\text{C}$ ) COSY experiments.

In the pulse sequence for a homonuclear COSY experiment, the first  $90^\circ$  pulse flips the  $z$  magnetization into the  $x$  direction and into the  $xy$  plane. Considering an AX spin system (one in which the nuclei have very different chemical shifts) having two doublets due to spin–spin coupling, the magnetization will include four components precessing at different frequencies. During a

time  $t_1$  the magnetization separates into two pairs of vectors due to the chemical shift difference, with the separation in each pair being due to the coupling interaction. The second  $90^\circ$  pulse swings the magnetization into the  $yz$  plane, where it starts to precess around the  $z$  axis. If the vectors were in the right half of the  $xy$  plane, they are swung into the  $-z$  direction; if they were in the left half of the  $xy$  plane, they are swung into the  $+z$  direction. Inversion of the components causes transfer of polarization, and the signal of either doublet will be a function of  $t_1$ . For each value of  $t_1$  the FID will contain the four frequencies of the two AX doublets distorted in both intensity and phase. The FIDs are collected at a number of  $t_1$  values, and then they are transformed to yield a stack of spectra with different degrees of distortion. These are then transformed at each frequency to obtain a two-dimensional (2D) map. Where there is no correlation between spins, the frequency is the same in both dimensions. These signals appear on a diagonal plot, which is the normal 1D spectrum. Where there is a correlation, the chemical shift and coupling constant mix in each resonance and a signal (peak) appears off the diagonal. The off-diagonal peaks (cross peaks) appear at the intersection of the A and X chemical shift resonances. Instantaneously, one can identify two peaks in the 1D spectrum that correlate to each other, that exhibit spin–spin coupling, and that are therefore located next to each other.

The heteronuclear COSY experiment investigates the connectivities imparted by coupling paths between two different nuclei, most commonly those between  $^1\text{H}$  and  $^{13}\text{C}$ . The experiment is facilitated by the fact that one-bond CH coupling constants are much larger than the two or three bond (CCH or CCCH) coupling constants. Decoupling of protons from  $^{13}\text{C}$  is accomplished in the usual way by broad-band decoupling during the  $t_2$  period while accumulating the  $^{13}\text{C}$  FID. Other, more complex pulse sequences are capable of decoupling the proton resonances as well. See Figure 3.16 for a simplified



**Figure 3.16** Pulse sequence used to produce a  $^1\text{H}$ – $^{13}\text{C}$  heteronuclear COSY spectrum. (Adapted with permission of Nelson Thornes Ltd. from Figure 8.9 of reference 21.)

COSY pulse sequence. The first  $90^\circ$  pulse at the  $^1\text{H}$  frequency produces transverse  $y$  magnetization. This evolves with  $t_1$  and depends on the frequency of the proton's chemical shift. Protons attached to carbon nuclei will have two magnetization components precessing at different frequencies. These components are refocused halfway through  $t_1$  by inverting the carbon magnetization with a  $180^\circ$  pulse. The proton magnetization at the end of  $t_1$  is a function of  $t_1$ . This magnetization is transferred to the carbon magnetization by a second proton  $90^\circ$  pulse placing the  $y$  magnetization into the  $z$  direction and permitting polarization transfer. The period  $\Delta_1$  allows the two magnetization components to come into phase before the second proton  $90^\circ$  pulse is applied. This delay is timed to coincide with  $1/2J(\text{CH})$ , one-half  $^1J(\text{CH})$ , the coupling constant between  $^{13}\text{C}$  and  $^1\text{H}$  nuclei. The  $90^\circ$  pulse applied at the carbon frequency at the same time as the second proton pulse creates transverse carbon magnetization and produces the output FID. The refocusing delay  $\Delta_2$  is usually the same as  $\Delta_1$ . The experiment is repeated for many values of  $t_1$ , and the stack of FIDs obtained (which contain both carbon and proton chemical shift information) are transformed and plotted. Usually the carbon spectrum appears on the  $x$  axis of the 2D plot with the proton spectrum plotted vertically. Contours appear at the intersection of carbon and proton resonances corresponding to the carbon and proton atoms bonded to each other.

Only protons that are two ( $^2J$ ) to three ( $^3J$ ) bonds apart will exhibit coupling (and therefore cross peaks) in a 2D experiment. Protons that are more than three chemical bonds apart give no cross signal because the  $^4J$  coupling constants are close to 0. Taking the example of amino acid residues in a protein, important cross signals arise between the proton associated with the main-chain amide group ( $\text{H}^N$ ) and the proton(s) attached to the  $\alpha$  carbon atom ( $\text{H}^\alpha$ ). These cross signals allow derivation of the *phi* torsion angle ( $\phi$ ) of the protein backbone from the  $^3J$  coupling constant between them (see Figure 2.8). Continuing with the protein example, the  $\text{H}^\alpha$  proton transfers magnetization to the beta protons,  $\text{H}^\beta$ ,  $\text{H}^\beta$  protons transfers to the  $\text{H}^\alpha$  and gamma protons,  $\text{H}^\gamma$ , if any are present, then the gamma proton(s) transfer to the  $\text{H}^\beta$  and  $\text{H}^\delta$  protons, and the process continues.

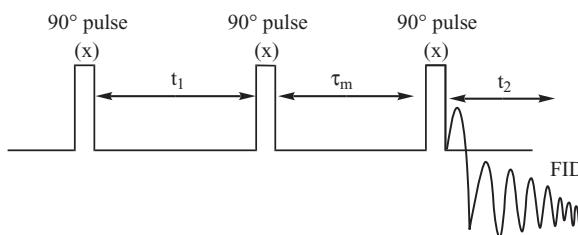
The TOCSY 2D NMR experiment correlates *all* protons of a spin system, not just those directly connected via three chemical bonds. For the protein example, the alpha proton,  $\text{H}^\alpha$ , and all the other protons are able to transfer magnetization to the beta, gamma, delta, and epsilon protons if they are connected by a continuous chain—that is, the continuous chain of protons in the side chains of the individual amino acids making up the protein. The COSY and TOCSY experiments are used to build so-called spin systems—that is, a list of resonances of the chemical shift of the peptide main chain proton, the alpha proton(s), and all other protons from each aa side chain. Which chemical shifts correspond to which nuclei in the spin system is determined by the conventional correlation spectroscopy connectivities and the fact that different types of protons have characteristic chemical shifts. To connect the different spin systems in a sequential order, the nuclear Overhauser effect spectroscopy

(NOESY) experiment must be used. See Section 3.4.10. The NOESY experiment transfers magnetization through space, and, therefore, it will show cross peaks to all protons that are close in space (within 4–5 Å of each other) regardless of whether they are in the same spin system or not. In a folded protein, aa residues are inherently close in space, so assignments can be made by comparing the peaks in the NOESY experiment with other spin systems determined by COSY and TOCSY.

### 3.4.10 Nuclear Overhauser Effect Spectroscopy (NOESY)

NOESY NMR spectroscopy is a homonuclear two-dimensional experiment that identifies proton nuclei that are close to each other in space. If one has already identified proton resonances in one-dimensional NMR spectroscopy or by other methods, it is then possible to determine three-dimensional structure through NOESY. For instance, it is possible to determine how large molecules such as proteins fold themselves in three-dimensional space using the NOESY technique. The solution structures thus determined can be compared with solid-state information on the same protein obtained from X-ray crystallographic studies. The pulse sequence for a simple NOESY experiment is shown in Figure 3.17.

The first  $90^\circ$  pulse places magnetization in the  $xy$  plane, and the two components then precess at their individual frequencies for  $t_1$ . At the end of  $t_1$ , each component will have its own  $x$  and  $y$  component of magnetization. The second  $90^\circ$  pulse places the magnetization into the  $xz$  plane, where the  $y$  components continue to precess about the  $z$  axis. A short mixing time,  $\tau_m$  (typically 0.05 s), follows, allowing an exchange of magnetization to take place. At the end of  $\tau_m$  the two component frequencies have been modulated through the process of exchange going on between the nuclei (this in turn is a function of  $t_1$ ). An FID is produced by a third  $90^\circ$  pulse, which moves the magnetization back into the  $xy$  plane. Transformation of the FID stacks yields the two-dimensional trace with cross peaks between exchanging nuclei. The pulse sequence is set up to detect two uncoupled spins having different chemical shifts but that undergo slow exchange. In nonexchanging nuclei, through-space



**Figure 3.17** A pulse sequence used for a NOESY NMR experiment. (Adapted with permission of Nelson Thornes Ltd. from Figure 8.12 of reference 21.)

relaxation can also produce an exchange of magnetization through the NOE effect. In either case, nuclei that are close together in space may produce cross peaks in the NOESY experiment that, along with additional structural information (see COSY and TOCSY in Section 3.4.9), leads to three-dimensional structure determination of large molecules in solution.

### 3.4.11 Multidimensional NMR

Over the last 10–15 years, multidimensional (3D and 4D) NMR spectroscopic and computational techniques have been developed to solve structures of larger proteins in solution. This technology has the advantage of resolving the severe overlap in 2D NMR spectra for proteins having >100 residues. Usually, multidimensional NMR experiments require isotopic labeling of some or all nuclei of interest. For example, signal overlap in a two-dimensional  $^1\text{H}^1\text{H}$  correlation spectrum can be resolved in the dimension of the heteronucleus—that is,  $^{15}\text{N}$  or  $^{13}\text{C}$  attached to the protons of relevance. If fast relaxation is a problem, all nonexchangeable protons can be deuterated as well, leading to sharper signals. In a year 2000 publication by the Kurt Wüthrich group, one finds this example.<sup>22</sup> Transverse relaxation-optimized spectroscopy (TROSY)-type triple resonance NMR experiments were used to determine the secondary structure of an octameric 110-kDa protein. The final structure determined from the uniformly  $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled 7,8-dihydroneopterinaldolase (DHNA), a lyase, from *Staphylococcus aureus*, showed 20-fold to 50-fold sensitivity gains when compared to the corresponding conventional triple resonance NMR experiments. Complete sequence-specific assignments of the 121-residue (13.7-kDa) polypeptide chain in the 110-kDa octamer were obtained in aqueous solution at 20°C. Additionally, the secondary structures determined in the solution by NMR were found to coincide nearly identically with those in the crystal structure of the DHNA octamer. It is important to note that secondary and tertiary structures of monomeric or multicomponent enzymes in the 110-kDa, ~900-amino acid size range have not been accomplished using multidimensional solution-NMR to date in 2007. Portions of these larger molecules have been studied by solution NMR, and examples will be found in Chapters 5, 6, and 7.

Currently the upper limit of applicability of multidimensional NMR methods may be for monomeric proteins up to 50 kDa, 250–300 amino acid residues, but this may increase with future technological advances. Using these methods, one is able to resolve protein structure at the same level as X-ray crystallographic data that has been resolved to approximately 2.5 Å. While it may be true that a carefully refined X-ray structure of a given protein (a solid-state picture) may not be identical to the “physiologically true” solution structure determined by NMR, in most cases where structures from both methods are available, strong agreement in secondary and tertiary structure has been found. Evidence reinforcing the comparison comes from calculations of three bond coupling ( $^3J_{\text{H}\alpha\text{HN}}$ ) constants from well-refined crystallographic data that

have been found to be in excellent agreement with experimentally determined values in solution. Methods that improve structure refinement in NMR solution studies include the use of a conformational database potential and direct refinement against three-bond coupling constants ( $^3J$  values), secondary  $^{13}\text{C}$  shifts,  $^1\text{H}$  shifts,  $T_1/T_2$  ratios, and residual dipolar couplings. Residual dipolar couplings (RDCs), yield long-distance restraints not accessible by other solution NMR parameters. The article by Clore and Gronenborn gives an excellent overview of the new refinement strategies that increase the accuracy of solution NMR structures.<sup>23</sup>

### 3.4.12 Descriptive Examples

Cytochromes c (see Section 7.7) are ubiquitous in nature. This electron transfer protein is relatively small (~12.5 kDa, ~100+ amino acid residues) and features high solubility, high helical content, thermodynamic stability, and a spectroscopically accessible iron-containing heme cofactor. In eukaryotes, cytochrome c mediates single electron transfer between mitochondrial inner membrane enzymes cytochrome bc<sub>1</sub> (Section 7.6) and cytochrome c oxidase (Section 7.8). Cytochromes bc<sub>1</sub> and cytochrome c oxidase are known as complexes III and IV, respectively, of the physiological respiratory chain. Cytochromes c are characterized by the attachment of the heme c cofactor to the protein chain through thioether linkages provided by two cysteinyl residues covalently bonded to vinyl substituents of the heme porphyrin ligand. (See Figures 7.25 and 7.32.) In addition to the four porphyrin nitrogen ligands, cytochrome c's heme iron ion carries one axial histidine ligand. Some variety exists for the iron ion's second axial ligand, although it is often the S<sub>δ</sub> atom of a methionine residue.

Recently, Sivakolundu and Mabrouk published an NMR solution structure of horse heart ferrocytochrome c in which the heme contains an Fe(II) ion.<sup>24</sup> This solution structure was the first in which the cytochrome c protein was dissolved in a nonaqueous solvent: a solvent mix of 70% water and 30% acetonitrile (ACN). The data obtained from the NMR study are deposited in the protein data bank (PDB) as: (1) 1LC1, the minimized average NMR structure; and (2) 1LC2, the 30 lowest energy NMR structures.

The ferrocytochrome c features a 104-amino-acid residue single chain with secondary structure consisting of five α-helices, two omega loops, and several random coil segments. The c-type heme is attached covalently to the polypeptide chain through cys14 and cys17. The signature CXXCH sequence found in cytochromes c exists in this ferrocytochrome as cys14–ala15–gln16–cys17–his18. His18 and met80 form the heme Fe(II)'s axial ligands. His 18 is bonded through its N<sub>ε2</sub> side-chain atom with a N<sub>ε2</sub>–Fe = 1.97 Å bond length. Met80 is bonded through its S<sub>δ</sub> atom with S<sub>δ</sub>–Fe = 2.68 Å, an atypically long Fe–S bond length. Much more detail on this and other cytochromes c is found in Section 7.7. Here we will discuss how the NMR structure was obtained experimentally.

Horse heart cytochrome c solutions were prepared at concentrations of ~4–6 mM in 30% CD<sub>3</sub>CN (deuterated acetonitrile) and 70% D<sub>2</sub>O or H<sub>2</sub>O solutions containing 50 mM phosphate buffer (pH 6.9). The external standard (<sup>1</sup>H, 0.00 ppm) used was 2,2-dimethyl-2-silapentane-5-sulfonate sodium salt (DSS), a commonly used standard in biological and pharmacological NMR experiments. Data were collected on one of two 600-MHz NMR instruments, and then they were exported and processed in Felix 98, offline data processing, visualization, and analysis software from Molecular Simulations (now Accelrys; see <http://www.accelrys.com/>). The ferrocytochrome c NOESY spectra were recorded with mixing times of 100 ms with 64 scans per FID using a spectral width of 20 ppm. A total of 2048 points were collected in the *F*2 dimension and 256 points in the *F*1 dimension. TOCSY spectra were recorded with mixing times of 20, 50, and 75 ms with the same acquisition parameters as the NOESY spectra. TOCSY, total correlation spectroscopy, is a COSY-like method that shows different <sup>1</sup>H connectivities at different mix times, providing more detailed structural information and thus reducing the number of possible structures. DQF-COSY spectra were recorded with 128 scans per FID with 2048 × 256 (*F*2 × *F*1) points acquired. DQF-COSY, double quantum filtered COSY, uses a pulsed-field gradient to assist in the separations of overlapping spectral peaks and thus to reduce the number of possible structures. In this case, DQF-COSY produced <sup>3</sup>J<sub>HαHN</sub> coupling constants (coupling constants between the proton attached to the aa α-carbon atom and the proton attached to the peptide main chain amide nitrogen) for 73 of the ferrocytochrome c's 104 amino acid residues.

The next calculations establish distance restraints using cross-peak volumes generated from the 100-ms 2D NOESY spectra. The first step involves manual assignment of cross peaks due to intra-residue and sequential proton-proton interaction (~500 NOESY peaks assigned). The second step is an automated NOESY cross-peak assignment using the NOAH module in the DYANA structure calculation program. The DYANA software program suite for structural analysis of NMR data results from the work of Kurt Wüthrich and fellow researchers.<sup>25</sup> Professor Wüthrich shared the 2002 Nobel Prize in Chemistry “for his development of nuclear magnetic resonance spectroscopy for determining the three-dimensional structure of biological macromolecules in solution.” Researchers Sivakolundu and Mabrouk performed 30 cycles until no further changes in the resonance assignments of cross peaks were observed in NOAH. This process resulted in assignment of 2232 NOESY cross peaks.

Next, the researchers performed structural calculations using the DYANA program suite. First, the ferrocytochrome c heme was constructed and designated as HES, using the heme and his18 coordinates from the cytochrome c solution structure deposited as PDB: 1GIW.<sup>26</sup> The HES residue is added so that the heme moiety and his18 are free to move independently during subsequent structural calculations. Because DYANA does not recognize the iron atom, it was represented by a dummy atom. The NOESY cross-peak volumes

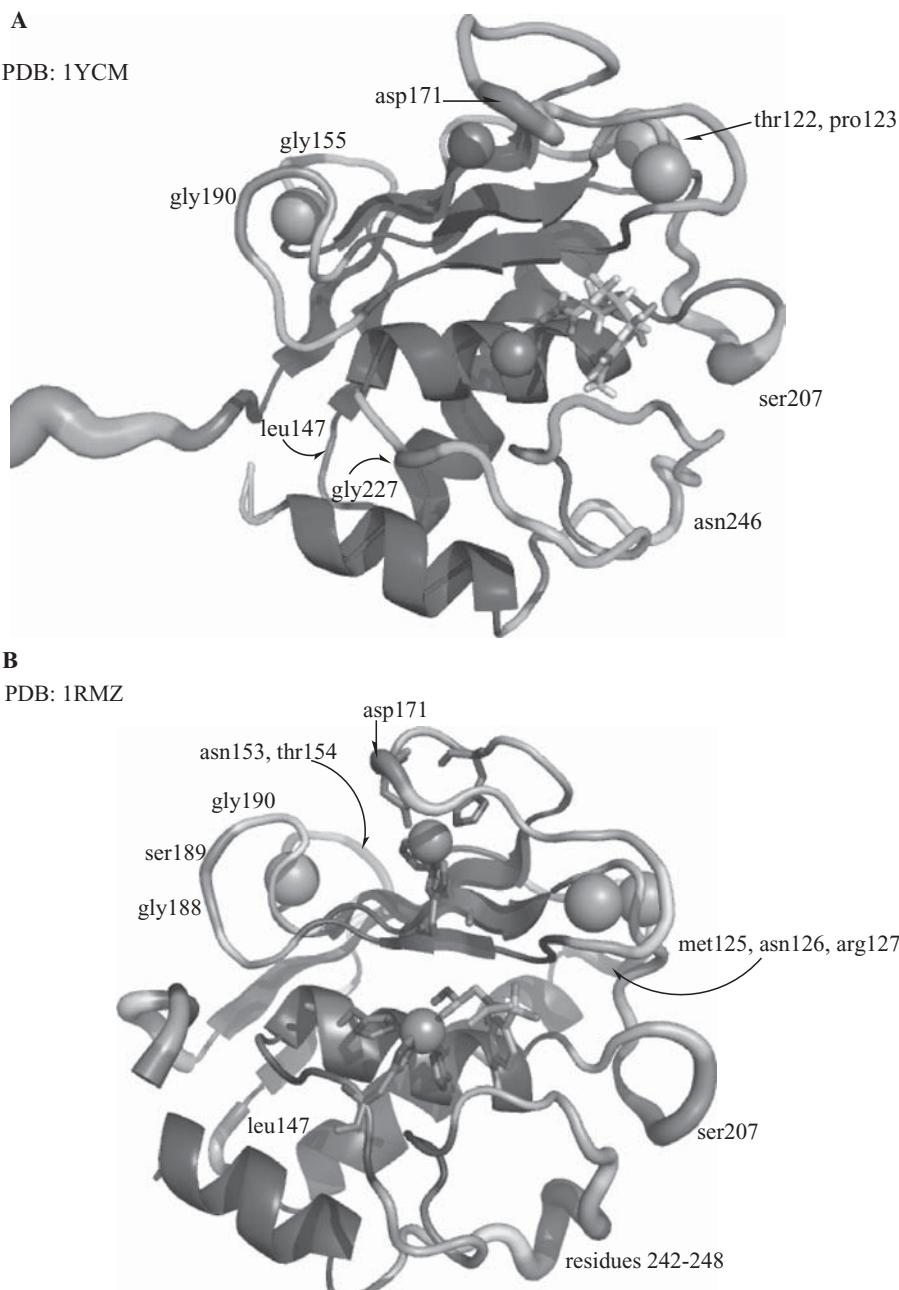
were converted into distance constraints using the CALIBA module in DYANA (see reference 25b). The structural calculation began by subjecting 200 randomly generated structures to a simulated annealing torsional angle dynamics procedure. The PBD: 1 GIW structural data were used again to set the Fe–S distance for the Fe(II) heme's met80 sulfur ligand at 2.50 Å. Consistent violation for this bond during DYANA structural calculation caused the researchers to relax the Fe–S bond length by 0.25 Å. Indeed, the final met80 sulfur ligand–Fe(II) bond length is reported as S<sub>δ</sub>–Fe = 2.68 Å, an atypically long Fe–S bond length. Stereospecific assignments for 41 diastereotopic ferrocytocrome c protons were made using the HABAS and GLO MSA modules in DYANA. The final calculation was performed with 2232 distance restraints and 73 angle restraints on 750 randomly generated conformers. A family of 30 structures was used for subsequent energy minimization.

Restrained energy minimization (REM) was carried out using the SANDER module in the AMBER 5.0 molecular modeling package. AMBER (Assisted Model Building with Energy Refinement) refers to two things: (1) a molecular mechanical force field for the simulation of biomolecules and (2) a package of molecular simulation programs which includes source code and demos.<sup>27</sup> The ferrocytocrome c heme was generated using the coordinates and angles for heme b, as available in AMBER, and was then modified to represent heme c. Force field parameters for all residues and the heme were taken from the AMBER database. The conformation of the met80 ligand was determined by the NMR restraints used in the DYANA structure calculation. The charges on the heme were obtained by an ab initio calculation using Gaussian 98. The hypothetical model complex used a diethyl thioether to represent met80, a 2-methylimidazole to represent his18 and CH<sub>3</sub>–S– groups attached to the CAB and CAC heme carbon atoms. (See Figures 7.25 and 7.32 to see representations of heme c and its atom identification system.) The coordinates used were again taken from the 1GIW NMR structure. The distance and angle restraints for the energy minimizations were imposed using a square-well penalty function with force constants of 20 kcal mol<sup>-1</sup> Å<sup>-2</sup> and 70 kcal mol<sup>-1</sup> rad<sup>-2</sup>, respectively. The energy minimizations were performed until the E<sub>rms</sub> was less than 0.4 kcal mol<sup>-1</sup>. The average structure (PDB: 1LC1) was calculated from the 30 REM structures (PDB: 2LC1) obtained by this procedure. Finally, structure validation for stereochemical quality was performed using PROCHECK and PROCHECK-NMR. Figure 7.34 in Section 7.7 visualizes the ferrocytocrome c structure resulting from these experiments and calculations.

To further extend the utility of structural methods, researchers compare solid state X-ray crystallographic and solution-state NMR structures to define important differences. For instance, the Bertini group has studied the enzyme matrix metalloproteinase 12 (MMP12), in the presence of its inhibitors.<sup>28</sup> Matrix metalloproteinases (MMPs) are involved in extracellular matrix degradation, a fundamental step in tissue remodeling and repair. There are a great variety of enzymes of this type, the one studied here is one of many found in humans. Most MMPs have three domains: (1) a prodomain that is removed

upon enzyme activation; (2) a catalytic domain responsible for the enzyme's hydrolase activity—the enzyme domain studied here; and (3) a hemopexin-like domain that may necessary for substrate recognition. Much pharmaceutical research has taken place to find inhibitors of MMP because misregulation or overexpression of the enzyme is known to be a factor in several disease states. For instance, MMP expression and activity increases in almost every type of human cancer. Its expression correlates with increased invasiveness, metatheses (spreading of the tumor to other parts of the body), and decreased survival rates in cancer patients. No MMP inhibitor has passed the clinical trial stage of drug development because of side effects caused by their low selectivity. Researchers believe that one reason for the lack of selectivity may be due to the inherent flexibility of the MMP protein's backbone. The Bertini group's research looked at X-ray structures of several different MMPs complexed with different inhibitors finding that certain loop regions in the protein show enhanced mobility and/or conformational heterogeneity. The researchers then looked at the structures of various MMPs studied by NMR, finding similar results—increased mobility and conformational heterogeneity in certain loop regions.

The matrix metalloproteinase discussed here—MMP12—consists of 159 aa residues (numbers 105–263) and has a molecular weight of 16.7 kDa. The inhibitor *N*-isobutyl-*N*-[4-methoxyphenylsulfonyl]glycyl hydroxamic acid, NNGH, is complexed to the protein. The X-ray crystallographic structure, at 1.34-Å resolution, is deposited as PDB: 1RMZ, and the NMR structures are deposited as PDB: 1YCM (20 structures of lowest energy) and 1Z3J (minimized average structure). The root mean square deviation (RMSD) between the backbone (BB) atoms of the X-ray and NMR structures differ with total BB RMSD = 1.37 Å. The RMSD is much less in  $\alpha$ -helix or  $\beta$ -sheet domains with BB RMSD = 0.65 Å. Dissimilarity increases substantially in the loop domains with BB RMSD = 1.64 Å. Dissimilarity in the loop regions is indicated in Figure 3.18 using PyMOL's preset "B factor putty" setting that shows how much structural uncertainty is occurring in two ways: (1) by thickness of ribbon—that is, the thicker the ribbon, the more the uncertainty or mobility of position; and (2) by color—that is, blue-green = least uncertainty or flexibility; yellow-orange = more flexibility; red = most flexibility. The MMP12 catalytic domain's secondary structure is that of a typical matrix metalloproteinase: three  $\alpha$ -helices, a twisted five-strand  $\beta$ -sheet, and eight connector loops (see Figure 3.18). Loops L5 and L8 are quite long. The resulting topology reads as follows: L1- $\beta$ 1-L2- $\alpha$ 1-L3- $\beta$ 2-L4- $\beta$ 3-L5- $\beta$ 4-L6- $\beta$ 5-L7- $\alpha$ 2-L8- $\alpha$ 3. Three calcium ions and their binding sites lie approximately in the plane of the  $\beta$ 3,  $\beta$ 5, and  $\beta$ 4 strands of the  $\beta$ -sheet and are bound by residues in the  $\beta$ -turns of the sheet. Two zinc(II) ions, Zn1 and Zn2, are found in the domain. Zn1, the catalytic zinc ion, is five-coordinate with ligands: N<sub>e2</sub> atoms of his218, his222, and his228 and two oxygen ligands of the NNGH inhibitor (see Figure 3.18). Zn2 is tetrahedrally coordinated with ligands: N<sub>e2</sub> atoms of his168, his183, and his196 and the O<sub>δ1</sub> atom of asp170.



**Figure 3.18** Secondary structure of matrix metalloproteinase 12 (MMP12): (A) PDB: 1YCM (structure determined using solution NMR) and (B) PDB: 1RMZ (structure determined using X-ray crystallography). Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft Corporation.) (See color plate)

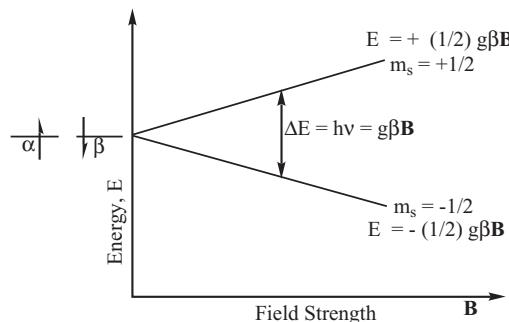
The reference 28 authors concluded that for the NMR solution study of NNGH inhibited MMP12 (MMP12-NNGH), the disorder exhibited arose from the lack of medium- to long-range NOEs due to the protein's backbone (BB) mobility. The side-chain NOEs were strongly reduced in intensity, indicating that these were very mobile as well. Residual dipolar couplings (RDCs), potentially sensitive indicators of mobility in NMR structures, were found, for the PDB: 1YCM NMR structure, to be reduced in their absolute values compared to calculated values. The changes occurred particularly at hinges between loops and  $\alpha$ -helices or  $\beta$ -sheets, and in loop regions. In Figure 3.18A, these regions can be identified as: (1) between  $\beta$ 1 and L2, and L2 and  $\alpha$ 1 (see orange-red residues 122–124); (2) before  $\beta$ 2 (residues 146–147); (3) after  $\beta$ 2 (residue 155); (4) in L5 (residue 171); (5) in L7 (residue 207); and (6) in L8 (residues 244–248). Importantly, comparison of the loop regions in solid state (X-ray) versus solution (NMR) indicated a high degree of similarity in regions exhibiting mobility (see Figures 3.18A and 3.18B). The reference 28 authors also concluded, based on comparisons to other MMPs that have had their NMR structures determined, that the overall profiles of the structures were similar and that similar regions of high mobility were found in loop regions. Overall, the authors concluded that comparisons of X-ray crystallographic and NMR structures, as well as NMR structures of similar molecules, yield important information about regions of instability in proteins. This information will be useful in many regards; in the case of this study of MMP inhibitors, the conclusions are important for researchers attempting to discover drug molecules that could be clinically useful.

### 3.5 ELECTRON PARAMAGNETIC RESONANCE

#### 3.5.1 Theory and Determination of g-Values

In electron paramagnetic resonance (EPR) spectroscopy, also called electron spin resonance (ESR), radiation of microwave frequency is absorbed by molecules, ions, or atoms (organic or inorganic) containing a paramagnetic center—that is, a system with one or more unpaired electrons. In EPR, the unpaired electron spin moment ( $m_s = \pm 1/2$  for a free electron) interacts with an applied magnetic field producing the so-called Zeeman effect. The effect is illustrated in Figure 3.19 for the electron spin functions  $\alpha$  and  $\beta$  corresponding to  $m_s = +1/2$  and  $m_s = -1/2$ .

Three basic equations (3.34–3.36) are needed to describe the technique. In the equations,  $\mu$  is the magnetic moment of the electron, sometimes also written as  $\mu_e$ ,  $g$  is called the  $g$  factor or spectroscopic splitting factor,  $\mathbf{S}$  is defined as the total spin associated with the electron (in bold type because it is considered as a vector),  $\mathbf{B}$  is the imposed external magnetic field (also defined as a vector quantity), and  $\beta = (e/2m) \times (h/2\pi)$  and is called the Bohr magneton.



**Figure 3.19** Removal of degeneracy of the  $\alpha$  and  $\beta$  electron spin states by a magnetic field. (Adapted with permission of John Wiley & Sons, Inc. from Figure 2.16 of reference 3. Copyright 1997 Wiley-VCH.)

$$\mu = -g\beta \mathbf{S} \quad (3.34)$$

$$E = \mu \cdot \mathbf{B} \quad (3.35)$$

$$\Delta E = h\nu \quad (3.36)$$

Replacing  $\mu$  by its equivalent operator from equation 3.34 and  $E$  by  $H$ , the Hamiltonian, equation 3.35 becomes

$$H = g\beta \mathbf{S} \cdot \mathbf{B} \quad (3.37)$$

or by replacing the dot product by  $m_s$ , the projection of  $\mathbf{S}$  onto  $\mathbf{B}$  and multiplying by the magnitude of  $\mathbf{B}$ , equation 3.38 is written as

$$E = g\beta g_s \mathbf{B} \quad (3.38)$$

Giving  $m_s$  its two values,  $\pm 1/2$ , one finds equation 3.39:

$$\Delta E = [E_{1/2} - E_{-1/2}]g\beta \mathbf{B} \quad (3.39)$$

If one then defines the resonance condition,  $B_R$  (alternately written as  $B_0$  or as  $B_L$  when referring to the laboratory field associated with a particular EPR instrument system), as the magnetic field at which the energy of the transition comes into resonance with the field, one finds equation 3.40 or, more usefully, equation 3.41.

$$\Delta E = h\nu = g\beta \mathbf{B}_R \quad (3.40)$$

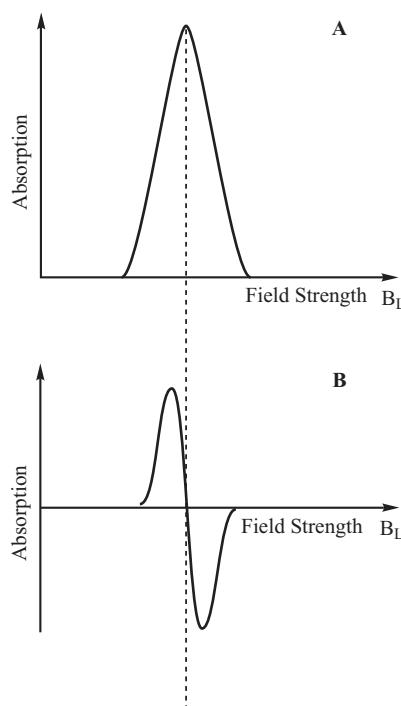
$$g = \frac{h\nu}{\beta B_R} \quad (3.41)$$

From the knowledge of the spectrometer's operating frequency (held constant) and the magnetic field intensity at which maximum EPR absorption occurs as one varies the magnetic field, one easily calculates  $g$  from equation 3.41.

One can calculate the ratio of populations of spin-up to spin-down electron orientations at room temperature ( $T=300\text{ K}$ ) from the Boltzmann formula finding that  $N^+/N^-$  is approximately equal to one (0.999), indicating that there is about a 0.1% net excess of spins in the more stable, spin-down orientation at room temperature. Using the same mathematical expression, this difference in populations can be shown to increase as the temperature is lowered. Actually the EPR signal will be linearly dependent on  $1/T$ , and this linear dependence is called the Curie law. Because of the excited state population's temperature dependence, most EPR spectra are recorded at temperatures between 4 and 77K.

The magnetic field strength at which  $\Delta E$  (proportional to the microwave frequency as seen from Equation 3.40) comes into resonance with the transition illustrated in Figure 3.19 will produce an intensity spike along the magnetic field abscissa. Because of line-broadening phenomena the typical absorption curve will look like Figure 3.20A. The absorption curve is routinely displayed as its first derivative (Figure 3.20B) as details of the spectrum are more easily detected.

This chapter's discussion does not treat inorganic and organic free radicals and triplet states (such as dioxygen,  $\text{O}_2$ ), which produce EPR spectra. Rather,

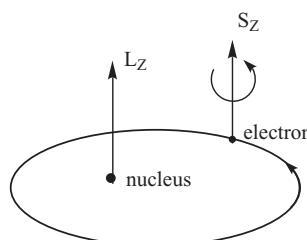


**Figure 3.20** (A) Typical EPR absorption curve. (B) First derivative EPR absorption curve.

the focus here will be on EPR behavior of transition metal centers that occur in biological species. An excellent presentation of the subject, written by Graham Palmer, is found in Chapter 3 of reference 29. The discussion here is summarized mostly from that source.

When an electron is exposed to a magnetic field,  $B_L$ , the electron can be either stabilized ( $m_s = -1/2$ ) or destabilized ( $m_s = +1/2$ ), with the magnitude of the effect varying linearly with the intensity of  $B_L$ . This interaction of magnetic moments with  $B_L$  is called the Zeeman effect or Zeeman interaction. Figure 1 on page 126 of reference 29, along with Figures 3.19 and 3.20, relates the energy level diagram for the electron to the absorption and first-derivative modes of EPR spectral presentation. Standard EPR instrumentation utilizes a fixed frequency (usually the 9 GHz “X band”) and a variable magnetic field. Other frequencies may also be used to enhance spectra, thereby increasing resolution of unresolved hyperfine structure (3 GHz) or resolution associated with  $g$  anisotropy (35 GHz).

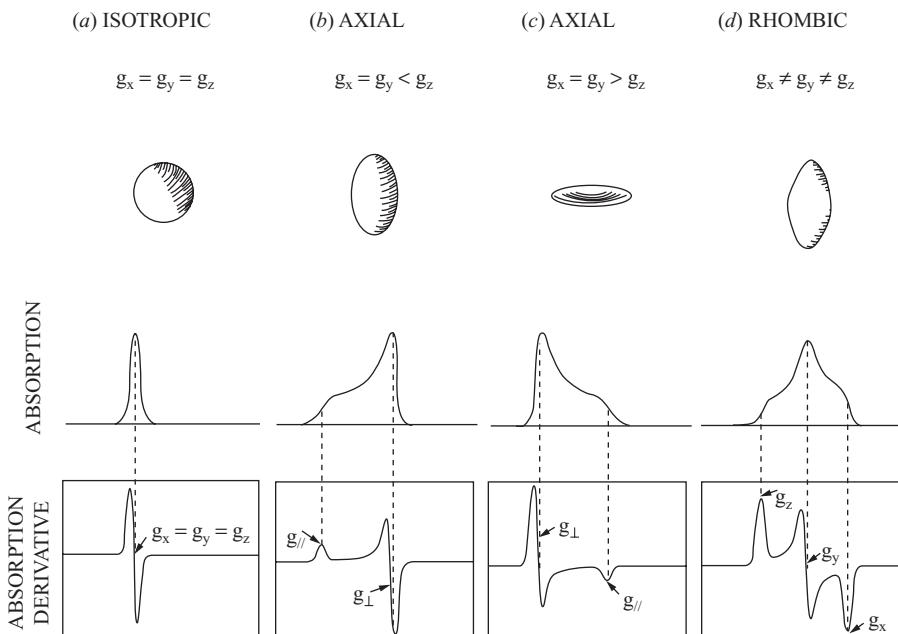
If one calculates  $g$  from equation 3.41, the measurable experimental quantity would appear to be a single number of approximately 2. Observed  $g$  factors for paramagnetic metal ions range from <1 to 18 (measured for some lanthanide ions). Two phenomena, known as spin-orbit interactions (spin-orbit coupling) and zero-field splitting, are responsible for  $g$  factor deviations from the free electron value. Spin-orbit coupling arises because the magnetic dipole associated with the orbital momentum of the electrons ( $L$ ) tends to align itself with the magnetic dipole due to the electrons’ intrinsic spin ( $S$ ). Spin-orbit coupling tends to be quenched if the metal ion exists in a ligand field that lifts the degeneracy of the  $d$  orbitals, and in practice the  $g$  factor value will lie somewhere between the free ion value (favored by spin-orbit coupling) and the free electron value (quenched spin-orbit coupling). The greater the lifting of the degeneracy of the  $d$  orbitals, the more effective the quenching of spin-orbit coupling, and the closer  $g$  will be to free electron value. A simple spin-orbit interaction is illustrated in Figure 3.21.



**Figure 3.21** The motion of an electron in orbit about a nucleus generates an orbital momentum ( $L$ ) adding a component to the magnetic field experienced by the electron spin ( $S$ ). (Adapted with permission of John Wiley & Sons, Inc. from Figure 2.18 of reference 3. Copyright 1997, Wiley-VCH.)

For the simple case of one unpaired electron ( $S=1/2$ ), the associated magnetic moment is not a simple number but is directionally oriented—that is, is anisotropic. Taking spin-orbit interactions into account for EPR spectra leads to four limiting cases:

1.  $g_x = g_y = g_z$ . The magnetic moment is independent of orientation. This case is called isotropic, and a single symmetric EPR absorption is obtained. The paramagnet can be represented by a sphere. The EPR spectrum will resemble Figures 3.20A, 3.20B, and 3.22A.
2.  $g_x = g_y < g_z$ . The paramagnet, represented by a football shape, exhibits a minor feature at low field (from  $g_z$ , often called  $g_{\parallel}$  or  $g$  parallel) and a major feature at high field (from  $g_x$  and  $g_y$ , often called  $g_{\perp}$  or  $g$  perpendicular). The spectrum is said to be axial. If Cu(II)  $d^9$  ions in octahedral ligand fields exhibit  $z$ -axis elongated  $d$ -orbital splitting because of the Jahn-Teller effect (see Section 1.6 for a discussion of these topics), the EPR envelope will appear as shown in Figure 3.22B. All  $g$  values will be  $>2$ .
3.  $g_x = g_y > g_z$ . The paramagnet, represented by a discuss shape, exhibits a minor feature at high field (from  $g_z$ ,  $g_{\parallel}$  or  $g$  parallel) and a major feature



**Figure 3.22** EPR Absorption curves. (A) Isotropic spectrum,  $g_x = g_y = g_z$ . (B) Axial spectrum,  $g_x = g_y < g_z$ . (C) Axial spectrum,  $g_x = g_y > g_z$ . (D) Rhombic spectrum,  $g_x \neq g_y \neq g_z$ . (Reprinted with permission from Figure 4 of reference 29. Copyright 2000, University Science Books.)

at low field (from  $g_x$  and  $g_y$ ,  $g_{\perp}$  or  $g$  perpendicular). The spectrum is also said to be axial. It will appear as shown below in Figure 3.22C.

4.  $g_x \neq g_y \neq g_z$ . The spectrum is said to be rhombic. Three different EPR values are recorded. Figure 3.22D shows this behavior.

When there is more than one unpaired electron in the paramagnetic center, zero-field splitting (zfs) will occur. Zero-field splitting is the separation in energy of the various  $m_s$  states in the absence of an applied magnetic field. It is the result of interelectronic interactions and ligand fields of low symmetry. The hamiltonian for zfs is written as equation 3.42, where  $D$  is the axial zfs parameter and  $E/D$  indicates the degree of rhombic distortion in the electronic environment. The zfs is applied as a correction to the energies of the individual spin states arising from spin-orbit coupling.

$$H_{\text{zfs}} = D \left[ S_z^2 - \frac{1}{3} S^2 + \frac{E}{D} (S_x^2 - S_y^2) \right] \quad (3.42)$$

The case of high-spin Fe(III) ( $S=5/2$ ) is important because it is found in high-spin heme systems of hemoglobin and myoglobin (see Section 7.2). The parameter  $D$  is a directed quantity; that is, it characterizes the magnitude and direction of axial distortion. In the high-spin Fe(III) heme case, the direction is normal to the heme plane. Actually, three cases are of interest for the  $S=5/2$  system. When  $D=0$  there is no zfs, the separation between all levels is the same, and each  $m_s$  level converges to a common origin at zero field. Therefore each level comes into resonance at the same value of  $B_L$ , and a single EPR line is seen. This simple behavior has never been observed in a biological system. When  $0 < D < \hbar v$ , the levels are split in zero field because the paramagnet experiences a small asymmetry in its environment which separates the  $m_s$  values. The levels converge to different origins. When the zero-field separation is small, the levels can come into resonance with  $B_L$  and five separate resonances are observed. In the third case,  $D \gg \hbar v$ , the zero-field splitting is very large. The separation between  $m_s = \pm 3/2$  and  $\pm 5/2$  is too large for the transitions to be observed. Only transitions within the  $m_s = \pm 1/2$  levels are observed, and these are very sensitive to the orientation of the paramagnet with respect to the applied field. The EPR spectrum would look like that of Figure 3.22C with greater separation of  $g_{\parallel}$  from  $g_{\perp}$  and a lower relative intensity for  $g_{\parallel}$ . It can be calculated that the  $m_s = \pm 1/2$  transitions correspond to a  $g_{\parallel}$  value of 2.0 and  $g_{\perp}$  values of 6.0 ( $2S+1$  with  $S=5/2$ ). This is the circumstance found for hemes where the square-planar array of four porphyrin nitrogen ligands gives rise to large  $D$  values. A more complete discussion is found on pages 145–152 of reference 29; Figures 11 and 14 are particularly instructive, and the interested reader is referred there.

### 3.5.2 Hyperfine and Superhyperfine Interactions

Hyperfine interactions in EPR spectra arise when the paramagnet finds itself in the vicinity of a nucleus having a nuclear spin and consequently a nuclear

magnetic moment. The interactions may be of two types: (1) the hyperfine interaction—the unpaired electron and the nucleus having a nuclear spin belong to the same atom; and (2) the superhyperfine interaction—the unpaired electron and the nucleus having a nuclear spin belong to different atoms in the molecule.

The resonance expression relating the magnetic field and energy (frequency) as seen in equation 3.40,  $\Delta E = h\nu = g\beta B_R$ , has its analog for hyperfine interactions written as equation 3.43, where  $A$  is the hyperfine coupling constant in Hz and  $a$  is the hyperfine splitting constant—that is, the distance between the split lines in the EPR spectrum:

$$hA = g\beta a \quad (3.43)$$

For  $g$  equal to 2,  $A$  will equal  $2.8a$  (MHz). Usually  $A$  is expressed in reciprocal centimeters using the relation that  $30\text{ GHz} = 1\text{ cm}^{-1}$ . For  $g=2$ , a hyperfine splitting of  $10\text{ mT}$  can be expressed as  $280\text{ MHz}$  or  $0.0093\text{ cm}^{-1}$ . (Use  $\beta = 9.2741 \times 10^{-21}\text{ erg G}^{-1}$ ,  $h = 6.626 \times 10^{-27}\text{ erg s}$ ,  $1\text{ G} = 10^{-4}\text{ T}$  to calculate that  $10\text{ mT} = 0.0093\text{ cm}^{-1}$ .) Provided that the  $z$  direction is parallel to the magnetic field, one can write

$$E = -\mu_z B_L \quad (3.44)$$

Assuming that  $B_{HF}$  (where HF=hyperfine) is much smaller than  $B_L$ , one can write equation 3.45 from equations 3.34, 3.38, and 3.44. The constant  $h$  is often omitted in this expression.

$$E = -\mu_z(B_L - am_I) = g\beta m_s B_L + A(h)m_s m_I \quad (3.45)$$

where  $m_s$  is electron spin and  $m_I$  is nuclear spin.

The hyperfine interaction is shown in Figure 21 of reference 29. The  $M_s = \pm 1/2$  states of an  $S = 1/2$  paramagnet interact with an  $I = 1/2$  nuclear moment to create the hyperfine interaction. Interactions from  $m_s = -1/2$  to  $M_I = -1/2$  and  $m_s = -1/2$   $M_I = +1/2$ , for instance, create the magnetic field specified as the hyperfine interaction  $A$ . Figure 21 of reference 29 describes the behavior for an  $I = 1/2$  nuclear moment. The number of hyperfine lines will be equal to  $2I+1$  for nuclear moments greater than  $1/2$ . Each hyperfine line will be of equal intensity when the electron is interacting with its own nucleus. For instance, the  $\text{Cu}^{2+}$ ,  $I = 3/2$  nucleus will produce four hyperfine lines as described in Section 3.5.4.

The superhyperfine interaction is observed for metal complexes in cases where the metal ligands have a nuclear moment. For instance, the nitrosyl ( $\text{NO}$ , with an unpaired electron) complexes of iron(II) heme proteins have two inequivalent axial nitrogen ligands. The  $^{14}\text{N}$  ( $I = 1$ ) nucleus of one ligand couples strongly to NO's unpaired electron, yielding a widely split triplet with each component of equal intensity and separated by  $2.1\text{ mT}$ . The second nitrogen of the histidine proximal ligand (see Figures 7.4 and 7.7 in Section 7.2) shows weaker coupling with the splitting about  $0.7\text{ mT}$ . Difficulties arise in trying to

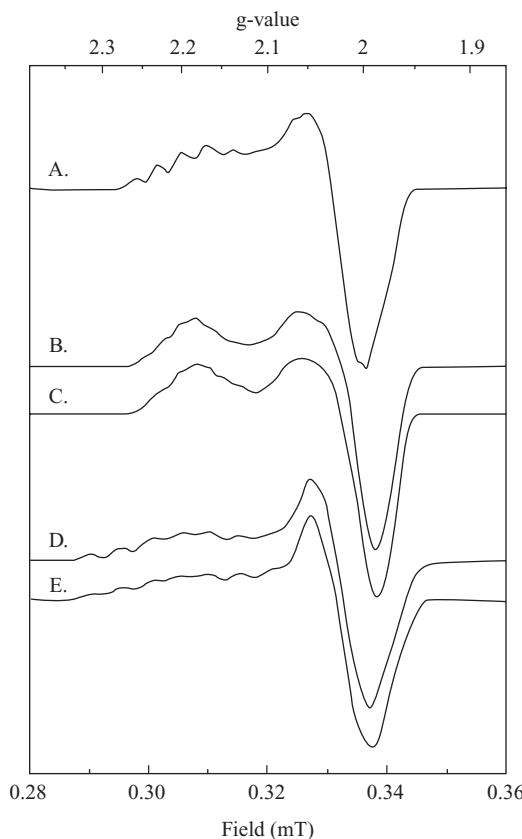
distinguish between hyperfine and superhyperfine coupling in a bioinorganic system where both may be present. Palmer discusses the solutions to this problem on pages 163–165 of reference 29.

### 3.5.3 Electron Nuclear Double Resonance (ENDOR) and Electron Spin-Echo Envelope Modulation (ESEEM)

Electron nuclear double resonance (ENDOR) and electron spin-echo envelope modulation (ESEEM) are two of a variety of pulsed EPR techniques that are used to study paramagnetic metal centers in metalloenzymes. The techniques are discussed in Chapter 4 of reference 1a and will not be discussed in any detail here. The techniques can define electron–nuclear hyperfine interactions too small to be resolved within the natural width of the EPR line. For instance, as a paramagnetic transition metal center in a metalloprotein interacts with magnetic nuclei such as  $^1\text{H}$ ,  $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{14}\text{N}$ ,  $^{15}\text{N}$ ,  $^{17}\text{O}$ ,  $^{31}\text{P}$ , or  $^{33}\text{S}$ , these interactions may be detected by ENDOR or ESEEM analysis. Using these techniques, the nucleus of a particular ligand atom complexed to the metal may be identified. In favorable circumstances, metal–ligand bond distances and bond angles may be determined as well.

### 3.5.4 Descriptive Examples

The enzyme cytochrome c oxidase provides an example of the utility of EPR in studying paramagnetic metal sites in metalloenzymes. The enzyme cytochrome c oxidase (CcO) is a member of an enzyme superfamily that couples oxidation of ferrocytochrome c to the  $4\text{ e}^-/4\text{ H}^+$  reduction of molecular oxygen in eukaryotes. Bacterial CcO complexes consist of two, or at most three, subunits—I, II, III—these contain the catalytic sites for all CcOs. Subunit I contains the ligating amino acid residues for the iron-containing heme a and heme a<sub>3</sub> sites and a monometallic copper ( $\text{Cu}_B$ ) site. CcO's subunit II features a 10-strand  $\beta$  barrel that contains the ligating aa residues for the bimetallic copper ( $\text{Cu}_A$ ) site. Much more detail on cytochrome c oxidase is found in Section 7.8. The metal ions in all sites in the enzyme undergo oxidation and reduction as CcO goes through its catalytic cycle. The bimetallic  $\text{Cu}_A$  site cycles between Cu(II)–Cu(II), Cu(II)–Cu(I), and, most importantly for this example, a Cu(1.5)–Cu(1.5), mixed-valence (MV) state in which the copper ions share an unpaired electron. Researchers P. M. H. Kroneck, W. E. Antholine, and co-workers first proposed the mixed-valence state for the bimetallic copper site,  $\text{Cu}_A$ , on the basis of their EPR measurements in the early 1990s.<sup>30</sup> Although controversial at the time it was first proposed, X-ray crystallography subsequently confirmed the bimetallic  $\text{Cu}_A$  site, as will be discussed in Section 7.8. In 1996, Kroneck and co-workers published an extended study of the dinuclear copper electron transfer center in several enzymes, including cytochrome c oxidase (referred to as COX by these and other authors), nitrous oxide reductase ( $\text{N}_2\text{OR}$ ), and a bimetallic copper site engineered into the soluble domain of subunit II of quinol oxidase (QOX) in *Escherichia coli* (*E. coli*), generating



**Figure 3.23** Experimental EPR spectra of N<sub>2</sub>OR V (spectrum A), PdII (spectrum B), and purple CyoA (spectrum D). Also shown are simulated spectra for PdII (spectrum C), and purple CyoA (spectrum E). (Reprinted with permission from Figure 2 of reference 31. Copyright 1996, American Chemical Society.)

what the authors called the purple CyoA protein.<sup>31</sup> This last protein exhibited different characteristics for the bimetallic center than for the first two and was called the Cu<sub>A</sub><sup>\*</sup> center (see Figure 3.23). First, these researchers noted similarities in the low-temperature magnetic circular dichroism (MCD) spectrum of COX and the room temperature MCD spectrum of N<sub>2</sub>OR that had its mixed-valence, Cu(1.5)–Cu(1.5) bimetallic Cu<sub>A</sub> center detected by EPR previously. Next, the researchers noted that COX's Cu<sub>A</sub> center is found in subunit II and that there was a clear primary sequence similarity between the COX and N<sub>2</sub>OR copper-binding domains—namely, copper ion binding ligands found in the sequences cys–X<sub>3</sub>–cys–X<sub>3</sub>–his–X<sub>2</sub>–his, where X stands for any amino acid residue. To eliminate interference from the other metal centers found in COX, EPR spectroscopy was carried out on a site-directed mutant form of the soluble domain of *Paracoccus denitrificans* COX subunit II expressed in *E.*

*coli* (labeled PdII in Figure 3.23). Figure 3.23 shows the experimental EPR spectra of N<sub>2</sub>OR V (spectrum A), PdII (spectrum B), and purple CyoA (spectrum D). Also shown are simulated spectra for PdII (spectrum C) and purple CyoA (spectrum E). The experimental spectra are axial with  $g_z, g_{\parallel}, > g_x$  and  $g_y, g_{\perp}$  (see Figure 3.22B). The seven-line splitting pattern, to be expected from the coupling of one unpaired electron ( $S=1/2$ ) with two  $I=3/2$  spin copper nucleus (number of lines =  $2I+1=2(6/2)+1$ ), are very similar for N<sub>2</sub>OR and COX, while the purple CyoA protein shows a more complex pattern. For the simulated cytochrome c oxidase subunit II spectra (PdII, spectrum C in Figure 3.23),  $g_{\max}=2.19$ ,  $g_{\text{mid}}=2.03$ , and  $g_{\min}=1.99$ . The small hyperfine splitting constant  $A$  is simulated at 3.1 mT and results in a classic seven-line splitting pattern with an intensity ratio of 1:2:3:4:3:2:1. The small 3.1 mT splitting indicates a high degree of delocalization of the electron between the copper ions and onto the copper ligands as well.

The reference 31 authors conclude that the Cu<sub>A</sub> center in both N<sub>2</sub>OR and COX should be described as a Cu(1.5)–Cu(1.5) dimer rather than as a mixed-valence Cu(II)–Cu(I) dimer. X-ray crystallographic and EXAFS studies on cytochrome c oxidase, as discussed in Section 7.8, have defined two bridging cysteine (Cu<sub>2</sub>S<sub>2</sub> nearly planar core) and two terminal histidine ligands, coordinated through their N<sub>δ1</sub> nitrogen atoms (as well as methionine sulfur and glutamate main chain oxygen ligands) for the Cu<sub>A</sub> center. The structural data agree with Cu<sub>A</sub>'s EPR identification as a bimetallic, delocalized, one-unpaired-electron center in its reduced state with the unpaired electron delocalized over the two copper ions as well as the cysteine sulfur ligands.

The ENDOR technique, described briefly in Section 3.5.3, has been applied to Cu<sub>A</sub> centers in nitrous oxide reductase (N<sub>2</sub>OR) and the recombinant water-soluble fragment of subunit II of *Thermus thermophilus* cytochrome c oxidase (COX), called M160T9, and a mutant M160QT0, where the weak axial sulfur ligand of methionine (M) at position 160 was replaced by a glutamine (Q).<sup>32</sup> The reference 32 authors employed high-field (95-GHz)<sup>1</sup>H ENDOR spectroscopy in a comparative study of the Cu<sub>A</sub> site present in the proteins, for instance, comparing their sulfur and copper spin densities. Comparison of their results to known X-ray crystallographic structures allowed some correlations between the hyperfine parameters of weakly (primarily involved with the histidine ligands) and strongly coupled protons (those involved with the cysteine ligands), the spin density distribution, and structural properties of the various Cu<sub>A</sub> sites. In other words, the Cu<sub>A</sub> ENDOR experiment formed relationships between the EPR spin couplings and couplings involving the nearby protons of the ligand atoms. For instance, the isotropic hyperfine coupling ( $A_{\text{iso}}$ ) of the strongly coupled cysteine β-protons (protons attached to cysteine's β carbon atom) were an important source of structural information because they depended on the H–C–S–S dihedral angle in Cu<sub>A</sub> (the angle formed by the β proton, the β-carbon atom, and the sulfur atom of one cysteine ligand, and the sulfur atom of the other bridging cysteine ligand). Similarly, the weaker hyperfine coupling of the histidine protons, specifically of the H<sub>ε1</sub> protons (protons

attached to the C<sub>e</sub> atom of histidine), provided spin density and structural information—how the histidine ligands are arranged in relation to the Cu<sub>2</sub>S<sub>2</sub> nearly planar core. The researchers found another weak coupling that was assigned to an amide proton in the main chain of one of the cysteine ligands. Their conclusions included the following: (1) The ENDOR data were consistent with the electron-spin density being similar for the two Cu ions ( $\rho_{\text{Cu}}$ ) and two sulfur atoms ( $\rho_{\text{S}}$ ) in the Cu<sub>2</sub>S<sub>2</sub> core, although the COX fragment M160T9 appeared to have a somewhat larger  $\rho_{\text{S}}$  than the other Cu<sub>A</sub> sites; (2) the more weakly coupled histidine H<sub>e1</sub> protons indicated their positions outside of the nearly planar Cu<sub>2</sub>S<sub>2</sub> core; and (3) the most weakly coupled amide proton of Cu ligand cys200 (in the bovine heart cytochrome c oxidase X-ray crystallographic structure, PDB: 1OCC discussed in Section 7.8) might suggest its significance for electron transfer from COX's redox partner, cytochrome c.

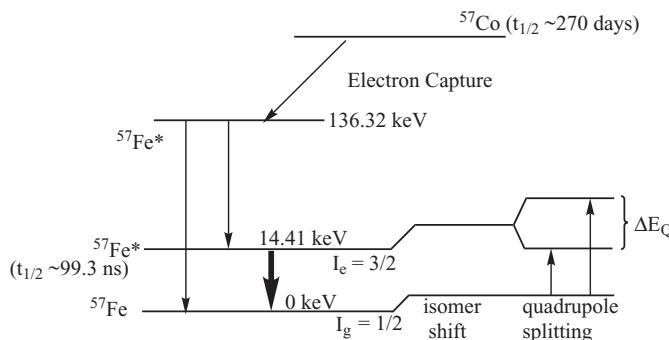
## 3.6 MÖSSBAUER SPECTROSCOPY

### 3.6.1 Theoretical Aspects

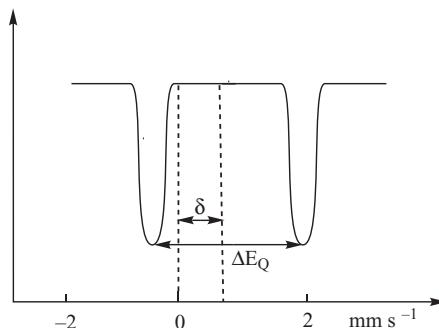
The Mössbauer effect as a spectroscopic method probes transitions within an atom's nucleus and therefore requires a nucleus with low-lying excited states. The effect has been observed for 43 elements. For applications in bioinorganic chemistry, the <sup>57</sup>Fe nucleus has the greatest relevance and the focus will be exclusively on this nucleus here. Mössbauer spectroscopy requires (a) the emission of  $\gamma$  rays from the source element in an excited nuclear state and (b) absorption of these by the same element in the sample under study. The Mössbauer phenomenon requires that the emission and absorption of the  $\gamma$  radiation take place in a recoil-free manner; this is accomplished by placing the nucleus in a solid or frozen solution matrix. Three main types of interaction of the nuclei of interest with the chemical environment surrounding it cause detectable changes in the energy required for absorption: (1) resonance line shifts from changes in electron environment—see discussion of the isomer shift,  $\delta$ , below; (2) quadrupole interactions—see discussion of  $\Delta E_Q$  below; and (3) magnetic interactions. The last type, magnetic interactions, is especially important in studying bioinorganic systems such as iron–sulfur clusters found in aconitase (Section 7.9.2.1), cytochrome bc<sub>1</sub> (Section 7.6), and cytochrome b(6)f (Section 7.5).

The source for the 14.41-keV  $\gamma$  radiation used in Mössbauer experiments is indicated by the boldface arrow in Figure 3.24.<sup>3</sup> Origin of the isomer shift and quadrupole splitting phenomena are indicated at the right-hand side of the diagram.

The electronic environment about the sample's nucleus influences the energy of the  $\gamma$  ray necessary to cause the nuclear transition from the ground to the excited state. The energies of the  $\gamma$  rays from the source can be varied by moving the source relative to the sample. In order to obtain the Mössbauer spectrum, the source is moved relative to the fixed sample, and the source



**Figure 3.24** Cobalt-57 source of 14.41-keV  $\gamma$  radiation used in Mössbauer experiments. Isomer shift and quadrupole splitting characteristics are shown at right. (Adapted with permission of John Wiley & Sons, Inc. from Figure 2.26 of reference 3 and from Figure 1 of reference 34.)



**Figure 3.25** Typical Mössbauer spectrum for a sample containing randomly oriented molecules.

velocity at which maximum absorption of  $\gamma$  rays occurs is determined. For a  $^{57}\text{Fe}$  source emitting a 14.4-keV  $\gamma$  ray, the energy is changed by  $4.8 \times 10^{-8}$  eV for every  $\text{mm s}^{-1}$  of velocity imposed on the source.<sup>33</sup> Detectors will be similar to those used for X-ray crystallography.

Figure 3.25 shows the Mössbauer spectrum that results from splitting of the  $^{57}\text{Fe}$  excited state, a quadrupole doublet, for a sample containing randomly oriented molecules such as found in polycrystalline solids or frozen solutions. The two doublets are separated in energy by the quadrupole splitting,  $\Delta E_Q$ , defined by the following equation:

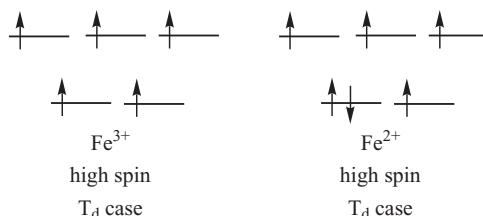
$$\Delta E_Q = \left( \frac{eQV_{zz}}{2} \right) \sqrt{\frac{1+\eta^2}{3}} \quad (3.46)$$

where  $e$  is the proton charge,  $Q$  is the nuclear quadrupole moment,  $V_{zz}$  is the electric field gradient (EFG) tensor (defined along the  $z$  axis by convention), and  $\eta$  is the asymmetry parameter (defined by  $(V_{xx}-V_{yy})/V_{zz}$ ).<sup>34</sup>

Asymmetry in the ligand environment, either geometric or in charge distribution (or both), affect the asymmetry parameter,  $\eta$ . An  $\eta=0$  value corresponds to complete axial symmetry, whereas  $\eta=1$  corresponds to pure rhombic symmetry. Electric monopole interactions between the nuclear charge distributions and the electrons at the nucleus cause a shift of the nuclear ground and excited states. These interactions are known as the isomer shift,  $\delta$ . Both the Mössbauer source and the absorber (the sample of interest) experience an isomer shift, and it is customary to quote  $\delta$  relative to a standard, usually Fe metal or  $\text{Na}_2[\text{Fe}(\text{CN})_5\text{NO}] \cdot 2 \text{H}_2\text{O}$  at 298 K.

### 3.6.2 Quadrupole Splitting and the Isomer Shift

The isomer shift,  $\delta$ , is a measure of the  $s$ -electron density at the iron nucleus. Influences on this parameter include changes in the  $s$  population of a valence shell or shielding effects caused by increasing or decreasing  $p$  or  $d$  electron density. Because the radial distributions of  $s$  and  $d$  electrons overlap,  $\delta$  can be a good measure of iron's oxidation state in the sample. Values for the isomer shift may also yield information on the spin state (several high- or low-spin states exist for,  $\text{Fe(IV)} d^4$ ,  $\text{Fe(III)} d^5$ , or  $\text{Fe(II)} d^6$ ), the coordination sphere of the metal ion, and the degree of covalency in the metal-ligand bond being studied. High-spin  $\text{Fe(II)}$  has a unique range of  $\delta$  values higher than those for low-spin  $\text{Fe(II)}$  and higher than those for high- or low-spin  $\text{Fe(III)}$  or  $\text{Fe(IV)}$ . The trend can be understood through the following reasoning. A decrease in the number of  $d$  electrons with increasing iron oxidation state decreases the screening of  $s$  electron density at the nucleus (increases  $s$  electron density at the nucleus). This, in turn, tends to decrease the value of the isomer shift,  $\delta$ .  $\Delta E_Q$  values depend on the spherical symmetry or asymmetry of the electric field gradients at the nucleus as explained in more detail below. In general, the higher the symmetry of electric field about the nucleus, the smaller the value of  $\Delta E_Q$ . For instance,  $\Delta E_Q(\text{Fe}^{2+}) > \Delta E_Q(\text{Fe}^{3+})$  for the high-spin case because the  $d^5$  high-spin  $\text{Fe}^{3+}$  ( $S=5/2$ ) ion has spherical symmetry and the  $d^6$  high-spin  $\text{Fe}^{2+}$  ( $S=2$ ) does not, as indicated in Figure 3.26. Comparison of  $\Delta E_Q$  values for heme ligands in Table 3.2 illustrates the trend: For  $\text{Fe}^{3+}$ -heme ( $S=5/2$ ) we have  $\Delta E_Q=0.5-1.5$ , while for  $\text{Fe}^{2+}$ -heme ( $S=2$ ) we have  $\Delta E_Q=1.5-3.0$ .



**Figure 3.26** Spherical symmetry and asymmetry of electric field gradients affecting  $\Delta E_Q$ .

Table 3.2 lists values for  $\Delta E_Q$  and  $\delta$  for some important oxidation and spin states found in bioinorganic molecules. Data are taken from reference 34 and from Table 1 of reference 35 for hemoglobin, Hb, myoglobin, Mb, and the picket-fence porphyrin model compound, FeTpivPP(1-MeIm).<sup>35</sup> The myoglobin and hemoglobin model compounds are discussed in Sections 7.2.6 through

**TABLE 3.2 Mössbauer Parameters for Some Biological Species<sup>a</sup>**

Oxidation State	Spin State	Ligands	$\Delta E_Q$ (mm s <sup>-1</sup> )	$\delta$ (mm s <sup>-1</sup> )
Fe(IV)	$S=2$	Fe-(O, N) <sup>b</sup>	0.5–1.0	0.0–0.1
Fe(IV)	$S=1$	Hemes	1.0–2.0	0.0–0.1
		Fe-(O, N)	0.5–4.3	-0.20–0.10
Fe(III)	$S=5/2$	Hemes	0.5–1.5	0.35–0.45
Fe(III)	$S=5/2$	Fe-S <sup>c</sup>	<1.0	0.20–0.35
Fe(III)	$S=5/2$	[FeS <sub>4</sub> ] cluster (rubredoxin)		0.25
Fe(III)	$S=5/2$	[Fe <sub>2</sub> S <sub>2</sub> ] <sup>2+</sup> cluster		0.27
Fe(III)	$S=5/2$	Fe(III) in [Fe <sub>2</sub> S <sub>2</sub> ] <sup>+</sup> cluster		0.30
Fe(III)	$S=5/2$	[Fe <sub>3</sub> S <sub>4</sub> ] <sup>+</sup> cluster		0.27
Fe(III)	$S=5/2$	[Fe <sub>3</sub> S <sub>4</sub> ] <sup>0</sup> cluster		0.32
Fe(III)	$S=5/2$	Fe-(O, N)	0.5–1.5	0.40–0.60
Fe(III)	$S=3/2$	Hemes	3.0–3.6	0.30–0.40
Fe(III)	$S=1/2$	Hemes	1.5–2.5	0.15–0.25
		Hb(O <sub>2</sub> ) <sup>e</sup>	2.19	0.26
		FeTpivPP(1-MeIm) (O <sub>2</sub> ) <sup>d</sup>	2.04	0.27
Fe(III)	$S=1/2$	Fe-(O, N)	2.0–3.0	0.10–0.25
Fe(II)	$S=2$	Hemes	1.5–3.0	0.85–1.0
		Hb <sup>e</sup>	2.22	0.92
		Mb <sup>e</sup>	2.17	0.91
		FeTpivPP(1-MeIm) <sup>d</sup>	2.32	0.88
Fe(II)	$S=2$	Fe-S	0.0–3.0	0.60–0.70
Fe(II)	$S=2$	[FeS <sub>4</sub> ] cluster (rubredoxin)		0.70
Fe(II)	$S=2$	Fe(II) in [Fe <sub>2</sub> S <sub>2</sub> ] <sup>+</sup> cluster		0.72
Fe(II)	$S=0$	Fe-(O, N)	1.0–3.2	1.1–1.3
Fe(II)	$S=0$	Hemes	<1.5	0.30–0.45
Delocalized	$S=1/2$ or $S=9/2$	[Fe <sub>3</sub> S <sub>4</sub> ] <sup>0</sup>		0.46
Fe <sup>2.5+</sup> –Fe <sup>2.5+</sup>	coupled to $S=5/2$ Fe <sup>3+</sup> site			

<sup>a</sup>Typical values at 4.2 K. Isomer shifts are taken relative to the standard value for Fe metal at 298 K.

<sup>b</sup>Fe-(O, N) hexacoordinate or pentacoordinate sites.

<sup>c</sup>Fe-S tetrahedral sulfur ligation.

<sup>d</sup>TpivPP = 5, 10, 15, 20-tetrakis-[*o*-(*p*-ivalamido)phenyl]porphyrinate(<sup>2-</sup>); 1-MeIm = 1-methylimidazole.

<sup>e</sup>Hb = hemoglobin, Hb(O<sub>2</sub>) = dioxygenated hemoglobin, Mb = myoglobin.

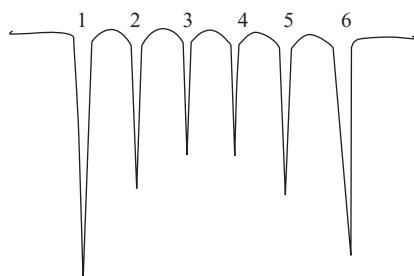
Source: References 34–36.

7.2.7. Reference 36 provides the Table 3.2 data on iron sulfur clusters found in many bioinorganic species.<sup>36</sup>

### 3.6.3 Magnetic Hyperfine Interactions

A nucleus with spin quantum number  $I$  has  $(2I+1)$  magnetic energy levels. The energy gaps between these levels cause splitting in Mössbauer spectra. This effect is known as the *magnetic hyperfine interaction*. At higher temperatures of observation (195 K), the splitting of energy levels may not be observed for paramagnetic species because the electrons may change spins so rapidly that the time-averaged effect is zero. Below a certain temperature, the Curie temperature, thermal agitation is insufficient to prevent the alignment of unpaired electrons so that magnetic hyperfine splitting will be exhibited for the paramagnetic centers. For diamagnetic species ( $S=0$ ), magnetic hyperfine splitting may be induced by the application of an external magnetic field, which is often done experimentally. The diagram in Figure 3.27 illustrates the magnetic hyperfine splittings of the nuclear ground state to the 14.4-keV level (see Figure 3.24) assuming  $\Delta E_Q=0$ . In the presence of a nuclear quadrupole, the hyperfine splittings will be unequally spaced as can be seen in Figure 2b of reference 34. Other splitting patterns are observed for  $S=1/2$  systems that have isotropic magnetic properties; these are illustrated in Figure 3 of reference 34.

For Figure 3.27, note that lines 1, 3, 4, and 6, obey the selection rule  $|\Delta m_I|=1$  for the allowed  $\gamma$  transitions between the nuclear sublevels, whereas lines 2 and 5 obey the  $\Delta m_I=0$  selection rule. For a isotropic ( $g_x=g_y=g_z=g$ ) sample in which the effective magnetic field is parallel to the observed  $\gamma$  rays, the intensity of the  $\Delta m_I=0$  lines vanishes so that only four lines are seen in the spectrum (Figure 3a of reference 34). The same lines that are missing in the isotropic case will be maximized when the effective magnetic field is perpendicular to the  $\gamma$  rays (Figure 3b of reference 34). For a uniaxial case ( $g_x=g_y=0$  and  $g_z\neq 0$ , Figure 3c of reference 34) or the extreme anisotropic case ( $g_z\gg g_x, g_y$ ), the intensities of the absorption lines are independent of the



**Figure 3.27** Magnetic hyperfine splittings of the nuclear ground state assuming  $\Delta E_Q=0$ . (Adapted from Figure 2 of reference 34.)

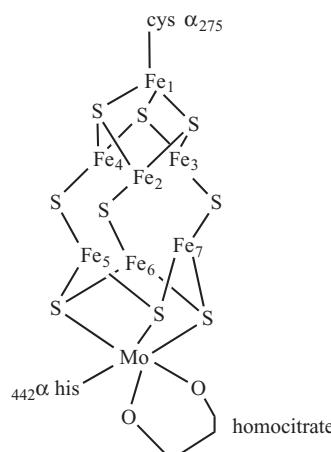
orientation of the applied field. In other words, the parallel versus perpendicular dependence of the Mössbauer spectrum becomes less pronounced as  $g$  becomes more anisotropic ( $g_x \neq g_y \neq g_z \neq 0$ ) and the spectrum looks more like Figure 3.27.

The six-line magnetic hyperfine splitting spectrum seen for the anisotropic cases is averaged over all molecular orientations and will have a 3:2:1:1:2:3 intensity pattern as shown in Figure 3.27. Because  $g$  values are determined by the electron paramagnetic resonance (EPR) technique as discussed above in Section 3.5, one can make reasonable predictions about the shape of the Mössbauer spectrum if the results of an EPR study are known. For instance, a magnetic-field-independent Mössbauer spectrum suggests an EPR silent state; and conversely, a magnetic-field-dependent Mössbauer spectrum implies that an EPR spectrum should be observed.

### 3.6.4 Descriptive Examples

Mössbauer spectroscopy, along with evidence gathered from EPR and ENDOR, has identified the behavior of iron atoms in the M center, FeMo cofactor, of the enzyme nitrogenase from *Azobacter vinelandii*. Figure 3.28 shows a schematic diagram of nitrogenase's M center. At the time of the EPR study, it was known that two M center clusters were contained in each nitrogenase MoFe-protein subunit, that each was of composition  $\text{Mo Fe}_{6-8}\text{S}_{9\pm 1}$ , and that each cluster contained three unpaired electrons in an  $S=3/2$  system.<sup>38</sup> The rhombic EPR signal with  $g$  factors  $g_1 (= g_x) = 4.32$ ,  $g_2 (= g_y) = 3.68$  and  $g_3 (= g_z) = 2.01$  would have the appearance of the rhombic spectrum of Figure 3.22D in Section 3.5.1. (Also see Figure 16 of reference 39.)

The electron-nuclear double resonance (ENDOR) technique mentioned in the previous paragraph is used to study electron–nuclear hyperfine



**Figure 3.28** The FeMo cofactor, M center, of nitrogenase's MoFe protein.

interactions (see Sections 3.5.2 and 3.5.3) that are too small to be resolved within the natural width of the EPR line. Usually, the hyperfine splitting values,  $A$ , detected by ENDOR, range from 2 to 40 MHz. In transition metal complexes and metalloproteins, magnetic nuclei such as  $^1\text{H}$ ,  $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{14}\text{N}$ ,  $^{15}\text{N}$ ,  $^{17}\text{O}$ ,  $^{31}\text{P}$ , and  $^{33}\text{S}$  can be detected by ENDOR as being in the vicinity of paramagnetic metal ions such as high-spin Fe(II) or high- or low-spin Fe(III). In the current example the  $^{57}\text{Fe}$  ENDOR spectra were obtained with protein from *Azobacter vinelandii* grown on  $^{57}\text{Fe}$  enriched media.<sup>37,38</sup> The high-field  $g_3$  and low-field  $g_1$  edges of the EPR spectrum (see previous paragraph) were used to generate the ENDOR spectra assuming that, in different ENDOR experiments, molecules were oriented with their  $g_3$  and  $g_1$  axes approximately parallel to the applied magnetic field.<sup>39</sup> The spectra revealed that there were five distinguishable (inequivalent) iron sites designated  $\text{A}^1$ ,  $\text{A}^2$ ,  $\text{A}^3$ ,  $\text{B}^1$ , and  $\text{B}^2$  for the *Azobacter vinelandii* M center. It is not possible to count the total number of irons in the spectrum using the ENDOR technique, however, subsequent X-ray crystallography<sup>40</sup> indicated seven iron atoms in the M center cluster (five of which are inequivalent).

In reference 41, the authors take the ENDOR data described above and extend its conclusions using Mössbauer spectroscopy. Specifically,  $^{57}\text{Fe}$  enrichment of nitrogenase's M center is used to identify all seven irons present. Between the time of the EPR and ENDOR studies described in the previous paragraphs and the Mössbauer study presented in reference 41, the X-ray crystallographic structure of nitrogenase was published.<sup>40</sup> The X-ray studies revealed that the M center consists of two cuboidal fragments, [Mo–3Fe–3S] and [4Fe–3S], these being linked by three sulfide bridges and attached to the nitrogenase protein via cysteine and a histidine ligands. The M center ( $S=3/2$ ) is further classified as being in the  $\text{M}^{\text{N}}$  state to distinguish it from the diamagnetic  $\text{M}^{\text{OX}}$  ( $S=0$ ) moiety not found *in vivo* and the  $\text{M}^{\text{R}}$  EPR-silent ( $S>1$ ) reduced state found *in vivo* in the presence of nitrogenase's Fe-protein subunit and MgATP. The reference 41 authors identified the seventh Fe site ( $\text{A}^4$ ) as having an unusually small and anisotropic magnetic hyperfine coupling constant,  $A \approx -4$  MHz. They also identified the previously identified (by ENDOR)  $\text{B}^1$  site as representing two equivalent Fe sites having the same hyperfine interactions. The following values for the isomer shifts,  $\delta$ , in mm/s are reported in reference 41 for  $\text{M}^{\text{N}}$ :  $\text{A}^1$ , 0.39;  $\text{A}^2$ , 0.48;  $\text{A}^3$ , 0.39;  $\text{A}^4$ , 0.41;  $\text{B}^1$ , 0.33; and  $\text{B}^2$ , 0.50; this yielded a  $\delta_{\text{avg}} = 0.41$  mm/s. The  $\delta_{\text{avg}} = 0.41$  mm/s for  $\text{M}^{\text{N}}$  differs by only 0.02 mm/s from that found for  $\text{M}^{\text{R}}$ , whereas  $\text{M}^{\text{N}}$  and  $\text{M}^{\text{OX}}$  differ by 0.06 mm/s. Because  $\text{M}^{\text{N}}$  and  $\text{M}^{\text{R}}$  have very similar high-field Mössbauer spectra, the authors believe that the iron oxidation and spin states remain the same in both forms while the molybdenum, Mo, ion becomes reduced in the  $\text{M}^{\text{R}}$  state. The conclusion is supported by the fact that the magnetic hyperfine interactions for  $\text{M}^{\text{R}}$  and  $\text{M}^{\text{N}}$  are quite similar. The  $\delta_{\text{avg}}$  difference for the  $\text{M}^{\text{N}}$  versus the  $\text{M}^{\text{OX}}$  pair led the researchers to conclude that the redox events (leading to nitrogenase's production of ammonia from dinitrogen) center on the iron ions of FeMo cofactor.

Putting together their Mössbauer analysis with previous ENDOR data, EXAFS analyses, and X-ray crystallographic evidence, along with comparison to a model Fe(II) complex with trigonal sulfur coordination, allowed the reference 41 authors to describe the M<sup>N</sup> center as having the oxidation state arrangement of Mo<sup>4+</sup>–3Fe<sup>3+</sup>–4Fe<sup>2+</sup>. As the preceding discussion of nitrogenase metal–sulfur clusters indicate, analysis of complex bioinorganic systems requires the use of multiple analytical techniques and the cooperative exchange of data and ideas of many researchers. A more complete discussion of the complex enzyme nitrogenase is found in Chapter 6 of the first edition of this text.<sup>42</sup>

A more recent X-ray crystallographic study of the nitrogenase MoFe protein at 1.16-Å resolution shows a previously unrecognized ligand coordinated to the six iron ions at the center of the FeMo cofactor (see Figure 3.28).<sup>43</sup> The data are deposited in the protein data bank, PDB, with accession number 1M1N. Previous X-ray crystallographic structures, such as data deposited in the PDB as 2MIN and 3MIN,<sup>40b</sup> were solved at lower resolutions, and they could not show a central ligand because the overwhelming electron density of the iron and sulfur atoms surrounding it hid any electron density due to a light (N, C, O) atom. The reference 43 authors, solving and refining the 1M1N structure, considered carbon, nitrogen, oxygen, and sulfur as plausible central atoms, finally assigning it tentatively as a hexacoordinated nitrogen atom. Interested readers may want to read the perspective on nitrogenase written by D. C. Rees and J. B. Howard and published in late 2006 in the *Proceedings of the National Academy of Sciences, USA*.<sup>44</sup> The article provides an overview of biological nitrogen fixation and introduces three other articles in the same issue that address central aspects of the mechanism and assembly of nitrogenase. Before leaving this topic, this author would like to address problems that arise in accessing data in the protein data bank (PDB). It is important that PDB users recognize the difference between the letter “I” and the number “1” when entering accession numbers into the PDB’s search routine (found at <http://www.rcsb.org/pdb/home/home.do>). For instance, if the user enters 1MIN using the letter “I,” the search routine returns 2MIN, the nitrogenase X-ray crystallographic data that replaced 1MIN in the structure database (see reference 40b). If the user enters 1M1N using the number “1,” the search routine returns information on the nitrogenase structure described in reference 43. The same problem occurs when searching PDB accession numbers containing the letter “O” and/or the number “0.”

### 3.7 OTHER INSTRUMENTAL METHODS

#### 3.7.1 Atomic Force Microscopy

Atomic force microscopy (AFM) is a microscopic method for chemists and biologists offering the magnification range of both the light and electron

microscope, but allowing imaging under the “natural” conditions usually associated with the light microscope. Images can be obtained for samples in air, water, or vacuum with typical resolutions on the order of 10 nm. AFM offers the prospect of high-resolution images of biological material, images of molecules and their interactions even under physiological conditions, and the study of molecular processes in living systems. Applications of AFM in the biosciences include analysis of (1) DNA and RNA, (2) protein–nucleic acid complexes, (3) chromosomes; (4) cellular membranes, (5) proteins and peptides, (6) molecular crystals, (7) biopolymers and biomaterials, and (8) ligand–receptor binding.

The atomic force microscope is one of about two-dozen types of scanned-proximity probe microscopes. All of these microscopes work by measuring a local property—height, optical absorption, or magnetism—with a probe or “tip,” typically made from silicon nitride,  $\text{Si}_3\text{N}_4$ , or elemental silicon, placed very close to the sample. The small probe–sample separation (on the order of the instrument’s resolution) makes it possible to take measurements over a small area. To acquire an image, the microscope raster-scans the probe over the sample while measuring the local property in question. The resulting image resembles an image on a television screen in that both consist of many rows or lines of information placed one above the other. Unlike traditional microscopes, scanned-probe systems do not use lenses, so the size of the probe (rather than diffraction effects) generally limits their resolution.

The concept of resolution in AFM is different from radiation-based microscopies because AFM is a three-dimensional imaging technique. There is an important distinction between images resolved by wave optics and scanning probe techniques. The former is limited by diffraction, whereas the latter is limited primarily by apical probe geometry and sample geometry. Usually the width of a DNA molecule is loosely used as a measure of resolution, because it has a known diameter of 2.0 nm in its B form.

Intermolecular forces largely govern many biological processes—DNA replication, protein synthesis, drug interactions, and others. AFM has the ability to measure these forces, some of which may be in the nanonewton ( $10^{-9}$  newton) range. This makes it possible to quantify molecular interactions in biological systems such as important ligand–receptor interactions. The dynamics of many biological systems depends on the electrical properties of the sample surface, and AFM is able to image and quantify electrical surface charges. In addition to measuring binding and electrostatic forces, the atomic force microscope can also probe the micromechanical properties of biological samples. Specifically, the AFM can observe the elasticity and, in fact, the viscosity of samples ranging from live cells and membranes to bone and cartilage.

The first highly reproducible AFM images of DNA were obtained in 1991. Four major advances that have enabled clear resolution of nucleic acids are (1) control of the local imaging environment including sample modification, (2) TappingMode<sup>TM</sup> scanning techniques, (3) improved AFM probes (such as standard silicon nitride probes modified by electron beam deposition and

oxide-sharpened nanoprobe), and (4) compatible substrates (such as salinized mica and carbon coated mica). The ability to generate nanometer-resolved images of unmodified nucleic acids has broad biological applications. Chromosome mapping, transcription, translation, and small molecule–DNA interactions such as intercalating mutagens provide exciting topics for high-resolution studies. For example, one study looked at the effect of Ni(II) ions on double-stranded, ds, GC-DNA (DNA strands having only guanine and cytosine base pairs, dG–dC·dG–dC) versus the effect on ds AT-DNA (DNA strands having only adenine and thymine base pairs, dA–dT·dA–dT).<sup>45</sup> It is known that many metal ions affect DNA structure, especially a B- to Z-DNA structural transformation in GC-DNA. Group I and II metal cations—Na<sup>+</sup>, K<sup>+</sup>, Ca<sup>2+</sup>, Mg<sup>2+</sup>—and, at much lower concentrations, transition metal divalent cations—Ni(II), Co(II), Zn(II)—are known to cause the B- to Z-DNA transformation. The reference 45 researchers used AFM to detect condensed forms of the two DNA types in the presence of NiCl<sub>2</sub> solutions of differing molarities. They detected condensed forms of GC-DNA—toroids, rods, and jumbles—in solutions with 0.5 M Ni(II) concentrations, whereas AT-DNA only began to form condensed structures at 6 M concentrations of Ni(II). The researchers concluded that the better ability of GC-DNA to form condensed structures was due to the formation of Ni(II)–N7 coordination with major-groove available guanine base N7 positions. In contrast, Ni(II) interaction with AT-DNA appeared to occur only electrostatically in the DNA minor groove and not with coordination positions on the adenine or thymine bases. It is known that the guanine base N7 position is the one most favored by transition metals in DNA binding as is well known for the anticancer drug, *cis*-diamminedichloroplatinum(II), *cis*DDP (see Figure 1.6). The reference 45 researchers also used the “electrostatic zipper model” for DNA aggregation to show that GC-DNA was more likely to “zip” itself into agglomerated structures than was the AT-DNA molecule. They concluded that AFM was an excellent technique for detecting condensed structures in dsDNA.

Researchers have also succeeded in imaging individual proteins and other small molecules with the AFM. Smaller molecules that do not have a high affinity for common AFM substrates have been successfully imaged by employing selective affinity binding procedures. Thiol incorporation at both the 5' and 3' ends of short PCR products (PCR, polymerase chain reaction, described in Section 2.3.5) has been shown to confer a high affinity for ultraflat gold substrates and therefore improved AFM imaging. Small proteins, like cytochrome c (see Section 7.7), have been imaged interacting with their redox partners. For example, Choi and Dimitriadi have used AFM to detect how cytochrome c binds to the lipids in the mitochondrial inner membrane.<sup>46</sup> As discussed in much more detail in Chapter 7, cytochrome c plays an essential role in the respiratory electron transport chain by shuttling electrons between cytochrome c reductase and cytochrome c oxidase. Also, release of cytochrome c from lipid interactions in the mitochondrial inner membrane is believed to be important first step in cell death (apoptosis). Experimentally, the researchers used AFM

to study the adsorption of membrane-associated protein cytochrome c to anionic lipid bilayers of dioleoyl phosphatidyl-glycerol in low ionic strength physiological buffer. The lipids were supported on polylysinated mica, forming stable single lipid bilayers. When low concentrations of cytochrome c were added to the mix, the protein molecules were not topographically visible on the lipid bilayer–buffer interface. However, the forces required to punch through the bilayer by indentation using the atomic force microscopy probe were significantly lower after protein adsorption, suggesting that the protein inserts into the bilayer. Interestingly, the apparent thickness of the bilayer remained unchanged after cytochrome c adsorption as well. The researchers knew that cytochrome c was present in the bilayer during their AFM experiments by confirming cytochrome c's presence using mass spectrometry and UV-visible absorption spectroscopy. The reference 46 researchers concluded that (1) cytochrome c inserts into the bilayer's hydrophobic core, (2) cytochrome c insertion changes the mechanical properties of the bilayer, and (3) atomic force microscopy is a useful tool for investigating lipid–protein interactions.

Cell biologists have applied the AFM's unique capabilities to study the dynamic behavior of living and fixed cells such as red and white blood cells, bacteria, platelets, cardiac myocytes, living renal epithelial cells, and glial cells. AFM imaging of cells usually achieves a resolution of 20–50 nm, not sufficient for resolving membrane proteins but still suitable for imaging other surface features, such as rearrangements of plasma membrane or movement of submembrane filament bundles.

It is informative to compare AFM with other techniques. Scanning tunneling microscopy (STM) is considered the predecessor technique to AFM. The scanning tunneling microscope has better resolution than the atomic force microscope, but the technique can only be applied to conducting samples while AFM can be applied to both conductors and insulators. Compared with scanning electron microscopy (SEM), AFM provides extraordinary topographic contrast, direct height measurements, and unobscured views of surface features (no coating is necessary). Compared with transmission electron microscopes, three-dimensional AFM images are obtained without expensive sample preparation and yield far more complete information than the two-dimensional profiles available from cross-sectioned samples. New approaches in AFM have provided a solid foundation from which research is expanding into more and more complex bioorganic and bioinorganic analyses. Conventional atomic force microscopes suffer from one disadvantage: 1–100 minutes is required to obtain a high-quality image in contrast to the millisecond to 1-minute image times of scanning electron microscopes. High-speed AFM could address this problem; however, commercial high-speed instruments are not currently available. Three innovation areas will address this disadvantage in the near future according to a perspective in *Science* magazine in October 2006<sup>47</sup>: (1) smaller cantilevers with resonant frequencies higher by a factor of 30; (2) increased scanner resonant frequencies with practical ranges of 13 μm

in  $x$  and  $y$  (horizontal) directions and  $4.3\text{ }\mu\text{m}$  in the  $z$  (vertical) direction; and (3) better control systems with increased speed of the feedback loop that controls the height of the AFM tip to maintain minimal imaging force and high image accuracy.

### 3.7.2 Fast and Time-Resolved Methods

**3.7.2.1 Stopped-Flow Kinetic Methods.** Enzyme kinetics happen on very fast time scales; for instance, it is known that the rate of reaction for copper-zinc superoxide dismutase (CuZnSOD),  $\sim 1 \times 10^9\text{ M}^{-1}\text{s}^{-1}$ , approaches the diffusion-controlled rate. Chemists use various methods to study fast reactions. One of the most frequently used rapid kinetic techniques is that of stopped-flow in which the reactants (enzyme and substrate) are rapidly mixed. The lower practical limit for mixing to take place is about  $0.2\text{ ms}$ . The stopped-flow principle of operation allows small volumes of solutions to be driven from high-performance syringes to a high-efficiency mixer just before passing into a measurement flow cell. As the solutions flow through, a steady-state equilibrium is established and the resultant solution is only a few milliseconds old as it passes through the measurement cell. The mixed solution then passes into a stopping syringe, which allows the flow to be instantaneously stopped. Some of the resultant solution will be trapped in the flow cell and as the reaction proceeds, the kinetics can be followed using the appropriate measurement technique. The most common method of following the kinetics is by absorbance or fluorescence spectrometry, and in these cases the measurement cell is an appropriate spectrometer flow cell. Many commercially available absorbance and fluorescence spectrometers may be modified to accept stopped-flow accessories.

In order to use the stopped-flow technique, the reaction under study must have a convenient absorbance or fluorescence that can be measured spectrophotometrically. Another method, called rapid quench or quench-flow, operates for enzymatic systems having no component (reactant or product) that can be spectrally monitored in real time. The quench-flow is a very finely tuned, computer-controlled machine that is designed to mix enzyme and reactants very rapidly to start the enzymatic reaction, and then quench it after a defined time. The time course of the reaction can then be analyzed by electrophoretic methods. The reaction time currently ranges from about  $5\text{ ms}$  to several seconds.

M. Fabian and co-workers have studied the protein's role in internal electron transfer to the catalytic center of cytochrome c oxidase using stopped-flow kinetics.<sup>48</sup> Mitochondrial cytochrome c oxidase, CcO, an enzyme that catalyzes the oxidation of ferrocyanochrome c by dioxygen, is discussed more fully in Section 7.8. In the overall process,  $\text{O}_2$  is reduced to water, requiring the addition of four electrons and four protons to the enzyme's catalytic center. Electrons enter CcO from the cytosolic side, while protons enter from the matrix side of the inner mitochondrial membrane. This redox reaction,

along with a second process known as “proton pumping,” results in the generation of a transmembrane proton gradient driven by the electron transfer (ET) reaction. The reference 48 researchers intended to study proton access to the catalytic site, at heme  $a_3$  and  $Cu_B$  (see Figures 7.39, 7.40, 7.41, and 7.42), that presumably limits the rate of electron transfer to heme  $a_3$  during anaerobic conditions. To do this, two forms of oxidized CcO were studied: (1) the so-called “resting” form (**O** in Figure 7.42) and (2) activated oxidized CcO produced immediately following reoxidation of the fully reduced enzyme with  $O_2$ —the so-called **O<sub>H</sub>** metastable state. The reduction kinetics of oxidized CcO (**O**) by a hexaruthenium complex (Ru) were measured under anaerobic conditions (Ar gas) using a stopped-flow apparatus. An anaerobic solution of (**O**) was mixed in a 1:1 ratio with the mixture of Ru and sodium dithionite (Ru-DT)—final concentrations of 2.5 mM Ru and 5 mM DT. The researchers also compared the kinetics of electron transfer (ET) to heme  $a_3$  of CcO in the (**O**) and (**O<sub>H</sub>**) states at pH 9.0 and 8.0. For measurement of (**O<sub>H</sub>**) anaerobic, the procedure mixed, in the stopped-flow apparatus, fully reduced CcO (in 5 mM Ru and 20 mM dithionite, DT) and air-saturated buffers maintaining the desired pH. In the approximately 2-ms dead time of the stopped-flow apparatus, CcO was oxidized with any excess  $O_2$  present consumed by the dithionite, both  $Cu_A$  and heme  $a$  components of CcO are reduced by the excess reducing agent, all within the dead time of the apparatus, so that only re-reduction of heme  $a_3$  was available to be monitored. This fact was demonstrated experimentally by monitoring the UV-visible absorbances of the mixtures at 1 ms (within the dead time of the stopped-flow apparatus) and at 1.9 s, at which time maxima at 446 and 604 nm correspond to the fully reduced CcO. The large increase in the Soret region absorbance band at 446 nm, which takes place in the interval between 1 ms and 1.9 s, is characteristic of heme  $a_3$ ’s reduction exclusively. These experiments isolate the kinetics of electron transfer to heme  $a_3$ . The researchers found that ET to heme  $a_3$  was controlled by the state of ionization of a single amino acid residue with a  $pK_a$  of  $6.5 \pm 0.2$ . This  $pK_a$  was attributed to glu60, located on the entrance to the so-called matrix side “K” channel (see Section 7.8 and Figure 7.41), a proton channel distinguished by important lysine (K) residues. Thus glu60 was thought to control proton entry into the channel and therefore into the catalytic site by coupled proton transfer. Additionally, a second factor was thought to be the rate of proton diffusion in the channel. The researchers also concluded, by experiments with both (**O**) and (**O<sub>H</sub>**) states of CcO, that rates of ET to heme  $a_3$  in the resting enzyme (**O**) and in CcO activated by reaction of the fully reduced enzyme (**O<sub>H</sub>**) with  $O_2$  were the same. This implied that the catalytic sites of the two forms of the enzyme, (**O**) and (**O<sub>H</sub>**), are essentially identical.

**3.7.2.2 Flash Photolysis.** Time-resolved spectroscopy techniques are a powerful means of studying materials, giving information about the nature of the excitations, energy transfer, molecular motion, and molecular environment, information that is not available from steady-state measurements. It is

a rapidly advancing field with applications in many areas of science and technology. Flash photolysis allows one to follow a reaction using fast (nanosecond to microsecond) laser excitation pulses to cause absorption in the species of interest. Following the excitation, one must use fast electronic devices to measure the light emission or absorption by the species of interest. A necessary criterium for the use of flash photolysis methods is that the molecule under study must show a detectable change upon laser excitation.

An example of experimental data collected using flash photolysis, along with conclusions reached as a result of the experiments, involves electron transfer between heme a and heme  $a_3$  in cytochrome c oxidase, CcO.<sup>49</sup> The experimental setup relies on the fact that reverse electron flow between the spatially close heme a and  $a_3$  sites can be initiated by flash photolysis of carbon monoxide (CO) from reduced heme  $a_3$  (heme contains a Fe(II) ion) under conditions when heme a is initially oxidized (heme contains a Fe(III) ion). See Section 7.8 for more information about CcO and Figure 7.41 for a view of the heme sites within the protein. The researchers followed the reaction using transient absorption spectroscopy, with femtosecond (fs,  $10^{-15}$  s) time resolution and a time window extending to 4 nanoseconds (ns,  $10^{-9}$  s). Experimentally, transient absorption pump probe spectroscopy was performed with a 55-fs pump pulse centered at 590 nm and a <30-fs white light continuum probe pulse. The probe continuum was generated by using the fundamental beam of the laser system, centered at ~615 nm. Absorbance changes after CO photolysis were recorded for fully reduced (FR) CcO enzyme—heme  $a_3$  and heme a both having Fe(II) ions—and for mixed-valent (MV) enzyme—heme  $a_3$  having Fe(II) and heme a having Fe(III). The experimental results showed that there were significant spectral changes after photolysis for the MV state, whereas for the FR state the spectrum hardly changed. This implied significant differences in the behavior of the hemes for the different redox states. The ultraviolet-visible absorption spectra of CcO were monitored for spectral evolution at the so-called Soret (~430–440 nm) and the  $\alpha$  (~590–605 nm) bands for CcO. Comparison of picosecond (ps,  $10^{-12}$  s) heme  $a_3$ -CO photodissociation showed significant spectral interaction between the hemes. The most important finding from the experiments were twofold: (1) The electron equilibration between heme  $a_3$  and heme a occurred in  $1.2 \pm 0.1$  ns, and (2) electron equilibration corresponded to a  $\Delta G^0$  of 45–55 meV. The data suggest very fast equilibration between the two hemes and demonstrate a low driving force (small  $\Delta G^0$ ) for the redox reaction. Figure 5 of reference 49 summarizes the findings for the MV CcO complex: (1) Flash photolysis removes CO from heme  $a_3$ , transferring it to Cu<sub>B</sub>; (2) the effective midpoint potential of heme  $a_3$  rises after removal of CO, bringing it closer to heme a's midpoint potential; (3) electron redistribution based on this new equilibrium (with small energy difference between the two hemes) occurs within 1.2 ns; and (4) CO is released from Cu<sub>B</sub> and migrates out of the catalytic site within 3  $\mu$ s, a previously measured time scale, then believed to be the fastest electron transfer reaction for CcO. The previous experiments were carried out using microsecond instrumentation,<sup>50</sup>

and the reference 49 authors believe that their faster electron transfer using femtosecond instrumentation may be more characteristic of the true reaction rate.

**3.7.2.3 Time-Resolved Crystallography.** Time-resolved crystallography (TC) uses an intense synchrotron X-ray source and Laue data collection techniques to greatly reduce crystallographic exposure times. Normal time resolution for X-ray crystallography has been in the range of seconds or tens of seconds. TC has the potential to take snapshots of protein structural changes on a nanosecond time scale. Consequently, multiple exposures may be taken that capture the evolution of the crystallographic unit cell as it reacts over time. Traditionally, crystallographers have applied several techniques to obtain detailed structural information on reaction intermediates. The most common approach has been to design a series of stable structures that mimic normally short-lived intermediates. However, these structures are stable precisely because they are not identical to the intermediates they seek to mimic, and key interactions are usually missing. Other experimental techniques and chemical intuition are called upon to supply the missing information, sometimes with only limited success. One successful attempt to understand how the attachment and release of carbon monoxide, and ultimately dioxygen, happens on a molecular scale is described in Section 7.2.7. Rodgers and Spiro studied the nanosecond dynamics of the R to T transition in hemoglobin.<sup>51</sup> Using pulse-probe Raman spectroscopy, with probe excitation at 230 nm, these workers were able to model the R–T interconversion of the hemoglobin molecule as it moved from the R state (HbCO) to the T state (Hb).

Time-resolved crystallography (TC), now has the potential to offer detailed structural information on short-lived intermediates in macromolecular reactions under near-physiological, crystalline conditions, and this aids elucidation of the underlying molecular mechanisms. Interpretation of TC data has been hindered, in part due to the difficulty in extracting structural information on intermediates from time-resolved electron density maps. Under certain assumptions, these maps are weighted averages of the electron density maps of the different structural species present at the experimental time points. That is, these time-dependent electron density maps are structurally heterogeneous. K. Moffat has published techniques for interpreting these maps and continues to publish reviews updating progress in this quick-changing and expanding field.<sup>52</sup>

In their 1996 *Science* article “Photolysis of the Carbon Monoxide Complex of Myoglobin: Nanosecond Time-Resolved Crystallography,”<sup>53</sup> Moffat, Wulff, and co-workers described the nanosecond time resolution of structural changes that occur in the carbon monoxide complex of myoglobin (MbCO) at room temperature on CO photodissociation by a nanosecond laser pulse. Myoglobin (Mb) is the dioxygen-binding protein found in muscle. It contains the same Fe(II)/Fe(III) heme cofactor as hemoglobin (Mb). Myoglobin has long been used by researchers to study ligand-binding ( $O_2$  versus CO versus NO,

for instance) and conformational changes accompanying ligand-binding and relaxation. MbCO is the most stable and easily photolyzed of the three possible systems—MbCO, MbO<sub>2</sub>, MbNO—and is the most popular model system. (See Section 7.2.7 for discussion of CO binding to myoglobin.) The Fe–CO bond was broken with a 10-ns laser pulse and X-ray data sets were collected at different time delays between the laser flash and the X-ray pulse (4 ns, 1  $\mu$ s, 7.5  $\mu$ s, 50  $\mu$ s, and 1.9 ms). Although the difference maps clearly showed release of the CO molecule from the heme, they also suggested that CO recombination in the crystal form contains a fast, geminate phase with a recombination rate comparable with or greater than the maximum photolysis rate applied by the laser pulse of  $10^9\text{ s}^{-1}$ . This result confirmed that it is much more difficult to photolyze MbCO molecules in the crystal form than in solution. A second prominent feature of the X-ray difference maps arose from the motion of the iron atom out of the heme plane and toward the proximal histidine ligand. A third feature indicated a transient “docking site” for the photodissociated CO; however, well-populated docking sites indicating CO exit from the binding pocket were not identified. A number of small electron density features indicated structural rearrangements of aa residues surrounding the heme, especially the residues of the E and F helices implicated by other methods in heme and protein relaxation effects, and in iron ion displacement in or out of the heme plane. Their data suggested complete iron displacement and heme relaxation occurred in <4 ns, in agreement with other spectroscopic results.

In 2001, the same research group published the article “Protein Conformational Relaxation and Ligand Migration in Myoglobin: A Nanosecond to Millisecond Molecular Movie from Time-Resolved Laue X-Ray Diffraction.”<sup>54</sup> Using the same techniques in the same system (MbCO), the researchers photodissociated the CO ligand by a 7.5-ns laser pulse and also probed the subsequent structural changes by 150-ps or 1- $\mu$ s X-ray pulses at 14 laser/X-ray delay times, ranging from 1 ns to 1.9 ms. Very fast heme and protein relaxation involving the E and F helices was evident from the data at a 1-ns time delay. The photodissociated CO molecules were detected at two locations: at a distal pocket docking site and at the Xe 1 binding site in the proximal pocket. The so-called Xe 1 site is the highest-affinity Xe binding site that is almost fully occupied when metMb (Mb with Fe(III) bound in the heme ligand) are equilibrated with seven atmospheres of Xe gas. Researchers believe that the Xe 1 site identifies a possible proximal site on the dioxygen exit pathway from myoglobin. The population by CO of the primary, distal site peaks at a 1-ns time delay and decays to half the peak value in 70 ns. The secondary, proximal docking site reaches its highest occupancy of 20% at approximately 100 ns and has a half-life of approximately 10  $\mu$ s. At approximately 100 ns, all CO molecules are accounted for within the protein, in one of these two docking sites or bound to the iron ion in the heme. Thereafter, the CO molecules migrate to the solvent from which they rebind to deoxymyoglobin in a bimolecular process with a second-order rate coefficient of  $4.5 \times 10^5\text{ M}^{-1}\text{ s}^{-1}$ . The results demonstrated that structural changes as small as 0.2 Å and populations of CO

docking sites of 10% could be detected by time-resolved X-ray diffraction. The GIF-formatted movie illustrating the CO photodissociation and movement to the proximal pocket is available from a link in the HTML online version of this *Biochemistry* 2001 article.

Schotte and co-workers demonstrated 150-ps time-resolved X-ray crystallography of the MbCO mutant system (leu29 replaced by phe, L29F) in 2003.<sup>55</sup> Their technique characterized the L29F myoglobin structure as it evolved from the carboxy to the deoxy state. One can access a QuickTime™ version of the movie from the supplemental material accompanying the online *Science* article. The movie is described in the supplemental materials in the following manner. A moving average of the time-resolved electron density maps was constructed from the image with laser off at time points: 100ps, 316ps, 1ns, 3.16ns, 31.6ns, 316ns, and 3.16μs (each frame represents a weighted average of the electron density from adjacent time points). The numerical time indicator advances at the midpoint between adjacent time points. The yellow circles enclose electron density likely arising from CO; the solid lines switch to dashed lines when CO departs from that site. Figure 4 of reference 55 contains an informative legend that decodes what is happening during the movie. The reference 55 authors believe that this work extends the time resolution of experimental crystallography into the same time domain available in molecular dynamics simulations (see Section 4.3.2). In the future, they believe that X-ray free electron lasers, now under development, will deliver intense X-ray pulses shorter than 100fs, allowing further experimental study of dynamic enzyme systems. More recent experimental results for MbCO systems studied by time-resolved crystallography can be found in reference 56.

### 3.7.3 Mass Spectrometry

Mass spectrometry (MS) is probably a familiar tool to chemistry and biology students as a technique commonly used to measure the molecular mass of a sample. Often, MS is used in tandem with other techniques for chromatic separation of the sample before mass measurement. Some common hyphenated techniques include: HPLC-MS, high-pressure liquid chromatography coupled to MS; GC-MS, gas chromatography coupled to MS; or CE-MS, capillary electrophoresis coupled to MS.

For large molecules such as biomolecules, molecular masses can usually be measured to an accuracy of 0.01% of the total molecular mass—that is, within 4Da for a sample of 40kDa. Structural information on biomolecules can be gained using tandem mass spectrometers—the MS–MS techniques. One fragments the sample inside the instrument and analyzes the products generated in the second mass spectrometer. This technique is used for protein, peptide, and oligonucleotide sequencing, as will be described in more detail below.

Mass spectrometry studies on proteins can determine the purity of the sample, verify amino acid substitutions in mutants, detect post-translational modifications, or calculate the number of disulfide bridges. Amino acid

sequences in proteins and nucleotide sequences in oligonucleotides can be determined. Protein folding can be monitored by hydrogen/deuterium (H/D) exchange. The principle applied to protein folding analyses is that backbone amide nitrogens undergo rapid H/D exchange when they are exposed to solvent (unfolded domains) but much slower exchange when buried in a hydrophobic protein domain (folded). For instance, researchers have designed methods for monitoring solvent accessibility of protein–ligand and protein–protein interfaces using mass spectrometry.<sup>57</sup> The H/D exchange reaction must be carried out with the isolated proteins as well as with the protein complex in the protein–protein interface experiment. The exchanging regions for the isolated versus the combined proteins are then compared. If a region is buried by protein–protein binding, the amides in this region will be protected and exchange slowly, revealing the interaction interface. Subsequently, the data can be used as input for a computer docking simulation to build a computer model of the protein–protein complex.

The instrumentation necessary for mass spectroscopic analysis of biomolecules does not differ substantially from that used for small organic molecules. The mass spectrometer must have the same three fundamental parts: ionization source, analyzer, and detector. However, the parts may be more sophisticated and different from the normal LC–MS or GC–MS spectrometer. For biomolecule analyses, the ionization source usually will be one of two types: (1) electrospray, known as ESI; or (2) matrix-assisted laser desorption, known as MALDI. The ESI source is well-suited to analysis of polar molecules of molecular mass 0.1–10<sup>3</sup> kDa. Nanospray ionization is a low-flow-rate version of ESI that allows use of small sample sizes and multiple experiments with limited sample supplies. MALDI is based on bombardment of sample molecules with a laser to bring about sample ionization. In this method, the sample is mixed with an excess of a highly absorbing matrix compound. The matrix transforms the laser energy into excitation energy that is then transferred to the sample molecules. Once the sample molecules have been ionized, they are transferred into the analyzer region of the spectrometer where they are separated according to their mass (*m*)-to-charge (*z*) ratio (*m/z*). (See below for more details about the *m/z* ratio.) Some mass analyzers currently used in biomolecular applications include (1) quadrupole, (2) time-of-flight (TOF) magnetic sector, (4) Fourier transform quadrupole ion trap, and (5) Fourier transform ion cyclotron resonance (FT-ICR). Quadrupole analyzers routinely analyze samples with *m/z* ratios up to 3000. This is useful for the ESI source as this ionization method produces ions with *m/a* ratios of <3000. The TOF analyzer depends on mass, with the smallest masses reaching the detector first. TOF is used for kinetic studies, for fast reactions, and for methods interfaced with chromatographs for sample separation preceding mass spectrometry. The magnetic sector analyzer depends on changes in voltage (and thus the magnetic field surrounding the individual ions). Each of the mass analyzers mentioned here has a different range of coverage or accuracy for *m/z* ratios, molecular mass determination, and achievable resolution. All analyzers

mentioned here may be interfaced with ESI, whereas MALDI is not usually coupled to a quadrupole analyzer. More term definitions can be found at [http://www.ibiblio.org/pub/academic/chemistry/iupac/Download/publications/analytical\\_compendium/Cha12sec2.pdf](http://www.ibiblio.org/pub/academic/chemistry/iupac/Download/publications/analytical_compendium/Cha12sec2.pdf). The mass spectrometer's detector monitors ion current, amplifies it, then transmits it to the data system where it is recorded as a mass spectrum. Normally, the  $m/z$  values are plotted against their intensities, showing the number of components in the sample, the molecular mass of each component and its relative abundance. The common detectors are the photomultiplier, the electron multiplier, and the micro-channel plate.

The  $m/z$  ratio for biomolecules, as they ionize and pass through the mass spectrometer's analyzer, depends on a number of factors. For instance, using ESI, samples with molecular masses ( $M$ ) up to  $\sim 1.2\text{ kDa}$  produce singly charged molecular ions, either  $(M + H)^+$  in positive ionization mode or  $(M - H)^-$  in negative ionization mode. In the singly charged case, the  $m/z$  value corresponds to the molecular mass. Proteins and peptides analyze better in positive ionization mode (facilitated by addition of a small amount of acid, usually formic acid), whereas oligonucleotides analyze better in negative ionization mode. In positive ionization mode,  $(M + \text{Na})^+$ ,  $(M + \text{K})^+$ , and  $(M + \text{NH}_4)^+$  ions may also be produced in which cases the molecular mass is higher than expected by 23, 39, and 18 atomic mass units, amu, respectively. A common negative ionization mode ion includes that having the chloride ion,  $\text{Cl}^-$ , +35 amu. Samples having molecular weights greater than  $\sim 1.2\text{ kDa}$  give rise to multiply charged molecular-related ions such as  $(M + nH)^{n+}$  in positive ionization mode and  $(M - nH)^{n-}$  in negative ionization mode. The mass-to-charge ratio,  $m/z$ , will be expressed as

$$m/z = \frac{(MW + nH^+)}{n} \quad (3.47)$$

where  $m/z$  is the mass-to-charge ratio marked on the mass spectrum's abscissa, MW is the molecular mass of the sample,  $n$  is the integer charges on the ions, and H is amu of a proton, 1.008 Da.

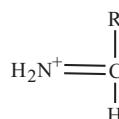
Usually the number of charges on an ion will not be known, but it can be calculated using a formula based on two different ions appearing in the spectrum. Actually, the molecular mass of a sample can be calculated automatically, or semiautomatically, by the processing software associated with the mass spectrometer. Experimentally, the automatic calculation of molecular mass is very helpful because a complex peptide or protein mixture will display an  $m/z$  spectrum with several overlapping series of multiply charged ions.

Peptide and protein sequencing and its importance in the proteomics field were discussed in Section 2.2.3. The following gives a brief description of the mass spectrometry methods used to achieve sequencing. First, to produce protein or oligonucleotide structural/sequence information by mass spectrometric techniques, one needs to use tandem mass spectrometry (MS-MS). In this technique, a sample is first fragmented and analyzed in one mass spec-

trometer, followed by fragmentation of specific sample ions inside a second spectrometer. The process results in identification of the fragment ions with the goal of identifying one large protein by its primary sequence or components of a complex protein mixture. MS–MS also enables specific compounds to be detected in complex mixtures because of their specific and characteristic fragmentation patterns. Usually, a tandem mass spectrometer will contain two analyzers, with the following common combinations: (1) quadrupole–quadrupole; (2) magnetic sector–quadrupole; (3) magnetic sector–magnetic sector; and (4) quadrupole–time-of-flight (TOF). In the peptide sequencing experiment, the first analyzer is used to select user-specified sample ions arising from a particular component, usually the  $(M + H)^+$  ions. The chosen ions pass into a collision cell where they are bombarded by gas molecules that cause the fragment ions to form. The process is known as collision-induced dissociation (CID). The second analyzer then separates the fragment ions according to their  $m/z$  ratios. Polypeptides may fragment along their side chains or along the polypeptide backbone. High-energy collisions, such as in a magnetic sector–magnetic sector MS–MS, will result in many different types of side-chain cleavage. Quadrupole–quadrupole or quadrupole–TOF mass spectrometers generate low-energy fragmentations with fewer side-chain fragmentations. Polypeptide backbone bonds may fragment at the backbone NH–CH, CH–CO, or CO–NH bonds, resulting in ions called *a*, *b*, or *c* if they have the charge retained on the N-terminal fragment or the *x*, *y*, or *z* ions if they have the charge retained on the C-terminal fragment. The most common cleavage site occurs at the CO–NH bonds, giving rise to “*b*” or “*y*” ions. Fragments containing more than one amino acid residue will be identified by subscripts; that is,  $b_2$  is an N-terminal fragment with two aa residues, and  $y_4$  is a C-terminal fragment of four aa residues. The mass difference between two adjacent “*b*” ions or “*y*” ions is indicative of a particular amino acid residue. Additionally, the so-called immonium ion is produced for many amino acid residues in fragmenting peptides, and these often have distinctive mass spectrum signatures as well (see Figure 3.29).

Tables have been produced that indicate the most common molecular masses found for amino acid residues in mass spectrometric analyses.

A protein identification study might proceed in the following manner. First, the protein is analyzed by mass spectrometry to determine its molecular mass to within 0.01%. Second, the protein is digested with an enzyme, commonly trypsin. The enzyme trypsin cleaves polypeptide chains at points following lysine and arginine residues. Using this proteolytic enzyme ensures that each



**Figure 3.29** The Immonium Ion.

fragment contains lys or arg, so that all fragments are suitable for positive ionization mass spectrometric analysis. Peptides of ~2.5 kDa or less in molecular mass produce the most useful data. Third, the digest mixture is analyzed by MS, producing a complex spectrum—a peptide map—from which the molecular masses of all the proteolytic fragments can be read. At this point, the researcher may compare results to a peptide map database and possibly confirm the protein's identity from this comparison. Up to this point, the tandem mass spectrometer has been operating in the MS mode. If the protein has not been identified at this point, the operator switches to the MS–MS mode. Using the quadrupole–TOF mass spectrometer as a model, the identification proceeds as follows. The digest mixture continues to spray into the mass spectrometer. The protonated molecular fragments of each digest fragment are independently selected and transmitted through the quadrupole analyzer, now being used to submit only the ions of interest to the collision cell lying between the first and second mass analyzers. The fragment—daughter or product—ions produced by the collision cell are then analyzed by the TOF analyzer. The MS–MS spectrum produced shows all the fragment ions arising directly from the precursor ions for a given peptide component. MS–MS spectra are produced for each component in the proteolytic digest. At this point, experience in interpretation and processing of the results becomes very important to accurately identify the protein(s) in the sample. Some peptides from the proteolytic digest may be completely sequenced, while others may yield only a partial sequence of four to five amino acid residues. Often, this sequence “tag” of four to five amino acids may be sufficient to search a protein database and confirm the protein's identity. Only about 4 μL of solution is required for the digest mixture analysis if the solution concentration is sufficient—that is, concentration of the original protein is ~1–10 pmol/L. More information on the techniques and results can be found at many websites. Two of these are: <http://www.astbury.leeds.ac.uk/facil/MStut/mstutorial.htm> and [http://mascot.proteomics.ucdavis.edu/mascot/help/fragmentation\\_help.html](http://mascot.proteomics.ucdavis.edu/mascot/help/fragmentation_help.html).

Small proteins and peptides can be sequenced using MALDI–TOF–MS with a “ladder” sequencing technique.<sup>58</sup> First, the researcher fragments the protein or peptide using either a time-dependent or concentration-dependent chemical degradation beginning from either the N- or C-terminus of the protein or peptide. The resulting protein or peptide fragments differ from each other by one amino acid residue. The mixture is mass analyzed in a single MALDI–TOF–MS experiment with mass differences between adjacent mass spectral peaks corresponding to a specific amino acid residue. This type of analysis can be thought of as simply determining the masses of a series of peptides or proteins that are present in a single MALDI sample. The order of occurrence in the mass spectrum defines the sequence of amino acids in the original protein or peptide.

As discussed in Section 2.2.3, proteomic investigations require powerful analytical techniques such as those continuing to be developed using mass spectrometry. As Aebersold and Mann declare in a 2003 review of mass spec-

trometry-based proteomics: “The mass analyzer is, literally and figuratively, central to the technology.”<sup>59</sup> From the perspective of proteomic and systems biology analyses, MS–MS spectrometry offers sensitivity, resolution, mass accuracy, and the ability to generate information-rich mass spectra from peptide fragments. The interested reader will find a wealth of information on applying mass spectrometry to proteomics in this publication.

In an accompanying article in *Nature* in 2003, researchers discuss some of the long-term challenges of proteomics and describe technologies and strategies, both in place and emerging, that rise to these challenges.<sup>60</sup> One method, intended to detect protein–protein interactions, involves the use of fluorescence resonance energy transfer (FRET) between fluorescent tags on interacting proteins. FRET is a nonradiative process whereby energy from an excited donor fluorophore is transferred to an acceptor fluorophore that is within ~60 Å of the excited fluorophore. After excitation of the first fluorophore, FRET is detected either by emission from the second fluorophore using appropriate filters, or by alteration of the fluorescence lifetime of the donor. According to the reference 60 authors, the potential of FRET is considerable, for two reasons. First, it can be used to make measurements in living cells, which allows the detection of protein interactions at the location in the cell where they normally occur, in the presence of the normal cellular milieu. Second, transient interactions can be followed with high temporal resolution in single cells. The reference 60 authors include several examples of FRET analyses in this publication. They conclude that FRET methods are becoming practical adjuncts in many proteomic arenas and believe, in addition to the above-mentioned applications, that FRET could be used to monitor post-translational modifications and other cellular behaviors of interest.

### 3.8 SUMMARY AND CONCLUSIONS

This chapter has provided an introduction to some instrumental methods used in bioinorganic chemistry with emphasis on methods referred to in later chapters. The structural methods of X-ray crystallography (solid state) and NMR (in solution) have proved indispensable to researchers study complex bioinorganic systems. Mössbauer spectroscopy is an invaluable aid for researchers studying bioinorganic systems containing iron. Electron paramagnetic resonance methods provide information on the many bioinorganic systems containing unpaired electrons. Fast and time-resolved methods help bioinorganic chemists identify short-lived catalytic intermediates and, thus, better understand the catalytic cycles of metalloenzymes. Mass spectrometry experiments have been developed to find and confirm many properties of proteins, including how it folds into its final, native conformation.

The descriptions in this chapter have attempted to give students some idea of the scope and complexity of instrumental techniques available to the bioinorganic chemist. It has not been intended to be either comprehensive or overly

theoretical in presentation. Students are encouraged to acquaint themselves further with the theory and practice of instrumental techniques, especially those that are important to their particular research interests.

## REFERENCES

1. (a) Que, L., ed. *Physical Methods in Bioinorganic Chemistry: Spectroscopy and Magnetism*, University Science Books, Sausalito, CA, 2000. (b) Solomon, E. I.; Hodgson, K. O., eds. *Spectroscopic Methods in Bioinorganic Chemistry*, Oxford University Press, Oxford, 1998.
2. Skoog, D. A.; Holler, F. J.; Nieman, T. A. *Principles of Instrumental Analysis*, 5th ed., W. B. Saunders, Philadelphia, 1998.
3. Cowan, J. A. *Inorganic Biochemistry, An Introduction*, 2nd ed., Wiley-VCH, New York, 1997.
4. Scott, R. A. In Que, L., ed., *Physical Methods in Bioinorganic Chemistry: Spectroscopy and Magnetism*. University Science Books, Sausalito, CA, 2000, pp. 465–503.
5. Kau, L.-S.; Spira-Solomon, D. J.; Penner-Hahn, J. E.; Hodgson, K. O.; Solomon, E. I. *J. Am. Chem. Soc.*, 1987, **88**, 595–598.
6. Penner-Hahn, J. E.; McMurry, T. J.; Renner, M. W.; Latos-Grazynski, L.; Eble, K. S.; Davis, I. M.; Balch, A. L.; Groves, J. T.; Dawson, J. H.; Hodgson, K. O. *J. Biol. Chem.*, 1983, **258**, 12761–12764.
7. Bukowski, M. R.; Koehntop, K. D.; Stubna, A.; Bominaar, E. L.; Halfen, J. A.; Münck, E.; Nam, W.; Que, L., Jr. *Science*, 2005, **310**, 1000–1002.
8. Del Rio, D.; Sarangi, R.; Chufán, E. E.; Karlin, K. D.; Hedman, B.; Hodgson, K. O.; Solomon, E. I. *J. Am. Chem. Soc.*, 2005, **127**(34), 11969–11978.
9. Nobbs, C. L.; Watson, H. C.; Kendrew, J. C. *Nature*, 1966, **209**, 339.
10. Fermi, G.; Perutz, M. F.; Shaanan, B.; Fourme, R. *J. Mol. Biol.*, 1984, **159**, 175.
11. Drenth, J. *Principles of Protein X-Ray Crystallography*, 2nd ed., Springer-Verlag, New York, 1999.
12. (a) Gilliland, G. L.; Ladner, J. E. *Curr. Opin. Struct. Biol.*, 1996, **6**, 595–603.  
(b) Gilliland, G. L. *Methods Enzymol.*, 1997, 277, 546–556.
13. Hahn, T., ed. *International Tables for Crystallography*, Vol. A, D. Reidel, Dordrecht, 1993.
14. Matthews, B. W. *J. Mol. Biol.*, 1968, **33**, 491–497.
15. Jameson, G. B.; Molinaro, F. S.; Ibers, J. A.; Collman, J. P.; Brauman, J. I.; Rose, E.; Suslick, K. S. *J. Am. Chem. Soc.*, 1980, **102**(9), 3224–3237.
16. Jameson, G. B.; Rodley, G. A.; Robinson, W. T.; Gagne, R. R.; Reed, C. A.; Collman, J. P. *Inorg. Chem.*, 1978, **17**, 850–857.
17. (a) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Crystallogr.*, 1993, **26**, 283–291. (b) Laskowski, R. A.; Moss, D. S.; Thornton, J. M. *J. Mol. Biol.*, 1993, **231**, 1049–1067. (c) MacArthur, M. W.; Laskowski, R. A.; Thornton, J. M. *Curr. Opin. Struct. Biol.*, 1994, **4**, 731–737. (d) Otwinowski, A. In Sawyer, L.; Issacs, N.; Bailey, S., eds. *Data Collection and Processing*, SERC Daresbury Laboratory,

- Daresbury, U.K. 1993, pp. 56–62. (e) Brünger, A. R.; Kuriyan, J.; Karplus, M. *Science*, 1987, **235**, 458–460.
18. (a) Hauptman, H. A., Xu, H. *Acta Crystallogr. D Biol. Crystallogr.*, 2006, **62**, 897–900. (b) Usón, I.; Sheldrick, G. M. *Curr. Opin. Struct. Biol.*, 1999, **9**, 643–648.
19. Toyoshima, C.; Nakasako, M.; Nomura, H.; Ogawa, H. *Nature*, 2000, **405**, 647–655.
20. Wilson, M. A.; Brunger, A. T. *J. Mol. Biol.*, 2000, **301**, 1237–1258.
21. Akitt, J. W. *NMR and Chemistry: An Introduction to Modern NMR Spectroscopy*, 3rd ed., Chapman and Hall, London, 1992.
22. Salzmann, M.; Pervushin, K.; Wider, G.; Senn, H.; Wüthrich, K. *J. Am. Chem. Soc.*, 2000, **122**, 7543–7548.
23. Clore, G. M.; Gronenborn, A. M. *Proc. Natl. Acad. Sci.*, 1998, **95**(11), 5891–5898.
24. Sivakolundu, S. G.; Mabrouk, P. A. *J. Biol. Inorg. Chem.*, 2003, **8**, 527–539. (PDB: 1LC1, 1LC2)
25. (a) Güntert, P.; Mumenthaler, C.; Wüthrich, K. *J. Mol. Biol.*, 1997, **273**, 283–298. (b) Güntert, P.; Braun, W.; Wüthrich, K. *J. Mol. Biol.*, 1991, **217**, 517–530.
26. Banci, L.; Bertini, I.; Huber, J. G.; Spyroulias, G. A.; Turano, P. *J. Biol. Inorg. Chem.*, 1999, **4**, 21–31. (PDB: 1GIW)
27. (a) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M. Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. *J. Comput. Chem.*, 2005, **26**, 1668–1688. (b) Ponder, J. W.; Case, D. A. *Adv. Prot. Chem.*, 2003, **66**, 27–85.
28. Bertini, I.; Calderone, V.; Cosenza, M.; Fragai, M.; Lee, Y.-M.; Luchinat, C.; Mangani, S.; Terni, B.; Turano, P. *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 5334–5339. (PDB: 1RMZ, 1YCM)
29. Palmer, G. In Que, L., ed., *Physical Methods in Bioinorganic Chemistry: Spectroscopy and Magnetism*, University Science Books, Sausalito, CA, 2000, 121–185.
30. (a) Kroneck, P. M.; Antholine, W. E.; Kastrau, D. H.; Buse, G.; Steffens, G. C.; Zumft, W. G. *FEBS Lett.*, 1990, **268**(1), 274–276. (b) Antholine, W. E.; Kastrau, D. H. W.; Steffens, G. C. M.; Buse, G.; Zumft, W. G.; Kroneck, P. M. H. *Eur. J. Biochem.*, 1992, **209**, 875–881.
31. Farrar, J. A.; Neese, F.; Lappalainen, P.; Kroneck, P. M. H.; Saraste, M.; Zumft, W. G.; Thomson, A. *J. J. Am. Chem. Soc.*, 1996, **118**, 11501–11514.
32. Epel, B.; Slutter, C. S.; Neese, R.; Kroneck, P. M. H.; Zumft, W. G.; Pecht, I.; Farver, O.; Lu, Y.; Goldfarb, D. *J. Am. Chem. Soc.*, 2002, **124**, 8152–8162.
33. Drago, R. S. *Physical Methods in Chemistry*, W. B. Saunders, Philadelphia, 1977.
34. Münck, E. In Que, L., ed. *Physical Methods in Bioinorganic Chemistry: Spectroscopy and Magnetism*, University Science Books, Sausalito, CA, 2000, pp. 287–319.
35. Suslick, K. S.; Reinert, T. *J. J. Chem. Ed.*, 1985, **62**, 974–983.
36. Beinert, H.; Holm, R. H.; Münck, E. *Science*, 1997, **277**, 653–659.
37. Venters, R. A.; Nelson, M. J.; McLean, P. A.; True, A. E.; Levy, M. A.; Hoffman, B. M.; Orme-Johnson, W. H. *J. Am. Chem. Soc.*, 1986, **108**, 3487–3498.
38. True, A. E.; Nelson, M. J.; Venters, R. A.; Orme-Johnson, W. H.; Hoffman, B. M. *J. Am. Chem. Soc.*, 1988, **110**, 1935–1943.
39. Chasteen, N. D.; Snetsinger, P. A. In Que, L., ed., *Physical Methods in Bioinorganic Chemistry: Spectroscopy and Magnetism*, University Science Books, Sausalito, CA, 2000, pp. 187–231.

40. (a) Chan, M. K.; Kim, J.; Rees, D. C. *Science*, 1993, **260**, 792–794. (PDB: 1MIN) (b) Peters, J. W.; Stowell, M. H. B.; Soltis, S. M.; Finnegan, M. G.; Johnson, M. K.; Rees, D. C. *Biochemistry*, 1997, **36**, 1181–1187. (PDB: 3MIN, 2MIN, replaces 1MIN)
41. Yoo, S. J.; Angove, H. C.; Papaefthymiou, V.; Burgess, B. K.; Münck, E. *J. Am. Chem. Soc.*, 2000, **122**, 4926–4936.
42. Roat-Malone, R. M. *Bioinorganic Chemistry: A Short Course* 2002, Wiley-Interscience, John Wiley & Sons, Hoboken, NJ.
43. Einsel, O.; Tezcan, F. A.; Andrade, S. L. A.; Schmid, B.; Yoshida, M.; Howard, J. B.; Rees, D. C. *Science*, 2002, **297**, 1696–1700. (PDB: 1M1N)
44. Howard, J. B.; Rees, D. C. *Proc. Natl. Acad. Sci. USA*, 2006, **103**(46), 17088–17093.
45. Sitko, J. C.; Mateescu, E. M.; Hansma, H. G. *Biophys. J.*, 2003, **84**(1), 419–431.
46. Choi, E. J.; Dimitriadis, E. K. *Biophys. J.*, 2004, **87**(5), 3234–3241.
47. Hansma, P. K.; Schitter, G.; Fantner, G. E.; Prater, C. *Science*, 2006, **314**, 601–602.
48. Antalik, M.; Jancura, D.; Palmer, G.; Fabian, M. *Biochemistry*, 2005, **44**, 14881–14889.
49. Pilet, E.; Jasaitis, A.; Liebl, U.; Vos, M. H. *Proc. Natl. Acad. Sci. USA*, 2004, **101**(46), 16198–16203.
50. (a) Oliveberg, M.; Malmström, B. G. *Biochemistry*, 1991, **30**, 7053–7057. (b) Verkhovsky, M. I.; Morgan, J. E.; Wikström, M. *Biochemistry*, 1992, **31**, 11860–11863.
51. Rodgers, K. R.; Spiro, T. G. *Science*, 1994, **265**, 1697–1699.
52. (a) Moffat, K. *Faraday Discuss.*, 2003, **122**, 65–77 and 79–88. (b) Moffat, K. *Chem. Rev.*, 2001, **101**(6), 1569–1581.
53. Srajer, V.; Teng, T.-Y.; Ursby, T.; Pradervand, D.; Ren, Z.; Adachi, S.-I.; Schildkamp, W.; Bourgeois, D.; Wulff, M.; Moffat, K. *Science*, 1996, **274**, 1726–1729.
54. Srajer, V.; Ren, Z.; Teng, T. Y.; Schmidt, M.; Ursby, T.; Bourgeois, D.; Pradervand, C.; Schildkamp, W. *Biochemistry*, 2001, **40**, 13802–13815.
55. Schotte, F.; Lim, M.; Jackson, T. A.; Smirnov, A. V.; Soman, J.; Olson, J. S.; Phillips, G. N. Jr; Wulff, M.; Anfinrud, P. A. *Science*, 2003, **300**, 1944–1947.
56. (a) Aranda, R., 4th; Levin, E. J.; Schotte, F.; Anfinrud, P. A.; Phillips, G. N.; Jr. *Acta Crystallogr. D Biol. Crystallogr.*, 2006, **62**(Pt 7), 776–783. (b) Bourgeois, D.; Vallone, B.; Arcovito, A.; Sciara, G.; Schotte, F.; Anfinrud, P. A.; Brunori, M. *Proc. Natl. Acad. Sci. USA*, 2006, **103**(13), 4924–4929. (c) Schmidt, M.; Nienhaus, K.; Pahl, R.; Krasselt, A.; Anderson, S.; Parak, F.; Nienhaus, G. U.; Srajer, V. *Proc. Natl. Acad. Sci. USA*, 2005, **102**(33), 11704–11709.
57. Mandell, J. G.; Baerga-Ortiz, A.; Falick, A. M.; Komives, E. A. *Methods Mol. Biol.*, 2005, **305**, 65–80.
58. Patterson, D. H.; Tarr, G. E.; Regnier, F. E.; Martin, S. A. *Anal. Chem.*, 1995, **67**, 3971–3978.
59. Aebersold, R.; Mann, M. *Nature*, 2003, **422**, 198–207.
60. Phizicky, E.; Bastiaens, P. I. H.; Zhu, H.; Snyder, M.; Fields, S. *Nature*, 2003, **422**, 208–215.

---

# 4

---

## COMPUTER HARDWARE, SOFTWARE, AND COMPUTATIONAL CHEMISTRY METHODS

### 4.1 INTRODUCTION TO COMPUTER-BASED METHODS

Chemists use computers for many purposes. As the chapter on instrumental methods (Chapter 3) has illustrated, every modern analytical instrument must include a computer interface. Chemical structure drawing, visualization, and modeling programs are important computer-supported applications required in academic, industrial, and governmental educational and research enterprises. Computational chemistry has allowed practicing chemists to predict molecular structures of known and theoretical compounds and to design and test new compounds on computers rather than at the laboratory bench.

### 4.2 COMPUTER HARDWARE

A basic review of computers and computing is given in reference 1. A short summary of this introductory material is presented here.

The basic elements of a computer are:

1. The central processor unit (CPU) that does the work
2. Memory locations where programs are controlled and results are stored

3. Input and output (I/O) devices to communicate with other computers or devices
4. Buses that provide communication between various computer elements

Computers all contain read-only memory whose contents are permanent—that is, can only be read and not written to by the user, along with random access memory that can both be read from and written to by the user. The basic computing unit is a bit (b), which stands for binary digit; 8 bits comprise a byte (B). Table 4.1 illustrates calculation of computer memory bytes—that is, the number of locations that can be addressed.

Computer storage capacities are constantly increasing to keep pace with requirements for data storage and manipulation; however, a recent model desktop or laptop computer will have 250 gigabytes (GB) of storage capability or more. Microcomputers, familiar to all as PC and Macintosh desktop or laptop computers, are the basic interface for many instrumental methods including all types of chromatography (LC, HPLC, GC), hyphenated methods (GC-MS, LC-IR), flame emission spectroscopy, and basic ultraviolet-visible and infrared spectroscopy. Microcomputers have operating systems (OS) such as Windows and Unix or Linux variants for the PC or have Mac OS X, Linux, and BSD (Berkeley Software Distribution or Berkeley Unix) for Apple Macintosh computers.

Many instrument manufacturers will have individualized dedicated operating systems, based on Windows, Unix, Linux, or Sun systems that interface with their instrumentation. Computer and instrument operators should become familiar with more than one OS and expect them to continually undergo change and upgrading. Microcomputers capable of performing the various analog-to-digital or digital-to-analog conversions, ADC or DAC respectively, in addition to PIO (parallel input output) functions are normally purchased as an integral part of the analytical instrument in question. Upgrades to software are to be expected during the instrument's lifetime, usually requiring hardware upgrades as well. Many molecular design, molecular modeling, visualization, and computational programs will run on high-end microcomputers as discussed in Sections 4.3 and 4.4.

Workstations, a step up in capability and cost from microcomputers, usually feature a multitasking operating system (OS) enabling the computer to run

**TABLE 4.1 Computing Units**

Power of 10	Number and Unit	Number of Locations (bytes)
$2^{10}$	1024 = 1 kilobyte	1024
$2^{20}$	$1024 \text{ k} = 1 \text{ M}$ (megabyte)	$1,048,576$ ( $1.045 \times 10^6$ )
$2^{30}$	$1024 \text{ M} = 1 \text{ G}$ (gigabyte)	$1,073,741,824$ ( $1.07 \times 10^9$ )
$2^{40}$	$1024 \text{ G} = 1 \text{ T}$ (terabyte)	$(1.099511628 \times 10^{12})$

more than one operation simultaneously. Workstations will feature at least 64 MB (megabytes) of RAM, built-in network support, and a graphical user interface (GUI). More sophisticated instrumentation such as high-field NMR require workstations for data acquisition, manipulation, and display of one-, two-, and three-dimensional NMR spectra. Low-end workstations may be powerful PCs and Macs, while high-end models are available from companies such as Silicon Graphics, Sun, Hewlett Packard, IBM, and so on. More sophisticated graphics, molecular modeling, molecular design, and analysis software such as SYBYL™, SYBYL CSCORE™, and UNITY™ components sold by Tripos, Inc. and discussed below in Section 4.5 require workstation capability for full usability.

Minicomputers are medium-sized machines usually having multiple users. They have one or more CPUs working together and large amounts of both RAM and storage. The OS will be some form of the Unix operating system. Most of the time, the chemist will interface through a PC or workstation with the minicomputer acting as a server. Manufacturers of these machines are Sun, Silicon Graphics, or Hewlett Packard, among others.

Mainframes are large computers comprised of a cluster of tightly coupled machines or having multiple processors. These units will often be set up for specific applications as database servers, or for handling calculations such as those generated by quantum mechanics-based computational chemistry methods.

Supercomputers often have hundreds of linked high-speed processors and very large memories. Very large calculations are performed across many processors at once, and the so-called parallel processing is very efficient for manipulating large mathematical data arrays. Supercomputers often have a minicomputer front end providing the user interface, and only the largest computing jobs are actually run on the supercomputer. These machines are produced by IBM, Intel, Cray Research, Silicon Graphics, Fujitsu, NEC, and Hitachi companies, among others.

A relatively new computer architecture called cluster or parallel computing is comprised of groups of small machines hooked together. The clusters can be assembled from off-the-shelf PCs, making their cost low. Chemists are particularly interested because they are useful for increasingly computing-intensive modeling and simulation projects. Chemists are engaged, alongside computing experts, in modifying chemistry software but also designing general tools to make clusters and their ilk run better. Bioinformatic researchers are interested because of the huge computing needs brought about by developments in genetics and proteomics. A recent book; *Parallel Computing for Bioinformatics and Computational Biology*, edited by Albert Zomaya, will be of interest to those in this field.<sup>2</sup>

Computer clusters contain a number of processors put together on a motherboard into a unit known as a node. The nodes are then hooked together with other boards via high-speed communications networks. Nodes can be hard-wired into supercomputers produced by companies like IBM and Compaq or

they can be composed of individual small PC's defining the low-end cluster. Next-generation powerful low-end and high-end computing systems may be cluster-type or parallel species. Another developing computer cluster architecture might place different types of clusters together so that one part of the cluster does super fast calculations (needed for ab initio computational chemistry methods) and another part carries out molecular dynamics simulations that require fast graphics engines. Either type of cluster needs an extremely fast communications network connecting the components. This, in turn, leads to the need for reformulated modeling software so that it is usable on the cluster-type computers. Many types of computer clusters have been developed, one of which is the so-called high-performance cluster (HPC). High-performance clusters provide increased performance by splitting a computational task across many different nodes (the individual computers) in the cluster. One of the most popular HPC implementations is a cluster with nodes running Linux as the open source operating system (OS). Free software is available to implement the parallelism in a system that may encompass from tens to hundreds of machines. HPC clusters are optimized for workloads that require jobs or processes happening on the separate cluster computer nodes to communicate actively during the computation. These include computations where intermediate results from one node's calculations will affect future calculations on other nodes.

Another high-performance computer architecture is called a PC grid. The PC grid operates by tapping the downtime in idle desktop computers across hundreds of thousands of computers. One such project, under the supervision of Professor Vijay Pande at Stanford University's Chemistry Department, is called Folding@home (FAH). This distributed computing project performs computationally intensive simulations of protein folding and other molecular dynamics simulations. The goal of the project is to understand protein folding, misfolding, and related diseases. For example, it is known that when proteins misfold, there can be serious consequences, including many well-known diseases such as Alzheimer's, Mad Cow (BSE), CJD (Creutzfeldt–Jakob disease, a rare and fatal neurodegenerative disease of unknown cause), ALS (amyotrophic lateral sclerosis or Lou Gehrig's Disease, a progressive disorder of the nervous system), Huntington's, Parkinson's, and many cancers and cancer-related syndromes. Any computer user can join this distributed computing project by downloading software found at Folding@home. As of January 2007, 597,500 non-anonymous donors had registered almost two million CPUs (central processing units) in the project.

## 4.3 MOLECULAR MODELING AND MOLECULAR MECHANICS

### 4.3.1 Introduction to MM

In the Summer 2001 issue of the ACS newspaper "Chemistry" computational chemistry or "cheminformatics" was described in the following manner. "Computational chemistry and quantum chemistry have enlisted the computer and

software in an entirely new kind of experimental methodology. Computational chemists, for example, don't study matter directly. In the past, chemists who wanted to determine molecular properties chose their instrumentation, prepared a sample, observed the reactions of the sample, and deduced the molecule's properties. Computational chemists now choose their computer and software packages and get their information by modeling and mathematical analyses." This portion of Chapter 4 will provide an introduction to this fast-growing and changing area of chemistry.

Theoretical models include those based on classical (Newtonian) mechanical methods—force field methods known as molecular mechanical methods. These include MM2, MM3, Amber, Sybyl, UFF, and others described in the following paragraphs. These methods are based on Hook's law describing the parabolic potential for the stretching of a chemical bond, van der Waal's interactions, electrostatics, and other forces described more fully below. The combination assembled into the force field is parameterized based on fitting to experimental data. One can treat 1500–2500 atom systems by molecular mechanical methods. Only this method is treated in detail in this text. Other theoretical models are based on quantum mechanical methods. These include:

1. Hückel and extended Hückel treatments that consider many fundamental ideas of orbitals and bonding but neglect electron repulsions in molecules. These methods treat systems up to 200 atoms.
2. Semiempirical (CNDO, MNDO, ZINDO, AM1, PM3, PM3(tm), and other) methods based on the Hartree–Fock self-consistent field (HF-SCF) model, which treats valence electrons only and contains approximations to simplify (and shorten the time of) calculations. Semiempirical methods are parameterized to fit experimental results, and the PM3(tm) method treats transition metals. These methods treat systems of up to 200 atoms.
3. Ab initio (nonempirical, from "first principles") methods also use the HF-SCF model but include all electrons and use minimal approximation. Basis sets of functions based on linear combinations of atomic orbitals (LCAO) increase in complexity from the simplest (STO-3G) to more complex (3-21G(\*)) to extended basis sets (6-311+G\*\*) for the most accurate (and most time-consuming) results. Ab initio methods treat systems up to 50 atoms.
4. Density functional methods treat larger molecules more successfully than ab initio methods and make use of explicit atomic basis sets as described for ab initio methods.

### 4.3.2 Molecular Modeling, Molecular Mechanics, and Molecular Dynamics

*Molecular modeling* seeks to answer questions about molecular properties—stabilities, reactivities, electronic properties—as they are related to molecular

structure. The visualization and analysis of such structures, their molecular properties, and their molecular interactions are based on some theoretical means for predicting the structures and properties of molecules and complexes. If an algorithm can be developed to calculate a structure with a given stoichiometry and connectivity, one can then attempt to compute properties based on calculated molecular structure and vice versa.

*Molecular mechanics* is defined as the calculation of the molecular structure and corresponding strain energy by minimization of the total energy. The energies of various minima are calculated using functions that relate internal coordinates to energy values. Molecular mechanics methods can provide excellent descriptions of equilibrium geometries and conformations, but do not supply thermochemical information. They cannot yield acceptable results outside the range of their parameterization. This important point, repeated and amplified below, says that a force field parameter set (also defined below), assembled for one coordination complex or bioinorganic molecule, may not be applied to other molecules unless some means of testing the parameter set (comparison of calculated bond distances and bond angles to those known experimentally from an X-ray crystallographic structure, for instance) is applied.

An excellent introduction to molecular modeling of inorganic compounds has been provided by Comba and Hambley in their 1995 book *Molecular Modeling of Inorganic Compounds*.<sup>3a</sup> The book has been updated by a new enlarged edition containing an interactive tutorial on CD-ROM. The newer edition includes descriptions of calculations of stereoselective interactions of metal complexes with biomolecules.<sup>3b</sup> Additionally, the authors have developed a inorganic compound-oriented molecular modeling system called MOMEc, which has been designed as an add-on to the HyperChem™ drawing and modeling program discussed below in Section 4.5. The reference 3 authors discuss the application of molecular mechanics to coordination and organometallic compounds, to inorganic compounds involved in catalysis (including stereospecific catalysts), to design of new metal-based drugs, and to the interaction of metal ions with biological macromolecules such as proteins or DNA. In all cases, modeling of transition metal species is complicated by the number of different metals of interest as well as the variety of coordination numbers, geometries, and electronic states they may adopt. Other pitfalls abound. If a calculation is performed on the wrong electronic state of a molecule (singlet rather than triplet oxygen for example), any energy minimization result will be incorrect as well. An incorrect starting geometry (the wrong conformer for instance) may yield the wrong energy minimum. The geometry optimization method defined may not produce an accurate geometric result.

In spite of its limitations, molecular mechanics (MM) is a technique that is widely used for the computation of molecular structures and relative stabilities. The advantage of MM over quantum mechanical methods is mainly based on the computational simplicity of empirical force field calculations, leading to a comparatively small computational effort for MM calculations. Therefore,

even large molecular assemblies with hundreds of atoms are tractable by MM, and energy surfaces with many minima may be screened successfully. MM has been used routinely for many years in organic chemistry applications. Studies on inorganic systems have begun more recently because the effects of variable coordination numbers, geometries (square planar, tetrahedral, octahedral), oxidation and spin states, and electronic influences based on partly filled *d* subshells (such as Jahn–Teller effects) have been difficult to model with conventional MM approaches.

In reference 3, Comba and Hambley introduce their topic in three parts: (1) basic concepts of molecular mechanics, (2) applications of the techniques and difficulties encountered, and (3) a guide to molecular modeling of a new system. Only the introductory section of reference 3a is summarized here.

In molecular mechanics calculations, the arrangement of the electrons is assumed to be fixed and the positions of the nuclei are calculated. Bonded atoms are treated as if they are held together by mechanical springs (vibrational frequencies), and nonbonded interactions are assembled from van der Waals attractive and repulsive forces (gas compressibility data). Finding suitable values for van der Waals parameters presents one of the greatest problems in force field development. Fortunately, van der Waals forces have less influence on the final molecular geometry than do bond-stretch and angle-bend parameters. The typical equations and parameters simulating the various interactions that describe the potential energy surface of a molecule are assembled from (1) a function that quantifies the strain present in all bonds, (2) a bond angle function, (3) a function that calculates all the dihedral strain, and (4) a number of nonbonded terms. To use the equations to calculate the total strain in the molecule, one needs to know (1) the force constants (*k*) for all the bonds and bond angles in the molecule, (2) all the ideal bond lengths (*r*) and bond angles (*q*), (3) the periodicity of the dihedral angles (*n*) and the barriers to their rotation (*V*), (4) the van der Waals parameters  $A_{ij}$  and  $B_{ij}$  between the *i*th and *j*th atoms to simulate the nonbonded van der Waals interactions, and (5) the point charges  $q_i$  and  $q_j$  and the effective dielectric constant ( $\epsilon$ ) to model the electrostatic potential.<sup>4</sup> In inorganic molecular mechanics, these parameters are empirically derived usually by fitting a number of crystal structures and are derived for the specific force field used. Parameters are transferable from one molecule to another (within limits described below) but are not transferable between force fields. A “force field” is defined as a collection of numbers that parameterize the potential energy functions. These functions include the force constants, ideal bond distances and angles, and parameters for van der Waals, electrostatic, and other terms.

To optimize the geometry of a molecule, the total energy from all forces is minimized by computational methods. This so-called “strain energy” is related to the molecule’s potential energy and stability. Because parameters such as bond length, bond angles, and torsional angles used to derive the strain energies are fitted quantities based on experimental data such as X-ray crystallographic structures, molecular mechanics is often referred to as empirical force

field calculations or just force field calculations. Because force field parameters are dependent on the potential energy functions, the entire set of functions and parameters are often referred to as “the force field.” In essence, the *molecular mechanics* method interpolates the structure and strain energy of an unknown molecule from a series of similar molecules with known structure and properties. As mentioned above, *molecular modeling* of transition metal compounds is complicated by the partially filled *d* orbitals of the metal ions. However, the structure of a coordination compound, and thus its thermodynamics, reactivity, and electronic behavior, is strongly influenced by the ligand structure. Because empirical force field calculations have been very successful in modeling organic ligands, it follows that these techniques should be extendable to transition metal coordination complexes as well. The problem remains that *molecular mechanics* interpolates the structure of an unknown inorganic complex from a set of parameters derived from fitting a number of X-ray crystallographic structures. If the complex being investigated differs significantly from the structures used to determine the parameters, poor results will be obtained.

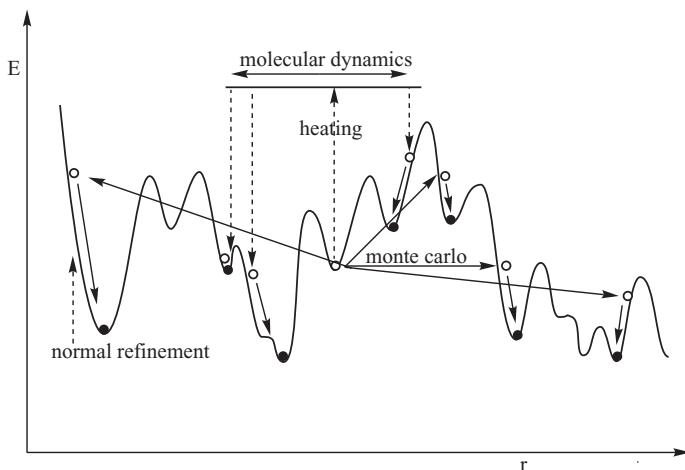
Specific *molecular mechanics* (MM) force fields have been assembled, and several of these are mentioned here. The MM2 *molecular mechanics* force field and its applications to inorganic chemistry are described in reference 5a,b and references therein.<sup>5</sup> The MM3 force field, a third-generation refinement, is one used extensively in molecular mechanical computations.<sup>6</sup> For applications in inorganic chemistry see reference 7. Information on the MM2 and MM3 force fields is available on the website <http://europa.chem.uga.edu/>. The MMFF94 Merck molecular force field is described on the website <http://server.ccl.net/cca/data/MMFF94/> and in reference 8. MMFF94 achieves MM3-like accuracy for small molecules, yet is as applicable to proteins and other systems of biological significance. The MMX force field<sup>9</sup> has been used in many inorganic applications, and one student exercise using MMX implemented in PCMOD-EL<sup>TM</sup> (Serena Software, Inc.; see Section 4.5) is described in reference 10. Many other force fields have been described. A few of these are as follows: MOLBLD<sup>11</sup>; MOLMECH, a molecular mechanics program containing an extended MM2 force field<sup>12</sup>; and SHAPES.<sup>13</sup> UFF, a molecular mechanics generic force field has been used in inorganic chemistry applications.<sup>14</sup> MOMECK97 is a molecular mechanics program adapted to HyperChem<sup>TM</sup> as described below in Section 4.5. More information is available at Professor Comba’s website <http://www.uni-heidelberg.de/institute/fak12/AC/comba/>. The current MOMECK force field is described on this website.

The reference 4 authors discuss criteria that should be applied when describing molecules with these molecular mechanics programs. Some of these are as follows: (1) Check the error file for interactions not in the parameter set, because some programs will assign a force constant of zero to unrecognized atom types; (2) check all interactions generating >5 kJ/mol of strain to determine, for instance, whether that bond or angle really is that strained or whether there is a parameterization or molecular structure problem; and (3) check the

hybridization about the central atom since a commercial program not written for inorganic molecules may not recognize metal-ligand bonds. In summary, the significant advantage of *molecular mechanics* calculations is that they are relatively rapid; however, caution is needed in interpretation of molecular mechanical results produced by empirical force field calculations. A good analogy is to that of the neural network; that is, the molecular mechanical result is completely dependent on the facts the system has been taught.

The types of input and output accepted by the computer program as well as algorithms used to achieve energy minimization are described here briefly. Generally, an energy minimization routine produces an optimized structure (conformer) most closely related to the input coordinates. That is, the routine falls into the *closest* energy minimum, which may or may not be the global energy minimum of a system. One carries out a conformational analysis by deriving the energy surface that results from changing a particular rotational value, often a set of dihedral angles. Conformations are then submitted to energy minimizing routines that include: (1) simplex (only the potential energy function is used), (2) gradient or first derivative methods (steepest descent), (3) conjugate-gradient methods where the history of the search can influence direction and step size (Fletcher-Reeves, Polak-Ribière), (4) second-derivative methods (Newton-Raphson and block-diagonal Newton-Raphson), and (5) least squares methods (Marquardt). In applying *molecular mechanics* methods, particularly for those structures that are predominant in solution, it is important to find the lowest-energy structure, that is, the global energy minimum. Several methods exist for carrying out the minimization, but only the torsional Monte Carlo method will be discussed here. The Monte Carlo method is a stochastic search in which the variation from one starting conformation to another is limited in magnitude—for instance, by limiting the starting geometries to those that conform to some energy requirement (perhaps 15 kJ/mol from the energy minimum). Using the internal coordinate of torsional angle causes significant differences in possible structure conformations—in contrast to changes in bond distances or bond angles, which do not. In each Monte Carlo step, a random number of torsional angles are varied by a random amount generating a new starting geometry that can be minimized.

*Molecular dynamics* involves the calculation of the time-dependent movement of each atom of a molecule, achieved by solving and applying Newton's equations of motion. Structures for starting geometries are sampled as a function of time or geometry during a molecular dynamics run of a few nanoseconds. Usually, an elevated temperature is used to favor faster and more complete molecular dynamics searching. A good conformational search systematically (deterministically) scans the entire potential energy surface, generates starting geometries, and then minimizes them. A combination of the above-described methods—deterministic, stochastic, and molecular dynamics, illustrated in Figure 4.1—screens the entire potential energy surface and produces the best possible results.



**Figure 4.1** Conformational search procedures. (Adapted with permission from Figure 4.2 of reference 3a. VCH Publishers, Inc., New York and VCH Verlagsgesellschaft mbH, Weinheim, Federal Republic of Germany. Copyright 1995, VCH, Verlagsgesellschaft mbH, Weinheim, Federal Republic of Germany.)

### 4.3.3 Biomolecule Modeling

The modeling of biomolecules is discussed in reference 15, which provides good background on protein folding, secondary protein structure prediction, sequence alignment for finding comparable natural or synthetic proteins, and modeling by homology to produce three-dimensional protein structures in advance of their experimental elucidation. The last topic makes use of the Ramachandran map that places protein structure into allowed conformational regions for amino acids. Fitting of the modeled protein into allowed regions— $\alpha$ -helical ( $\alpha$ ),  $\beta$ -strand ( $\beta$ ) and left-handed helical (L)—indicates that the protein conforms to known protein secondary structural constraints. Reference 15 includes a short discussion of theoretical investigations of enzyme mechanism, useful for understanding the molecular basis of enzyme activity. Karplus and Petsko have provided a review of the application of molecular dynamics to biological problems.<sup>16</sup> Bioinorganic applications and examples of molecular modeling studies discussed in reference 3 often are limited by the lack of suitable small molecule analogs to the metal ion environments found in metalloproteins. Two examples are the grossly distorted tetrahedral type I Cu(II) sites in blue copper proteins and three-coordinate iron sites found in the enzyme nitrogenase. Models for metalloporphyrin (hemoglobin, myoglobin, cytochrome) active sites have been derived by modeling small molecules and extending the AMBER<sup>17</sup> and MM2<sup>18</sup> force-field parameterization schemes. Learn more about AMBER at the website <http://amber.scripps.edu/>. A molecular dynamics simulation of hydrogen peroxide binding to the heme iron in cytochrome c peroxidase has been modeled using a modified AMBER force

field.<sup>19</sup> The CHARMM force field (more information at <http://yuri.harvard.edu/>) was used in a study of the temperature dependence of both the structure and the internal dynamics of (carbonmonoxy)myoglobin.<sup>20</sup> The geometry about the iron center was assumed to be octahedral. Analysis of the movements of the iron center with respect to the heme group indicated that the largest-amplitude motions were perpendicular to the heme plane.

A major problem in modeling of large biomolecules arises from the flexibility of proteins and DNA with corresponding numbers of adoptable geometries. The experimental system modeled is therefore only one possible representation of the many possible geometries. A second difficulty is that the accuracy of molecular mechanics models of biomolecules is substantially lower than that of small molecules. Large numbers of independent parameters are needed, but there are few experimentally known structures and these are often of low precision. Electrostatic considerations and solvent effects cause further limitations. The real value in molecular modeling of macromolecular systems, as stated by the reference 3a authors, emerges when the models make predictions that can be tested experimentally. Qualitatively, the models can be used to visualize molecules whose structures are not accessible by any other means.

In Part 3 of reference 3a, the authors give much needed advice on developing a force field, taking into account bond length deformation, valence angle deformation, torsion angle deformation, out-of-plane deformation, van der Waals and electrostatic interactions, and hydrogen bonding interactions. All of these parameters are interrelated, and modification of one must lead to further testing of all. They reiterate that the force field parameter set must model the molecule under consideration as accurately as possible and that results of the calculation should be compared to experimental data whenever this is available. The authors then discuss carrying out the calculation, first listing the important considerations. These include: having an adequate starting model, choosing an appropriate energy minimization method, and considering the probability that there are many possible energy minima.

#### 4.3.4 A Molecular Modeling Descriptive Example

Ubiquinol/cytochrome c oxidoreductase–cytochrome bc<sub>1</sub> complex, a protein found in electron transport chain of mitochondrial membranes, bacteria, and chloroplasts, is the subject of Section 7.6. X-ray crystallographic structures of the native protein, and those with inhibitors in place, are discussed in Section 7.6.2. One set of crystallographic studies, carried out by the E. A. Berry group,<sup>21</sup> (PDB: 1BCC, native, and 3BCC, stigmatellin bound) showed substantial movement of the soluble head of the Rieske iron–sulfur protein (ISP), containing an [Fe<sub>2</sub>S<sub>2</sub>] cluster, between reaction domains in cytochrome b and cytochrome c<sub>1</sub> protein subunits. In a 1999 publication, results of steered molecular dynamics (SMD) simulations of the ISP soluble domain of native and stigmatellin bound cytochrome bc<sub>1</sub> were reported.<sup>22</sup> The SMD technique overcomes a time-scale limitation (of a few nanoseconds) for molecular dynamics simulations of

large proteins. In the simulations, a solvated bc<sub>1</sub> complex in a phospholipid layer simulating a membrane was constructed, having a total of 206,720 atoms. A subset of 91,061 atoms was actually included, with 45,131 moving atoms, in an SMD simulation. First, an algorithm was assembled to compare two known conformations of the ISP protein in relatively rigid domains. Rotations, based on the  $\alpha$ -carbon atoms of the ISP in its two forms (native and stigmatellin bound), characterize the domain movements relative to each other. (See Figure 3 of reference 22.) The algorithm identified a nearly rigid-body rotation of the water-soluble portion of the ISP (amino acid residues 73–196 of PDB: 1BCC and 3BCC) with respect to the ISP's transmembrane helix (aa residues 1–67). The reference 22 molecular dynamics simulations were performed with the parallel molecular dynamics program NAMD<sup>23</sup> using the CHARMM22 force field.<sup>24</sup>

Before the SMD simulation can be carried out, the researchers must build a complex cytochrome bc<sub>1</sub> model. The next paragraphs describe this process. First, the reference 22 authors generated charge distribution parameters for the cytochrome bc<sub>1</sub> heme cofactors. They assumed that all three hemes in the cytochrome bc<sub>1</sub> complex were in the oxidized state and carried the same point charge distribution. The carboxyl chains of the hemes (see Figure 7.25) were assumed to be deprotonated with a net charge of -2, so that the oxidized state heme carries a -1 charge, or in the reduced (Fe(II)) state, a -2 charge. The distributions of point charges on the oxidized and reduced hemes were calculated using the program GAUSSIAN-94<sup>25</sup> at the Hartree–Fock level with a 6-311G basis set. A radius of 1.22 Å was used for the Fe atom in electrostatic potential fitting calculations. The reference 22 authors compared resulting charge distributions on several groups of heme atoms for deprotonated and protonated hemes in the reduced state, and these in turn allowed them to identify changes in charge distribution for protonation of the oxidized heme state. Continued calculation and manipulation of heme charge distribution resulted in calculated charges for each heme atom, collected in reference 22's Table 1. The authors stipulate that their calculation of heme charge distribution was adequate for the SMD simulations conducted in their work, but added that geometry optimization and more structural information (coordinates of the histidine coordinating to the heme iron, for instance) should have been included in the heme charge distribution calculation.

The next problem was parametrization of the cytochrome bc<sub>1</sub> [Fe<sub>2</sub>S<sub>2</sub>] cluster. The Rieske [Fe<sub>2</sub>S<sub>2</sub>] cluster in the cytochrome bc<sub>1</sub> complex's iron–sulfur protein, ISP, is coordinated by two cysteine residues (cys139 and cys158 in PDB: 1BCC and 3BCC) and two histidine residues (his141 and his161). In the oxidized state, both [Fe<sub>2</sub>S<sub>2</sub>] cluster irons are in the Fe(III) state (overall charge is 0) while in the reduced state, one iron is in the Fe(II) state and one remains in the Fe(III) state (overall charge -1). The reference 22 authors calculated the charge distribution on the [Fe<sub>2</sub>S<sub>2</sub>] cluster and the coordinating histidine and cysteine residues. The atomic coordinates from the 1.5-Å resolution structure of the water soluble portion of the ISP from bovine heart cytochrome bc<sub>1</sub>

complex<sup>26</sup> (PDB: 1RIE) were used for equilibrium bond lengths, angles, and torsional angles. A method similar to that used for calculating the heme charge distribution (Gaussian-94 at the Hartree–Fock level with a 3-21G basis set), with further manipulation to include 4-methyl-imidazole (representing histidine) and a methylthiolate (representing cysteine), resulted in a charge distribution set for the [Fe<sub>2</sub>S<sub>2</sub>] cluster presented in Table 2 of reference 22. Any missing force constants for the [Fe<sub>2</sub>S<sub>2</sub>] cluster and for his and cys groups were derived from existing parameters for similar structures in CHARMM22 and CHARMM19 force fields.

The protein structure was prepared in the following manner. Atomic coordinates from the cytochrome bc<sub>1</sub> complex from chicken heart mitochondria (PDB: 1BCC and 3BCC) were used with hydrogen atom coordinates added using the HBUILD feature of X-PLOR.<sup>27</sup> The bc<sub>1</sub> complex structure with stigmatellin bound at the Q<sub>0</sub> site (see Section 7.6.2 and Figure 7.30, PDB: 3BCC) was used to build a model in which the reduced [Fe<sub>2</sub>S<sub>2</sub>] cluster is in the proximal position (closer to heme b<sub>L</sub>, its redox partner in cytochrome bc<sub>1</sub>). The hemes b<sub>L</sub>, b<sub>H</sub>, and c<sub>1</sub> were modeled in their oxidized forms by assigning the appropriate charge distributions to each. Stigmatellin was then removed from the model. (Figures 1 and 2 of reference 22 illustrate the heme and [Fe<sub>2</sub>S<sub>2</sub>] cluster positions clearly.) The resulting structure, comprised of 32,310 atoms, was refined by 500 steps of energy minimization of hydrogen atoms, followed by 1500 steps of minimization of all atoms.

Next, water molecules were placed in the protein following a modified procedure of the software program DOWSER.<sup>28</sup> After use of the HBUILD routine and an additional 500 steps of energy minimization, 121 water molecules were selected and refined into the structure in 500 more energy minimization steps. This final cytochrome bc<sub>1</sub> complex, including internal water molecules, was used in further modeling. The reference 22 researchers also modeled a lipid bilayer to mimic the *in vivo* position of cytochrome bc<sub>1</sub> in a membrane bilayer. A 3-Å layer of solvent water was energy-minimized into a 120-Å × 155-Å × 35-Å box; also, the protein, in its membrane bilayer, was positioned in the water solvent box so that the lipid bilayer was bounded by a solvent layer on either side. Additional solvent was modeled to interact with the protein surrounding heme c<sub>1</sub>. Figure 6 of reference 22 visualizes the final model.

Now the model can be subjected to the steered molecular dynamics simulation (SMD). The researchers applied external forces generating a torque to induce rotation of the mobile head of the ISP about its rotation axis as described in the initial algorithm. In the SMD simulation, the forces were exerted on the  $\alpha$ -carbon atoms of the ISP residues 73–196. As the forces are applied, the system must overcome potential energy barriers, such as the breaking of a hydrogen bond between the ISP and another subunit. Peaks are generated that correspond to the energy minima along the ISP rotation path. The SMD trajectory was analyzed by calculating the torque applied to each of the restrained  $\alpha$ -carbon atoms relative to the assumed axis of rotation and

the actual angle of rotation of each atom every 2 ps. Lastly, the applied torque and rotation angles were averaged over all 126 residues (ISP residues 73–196), and they were analyzed for individual residues as well.

The reference 22 researchers found that a water channel opened up, involving a number of critical amino acid residues, between the internal water molecules and the bulk, solvent water layer. As the ISP protein rotated during the simulation, many critical hydrogen bonds were broken and reformed between different amino acid residues. The authors suggest that their simulations have relevance to the mechanism by which cytochrome bc<sub>1</sub> complex carries out binding and reduction of substrate ubiquinone at the Q<sub>0</sub> site. Removal of stigmatellin from the model, followed by the SMD simulation, mimics the process by which product would be removed from the active site. Conversely, docking of the ISP in its proximal conformation (close to heme b<sub>L</sub>) in the presence of stigmatellin mimics the formation of the active complex with substrate. The authors also believe that hydrogen bonding within the active site plays a major role in positioning the ISP and substrate. As the reader will note from the preceding discussion, the major portion of a molecular dynamics study must be the laborious preparation of the model, so that possible, or probable, mechanistic details can be discovered for complex protein systems such as cytochrome bc<sub>1</sub>.

## 4.4 QUANTUM MECHANICS-BASED COMPUTATIONAL METHODS

### 4.4.1 Introduction

Many other approaches for finding a correct structural model are possible. A short description of ab initio, density functional and semiempirical methods are included here. This information has been summarized from the paperback *Chemistry with Computation: An Introduction to Spartan*.<sup>29</sup> The Spartan program is described in Section 4.5. Another description of computational chemistry including more mathematical treatments of quantum mechanical, molecular mechanical, and statistical mechanical methods is found in the Oxford Chemistry Primers volume: *Computational Chemistry*.<sup>15</sup>

### 4.4.2 Ab Initio Methods

Ab initio calculations are based on first principles using molecular orbital (MO) calculations based on Gaussian functions. Combinations of Gaussian functions yield Slater-type orbitals (STOs), also called Slater determinants. STOs are mathematical functions closely related to exact solutions for the hydrogen atom. In their ultimate applications, ab initio methods would use Gaussian-type wave functions rather than STOs. The ab initio method assumes that from the point of view of the electrons the nuclei are stationary, whereas

*molecular mechanical* methods assume that the motions of the nuclei of a molecule are independent of the motions of surrounding electrons. In ab initio calculations, the Born–Oppenheimer approximation—separation of the movement of atom nuclei and electrons is possible because electrons move much more rapidly than nuclei—is used to solve the Schrödinger equation ( $H\Psi = E\Psi$ ) with a large but finite basis set of atomic orbitals. In practice, most ab initio methods use the Hartree–Fock approximation, which represents the many-electron wavefunction as a sum of products of one-electron wavefunctions, termed molecular orbitals. Many different basis sets of orbitals have been generated for use with ab initio calculations. A practical minimal basis set (such as the popular STO-3G representation, STO=Slater-type orbitals or Slater-type basis functions) for lithium and beryllium would contain 1s, 2s, and 2p occupied atomic orbitals supplemented by a set of unoccupied (in the atom) but energetically low-lying p-type functions. Minimal basis sets do not adequately describe nonspherical (anisotropic) electron distributions in molecules. One finds this basis set in computer modeling programs, but it should not be expected to yield realistic results for any inorganic complexes. To remedy this, one “splits” the valence description into “inner” and “outer” components. The result is the split-valence basis set, and the valence manifolds of main-group elements for instance are represented by two complete sets of s- and p-type functions. A simple split-valence basis set is called 3-21G. For heavier main-group elements, unoccupied (in the atom) but energetically low-lying d-type functions are added to create the 3-21G(\*) basis set. Polarization basis sets account for the displacement in molecular orbitals resulting from hybridization (an sp hybrid would be an example). Two popular polarization basis sets are the 6-31G\* and 6-31G\*\* representations. Ab initio methods provide excellent accounts of equilibrium and transition state geometries and conformations as well as reaction thermochemistry. They are computationally very intensive and are usually limited to molecules containing 50 atoms or less.

#### 4.4.3 Density Function Theory

Density functional models attempt to describe the total energy of a molecular system also from the standpoint of Hartree–Fock theory—the many-electron wavefunction as a sum of products of one-electron wavefunctions, termed molecular orbitals. These methods accurately describe equilibrium and transition-state geometries as well as thermochemistry. For density functional models, one must calculate the kinetic energy (KE) of the individual electrons (nuclear KE is zero in the Born–Oppenheimer approximation), the attractive potential energy between nuclei and electrons, and the repulsive potential energy between electrons (the Coulomb electron repulsive term in Hartree–Fock theory adjusted by the exchange term which takes into account a repulsive overestimation). The exchange and correlation terms ( $E_{\text{correlation}} = E_{\text{true}} - E_{\text{Hartree–Fock}} - E_{\text{relativistic}}$ ) in so-called “local” density functional models

originate from the exact solution of an idealized many-electron problem, mainly an electron gas of constant total electron density. In practice, one establishes functional relationships between the exchange and correlation energies for the “idealized gas” and the total electron density. However, such relationships are not unique and do not as yet lead to systematic progression to a low-energy structure. Practical density functional calculations make use of explicit atomic basis sets as described for ab initio methods. Additionally, numerical integration steps are necessary. These may lead to loss of precision and slower calculations compared to ab initio methods, especially for small molecules. Costwise density function methods are particularly useful for large systems when compared to other Hartree–Fock methods. In Section 4.5, density functional methods are described for high-end programs, usually run on workstation-level computers.

Application of density function methods to bioinorganic systems was the subject in the American Chemical Society Inorganic Section at the August 2000 national meeting. In introductory remarks, David A. Dixon of the Pacific Northwest National Laboratory stated that density functional theory has been shown to be an effective way—in accuracy and computational cost—to predict structures and vibrational spectra of inorganic compounds in comparison to other methods. He stated that DFT can be used to predict other properties including reaction energies, electron densities, excited-state spectra, and NMR chemical shifts. Subsequent speakers in the section told of applying localized perturbation approaches and density functional theory to systems with hundreds of atoms, allowing accurate calculations including electron correlation to be carried out. One researcher, M. B. Hall, reported density functional calculations on models for [Fe]-hydrogenase. He related these to structures and vibrational frequencies of the observed redox forms of the enzyme, as well as to the reaction mechanism at the enzyme’s diiron active center.<sup>30</sup> R. A. Friesner spoke on large-scale ab initio quantum chemical calculations on biological systems.<sup>31</sup> More recently, Friesner and Guallar reported a hybrid quantum mechanics/molecular mechanics study of the cytochrome P450<sub>CAM</sub> enzymatic catalysis cycle.<sup>32</sup> Cytochrome P450<sub>CAM</sub>, a monooxygenase, is discussed in Section 7.4 (see especially Section 7.4.2 and Figures 7.12 and 7.13), and the catalytic cycle is shown in Figure 7.14. The cytochrome P450<sub>CAM</sub> enzyme, one of the most studied P450s, catalyzes the oxidation of camphor (see Figure 7.12). The reference 32 study’s results showed an active role for cytochrome P450<sub>CAM</sub> in several catalytic steps. The enzyme is involved in controlling the energy gap between high- and low-spin states in the substrate binding process. At the enzyme’s active site, selective interaction of thr252 and the distal dioxygen molecule causes O–O bond cleavage. The protein environment was found to catalyze substrate (camphor) hydrogen atom abstraction with a very low free-energy barrier (~8 kcal/mol). The reference 32 authors believe that their results provide an explanation as to why the proposed catalytically competent intermediate—a ferryl iron–oxo species, Fe<sup>IV</sup>=O, also called “compound I”—cannot be trapped experimentally.

Many researchers have synthesized cytochrome P450 model compounds. Some of these are discussed in Section 7.4.5. One goal of cytochrome P450 model compound research has been to synthesize an oxoiron(IV) species with an iron coordination sphere similar to that of the supposed cytochrome P450 catalytic intermediate—that is, four planar nitrogen ligands and axial oxo and thiolate ligands. This goal has been partially achieved recently by a collaborative effort headed by Jason Halfen, Eckard Münck, and Larry Que.<sup>33</sup> The model compound,  $[\text{Fe}^{\text{IV}}(\text{O})(\text{TMCS})]^+$ , is discussed in Section 7.4.5.3 (see Figure 7.23). In addition to electronic and Mössbauer spectroscopy, along with extended X-ray absorption fine structure (EXAFS) structural data, the researchers optimized the geometry of the model compound using DFT calculations. The supplement to the 2005 Science paper describes the DFT procedures used in great detail. Only the first procedural remarks are given here. The density functional calculations were performed using Becke's three-parameter hybrid functional (B3LYP) and basis set 6-311G provided by the Gaussian 03 (Revision-B.05) software package.<sup>34</sup> The Mössbauer  $^{57}\text{Fe}$  hyperfine parameter,  $\Delta E_Q$ , was calculated using the properties keyword of the Gaussian code. The Mössbauer  $^{57}\text{Fe}$  isomer shifts,  $\delta$ , were evaluated from the DFT charge density at the Fe nucleus using the calibration given by Vrajmasu.<sup>35</sup> The SCF procedures and geometry optimizations were terminated upon reaching the default convergence criteria. Geometry optimizations were performed for the  $S=1$  and  $S=2$  states of  $[\text{FeO}(\text{TMCS})]^+$ . The interested reader should consult reference 33 and its accompanying supplemental material for more information.

#### 4.4.4 Semiempirical Methods

Semiempirical molecular orbital (MO) methods follow from (a) the Hartree–Fock models with inclusion of additional approximations and (b) the introduction of empirical parameters. These methods provide acceptable equilibrium and transition-state geometries but fail to account reliably for known thermochemistry. The single additional approximation used in semiempirical methods, termed NDDO, eliminates overlap of atomic basis functions on different atoms, a severe limitation. However, it leads to a great reduction in computation effort. Most currently used semiempirical models are restricted to a minimal valence basis set of atomic functions. Because of this, calculations involving a heavy main-group element (e.g., gallium) take place in the same time frame as those for a first-row element (e.g., boron). Slater-type basis functions (STOs closely related to exact solutions for the hydrogen atom) are used in place of Gaussian functions (bell-shaped wavefunctions) employed for ab initio calculations. The adjustable parameters introduced to reproduce experimental data in semiempirical calculations can improve the model obtained only if the correct method is chosen for the molecule in question. Three of the commonly used methods are called AM1 (limited data set of parameters mostly applicable to organic molecules and ligands),<sup>36</sup> PM3 (larger and more

diverse “training set” of parameters,<sup>37</sup> and MNDO (the oldest method using fewer parameters and slightly different functional forms).<sup>38</sup> A newer MNDO/d method includes *d*-type functions for second-row and heavier main-group elements (much like the 3-21G(\*) basis set for second-row and heavier main-group elements vs. 3-21G for first-row elements).<sup>39</sup> A newer PM3(tm)<sup>40</sup> describes transition metals in terms of a minimal valence basis set [*nd*, (*n*+1)s, (*n*+1)*p*, where *n* is the principal quantum number]. PM3(tm) is normally used in conjunction with PM3. Semiempirical calculations on molecules containing up to 200 atoms are practical.

One example of the use of semiempirical methodology is provided in an article detailing a molecular-dynamics simulation of the beta domain of metallothionein with a semiempirical treatment of the metal core.<sup>41</sup> The beta domain of rat liver metallothionein-2 contains three-metal centers. In this study, three molecular variants with different metal contents—(1) three cadmium ions, (2) three zinc ions, and (3) one cadmium ion and two zinc ions—were investigated using a conventional molecular dynamics simulation, as well as a simulation with a semiempirical quantum-chemical description (MNDO and MNDO/d) of the metal core embedded in a classical environment. For the purely classical simulations, the standard GROMOS96 force-field parameters were used, and parameters were estimated for cadmium. The results of both kinds of simulations were compared to each other and to the corresponding experimental X-ray crystallographic and NMR solution data. The purely classical simulations were found to produce a too compact metal cluster with partially incorrect geometries, which affected the enfolding protein backbone structure. The inclusion of MNDO/d for the treatment of the metal cluster improved the results to give correct cluster geometries and an overall protein structure in agreement with experimental results. The metal cluster and the cysteine residues bound to it were found to be structurally stable, while the irregular polypeptide backbone loops between the cysteines exhibited a considerable flexibility. MNDO without extension to *d* orbitals failed to maintain the structure of the metal core.

## 4.5 COMPUTER SOFTWARE FOR CHEMISTRY

This section provides a snapshot of chemistry software and websites, arranged alphabetically, available in early 2007. Any software listing will necessarily be incomplete and out-of-date before this text’s publication. The emphasis in the following listing is on software that contains algorithms suitable for use with transition metals and biomolecules. More sophisticated software programs allowing quantum mechanical (ab initio or density functional theory calculations) may require workstation-level computer hardware, although requirements change frequently as desktop and laptop machines become more powerful. At the current time, most software suppliers offer very capable

programs suitable for use with desktop PCs and Macintosh computers. This author suggests that prospective users evaluate as many programs as possible by visiting websites or vendor booths at professional society meetings. Many vendors offer useful software demonstrations at professional meetings and provide online demonstration software or evaluation CDs for further evaluation. Many other programs not discussed here are available for drawing, visualization, computation, conformational analysis, combinatorial chemistry analysis, quantitative structure–activity relationship (QSAR) searching, and other database searching. Not all capabilities of the software programs discussed here are mentioned. The idiosyncratic nature of the information below arises from its accessibility on various software company websites. Some product description pages were well-organized and very accessible, while others were difficult and required unnecessarily long download times for complex visualizations better left to demonstration downloads or CDs. In the listing below, each software supplier or program mentioned will list a website that should be consulted for the most accurate and up-to-date information. Most software operating in Microsoft Windows environment will require, at a minimum, Windows 2000, XP Home or Professional, 1GB of available disk space, 256MB of system memory, a  $1024 \times 768$  resolution-capable display monitor, and a 1-GHz Pentium or AMD equivalent CPU. For modeling and calculations involving proteins, double all of the above. A useful website with a plethora of information about scientific software is found at <http://www.ks.uiuc.edu/Development/biosoftdb/>. The category that gives short description of many of the programs described below, and more, is “Quantum Chemistry Calculation.”

An interesting research article published in 2005 compared a number of different quantum mechanical and molecular mechanics methods—the programs and methods to be described in the following paragraphs.<sup>43</sup> The test molecule, analyzed by the different methods, which were then compared to known experimental results, was (2-amino-5-thiazolyl)-alpha-(methoxyimino)-*N*-methylacetamide, a model of the aminothiazole methoxime (ATMO) side chain of third-generation cephalosporin antibacterial agents. Among the quantum methods examined were the semiempirical MNDO, AM1, and PM3 methods, Hartree–Fock (ab initio) at a range of basis set levels, density functional theory (DFT) at a range of basis sets, and a post-Hartree–Fock method, local Möller–Plesset second-order perturbation theory (LMP2). Among the force fields compared were AMBER, MMFF94, MMFF94s, OPLS/A, OPLS-AA, Sybyl, and Tripos. Programs used were Spartan, MacroModel, SYBYL, and Jaguar. For the computational versus experimental comparisons made, the MMFF94 force field such as implemented in MacroModel was the best overall computational chemistry method at reproducing crystallographic data and conformational properties of the ATMO moiety. This work demonstrates that going to a higher level of quantum theory does not necessarily give better results and that quantum mechanical results are not necessarily better than molecular mechanics results, at least for small organic molecules.

CambridgeSoft at <http://www.cambridgesoft.com/software/> offers a variety of tools including the software suite, ChemOffice™. ChemOffice consists of the following: ChemDraw, Chem3D, ChemFinder, and ChemACX for chemists; BioOffice, BioAssay, BioViz, and BioDraw for biologists; and Inventory, E-Notebook, and The Merck Index for all scientists. ChemOffice and BioOffice are available for Microsoft Windows. Chem3D provides visualization and display of molecular surfaces, orbitals, electrostatic potentials, charge densities, and spin densities. Chem3D utilizes MOPAC, Gaussian, GAMESS, and extended Hückel to compute molecular properties. Every biology or chemistry student should be familiar with the chemical structure drawing program ChemDraw. Many undergraduate organic chemistry texts bundle ChemOffice, as a course resource. The figures in this text have been executed in ChemDraw Ultra 10.0. Some figures with colored visualizations have been rendered from Protein Data Bank (PDB) files in Chem3D Ultra 10.0, then imported into and labeled using ChemDraw.

The Open-Source visualization program, PyMOL™ is available from DeLano Scientific at <http://www.delanoscientific.com/> and <http://pymol.sourceforge.net/>. DeLano Scientific, a successful Open-Source software company, continues to create free and affordable software tools, but also provides commercial software services, added convenience, and other benefits to paying customers who sponsor their efforts through monetary support or value-for-value exchanges. PyMOL is available from Delano Scientific under a traditional Open-Source license without mandatory usage fees. PyMOL will run on Linux (or any Unix) operating system, Mac OS X and Windows. PyMOL™ is a molecular graphics system with an embedded Python interpreter designed for real-time visualization and rapid generation of high-quality molecular graphics images and animations. It imports and edits PDB files. The PyMOL™ user's and reference manuals can be downloaded as .html or .pdf files. Some figures with colored visualizations have been rendered from Protein Data Bank (PDB) files in PyMOL, then imported into and labeled using ChemDraw Ultra, version 10.0.

Fujitsu offers the CAChe™ software systems at [http://www.computers.us.fujitsu.com/www/products\\_bioscience.shtml?products/bioscience/cache](http://www.computers.us.fujitsu.com/www/products_bioscience.shtml?products/bioscience/cache). A variety of programs are available, depending on the user's needs and pocket-book. Model systems can be generated within CAChe™ or imported from the Protein Data Bank (PDB) and other chemical drawing programs. CAChe™ provides a wide range of computational models, from molecular mechanics through ab initio electronic structure techniques. Small organic and inorganic molecules, polymers, materials systems, and whole proteins may be treated. Electronic structure methods include Extended Hückel (for all elements), ZINDO (primarily for UV-visible spectra), MOPAC, and Dgauss (density functional theory). MOPAC2002, a refined MOPAC, covers all main group elements and many metals using many different models including PM5, AM1,<sup>36</sup> PM3,<sup>37</sup> MINDO/3, and MNDO.<sup>38</sup> It includes the COSMO solvent model and *d* orbitals for transition elements. MOPAC2002 includes a linear scaling

method, Mozyme, that can optimize the geometry of large systems—proteins included—up to 20,000 atoms, with or without a solvent field. CAChe™ can run and visualize results for Gaussian™ software (requires separate installation of Gaussian™).

Gaussian, Inc.'s newest software program, Gaussian 03™, is available at <http://www.gaussian.com/>. Starting from the basic laws of quantum mechanics, Gaussian 03™ predicts the energies, molecular structures, and vibrational frequencies of molecular systems, along with numerous molecular properties derived from these basic computation types. The program can be used to study molecules and reactions under a wide range of conditions, including both stable species and compounds that are difficult or impossible to observe experimentally such as short-lived intermediates and transition structures. The Gaussian 03™ ONIOM method can treat proteins and other large biological molecules, traditionally out of the reach of electronic structure methods. ONIOM first appeared in Gaussian 98™, and several significant innovations in Gaussian 03™ make it applicable to much larger molecules. The ONIOM facility in Gaussian 03™ provides substantial performance gains for geometry optimizations via a quadratic coupled algorithm and the use of micro-iterations. In addition, the program's option to include electronic embedding within ONIOM calculations enables both the steric and electrostatic properties of the entire molecule to be taken into account when modeling processes in the high-accuracy layer (e.g., an enzyme's active site). These techniques yield molecular structures and properties results that are found to be in very good agreement with experiment. Gaussian 03™ can predict spin–spin coupling constants in addition to the NMR shielding and chemical shifts available previously. Computing these constants for different conformations and then comparing predicted and observed spectra makes it possible to identify the specific conformations that were observed. In addition, the assignment of observed peaks to specific atoms is greatly facilitated. Gaussian 03™ also can compute a very wide range of spectra and spectroscopic properties. The full Gaussian 03™ program requires a Unix workstation platform. GaussianW™ is available for PC, Windows 98 or above OS, 300 MHz or higher processor, 256 MB memory, and 200 MB disk storage available. GaussianM™ is available for Macintosh machines with G3, G4, or G5 processor, OS X 10.2 or higher operating system, 256 MB memory, and 200 MB disk storage available.

Hypercube, Inc. at <http://www.hyper.com> offers molecular modeling packages under the HyperChem™ name. HyperChem's newest version, HyperChem Release 7.5, is a full 32-bit application, developed for the Windows 95, 98, NT, ME, 2000, and XP operating systems. Density Functional Theory (DFT) has been added as a basic computational engine to complement Molecular Mechanics, Semiempirical Quantum Mechanics and ab initio Quantum Mechanics. The DFT engine includes four combination or hybrid functions, such as the popular B3-LYP or Becke-97 methods. The Bio+ force field in HyperChem represents a version of the Chemistry at HARvard using Molecular Mechanics (CHARMM) force field. Release 7.5 of HyperChem updates

this force field with new functional terms and new parameters to represent the latest science from the CHARMM community. Two HyperChem versions are available for Macintosh. HyperChem for Mac B.1 is supported for the Mac mini, iMac, iBook, PowerBook, and PowerMac, but not the G5 machine with its 64-bit processor. HyperChem Mac B.2 runs on the new Intel Macintosh machines.

MOMEC97, available at <http://www.uni-heidelberg.de/institute/fak12/AC/comba/>, is a program for strain energy minimization, designed to be used in conjunction with HyperChem in a Windows environment. MOMEC97 is a program for strain energy minimization. The program was developed by the Peter Comba and Trevor Hambley for modeling transition metal compounds, especially transition metal coordination compounds, but may be used in many areas of inorganic and organic chemistry. MOMEC97, used in conjunction with HyperChem<sup>TM</sup>, carries out optimization of geometries in MOMEC, then model building and visualization in HyperChem<sup>TM</sup>. The program contains an extensive force field for transition metal and rare earth compounds, along with a module for the refinement of Jahn–Teller distortions in hexacoordinate coordination compounds (see Sections 1.5.2 and 1.6). The Jahn–Teller module's approach is based on a first order harmonic model that minimizes the sum of steric strain and electronic stabilization. The strain energy calculation uses the MOMEC *Energy* module with the metal-donor distances following a Jahn–Teller mode—that is, elongation of the metal–ligand distances on the *z*-axis by  $2d$  and compression of the in-plane metal–ligand bonds by  $d$ . The electronic term depends on the ideal metal-donor distance, the type of ligand ( $\sigma$  or  $\pi$  bonding) and the ligand field strength. Several applications of the MOMEC program by researchers can be found in reference 42.

Schrodinger, Inc. at <http://www.schrodinger.com/> offers a number of modeling and calculation packages. Jaguar<sup>TM</sup>, running on Unix workstations, is a high-performance ab initio quantum mechanical application for gas and solution phase simulations with particular strength in metal-containing systems. The program uses methods including Hartree–Fock theory, generalized valence bond theory (GVB), density functional theory (DFT), and second-order local Möller–Plesset perturbation theory (LMP2). The Jaguar<sup>TM</sup> program can model important solvent effects by applying a self-consistent reaction field (SCRF). Jaguar<sup>TM</sup> computes molecular properties including NMR, IR,  $pK_a$ , partial charges, multipole moments, polarizabilities, molecular orbitals, electron density, and electrostatic potentials. Jaguar accepts a wide variety of input formats. MacroModel<sup>TM</sup>, Schrodinger's molecular modeling program, also requiring workstation architecture, has a large user base in the scientific community. The program features molecular graphics, molecular mechanics, and a large selection of force fields and advanced methods for conformational analysis, molecular dynamics, and free-energy calculations. Use of a wide array of force fields—MM2, MM3, AMBER, AMBER94, MMFF, OPLS, OPLS-AA—is possible. The program handles conformational analyses of entire proteins as well as small molecules. The Maestro<sup>TM</sup> graphical user interface (GUI), also a

Unix workstation platform user, visualizes results that can be displayed across many computer platforms including Macintosh and Windows.

SemiChem products are available at <http://www.semichem.com>. AMPAC<sup>TM</sup>, available as a stand-alone product with Windows-based and workstation level interfaces, is a semiempirical quantum mechanical program featuring SAM1, AM1, MNDO, MNDO/d, PM3, MNDO/C, MINDO/3 semiempirical methods. AMPAC<sup>TM</sup> also includes a graphical user interface (GUI) that builds molecules and offers full visualization of results. The SYBYL/Base<sup>TM</sup> program offered by Tripos, Inc. (products described below) provides an interface with interactive graphing and structural display tools that can be used to access AMPAC's calculation tools.

Serena Software, Inc. at <http://www.serenasoft.com> has released PcModel 9.1. The program supports OpenGL rendering of molecules (high-quality graphics), as well as the Q-Chem ab initio quantum chemistry program. Geometry optimization is available using a wide variety of force fields—MMX, MM3, MMFF94, AMBER, or OPLS-AA—and includes a central force field model for organometallic molecules. One can carry out conformational searching using any of the included force fields, as well as compute normal vibrational modes. PcModel 9.1 interfaces to many quantum chemistry programs including Gaussian, Gamess, PQS, ADF, Turbomole, and Q-Chem and can read from and write to many types of chemical information files including: PDB (Protein Data Bank), SDF, MM3, Macromodel (Schrodinger, Inc.), SYBYL (Tripos, Inc.), Alchemy (Tripos, Inc.), MDL-Mol, Gaussian (Gaussian, Inc.), Gamess, Jaguar, MOPAC, AMPAC, Hondo, ADF, PQS, Turbomole, Chem3D (CambridgeSoft), Tinker, and Smiles. Transition metal complexes can be built with explicit sigma bonding, lone-pair coordination, and pi-system coordination. Parameters are available for all transition metals. The software is supported on Macintosh, Windows, and Linux platforms.

Tripos offers a wide array of computational and modeling software programs available at [http://www.tripos.com/index.php?family=Modules,SimplePage,discovery\\_info](http://www.tripos.com/index.php?family=Modules,SimplePage,discovery_info). Those of most interest to bioinorganic chemists are described briefly here. The SYBYL/Base<sup>TM</sup> program includes a comprehensive set of molecular modeling tools for structure building, optimization, comparison, visualization of structures and associated data, annotation, hard-copy and screen capture capabilities, and a wide range of quantum mechanical methods and force fields. Its use requires a workstation environment. The SYBYL/Base<sup>TM</sup> program includes implementations of the AMBER united-atom and all-atom force fields as well as MMFF94 and AMPAC (SemiChem). In late 2006, Tripos offered SYBYL<sup>TM</sup> 7.3. In its advanced protein modeling section, the software suite provides researchers with protein homolog searching, sequence alignment, and comparative modeling capabilities. A new graphical interface provides a streamlined workflow while allowing expert users full control over the protein modeling and refinement process. Other improvements in the suite will be of interest in protein informatics and modeling.

Wavefunction, Inc. at <http://www.wavefun.com> offers molecular modeling software under the Spartan name. A wide variety of Spartan software versions are available, ranging from a basic program for student use to more complete software suites. Windows and Macintosh operating systems will run Spartan-Model, Spartan Student, Spartan P-Chem, and Spartan 04 Essential versions of the software. Spartan 04 runs on Linux- or Macintosh-based machines, and Spartan 06 runs on Windows. The latest Spartan™ 06 release offers all features and methods included in the Spartan™ Essential Edition and, in collaboration with Q-Chem, provides a full range of post-Hartree–Fock methods including Density Functional, Möller Plesset, thermochemistry recipes, and an assortment of advanced correlated approaches. Q-Chem 3.0 is a comprehensive ab initio quantum chemistry package. It is available fully integrated as the back end of Spartan™ 06 and stand-alone (LINUX, Mac OS and other UNIX platforms). Spartan users may import and export files from the Protein Data Bank (PDB) and other drawing and modeling programs. For molecular mechanics energy minimizations, the programs use the SYBYL force field. Semiempirical methods included are AM1 and AM1-SM2 for solvation, along with PM3 and PM3(tm). Possible Hartree–Fock ab initio calculations use the basis sets 3-21G, 3-21G(\*), and 6-31G\*.

Some of the many other drawing, modeling, and computation programs are listed below. Some of these have been already mentioned or referenced in previous material in this section. AMBER carries out molecular simulations, particularly of large biomolecules.<sup>44</sup> Website: <http://amber.scripps.edu/>. CHARMM also carries out molecular simulations, particularly of large biomolecules.<sup>45</sup> Website: <http://yuri.harvard.edu/>. The GROMOS program carries out molecular simulations.<sup>46</sup> Website: <http://www.igc.ethz.ch/gromos>. MOBY is a molecular modeling package for the PC. For applications in inorganic chemistry see reference 47.

#### 4.5.1 Mathematical Software

Information on the useful graphing program SigmaPlot is found at <http://www.systat.com/products/sigmaplot/> (for Windows-based operating systems) and DeltaGraph (for Macintosh or Windows) at <http://secure.redrocksw.com/store/home.php>. SigmaPlot will display multiple graphs, bar charts, 3D mesh and line plots and ternary plots. Both programs import data entered into Microsoft Office Excel™ files.

Design Science, Inc. provides MathType (<http://www.dessci.com/en/>), the professional version of the Microsoft Equation Editor. It is an interactive tool for Windows and Macintosh that lets users create mathematical notation for word processing and desktop publishing documents, web pages, and presentations. Their product WebEQ builds interactive mathematics on the web.

Wolfram Research at <http://www.wolfram.com/> supplies Mathematica, now in version 5.2. This program does everything mathematical from simple calculator operations to large-scale programming and interactive document

preparation. Mathematica combines interactive calculation (both numeric and symbolic), visualization tools, and a complete programming environment. Mathematica runs on Windows-based PC's, Macintosh with OS X 10.2 and higher, and workstations. A Web-based version is available.

## 4.6 WORLD WIDE WEB ONLINE RESOURCES

Any listing of World Wide Web online resources becomes obsolete and incomplete as soon as it is written down. Sites come and go daily, probably hourly or by the millisecond. Two web-based lists of lists that might be helpful are found at <http://www.chem.ucla.edu/VL/InfoRes.html> (a list of information resources for biologists and chemists) and <http://www.chem.ucla.edu/VL/Commercial.html> (a list of WWW Chemistry Sites at Commercial Organizations).

### 4.6.1 Nomenclature and Visualization Resources

Nomenclature resources help the user give correct names to chemical structures. Visualization resources allow the user to build molecules from scratch or display molecular structures imported from databases or other software programs (like the Protein Data Bank, PDB, for example).

MDL Information Services, Inc. offers free software downloads for chemists at [http://www.mdl.com/downloads/search\\_by\\_si/si\\_chemistry.jsp](http://www.mdl.com/downloads/search_by_si/si_chemistry.jsp). The AutoNom 2000 Add-in is a no-fee download for licensed users of the MDL™ products. AutoNom 200 requires MDL ISIS or Isentris products running on Windows XP. AutoNom generates IUPAC chemical names directly from graphical chemical structures. It provides names for stereoisomers, designating *R,S* configurations for chiral compounds, and *E,Z* descriptors for geometrical isomers. AutoNom is offered under license from MDL Information Systems GmbH. MDL™ ISIS/Draw 2.5 is available as a no-fee download for academic and personal home use only. ISIS/Draw enables you to electronically draw and insert your sketches into documents, Web pages, spreadsheets, and presentations. You can also use ISIS/Draw to create structures to register into 2D and 3D molecule, polymer, and reaction databases, as well as to create queries for searching these databases. This version of MDL ISIS/Draw does not include enhanced stereochemistry features. Enhanced stereochemistry requires the commercial version of ISIS/Draw 2.5 and MDL ISIS/Host 5.0. It operates with the Windows operating system. A Macintosh version is available but not supported by MDL. MDL™ Chime, version 2.6 SP3 is a plug-in that interactively displays 2D and 3D molecules directly in Web pages. You can rotate, reformat, and save the molecules for use in other programs. You will be required to accept a license agreement when you install the program. MDL makes selective grants of licenses to MDL Chime users operating over a network, or for redistribution to other users. Macintosh or Windows operating systems are supported.

ACD/Name by Advanced Chemistry Development working in Windows generates accurate systematic names according to IUPAC (International Union of Pure and Applied Chemistry) and IUMBM recommendations on nomenclature of organic chemistry and selected classes of natural products, biochemical, organometallic, and inorganic compounds. A free download of the Web version is available at <http://www.acdlabs.com/products/java/sda/>. ACD/Structure Drawing Applet (ACD/SDA) is a complete structure drawing, editing, and visualization tool written in pure Java that can be incorporated into HTML documents. The applet can be used for composing substructure queries to databases and visualizing results. Unlike platform-dependent plug-ins or ActiveX controls, the applet runs in the same way on any hardware platform or operating system—the only requirement is a Java-compatible browser such as Netscape 4.0 or later or MS Internet Explorer 4.0 or later.

The Research Collaboratory for Structural Bioinformatics' Protein Data Bank (RCSB-PDB at <http://www.rcsb.org/pdb/home/home.do>) is the online source for X-ray and NMR structural data. Many software programs mentioned in Section 4.5 include the facility to visualize imported data; however, two free software programs operate well in this regard. One is MDL™ Chime described previously in this section. Chime, a chemical structure visualization plug-in for Internet Explorer and Netscape Communicator, supports a wide variety of molecule coordinate formats, including PDB (protein data bank), Molfile (from ISIS/Draw), MOP (MOPAC input files), and GAU (Gaussian™ Input files).

A good website that gives an overview of free programs available for molecular visualization is maintained by Eric Martz at <http://www.umass.edu/molvis/martz/index.htm>. RasMol, originally developed by Roger Sayles, is still available at <http://www.umass.edu/microbio/rasmol/index.html>. It has mostly been supplanted by Protein Explorer, a free Rasmol derivative with extended capabilities, also available at the website addresses given in this paragraph. The program allows the user to visualize macromolecular structure in relation to function; also, it is easier to use than Rasmol, and it is more powerful and faster.

The National Institutes of Health (NIH) Center for Molecular Modeling (CMM), part of the Division of Computation Bioscience (DCB/CIT), displays a molecular modeling page at <http://cmm.info.nih.gov/modeling/>. The CMM software list is available at <http://cmm.info.nih.gov/modeling/software.html>.

One important visualization technique that users should acquire is the ability to transform two-dimensional stereoviews into three dimensions. The stereoviews may be found in visualization software such as MDL Chime and RasMol. They are also found in journal articles in *Science* or *Nature* and others, especially those describing new biomolecule X-ray crystallographic or NMR solution structures. The following hints for learning to transform the stereoviews come from the extremely helpful website <http://www.usm.maine.edu/~rhodes/0Help/StereoView.html>.

Computer and projected stereo images require convergent (cross-eye) viewing—that is, looking at the left-hand image with the right eye and the right-hand image with the left eye. Gaze at a projected stereo pair with your head level. Cross your eyes slightly and slowly so that the two center images (of the four you see) come together. When they fuse, you will see them as a single 3-D image. Ignore the other images at the periphery of your vision. Another approach: With your head level and about 2.5 feet from the computer screen, hold up a finger, with its tip about 6 inches from your face, centered between the stereo pair on the screen. Focus on your finger tip. If you see four images, move your finger slowly back and forth until the middle images converge, then change your focus to the screen. Try removing the finger.

Texts and journal articles require divergent viewing—that is, viewing the left-hand image with the left eye and right-hand image with the right eye. Put your nose on the page between the two views. With both eyes open the two images should be superimposed, but blurred. Slowly move the paper away from your face, trying to keep the images superimposed until you can focus on them. At this point, the middle image should appear three-dimensional and the two peripheral ones should be ignored. Another technique suggests that you tape a divergent stereopair to a mirror, just below eye level. Look at your eyes in the mirror above the image then bend your knees so that your view passes through the stereo pair on the way to looking at your eyes below the image. Rise, then repeat. At some point, the images should fuse. If not, consider the practice a good leg-strengthening exercise. Practice with any or all of the techniques helps.

#### **4.6.2 Online Societies, Online Literature Searching, and Materials and Equipment Websites**

The American Chemical Society maintains the chemistry.org website at <http://acswebcontent.acs.org/home.html>. One can access Chemical Abstracts (CAS) from this site. The site lists meetings, publications, and includes a careers and jobs site and an online store. American Chemical Society (ACS) publications at <http://pubs.acs.org> provides its members information products and services. Currently, 35 peer-reviewed journals are published or co-published by the Publications Division.

The website <http://www.annualreviews.org/> offers online searching capability to the entire Annual Reviews series. Annual Reviews publishes authoritative, analytic reviews in 32 focused disciplines within the Biomedical, Physical, and Social Sciences. The Annual Reviews of Biochemistry and Annual Reviews of Biophysics and Biomolecular Structure are of most pertinence to bioinorganic chemists. This nonprofit, by-subscription, scientific publisher provides free searching of the site and no-cost abstract retrieval.

The Library of Congress site at <http://catalog.loc.gov/> provides free searching access to the Library of Congress' huge collection of books and journals. Some journal articles are available online.

Chemical Abstracts, a subscription service, at <http://www.cas.org/> offers access, by subscription, to CAS, a large and comprehensive database of chemical information.

STN International at <http://www.cas.org/stn.html> is a full-featured online service that offers information for a broad range of scientific fields, including chemistry, engineering, life sciences, pharmaceuticals, biotechnology, regulatory compliance, patents, and business. SciFinder at <http://www.cas.org/SCIFINDER/scicover2.html> is an easy-to-use desktop research tool that allows the user to explore research topics, browse scientific journals, and access information on the most recent scientific developments. SciFinder Scholar is a desktop research tool designed especially for use by students and faculty to easily access the information in the CAS databases. With either tool, one can search Chemical Abstracts and the CA Registry by author name, research topic, substance identifier, chemical structure, or chemical reaction.

Scirus at <http://www.scirus.com/srsapp/> is a comprehensive, commercial search engine designed specifically for finding relevant scientific information everywhere on the World Wide Web. Currently, it indexes more than 300 million science-related Web pages.

The Royal Chemical Society at <http://www.rsc.org/chemsoc/> provides information made available by about 30 national chemistry societies worldwide. It features events, links, educational materials and the popular Visual Elements Periodic Table and Timeline. The site includes a Careers and Job Center, a Chembytes Infozone section providing science, industry and product news, a guide to funding sources, publications updates, and an online magazine. The Conferences and Events section allows the user to search for events, advertise meetings, and view online abstracts from major conferences. The Learning Resources center presents Web tutorials and videotapes for teaching and learning; it also offers tutorial chemistry texts and the Visual Elements periodic table.

The National Cancer Institute (NCI) Research Resources site lists scientific information, tools, reagents, and services for cancer researchers on its website at <http://resresources.nci.nih.gov/>.

Access to Medline and PubMed through the National Library of Medicine at <http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed> provides free searching of the chemical and biological literature. MEDLINE/PubMed contains bibliographic citations and author abstracts covering the fields of medicine, nursing, dentistry, veterinary medicine, the health care system, and the preclinical sciences. MEDLINE contains bibliographic citations and author abstracts from more than 5000 biomedical journals published in the United States and 80 other countries. The database contains over 15 million citations dating back to the mid-1950s. The service provides titles and abstracts for over 130 chemistry journals.

The ISI web of science at <http://www.isinet.com/> is a multidisciplinary database that provides web access to current and retrospective journal literature. Thomson Scientific produces ISI Web of Knowledge<sup>TM</sup>, a tool to access, analyze,

and manage research information. Web of Knowledge<sup>TM</sup> covers journals, books, proceedings, patents, chemical structures, WWW content, grant funding, and preprints. Current Contents Connect<sup>TM</sup> offers electronic full text downloads and the ability to search more than 140,000 full-text world-wide-web documents.

The Scientific World at <http://www.thescientificworld.com/> offers literature searching through SciBase, a collection of databases of scientific, technical and medical research literature. SciBASE content is derived from databases created by the National Library of Medicine (MEDLINE), the British Library, BIOSIS, and PASCAL, as well as CAB ABSTRACTS. Abstracts are sometimes available free, and individual articles are available for purchase through this commercial site. TSW's OpenChoice policy allows access to a wealth of free content on the TSW website, including selected recently published articles through open access publishing and free publishing for authors whose institutions cannot afford publishing fees.

WWW chemicals at <http://www.chem.com> enables chemists to search catalogs and directories of suppliers, distributors, and manufacturers to find chemicals and equipment. Its Structures database offers access to 127,000 3-D structures from the NCI (National Cancer Institute) database.

## 4.7 SUMMARY AND CONCLUSIONS

This introduction to computer hardware and software provides a snapshot of some systems available in late 2006. The brief survey of some online resources should prove useful to students and researchers. Of all the fields one could address in this short text, those of computer hardware, software, and online resources will be the one most quickly outdated. Fortunately, updates are readily available through a search of the World Wide Web.

## REFERENCES

1. Biggs, P. *Computers in Chemistry*, Oxford University Press, New York, 1999.
2. Zomaya, A. Y. *Parallel Computing for Bioinformatics and Computational Biology*, John Wiley & Sons, Hoboken, NJ, 2006.
3. (a) Comba, P.; Hambley, T. W. *Molecular Modeling of Inorganic Compounds*, VCH Publishers, New York, 1995. (b) Comba, P.; Hambley, T. W. *Molecular Modeling of Inorganic Compounds*, John Wiley & Sons, New York, 2001.
4. Comba, P.; Zimmer, M. *J. Chem. Ed.*, 1996, **73**, 108.
5. (a) Hay, B. P. *Coord. Chem. Rev.*, 1993, **126**, 177. (b) Yates, P. C.; Marsden, A. K. *Comput. Chem.*, 1994, **18**, 89.
6. (a) Allinger, N. L.; Yuh, Y. H.; Li, J.-H. *J. Am. Chem. Soc.*, 1989, **111**, 8551–8566. (b) Bowen, J. P.; Allinger, N. L. *Rev. Comput. Chem.*, 1991, **2**, 81.

7. (a) Albinati, A.; Lanza, F.; Berger, H.; Pregosin, P. S.; Rüegger, H.; Kunz, R. W. *Inorg. Chem.*, 1993, **32**, 478. (b) Hay, B. P.; Rustad, J. R. *J. Am. Chem. Soc.*, 1994, **116**, 6313.
8. (a) Halgren, T. A. *J. Am. Chem. Soc.*, 1992, **114**, 7827. (b) Halgren, T. A. *J. Comp. Chem.*, 1996, **17**, 490, 520, 553.
9. Gajewski, J. J.; Gilbert, K. E.; McKelvey, J. In *Advances in Molecular Modeling*, Vol. 2, Liotta, D., ed., JAI Press, Greenwich, CT, 1990, p. 65.
10. Lipkowitz, K. B.; Pearl, G. M.; Robertson, D. H.; Schultz, F. A. *J. Chem. Ed.*, 1996, **73**(2), 105.
11. Hancock, R. D.; Dobson, S. M.; Evers, A.; Wade, P. W.; Ngwenya, M. P.; Boeyens, J. C. A.; Wainwright, K. P. *J. Am. Chem. Soc.*, 1988, **110**, 2794.
12. Adam, K. R.; Antolovich, M.; Brigden, L. G.; Lindoy, L. F. *J. Am. Chem. Soc.*, 1991, **113**, 3346.
13. Allured, V. S.; Kelly, C. M.; Landis, C. R. *J. Am. Chem. Soc.*, 1991, **113**, 1.
14. (a) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skiff, W. M. *J. Am. Chem. Soc.*, 1992, **114**, 10024. (b) Casewit, C. J.; Colwell, K. S.; Rappé, A. K. *J. Am. Chem. Soc.*, 1992, **114**, 10046. (c) Rappé, A. K.; Colwell, K. S.; Casewit, C. J. *Inorg. Chem.*, 1993, **32**, 3438.
15. Grant, G. H.; Richards, W. G. *Computational Chemistry*, Oxford University Press, New York, 1996.
16. Karplus, M.; Petsko, G. A. *Nature*, 1990, **347**, 631–639.
17. Lopez, M. A.; Kollman, P. A. *J. Am. Chem. Soc.*, 1989, **111**, 6212.
18. Charles, R.; Ganly-Cunningham, M.; Warren, R.; Zimmer, M. *J. Mol. Struct.*, 1992, **265**, 385.
19. Collins, J. R.; Du, P.; Loew, G. H. *Biochemistry*, 1992, **31**, 11166.
20. Kuczera, K.; Kuriyan, J.; Karplus, M. *J. Mol. Biol.*, 1990, **213**, 351.
21. Zhang, Z.; Huang, L.; Shulmeister, V. M.; Chi, Y.-I.; Kim, K. K.; Hung, L.-W.; Crofts, A. R.; Berry, E. A.; Kim S.-H. *Nature*, 1998, **392**, 677–684. (PDB: 1BCC, 3BCC)
22. Izrailev, S.; Crofts, A. R.; Berry, E. A.; Schulten, K. S. *Biophys. J.*, 1999, **77**, 1753–1768.
23. Nelson, M.; Humphrey, W.; Gursoy, A.; Dalke, A.; Kalé, L.; Skeel, R. D.; Schulten, K. *Int. J. Supercomput. Appl. High Perform. Comput.*, 1996, **10**, 251–268.
24. (a) MacKerell, A. D., Jr.; Wiorkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.*, 1995, **117**, 11946–11975. (b) MacKerell, A. D., Jr.; Bashford, D.; Bellott, D.; Dunbrack, R. L., Jr.; Evanseck, J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, I. W. E.; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B*, 1998, **102**, 3586–3616. (c) Schlenkrich, M.; Brickmann, J.; MacKerell, A. D., Jr.; Karplus, M. In Merz, K. M.; Roux, B., eds., *Biological Membranes: A Molecular Perspective from Computation and Experiment*, Birkhauser, Boston, 1996, pp. 31–81.
25. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Peng, C. Y.;

- Ayala, P. Y.; Chen, W.; Wong, M.W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. Gaussian 94, Revision B.3, 1995, Gaussian Inc., Pittsburgh, PA.
26. Iwata, S.; Saynovits, M.; Link, T. A.; Michel, H. *Structure*, 1996, **4**, 567–579. (PDB: 1RIE)
27. Brünger, A. T. X-PLOR, Version 3.1: A System for X-Ray Crystallography and NMR. The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, 1992.
28. Zhang, L.; Hermans, J. *Proteins: Struct., Funct., Genet.*, 1996, **24**, 433–438.
29. Hehre, W. J.; Huang, W. W. *Chemistry with Computation: An Introduction to Spartan*, Wavefunction, Inc., Irvine, CA, 1995.
30. (a) Fan, H. J.; Hall, M. B. *J. Am. Chem. Soc.*, 2001, **123**, 3828–3829. (b) Cao, Z.; Hall, M. B. *J. Am. Chem. Soc.*, 2001, **123**, 3734–3742.
31. Friesner, R. A.; Dunietz, B. D. *Acc. Chem. Res.*, 2001, **34**, 351–358.
32. Guallar, V., Friesner, R. A. *J. Am. Chem. Soc.*, 2004, **126**, 8501–8508.
33. Bukowski, M. R.; Koehntop, K. D.; Stubna, A.; Bominaar, E. L.; Halfen, J. A.; Münck, E.; Nam, W.; Que, L., Jr. *Science*, 2005, **310**, 1000–1002.
34. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford CT, 2004.
35. Vrajmasu, V.; Münck, E.; Bominaar, E. L. *Inorg. Chem.*, 2003, **42**, 5974–5988.
36. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.*, 1985, **107**, 3902.
37. Stewart, J. J. P. *J. Comput. Chem.*, 1989, **10**, 209.
38. Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.*, 1977, **99**, 4899.
39. (a) Thiel, W. *Adv. Chem. Phys.*, 1996, **93**, 703. (b) Thiel, W.; Voityuk, A. A. *J. Phys. Chem.*, 1996, **100**, 616. (c) Thiel, W.; Voityuk, A. *Theor. Chem. Acta*, 1992, **81**, 391. (d) Thiel, W.; Voityuk, A. *Int. J. Quantum Chem.*, 1985, **44**, 807.
40. Yu, J.; Hehre, W. J. *Polym. Mater. Sci. Eng. (Washington)*, 1996, **74**, 439.
41. Berweger, C. D.; Thiel, W.; van Gensteren, W. F. *Proteins*, 2000, **41**(3), 299–315.
42. (a) Bernhardt, P. V.; Comba, P. *Inorg. Chem.*, 1992, **31**, 2638. (b) Comba, P.; Hambley, T. W.; Ströhle, M. *Helv. Chim. Acta*, 1995, **78**, 2042. (c) Bol, J. E.; Buning, C.; Comba, P.; Reedijk, J.; Ströhle, M. *J. Comput. Chem.*, 1998, **19**, 512. (d) Comba, P.; Gloe, K.;

- Inoue, K.; Krueger, T.; Stephan, H.; Yoshizuka, K. *Inorg. Chem.*, 1998, **37**, 3310. (e)  
Comba, P.; Gyr, T. *Eur. J. Inorg. Chem.*, 1999, 1787–1792.
43. Ercanli, T.; Boyd, D. B. *J. Chem. Inf. Model.*, 2005, **45**(3), 591–601.
44. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M. Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.*, 1995, **117**, 5179.
45. (a) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.*, 1983, **4**, 187. (b) Momany, F. A.; Rone, R. *J. Comp. Chem.*, 1992, **13**, 888.
46. van Gunsteren, W. F.; Daura, X.; Mark, A. E. “The GROMOS force field,” in Schleyer, P. v. R.; Allinger, N. L.; Clark, T.; Gasteiger, J.; Kollman, P. A.; Schaefer, H. F., III; Schreiner, P. R., eds., *The Encyclopedia of Computational Chemistry*, John Wiley & Sons, Chichester, 1998.
47. Wiesemann, F.; Teipel, S.; Krebs, B.; Höweler, U. *Inorg. Chem.*, 1994, **33**, 1891.

---

# 5

---

## GROUP I AND II METALS IN BIOLOGICAL SYSTEMS: HOMEOSTASIS AND GROUP I BIOMOLECULES

### 5.1 INTRODUCTION

Several group I and II metal ions belong to the essential chemical elements “macrominerals and ions category” ( $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Mg}^{2+}$ ,  $\text{Ca}^{2+}$ ,  $\text{Cl}^-$ ,  $\text{PO}_4^{3-}$ ,  $\text{SO}_4^{2-}$ ) defined in Section 1.2. Group I and II metal ions have roles in maintaining osmotic and electrolytic control in biological species. They are involved in physiological signaling and triggering as well as photosynthetic events. As with all chemical elements and ions, they must be made available when and where they are needed—homeostatic control. The cations passively diffuse and are actively pumped across cell membranes, many times against concentration gradients. Table 5.1 summarizes the characteristics of Group I and II ions, lists some biological molecules that feature them, and indicates the toxicities that arise in nonbiological members of the groups. Reference 1 contains more information on the transport and storage of metal ions in its Chapter 3; also see Chapters 6 and 7 of reference 1 for more information on Groups I and II metals in biological systems. Reference 2 comprehensively reviews the biological chemistry of Groups I and II metals in its Chapters 8, 9, and 10.

In this chapter, homeostasis of Groups I and II metals will be described, and several systems involving the Group I metal ions sodium and potassium will be discussed in detail. In Chapter 6, several topics involving Group II metal ions will be presented.

**TABLE 5.1 Characteristics of Group I and II Ions in Biological Species**

Ion	Concentration (mM)			Function/Toxicity
	Radius (Å)/ Coordination No.	Extracellular (Blood Plasma)	Intracellular (Red Blood Cells)	
H <sup>+</sup>				Maintain pH
Li <sup>+</sup>	0.60			Medicinal usage against hyperactivity
Na <sup>+</sup>	0.95/6(+)	160	11	Electroneutrality, nerve action activation, osmotic pressure control, maintain ionic strength outside cell (as NaCl), maintain cell potential
K <sup>+</sup>	1.33/8(±)	10	92	Electroneutrality, nerve action activation, osmotic pressure control, maintain ionic strength inside cell (as KCl), maintain cell potential
Rb <sup>+</sup>	1.48			Toxicity—may compete with K <sup>+</sup> in potassium channels
Cs <sup>+</sup>	1.67			Cytotoxic, chromosomal aberrations in mice <sup>3</sup>
Be <sup>2+</sup>	0.31			Pulmonary tumors, <sup>4</sup> may impair some vascular endothelial functions and alter the interaction between platelet and endothelial mediators <sup>5</sup>

$Mg^{2+}$	0.65/6	2	Ionophores, ATPases, kinases, enolases, glutamine synthetase, methyl aspartase, ribulose bisphosphate carboxylase, alkaline phosphatase, xylose isomerase, chlorophyll	Nucleic acid stabilization, electrostatics, structure and potential stabilization or catalysis in enzymes
$Ca^{2+}$	0.99/7( $\pm$ )	$10^{-4}$	Ionophores, Ca-ATPases, EF-hand proteins, extracellular digestive enzymes, structural/storage proteins	Stabilize skeletal matter, muscle and nerve triggering, activation of kinases, catalysis
$Sr^{2+}$		2		Toxicity—replaces calcium in bone, <sup>6</sup> treatment of androgen-independent prostate cancer <sup>7</sup>
$Ba^{2+}$		1.13		$BaSO_4$ used as contrast agent for radiography and MRI and in bone cements used for the fixation of joint prostheses, $Ba^{2+}$ inhibits K <sup>+</sup> flux through potassium ion channels
$Ra^{2+}$		1.43		Radiation therapy for carcinomas

Sources: References 1 and 2.

## 5.2 HOMEOSTASIS OF METALS (AND SOME NONMETALS)

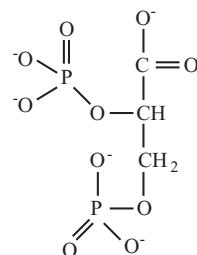
In a living system, homeostasis refers to a set of related and controlled steady states in processes involving continuous changes in material, charge, and energy flows into, out of, and within cells. Reagents must enter cells and products must leave. Inside cells a vast number of chemical reactions and transformations take place. The cell must maintain its structural organization, its osmotic balance, and its material and electrical potential gradients. All biochemically active inorganic species are involved in determining homeostasis; in this chapter we will discuss the involvement of  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ , and  $\text{Zn}^{2+}$  with shorter discussions of P as  $\text{PO}_4^{3-}$ ,  $\text{Cl}^-$ , and  $\text{H}^+$ . Because all species noted here are ions and interdependent upon one another for electroneutrality, osmotic balance, and maintenance of concentration gradients, it will be difficult to separate discussions of each species.

### 5.2.1 Phosphorus as Phosphate

In biological systems, phosphorus does not change oxidation state and is normally found as the phosphate ion,  $\text{PO}_4^{3-}$ , often indicated, mostly by nonchemists, as  $\text{P}_i$ . In the human body, 80–85% of phosphorus is found in the skeleton and 15–20% in organic compounds. The kidney plays a major role in phosphorus homeostasis. Interested readers will find much more information at the website [http://www.wramc.amedd.army.mil/departments/medicine/Nephrology/education/Lectures/CalciumAndPhosphorusHomeostasis\\_files/frame.htm](http://www.wramc.amedd.army.mil/departments/medicine/Nephrology/education/Lectures/CalciumAndPhosphorusHomeostasis_files/frame.htm).) Phosphate ion concentration in blood serum is maintained within a narrow range through a complex interplay between intestinal absorption, exchange with intracellular and bone storage pools, and renal tubular reabsorption. The kidney is a major regulator of  $\text{P}_i$  homeostasis and can increase or decrease its  $\text{P}_i$  reabsorptive capacity to accommodate  $\text{P}_i$  need. The crucial regulated step in  $\text{P}_i$  homeostasis is the transport of  $\text{P}_i$  across the renal (kidney) proximal tubule. The type IIa sodium-dependent  $\text{Na}^+/\text{phosphate}$  ( $\text{P}_i$ ) cotransporter (Npt2a), as regulated by hormonal [parathyroid hormone (PTH), for example] and nonhormonal factors, plays a major role in  $\text{P}_i$  homeostasis within renal (kidney) tubules.<sup>8</sup>

Phosphate ions are constituent parts of two universally found biopolymers, DNA and RNA. Phosphate ion is found in membrane lipids (phospholipids) and associated with the metabolism of many small molecules. The binding of dioxygen by hemoglobin is regulated by local concentrations of  $\text{H}^+$  (known as the Bohr effect),  $\text{CO}_2$  concentration, and organic phosphates such as diphosphoglycerate (DPG), whose structure is shown in Figure 5.1.<sup>9</sup>

Phosphate ion is a major participant in the biological energy cycle through the reactions of mono-, di-, and triphosphates, including one of the most important of these reactions, producing adenosine diphosphate from adenosine triphosphate (see structures in Section 2.3.1, Table 2.2):



**Figure 5.1** Diphosphoglycerate (DPG), regulator of heme dioxygen ( $O_2$ ) binding.



Other important biological manifestations of phosphorus include the pyrophosphate anion,  $P_2O_7^{4-}$  or  $PP_i$ , and inositol 1,4,5-triphosphate,  $IP_3$ . Pyrophosphate anions can arise through the following reaction:



Note that equation 5.2 is irreversible and the product AMP will require two phosphorylation steps to reconstitute the high-energy adenosine triphosphate, ATP. Inositol 1,4,5-triphosphate is an important molecule in the cytosol, where it releases calcium ions from storage. It forms part of a series of inositol-phosphate species that mediate calcium ion concentrations inside and outside the cell.

### 5.2.2 Potassium, Sodium, and Chloride Ions

It is difficult to separate the discussion of potassium, sodium, and chloride ions in a physiological context, because the activity of these ions is intimately interconnected. These ions maintain osmotic pressure between the inside and outside of cells, and their passage across cell membranes (as well as that of the proton) comprises a huge area for discussion. First it is important to know something about passive diffusion and active transport across cell membranes and to understand something of the complex chemistry that takes place within the cell in various vesicles. In general, physiological ionic strength is dependent on sodium chloride (NaCl) concentration outside the cell and on potassium chloride (KCl) concentration inside the cell. A deviation of less than 1% in the concentration of these ions in their proper locations leads to weakness within any organism. Reference 2 contains much more detailed discussions on the following topics.

Cell membranes contain selective ion channels that are highly discriminatory for potassium ions, sodium ions, calcium ions, and the proton. For instance, the highly selective potassium channels of nerves show selectivity for ions as  $Li < Na \ll K > Rb > Cs$ , and calcium channels show selectivity as  $Mg \ll Ca > Sr > Ba$ . Ion selection operates on the basis of size and repulsion, not

attraction, because ions must pass through the channel and not be bound to it. Ions may not necessarily be too small to enter the channel but rather may carry a large shell of hydration that is energetically disadvantageous to remove. Larger ions may not enter a given channel based directly on their too-big-to-fit size. If neither the hydration shell nor size considerations prevent ions from entering a channel, then other factors such as gates come into play. Gates, also known as selectivity filters, are kinetic devices allowing or preventing the free flow of ions. Gates may also be charge movement facilitators based on electrostatic cell potentials. A theme common to the hundreds of known ion channels—gates, selectivity filters—is a central cavity or pore(s) through which, during gating, ions flow depending on their charge, size, and concentration. Protons have their own selective channels that maintain an optimum pH for various cellular systems. Proton channels will not be discussed further here. Interested readers should read recent publications in *Science* magazine for further details.<sup>10</sup>

### 5.2.3 Calcium Homeostasis

Phosphorus homeostasis (see Section 5.2.1) is intimately involved with that of calcium. The most important reservoir of calcium and phosphorus within the mammalian body is in bone—85% of the body's calcium and 85–90% of phosphorus is found there. Ninety-nine percent of bone calcium remains in the mineral phase as  $\text{Ca}_3(\text{PO}_4)_2$ , and so on, but the other 1% can rapidly exchange with extracellular calcium.

Free calcium ions act as secondary messengers and function within the cell in response to an extracellular agent. Other secondary messengers include cyclic nucleotides such as cAMP (cyclic adenosine monophosphate) and cGMP (cyclic guanosine monophosphate). Primary messengers include hormones—insulin, for example—and metabolites such as glucose. Primary messengers operate in the extracellular environment and effect communication between cells or respond to changes in the extracellular environment. Chapter 7 of reference 1 contains a more complete discussion.

In extracellular fluids, about half of  $\text{Ca}^{2+}$  ions are bound to proteins. Free calcium ions have a concentration of approximately 2 mM in blood as seen in Tables 5.1 and 5.2, with the concentration of phosphate ions being similar. Intracellularly,  $\text{Ca}^{2+}$  is held within mitochondria and in the endoplasmic reticulum. The intracellular concentration of calcium ions in blood is about  $10^{-4}$  mM ( $10^2$  nM) as seen in Tables 5.1 and 5.2. Intracellular calcium ion concentration may fluctuate between  $10^2$  and  $10^3$  nM as  $\text{Ca}^{2+}$  is released from cellular stores or influxes from the extracellular fluid. These fluctuations are integral to calcium's role in intracellular signaling, enzyme activation, and muscle contraction. Some proteins and enzymes involved in binding and releasing calcium ions are, for example: (1) Troponin C (binding of  $\text{Ca}^{2+}$  results in muscle contraction), (2) calmodulin (binding of  $\text{Ca}^{2+}$  activates enzymes such as protein kinases), and (3)  $\text{Ca}^{2+}$ -ATPases (cross-membrane calcium pumps binding and

releasing calcium ions to foster muscle relaxation). Calmodulin and calcium ATPases will be discussed in more detail in Sections 6.3.2 and 6.4.2. Many other proteins and enzymes require calcium, but these will not be discussed in any detail here. Extracellular digestive enzymes requiring calcium ions include staphylococcal nuclease (one calcium site), phospholipase A<sub>2</sub> (two calcium sites), and trypsin (one calcium site). Structural and storage enzymes that sequester many calcium ions per protein unit include thrombin, phosphodentine (material in teeth), and calsequestrin (calcium storage in the sarcoplasmic reticulum). Table 6.6 of reference 1 collects some information on calcium-binding proteins.

Calcium ion movements are sensitive to the concentration of other cations. In one study, it was found that Ca<sup>2+</sup> influx into cells might occur through K<sup>+</sup> inward rectifier channels when extracellular K<sup>+</sup> ion concentration fell below 1 mM. These channels became Ca<sup>2+</sup>-permeable only when the extracellular K<sup>+</sup> concentration decreased to 1 mM or below. The same study found that the addition of different divalent cations revealed that Ba<sup>2+</sup>, but not Ni<sup>2+</sup>, Cd<sup>2+</sup>, Sr<sup>2+</sup>, or Mg<sup>2+</sup>, reversibly blocked the Ca<sup>2+</sup> influx into cells during low external K<sup>+.11</sup> Intracellular proteins requiring calcium ion must be very selective for Ca<sup>2+</sup> because the concentration of free Mg<sup>2+</sup> (2.5 mM) and K<sup>+</sup> (~100 mM) are much higher. Intracellular Ca<sup>2+</sup> concentrations must be kept at very low levels because calcium ions inhibit the activity of Mg<sup>2+</sup>-dependent enzymes. In addition, the precipitation of sparingly soluble calcium salts may cause serious problems within a cell.

## 5.3 MOVEMENT OF MOLECULES AND IONS ACROSS MEMBRANES

### 5.3.1 Passive Diffusion

At equilibrium there is a zero free-energy change,  $\Delta G = 0$ , that takes place between compartments separated by a membrane, with the free-energy change being dependent on the difference in concentration of various ions and the electrical potential difference that exists across the membrane. The relationships among sodium, potassium, and chloride ions, pH, and electrolytic potential have become known as Donnan equilibria. The concentrations and electrolytic potentials are related by the following equation:

$$\Delta G = RT \ln \frac{C_2}{C_1} + zF\psi = 0 \quad (5.3)$$

where  $z$  is the charge on the ion,  $F$  is the the Faraday constant,  $\psi$  is the electrolytic potential.

If going down the concentration gradient— $C_2$  is less than  $C_1$ —the first term in equation 5.3 will be negative (remember  $\Delta G = -RT \ln K_{eq}$ ), and if going up the concentration gradient— $C_1$  is less than  $C_2$ —the first term will be positive.

The second term will be positive for a cation ( $K^+$ , for instance) and negative for an anion ( $Cl^-$ ). To understand how this might be applied to any cell, consider a simple “cell” that contains potassium ions, chloride ions, and protein inside the cell with concentrations described in Figure 5.2: 50 mM  $K^+$ , 25 mM  $Cl^-$ , and 25 mM protein (that cannot diffuse through the membrane). For the purposes of this example, we will consider that the cell is bathed in a 50 mM KCl isotonic solution—that is, the concentration of  $K^+$  and  $Cl^-$  outside the cell is 50 mM. For this cell the following equations are established:

$$\Delta G = \frac{RT}{F} \ln \frac{[K^+]_{in}}{[K^+]_{out}} + z\psi = 0 = \frac{RT}{F} \ln \frac{[Cl^-]_{in}}{[Cl^-]_{out}} - z\psi \quad (5.4)$$

$$-\left[ \frac{RT}{F} \ln \frac{[K^+]_{in}}{[K^+]_{out}} \right] - z\psi = \frac{RT}{F} \ln \frac{[Cl^-]_{in}}{[Cl^-]_{out}} - z\psi \quad (5.5)$$

$$\ln \frac{[K^+]_{out}}{[K^+]_{in}} = \ln \frac{[Cl^-]_{in}}{[Cl^-]_{out}} \quad (5.6)$$

and therefore

$$[K^+]_{in} [Cl^-]_{in} = [K^+]_{out} [Cl^-]_{out} \quad (5.7)$$

It is also true that:

$$[K^+]_{in} + [K^+]_{out} = 100 \quad (5.8)$$

$$[Cl^-]_{in} + [Cl^-]_{out} = 75 \quad (5.9)$$

$$[K^+]_{in} = [Cl^-]_{in} + [Prot^-]. \quad (5.10)$$

Using these equations, one finds that the equilibrium concentrations of  $K^+$  and  $Cl^-$  and protein inside and outside the cell are as shown in Figure 5.2. If one then uses the equilibrium concentrations of  $[K^+]_{in}$  and  $[K^+]_{out}$  to calculate the potential generated across the membrane, one finds

$$\psi = \frac{RT}{F} \ln \frac{[K^+]_{out}}{[K^+]_{in}} = 0.059 \log \frac{43}{57} \approx -7 \text{ mV} \quad (5.11)$$

Before Donnan Equilibrium			After Donnan Equilibrium
$[K^+] = 50 \text{ mM}$	$[K^+] = 50 \text{ mM}$		$[K^+] = 57 \text{ mM}$
$[Cl^-] = 25 \text{ mM}$	$[Cl^-] = 50 \text{ mM}$		$[Cl^-] = 32 \text{ mM}$
$[Prot^-] = 25 \text{ mM}$			$[Prot^-] = 25 \text{ mM}$
inside cell	outside cell		inside cell
			outside cell

**Figure 5.2** Ion concentrations inside and outside a hypothetical cell.

**TABLE 5.2 Concentration of Some Ions in Living Cells (mM)**

System	Na <sup>+</sup>	K <sup>+</sup>	Ca <sup>2+</sup>	Mg <sup>2+</sup>	Cl <sup>-</sup>	HPO <sub>4</sub> <sup>2-</sup>
Red blood cells	11	92	10 <sup>-4</sup>	2.5	50	3
Blood plasma	160	10	2	2	100	~3

Source: Reference 2.

This calculation ignores the effect of the proton (pH) and other ions that may be present—mainly sodium ion since the concentrations of calcium and magnesium ions usually are much lower (see Table 5.2).

### 5.3.2 Facilitated Diffusion

Facilitated diffusion within organisms takes place when carriers or proteins residing within membranes—ion channels, for instance—organize the movement of ions from one location to another. This diffusion type is a kinetic, not thermodynamic, effect in which  $E_A$  for the transfer is lowered and the rate of diffusion is accelerated. Facilitated diffusion channels organize ion movements in both directions, and the process can be inhibited both competitively and noncompetitively. It is known that most cells maintain open channels for K<sup>+</sup> most of the time and closed channels for other ions. Potassium-ion-dependent enzymes include Na<sup>+</sup>/K<sup>+</sup> ATPases (to be discussed in Section 5.4.1), pyruvate kinases, and dioldehydratases (not to be discussed further).

**5.3.2.1 Gated Channels.** The usual situation for gated channels is that the gate has a “receptor” species for an “effector” species. The receptor binds in the channel, generating a binding energy that causes a conformation change in the effector to open the gate. The gate opens, ions enter, and a change in concentration ( $C$ ) or electrolytic potential ( $\psi$ ) sends the desired message. Voltage-gated channels that open or close by voltage changes, as well as chemically gated channels that open or close by concentration changes, exist. The proteins forming gates are usually helical bundles and therefore have primary sequences that favor  $\alpha$ -helix formation. Membranes containing the channels have shapes controlled by filamentous networks or structures within the cell. Two important types of membranes for our discussions are (1) cytoplasmic membranes and (2) vesicle or organelle membranes for entities within the cell.

### 5.3.3 Active Transport—Ion Pumps

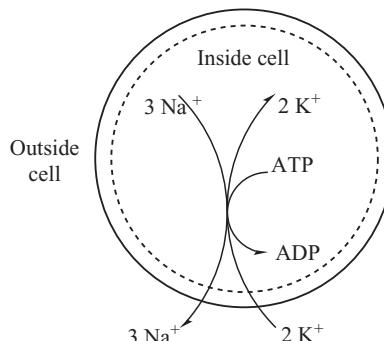
The action of so-called active transport, also known as ion pumps, facilitates larger Na<sup>+</sup>/K<sup>+</sup> gradients than those possible considering calculations of Donnan equilibria. For instance, the concentration of K<sup>+</sup> in red blood cells equals 92 mM versus 10 mM in blood plasma. Calculation of the membrane potential using equation 5.11 would lead to a large negative potential:

$$\psi = \frac{RT}{F} \ln \frac{[K^+]_{out}}{[K^+]_{in}} = 0.059 \log \frac{10}{92} \approx -57 \text{ mV} \quad (5.12)$$

What we have calculated here is the equilibrium potential—the voltage at which the inward and outward flows of the ion are balanced. This calculation has not considered other ions that can ( $\text{Na}^+$ ) or cannot (organic phosphates or ionized proteins) diffuse. In general, one expects that potassium ions can enter cells due to the negative potentials as calculated above. However, the same sort of calculation for sodium ions with concentrations of 11 mM inside red blood cells and 160 mM outside would lead to a large positive potential of about +70 mV. The concentration gradient for potassium is directed out of the cell, while the concentration gradient for sodium is directed into the cell. To pump  $\text{Na}^+$  against a concentration gradient, organisms must utilize active transport (ion pumps) coupled to energy-yielding reactions such as



The cells we will discuss are those having a known potassium ion resting potential, although this is only one type of many. Suppose energy input stops. Sodium ions may still enter cells slowly through diffusion, leaks, and membrane channels. When this happens,  $\text{Na}^+$  must be pumped out of the cell by active transport. In addition, potassium ions may move out of the cell by diffusion and need to be pumped back in by active transport. In the resting state for these cells, the  $\text{Na}^+$  channel is closed, the  $\text{K}^+$  channel is open, and the pump is working. (See Figure 5.3 as adapted from Figure 8.5 of daSilva.) Essentially the resting cells are impermeable to  $\text{Na}^+$ ,  $\text{Ca}^{2+}$ ,  $\text{Cl}^-$ , and  $\text{HPO}_4^{2-}$  but permeable to  $\text{K}^+$  because the potassium ion channels are open. Specifically, for the  $\text{Na}^+/\text{K}^+$  ATPase pump, two  $\text{K}^+$  move in for every three  $\text{Na}^+$  that move out, creating an



**Figure 5.3** Pumped movement of sodium ions out of and potassium ions into a cell assisted by ATP hydrolysis. (Adapted by permission of Oxford University Press from Figure 8.5 of reference 2. Clarendon Press, Oxford and Oxford University Press Inc., New York, 1991.)

electrogenic system. The large membrane potential thus generated yields hydrolysis energy of ATP—the free energy stored in the living cell. This energy will be used for impulses (messages) sent by nerve cells, for exchange reactions to accumulate or dissipate other chemicals (amino acids perhaps), for coupling to other reactions such as oxidations or phosphorylations, and for the synthesis of new raw materials needed by the cell.

In summary, homeostasis for potassium, sodium, calcium, and chloride ions must be controlled carefully so that current (life) is always flowing. Critical balances among these ions, as well as pH control, maintains the ionic characteristics of physiological systems in (1) osmotic pressure balance, (2) electrical potentials involving membranes, (3) interaction with polyelectrolytes (proteins and DNA), and (4) connections to the uptake of organic metabolites. All systems are linked through their energy requirements, normally stored in the pyrophosphate of ATP. Figure 21.20 of reference 2 shows a schematic diagram connecting extracellular levels of  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Ca}^{2+}$ , and  $\text{Cl}^-$ , ATP energy used by ion pumps,  $\text{Zn}^{2+}$  levels inside cells,  $\text{Fe}^{2+}/\text{Fe}^{3+}$  levels inside mitochondria, metabolic pools, and hormones, and as the authors state: “Protons are involved everywhere.”

## 5.4 POTASSIUM-DEPENDENT MOLECULES

This section will concentrate on requirements for potassium ions as cofactors in biological molecules. In doing so, one involves another Group I metal ion—sodium—in discussing the sodium pump,  $\text{Na}^+/\text{K}^+$  ATPase (Section 5.4.1). Sodium ions are intimately involved in the activity of other biomolecules—the voltage-gated sodium ion channels, for instance. We will not discuss the sodium ion channels here, but will concentrate on potassium ion channels in Section 5.4.2. Voltage-gated potassium ion channels are implicated in genetic and autoimmune diseases, making potassium ion channels therapeutic targets in drug design research. The biological activity of the lithium ion rests mostly in (a) its inhibitory effect on many enzymes and (b) its multiple effects on some physiological processes. Lithium is highly effective in the treatment of bipolar disorder, however, the mechanism of lithium action in the treatment of this psychiatric disorder is still unknown. The Group I ion’s biological activity will not be discussed further here.

### 5.4.1 $\text{Na}^+/\text{K}^+$ ATPase: The Sodium Pump

The enzyme  $\text{Na}^+/\text{K}^+$ -ATPase is a highly conserved integral membrane protein that is expressed in virtually all cells of higher organisms.  $\text{Na}^+/\text{K}^+$ -ATPase has been estimated to hydrolyze nearly 25% of all cytoplasmic ATP in resting humans. In nerve cells, approximately 70% of the ATP is consumed to fuel this enzyme. More basic information about  $\text{Na}^+/\text{K}^+$ -ATPase can be found at the website [http://arbl.cvmbs.colostate.edu/hbooks/molecules/sodium\\_pump](http://arbl.cvmbs.colostate.edu/hbooks/molecules/sodium_pump).

html.  $\text{Na}^+/\text{K}^+$ -ATPase belongs to the  $P_2$  subtype of the P-type ATPase superfamily and is composed of two subunits.<sup>12</sup> The informative database website <http://www.patbase.kvl.dk/>, classifies  $\text{Na}^+/\text{K}^+$ -ATPase as a  $P_{2C}$  ATPase. This class also includes the  $\text{H}^+/\text{K}^+$ -ATPases that will not be discussed here. The  $\text{Na}^+/\text{K}^+$ -ATPase catalytic alpha subunit (~113 kDa) contains 10 transmembrane segments, binds ATP as well as both sodium and potassium ions, and includes the amino acid phosphorylation site. The alpha subunit also contains three cytoplasmic domains: (1) the A- (anchor or actuator) domain, (2) the N-domain that binds the ATP nucleotide, and (3) the P-domain that contains the aspartic acid residue to be phosphorylated. The  $\text{Na}^+/\text{K}^+$ -ATPase amino acid residue numbering system used here is that from pig. The domains include the following: (1) N-domain, residues arg378–arg589; (2) P-domain, residues leu354–asn377 and ala590–leu773; (3) phosphorylation site, asp369. The smaller beta subunit (a ~55-kDa glycoprotein) is necessary for activity of the complex. It appears to be critical in facilitating plasma membrane localization and activation of the alpha subunit. Potassium and sodium cation transport occurs during a cycle of conformational changes in the enzyme that involve phosphorylation of a specific amino acid—asp369 in the P-domain. As currently understood, the sequence of events can be summarized as follows: (1) The pump, with bound ATP, binds three intracellular  $\text{Na}^+$  ions; (2) ATP is hydrolyzed, leading to phosphorylation of an amino acid residue within the cytoplasmic loop of the pump; (3) conformational changes in the pump expose the  $\text{Na}^+$  ions to the cell's exterior, where they are released; (4) the pump binds two extracellular  $\text{K}^+$  ions, perhaps concurrently with dephosphorylation of the alpha subunit; (5) ATP binds and the pump reorients to release  $\text{K}^+$  ions into the cytoplasm (the inside of the cell); and (6) the cycle begins again. The details of the cycle are better known for the  $\text{Ca}^{2+}$ -ATPase system and these are discussed in Section 6.4.2.

The first X-ray crystallographic structure of the  $\text{Na}^+/\text{K}^+$ -ATPase N-domain (at 2.6-Å resolution) was published in 2003, and the data deposited in the Protein Data Bank as PDB: 1Q3I.<sup>13</sup> (Note the PDB accession number contains number one “1” as its first digit and the letter “I” as its last.) This author’s explanation of the enzyme’s cycle includes the following details: (1) The A-domain is formed by a polypeptide segment between transmembrane  $\alpha$ -helices 2 and 3 and includes the enzyme’s N-terminal end; and (2) a larger loop (extending into the cell’s cytoplasm) between transmembrane  $\alpha$ -helices 4 and 5 forms the P-domain, into which the N-domain is inserted. As stated previously, the N-domain contains the ATP binding site and the P-domain contains the aspartic acid residue that is phosphorylated (asp369). Subsequently in the cycle, ADP is released and hydrolysis of the phosphorylated residue takes place. These events are coupled to the active transport of three sodium ions ( $3 \text{ Na}^+$ ) out of the cell and two potassium ions ( $2 \text{ K}^+$ ) into the cell. (See Figure 5.3.) The transmembrane helices are intimately involved in this transport. The A-, N-, and P-domains undergo large conformational movements during the transport cycle, alternating between the so-called  $E_1$  and  $E_2$  states. (See more

details in the discussion of  $\text{Ca}^{2+}$ -ATPase, Section 6.4.2.) An informative ribbon display of the PDB: 1Q3I  $\text{Na}^+/\text{K}^+$ -ATPase molecule is found in the <http://www.patbase.kvl.dk/> website, click on “Crystal Structures of P-Type ATPases.” The same diagram and others are found in Figure 2 of reference 13. Figure 3 of the same reference shows the primary and secondary structure alignment of the pig  $\text{Na}^+/\text{K}^+$ -ATPase with the rabbit sarcoplasmic reticulum  $\text{Ca}^{2+}$ -ATPase N-domain discussed in Section 6.4.2.

The PDB: 1Q3I  $\text{Na}^+/\text{K}^+$ -ATPase N-domain fold is described as a seven-stranded antiparallel  $\beta$ -sheet with an additional hairpin consisting of two  $\beta$ -strands. Five  $\alpha$ -helices of varying lengths and connecting loops complete the secondary structure. Several important motifs are conserved when the  $\text{Na}^+/\text{K}^+$ -ATPase N-domain is compared to that of  $\text{Ca}^{2+}$ -ATPase (PDB: 1EUL; see Section 6.4.2). Many of these motifs, conserved in most P-type ATPases, are important in forming the ATP-binding pocket, and many residues within these motifs are involved in ATP hydrophobic and electrostatic binding. First, in the  $\beta$ 1 strand, residues met379, val381, and met384 are conserved in both  $\text{Na}^+/\text{K}^+$ -ATPase and  $\text{Ca}^{2+}$ -ATPase (the  $\text{Na}^+/\text{K}^+$ -ATPase numbering system for PDB: 1Q3I is used here). These hydrophobic residues are mostly directed away from the enzyme’s surface. The second patch of conserved residues in  $\alpha$ -helix 1 include leu414, ile417, and leu420. These residues pack against the beta strands  $\beta$ 4,  $\beta$ 5,  $\beta$ 6, and  $\beta$ 7. The conserved asn422 residue, near the C-terminal end of  $\alpha$ -helix 1, hydrogen bonds to the likewise conserved arg464 side chain. The arg464 NH1 bonding distance to asn422’s main chain O equals 3.3 Å. At the beginning of  $\alpha$ -helix 2, the hydrophilic residues asp(glu)443, ser(thr)445, and glu446 are conserved (residues in parentheses are those found in  $\text{Ca}^{2+}$ -ATPase). These comprise the so-called DASE (or GD/(E)AS/(T)E) motif found in many P-type ATPases. Unfortunately,  $\text{Na}^+/\text{K}^+$ -ATPase residues phe475 through lys480 are disordered in the PDB: 1Q3I structure. It would be helpful to be able to visualize this region because it contains a conserved and essential phenylalanine residue that is invoked in many other ATPase structures in forming  $\pi-\pi$  interactions with the adenine ring of ATP. A highly conserved aa sequence, comprised of residues lys501, gly502, ala503, pro504, and glu505 (the so-called KGAPE motif), is located at the end of strand  $\beta$ 5 carrying on through the beginning of short  $\alpha$ -helix 4. In other N-domain structures, lys501 interacts directly with ATP. Finally,  $\beta$ 6 carries the motif RXL, with residues arg544 and leu546 defining the surface of the ATP-binding pocket. Despite several attempts, the reference 13 author found that nucleotides did not bind to the  $\text{Na}^+/\text{K}^+$ -ATPase N-domain described here; however, a possible ATP binding scheme agrees with that found in other P-type ATPases (see discussion for  $\text{Ca}^{2+}$ -ATPase found in Section 6.4.2). A hydrophobic binding site around leu546 has a suitable size to accommodate ATP’s adenine ring, a hydrophilic site near the N-terminal end of  $\alpha$ -helix 2 is suitable to interact with its ribose ring, and a positively charged site at arg544 may interact with the phosphate chain.

The reference 13 author compared the N-domain  $\text{Na}^+/\text{K}^+$ -ATPase X-ray crystallographic structure with NMR solution structures also published in 2003

using  $\text{Na}^+/\text{K}^+$ -ATPase from rat (PDB: 1MO7, without ATP, 1MO8, with ATP).<sup>14</sup> (Note that the PDB accession numbers contain the letter “O” and not the number zero “0.”) Comparison of the solid-state versus solution state structures reveals similarities and differences. The NMR structure features a six-stranded  $\beta$ -sheet rather than the seven-stranded one found for the X-ray structure, and the NMR structure  $\alpha$ -helices are shorter. Disorder found in the same regions of both structures (residues glu397–ala402 and two residues in the phe475–lys480 region) indicates that the methods used to study the system do not produce these structural artifacts. However, ATP was found to bind to the  $\text{Na}^+/\text{K}^+$ -ATPase N-domain, albeit weakly, in the solution structure (PDB: 1MO8). Nuclear Overhauser enhancements (NOEs) showed interactions between the ATP molecule and residues phe475, lys480, ala503, and leu546, indicating that these residues firmly anchor the ATP ribose moiety in the N-domain. In addition, phe475 provides good stacking interaction with ATP’s adenine ring. Only one hydrogen bond between ATP and the N-domain could be unequivocally assigned, namely between amino protons of the adenine base and carboxy oxygen Oe1 of gln482 (3.6 Å). The largest N-domain conformational changes occur in the beta sheets  $\beta$ 1 and  $\beta$ 6 (the so-called hinge region) that are bent in the native protein (PDB: 1MO7) and straight in the ATP-bound form (PDB: 1MO8). Since the  $\beta$ 1 and  $\beta$ 6 regions connect the N-domain to the P-domain, change in this region may affect the conformations of both the N- and P-domains in the intact  $\text{Na}^+/\text{K}^+$ -ATPase enzyme. Figure 2(d) of reference 14 illustrates the change. Unfortunately, the numbering system found in the *Nature Structural Biology* article (reference 14) does not agree with that found in the downloaded PDB: 1MO8 data. However, if one adds the number 7 to each residue number given in the *Nature Structural Biology* article, one finds the referred-to residues—that is, phe475 becomes phe482 and gln482 becomes gln489 when comparing the publication’s numbering to the PDB data downloaded for visualization. Hilge and co-workers believe that the nucleotide-free N-domain (1MO7) corresponds to the E<sub>2</sub> state in the intact enzyme in view of its structural similarity to the E2(TG) state (PDB: 1IWO) for  $\text{Ca}^{2+}$ -ATPase. (See discussion of this enzyme in Section 6.4.2 and Figure 6.28.) Additionally, they believe that ATP binding causes a shift toward the E1 state (PDB: 1SU4 in Section 6.4.2.1 and Figure 6.28). Finally, the authors speculate that, in analogy to  $\text{Ca}^{2+}$  binding in  $\text{Ca}^{2+}$ -ATPase, binding of  $\text{Na}^+$  ions causes the final N-domain conformational change that brings the  $\gamma$ -phosphate into the shallow cavity where it binds to the invariant aspartate residue.

In a molecular modeling study, Amler and co-workers established eight amino acid residues that form the ATP binding site in the intact  $\text{Na}^+/\text{K}^+$ -ATPase enzyme.<sup>12</sup> These workers expressed point-mutated versions of a part of the H<sub>4</sub>–H<sub>5</sub> loop (residues leu354–ile604) of the  $\text{Na}^+/\text{K}^+$ -ATPase enzyme that contains the ATP binding residues. The so-called H<sub>4</sub>–H<sub>5</sub> loop contains N- and C-terminal subdomains of the  $\text{Na}^+/\text{K}^+$ -ATPase P-domain and the entire N-domain. This loop has a self-supporting structure, retains the ability to bind ATP, but lacks interactions with other  $\text{Na}^+/\text{K}^+$ -ATPase domains, and is not

affected by  $\text{Na}^+$ - or  $\text{K}^+$ -dependent conformational changes. A fluorescent analog of ATP, TNP-ATP [ $2',3'-O$ -(2,4,6-trinitrophenyl)adenosine 5'-triphosphate, trisodium salt], was used so that binding could be visualized using fluorescence spectroscopy. FT-IR spectroscopy was used to show that mutations did not change the secondary loop structure. The following mutants were studied: the double mutant I417N-N422A and single mutants R423L, Q482L, E505Q, F548G, and F548Y. It was found that the dissociation constants for TNP-ATP binding were significantly increased for nearly all point mutations but not for ATP binding to the I417N-N422A and E505Q mutants. Previously, these workers and others had shown that mutations of phe475 and glu446 resulted in inhibition of ATP or TNP-ATP binding to the  $\text{H}_4\text{-H}_5$  loop. Molecular modeling carried out by researchers in reference 12 showed that the aromatic ring of phe475 and ATP's adenine ring are parallel to each other at a distance of 0.3 nm, resulting in  $\pi$ - $\pi$  stacking interactions between them. Negatively charged glu446 forms a hydrogen bond to the  $\text{NH}_2$  donor of the adenosine moiety. Gln482 was shown to stabilize a side chain of glu446 through hydrogen bonding, thus stabilizing glu446's position. Phe548 was mutated to both tyrosine (Y) and glycine (G) with the observations that there was strong inhibition of ATP binding for the F548Y mutant but weak inhibition for the F548G mutant. Steric hindrance in the ATP-binding site by the large tyrosine (Y) residue was invoked to explain this observation. In addition, the authors observed that phe548, although playing a minor role in direct interaction with ATP, does help support and maintain the ATP-binding pocket structure. Surprisingly, the R423L mutant showed the strongest inhibition of ATP binding, although the arginine residue lies outside the ATP-binding pocket. The authors concluded that the arginine residue forms a strong hydrogen bond with glu472, and this bonding contributes to pocket stability and orientation of ATP-binding residues phe475, lys480, and gln482 into their proper positions. The S445A and E505Q mutations had no significant effect on ATP binding, indicating their minor role in ATP recognition or binding. In conclusion, the authors showed that the ATP recognition and binding pocket includes not only the previously identified important residues of lys480, lys501, gly502, and cys549, but also includes residues glu446, phe475, gln482, and phe548.

In comparing  $\text{Na}^+/\text{K}^+$ -ATPase structures to those of other P-type ATPases, we find that available structures for the  $\text{Na}^+/\text{K}^+$ -ATPase system are not as robust and informative as those for  $\text{Ca}^{2+}$ -ATPase. The reader is referred to Section 6.4.2 for further discussion of the  $\text{Ca}^{2+}$ -ATPase system.

## 5.4.2 Potassium ( $\text{K}^+$ ) Ion Channels

**5.4.2.1 Introduction.** Studies on ion channels, potassium ion channels in particular, carried out by Roderick MacKinnon and his research group, resulted in his receiving the 2003 Nobel Prize in chemistry "for structural and mechanistic studies of ion channels." Thanks to this contribution and research continuing in the MacKinnon and many other laboratories, it is now possible to

visualize ions flowing through channels that can be opened and closed by different cellular signals. In 1998, the MacKinnon group published the first structural study of a potassium ion channel, a feat welcomed by the scientific community eager to learn the physical details of ion channel workings (PDB: 1BL8).<sup>15</sup> Since then, a variety of potassium and other ion channel structures have been published, leading to a greater understanding of these important systems. The potassium ion channel structural studies will be detailed in the following sections. Following their initial success, the MacKinnon group published the structure of the cytoplasmic  $\beta$  subunit-T1 complex of voltage-dependent K<sup>+</sup> ion channels in 2000 showing the connection between the ion channel pore and these larger structures within the cytoplasm. The article also explained how inactivation peptides reach their action site through lateral, negatively charged openings between the cytoplasmic  $\beta$  subunit-T1 complex and the ion channel pore (PDB: 1EXB).<sup>16</sup> In 2001, MacKinnon's group published the structure of a potassium ion channel–monoclonal antibody (Fab) complex that further explained how ions to be transported have their hydration shells removed or reattached and how the ion channel selectivity filter changes its coordination pattern in response to high or low K<sup>+</sup> ion concentration (PDB: 1K4C, 1K4D).<sup>17</sup> The MacKinnon group next turned its attention to voltage-dependent K<sup>+</sup> ion channels that open to allow ion passage in response to changes in cell membrane voltage. The research identified “voltage-sensor paddles” that move in response to membrane voltage changes and that carry positive charge across membranes (PDB: 1ORS, 1ORQ).<sup>18</sup> More recently, in 2005, MacKinnon and co-workers have published details of *Shaker* family voltage-dependent K<sup>+</sup> (K<sub>v</sub>) ion channels to show how voltage sensors operate to open and close the ion pore and how the complex's  $\beta$  subunits help regulate the channels (PDB: 2A79).<sup>19</sup> Also in 2005, the MacKinnon group presented two structures of the voltage-dependent channel KvAP with and without monoclonal antibody fragments (PDB: 2A0L).<sup>20</sup> In 2006, the group published findings on ion selectivity in a semisynthetic K<sup>+</sup> channel locked into a conductive conformation (PDB: 2IH1 and 2IH3).<sup>21</sup>

Physically, ion channels are tiny pores that stud the surface of all cells. The ion channels are important for, among other things, the function of muscles and the nervous system. These channels allow the passage of potassium, calcium, sodium, and chloride ions. Through a balance of electrical forces and chemical bonds, ion channels are specific for one ion; for instance, a potassium ion channel will reject a sodium ion trying to enter its channel. An excellent visualization of the overall process is found at the website <http://www.rockefeller.edu/pubinfo/howkion.html>. It will be helpful to look at this website before going any further in the discussion.

Potassium channels are part of a complex system that helps maintain the normal ionic balance across the cell membrane. In excitable cells, like those in nerves and muscles, the channels also help reestablish the electrical difference between the inside and outside of the cells after excitation. In the case of neuron firing, potassium ions, and thus positive charge, builds up inside the

cell. The charged ions create an electrical voltage across the cell membrane. This buildup in turn causes the potassium channel to open, spilling out potassium ions and restoring the cell to its resting state. The proteins that manage the neuronal system are called potassium voltage-gated ion channels. The so-called “action potential” of nerve cells is generated when an ion channel on the surface of a nerve cell is opened by a chemical signal sent from an adjacent nerve cell, whereupon an electrical pulse is propagated along the surface of the nerve cell through the opening and closing of other ion channels in the course of a few milliseconds. Rapid-fire opening and closing of these channels releases ions, moving electrical impulses from the brain in a wave to their destination in the body. Without potassium and sodium channels, neurons could not generate electrical signals and hearts could not beat rhythmically. In this discussion, we will concentrate on several potassium ion channel proteins whose function it is to generate nerve impulses—the electrical activity that underlies all movement, sensation, and perhaps even thought. Malfunctioning ion channels contribute to disease states such as diabetes, epilepsy, arrhythmia (irregular heart beat), and others. The research now being carried out by the many groups studying ion channel structure and function may play an important role in the development of drugs to deal with these diseases. Finding protein sequence information (the primary sequence of amino acids in a protein chain) is often important to understand researchers’ explanations of their experimental work, especially for the large variety of potassium ion channels being discussed in the next section. An excellent database for finding sequence information is the ExPASy (Expert Protein Analysis System) proteomics server of the Swiss Institute of Bioinformatics (SIB) at <http://www.expasy.org/>. At this site, search the protein knowledgebase Swiss-Prot and TrEMBL.

**5.4.2.2 X-Ray Crystallographic Studies.** In 1998 when the first X-ray crystallographic structure of a K<sup>+</sup> channel was published by his group (PDB: 1BL8), MacKinnon and fellow researchers had been studying potassium ion channels for at least a decade. They had used site-directed mutagenesis to determine a “signature sequence” of eight amino acids that were essential to the K<sup>+</sup> channel’s function. The mutations were carried out on the so-called *Shaker* potassium channel from fruit flies whose DNA (and therefore amino acid) sequence had been determined in 1987.<sup>22</sup> At this time, researchers also knew that K<sup>+</sup> channels consisted of four subunits, with each one contributing its five amino acid signature sequence—TVGYG—to form a selectivity filter for potassium ions to the exclusion of all others. However, to know with certainty what the chemistry and mechanism of potassium ion selectivity could be, structural analysis at atomic scale was needed. The potassium ion channel history is complex and somewhat confusing. It will be helpful if the reader has a full complement of the figures contained in this section and those recommended for viewing from the publications referenced here. Information on the various potassium ion channel structural studies is collected in Table 5.3. The

**TABLE 5.3 X-Ray Crystallographic Structures of K<sup>+</sup> Channels**

Species	K <sup>+</sup> Channel Type	Number of a Residues/ Molecular Mass (kDa)	PDB Accession Number, Resolution (Å)	Chapter 5 Reference Number	Year of Publication	Reference Figure Numbers	Activation Gate Open/ Closed
<i>Streptomyces lividans</i> (KcsA)	Weakly voltage-dependent	K <sup>+</sup> channel 97/10.3	1BL8, 3.02	15	1998	5.5A, 5.5B	Closed
<i>Rattus norvegicus</i> (Kv1.1)	Voltage-dependent	Kv β2 subunit 332/372, T1 domain	1EXB, 2.10	16	2000	5.6, Figure 5, reference 16 <sup>a</sup>	Closed
<i>Streptomyces lividans</i> (KcsA)	Weakly voltage-dependent	K <sup>+</sup> channel <sup>b</sup> 103/13.2	1K4C, 2.00 1K4D, 2.30	17	2001		Closed
<i>Methanothermobacter thermotrophic</i> (MthK)	Ca <sup>2+</sup> -gated	K <sup>+</sup> channel 124/13.2 336/37.4	1LNQ, 3.30	23	2002		Open
<i>Methanothermobacter thermotrophic</i> (MthK)	Ca <sup>2+</sup> -gated	K <sup>+</sup> channel 234/25.9	2AEF, 1.70 (with Ca <sup>2+</sup> ) 2AEJ, 2.10 (without Ca <sup>2+</sup> ) 2AEM.		Footnote c <sup>c</sup>	2005	
<i>Aeropyrum pernix</i> (KvAP)	Voltage-dependent	K <sup>+</sup> channel <sup>b</sup> 223/24.3	2.80 (RCK domain) 1ORQ, 3.20 (contains Y46C mutation)	18, 25	2003	5.7, 5.8 Figures 3a, 3c, and 8, reference 18 Figures 5a and 5b, reference 25	Open

<i>Aeropyrum pernix</i> (KvAP)	Voltage-dependent Inward-rectifier	K <sup>+</sup> channel <sup>b</sup> 132/14.6	1ORS, 1.90	18, 25	2003	Open
<i>Burkholderia pseudomallei</i> (KirBac1.1)		K <sup>+</sup> channel with cytosolic domains	1P7B, 3.65		2003	Closed
<i>Rattus norvegicus</i> (Kv1.2)	Voltage-dependent	K <sup>+</sup> channel with other subunits (T1 domain) 499/56.8; β2 subunit	2A79 337/37.2	19, 37, 38	2005	Figures 2c and 4b, reference 37 Figures 3c, 3d, and 4, reference 38
<i>Aeropyrum pernix</i> (KvAP)	Voltage-dependent	K <sup>+</sup> channel <sup>d</sup> 241/26.2	2A0L	20	2005	Open
<i>Streptomyces lividans</i> (KcsA) synthetic gly77-d-ala mutant	Voltage-dependent	K <sup>+</sup> -channel <sup>b,e</sup> mutations	2IH1, 2.40 S69A and V70C	21	2006	Open
<i>Streptomyces lividans</i> (KcsA) synthetic gly77-d-ala mutant	Voltage-dependent	K <sup>+</sup> -channel <sup>b,e</sup> mutations	2IH3, 1.72 S69A and V70C	21	2006	Open

<sup>a</sup>In Figure 5 of reference 16, the pore region is based on the KcsA K<sup>+</sup> channel from PDB: 1BL8.

<sup>b</sup>Co-crystallized with Fab monoclonal antibodies from mouse.

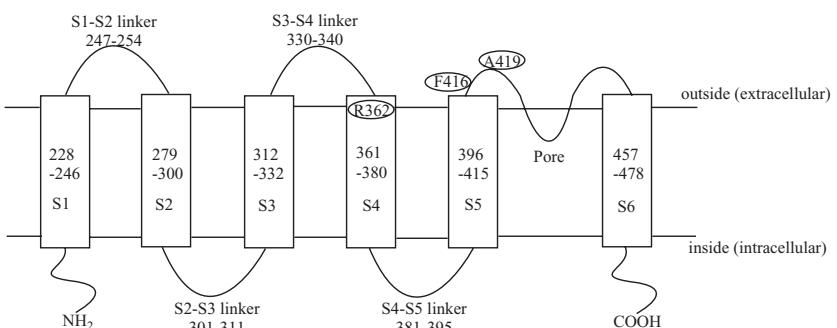
<sup>c</sup>Dong, J.; Shi, N.; Berke, I.; Chen, L.; Jiang, Y. *J. Biol. Chem.*, 2005, **280**, 41716–41724.

<sup>d</sup>Co-crystallized with Fv monoclonal antibody fragments from mouse.

<sup>e</sup>Low [K<sup>+</sup>] for 2IH1; high [K<sup>+</sup>] for 2IH3.

Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB-PDB) website at <http://www.rcsb.org/pdb/home/home.do> contains a wealth of information about each biomolecule whose structural data have been deposited. Enter the PDB accession number in the search routine, then click on the “biology and chemistry,” “material and methods,” “sequence details,” and “geometries” tabs to obtain more information. A schematic diagram of the membrane topology of the Shaker potassium ion channel is shown in Figure 5.4. The diagram shows the S1–S6  $\alpha$ -helices that surround the potassium ion channel pore. More detail about S1–S6 is given in the discussion of X-ray crystallographic determinations and other methods for studying the structure and function of the pore domain that follows.

The *Shaker* potassium channel did not yield crystals for the first potassium ion channel X-ray crystallographic structure. Rather, it was the prokaryote bacterium *Streptomyces lividans*, abbreviated as the KcsA  $K^+$  channel, that first produced crystals suitable for crystallography. This structure, published in 1998 by MacKinnon’s group in *Science* magazine, was received with great praise from the scientific community. It was known at that time that the amino acid sequence of KcsA was similar to that of other  $K^+$  channels, including vertebrate and invertebrate voltage-dependent  $K^+$  ( $K_v$ ) channels (such as the



#### Shaker $K^+$ Channel Protein

Residue numbers given inside each transmembrane segment S1 - S6

Pore contains selectivity filter residues 442 - 447

Shaker residue numbers from <http://www.expasy.org/uniprot/P08510>

N-terminal residues 1-227

C-terminal residues 479-616

#### Comparable KvAP residues (PDB: 1ORQ, 1ORS)

(differs from PDB residue numbering system)

from <http://www.expasy.org/uniprot/Q9YDF8>

S1 39-63 (PDB:26-50)

S2 68-92 (PDB:55-79)

S3a 97-105 (PDB:84-92)

S3b 109-125 (PDB:96-112)

S4 129-145 (PDB:116-132)

S5 160-184 (PDB:147-171)

S6 222-253 (PDB:209-240)

N-terminal 1-38

C-terminal 254-295

pore 196-208 (PDB:183-195)

selectivity filter 209-214 (PDB: 196-201)

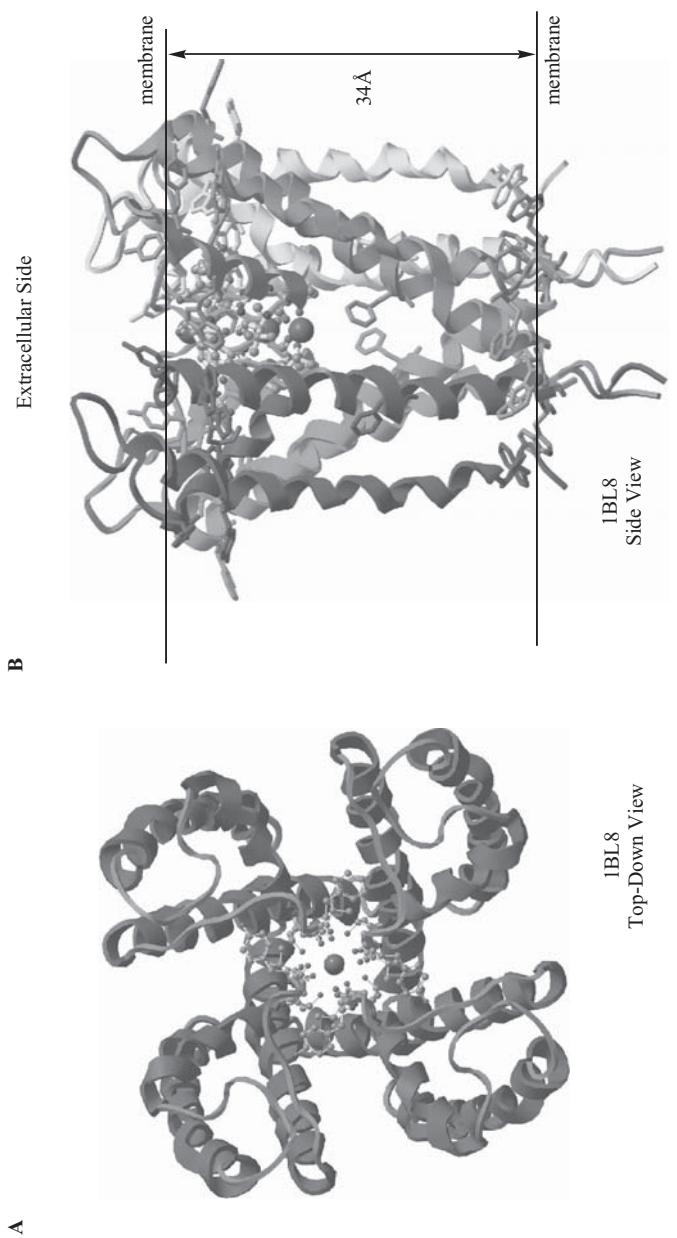
**Figure 5.4** Membrane topology for the Shaker potassium ion channel with important residues circled. (Adapted with permission from Figure 1A of reference 29. Copyright 2003, with permission from Elsevier.)

*Shaker* potassium channel), vertebrate inward rectifier K<sup>+</sup> channels (K<sub>ir</sub>), Ca<sup>2+</sup>-activated K<sup>+</sup> channels, and other potassium ion channels. It was also common knowledge that there are two closely related varieties of K<sup>+</sup> channels, with one type containing two membrane-spanning segments per subunit and the other containing six. In all cases, the functional K<sup>+</sup> channel is a tetramer of four usually identical subunits. KcsA is a prokaryotic two-membrane-spanning K<sup>+</sup> channel, although its amino acid sequence is closer to that of the eukaryotic six-membrane K<sup>+</sup> channel proteins. Comparisons of amino acid sequences for various species indicate that in the potassium ion pore—the region where the K<sup>+</sup> channel crosses the cellular membrane—the K<sup>+</sup> channel signature sequence (TVGYG) is identical to that found for *Shaker* proteins and for vertebrate voltage-gated K<sup>+</sup> channels. Additional research has shown that the KscA potassium ion pore structure is nearly identical for all K<sup>+</sup> channel proteins. K<sup>+</sup> channels are classified as “long pore channels” because multiple ions can queue inside the long narrow pore in single file. The pores of all K<sup>+</sup> channels can be blocked by tetraethylammonium (TEA) ions.

KcsA crystals suitable for X-ray crystallographic analysis using synchrotron radiation were obtained and the data collected and analyzed for multiple crystals and six different data sets as described in the 1998 *Science* publication (reference 15). The final KcsA pore structure, including amino acid residues 23 to 119 of the K<sup>+</sup> channel, refined to 3.2 Å. The X-ray data were deposited in the Protein Data Bank with the accession number 1BL8.

A top-down view of the tetrameric pore helices is shown in Figure 5.5A. The tetramer of four identical subunits created an inverted teepee, or cone, shape with the important ion selectivity filter in its extracellular facing end as shown in Figure 5.5B. The four chains of the tetramer are shown in different colors. The colors emphasize that each chain contains an outer helix and an inner helix. Three potassium ions (space fill gray) and one water molecule (red small sphere) are shown in the selectivity filter. The potassium ions in the selectivity filter are surrounded by the backbone oxygen atoms of the residues gly77, tyr78, and gly79, shown in Figure 5.5A and B in ball-and-stick form. The entire ion conduction pore in the PDB: 1BL8 structure was found to be 45 Å long with a variable diameter, while the selectivity filter region had a length of 12 Å. The membrane thickness is ~34 Å. Note that aromatic, mostly hydrophobic, amino acid residues—tryptophan, phenylalanine, and tyrosine—collect near the membrane surfaces. The selectivity filter region separating the rest of the pore from the extracellular solution was found to be so narrow that K<sup>+</sup> ions need to shed their hydration shell to enter it. Following the selectivity filter, potassium ions enter the internal pore and cavity that are lined with predominantly hydrophobic residues, a general feature of K<sup>+</sup> channels. These sections of the pore have diameters that will accommodate hydrated K<sup>+</sup> ions. More details are found below.

Two aspects of potassium ion conduction across membranes were of particular interest to these researchers and to many other biophysicists and biochemists working in this area: (1) What is the chemical basis for the ion



**Figure 5.5** (A) Top-down view of the  $\text{K}^+$  ion channel, PDB: 1BL8. (B) Side view of the  $\text{K}^+$  ion channel, PDB: 1BL8. Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. (Printed with permission of CambridgeSoft Corporation.) (See color plate)

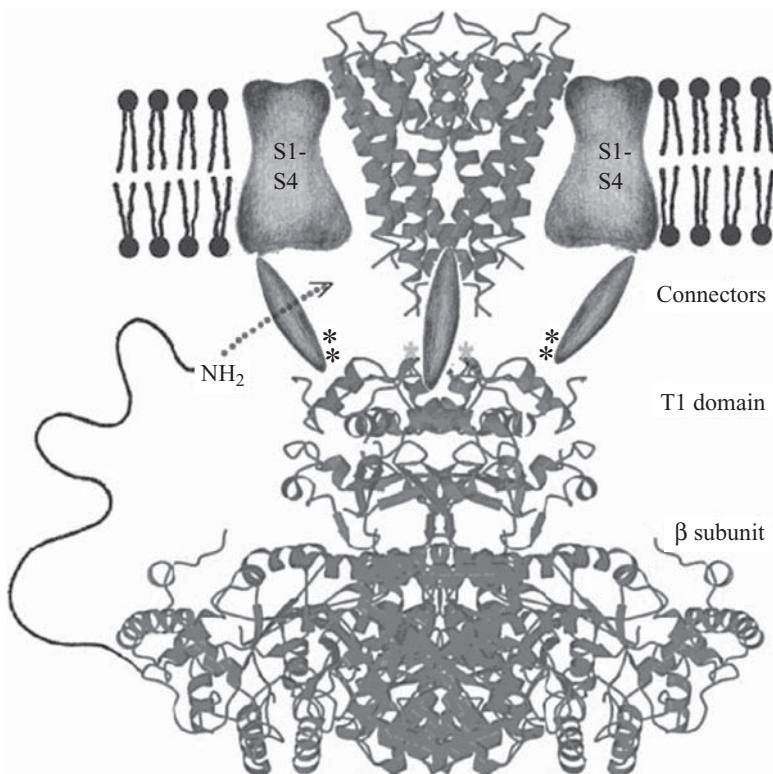
selectivity exhibited by these proteins [faithful discrimination between K<sup>+</sup> (radius 1.33 Å) and the smaller Na<sup>+</sup> (radius 0.95 Å) ions, for instance], and (2) how can the K<sup>+</sup> channels be highly selective and at the same time move ions through membranes at extremely rapid rates (approaching the diffusion limit). Potassium ion channels exclude smaller alkali metal cations Li<sup>+</sup> (radius 0.60 Å) and Na<sup>+</sup> (radius 0.95 Å) but allow permeation of the larger Rb<sup>+</sup> (radius 1.48 Å) and Cs<sup>+</sup> (radius 1.69 Å). In fact, Rb<sup>+</sup> is used as a K<sup>+</sup> analog within potassium ion channel pores in this KcsA study because it is more electron-dense and thus easier to visualize by X-ray diffraction. Both Rb<sup>+</sup> and Cs<sup>+</sup> were used to show that the selectivity filter contains two ions about 7.5 Å apart. In this study, diffuse electron density indicates a third ion, perhaps surrounded by a hydration shell, in the large diameter cavity (~10 Å across) below the selectivity filter and one or more poorly defined (by electron density) cations continuing through the pore toward the cell's interior. The ion found in the central cavity overcomes the electrostatic destabilization resulting from its hydrophobic environment by assuming a shell of polarizable water. In addition, four pore helices point toward the center of the cavity in an orientation that imposes a negative electrostatic potential—attractive to the cation—at the edges of the cavity. These two features, the aqueous cavity and the helical orientation, solve the problem of the electrostatic barrier to a cation crossing a lipid bilayer (the membrane). The authors then ask an important question: What is the significance of the internal pore's hydrophobic lining? They believe the answer lies in the high ion throughput necessary for the pore's efficiency. If the potassium, or any other ion being transported, were to be attracted and held by hydrophilic residues in the pore, rapid ion transport would be impossible. In other words, the chemical and structural design of this part of the pore facilitate high ion throughput in this longer segment of the pore. The rate-limiting step for K<sup>+</sup> ions traversing the channel is limited to the shorter 12-Å length and ~2.5-Å diameter of the selectivity filter. The selectivity filter itself exhibits two essential features: (1) The main chain carbonyl oxygens of the amino acid residues (thr75, val76, gly77, tyr78, gly79, TVGYG) lining the filter form an oxygen atom stack suitably positioned to coordinate the dehydrated K<sup>+</sup> ions as they pass through, and (2) the valine and tyrosine residues in the conserved sequence TVGYG have their side chains pointing out of the filter toward tryptophan residues in the pore helices surrounding the filter. The aromatic residues interact through van der Waals contacts and through hydrogen bonding—for instance, between tyr hydroxyls and tryptophan nitrogens. The reference 15 authors state that the structure appears to behave "like a layer of springs stretched radially outward to hold the pore open at its proper diameter." In fact, this structure prevents the accommodation of the smaller Li<sup>+</sup> and Na<sup>+</sup> ions with their smaller radii. How does the selectivity filter move K<sup>+</sup> ions along when attractive forces would keep them in the filter? The authors believe that a single K<sup>+</sup> ion would be held tightly but that a second K<sup>+</sup> ion entering causes a repulsion that allows the first K<sup>+</sup> to move along.

Conclusions from the 1998 PDB: 1BL8 structural study are as follows: (1) The K<sup>+</sup> channel pore is constructed of an inverted teepee with the selectivity filter at the wide (extracellular-facing) end; (2) the narrow selectivity filter is 12 Å long (limiting the distance of strong K<sup>+</sup> attraction), while the remainder of the pore is wider and lined with hydrophobic residues; (3) a large water-filled central cavity and helix dipole orientation help overcome a high electrostatic energy barrier facing cations in the hydrophobic environment of the membrane center; (4) the K<sup>+</sup> selectivity filter is lined with carbonyl oxygen atoms spaced for proper coordination of potassium ions but too far apart to coordinate smaller sodium ions; and (5) repulsion between two K<sup>+</sup> ions close to each other in the selectivity filter overcomes the strong attraction between the potassium ions and carbonyl oxygens of filter amino acid residues, allowing rapid ion conduction.

The MacKinnon group next published the X-ray crystallographic structure (at 2.1-Å resolution) of a eukaryotic (rat) voltage-dependent K<sup>+</sup> channel β subunit: T1 cytoplasmic (cell interior) assembly attached to the pore subunit described in reference 16 (PDB: 1EXB). First, it is important to describe the two subunits of the K<sup>+</sup> channel. The α subunit includes the K<sup>+</sup> channel ion conduction pore residing in the cell membrane and the so-called T1 domain that resides in the cell interior. The pore and T1 domain are connected by the voltage-sensing region and other connectors of unknown structure at the time of this research. The β subunit is a tetramer of oxidoreductase proteins with fourfold rotational symmetry as was found for the K<sup>+</sup> channel ion conduction pore discussed above. Each β subunit contains an active site with an NADPH (nicotinamide adenine dinucleotide phosphate) cofactor, but its substrate and biological function were unknown. It was known that the α and β subunits assemble in the endoplasmic reticulum and remain together throughout their active life. The α subunit's T1 domain, consisting of approximately 100 amino acids, resides between the membrane-situated ion conduction pore and the β subunit. It forms a tetrameric ring with a narrow positively charged central core. Figure 5 of reference 16 (Figure 5.6) shows an illustrative composite model of the entire assembly. The model of the pore region is modeled using the PDB: 1BL8 structure.

It was known that large organic cations such as the tetraethylammonium (TEA) ion and large inactivation peptides (part of the β subunit—T1 assembly) enter the transmembrane pore and the question arises as to how that would take place given the T1 domain's narrow central core. These researchers answer the question in reference 16 through their analysis of the structure and function at the cytoplasmic interface (the connection region between the α and β subunits).

First, the researchers confirmed that the β subunit's association with the α subunit is disrupted if the T1 domain is removed. The X-ray crystallographic structure of the assembled T1<sub>4</sub>β<sub>4</sub> complex showed why this is the case. The structure shows that large flat surfaces of the β subunit interact with four prominent loops—called contact loops—that extend from the T1 tetramer's



**Figure 5.6** Composite model of a voltage-dependent  $K^+$  channel. (From Figure 5 of reference 16. Reprinted with permission of AAAS.) (See color plate)

N-terminal end. The researchers also found that inactivation peptides are effective when attached to the N-terminal ends of either the  $\alpha$  or  $\beta$  subunits of the  $T1_4\beta_4$  complex apparently because these sites are close to each other in the complex. Through several experiments co-expressing elements of the different  $T1_4\beta_4$  complexes the authors determined that the T1 domain does not participate directly in activation but serves to hold the  $\beta$  subunit in place.

Single-site amino acid alterations were carried out to show that mutations in the T1 domain loop region disrupt  $\beta$  subunit binding to the  $\alpha$  subunit and abolish inactivation. Similar mutations on the  $\beta$  subunit surface that contacts the T1 domain also affected inactivation. The conclusion is that both subunits are necessary for inactivation peptides to carry out their function.

The aqueous channel down the center of the  $T1_4\beta_4$  complex is narrow ( $\sim 4 \text{ \AA}$ ) and positively charged. This channel could not allow passage of the large TEA ion or an inactivation peptide. So how do these species navigate through the  $T1_4\beta_4$  complex to reach the ion conduction pore? Or “How does the ion pathway connect to the cytoplasm?” in the words of the reference 16 authors.

Inactivation peptide interaction with the assembly was used to find answers to this question. It was known that mutation of basic (positively charged) amino acids to neutral or hydrophobic residues on the inactivation peptide influenced the extent of inactivation. Therefore, the researchers looked for areas of acidic (negatively charged) amino acids at the known site of inactivation peptide binding using another voltage-dependent K<sup>+</sup> channel protein, Kv1.4. They found a cluster of acidic residues (glu273, asp274, and glu275 for Kv1.4) at the T1–S1 linker region—the area where the C-terminal end of the T1 domain interfaces with connectors to the membrane spanning helices (S helices) that surround the ion conduction pore. (See black asterisks on Figure 5.6.)

When these acidic residues were mutated to alanine (neutral) or lysine (positive), rates of inactivation were affected. The largest effect took place with the positive amino acid mutation. The authors found that mutations in the inactivation peptide are tied to those in the T1–S1 linker near the T1 domain; in other words, the inactivation peptide must be near the T1–S1 linker when it inactivates the pore. The conclusion reached is that since the inactivation peptide cannot fit through the center of the T1<sub>4</sub>β<sub>4</sub> complex, it must reach the ion conduction pore through lateral openings above the T1 tetramer.

From their work reported in reference 16, the research group reached the following conclusions: (1) The T1 domain forms a docking platform for the necessary β subunit through loops at the N-terminal end of the T1 domain and amino acid residues on the T1-facing surface of the β subunit; (2) lateral openings between the T1 domain and the cell membrane-spanning region of the α subunit form a conduit for an inactivation peptide (and probably K<sup>+</sup> ions as well); and (3) the central question of why K<sup>+</sup> channels contain an oxidoreductase enzyme subunit (the β subunit) remains unanswered, although the authors suggest that interactions between α and β subunits could allow cellular redox regulation of the channel.

The MacKinnon group continued its study of K<sup>+</sup> channels and presented the structure of KvAP, the voltage-dependent K<sup>+</sup> channel from the thermo-philic archaeabacteria *Aeropyrum pernix* in 2003.<sup>18</sup> In this publication, the process known as gating is explained. The control of pore opening or closing is related to the membrane voltage—that is, we have a voltage-dependent channel. First, it is known that the cell membrane can undergo transient changes in permeability to its selected ion and that these changes depend on membrane voltage. As this article puts it: “Selective permeability to ions determines the membrane voltage and the voltage determines the permeability.” The question then is, By what means does the voltage-dependent channel know the membrane voltage and open (or close) its gate as a function of the voltage value? It is believed that charged amino acids called “gating charges” move through the membrane electric field and that the transient electric currents generated in this manner affect pore opening and closing. K<sup>+</sup> channels have a large gating charge, giving rise to an “open” probability as a function of membrane voltage.

It is known that all voltage-dependent ion channels contain six hydrophobic segments per subunit, referred to as S1–S6. In K<sup>+</sup> channels four identical S1–S6 subunits surround the central ion-conduction pore (the pore is structurally detailed for the KscA K<sup>+</sup> channel in reference 15 as discussed earlier). Segments S5–S6 line the pore and determine ion selectivity while S1–S4 form the voltage sensors. Charged amino acids, particularly the first four arginines in S4 (arg117, arg120, arg123, and arg126 in KvAP), account for most of the gating charge. The reference 18 researchers have studied the mechanism of voltage-dependent gating using biochemical, X-ray crystallographic, and electrophysiological methods with special attention to KvAP whose structure is described in the 2003 *Nature* publication. It should be said that the KvAP amino acid sequence and its electrophysical properties are closely related to those of other K<sup>+</sup> channels. Reference 18 describes the KvAP structure as well as two conformations of the important voltage sensor.

First, the researchers found that the voltage sensor region was very difficult to crystallize, leading to the idea that it might be a mobile structure within the K<sup>+</sup> channel assembly. A more stable (and therefore less mobile) voltage sensor domain might be found for the thermophilic KvAP species; however, even in this case monoclonal antibodies (Fab) raised against the K<sup>+</sup> channel were necessary to “hold” the channel in an orientation to foster crystallization. This approach proved successful for two structures: (1) the full-length channel in complex with Fab determined at a resolution of 3.2 Å (PDB: 1ORQ); and (2) the isolated voltage sensor domain in complex with Fab determined at a resolution of 1.9 Å (PDB: 1ORS).

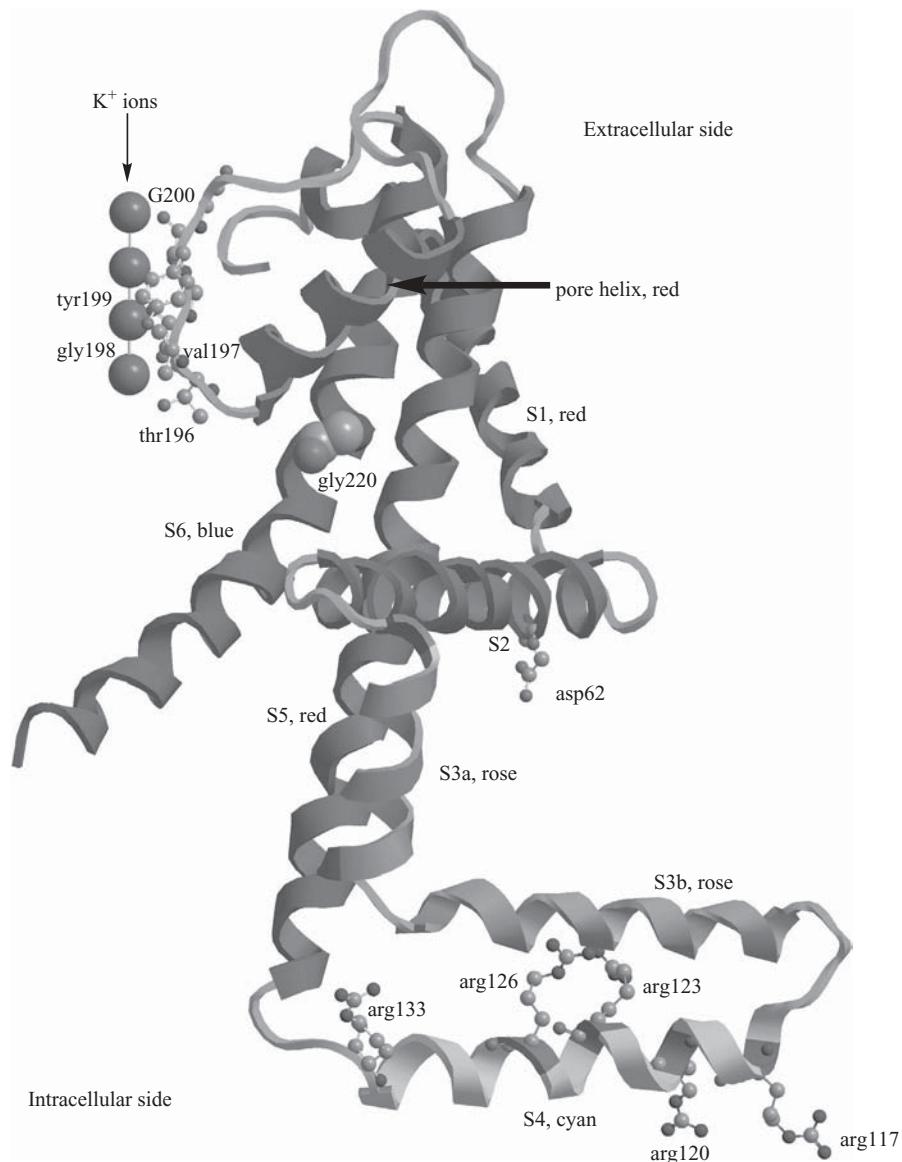
In the 1ORQ structure, the ion conduction pore with its selectivity filter (conserved sequence TVGYG) looks virtually identical to that of KcsA (PDB: 1BL8) discussed above. In the hydrophobic core surrounding the selectivity filter, some deviation in amino acid sequence from KcsA is found, but the secondary structure overlap between KvAP and KscA is essentially complete. These findings are not surprising in that the filter serves the same purpose in both proteins—the selective and rapid conduction of potassium ions. The structures of KvAP (PDB: 1ORQ) and KscA (PDB: 1BL8) begin to deviate in the area of the S6 helices (inner helices), diverging most notably at a point referred to as the glycine-gating hinge. The gating hinge appears to affect the openness of the conduction pore, as can be seen when comparing the structures of different pores. These authors observe that the “inner helices seem to open like the aperture of a camera by bending at a glycine residue.” When comparing the KcsA (closed K<sup>+</sup> channel) with that of MthK (an opened K<sup>+</sup> channel), the researchers find that the KvAP channel is opened wider than the closed KscA channel and is nearly as wide open as the MthK channel. MthK is a calcium-gated potassium channel whose structure was determined by the MacKinnon group in 2002.<sup>23</sup> The S6 (inner relative to the pore) and S5 (outer) helices are in an antiparallel conformation to one another and probably move together as a single unit. This is important in that the voltage sensor domain (to be discussed next) is directly attached to the outer S5 helix so that

the voltage sensor could pull on the outer helix, thereby affecting the inner S6 helix and opening the pore.

The voltage sensor, S1–S4 helices, are attached to the pore by forming concentric helices around it. Looking at Figure 3a of reference 18, one sees that S5 and S6 are closest to the pore while S1 and S2 form another concentric layer of helices outside S5, and helices S3 and S4 are located at the K<sup>+</sup> channel's outer perimeter. Additionally, helix S3 appears to be two individual helices, designated S3a and S3b. Figure 3c of reference 18 and Figure 5.7 show that the pore helix is close to the external (extracellular) side of the membrane while helices S3b and S4 straddle the membrane at its internal side facing the cytoplasm.

Helices 3b and the N-terminal end of the S4 helix pack tightly against each other and form a mostly hydrophobic helix–turn–helix structure called the “voltage sensor paddle.” In contrast to its mostly hydrophobic amino acid composition, the S4 helix contains four important, neatly spaced, basic arginine residues (arg117, arg120, arg123, and arg126) that are highly conserved among voltage-dependent K<sup>+</sup> channels. These four arginine residues are believed to account for most of the gating charge responsible for opening and closing the pore. Finding the S4 helix arranged perpendicular to the pore helix at the *intracellular* side of the membrane was an unexpected finding because it contradicts many previously established electrophysiological measurements. The previous measurements show that toxins and thiol-reactive compounds react with the N-terminal portion of S4 at the membrane's extracellular side but not from the intracellular side.

To address these contradictions the researchers crystallized only the S1–S4 portion of the protein (PDB: 1ORS) in the presence of a Fab different from that used in the 1ORQ structure. The X-ray crystallographic structure of the voltage sensor portion only (PDB: 1ORS), at a resolution of 1.9 Å, has a similar conformation to that in 1ORQ and behaves as a normal voltage sensor. However, the atomic resolution shows more detail of the interactions among the helices in the voltage sensor paddle. The first four arginine residues (117, 120, 123, and 126) on S4 are exposed to solvent in the crystal. The fifth arginine residue (arg133) makes a salt bridge with asp62 in S2. This has the effect of bringing the S2 and S4 helices closer together right beside the S3 loop connecting S3a and S3b. Further salt bridges form between arg76 in S2 with asp72 (also in S2) and glu93 in S3a bridging S3a to the C-terminal end of S2. These residues are all highly conserved in voltage-dependent K<sup>+</sup> channels. The salt bridges found in the isolated S1–S4 voltage-sensor paddle structure (1ORS) are not found in the full-length channel structure (1ORQ). Other differences found between the two voltage sensor paddle structures are as follows: (1) S1 is folded back; and (2) S4 is straight in the isolated domain and bent in the full channel structure, the bend occurring just after the voltage-sensor paddle. In the full channel structure the S4 bends at a glycine residue to redirect S4 into S5 while S4 can continue straight in the isolated domain (S5 is not present). Significantly, the N-terminal half of S4 has residues that are

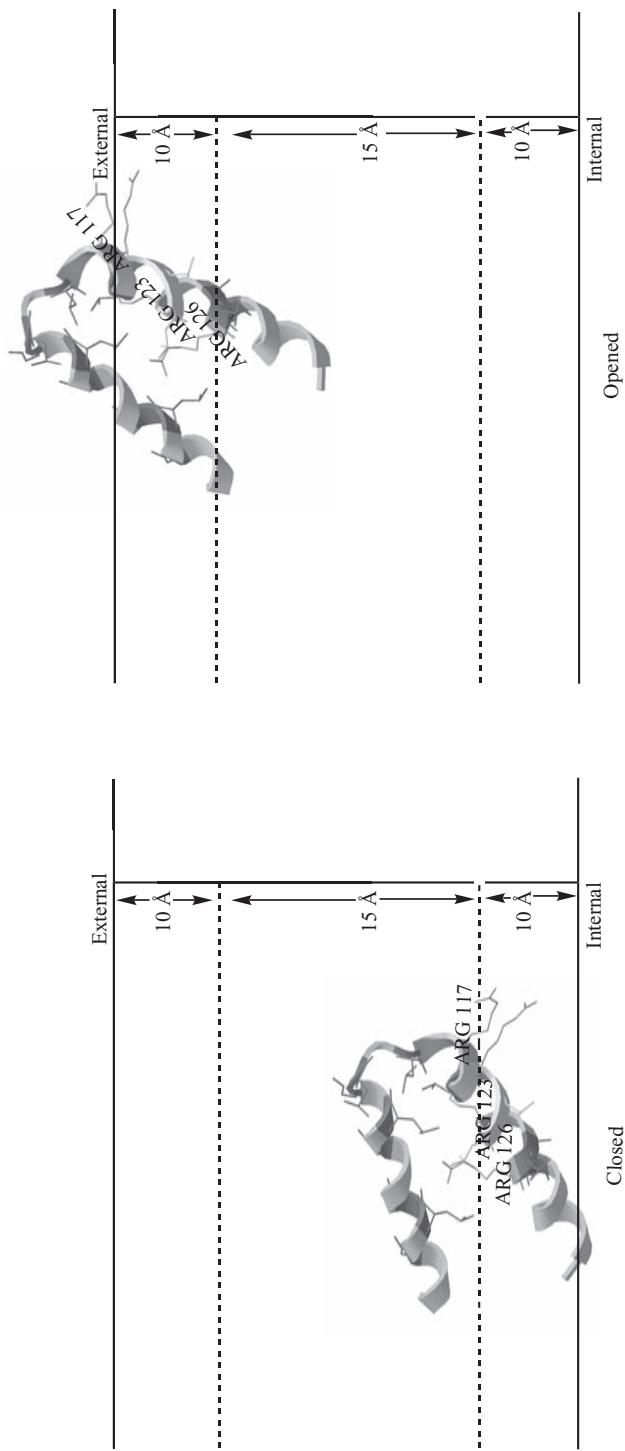


**Figure 5.7** View of the  $K^+$  ion channel, the pore helix, and helices S1–S6, PDB: 1ORQ. Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. (Printed with permission of CambridgeSoft Corporation.) (See color plate)

conserved in many K<sup>+</sup> channels as is shown in reference 18 for 15 different channels. The conclusion reached here is that the voltage-sensor paddle is a conserved unit in voltage-dependent K<sup>+</sup> channels and that the S4–S5 region can bend easily depending on the relative gating state of the channel. The S4–S5 region contains many hydrophilic residues and is comfortable at the interface with aqueous solution on the intracellular side of the membrane. Additionally, voltage-sensor paddles appear to be flexibly attached to S2 and S5, and S1 and S2 are loosely packed against the channel pore. Indeed the voltage sensor seems to float within the membrane, leading the authors to propose a new model for gating charge movements in which the arginine gating charge carriers can move the voltage-sensor paddle through the membrane from the intracellular side to the extracellular side. (See Figure 8 of reference 18, a news focus article in *Science* a month after reference 18 was published,<sup>24</sup> and Figure 5.8 for another view of the newly proposed movements of the voltage sensor paddle.)

Conclusions reached in reference 18 include: (1) Voltage-sensor paddles are helix–turn–helix structures made from helices S3b and the N-terminal half of S4; (2) voltage-sensor paddles' amino acid composition is mainly hydrophobic with the important exceptions of S4 arginine residues; (3) paddles are movable with respect to the pore because of their flexible attachments to the rest of the voltage sensor; (4) movements of the paddle in response to changes in membrane voltage could open the pore; and (5) observations and experimentation led to a proposal that the voltage-sensor paddles move across the membrane carrying their gating charges (arginine residues) through the electric field. This last proposal led to experiments described in the group's accompanying *Nature* publication.<sup>25</sup>

The MacKinnon group next exploited the Fab attachments to the full-length KvAP channel (Fab6E1) and to the isolated voltage sensor (Fab33H1) to examine the position of voltage-sensor paddles with the KvAP channel functioning in a lipid membrane. First, it should be said that the Fabs attach themselves to the same sector of the voltage-sensor paddle (the same epitope) between S3b and S4 in both structures. The experiments sought to assess whether the voltage-sensor paddles change their position when the channel gates open under conditions of depolarization (change from a holding voltage of -100 mV to +100 mV). The researchers could show that both Fabs inhibited channel function when applied to the solution external to the lipid membrane, but not when the Fabs were applied internal to the membrane. It was also shown that inhibition by the externally applied Fabs first required membrane depolarization. Experimentally the channel was held closed at -100 mV for 10 minutes, during which time little or no inhibition was observed. Additionally, no inhibition was observed in the absence of a Fab when the membrane was depolarized from -100 to 100 mV (typically for 200 ms). Then a Fab was added and depolarizations from -100 to 100 mV were continued. Inhibition began after some time lag—perhaps because the channel had to open before Fabs could bind and begin to cause inhibition. The researchers concluded that the

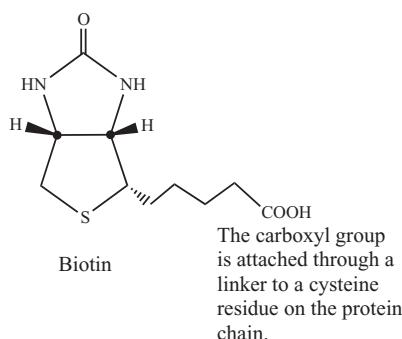


**Figure 5.8** Closed and opened views of the potassium ion channel pore according to reference 25. S4 helix of PDB: 1ORQ. Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. (Printed with permission of CambridgeSoft Corporation.) (See color plate)

voltage-sensor paddles remain inaccessible as long as the channel is held closed by a negative membrane voltage and that the Fab binding site on two helical turns of S3b and one turn of S4 (the epitope) becomes exposed to the membrane's external side on depolarization.

Researchers then used a sophisticated biochemical test to place individual amino acids into positions within the channel membrane under conditions of polarization and depolarization—that is, under conditions when the pore would be closed (positive charge on the membrane's exterior, negative on the interior) or open (negative charge on the membrane's exterior, positive on the interior). Finkelstein and co-workers had used this test earlier.<sup>26</sup> The procedure is as follows: (1) Using site-directed mutagenesis, place cysteine residues one at a time at specific amino acid locations of interest within the channel protein (G112C, I127C, and L103C, for instance); (2) react cysteine residues with biotin (biotinylation)—see Figure 5.9; (3) reconstitute the K<sup>+</sup> channel into planar lipid membranes; and (4) determine whether avidin, a rigid glycoprotein molecule incapable of entering the membrane but binding to biotin with strong affinity, binds from the external or internal side of the membrane under depolarization conditions. Taking into account the length of the biotin molecule and where it binds to avidin (that cannot enter the membrane), one can place the cysteine  $\alpha$ -carbons within 10 Å of the bulk aqueous solution on either side of the membrane. The system functions as a molecular ruler.

Wild-type K<sup>+</sup> channels are not affected or inhibited by external avidin, showing a small reduction in current when avidin is applied internally. The G112C mutant, positioned at the end of membrane segment S3b on the voltage-sensor paddle, was inhibited completely by external avidin but not internal avidin. Mutant I127C, located near the center of membrane segment S4 on the voltage-sensor paddle and next to important gating-charge mover arg126, was inhibited completely by internal avidin but not by external avidin. Mutant L102C, located near the N-terminal end of segment S3b, showed a different behavior. External avidin reduces the current but does not abolish it



**Figure 5.9** The biotin molecule.

and changes the kinetics of current activation. The authors believe that incomplete biotinylation of interior residues is responsible for this behavior with this and some other mutated residues.

Now the reference 25 researchers conducted the cysteine mutant–biotin–avidin test on selected residues of the voltage-sensor paddle. Residues gly101, leu103, ala104, gly108, leu110, gly112, leu113, leu115, and val119 (red stick form in Figure 5.8) include all of S3b, the tip of the paddle (S3–S4 loop) and the first helical turn into S4. They behave in the following manner: (1) At negative voltages—channel closed—none of these positions test positively—that is, they lie at least 10 Å from the membrane’s edge internally and externally; (2) at positive membrane voltage—channel open—these residues become accessible to external avidin—that is, they are within 10 Å of the membrane surface. Residues leu125 and leu127 (royal blue in Figure 5.8) in the S4 segment of the voltage-sensor paddle behave as follows: (1) At negative voltages—channel closed—these residues have a positive avidin test only internally—that is, they lie within 10 Å of the internal membrane surface; (2) at positive membrane voltage—channel open—these residues test negatively—that is, they are more than 10 Å from the internal and external membrane surface. Residues leu121 and leu122 (yellow in Figure 5.8), near the center of segment S4, react positively to internal avidin at negative voltages and positively to external avidin at positive voltages—that is, they are within 10 Å of the internal membrane at negative voltages and within 10 Å of the external membrane when depolarization takes place. Figures 5a,b of reference 25 and Figure 5.8 indicate the changes graphically, showing that under conditions of depolarization the tip (S3b–S4 connector loop) of the voltage-sensor paddle is actually pulled up out of the membrane into the external solution, whereas under negative voltage conditions the voltage-sensor paddle lies deep within the membrane and portions of S4 actually reach through the internal membrane surface into the cell cytoplasm. The authors estimate that the voltage-sensor paddle’s center of mass translates approximately 20 Å through the ~30- to 35-Å-thick membrane from inside to out, while the orientation with respect to the membrane changes from nearly horizontal—pore closed—to nearly vertical—pore open.

The authors’ conclusions from references 18 and 25 are the following: (1) Gating charges (principally on residues arg117, arg120, arg123, arg126) are carried on the voltage-sensor paddles; (2) voltage-sensor paddles are helix-turn–helix structures comprised of helix S3b, loop S3b–S4, and the N-terminal half of helix S4; (3) voltage-sensor paddles are attached to the K<sup>+</sup> channel through flexible S3 loops and S4–S5 loop linkers; (4) the paddles are located at the K<sup>+</sup> channel’s outer perimeter and can move through the lipid membrane; (5) S2 helices lie beside the K<sup>+</sup> channel pore and contain acidic amino acid residues, glu53, asp62, and asp72, that could help stabilize positive arginine residues as they traverse the membrane; (6) voltage-sensor paddles move approximately 20 Å across the membrane and perpendicular to it; and (7) the large displacement of the voltage-sensor paddles could open the pore by pulling on the S4–S5 linker.

The preceding conclusions were presented as a new theory about channel pore opening and closing. The tilted positions of the voltage sensor paddles and the position of S4 exposed to the lipid bilayer at the periphery of the K<sup>+</sup> channel ion conduction pore contradicted many other experimental results. In fact, the proposal as to the position of the S4 helix ignited controversy among scientists in the field that continues to the current time. Earlier theories had been put forward and backed by large bodies of experimental data from many different sources. For instance, Bezanilla and co-workers and Isacoff had used fluorescence resonance energy transfer (FRET) experiments to show much smaller movements of S4 across the membrane (3 Å rather than 15–20 Å) and voltage sensors in more upright positions than indicated by the MacKinnon's group work.<sup>27</sup>

One theory that differs from the MacKinnon proposals places the S4 helix in a more upright position, at the edge of the K<sup>+</sup> channel pore in a so-called canaliculi, a canal or aqueous gating pore, through which helix S4 (or the charges themselves) could move in response to changes in the electric field of the membrane. A protein wall consisting of helices S1, S2, and S3 would surround the positively charged arginine gating charges of S4 (and the rest of this helix), shielding it from the lipid membrane. In early 2004, Laine, Papazian, and Roux published a minireview in *FEBS Letters* that compared their K<sup>+</sup> channel voltage sensor domain structural theories with those of the MacKinnon group.<sup>28</sup> These researchers proposed a model for the Shaker K<sup>+</sup> channel in its activated state.<sup>29</sup> (see Figures 9 and 10 of reference 29 or Figure 2 of reference 28). The model shows the voltage sensor domain S1–S4 segments as transmembrane  $\alpha$ -helices. In the tetrameric complex of S1–S6 helices, the S4 segment is located near a groove formed in the interface between the ion conduction pores of adjacent subunits. The S4 helices of the tetramer appear at its four corners interior to the S1, S2, and S3 helices and almost halfway between neighboring S5 helices. The overall organization of the domains is similar to that seen in the KvAP crystal structure (PDB: 1ORQ, 1ORS); however, the position of the S4 helix, in particular, differs.

The reference 28 authors continue to detail experimental observations that place voltage sensor helices in positions within the membrane. Miller and coworkers conducted site-directed mutagenesis for all residues of helices S1–S3.<sup>30</sup> In these experiments, tryptophan (trp) residues were substituted for each amino acid in turn to determine which residues would be trp-tolerant. These experiments confirmed  $\alpha$ -helical conformations for S1 and S2 and showed that K<sup>+</sup> channel function was altered when trp residues were placed in some (labeled non-trp-tolerant), but not all, positions. The same treatment for helix S3 yielded complex results. At S3's N-terminal end the distribution of trp-tolerant positions were consistent with an  $\alpha$ -helical structure, however, this was not the case at S3's C-terminal end. Other tests indicated that S3 might be helical for its entire length and that the N-terminal end interfaces with both lipid and protein while the C-terminal end interfaces with water. Comparisons of trp-tolerant or trp-intolerant residues over several different Kv channel

family sequences indicated that while trp-tolerant sites varied among the species, trp-intolerant sites were highly conserved. These observations reinforced the suggestion that trp-tolerant sites in the Shaker K<sup>+</sup> channel are exposed to the lipid bilayer while trp-intolerant residues make the protein-protein contacts. In the Laine model being discussed here, the trp-tolerant residues all point toward the hydrophobic lipid bilayer arguing for S3's position outside of S4. In the Mackinnon KvAP model (1ORQ, 1ORS), the S3 helix is broken down into S3a and S3b sectors connected by a short loop while S1 and S2 are modeled as  $\alpha$ -helical segments. The secondary structure elements are in agreement for both models, it is their position within the K<sup>+</sup> channel that differs.

Charge reversal mutations have been shown to affect maturation of the protein and indicate positions of strong electrostatic interactions.<sup>31</sup> Particular Shaker residues involved are glu283 in S2 with arg368 in S4 in an intermediate closed pore conformation and glu283 in S2 with arg371 in S4 in the activated (open pore) conformation. Other interactions involve lys374 in S4 with glu293 in S2 and asp316 in S3. These interactions are postulated to constrain the S2, S3, and S4 helices into the positions that this theory advocates. Conserved basic and acidic residues were found to form the same sort of salt bridges in the isolated S1–S4 voltage-sensor paddle structure of KvAP (1ORS), but the same sort of interactions were not found in the full-length channel structure (1ORQ). Additionally, KvAP interactions do not occur in precisely the same register as those seen for the Shaker closed or open pore interactions. As stated above in the discussion of the KvAP X-ray crystallographic structures, in the 1ORS isolated voltage sensor paddle, arg133 in S4 (Shaker position arg365) interacts with asp62 in S2 (Shaker position glu283) to bring the S2 and S4 helices together. KvAP residue lys136 in S4 (comparable to Shaker lys374) is within a long hydrogen bonding distance (4.1 Å) of asp72 (comparable to Shaker glu293), again placing S2 and S4 in proximity. KvAP residue lys136 also is within hydrogen bonding distance of glu93 in S3a (comparable to Shaker asp316), bringing the S3a and S4 helices into close proximity as seen in the Shaker experiments. It does not appear that these interactions per se will argue for one S1–S4 conformation versus the other.

Various experiments have attempted to place the S4 helix in relation to the ion conduction pore's position. Using lanthanide-based resonance energy transfer (LRET), Bezanilla and co-workers measured a distance of 45 Å between val363 residues in S4 helices across the tetrameric structure from each other.<sup>32</sup> This places the N-terminal end of S4 near the groove between adjacent subunits. Other experiments using tethered quaternary ammonium pore blockers localized S1–S4 segments relative to the central pore.<sup>33</sup> These experiments placed the extracellular ends of S1 and S3 approximately 30 Å from the extracellular pore opening and the S3–S4 linker at approximately 17–18 Å from the pore opening. Formation of disulfide bonds between cysteines substituted for Shaker arg362 (in S4) and phe416 and ala419 (in S5) and other confirming experiments for the two residues' proximity place the S4 and S5 helices close

to each other. Actually, these experiments and others place the N-terminal end of one subunit's S4 helix in close proximity to the C-terminal end of S5 in the adjacent subunit of the tetrameric complex. The reference 28 authors believe that this foregoing evidence confirms their placement of S4 in its position between the pore and helices S1–S3.

Accessibility measurements have placed residues on voltage sensor helices S1–S4 in the following varying positions: (1) close to the intracellular side of the membrane or in the cytoplasm; (2) close to the extracellular side of the membrane or outside the membrane extracellularly; (3) being wholly within the membrane. This type of experiment, carried out by the MacKinnon group, has already been described in the discussion of the KvAP structure (1ORQ, 1ORS) previously. (See the discussion of reference 25 above.) Laine, Papazian, and Roux<sup>28</sup> describe further fluorescence resonance energy transfer (FRET) experiments in which cysteine residues are introduced into S1–S3 sites by site-directed mutagenesis.<sup>34</sup> The cysteines then are labeled with tetramethylrhodamine maleimide, a fluorescent molecule, to probe the position's accessibility extracellularly. Since the fluorescent reagent cannot pass through the membrane, voltage-dependent changes in fluorescence relative to wild-type K<sup>+</sup> channel indicates whether the residue is accessible to solvent. Residues that showed important fluorescent changes were observed at the C-terminal end of S1 and the N-terminal end of S2 (the S1–S2 linker) and the C-terminal side of S3 (S3b) (S3–S4 linker), indicating that these residues were exposed to the external solution. No significant fluorescent changes were seen for helical segments within the membrane (235–246 for S1 and 281–329 in S2–S3). Bezanilla and co-workers used histidine scanning mutagenesis of the S4 helix to position important S4 residues through their accessibility from the external or internal side of the membrane.<sup>35</sup> Their experiments showed that Shaker S4 residues arg365, arg368, and arg371 traverse the membrane from the internal side at hyperpolarized potentials (membrane negative inside) to the external side at depolarized potentials (membrane negative outside). These results are similar to those found by the MacKinnon team for KvAP using the biotinylation experiments described previously. In these experiments, with the channel in its activated state (pore open, depolarized), extracellular accessible positions ranged from gly101 (Shaker phe324) in helix S3b to leu122 (Shaker val367) approaching the center of the S4 helix but closer to its N-terminal end. Leu121 and leu122 were also accessible from the internal side with the channel in its resting state (pore closed) in the MacKinnon group experiments.

Conclusions from the reference 28 review article are as follows: (1) Several models backed by experiment place S4 near the groove between adjacent subunits, while the MacKinnon group model places S4 near the periphery of the protein; (2) lanthanide-based resonance energy transfer (LRET) places two S4 residues in segments across the tetramer from each other at a distance of 45 Å; (3) a method based on tethered quaternary ammonium pore blockers places the extracellular ends of the S1 and S3 helices further away from the ion conduction pore than the S3–S4 linker, arguing that the S4 helix resides

closer to the pore than S1 or S3; (4) intersubunit disulfide bonds—formed after residues in the S3–S4 linker, within N-terminal S4 and near the C-terminal end of the S5 helix, underwent site-directed mutagenesis to cysteine residues—indicate that S4 lies in close proximity to the pore region; and (5) patterns of accessibility in S1–S3 and S4 helices indicate that their ends all lie close to the extracellular surface under pore open conditions. Finally, the reference 28 authors argue that more information is needed about the resting state of voltage-dependent K<sup>+</sup> channels before definitive mechanisms for reaching the activated state from the resting state can be elucidated.

Molecular modeling of voltage sensor–pore interactions were undertaken by the reference 29 authors using many of the experimental constraints discussed in the preceding paragraphs. The results placed the S4 segment in the groove formed at the interface between neighboring S5 helices from different tetrameric domains. The researchers found that this was the only S4 position that simultaneously satisfied all the starting constraints. Another theoretical model for voltage gating has been developed in the laboratory of H. Robert Guy.<sup>36</sup> These researchers described the general consensus concerning the voltage-sensing mechanism for K<sup>+</sup> channels before the KvAP X-ray crystallographic structure (1ORQ, 1ORS) became available. First, it was believed that the voltage-sensor (helices S1–S4) spanned the membrane in all conformations (open, closed, or in transition). Second, accessibility studies suggested that positively charged residues of S4 would be found in water-filled crevices and that voltage dependency was due to these charges being moved a short distance across some barrier. The distance that S4 would move was in dispute.

The publication of the KvAP crystal structures of the full channel (1ORQ) and the isolated voltage sensor domain (1ORS) and the model arising from the publications (see discussion of references 18 and 25) did not agree with the Guy and co-workers model. Because of similarities in sequence and conservation of key amino acid residues, the Guy research group did not believe that the KvAP voltage sensor domain should have a very different structure from that of the Shaker protein. Therefore, they believed that either (a) the previous models were incorrect, allowing the crystal structure to yield a correct interpretation, or (b) the crystal structure, especially of the full channel, did not represent a native conformation of the protein. In the view of these researchers, one of the main problems with the new interpretation based on the KvAP crystal structure is the evidence that the voltage sensor paddle (helices S3b and S4) moves through the lipid during the transition from the resting state to the activated state. This movement appears to be inconsistent with previously developed models and could possibly be an artifact arising from a protein structure distorted from its native conformation. Possible distortions postulated for the full-length KvAP channel include: (1) In native KvAP channels, antibody fragments (Fab) bind to the S3–S4 loop only from the extracellular side, whereas the S3–S4 loop is in the cytoplasm (internal to the cell) in 1ORQ; (2) in Shaker channels, the C-terminal ends of helices S1

and S3, N-terminal ends of S2 and S4, and S1–S2 and S3–S4 loops are accessible from the extracellular side of the membrane in all conformations, whereas the same regions in the PDB: 1ORQ structure appear to be on or near the internal side of the membrane; (3) in Shaker channels, residues immediately preceding helix S1 are located in the cytoplasm, whereas in PDB: 1ORQ these are located in the transmembrane region; and (4) except for the S3–S4 hairpin turn, the tertiary structure of the voltage-sensing domain of PDB: 1ORQ deviates substantially from that of the isolated domain (PDB: 1ORS).

The model put forward by the reference 36 researchers is a more traditional one in which each of the S1–S6 segments traverses the entire transmembrane region and in which the S4 helix moves through the membrane via the “helical screw mechanism.” The S4 helical screw model is believed to be more energetically favorable than that put forth by the MacKinnon group because the positively charged S4 groups in the transmembrane region are always close to a negatively charge residue on S1, S2, or S3. To build the model they advocate, these authors began with the KvAP crystal structures PDB: 1ORQ (full-length channel) and PDB: 1ORS (voltage-sensor domain only), combined these using the pore-forming domain of PDB: 1ORQ and the voltage-sensor domain of PDB: 1ORS, and imposed constraints based on experimental results and physiochemical principles. In their model the S4 helix and S4–S5 linker, with their hydrophilic residues, are positioned to interact with other residues in the hydrophilic core of the voltage-sensor domain, with polar lipid headgroups (at or near the membrane surface), or with water at the membrane–water interface. Also predicted from past studies are hydrophilic cavities or crevasses that face toward the pore-forming domain and isolate these structures from the hydrophobic membrane surroundings.

Next, models were made for the voltage-sensor domain in open, transition, and resting (closed) configurations. Viewing the movies supplemental to the Guy publication (reference 36) will help greatly in visualizing these models. Go to the *Biophysical Journal* website at <http://www.biophysj.org>, enter the volume number (87) and first-page number (2255) for this article, choose the full text option, and then click on supplemental materials. There are two movies, and they are the last two items on the supplemental materials list. In the open (activated) conformation, most of the positively charged S4 arginine residues are on the extracellular side, as expected from many experimental results. (See discussion of references 34 and 35b.) In order to close the ion conduction pore, the positive charges on S4 must cross the membrane while remaining charged. This model asserts that negatively charged amino acids existing on the S1, S2, and S3 segments in the voltage sensor’s core domain facilitate this process. In KvAP, the residues are glu28 (*E1a* in the terminology of Guy, reference 36) and glu45 (*E1b*) in S1, asp62 (*D2a*) and asp72 (*D2b*) in S2, glu93 (*E3a*) in S3a, and glu107 (*E3b*) in S3b. In the Shaker protein, the comparable residues would be glu247 (KvAP glu45) in the S1–S2 linker, glu283 (KvAP asp62) and glu293 (KvAP asp72) in S2, and asp316 (KvAP glu93) in S3. Because of their positions in this model, glu45, asp62, and glu107

should be accessible from the membrane's exterior, while glu28, asp72 and glu93 should be accessible from inside the membrane for all conformations (see Figure 5.4). In these models, KvAP arginine residues that will interact with the acidic residues are numbered *R1* (arg117), *R2* (arg 120), *R3* (arg123), *R4* (arg126), *R6* (arg133), and *R7* (lys136). *R5* is not included because KvAP has no charged residue at this position. The comparable positions in the Shaker protein would be *R1* (arg362), *R2* (arg365), *R3* (arg368), *R4* (arg371), *R5* (lys374), *R6* (arg377), and *R7* (lys380). In the first model—pore open—*R1* and *R2* reside in the external aqueous phase, *R3* is exposed but near *E3b*, *R4* is near *E1b*, *R6* salt bridges to *D2a*, and *K7* is near *D2b*. Now the helical screw model comes into play as one inward helical screw step translates the S4 helix by ~4.5 Å further into the membrane and rotates it ~60° about its axis. (See the movie and Figure 4 of reference 36.) After two helical screw steps the following interactions occur: *R1-E3b*, *R2-E1b*, *R3-D2a*, *R4-E3a*, *R6-D2b*, and *K7-E1a*. The pattern results in the positively charged S4 residues remaining in the core of the voltage sensor domain where continued interaction with the electronegative core residues lowers the electrostatic barrier to the movement and allows the arginine residues to remain protonated and to carry the positive gating charge through the membrane. Two more helical screw steps move S4 toward the intracellular side of the membrane as the resting state (pore closed) configuration is achieved. The transition states are not meant to be energy-minimized conformations but rather possible steps along a continuum.

To carry their model further, Guy and co-workers now docked the voltage sensor domain with the ion conduction pore domain forming a covalent linkage between S4 and S5 through the S4-S5 linker (Loop 4-5). The pore forming domain from the X-ray crystallographic structure of the full-length KvAP channel (PDB: 1ORQ) was used for the open conformation. Modeling for the closed conformations of KvAP were taken from the KcsA structure (PDB: 1BL8). It was assumed that interactions between S4 and the pore-forming domain for KvAP would be similar to those for the Shaker protein, as experimental results exist for Shaker. The experimental results indicate that interacting residues between domains are more likely to be conserved among different proteins than those exposed to lipid. These are the same concepts discussed in determining tolerant versus nontolerant residues in the reference 28 model discussed above. Experiments indicate that the voltage sensor domain docks on the pore-forming domain so that water-filled crevasses or clefts around the central barrier (crossing point of S4 and S2 helices in the sensor domain) form between the two domains. The conclusion from many such experiments predicts that the voltage sensor domain X-ray structure (PDB: 1ORS) shows this domain in a native conformation. The open conformation PDB: 1ORS domain was docked with the open conformation pore-forming domain so that conserved residues interacted to the maximum amount and in a manner consistent with the experimental studies on Shaker channels. The docking arrangement caused residues near the S4 helix' N-terminal end to interact with residues in the C-terminal end of the S5 helix of an adjacent tetramer subunit. The authors

propose that S4 docks in the groove between the tetramer subunits with a tilt of  $\sim 60^\circ$  with the membrane plane and interacts with the S5 segment on its counterclockwise side (viewed from outside the membrane). This view differs from that of reference 28 and 29 researchers, who say that S4 docks in the groove between the tetramer subunits with a tilt of  $\sim -15^\circ$  and interacts with S5 of an adjacent subunit in a clockwise manner.

The Guy open conformation model docked structure was minimized in vacuo followed by a 1-ns molecular dynamics simulation of the complex embedded in a phosphatidylethanolamine (POPE) lipid bilayer. Adjustments were made to the model, and simulations were repeated so that very little movement occurred during the final iterations. Similar methods were used to dock the two domains in transitional and resting states. However, these results are more tenuous as little experimental data is available. In particular, the position of the S4–S5 linker and its role in opening and closing the pore are uncertain. The supplemental movie accompanying reference 36 illustrates the open-to-close-to-open cycle resulting from the simulations.

Conclusions reached by the Guy research group include the following: (1) A conventional helical screw model can be developed from the two KvAP crystallographic structures (PDB: 1ORQ, 1ORS) in a manner consistent with experimental results and theoretical constraints; (2) the model uses the open conformation KvAP pore-forming domain from PDB: 1ORQ and uses the voltage sensor domain from PDB: 1ORS, because these appear to be in their native conformational folds; and (3) energetic, evolutionary, and experimental criteria were combined to create the models they describe. Energetically, the root-mean-square deviation (RMSD) of the model structures are low during simulations, polar atoms (especially the charged atoms of S4 arginines) form hydrogen bonds or salt bridges with polar atoms of water, protein, or lipid headgroups, few hydrophobic residues are exposed to water, and most residues have energetically favorable conformations. The energetic considerations are in contrast with the MacKinnon group voltage sensor paddle model in which the helix S4 charged residues are postulated to move through the hydrophobic lipid phase of the membrane during activation.

New structures published by the MacKinnon team in 2005 eliminated the need for the antibody attachment (Fabs) to achieve protein crystallization and show the voltage sensors in the expected more upright position. This research is described in the following paragraphs.

The Mackinnon group published the crystal structure of a mammalian voltage-dependent *Shaker* family K<sup>+</sup> channel protein (Kv1.2) in *Science* magazine in 2005 (PDB: 2A79).<sup>37</sup> In an accompanying article, the team discussed the voltage sensor of Kv1.2 specifically attempting to link structural details to electromechanical coupling in K<sup>+</sup> channels.<sup>38</sup> First, this study of a eukaryotic Kv channel from rat brain, PDB: 2A79, may be compared to earlier studies of prokaryotic cousins—KvAP, in PDB: 1ORQ, 1ORS; and KcsA, in PDB: 1BL8, 1K4C, 1K4D—because there are many similarities in their structures. The selectivity filter domain is so conserved as to be essentially the same for all K<sup>+</sup>

channels. The K<sup>+</sup> channel's pore domain, with the “inverted tepee” arrangement of inner helices holding the selectivity filter at the extracellular membrane surface, is another mostly conserved feature. However, there are important differences between prokaryotic and eukaryotic K<sup>+</sup> channels. For instance, in eukaryotic *Shaker* family Kv channels, the S6 helix on the intracellular side of the selectivity filter has a highly conserved triplet sequence—Pro-X-Pro—where X is any amino acid. This sequence is not present in prokaryotic channels. This sequence has been shown, by mutation, to be important for K<sup>+</sup> channel gating but the reasons are unknown at this time. More importantly, as one looks down from the extracellular side of the membrane, the so-called T1 domain, inside the cell in *Shaker* Kv channels, is located directly under the pore domain entryway. Four T1 domains form a constrictive tetrameric assembly at the intracellular membrane surface. This means that in *Shaker* Kv channels the transmembrane pore must communicate with the cytoplasm (the cell interior) through side portals as described previously for PDB: 1EXB in reference 16. The side portals allow passage of K<sup>+</sup> ions as well as inactivation peptides that are part of the K<sup>+</sup> channel's N-terminal end. These peptides function as inactivation gates in some *Shaker* Kv channels. The T1 domains also forms a platform for four intracellular β subunits arranged in a tetrameric assembly whose function is unknown at this time. The β subunit was described previously in the discussion of reference 16 and PDB: 1EXB above. A possible function for this oxido-reductase, NADPH cofactor-containing subunit may be to serve as a sensor for the Kv channel and as a mediator of the redox state of the cell.

Experimentalists agree that voltage-dependent gating is fundamentally similar for prokaryotic and eukaryotic Kv channel systems. The all-important voltage sensor domain should be structurally similar as well. As documented above for the prokaryote KvAP channels, the flexible, mobile voltage sensor domain could not be crystallized without the antibody anchoring system, and this may have affected the orientation of the voltage sensor domain in the KvAP system (PDB: 1ORQ and 1ORS). In crystallizing the Kv1.2 K<sup>+</sup> channel protein reported in the *Science* 2005 articles (references 37 and 38), MacKinnon and co-workers used a mixture of detergent and lipids in the crystallizing liquid and maintained a strongly reducing environment, eliminating the need for anchoring antibodies. In the crystals that resulted, membrane-spanning pore and voltage sensor domains alternated with T1 and β subunit domains in an arrangement that closely mimics a native membrane structure. The T1 and β subunit domains appeared as rigid and well-packed structures, whereas the pore appeared somewhat more flexible and the voltage sensors appeared as mobile domains making few contacts with neighboring protein domains. Some parts of the crystalline structure were built as models with reference to previously determined structures (KvAP for instance) because electron density in these regions was weak.

The complete Kv1.2 channel–T1–β subunit complex is a tetramer of approximate dimensions 135 Å by 95 Å by 95 Å with pore and voltage sensors of

~30 Å in length across the membrane. The T1 domain and β subunit are each ~40 Å in length along the fourfold axis, and the T1 domain is offset from the intracellular pore opening by ~15–20 Å. The α-helical linker T1–S1 functions as a spacer between the transmembrane pore and the intracellular regions. The ion conduction pore is made up of the α-helices S5, pore helix, and S6, these being related to the voltage sensor α-helices of the adjacent subunit by wrapping around them. (See Figure 2C on p 899 of reference 37.) In a given subunit, the S4–S5 linker that runs parallel to the intracellular membrane surface connects the conduction pore and the voltage sensor. The S4–S5 linker lies at the level of the inner-helix bundle crossing, defined as the right-handed bundle of four inner helices, one from each of the tetramer subunits. In Kv1.2, the S6 helices line the ion conduction pore just on the intracellular side of the selectivity filter. The inner-helix bundle crossing is important because it forms an expandable constriction (the crossing) for opening and closing the pore. This construction is known as the activation gate.

The ion conduction pore structures of four different prokaryotic K<sup>+</sup> channels, including KcsA, MthK, KirBac, and KvAP, have been determined by X-ray crystallography as discussed above and described in Table 5.3. These four structures indicate a high level of conserved amino acid sequence and structure for the K<sup>+</sup> selectivity filter, while the inner pore (between the selectivity filter and intracellular solution) varies in sequence and conformation. Inner pore conformational changes open and close the pore. Some of the mentioned structures, those for KcsA and KirBac, have the pore in the closed position while MthK and KvAP pores appear to be open. These determinations are made by considering the pore's dimension at its narrowest point—the inner-helix bundle or bundle crossing point. Comparison to the current structure under discussion, Kv1.2, PDB: 2A79, indicates that this pore is open and has a diameter of approximately 12 Å. Other experiments have indicated that when the Shaker K<sup>+</sup> channel is closed, not even a small Ag<sup>+</sup> ion will pass the bundle crossing. When it is open, however, large organic quaternary ammonium cations, and even hydrophobic polypeptides, can enter and plug Kv channels. The inner helices (S6) of Kv1.2 contain the pro-val-pro sequence (pro405, val406, and pro407 in PDB: 2A79), highly conserved among *Shaker* family Kv channels, and is responsible for curving or bending the S6 inner helices so that they are almost parallel to the membrane near the intracellular membrane surface. In KvAP (PDB: 1ORQ, 1ORS) a glycine residue is responsible for a similar bend. The α-helical linkers connecting the T1 domain through S1 radiate out to create a wide space between T1 and the pore. This serves two purposes: (1) It allows space for the inner helices to undergo large gating movements to open the pore and (2) provides diffusion pathways for ions and inactivation peptides between the cytoplasm and the pore entryway. These so-called side portals have been described here previously in discussing the T1<sub>4</sub>β<sub>4</sub> complex of Kv channel of PDB: 1EXB as described in reference 16. The four side portals above the T1 domain each have a diameter of 15–20 Å and are lined with negatively charged amino acids (glu128, asp129, and glu130 for

Kv1.2) for their job of attracting cations. Inactivation peptides (also called an inactivation gate) are believed to exist either on the N-terminus of an intracellular domain or on the N-terminal end of an associated  $\beta$  subunit. In either case, inactivation peptides have (a) a hydrophobic region that can reach into the inner pore to plug it and (b) a positively charged hydrophilic region that is attracted to hydrophilic and negatively charged regions on the T1 side portal. Figure 4B of reference 37 shows a model inactivation peptide reaching into a Kv1.2 side portal. Questions remain as to the role of the  $\beta$  subunit domains of *Shaker* Kv channels. As discussed above, they may confer inactivation. It has also been observed that  $\beta$  subunits can influence levels of channel expression, leading to the hypothesis that they may be channel chaperones. Conserved catalytic residues for hydride transfer leads to another hypothesis, namely, that  $\beta$  subunits have a catalytic function. Perhaps the Kv channel regulates the activity of the  $\beta$  subunits; conversely, and perhaps more likely, the  $\beta$  subunits regulate the activity of the Kv channel. In the Kv1.2 structure reported here (PDB: 2A79), the NADP<sup>+</sup> cofactor is present in the  $\beta$  subunit active cleft along with diffuse electron density that possibly represents either a large organic molecule or polypeptide chain. More definitive experiments or structural detail are necessary before the true purpose of these  $\beta$  subunit features will be known.

In summation for the work of reference 37: (1) Important factors for obtaining crystals and determining structure for Kv channel proteins are the presence of lipids and detergent during crystallization as well as the maintenance of a nonoxidizing atmosphere; (2) relationships among integral membrane components, the intracellular T1 domain, and the  $\beta$  subunit bound to T1 are consistent with other electrophysiological studies of inactivation processes in *Shaker* K<sup>+</sup> channel proteins; (3) the ion conduction pore structure of the mammalian Kv1.2 studied here is similar to that of prokaryotic K<sup>+</sup> channels; (4) curved or bent inner, S6, helices result from the amino acid sequence Pro-X-Pro in *Shaker* K<sup>+</sup> channel proteins and from a glycine residue in KvAP and probably other Kv channels; (5) the Kv1.2 ion conduction pore is in an open conformation in the PDB: 2A79 structure; (6) 15- to 20-Å-wide side portals between the T1 domain and the ion conduction pore connect the pore to the cytoplasm allowing passage of K<sup>+</sup>, other larger cations, and inactivation peptides attracted by negatively charged glutamate and aspartate residues on the portal's surface; (7) additional experiments are necessary to test the influence of the  $\beta$  subunit's NADP<sup>+</sup> cofactor, active site cleft, and probable resident substrate molecule of unknown character on K<sup>+</sup> channel protein activity; and (8) the voltage sensor domain of Kv1.2 resides wholly in the membrane layer and appears unperturbed by the crystallization process.

In an accompanying article in *Science* (reference 38) Mackinnon and co-workers continue their study of the Kv1.2 voltage sensor (PDB: 2A79), trying to answer the following questions: (1) How do gating charges move within the membrane electric field? and (2) How do the gating charge movements couple to opening and closing of the ion conduction pore? Any model must account

for the fact that when a *Shaker K<sup>+</sup>* channel opens, the equivalent of 12–14 positive charges cross the membrane electric field from inside to outside and most of this charge is carried by four S4 arginine residues (arg294, arg297, arg300, and arg303 in Kv1.2) on each channel tetramer subunit.

Existent voltage sensor models postulate that the positively charged S4 helix (that must be protected from the membrane's lipid environment) rests in a groove at the interface between adjacent subunits of the K<sup>+</sup> channel tetramer so that pore  $\alpha$ -helices S5 and S6 form a wall on one side of S4 and voltage sensor  $\alpha$ -helices S1 and S3 form a wall on the other side. The groove or canal forms a gating channel and allows S4 to move its charged arginine residues across the membrane without exposing them to the lipid environment. How far the S4 helix moves has been a matter for much debate with hypotheses of large translations (~15 Å) and experimentally determined (by fluorescence resonance energy transfer, FRET) translations of less than 3 Å. All models have two features in common: (1) The S4 helix must be sequestered from the lipid membrane, and (2) voltage sensor helices S1–S4 are packed tightly against the S5 and S6 helices surrounding the pore. The MacKinnon research team believes that the voltage sensor may be an almost independent domain, and site evidence that indicates this. One piece of evidence comes from the study of KvAP and the PDB: 1ORQ and 1ORS structures discussed above in references 18 and 25. The PDB: 1ORS structure is that of the isolated voltage sensor domain itself—expressed without the pore—arguing that this may be an independent domain. The full-length channel structure, PDB: 1ORQ, showed the voltage sensor domain in a strange position, pulled toward the cytoplasmic side of the pore. Other evidence, cited above in the discussion of reference 18 and 25, indicated large movement of the helix-turn-helix voltage sensor paddle through the membrane and the possibility that some arginine residues were exposed to the hydrophobic membrane. However, there were major weaknesses in the KvAP study directly related to the voltage sensor's mobility: (1) Distortions arising from extracting the protein from its membrane left many aspects of the structure uncertain; and (2) the connections between the voltage sensors and the ion conduction pore were disrupted, leading to a possible incorrect placement of the voltage sensor within the membrane.

The voltage sensor domain structural problems are further discussed in reference 38 by using the knowledge gained from the reference 18 KvAP structural study (PDB: 1ORQ, 1ORS) in concert with another KvAP structure (PDB: 2A0L) published more recently.<sup>39</sup> In the current structural study of Kv1.2 (PDB: 2A79), the T1 domain and its connection to the S1 helix (part of the voltage sensor assembly) help to hold the voltage sensor in its native conformation. Similarities and differences between Kv1.2 (PDB: 2A79) and the KvAP (PDB: 1ORQ, 1ORS) structures are as follows: (1) Helices S3 and S4, called the voltage sensor paddle, are in an antiparallel arrangement for both Kv1.2 and KvAP; (2) in KvAP, helix S3 appears to have two subunits, S3a and S3b, whereas in Kv1.2 S3 appears to be a single helix with a bend; (3) voltage

sensor paddles in Kv1.2 are in a slightly different position with respect to helices S1 and S2, understandable if the voltage sensor paddle is mobile; (4) S4 and S3 are nearly two helical turns longer in Kv1.2 than in KvAP, meaning that the Kv1.2 paddle will project further into the extracellular solution and possibly react differently with molecules that interact with the voltage sensor from outside the cell (see experiments on biotinylation described in discussion of reference 25 above).

The S4–S5 linker helix is an amphipathic (containing both polar and non-polar amino acid residues)  $\alpha$ -helix that runs parallel to the membrane plane inside the cell. The helix's hydrophobic surface faces the membrane, and the polar side faces the cytoplasm. The S4–S5 linker helix rests against the curved part of helix S6 (after the Pro–X–Pro sequence) in a parallel conformation, and this interaction appears to be essential for coupling of voltage sensor movements to pore open or closed positions (see Figure 3C and 3D of reference 38). Mutations to the Pro–X–Pro sequence have large effects on  $K^+$  channel gating that appear to uncouple the voltage sensor from the pore. In experiments conducted by Lu and co-workers, transfer of the entire S1–S4 voltage sensor, the S4–S5 linker, and the C-terminal end of S6 were all required to engineer voltage dependence into KcsA, an otherwise voltage-independent (or weakly voltage-dependent)  $K^+$  channel.<sup>40</sup> The required S6 segment exactly corresponded to that contacting the S4–S5 linker in Kv1.2, indicating that these domains are necessary and sufficient to construct a functioning voltage sensor. Further description of the voltage sensor arrangements in the tetrameric  $K^+$  channel structure benefits greatly from access to Figure 4 of reference 38. As this figure illustrates for the overall structure, and in detail for the voltage sensor domain, the voltage sensors appear to be floating as separate domains at the corners of a square surrounding the ion conduction pore. In addition, the S4–S5 linker runs across to a neighboring subunit with the S4 helix adjacent to the S5 helix of the neighboring subunit. Experimental measurements have placed S4 arginine residues from one subunit at distances (between  $C_\alpha$  carbons) of 45 Å to the adjacent subunit and 64 Å diagonally across the pore, reinforcing the idea that S4 helices occupy the outer corners of the ion conduction pore helical structure. Other than the contacts discussed here, contacts between the S1–S4 voltage sensor helices and S5–S6 pore helices are found to be insubstantial, reinforcing the idea of mobile voltage sensor domains that undergo voltage-dependent conformational changes within the membrane.

The gating-charge arginine residues on the voltage sensor subunits (arg294, 297, 300, 303 in Kv1.2) reside near the N-terminal end of the S4 helix. These arginine residues are the most conserved among all Kv channel proteins and account for most of the gating charge. In Kv1.2, arg294 and arg297 are located on the voltage sensor's lipid facing surface, and arg294 may actually extend into the phospholipid headgroup layer. Meanwhile, arg300 and arg303 face helices S1 and S2 and make salt bridges with acidic amino acid residues on those helices. Similar salt bridges were found for the KvAP voltage sensor

discussed previously in reference 25 for the PDB: 1ORS structure. EPR data gathered by Perozo and co-workers for KvAP show a lipid environment for arg123, a lipid–water mixed environment for arg126 and a protein (neither lipid nor aqueous) for arg133 and arg136 [arginine numbering system is that found for KvAP (PDB: 1ORQ and 1ORS)].<sup>41</sup> These data agree with the X-ray structural data, indicating that the first two S4 arginine residues are exposed to lipid when the pore is in the open conformation—as is the case for Kv1.2.

Indications that the Kv channel is open in the Kv1.2 crystal structure (PDB: 2A79) are twofold: (1) The voltage sensors appear to be in an open position—that is, the voltage sensors have moved the gating-charges of two arginine residues nearer to the extracellular side of the membrane and all four arginine residues are above the membrane midpoint; (2) the inner helix bundle (activation gate) of the ion conduction pore is opened to a diameter of ~12 Å. The researchers (see references 37 and 38) now speculate as to how the pore might close. In the closed conformation the inner helix bundle might look like the closed pore of KcsA (PDB: 1BLB, 1K4C, and 1K4D, references 15 and 17). Downward movement of S4 will bring the arginine residues closer to the intracellular side of the membrane and push down on the S4–S5 linker helices. This movement in turn compresses the inner helices (S5–S6) and closes the pore. Qualitatively, the positively charged arginine residues are pushed out when the membrane is positive inside (opening the pore) and pulled in when the membrane is negative inside (closing the pore). This explanation is consistent with the biotinylation experiments described above in discussing KvAP structures PDB: 1ORQ and 1ORS (reference 25), and it indicates that a segment of the S4 helix can move up to 15 Å through the membrane on pore opening or closing. It is also consistent with observations of antiparallel alignment of S3b and S4 in KvAP and the same alignment between S3 and S4 in Kv1.2. These observations lead to the hypothesis that S3 and S4 move together as a voltage sensor paddle unit with S4 closer to the membrane’s intracellular side and S3 on top of S4 allowing S3 to remain closer to the membrane’s extracellular side.

Conclusions stated by the authors for the reference 38 studies include the following: (1) Voltage sensors are independent self-contained domains, except for the S4–S5 linkage to the inner pore helices, allowing them to perform mechanical work opening and closing the pore; (2) voltage sensor domains (helices S1–S4) are membrane-spanning with arginine residues on the S4 helix exposed to lipid on one face and salt-bridging with acidic residues from S1 and S2 on its other face; (3) the voltage sensor domain’s position with respect to the pore may vary among different Kv channels; and (4) motions of the Kv1.2 helix are transmitted to the inner helix bundle (activation gate) via the S4–S5 linker helices affecting the movements of the S5 and S6 helices to open or close the pore. The most important conclusion involves the positions of and interactions between the voltage sensor S1–S4 helices as they affect opening and closing of the pore. In the Kv1.2 structure, with the pore in the open conformation, S4 helix residues arg294 and arg297 appear to be on the lipid-

exposed surface of the voltage sensor, while arg300 and arg303 interact with acidic residues on helices S1 and S2 inside the voltage sensor. The authors believe that the energetic balance between the just-described electrostatic and hydrophobic forces is very important for voltage sensor function.

This section on describing potassium channels has relied heavily on structural analyses and ensuing models based on X-ray crystallographic studies. It must be remembered that all X-ray structures are static snapshots capturing a protein, or any other molecule, in one conformation. As the discussions of potassium ion voltage-dependent channels above has indicated, structural snapshots can lead to lengthy discussions of native versus non-native protein conformations and the relevance of any given protein structure to the function of said protein. Consequently, the reader should be on the lookout for new information sure to be published in the primary literature concerning potassium ion channels in the future.

## 5.5 CONCLUSIONS

Chapter 5 has introduced topics in metal and nonmetal ion homeostasis, concentrating on Groups I and II metal ions. Some biomolecules containing potassium (and sodium) ions have been described in detail, especially the sodium/potassium ion pump and voltage-gated potassium channels. Potassium ion channels are therapeutic targets in autoimmune and some neurodegenerative diseases and thus are of interest to bioinorganic chemists. To design drug molecules that would successfully target these biomolecules, researchers must understand their three-dimensional structure; thus the emphasis here has been on the X-ray crystallographic studies that have strived toward that goal.

## REFERENCES

1. Cowan, J. A. *Inorganic Biochemistry, An Introduction*, 2nd ed., Wiley-VCH, New York, 1997.
2. Frausto Da Silva, J. R. R.; Williams, R. J. P. *The Biological Chemistry of the Elements: The Inorganic Chemistry of Life*, Clarendon Press; New York, 1991.
3. Ghosh, A.; Ghosh, A. K.; Sharma, A.; Talukder, G. *Biol. Trace Elem. Res.*, 1991, **31**(2), 139–145.
4. Hart, B. A.; Harmsen, A. G.; Low, R. B.; Emerson, R. *Toxicol. Appl. Pharmacol.*, 1984, **75**, 454–465.
5. Togna, G.; Togna, A. R.; Russo, P.; Caprino, L. *Toxicol. Appl. Pharmacol.*, 1997, **144**, 262–267.
6. Nielsen, S. P. *Bone*, 2004, **35**, 583–588.
7. Tu, S.-M.; Delpassand, E. S.; Jones, D.; Amato, R. J.; Ellerhorst, J.; Logothetis, C. J. *Urol. Oncol.: Semin. Orig. Invest.*, 1996, **2**, 191–197.

8. (a) Takeda, E.; Taketani, Y.; Sawada, N.; Sato, T.; Yamamoto, H. *Biofactors*, 2004, **21**(1–4), 345–355. (b) Tenenhouse, H. S. *Annu. Rev. Nutrition*, 2005, **25**, 197–214.
9. Theil, E. C. *Adv. Inorg. Biochem.*, 1983, **5**, 1–38.
10. (a) Miller, C. *Science*, 2006, **312**, 534–535. (b) Sasaki, M.; Takagi, M.; Okamura, Y. *Science*, 2006, **312**, 589–592.
11. Dallwig, R.; Vitten, H.; Deitmer, J. W. *Cell Calcium*, 2000, **28**, 247–259.
12. Kubala, M.; Teisinger, J.; Ettrich, R.; Hofbauerova, K.; Kopecky, V., Jr.; Baumruk, V.; Krumscheid, R.; Plasek, J.; Schoner, W.; Amler, E. *Biochemistry*, 2003, **42**, 6446–6452.
13. Hakansson, K. O. *J. Mol. Biol.*, 2003, **332**, 1175–1182. (PDB: 1Q3I)
14. Hilge, M.; Siegal, G.; Vuister, G. W.; Guentert, P.; Gloor, S. M.; Abrahams, J. P. *Nat. Struct. Biol.*, 2003, **10**, 468–474. (PDB: 1MO7, 1MO8)
15. Doyle, D. A.; Morais-Cabral, J.; Pfuetzner, R. A.; Kuo, A.; Gulbis, J. M.; Cohen, S. L.; Chait, B. T.; MacKinnon, R. *Science*, 1998, **280**(5360), 69–77. (PDB: 1BL8)
16. Gulbis, J. M.; Zhou, M.; Mann, S.; MacKinnon, R. *Science*, 2000, **289**(5476), 123–127. (PDB: 1EXB)
17. Zhou, Y.; Morais-Cabral, J. H.; Kaufman, A.; MacKinnon, R. *Nature*, 2001, **414**, 43–48. (PDB: 1K4C, 1K4D)
18. Jiang, Y.; Lee, A.; Chen, J.; Ruta, V.; Cadene, M.; Chait, B. T.; MacKinnon, R. *Nature*, 2003, **423**, 33–41. (PDB: 1ORS, 1ORQ)
19. (a) Long, S. B.; Campbell, E. B.; MacKinnon, R. *Science*, 2005, **309**, 897–902. (b) Long, S. B.; Campbell, E. B.; MacKinnon, R. *Science*, 2005, **309**, 903–908. (PDB: 2A79)
20. Lee, S. Y.; Lee, A.; Chen, J.; MacKinnon, R. *Proc. Natl. Acad. Sci. USA*, 2005, **102**(43), 15441–15446. (PDB: 2A0L)
21. Valiyaveetil, F. I.; Leonetti, M.; Muir, T. W.; MacKinnon, R. *Science*, 2006, **312**, 1004–1007. (PDB: 2IH1 and 2IH3)
22. (a) Papazian, D. M.; Schwarz, T. L.; Tempel, B. L.; Jan, Y. N.; Jan, L. Y. *Science*, 1987, **237**(4816), 749–753. (b) Tempel, B. L.; Papazian, D. M.; Schwarz, T. L.; Jan, Y. N.; Jan, L. Y. *Science*, 1987, **237**(4816), 770–775.
23. Jiang, Y.; Lee, A.; Chen, J.; Cadene, M.; Chait, B. T.; MacKinnon, R. *Nature*, 2002, **417**, 515–522. (PDB: 1LNQ)
24. Miller, G. *Science*, 2003, **300**, 2020–2022.
25. Jiang, Y.; Ruta, V.; Chen, J.; Lee, A.; MacKinnon, R. *Nature*, 2003, **423**, 42–48.
26. (a) Qiu, X. Q.; Jakes, K. S.; Kienker, P. K.; Finkelstein, A.; Slatin, S. L. *J. Gen. Physiol.*, 1996, **107**, 313–328. (b) Qiu, X. Q.; Jakes, K. S.; Finkelstein, A.; Slatin, S. L. *J. Biol. Chem.*, 1994, **269**, 7483–7488.
27. (a) Cha, A.; Snyder, G. E.; Selvin, P. R.; Benzanilla, F. *Nature*, 1999, **402**(6763), 809–813. (b) Glauner, K. S.; Mannuzzu, L. M. Gandhi, C. S.; Isacoff, E. Y. *Nature*, 1999, **402**(6763), 813–817.
28. Laine, M.; Papazian, D. M.; Roux, B. *FEBS Lett.*, 2004, **564**, 257–263.
29. Laine, M.; Lin, M. C.; Bannister, J. P.; Silverman, W. R.; Mock, A. F.; Roux, B.; Papazian, D. M. *Neuron*, 2003, **39**, 467–481.
30. (a) Hong, K. H.; Miller, C. *J. Gen. Physiol.*, 2000, **115**, 51–58. (b) Monks, S. A.; Needleman, D. J.; Miller, C. *J. Gen. Physiol.*, 1999, **113**, 415–423.

31. Tiwari-Woodruff, S. K.; Schulteis, C. T.; Mock, A. F.; Papazian, D. M. *Biophys J.*, 1997, **72**, 1489–1500.
32. Cha, A.; Snyder, G. E.; Selvin, P. R.; Bezanilla, F. *Nature*, 1999, **403**, 809–813.
33. Blaustein, R. O.; Cole, P. A.; Williams, C.; Miller, C. *Nat. Struct. Biol.*, 2000, **7**, 309–311.
34. Ghandi, C. S.; Clark, E.; Loots, E.; Pralle, A.; Isacoff, E. Y. *Neuron*, 2003, **40**, 515–525.
35. (a) Starace, D. M.; Stefani, E.; Bezanilla, F. *Neuron*, 1997, **19**, 1319–1327. (b) Starace, D. M.; Bezanilla, F. *J. Gen. Physiol.*, 2001, **117**, 469–490.
36. Shrivastava, I. H.; Durell, S. R.; Guy, H. R. *Biophys. J.*, 2004, **87**, 2255–2270.
37. Long, S. B.; Campbell, E. B.; Mackinnon, R. *Science*, 2005, **309**, 897–903. (PDB: 2A79)
38. Long, S. B.; Campbell, E. B.; Mackinnon, R. *Science*, 2005, **309**, 903–908. (PDB: 2A79)
39. Lee, S. Y.; Lee, A.; Chen, J.; Mackinnon, R. *Proc. Natl. Acad. Sci. USA*, 2005, **102**(43), 15441–15446. (PDB: 2A0L)
40. (a) Lu, Z.; Klem, A. M.; Ramu, Y. *Nature*, 2001, **413**, 809–813. (b) Lu, Z.; Klem, A. M.; Ramu, Y. *J. Gen. Physiol.*, 2002, **120**, 663–676.
41. Cuello, L. G.; Cortes, D. M.; Perozo, E. *Science*, 2004, **306**(5695), 491–495.

---

# 6

---

## GROUP I AND II METALS IN BIOLOGICAL SYSTEMS: GROUP II

### 6.1 INTRODUCTION

Chapter 5 discussed homeostasis for Groups I and II metal ions and described a number of biomolecules containing Group I metal ions, concentrating on  $\text{Na}^+/\text{K}^+$  ATPases and potassium ion channels. Chapter 6 will discuss Group II metal ions in biomolecules, concentrating on magnesium ions in catalytic RNA and on two calcium-containing biomolecules—calmodulin and  $\text{Ca}^{2+}$ -ATPase.

### 6.2 MAGNESIUM AND CATALYTIC RNA

#### 6.2.1 Introduction

Prior to 1982, it was believed that biological catalytic species were solely constituted of amino acid chains—that is, proteins or, more correctly, enzymes. However, in that year T. R. Cech and colleagues reported the first catalytic RNA, calling it a ribozyme.<sup>1</sup> It is now known that ribozyme sequences are small parts of larger stretches of messenger or ribosomal RNA. Many reactions carried out by ribozymes are cleavage or ligation reactions and are part of post-transcriptional modification processes that lead to the mature RNA, DNA, or protein sequences—that is, maturation. The first characterized ribozyme was the self-splicing intron of the *Tetrahymena* pre-rRNA.<sup>2</sup> Researchers discovered that the *Tetrahymena* pre-rRNA spliced out its intron and

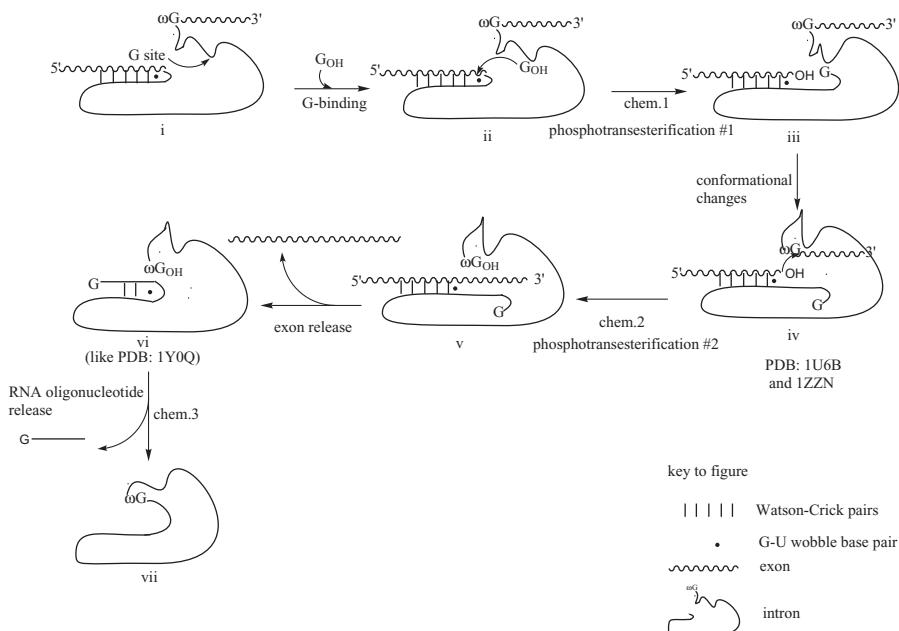
connected the flanking exon sequences. In Section 2.3.5, introns were defined as sections of DNA within a gene that do not encode part of the protein that the gene produces. Usually, introns are spliced out of the mRNA that is transcribed from the gene before it is translated. Exons were described as regions of DNA within a gene that are not spliced out from the transcribed RNA and are retained in the final mRNA molecule. Most, but not all, exon regions are translated to protein. (See Section 2.3.5.) With experiments carried out in vitro duplicating the activity seen in vivo in *Tetrahymena* pre-rRNA cells, the ribozyme chemistry taking place was described as follows: (1) Highly purified pre-rRNA (precursor ribosomal RNA) was mixed with guanosine or guanosine triphosphate, GTP, in a solution containing magnesium ions; (2) precise cutting and joining of phosphodiester bonds took place; and (3) during the process the added guanosine became covalently linked to the 5' end of the excised intron.<sup>3</sup> After a year's work attempting to inactivate the splicing reaction with protein-destroying treatments, the researchers decided to follow another hypothesis—that the RNA was providing the active site for its own splicing. This hypothesis was confirmed by experiments reported in reference 2.

An entirely new field of catalytic biochemistry would not have developed if the splicing found for *Tetrahymena* pre-rRNA (classified as a group I intron) had been the only example of a ribozyme. In 1983 and 1984, however, it was revealed that RNase P (ribonuclease P) was also a ribozyme. RNase P processes the 5' ends of transfer RNA (tRNA) precursors in all organisms. Ribonuclease P (RNase P) is a ubiquitous and essential ribonucleoprotein (RNP), containing both RNA and protein segments. It has been found that the RNA subunit by itself can accomplish the required chemistry under conditions of elevated counterion concentration in vitro.<sup>4</sup> Introns in fungal mitochondrial mRNA and rRNA (called group I introns) exhibit RNA splicing that is protein-dependent. The proteins in these systems are called *maturases*. It was found that *cis*-acting mutations in these introns prevented splicing, indicating that the RNA structure was important. The reference 4 and 5 publications made a key connection: The group I introns have secondary structure elements that are conserved not only with each other but also with *Tetrahymena* pre-rRNA itself. This led to the hypothesis that the group I introns would be self-splicing using the same mechanism as *Tetrahymena* pre-rRNA.<sup>5</sup> Work on bacteriophage T4 mRNA introns showed that they too were self-splicing tRNAs, contradicting earlier predictions that RNA splicing was confined to eukaryotic species.<sup>6</sup> Group II self-splicing RNAs differ from group I examples in both RNA structure and splicing mechanism.<sup>7</sup> The hammerhead ribozyme was discovered as the first small self-cleaving RNA motif.<sup>8</sup> The hammerhead ribozyme catalyzes self-cleavage at a rate some  $10^6$ -fold greater than that of uncatalyzed non-sequence-specific RNA self-cleavage. The hammerhead ribozyme will be discussed in more detail in Section 6.2.4.

Structure–function relationships were established for ribozymes by first considering how their tertiary structure must be dependent not only on the

nucleic acid–base interactions but also on interactions between ribose sugars and phosphates as well. One author pointed out that protein enzymes and ribozymes seem “inside out” with respect to each other. Ribozyme secondary structure is dependent on hydrogen bonding between nucleic acid bases that point inward with only the repetitive sugars and phosphates pointing outward, whereas protein enzyme secondary structure involves the peptide backbone with diverse amino acid side chains pointing outward for higher-order interactions.<sup>9</sup> To show that sugar positions must be used in assembling a ribozyme’s active site, researchers substituted 2'-deoxy, 2'-methoxy, or 2'-fluorine site-specifically for RNA’s 2'-OH group and measured the activity of the resulting mutant. These experiments revealed specific sugar backbone contacts necessary for catalysis.<sup>10</sup> Installation of a photocrosslinkable group on the 5' phosphate of the so-called internal guide sequence (IGS) of Group I introns showed that conformation changes at this site must accompany substrate binding and reaction.<sup>11</sup> In these systems, a specific guanosine binding site (the G site) is found on the intron RNA, an exogenous (from solution) guanosine ( $G_{OH}$ ) binds there, and the intron provides an “internal guide sequence” (IGS) to bring the two exons (3' exon region and 5' exon region) close together. (See Figure 6.1.)

Important questions regarding the individual steps in the reaction mechanism were studied during the 1990s by substituting DNA (deoxy at the 2'-sugar



**Figure 6.1** The group I intron exon-splicing mechanism. (Adapted with permission from Figure 1 of reference 24. Copyright 2004, with permission from Elsevier.)

position) for RNA as the substrate segment. These experiments teased out some mechanistic details such as the following: (1) Tertiary interactions with certain 2'-OH groups are important for docking the substrate–internal guide sequence, IGS, helix into the active site; (2) the rate of chemical cleavage step is reduced 40,000-fold, indicating that a lack of functional groups (2'-OH for one) that could protonate the leaving group is necessary for activity. Observations that magnesium ions facilitated both the stabilization of the leaving group and the activation of the guanosine nucleophile in group I intron ribozymes provided analogies to a number of protein metalloenzymes such as phosphodiesterases and RNA and DNA polymerases.<sup>12</sup>

Three-dimensional structures of ribozymes were highly desired at this time, but in the early 1990s no RNA larger than tRNA had ever been crystallized, let alone had its structure determined. The *Tetrahymena* ribozyme 160 nucleotide (nt) P4–P6 domain proved amenable to crystallization, and in 1996 its structure was solved by isomorphous replacement of heavy atoms and MAD (multiwavelength anomalous diffraction) phasing (PDB: 1GID).<sup>13</sup> See Section 3.3.3 for a discussion of these X-ray crystallographic techniques. New tertiary structure motifs and metal ion interactions in the relatively solvent-inaccessible RNA core validated biochemical data that had been previously determined, and much of the solvent phase biochemistry agreed with the solid-phase crystal structure. More will be said about the *Tetrahymena* ribozyme structure after this introduction. Next the so-called hammerhead ribozyme structure was solved (PDB: 1HMH),<sup>14</sup> followed by the structure of a 247-nt catalytically active group I ribozyme revealing an active site much like that found in a protein enzyme—that is, pre-organized for activity (PDB: 1GRZ).<sup>15</sup>

In 2002, two particular areas of interest to the reference 1 author T. R. Cech were: (1) interactions of proteins with ribozymes, as proteins enable or extend RNA activity; and (2) solving structures of ribozymes at atomic resolution. As will be seen in the following paragraphs, both of these areas are still being addressed by researchers.

### 6.2.2 Analyzing the Role of the Metal Ion

An informative review article, written in 2003, forms the basis for much of the following discussion.<sup>16</sup> Metal ions, monovalent and divalent, are known cofactors in catalytic RNA (ribozyme) activity, and they are critical for both their structural and catalytic properties. Most ribozymes are very sensitive to ionic conditions, and all require divalent cations such as Mg<sup>2+</sup> for activity. Since RNA folding is highly dependent on electrostatic interactions, metal ions are believed to be required for formation of correct structural conformations. Cations can also enable catalysis in the following ways: (1) as general bases perhaps as metal-hydroxo species, activating the nucleophile; (2) as general acids perhaps as metal-aqua species, aiding in the protonation of the leaving group; and (3) as charge neutralizers, stabilizing transition states.

Reference 16 discusses biochemical methods to predict specific ribozyme metal-binding sites and spectroscopic methods to identify and examine specific sites.

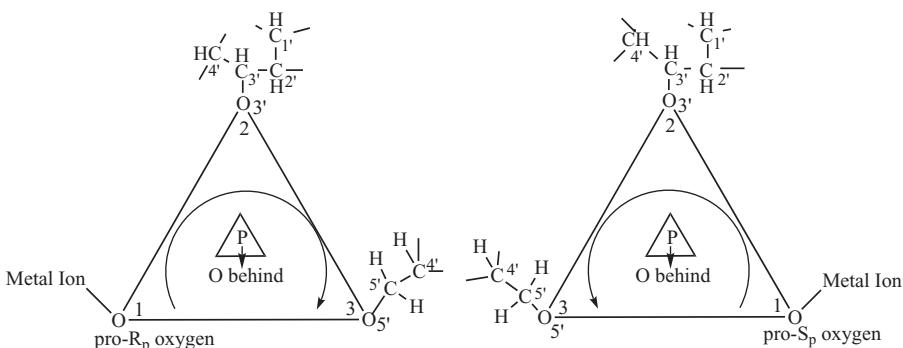
It is known that low concentrations of monovalent ions ( $\text{Na}^+$ ,  $\text{K}^+$ ) do not, by themselves, catalyze ribozyme reactions. For instance, in “small” ribozymes such as hammerhead, hairpin, Neurospora VS, and hepatitis delta virus (HDV), which catalyze a single site-specific phosphodiester bond cleavage reaction,  $\sim 1\text{ M}$  concentrations of monovalent ions will not promote catalytic activity unless micromolar concentrations of  $\text{Mg}^{2+}$  or certain other divalent cations are also present. Very high concentrations of monovalent ions can catalyze the activity of the small ribozymes but not that of “large” group I and II introns or ribonuclease P (RNase P). Figure 1a of reference 16 illustrates the cleavage reactions catalyzed by the small ribozymes as well as RNase P. Figure 1b shows the splicing reactions catalyzed by group I intron ribozymes such as *Tetrahymena* pre-rRNA using an exogenous guanosine as a nucleophile. Group II intron ribozymes catalyze a reaction similar to that of group I using a different extrinsic nucleophile.

A major challenge in studying the role of metal ions in ribozymes is deciding whether a cation is necessary only for structural purposes, only for catalytic purposes, or for both purposes. In other words, one would like to correlate metal ion sites with either the folding, or the activity, of the ribozyme. It is possible to find out that metal ions have structural purposes only by using ribozymes prohibited from undergoing their catalytic reaction by single atom changes at the active site. Ion-dependent structural changes in ribozymes have been illuminated using fluorescence resonance energy transfer (FRET),<sup>17</sup> hydroxyl radical footprinting,<sup>18</sup> small-angle X-ray scattering (SAXS),<sup>19</sup> and other experimental techniques. These techniques usually do not place the metal ion at a specific binding site but do explain the ion’s overall affect on ribozyme folding. Conversely, few experimental techniques can report on the metal ion’s effect on ribozyme activity without the metal ion concurrently causing structural alterations.

The main technique used by researchers in attempting to isolate catalytic activity from structural changes is one called the phosphorothioate metal-rescue (PS-rescue) experiment. This experiment tests the hypothesis that a metal ion binds to a particular phosphate oxygen on a specific nucleotide and that the metal ion is important to ribozyme catalytic activity. A successful PS-rescue experiment does the following: (1) substitutes one nonbridging phosphate oxygen with a sulfur atom creating a stereospecific thiophosphate; (2) tests that the substituted site has lost affinity for the “hard” magnesium cation,  $\text{Mg}^{2+}$ , but has enhanced affinity for “soft” cations such as  $\text{Cd}^{2+}$ ,  $\text{Zn}^{2+}$ , or  $\text{Tl}^+$  (see Section 1.4); and (3) tests that  $\text{Mg}^{2+}$ -dependent activity is lost but that the soft cations can “rescue” the activity thus predicting a functional catalytic cation site. The PS-rescue experiment can be successful in predicting sites of catalytic activity, although the situation is complicated by the fact that ribozyme active sites undergo several conformational changes between ground, pre-transition,

and transition states. Therefore, one might have an active site in one conformation that uses metal ion cofactor  $Mg_A^{2+}$ , while another conformation uses metal ion cofactor  $Mg_B^{2+}$  in its active site. DeRose, in reference 16, suggests that a metal ion may be considered to be functionally involved in catalysis if it (1) contributes to stabilizing the transition state and (2) is associated with a group that changes bond order during the reaction. Conversely, if the metal ion can be predicted to be far (and stay far) from the active site, then it should be considered as playing only a structural role. It should also be remembered that sulfur atoms are larger and more polarizable than oxygen atoms and that they may perturb ribozyme structure, in some cases to an unacceptable degree. In one case, that of the hammerhead ribozyme, to be discussed in more detail below, PS-rescue experiments have been confirmed using  $^{31}P$  NMR.<sup>20</sup> Electron paramagnetic resonance (EPR), electron nuclear double resonance (ENDOR), and electron spin-echo envelope modulation (ESEEM) are other spectroscopic methods used successfully to determine metal ion occupation at the active site in the hammerhead ribozyme.<sup>21</sup> (See Section 3.5 for more information on EPR techniques and see Section 3.5.3 for more information on ENDOR and ESEEM.) Use of EPR requires that a paramagnetic ion take the place of magnesium ions in the ribozyme. Fortunately, the paramagnetic  $Mn^{2+}$  ion is similar in ionic radius, enthalpy of hydration, and rate of catalytic activity when substituted for magnesium ions in the hammerhead ribozyme or in the group I intron. Hyperfine interactions between the  $Mn^{2+}$  electron spin and nuclei of the RNA ligands are observed using ENDOR and ESEEM. These help identify  $Mn^{2+}$  ions directly coordinated to phosphate and aqua ligands. These results for the hammerhead ribozyme will be discussed further in Section 6.2.4. Conclusions from the reference 16 review paper include the following: (1) Separating cation influences on RNA structure from those on catalytic activity constitutes a significant experimental challenge; (2) high- and low-affinity cation interactions have been identified as being required for ribozyme activity (high-affinity sites are involved in folding while low-affinity sites are involved in catalysis); and (3) although cations influence ribozyme function, the majority of catalytic power is probably derived from the nucleobases themselves.

In discussing the phosphorothioate metal-rescue (PS-rescue) experiment in the previous paragraph, it was mentioned that replacing one of the phosphate oxygens in an RNA polymer with a sulfur atom produces a chiral center. We say that the phosphorus atom in the phosphate connector is pro-chiral because changing one of two oxygen atoms (substituting a sulfur for instance as in the P-S rescue experiment) gives rise to a chiral center. Replacing one specific oxygen with a sulfur will give rise to an S configuration (pro-S oxygen), while changing the other one would yield an R configuration (pro-R oxygen) according to Cahn-Ingold-Prelog rules. For a phosphate group in an RNA chain, in which one goes toward the 5'-carbon atom of a sugar moiety in one direction and towards the 3'-end of another sugar in the opposite direction, the procedure for identifying pro-R and pro-S oxygens is outlined in Figure 6.2. Identifi-



**Figure 6.2** Identifying pro- $R_p$  and pro- $S_p$  oxygen atoms in RNA polymers.

Identifying pro-chiral atoms will be important in the discussion of the group I intron (Section 6.2.3) and the hammerhead (Section 6.2.4) ribozymes.

### 6.2.3 The Group I Intron Ribozyme

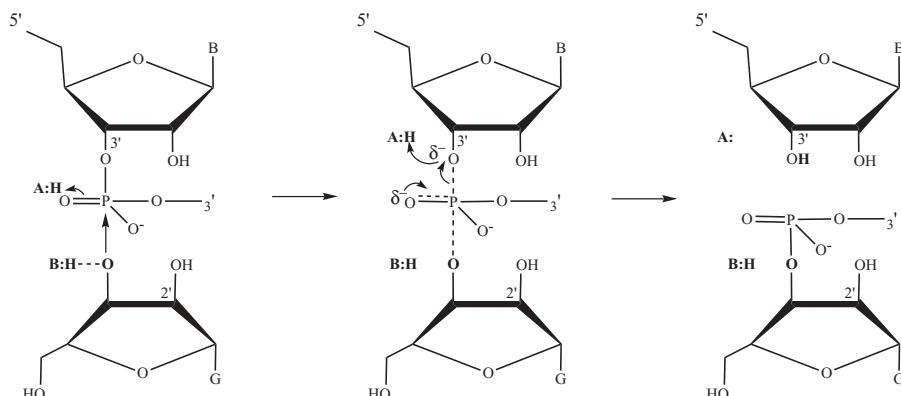
Some group I introns contain an active site that allows them to cut themselves out from a precursor RNA. In this self-splicing process, two exons are ligated and ejected while the remaining intron becomes an intact messenger RNA, transfer RNA, or ribosomal RNA.<sup>15</sup> While most introns are excised by the spliceosome, a large ribonucleoprotein complex, the introns we discuss here catalyze their own excision without assistance from any protein cofactors.<sup>22</sup> Divalent metal ions play a role in both conformation and catalysis; for instance, group I self-splicing intron ribozymes require sufficient magnesium ion (approximately 2 mM) to promote efficient folding. An exogenous guanosine nucleoside cofactor is also required for the self-splicing reaction. The introns are complex three-dimensional structures whose folding patterns are not completely understood. Most RNA ribozyme base pairs follow the Watson–Crick hydrogen bonding pattern; however, Hoogsteen base pairing, mismatched base pairs, and base triples (three bases connected by hydrogen bonds in several specific, well-known patterns) are found, often near active sites.

The intron group I ribozymes feature common secondary structure and reaction pathways. Active sites capable of catalyzing consecutive phosphodiester reactions produce properly spliced and circular RNAs. Ribozymes fold into a globular conformation and have solvent-inaccessible cores as quantified by Fe(II)-EDTA-induced free-radical cleavage experiments.<sup>23</sup> The *Tetrahymena* group I intron ribozyme catalyzes phosphoryl transfer between guanosine and a substrate RNA strand—the exon. This ribozyme also has been proposed to use metal ions to assist in proper folding, to activate the nucleophile, and to stabilize the transition state.<sup>27a</sup>

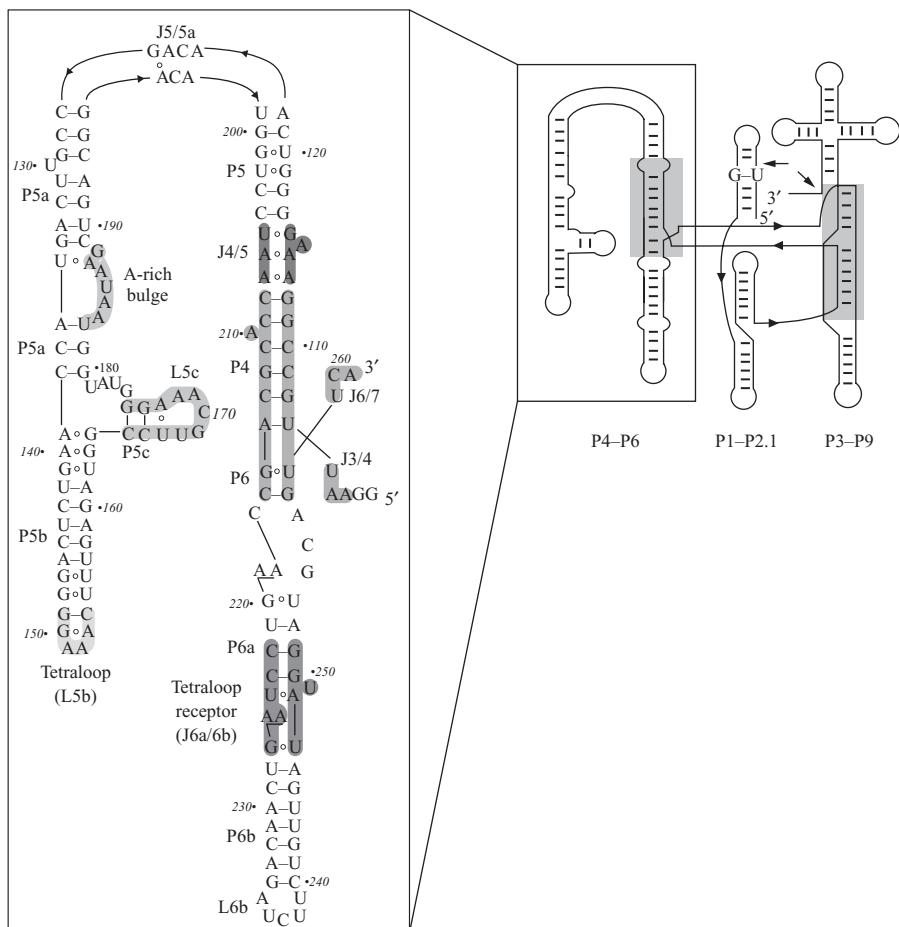
As stated previously in the introductory section, T. R. Cech and co-workers reported on the first catalytic RNA or ribozyme, the self-splicing intron of the

*Tetrahymena* pre-rRNA, beginning in 1982. Subsequent work showed that a specific guanosine binding site (the G site) is found on the intron RNA, an exogenous (from solution) guanosine ( $G_{OH}$ ) binds there, and the intron provides an “internal guide sequence” (IGS) to bring the two exons (3' exon region and 5' exon region) close together. (See Figure 6.1.) In these processes, the exon is considered to be the reaction substrate. In the first of two transesterification reactions,  $G_{OH}$  attacks at the 5' splice site, becoming covalently attached to the 5' end of the intron. Subsequently, there is a conformational change and the ribozyme's 3'-terminal guanosine (called the  $\omega G$ ) binds at the “G site.” This positions the 3' splice site for nucleophilic attack by the 3'OH of the 5' exon; and in the second transesterification the two exons are joined and released, and the intron is liberated. After release of the ligated exons, the intron's 5' region folds back to the active site and a third nucleophilic attack (by  $\omega G$ ) takes place to produce a circularized intron. At the same time an RNA oligonucleotide is released carrying the exogenous  $G_{OH}$ .<sup>24</sup> This splicing mechanism has become known as the G-dependent mechanism. (See Figure 6.1 and Figure 6.3.) Figure 6.3 shows the cleavage phosphotransesterification—labeled chem.1 in Figure 6.1. The secondary structure of an entire group I intron is shown in Figure 6.4, and a close-up view of a transesterification step is illustrated in Figure 6.8 for PDB: 1ZZN.

Group I intron phosphotransesterification reactions are carried out by a conserved active site that contains a set of imperfect double helices named P1 through P9. (See Figure 6.4.) P1–P9 helices are organized into three domains: P1–P2, P4–P6, and P3–P9. Specifically, the *Tetrahymena thermophila* intron contains two sets of coaxially stacked helices that overlap to create the active site. These helices reside in two domains of approximately equal size: P4–P6 and P3–P9. P domains are defined as base-paired regions, whereas J domains



**Figure 6.3** Mechanistic detail of the G-dependent mechanism—the cleavage phosphotransesterification. (Adapted with permission from Figure 1b of reference 16. Copyright 2003, with permission from Elsevier.)



**Figure 6.4** Secondary structure of a group I intron ribozyme. (From Figure 1 of reference 13. Reprinted with permission of AAAS.) (See color plate)

are defined as joining regions. The catalytic core is located at the junction of the P4–P6 and P3–P9 domains. The 5' exon splice site is located in the P1–P2 (substrate) domain. The internal guide sequence (IGS) guides the substrate to form an open complex with the enzyme to form the so-called P1 duplex that subsequently docks via tertiary interactions to form the active, closed ribozyme-substrate complex. Note that the final G of the IGS forms a mismatch base pair with a uracil, U; this is called the G–U wobble base. Wobble bases are a common motif in RNA secondary structure, and they have important and specific functions in all types of RNA. All group I introns contain a G–U wobble base pair at the cleavage site that is almost universally conserved.<sup>25</sup>

The guanosine binding site of group I introns is located within P7—a six-base-pair helix with a single base bulge. The conserved bulge forms a base triple (three bases connected by hydrogen-bonds in several specific, well-known patterns, a very common motif in introns) that stacks with the terminal guanosine. For instance, in the *Tetrahymena* group I intron, A263 forms a base triple with the C262–G312 base pair that stacks on top of the terminal guanosine ( $\omega$ G) in the PDB: 1X8W structure to be discussed below. (See Figures 6.5 and 6.6.) The guanosine-binding pocket is complex, bringing together nucleotides of the J6/7 and J8/7 joining regions within the major groove of P7.<sup>32</sup> The binding pocket will be discussed for several group I intron crystal structures in the following paragraphs.

The first crystal structure of a group I ribozyme, a 160 nucleotide P4–P6 domain of the self-splicing *Tetrahymena thermophila* intron, was published in 1996 at a resolution of 2.8 Å (PDB: 1GID).<sup>13</sup> The crystal structure was solved using multiwavelength anomalous diffraction (MAD) and single isomorphous replacement (SIR) using an osmium derivative to pinpoint metal-ion binding sites. The final model consisted of 154 nucleotides (nt), a total of 28 metal ions and six waters. The complete secondary structure of the intron–exon complex is shown in Figure 1 of reference 13 and reprinted as Figure 6.4. The *Tetrahymena* P4–P6 domain in the crystal structure PDB: 1GID contained about half of the intron’s active site—the base-paired segments P4, P5, and P6 and joining regions J3/4, J6/7, and J4/5—conserved elements found in the catalytic cores of all group I introns. The domain contained the P5abc extension (P5a, P5b, and P5c), required for efficient catalysis in the *Tetrahymena thermophila* intron. It also featured noncanonically paired or “bulge” regions responsible for interactions between helices (backbone–backbone interactions), metal-ion binding sites, and interactions involving a GAAA tetraloop. In the crystal structure, the P4 helix (PDB: 1GID residues C208–C216, G112–U107) interacts with the A-rich bulge (PDB: 1GID residues G188–U182) and the tetraloop receptor helix J6a/6b (PDB: 1GID residues C222–G227, G251–U247) interacts with the GAAA tetraloop (PDB: 1GID residues G149–C154). Within the A-rich bulge, A184 and A186 are invariant in group I intron subclasses IB and IC and A183 is conserved—deletion of the A-rich bulge or a point mutation of A186 to U disrupts the domain’s global structure. In fact, the global P4–P6 domain structure is sensitive to numerous nucleotide mutations in the A-rich bulge. The A-rich bulge starts the folding of the domain’s substructure to include the three-helix junction of P5a, P5b, and P5c. Phosphate oxygens in the A-rich bulge are unusually close to each other, but charge density is alleviated by the presence of several magnesium ions. The distances are collected in Table 6.1.

These interactions, plus others, stabilize the P4–P6 domain’s conformation. Site-directed mutagenesis and chemical protection studies indicate that G212 in the P4 helix interacts with residues in the A-rich bulge, and the crystal structure shows interactions with A184 (O<sub>2'</sub> G212–N1 A184 = 2.8 Å, for example). The same sorts of interactions are found between the GAAA

**TABLE 6.1 Selected Distances for Group I Intron PDB: 1GID**

Nucleotide (PDB: 1GID Residue Number)	Structural Element	Ligand Atom (PDB: 1GID Symbol, Atom Number)	Bond Distance (Å) to
A183 <sup>a</sup>	A-rich bulge	<i>pro</i> -S <sub>p</sub> -O (O2P, 1729) <i>pro</i> -R <sub>p</sub> -O (01P, 1728)	Mg <sub>2</sub> <sup>2+</sup> (6799), Mg <sub>2</sub> <sup>2+</sup> (6800), Mg <sub>3</sub> <sup>2+</sup> (6804), Mg <sub>4</sub> <sup>2+</sup> (6806), Mg <sub>5</sub> <sup>2+</sup> (6808), Mg <sub>6</sub> <sup>2+</sup> (6797)
A184 <sup>b</sup>	A-rich bulge	<i>pro</i> -S <sub>p</sub> -O (02P, 1751) <i>pro</i> -R <sub>p</sub> -O (01P, 1750)	Mg <sub>2</sub> <sup>2+</sup> -O 2.1 Mg <sub>1</sub> <sup>2+</sup> -O 2.3
A186 <sup>b</sup>	A-rich bulge	<i>pro</i> -S <sub>p</sub> -O (02P, 1793) <i>pro</i> -R <sub>p</sub> -O (01P, 1792)	Mg <sub>2</sub> <sup>2+</sup> -O 2.3 Mg <sub>1</sub> <sup>2+</sup> -O 2.5
A187	A-rich bulge	<i>pro</i> -R <sub>p</sub> -O (01P, 1814)	Mg <sub>2</sub> <sup>2+</sup> -O 2.1
G163	P5bc connector	N <sub>7</sub> (1309)	Mg <sub>3</sub> <sup>2+</sup> -N 2.6
G164	P5bc connector	O <sub>6</sub> (1335)	Mg <sub>3</sub> <sup>2+</sup> -O 2.1
A171	L5c (loop 5c)	<i>pro</i> -R <sub>p</sub> -O (01P, 1465)	Mg <sub>4</sub> <sup>2+</sup> -O 2.3
A173	L5c (loop 5c)	<i>pro</i> -R <sub>p</sub> -O (01P, 1509)	Mg <sub>5</sub> <sup>2+</sup> -O 2.4
G188	A-rich bulge	O <sub>6</sub> (1852)	Mg <sub>6</sub> <sup>2+</sup> -O 2.0

<sup>a</sup>Conserved in group I intron ribozymes.

<sup>b</sup>Invariant in group I intron ribozymes.

tetraloop and its receptor. The end result is that a number of riboses of the isolated P4–P6 domain are inaccessible to Fe(II)-EDTA cleavage—that is, they are protected from solvent. Free radicals produced by Fe(II)-EDTA in solution attack the C4' ribose carbons at each unprotected—exposed to solvent—nucleotide residue and therefore probe the solvent accessibility of domains within folded RNA.<sup>23</sup> Close packing of phosphates is relieved by outer-sphere contacts with hydrated magnesium ions (at distances < 4 Å) at several locations—between P4 and J6/6a, between P5 and P5a, and between P5b and P6a. The reference 13 authors believe that the divalent magnesium ions fulfill a role in intron (and other RNA ribozyme) folding. The following conclusions can be derived from the PDB: 1GID crystal structure observations: (1) Catalytic RNA appears to require a closely packed structure much different from the classical view of stringlike messenger RNA; (2) metal-phosphate interactions are abundant in the structure, allowing polyanionic helices to come into close contact; and (3) ribose 2'-OH groups appear to be important players in the intron structure, as a donor or acceptor of hydrogen bonds or in forming metal ion–ribose interactions.

The Cech group described an X-ray crystallographic structure of the group I intron from *Tetrahymena thermophila* in a *Science* magazine research article published in 1998 (PDB: 1GRZ).<sup>15</sup> The 5.0-Å resolution crystal structure included 247 nucleotides comprising most of the *Tetrahymena thermophila* intron. At this resolution, clear density for the ribozyme backbone was seen, and stacked bases were visualized as continuous tubes of electron density.

However, hydrogen-bonding characteristics could not be defined, nor could magnesium ions be placed with confidence. In the PDB: 1GRZ structure, the P4–P6 domain appears largely unchanged from that seen in the previous crystal structure, PDB: 1GID, an encouraging note for researchers determining structures of large RNA molecules. As before, the P4–P6 domain is hairpin-shaped with two coaxially stacked helical regions separated by a sharp bend. The P3–P9 domain, which had not previously crystallized, was found to be wrapped around the P4–P6 domain. Many interdomain interactions between P4–P6 and P3–P9 were found, including at the catalytic active site. The P3–P9 domain consists of five helices, four of which (P8, P3, P7, and P9.0) are coaxially stacked. The P9 helix makes an approximate 90° bend and interacts with the minor groove face of P5 in the P4–P6 domain. The large size of the intron—~110 × 65 Å—is compared with that of the restriction endonuclease Eco RI, which measures ~40 to 60 Å in all directions. The researchers suggest that the size difference indicates that a more efficient active site may be built with amino acids than with nucleotides. This last statement brings questions of evolutionary biology into play as RNA catalysis is thought to have preceded catalysis by amino acid-containing proteins and enzymes. Unfortunately, there is neither time nor space to discuss this fascinating topic here.

Previous experimental results indicate that the secondary structures of group I introns show a junction between the P4–P6 and P3–P9 domains that involves the close approach of the P4, P6, P3, and P7 helices. Figure 4 of reference 15 shows the ribozyme catalytic core being formed in the PDB: 1GRZ structure by the junction of these four helices: P3 (orange), P4 (yellow), P6 (pink), and P7 (green). Assembly of the helices into a catalytic conformation involves base triples formed in J3/4 and J6/7. For instance, J3/4 contains three bases: A104, A105, and A106. The nucleotide A104 stacks directly onto the P3 helix and appears to form an interacting quartet with nucleotides A270 (mismatches with A104), C217 (in P6), and A256 (mismatches with C217 in P6). Base triples forming in the J6/7 region also help to orient the P7 helix. Tetraloops are secondary structure domains that often mediate tertiary interactions in RNAs. In the PDB: 1GRZ structure, a standard GAAA tetraloop, L9 residues G323, A324, A325, A326, is modeled. This tetraloop docks into the minor groove of P5 at base pair C203–G118 and wobble base pair U202–G119.

The all-important exogenous guanosine-binding site (G site) of the intron, G414 in the PDB: 1GRZ structure, is located within the P7 helix. An adjacent A265–U310 base pair has been implicated in guanosine binding, and residues A306 (a strongly conserved nucleotide) and A261 are identified as possibly forming a “gateway” to the G site. Interactions within the P7 helix and backbone–backbone interactions with P4 in the vicinity of U106 and U107 distorts this helix and compresses its major groove in such a way to provide a unique binding site for the guanosine substrate (G414). The second group I intron substrate, the 5' splice site, located within the P1 duplex, is not present in the PDB: 1GRZ crystallized ribozyme. However, the reference 15 authors do

present a model for the 5' splice site based on previous functional group substitution experimental data and the structure of the PDB: 1GRZ ribozyme. The P1 duplex is hypothesized to fit into a concave binding site provided in the ribozyme's core. The "sides" of the pocket would be J4/5 and P3, while the "floor" would be J8/7. (See Figure 6.4.) It is known that J8/7 is involved in binding P1, and its other interactions with the backbone of P3 and helix P4, for instance, defines a space in which the phosphate at the 5' splice site is well placed with respect to the guanosine nucleophile. The authors also speculate as to which residues might coordinate the divalent metal ions required for catalytic activity. Four phosphate positions are named—those at positions A306, A261, A207, and C208—that span the "gateway" to the G site. The phosphate groups of these nucleotides have been previously identified as sites of phosphorothioate interference.<sup>26</sup> Briefly, substitution of sulfur at phosphate oxygen positions thought to be important metal-ion binding sites involved in ribozyme folding or catalysis causes a change in catalytic activity that confirms or denies the importance of the substituted site. This type of experiment has been described earlier in Section 6.2.2.

Other research groups have conducted biochemical experiments intended to place metal ions in the catalytic site of ribozymes. Groups led by Piccirilli and Herschlag proposed a three-metal model based on the difference in Mn<sup>2+</sup> concentrations needed to rescue different sulfur or amino substitutions of substrate functional groups.<sup>27</sup> The experiments were intended to quantitatively determine metal ion affinities that in turn would allow individual metal ions to be distinguished from one another. The functional group substitution also identified the important groups involved in catalysis. In the three-metal ion model for the *Tetrahymena* ribozyme, two different metals (M<sub>B</sub> and M<sub>C</sub>) were proposed to coordinate to the O2' and O3' sites of ωG with M<sub>C</sub> also coordinating to the pro-S<sub>p</sub> O of the 3'-exon terminus. A third metal (M<sub>A</sub>) was proposed to coordinate the pro-S<sub>p</sub> oxygen of the 3'-exon terminus as well as to the O3' of the 5'-exon residue. (dt<sub>-1</sub> in the PDB: 1U6B structure, see Figure 6.7). In this model, none of these three metal ions bridge between the scissile phosphate and the leaving group—the O3' of ωG. Figure 8 of reference 27b illustrates the three-metal models in two- and three-dimensional forms.

In the reference 15 conclusions, the Cech group mentions the good agreement between the group I intron ribozyme crystal structures achieved as of this date (PDB: 1GID and 1GRZ) and the constructed model of Michel and Westhof. The latter model was based mainly on comparative sequence analysis, stereochemical constraints, and biochemical testing of proposed ribozyme structural interactions.<sup>28</sup> The reference 15 researchers were concerned that the crystallization construct may deviate from the catalytically significant structure, a concern that applies to any RNA crystal structure, especially since RNA enzymes appear to be much more dynamic in solution (and perhaps in the solid state as well) compared to their protein enzyme counterparts. In contrast to the hammerhead ribozyme to be discussed in Section 6.2.4, the

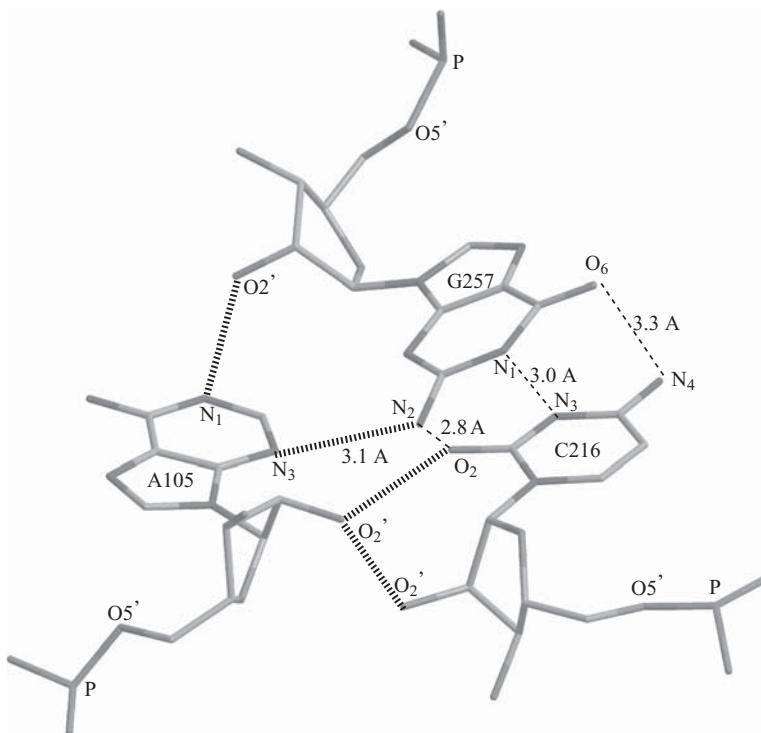
group I intron crystal structures exhibit much more agreement between biochemical and spectroscopic studies in solution and the structure observed in the solid state. The researchers believe that the difference may lie in the fact that, in contrast to the hammerhead, group I introns appear to be largely pre-organized to bind their substrates and carry out the catalyzed reaction upon them.

Next, the Cech group studied mutants of the *Tetrahymena thermophila* intron P4–P6 domain that formed a more stable tertiary structure in solution, hoping that this would lead to more easily crystallized ribozymes and more stable crystals. One mutant,  $\Delta$ C209 with the C209 residue deleted, formed crystals suitable for X-ray crystallography that diffracted to a resolution of 2.25 Å (PDB: 1HR2).<sup>29</sup> The  $\Delta$ C209 RNA construct contained 159 nucleotides with 27 octahedrally coordinated magnesium ions, 15 additional magnesium cations, and 129 water molecules distributed in two molecules in the asymmetric unit of the crystal. The better-diffracting crystals allowed assignment of innersphere andoutersphere coordination contacts for 27 magnesium ions and established the positions of numerous coordinated water molecules found on the inside rather than at the surface of the folded RNA domain. The authors proposed that the water molecules might stabilize noncanonical base-pair interactions—U–G or A–G pairs, for instance—in the ribozyme. The PDB: 1HR2 global structure was relatively unchanged compared to the wild-type structure (PDB: 1GID), and the two molecules in the asymmetric unit were essentially equivalent. However, changes were noted in the immediate vicinity of the deleted residue. Removal of the C209 residue in helix P4 causes A210 (bulging out in the wild-type ribozyme) to move into the helix and make a mismatch base pair with G111. The purine–purine base pair pushes the phosphate backbone out slightly and the helix is widened. The base pair itself is somewhat twisted, but hydrogen bonding distances are still maintained—for instance, G111 O<sub>6</sub>–A210 N<sub>6</sub> = 2.9 Å. The G–A motif has been observed in other RNAs. In the same P4 region, neighboring hydrogen-bonding base pairs form between G112–C208 and G110–C211. The authors believe that the elimination of the A210 P4 helix bulge and formation of the G111–A210 base pair caused the increased stability of the  $\Delta$ C209 mutant. However, the catalytic activity of the intron for self-splicing decreases 91-fold with the  $\Delta$ C209 mutation because the C209–G111–U305 base triple that is part of the ribozyme's active site can no longer form.

The PDB: 1HR2 crystal structure showed electron density corresponding to well-occupied binding sites for several magnesium ions. Inner-sphere metal ion coordination distances were defined to be within 1.9–2.5 Å of appropriate ligands, and outer-sphere coordination distances to be 2.5–3.3 Å from coordinating water molecules or other intervening ligands. The coordination sphere for the innersphere magnesium ions agreed with those found in the PDB: 1GID structure both in nucleotides atoms involved in bonding and in bonding distances although in most cases, bond lengths are 0.1–0.2 Å longer for the

PDB: 1GID structure. (See Table 6.1.) The PDB: 1HR2 bond length data are collected in Table S1 of reference 29. The majority of the directly coordinated magnesium ions were found near the A-rich bulge ( $Mg_1^{2+}$ ,  $Mg_2^{2+}$ , and  $Mg_6^{2+}$ ) or the three-helix junction ( $Mg_3^{2+}$ ,  $Mg_4^{2+}$ , and  $Mg_5^{2+}$ ). The metal ions appear to help fold important structural motifs, stabilize the structures formed by P5a and P5c, and facilitate the group I intron's tertiary structure. Other fully hydrated magnesium ions, with six aquo ligands, also stabilize the intron's tertiary structure mainly in the major groove of an A-form helix. The majority of ordered water molecules found in the structure, whether or not they are attached to magnesium ions, appear in the group I intron's interior. The authors believe that the water molecules may work independently or cooperatively with magnesium ions, but in either case they assist the RNA bases to extend their hydrogen bonding capability and further stabilize the intron's tertiary structure. In conclusion, the reference 29 researchers believe that the higher-resolution model established for PDB: 1HR2 provided new, important details of metal ion–RNA interactions as well as identifying a core of ordered water molecules. Both of these could be integral to RNA tertiary structure formation.

In 2004, the Cech group published the 3.8-Å resolution X-ray crystallographic structure of a catalytically active *Tetrahymena* ribozyme containing stabilizing mutations that helped achieve greater thermal stability and better diffracting crystals (PDB: 1X8W).<sup>24</sup> In solution tests on the enzymatic form of the mutant, the ribozyme's ability to cleave its RNA substrate remained intact. However, the crystallized species did not contain the exon but corresponded to the enzymatic form in the absence of its exon substrate. Even though the crystal structure did not include the ribozyme's RNA substrate, it did yield more information about the ribozyme's tertiary structure both close to and away from the catalytic pocket. Many of the interactions found in the P6 domain involved base triples. Base triples are a common stabilizing motif in RNA tertiary structures and have been described fully in an article by Jennifer Doudna and co-workers.<sup>30</sup> Essentially, base triples most often involve an adenosine nucleotide interacting with a traditional base pair and are found in two forms—Type I and Type II. In the P4–P6 domain of the *Tetrahymena* group I intron, Type I base triples involve the N<sub>1</sub>, C<sub>2</sub>, N<sub>3</sub>, and 2'-OH atoms or groups along the minor groove face of the adenosine contacting the minor groove surface of a base pair (including its sugars) burying ~195 Å<sup>2</sup> of molecular surface area. Type II base triples involve the N<sub>1</sub>, C<sub>2</sub>, N<sub>3</sub>, and 2'-OH atoms or groups along the minor groove face of the adenosine contacting about one-half of the minor groove surface of a base pair burying less surface area (~145 Å<sup>2</sup>). One PDB: 1X8W structure type I base triple in the P6 domain uses the minor groove side of the A105 base and ribose to contact the entire minor groove surface of the C216–G257 pair. This is shown in Figure 6.5 as adapted from Figure 3D of reference 24. Note that in Figure 6.5, Watson–Crick base pairing between C216 and G257 are shown in dashed bond format, while the triple base interactions with A105 are shown in hashed bond format.

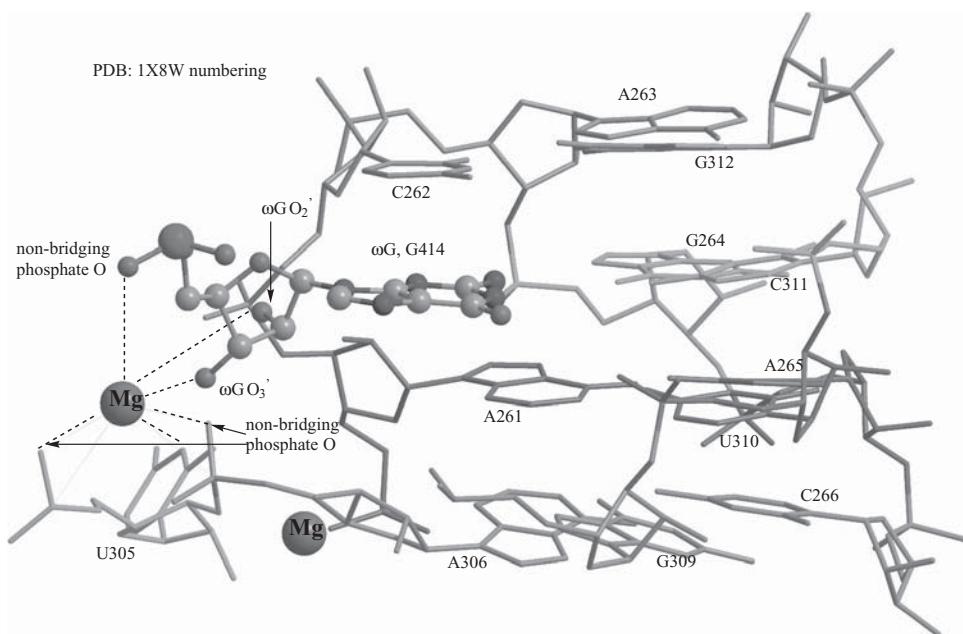


**Figure 6.5** Triple base pair for the *Tetrahymena* ribozyme (PDB: 1X8W). Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. (Printed with permission of CambridgeSoft Corporation.)

Another base triple between the bulged-out A218 in P6 and the P3 domain base pair C102–G272 ties together the two major domains of the *Tetrahymena* ribozyme using a type II base triple (see magenta lines in Figure 3A of reference 24).

In the P7 domain G binding site, the 3'-terminal guanosine ( $\omega$ G, G414 in PDB: 1X8W) is part of a base triple with the G264–C311 pair. Since a conserved G–C pair has been shown to be important in establishing the G binding site in group I introns, the authors believe that the particular G264–C311 pair can be specified as the G binding site facilitator. Three interactions form between  $\omega$ G and the G264–C311 pair:  $\omega$ G N<sub>1</sub>–G<sub>264</sub> O<sub>6</sub> = 3.3 Å,  $\omega$ G N<sub>2</sub>–G<sub>264</sub> N<sub>7</sub> = 3.0 Å, and  $\omega$ G O<sub>6</sub>–C<sub>311</sub> N<sub>4</sub> = 4.0 Å. The  $\omega$ G–G264–C311 base triple is sandwiched by three other layers of base triples. Above  $\omega$ G–G264–C311 is a base triple composed of the C262–G312 pair with A263, while below are two base triples: the A265–U310 pair with A261 and the C266–G309 pair with A306 (see Figure 4 of reference 24, and Figures 6.5 and 6.6).

The resolution of the PDB: 1X8W crystal structure does not allow identification of magnesium ions. However, heavy atom [europium (Eu) and iridium,

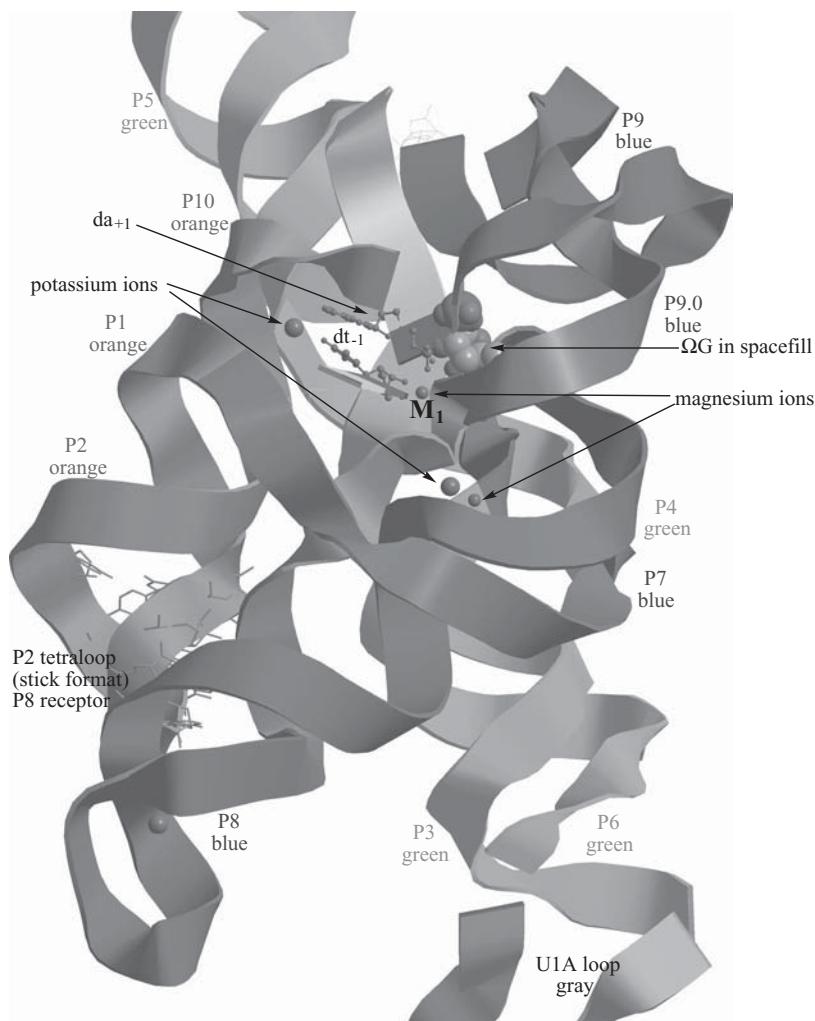


**Figure 6.6** Base triples surrounding  $\omega$ G in PDB: 1X8W. Magnesium ions in position to facilitate nucleophilic attack. Visualized using Chem3D Ultra and ChemDraw Ultra, version 10.0. (Printed with permission of CambridgeSoft Corporation) (See color plate)

(Ir)] models were used to place magnesium ions in the final native structural model. Each of the four molecules in the PDB: 1X8W asymmetric unit showed a magnesium ion in close contact with at least one of the two ribose hydroxyl groups of  $\omega$ G, G414 in PDB: 1X8W. The PDB: 1X8W model shown in Figure 6.6 shows a  $Mg^{2+}$ –G414 O3' bond equal to 2.7 Å and a  $Mg^{2+}$ –G414 O2' bond equal to 4.4 Å. G414 is rendered in ball-and-stick format with CPK atom coloring—red, oxygen; blue, nitrogen; gray, carbon; magenta, phosphorus. As stated previously, the  $\omega$ G 3'-OH is the attacking nucleophile in the cleavage reaction and the 2'-OH is believed to coordinate one of the magnesium ions involved in the three-metal mechanism for group I introns.<sup>27</sup> The same magnesium ion contacts nonbridging phosphate oxygens of orange-brown C262 (3.6 Å), green U305 (2.5 Å), and magenta A306 (2.5 Å). These specific oxygen atoms were previously identified in biochemical experiments as important to the guanosine addition step (Figure 6.1, parts ii–iii; Figure 6.3). As will be seen in the paragraphs below, similar positions for magnesium ions are found in the PDB: 1U6B and 1ZZN structures. In Figure 6.6, one sees a stack of base triples surrounding the  $\omega$ G base (G414 in PDB: 1X8W) with G414 itself as part of a base triple with the G264–C311 base pair. The coloring system in Figure 6.6 is similar to that of the informative Figure 4 of reference 24.

Another X-ray crystallographic structure, at 3.1-Å resolution, of the purple bacterium *Azoarcus sp.* group I self-splicing intron was published in 2004 by

the Scott Strobel research group (PDB: 1U6B).<sup>22</sup> This 222 nucleotide structure included the *Azoarcus* intron in complex with both of its exons in a conformational state preceding the second step of splicing. This is called the pre-2S state, in Figure 6.1 represented as the part iv structure, preceding the chem.2 step. The crystallized pre-2S state involved the introduction of 2'-deoxy substitutions at U<sub>-1</sub> (dt<sub>-1</sub>), A<sub>+1</sub> (da<sub>+1</sub>), A205 (da<sub>205</sub>), and ΩG (Ωdg, the 3'-OH terminus, PDB: 1U6B residue G206). These substitutions reduce the intron's splicing activity 10<sup>6</sup>-fold so that the structure may be visualized using X-ray crystallography. The intron-exon complex was prepared by annealing the RNA transcript—residues αG1–A190—with two RNA–DNA chimerical oligonucleotides (CAT, 3 nt, PDB: 1U6B chain D and dCIRC, 22 nt, PDB: 1U6B chain C containing the ΩG, G206 residue). The chimerical oligonucleotides are represented by lowercase letters—that is, cadt<sub>-1</sub> for the CAT nucleotide and da<sub>+1</sub> through Ωdg for the dCIRC nucleotide. The all-ribose version of this complex produced ligated exons, whereas the crystallized, deoxy-substituted complex showed no splicing activity after several days. The pre-2S intron appears as three coaxially stacked helical elements. The first of these includes P10 (PDB: 1U6B residues αG1, 5–9), P1 (PDB: 1U6B residues 10–12), and P2 (PDB: 1U6B residues 13–37, including a GAAA tetraloop). In Figure 6.7, these are colored orange. The second element includes P5 (PDB: 1U6B residues 60–84, including a GCCC tetraloop), P4 (PDB: 1U6B residues 51–56 and 88–93), P6 (PDB: 1U6B residues 94–96 and 122–124), and P6a (PDB: 1U6B residues 100–119, including a large loop that interfaces with the co-crystallized RNA binding protein U1A). These elements are colored green in Figure 6.7. The third element includes P9.0 and P9 (PDB: 1U6B residues 180–205, including a GAaa tetraloop), P7 (PDB: 1U6B residues 173–178 and 128–135), P3 (PDB: 1U6B residues 41–47 and 136–143) and P8 (PDB: 1U6B residues 144–166, including a CAAA tetraloop). These elements are colored blue in Figure 6.7. Additionally, a number of connectors join helical segments. These include: J2/3 (residues 38–40); J3/4 (residues 47–50); J4/5 (57–59); J5/4 (85–87); J6/7 (126–128); and J8/7 (residues 167–172). This last important joint is colored violet in Figure 6.7. Chimerical internal guide sequence (IGS) deoxy residues a<sub>+1</sub>–c<sub>+6</sub> and dt<sub>-1</sub>–c<sub>-3</sub> are colored yellow except for a<sub>+1</sub> and dt<sub>-1</sub>, which are colored red and displayed in ball-and-stick format. This pair is in close proximity to the ΩG (PDB: 1U6B G206) residue, visualized in space fill format in Figure 6.7. The structure also visualizes 18 metal ions, some of which are shown in Figure 6.7. Note that a potassium ion as well as a magnesium ion are positioned to bond to atoms in ΩG, a<sub>+1</sub>, and dt<sub>-1</sub>. The P2 GAAA tetraloop forms contacts with tetraloop receptors in helix P8 (GAAA tetraloop residues in orange stick format in Figure 6.7), while the P9 tetraloop (GAaa) forms interactions with residues in helix P5. The relative placement of helices P4–P6 and P3–P9.0 approximates that observed in the *Tetrahymena* structures discussed previously in this section. In the PDB: 1U6B structure, the 5'-exon's 3'-OH group (residue dt<sub>-1</sub>) is positioned for inline nucleophilic attack on the scissile phosphate at position da<sub>+1</sub> of the intron-3'-exon junction (see below).



**Figure 6.7** A bacterial self-splicing group I intron with both exons (PDB: 1U6B). Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. (Printed with permission of CambridgeSoft Corporation.) (See color plate)

In the PDB: 1U6B structure the 3'-exon is held in place through interaction with the internal guide sequence (IGS) in helix P10 (PDB: 1U6B residues  $\alpha$ G1, G5, C6, C7, G8, and U9). The 5'-exon 3-residue fragment is also held in place by IGS sequence G10, U11 and G12. The dt<sub>-1</sub> residue in the cadt<sub>-1</sub> nucleotide forms a wobble base mispair with G10 of the IGS, which is in turn recognized by the conserved A-rich J4/5–J5/4 symmetric loop region. The hydrogen bonding between the IGS sequence G10, U11, and G12 and the cadt<sub>-1</sub> nucleo-

tide forms the only interaction holding the 5'-exon in place because any covalent bond between it and the intron has been broken during the first step of the splicing sequence (chem.1 step of Figure 6.1, parts ii–iii). In the pre-2S structure the 3'-splice site is formed primarily by interactions with the ΩG, PDB: 1U6B G206 residue that is in turn bound into a pocket created by four nucleotides: A127, G128, A129, and G130. A127 stacks under ΩG, G128, and A129 stack over ΩG, and G130 is coplanar and makes a Hoogsteen base pair with it. A129 resides at the P7 helix apex and makes a base triple with the G128–C178 base pair. J6/7 residues U126, A127, and G128 span 22 Å across the active site, whereas the usual distance between three stacked bases would be approximately 9 Å. These residues are important in that they stack with ΩG (A127, G128) and the extended backbones of A127 and G128 coordinate an active site metal ion ( $3'\text{O G128-K}^+ = 2.7\text{ \AA}$ ).

The Strobel research group compared the metal ions found in the PDB: 1U6B crystal structure to those found using biochemical methods, specifically the metal ions  $M_A$ ,  $M_B$ , and  $M_C$  described by the Herschlag group using sulfur or amino substitutions rescued with soft divalent metal ions.<sup>27</sup> As described previously, the metal ions have specific purposes: (1)  $M_A$  activates the nucleophile—the O<sub>3'</sub> of U<sub>-1</sub> or dt<sub>-1</sub> in the PDB: 1U6B nomenclature; (2)  $M_B$  stabilizes the leaving group—the O<sub>3'</sub> of ΩG; and (3)  $M_C$  helps neutralize the negative charge on the trigonal bipyramidal phosphorus transition state as well as coordinating to the ΩG O<sub>2'</sub>. In the PDB: 1U6B structure, only two metal ions are positioned near the scissile phosphate (the a<sub>+1</sub> phosphate). The metal ion labeled as  $M_1$  was found to be a magnesium ion through heavy metal substitution experiments. This Mg<sup>2+</sup> coordinated to the nonbridging oxygens of four phosphates and had an outer sphere contact with a fifth. In Figure 6.7, this magnesium ion is identified as  $M_1$ . In the PDB: 1U6B structure,  $M_1$  coordinated to the pro-R<sub>p</sub>-oxygen of dt<sub>-1</sub> at a distance of 2.4 Å.  $M_1$  also coordinated to the O<sub>3'</sub>-oxygen of dt<sub>-1</sub> ( $\text{Mg}^{2+}\text{-O3'}\text{ dt}_{-1} = 2.4\text{ \AA}$ ), leading these researchers to identify it as comparable to the biochemically defined  $M_A$  described in reference 27b. A second metal ion in the PDB: 1U6B structure was defined as a K<sup>+</sup> ion, again using heavy metal substitution experiments. The potassium ion coordinates to a modeled-in O<sub>2'</sub> of ΩG ( $K^+\text{-O2'}\text{ dΩg} = 2.7\text{ \AA}$ ) and therefore was compared to the biochemically identified  $M_C$ . In biochemical experiments,  $M_C$  was found to coordinate to the pro-S<sub>p</sub>-oxygen of the scissile phosphate; however, this coordination is not found in the PDB: 1U6B structure. The closest approach of the potassium ion to the scissile phosphate occurs with the pro-R<sub>p</sub> nonbridging oxygen with a bonding distance of approximately 5 Å. The authors believe that the absence of O<sub>2'</sub> in the ΩG deoxy nucleotide may have contributed to the metal ion positions found in the PDB: 1U6B crystal structure, a proposal to keep in mind during analysis of the PDB: 1ZZN crystal structure to be discussed next. Importantly, the magnesium ion identified as  $M_1$  in the PDB: 1U6B structure is bringing the scissile phosphate and O<sub>3'</sub> oxygens of ΩG and dt<sub>-1</sub> into alignment for a S<sub>N</sub>2 nucleophilic attack—that is, the O<sub>3'</sub>-ΩG-P-O<sub>3'</sub>-dt<sub>-1</sub> bond angle is equal to 150° and the O<sub>3'</sub>-dt<sub>-1</sub>-scissile P distance equals 3.4 Å.

A more detailed crystal structure for the *Azoarcus sp.* ribozyme was published by Stahley and Strobel in *Science* magazine in 2005 (PDB:1ZZN).<sup>31</sup> This structure visualized a catalytically active group I intron splicing intermediate. The construct crystallized retained the ability to catalyze exon ligation at a slow rate and displayed a structure attained just before the second chemical step (pre-2S, Figure 6.1, part iv). It contained the complete intron, both exons, and the scissile phosphate, all the functional groups previously associated with catalytical metal ion coordination and the 2'-OH of the terminal guanosine ( $\omega$ G, PDB: 1ZZN residue G206). Inclusion of the  $\omega$ G 2'-OH normally would increase the rate of splicing 10<sup>6</sup>-fold; however, the rate is slowed (to encourage crystallization) by a 2'-deoxy substitution at the 5'-exon's last nucleotide, U<sub>-1</sub> (5MU1 in the terminology of PDB: 1ZZN). Two metal ions visualized in the structure serve to position the substrates (the 3'- and 5'-exons), activate the nucleophile (O3' of 1ZZN residue variously identified as 5MU1, U<sub>-1</sub>, dt<sub>-1</sub> (in PDB: 1U6B)), and stabilize the charge on the leaving group (O3' of 1ZZN residue  $\omega$ G) and the scissile phosphate. The two visualized magnesium ions in the active site, placed 3.9 Å apart, coordinate 10 oxygen ligands including those of the 5'-exon attacking nucleophile (O3' of U<sub>-1</sub>-Mg<sup>2+</sup> = 2.1 Å) and the intron leaving group (O3' of  $\omega$ G-Mg<sup>2+</sup> = 2.1 Å). See Figure 3C of reference 31. Other bond distances and angles are collected in Table 6.2.

The *pro-S<sub>p</sub>* oxygen's of nucleotides U<sub>173</sub> for Mg<sub>1</sub><sup>2+</sup> and A<sub>87</sub> for Mg<sub>2</sub><sup>2+</sup> make coordination sphere contacts to complete the magnesium ion octahedral coordination sphere in this structure. The two magnesium ions appear to be well-placed to catalyze the exon ligation reaction with Mg<sub>1</sub><sup>2+</sup> coordinated to the nucleophile (O3' of U<sub>-1</sub>) and the A<sub>+1</sub> scissile phosphate *pro-R<sub>p</sub>* oxygen. Mg<sub>2</sub><sup>2+</sup> also coordinates the A<sub>+1</sub> scissile phosphate *pro-R<sub>p</sub>* oxygen, the O2' oxygen, and the O3' leaving group of  $\omega$ G (G206 in PDB: 1ZZN). In addition, the metal ions coordinate all biochemically predicted metal ion ligands. In the PDB: 1ZZN structure the O3' U<sub>-1</sub> nucleophile and A<sub>+1</sub> scissile phosphate group are ideally aligned for the predicted nucleophilic attack by a S<sub>N</sub>2 mechanism with the O3' (U<sub>-1</sub>)-P (A<sub>+1</sub>) distance equal to 3.2 Å and the O3' (U<sub>-1</sub>)-P (A<sub>+1</sub>)-O3' ( $\omega$ G) angle equal to 175°. The catalytic site is illustrated in Figure 6.8, with nucleophile O3', phosphorus of scissile phosphate group, and leaving group O3' in yellow. Other magnesium ion ligand atoms are shown in orange. Exon-exon bond formation is indicated.

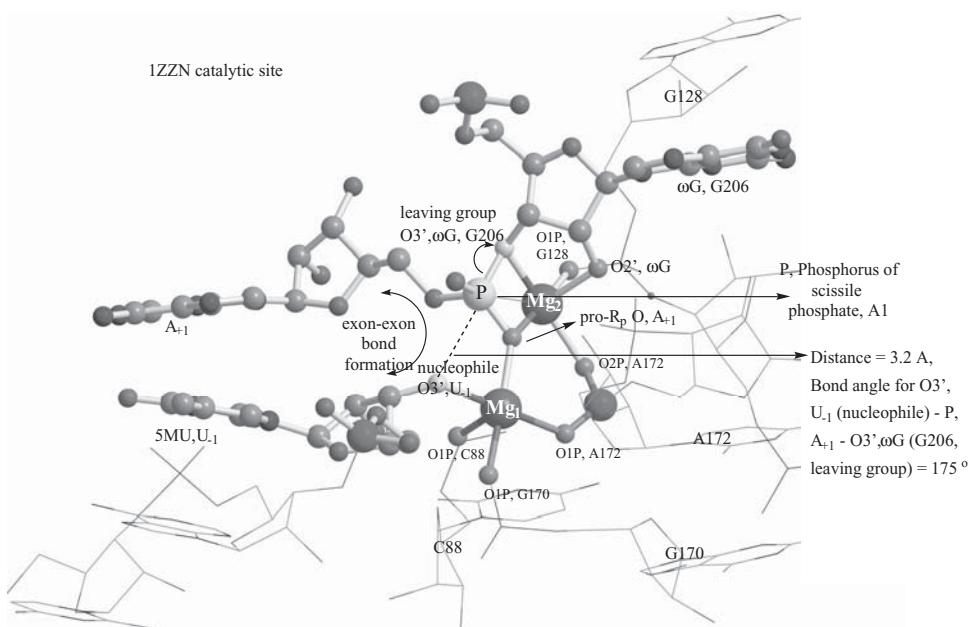
The authors Stahley and Strobel<sup>31</sup> believe that this two-metal ion intermediate describes the catalytic site bonding characteristics better than that of the three-metal ion described by Herschlag and Piccirilli.<sup>27</sup> Although Stahley and Strobel's results for PDB: 1ZZN do not rule out a disordered third metal ion in their crystal structure, they believe that the great majority of the biochemical data is explained by their two-metal model. Consequently, they would equate their Mg<sub>1</sub><sup>2+</sup> (M<sub>1</sub>) with reference 27's M<sub>A</sub> and their Mg<sub>2</sub><sup>2+</sup> (M<sub>2</sub>) with reference 27's M<sub>B</sub> plus M<sub>C</sub>. These authors also refer to other group I intron structures published in 2004 and 2005. In the PDB: 1X8W structure published

**TABLE 6.2 Selected Distances for Group I Intron PDB: 1ZZN**

Nucleotide (PDB: 1ZZN Residue Number)	Intron or Exon (Structural Element)	Bonding Atoms (PDB Symbol, Atom Number)	Bond Distance to $Mg_{\text{2+}}^{2+}$ (5550) or $Mg_{\text{i+}}^{2+}$ (5552)(Å)	
G128	Intron (P9.0) part of base triple G128–C178–A129	<i>pro</i> -S <sub>p</sub> -O (O1P, 3655)	$Mg_{\text{2+}}$ –O	2.0
A172	Intron (J8/7) stacks with U173–A134 base pair	Two nonbridging phosphate oxygens (4603, 4604)	$Mg_{\text{2+}}$ –O $Mg_{\text{i+}}$ –O	2.2 2.2
G170	Intron (J8/7) stacks with U173–A134 base pair	<i>pro</i> -S <sub>p</sub> -O (O1P, 4559)	$Mg_{\text{i+}}$ –O	2.1
C88	Intron (P4) base pair with G56	<i>pro</i> -S <sub>p</sub> -O (O1P, 2590)	$Mg_{\text{i+}}$ –O	2.2
A <sub>+</sub> 1	3'-Exon scissile phosphate	<i>pro</i> -R <sub>p</sub> -O (O2P, 5360)	$Mg_{\text{2+}}$ –O $Mg_{\text{i+}}$ –O	2.0 2.2
5MU1	5'-Exon	O3'-O (5540)	$Mg_{\text{i+}}$ –O3'	2.1
U <sub>-1</sub> (dt <sub>-1</sub> in PDB: 1U6B)	Nucleophile	O3'-O (5540)–P A <sub>+</sub> 1 (5358)	O3'-O–P	3.2
G206 ( $\omega$ G)	Intron terminus, leaving group O3'-O	O3'-O (5343) O2'-OH (5345)	$Mg_{\text{2+}}$ –O3' $Mg_{\text{2+}}$ –O2'	2.1 2.2
<i>Bond Angle and Distances for S<sub>N</sub>2 reaction at Catalytic Site</i>				
O3'-O (nucleophile, U <sub>-1</sub> )–P (A <sub>+</sub> 1)–O3'-O (leaving group, $\omega$ G)			Bond angle = 175°	
O3'-O (nucleophile, U <sub>-1</sub> )–P (A <sub>+</sub> 1)				3.2
$Mg_{\text{2+}}$ (5550)– $Mg_{\text{i+}}$ (5552)				3.9

by the Cech group in 2004, one metal ion (possibly  $Mg_{\text{2+}}$ ) coordinated to the O2' and O3' of  $\omega$ G.<sup>24</sup> Although this structure included  $\omega$ G, it did not include the scissile phosphate, the 3'- or 5'-exons, or the internal guide sequence (IGS) to which the exons bind. The PDB: 1Y0Q group I intron crystal structure published in 2005 also included a single metal ion that appeared to be equivalent to  $Mg_{\text{i+}}$ .<sup>32</sup> This metal coordinated the O3' of U<sub>-1</sub>. The construct in the latter structure included the nucleophile O3' and the  $\omega$ G but not the scissile phosphate or the 3'-exon.

The following conclusions were made by Stahley and Strobel for group I intron studies: (1) The group I intron reaction is reversible from that seen in the PDB: 1ZZN structure, with U<sub>-1</sub> becoming the leaving group and the exogenous G O3' becoming the nucleophile (see Figures 6.1 and 6.3); (2) the two-



**Figure 6.8** A group I intron transition state stabilized by two magnesium ions (PDB: 1ZZN). Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. (Printed with permission of CambridgeSoft Corporation.) (See color plate)

metal ion mechanism proposed for PDB: 1ZZN is similar to that seen for a wide variety of protein enzymes and for promotion of phosphoryltransfer reactions (many of these having a metal-metal distance around 3.9 Å); and (3) a marked similarity exists between the proposed PDB: 1ZZN active site and the active sites of RNA and DNA polymerases that feature two metal ions and very similar coordination spheres.

In early 2005, before publication of the PDB: 1ZZN structure just discussed, the research group of Barbara Golden published the structure of a third group I intron [PDB: 1Y0Q; please note that the third character in the PDB accession number is a zero (0), and not a capitol “oh” (O)].<sup>32</sup> This structure visualizes the second group I intron in the *orf* 142 gene of *Staphylococcus aureus* bacteriophage Twort. The 242 nucleotide (nt) ribozyme from the intron was co-crystallized with a 4-nt product analog RNA, and the structure was determined by X-ray crystallography at a resolution of 3.6 Å. The Twort ribozyme-product complex contained an active enzyme, including both 2'-OH groups—on ωG and U<sub>-1</sub>—at the splice site; however, it lacks the scissile phosphate-containing nucleotide—A<sub>+1</sub>.

The tertiary structure of this ribozyme was found to be very similar to that of *Azoarcus* sp. (PDB: 1U6B and 1ZZN). Many of the tertiary interactions found in the PDB: 1Y0Q structure are similar or identical to those seen as

conserved features in other RNA–RNA or RNA–protein structures. For instance, the Twort guanosine-binding site for the ωG nucleotide adopts a tertiary structure very similar to that found in other group I introns only changing the specific nucleotide found at a given position. Reference 32 lists these homologous positions for the Twort, *Tetrahymena thermophila* and *Azotarcus* introns in its Table 1. These researchers describe the guanosine-binding site as located in the six-base-pair P7 helix with a single-base bulge at residue C121. Contributions from nucleotides in the J6/7 and J8/7 junctions are also found in the P7 major groove. Notably in the PDB: 1Y0Q structure, ωG (G252) forms a base triple interaction with the G122–C192 base pair within helix P7. The G122–C192 base pair also interacts with the C121 bulge base. A second base triple composed of A120–U193–C121 stacks with the ωG–G122–C192 triple. Similar base triple interactions for PDB: 1X8W are shown in Figure 6.6. The ωG (G206) base triple is formed by ωG–G128–C178 (PDB: 1Y0Q numbering) with a continuing interaction to A129.

Next, the Golden research group co-crystallized the Twort ribozyme with manganese ions (as better diffracting magnesium ion analogs) to show interactions of metal ions within the PDB: 1Y0Q structure. A manganese-binding site (Mn1) was positioned well to be M<sub>A</sub> of the three-metal-ion mechanism based on biochemical experiments, especially those of Herschlag and co-workers as discussed for the other group I introns.<sup>27</sup> In the PDB: 1Y0Q structure, Mn1 was placed within 4 Å of the O3' of ωG and the O3' of U<sub>-1</sub> (U4, chainB in PDB: 1Y0Q) at the 5' splice site and exhibited innersphere coordination to other nonbridging phosphates as predicted by the biochemical experiments. The same metal ion showed coordination to a scissile phosphate oxygen when this group was modeled into the structure. The biochemically predicted M<sub>B</sub> and M<sub>C</sub> metal ions were not observed in the electron density maps for PDB: 1Y0Q. Modeling of M<sub>C</sub> into the structure placed it in a position for interaction with the scissile phosphate and other biochemically predicted phosphate oxygens (as discussed above for *Tetrahymena* and *Azotarcus* introns). In comparison to the *Azotarcus* structures PDB: 1U6B and 1ZZN, the PDB: 1Y0Q structure is similar with respect to the placement of M<sub>A</sub> (M<sub>1</sub> in PDB: 1U6B and 1ZZN) and the PDB: 1Y0Q modeled M<sub>C</sub> (M<sub>2</sub> in PDB: 1U6B and 1ZZN). However, none of these crystal structures visualize a metal ion at the biochemically predicted position M<sub>B</sub>. The reference 32 authors agree with the statements of other researchers in this area that more work will be necessary to determine the position of metal ions in the ground and/or transition states of group I introns. It should be said before beginning discussion of the hammerhead ribozyme that there is much more agreement on structure–function details for the group I introns than will be found for the hammerhead ribozyme.

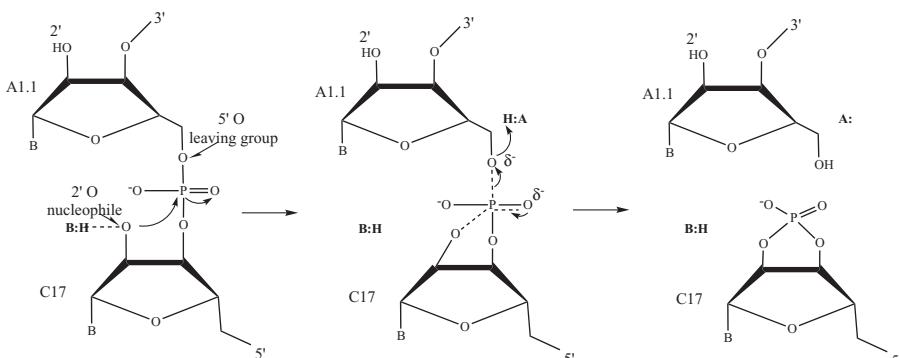
#### 6.2.4 The Hammerhead Ribozyme

The hammerhead ribozyme, so named because of its claw hammer shape, is a small catalytic motif conserved in plant viroids. Viroids are infectious agents

consisting of single-stranded RNA. They are smaller and simpler than viruses and lack the typical viral protein cover. Hammerhead ribozymes also occur in viroid-like satellite RNA (associated mostly with plant viruses) and in a few eukaryotic species.<sup>33</sup> The hammerhead ribozyme motif efficiently and site-specifically cleaves RNA through a transesterification reaction. The reaction is initiated by a nucleophilic attack by the 2'-oxygen of the nucleotide 5' to the cleavage site (in the hammerhead case, this is residue C17) on the cleavage site phosphorus, yielding a 2',3'-cyclic phosphate and 5'-hydroxyl product. (See Figure 6.9.)

The nucleophilic S<sub>N</sub>2 mechanism requires that reactants be in line with each other; that is, attack takes place along an approximate 180° axis. As will be seen in descriptions of X-ray crystallographic structures below, this arrangement does not appear to be the case in the solid-state structures, perhaps because additional conformation changes are needed from the state visualized by X-ray techniques. One or more cations may be required as catalysts for ribozyme cleavage reactions. One cation may ionize the 2'-hydroxyl group, while another (or the same) cation may stabilize the pentacoordinate transition state by coordinating to the pro-R phosphate oxygen. During the transesterification the chirality of the phosphate group is inverted in an S<sub>N</sub>2(P) mechanistic process. The phosphorus atom in the phosphate connector is pro-chiral since changing one of two oxygen atoms (substituting a sulfur for instance as in the P-S rescue experiment) gives rise to a chiral center. Changing one particular oxygen will give rise to an S configuration (pro-S<sub>p</sub> oxygen), while changing the other one would yield an R configuration (pro-R<sub>p</sub> oxygen) according to Cahn–Ingold–Prelog rules. (See Figure 6.2.)

The hammerhead motif has a conserved secondary structure consisting of a three-way helical junction. The helical elements may vary in base sequence among species but thirteen bases at the three-way helical junction are conserved and essential for ribozyme activity. X-ray structures to be discussed below define a domain organization based on the tertiary folding observed in



**Figure 6.9** Hammerhead mechanism producing a 2',3'-cyclic phosphate.

hammerhead ribozyme crystallographic studies. Domain 1 (or I) contains a U-turn motif (similar in tertiary fold to the uridine turn observed in tRNA structures) comprising the first four residues 3' to stem (helix) 1, these are conserved residues C<sub>3</sub>, U<sub>4</sub>, G<sub>5</sub> and A<sub>6</sub>. In addition, domain I includes the residues 3' and 5' to the cleavage site—that is, C<sub>17,0</sub> and X<sub>1,1</sub>. Domain 2 (or II) consists of parts of stems (helices) 2 and 3 that are coaxially stacked upon one another and is composed of eight conserved residues. It contains a connection between stems 2 and 3 that consists of the three noncanonical (non-Watson-Crick) pairs (G<sub>12</sub>–A<sub>9</sub>, A<sub>13</sub>–G<sub>8</sub> and A<sub>14</sub>–U<sub>7</sub>). The G-A/A-G mismatches are a highly conserved motif in many ribosomal RNA secondary structures. A three-way junction connects A<sub>6</sub>, U<sub>7</sub>, U<sub>16,1</sub>, and C<sub>17</sub>. The overall shape resembles a Y with stems 2 and 3 stacked, and stems 1 and 2 subtended by an acute angle. (See Figures 6.10, 6.11, and 6.12.)

Many researchers refer to stems 1, 2, and 3 using their Roman numeral equivalents—that is, stems I, II, and III. These motifs are also denoted as helices I, II, and III. It should be noted at the beginning of this hammerhead ribozyme discussion that structure–function relationships, the role of various nucleobases, metal ion participation in catalysis, and other features of the system have not been completely delineated and in some cases remain controversial. Globally, the hammerhead fold appears to be similar in both solution and solid-state studies. In solution, however, the central core of the hammerhead construct appears to be highly dynamic. This may account for different experimental results among the analytical techniques used in solution and certainly explains some distinct differences seen between solution and solid-state (X-ray crystallographic) structures.

Several hammerhead secondary structural motifs (constructs) are shown in Figures 6.10, 6.11, and 6.12. An arrow indicates the cleavage site at the C<sub>17,0</sub>–X<sub>1,1</sub> position in the substrate (X stands for nucleotide found at this position in the different constructs). The numbering system follows that of Hertel et al.<sup>34</sup> and has been applied to many different hammerhead constructs. The basic system starts numbering at the nucleotide 3' to the cleavage site—that is, X<sub>1,1</sub>. The decimal notation indicates that the nucleotide is part of a helix and is paired with another nucleotide. In Figure 6.11 for instance, G<sub>1,1</sub> of the substrate strand pairs with C<sub>2,1</sub> in the ribozyme strand. Seventeen nucleotides in the central core are numbered in a clockwise fashion: G<sub>1,1</sub>, C<sub>2,1</sub>, C<sub>3</sub>, U<sub>4</sub>, G<sub>5</sub>, A<sub>6</sub>, U<sub>7</sub>, G<sub>8</sub>, A<sub>9</sub>, G<sub>10,1</sub>, C<sub>11,1</sub>, G<sub>12</sub>, A<sub>13</sub>, A<sub>14</sub>, A<sub>15,1</sub>, U<sub>16,1</sub>, and C<sub>17</sub>. Nucleotides in loops are numbered by the helix they are contained in, for example: G<sub>L2,1</sub>, A<sub>L2,2</sub>, A<sub>L2,3</sub>, A<sub>L2,4</sub>, in Figure 6.11. In the hammerhead, cleavage occurs between nucleotides 17 and 1,1, resulting in phosphate 1,1 attached to ribose 17 as a 2',3' cyclic phosphodiester (see Figure 6.9). Mutations are denoted, for example, by G5A, where a guanine residue at position 5 is changed to an adenine. Two methods for illustrating the constructs are shown in Figures 6.11 and 6.12. The second emphasizes nucleotide numbering in the central core. Bimolecular hammerhead (ribozyme plus substrate) catalysis can be described by three principal steps: (1) association of the ribozyme to the substrate strand; (2) the chemical cleavage

Hammerhead construct RNA 6: 16 nt ribozyme, 25 nt substrate.  
Used for Scott group crystal structures and  
biochemical experiments by Hampel and Burke, references 33, 56

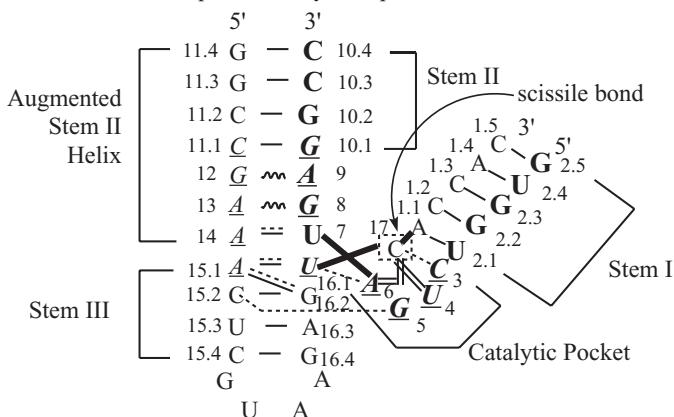


Figure key

Conserved residues underlined and italicized.

Enzyme residues in bold and larger font size.

Substrate residues in plain text and smaller font size.

Cleavage site nucleotide, C17, in a dashed box.

Scissile C17-A1.1 phosphate bond marked with arrow.

----- single hydrogen bonds between bases and backbone riboses

===== single hydrogen bonds between these bases

===== aromatic stabilization interactions between C17 and uridine turn of catalytic pocket

~~~~~ noncanonical bases

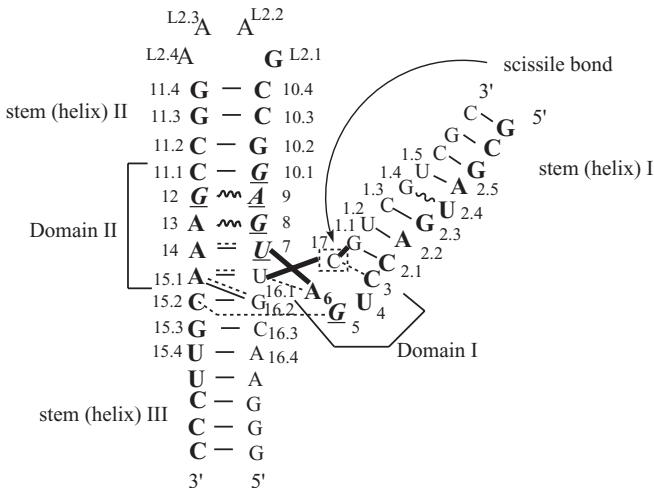
**Figure 6.10** Secondary structure of the RNA 6 hammerhead ribozyme as found in the X-ray crystallographic structures PDB: 1HMH, 1MME, 299D, 300D, 301D, 359D, 379D, and 488D. A modified RNA6 was used for crystal structures PDB: 1NYI, and 1Q29.

step; and (3) the dissociation of the ribozyme products. The cleavage step is the same for all hammerhead enzymes, and its mechanism is illustrated in Figure 6.9. The catalytic rate of hammerhead transesterification is  $10^6$ -fold that of the non-enzymatic process.

Data on the X-ray crystallographic structures discussed in Section 6.2.4 are collected in Table 6.3.

McKay and co-workers determined the X-ray crystallographic structure of a hammerhead RNA–DNA ribozyme complex at 2.6-Å resolution in 1994, deposited in the Protein Databank (PDB) as 1HMH.<sup>14</sup> The substrate DNA strand ensured that the complex would remain in the ground state since the DNA 2'-deoxy position could not undergo the cyclization/cleavage reaction.

Hammerhead ribozyme HH16, 38 nt ribozyme, 17 nt substrate.  
Used in biochemical experiments. Two representational methods shown.



#### Figure Key

Conserved residues underlined italicized.

..... single hydrogen bonds between bases  
and backbone riboses

Enzyme residues in bold and larger font size.

----- single hydrogen bonds between these bases

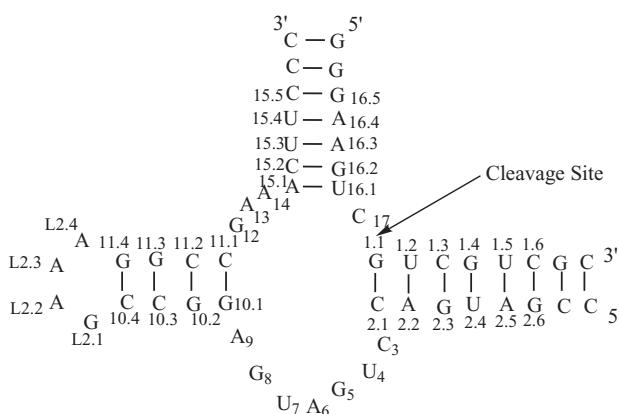
Substrate residues in plain text and smaller font size.

~~~~~ noncanonical bases

Cleavage site nucleotide, C17, in a dashed box.

Scissile C17-G1.1 phosphate bond marked

with arrow.

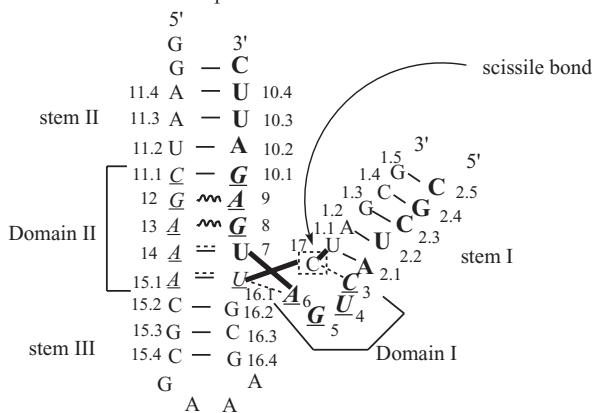


**Figure 6.11** Secondary structure of the HH16 hammerhead ribozyme used for biochemical, spectroscopic, and kinetic studies.

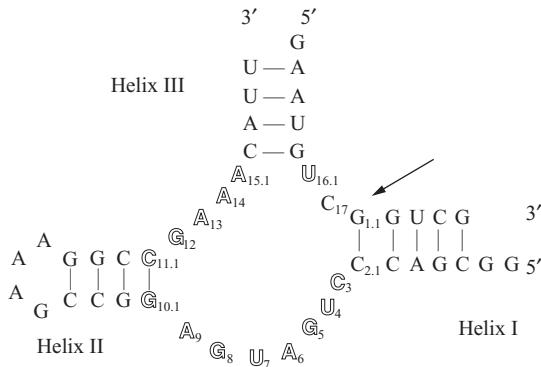
The structure showed that the three base-paired stems I, II, and III form type A-DNA helices and that the core contained the two conserved structural domains described in the previous paragraph. A single  $\text{Mg}^{2+}$  ion was bound in close proximity to the pro- $R_p$  oxygen of  $\text{A}_9$ 's phosphate and the  $\text{N}_7$  position of  $\text{G}_{10.1}$ . The phosphodiester backbone of the DNA inhibitor strand was splayed

**A**

Hammerhead Ribozyme HH $\alpha$ 1, 17 nt ribozyme, 27 nt substrate  
Used in biochemical experiments.

**B**

HH8 hammerhead ribozyme used in biochemical experiments.



**Figure 6.12** Secondary structure of hammerhead ribozyme constructs used for biochemical, spectroscopic, and kinetic studies. (A) the HH $\alpha$ 1 ribozyme and (B) the HH8 ribozyme. (Reprinted with permission from Figure 1B of reference 47.)

out at the phosphate 5' to the cleavage site, but this site was at least 20 Å from the A<sub>9</sub>–G<sub>10.1</sub> Me<sup>2+</sup> binding site. The structure indicated that a destabilized substrate strand might twist to put the cleavage site into position for cleavage of the scissile phosphate bond.

In the following year, Scott and co-workers solved the X-ray crystallographic structure of an all-RNA hammerhead ribozyme with a 2'-OCH<sub>3</sub> group incorporated at the active site cytosine (C<sub>17</sub>) to prevent cleavage (PDB: 1MME).<sup>35</sup> This structure differed from that of 1HMH in several important ways: (1) it was an all-RNA ribozyme rather than an RNA–DNA hybrid; (2) the connectivity of the ribozyme backbone strands was different (for instance

**TABLE 6.3 Description of Hammerhead Crystal Structures**

| PDB ID | Description   | Metal Ions (Locations)   | Resolution/<br>pH | Reference/<br>Year |
|--------|---|--|-------------------|--------------------|
| 1HMH   | RNA–DNA construct   |  | 2.60              | 14/1994            |
| 1MME   | RNA–RNA 2'-OCH <sub>3</sub> at cleavage site                  | Mg <sup>2+</sup> (near G <sub>2,2</sub> , G <sub>2,3</sub> , near cleavage site–CUGA loop, bonds to <i>pro-R</i> -oxygen A <sub>9</sub> , N <sub>7</sub> of G <sub>10,1</sub> )  | 3.10              | 35/1995            |
| 299D   | Freeze-trapped RNA–RNA  | No metals  | 3.00/6.0          | 39/1996            |
| 300D   | Freeze-trapped RNA–RNA  | 6 Mn <sup>2+</sup> (Mn <sup>2+</sup> – <i>pro-R</i> -oxygen A <sub>9</sub> = 2.3 Å, Mn <sup>2+</sup> –N <sub>7</sub> G <sub>10,1</sub> = 2.2 Å, Mn <sup>2+</sup> –N <sub>7</sub> G <sub>5,0</sub> = 3.1 Å, Mn <sup>2+</sup> –O <sub>6</sub> G <sub>5,0</sub> = 4.3 Å)  | 3.00/5.0          | 39/1996            |
| 301D   | Freeze-trapped RNA–RNA  | 5 Mg <sup>2+</sup> (Mg <sup>2+</sup> – <i>pro-R<sub>p</sub></i> -oxygen A <sub>1,1</sub> = 2.4 Å, Mg <sup>2+</sup> –N <sub>2</sub> G <sub>5,0</sub> = 3.8 Å, Mg <sup>2+</sup> –N <sub>1</sub> G <sub>5,0</sub> = 3.9 Å, Mg <sup>2+</sup> within 7–11 Å of A <sub>9,0</sub> –G <sub>10,1</sub> )  | 3.00/8.5          | 39/1996            |
| 359D   | RNA–RNA   | 5 Tb <sup>3+</sup> (Tb <sup>3+</sup> –N <sub>2</sub> G <sub>5,0</sub> = 3.9 Å, Tb <sup>3+</sup> –N <sub>1</sub> G <sub>5,0</sub> = 3.8 Å, Tb <sup>3+</sup> –O <sub>6</sub> G <sub>5,0</sub> = 4.7 Å)   | 2.90/6.0          | 43/1998            |
| 379D   | RNA–RNA <i>talo-</i> 5'-C–CH <sub>3</sub> at A <sub>1,1</sub> | 8 Co <sup>2+</sup> (Co <sup>2+</sup> – <i>pro-R<sub>p</sub></i> -oxygen A <sub>9</sub> = 2.5 Å, Co <sup>2+</sup> –N <sub>7</sub> G <sub>10,1</sub> = 2.0 Å, Co <sup>2+</sup> –N <sub>7</sub> G <sub>5,0</sub> = 2.2 Å, Co <sup>2+</sup> –N <sub>7</sub> A <sub>1,1</sub> = 2.6 Å Co <sup>2+</sup> –O <sub>6</sub> G <sub>16,2</sub> = 2.1 Å, Co <sup>2+</sup> –N <sub>7</sub> G <sub>16,4</sub> = 2.6 Å) | 3.1/8.5           | 44/1998            |
| 488D   | Lattice trapped RNA–RNA ribozyme–product                      | 8 Cd <sup>2+</sup> (Cd <sup>2+</sup> – <i>pro-R</i> -oxygen A <sub>9</sub> = 2.2 Å, Cd <sup>2+</sup> –N <sub>7</sub> G <sub>10,1</sub> = 2.0 Å, Cd <sup>2+</sup> –N <sub>7</sub> G <sub>5,0</sub> = 2.0 Å, Cd <sup>2+</sup> ions near A <sub>6,0</sub> , G <sub>2,3</sub> , U <sub>2,4</sub> , G <sub>2,5</sub> )  | 3.1/8.5           | 52/2000            |
| 1NYI   | RNA–RNA, base pair tether stem I to stem II                   | 4 Co <sup>2+</sup> (Co <sup>2+</sup> –N <sub>2</sub> G <sub>5,0</sub> = 3.5 Å, Co <sup>2+</sup> –N <sub>1</sub> G <sub>5,0</sub> = 3.6 Å, Co <sup>2+</sup> –O <sub>6</sub> G <sub>5,0</sub> = 4.7 Å, Co <sup>2+</sup> –O <sub>2'</sub> U <sub>16,1</sub> = 1.8 Å, Co <sup>2+</sup> – <i>pro-R<sub>p</sub></i> oxygen C <sub>17,0</sub> = 3.9 Å)  | 2.85/6.0          | 58/2003            |
| 1Q29   | RNA–RNA, base pair tether stem I to stem II                   | 5 Co <sup>2+</sup> (Co <sup>2+</sup> – <i>pro-R</i> oxygen A <sub>9</sub> = 2.3 Å, Co <sup>2+</sup> –N <sub>7</sub> G <sub>10,1</sub> = 2.3 Å, Co <sup>2+</sup> –N <sub>7</sub> G <sub>5,0</sub> = 2.6 Å, Co <sup>2+</sup> –O <sub>6</sub> G <sub>5,0</sub> = 4.3 Å, Co <sup>2+</sup> –N <sub>7</sub> G <sub>16,2</sub> = 2.0 Å)   | 2.99/8.5          | 58/2003            |

coordination of a metal cation to the A<sub>9</sub>-G<sub>10.1</sub> Me<sup>2+</sup> binding site appeared to be through the pro-S<sub>p</sub> oxygen rather than the pro-R<sub>p</sub> oxygen of A<sub>9</sub>'s phosphate); (3) the crystals were grown in a much lower concentration of salt; and (4) the crystal packing scheme was very different. In spite of these differences, the overall structure appeared similar. Magnesium ions were soaked into the PDB: 1MME hammerhead crystal allowing the identification of five Mg<sup>2+</sup>-binding sites, including one near the ribozyme–substrate complex catalytic pocket. In helix I a Mg<sup>2+</sup> interacts with N7 and O6 of both G2.2 and G2.3. Another Mg<sup>2+</sup> is near the cleavage site, in the CUGA loop, where it interacts with several bases, including A<sub>1.1</sub>, C<sub>3</sub>, U<sub>4</sub>, U<sub>7</sub>, G<sub>8</sub>, and C<sub>17</sub>. This magnesium ion, close to the cleavage site, may have a catalytic role. Another two sites were observed in domain II. At one of these, the Mg<sup>2+</sup> interacts primarily with the pro-R-oxygen of the A<sub>9</sub> phosphate and N7 of G<sub>10.1</sub>. Neither X-ray structure (PDB: 1HMH nor 1MME) showed the proper inline positioning of the scissile phosphate group for the S<sub>N</sub>2 cleavage mechanism to take place.

Other researchers attempted to clarify the differences reported between the 1HMH and 1MME hammerhead ribozyme structures. Herschlag's group first traced the history of biochemical data accumulated for the hammerhead ribozyme and then continued with their own experiments. It had been shown previously in phosphorothioate interference experiments that the pro-R<sub>p</sub> oxygen of A<sub>9</sub>'s phosphate was important for hammerhead substrate cleavage.<sup>36</sup> Nucleotide substitution experiments had previously implicated G<sub>10.1</sub> as important for ribozyme catalytic activity.<sup>37</sup> In their 1997 *Journal of Biological Chemistry* article, the Herschlag group substituted sulfur into the pro-R<sub>p</sub> and pro-S<sub>p</sub> phosphate oxygen positions at A<sub>9</sub> and then followed catalytic activity in the absence and presence of the thiophilic, soft Cd<sup>2+</sup> ion.<sup>38</sup> The experiments were conducted on the HH $\alpha$ 1 (Figure 6.12) and HH16 (Figure 6.11) hammerhead ribozyme constructs, constructs that differ in some residue positions but that have the same set of conserved residues as those used to determine the crystal structures PDB: 1HMH and 1MME (the RNA 6 construct shown in Figure 6.10). Both the HH $\alpha$ 1 and HH16 ribozymes showed loss of catalytic activity in the A<sub>9</sub> pro-R<sub>p</sub>-phosphorothioate in the presence of Mg<sup>2+</sup> that could be rescued by addition of 100  $\mu$ M Cd<sup>2+</sup>. These results indicated that a metal ion coordinated at the pro-R<sub>p</sub> oxygen of position 9 is critical for efficient catalysis. By studying the concentration dependence of Cd<sup>2+</sup> with respect to observed rates of the catalytic reaction, the researchers could calculate an apparent dissociation constant of  $K_d^{\text{Cd}} = 25 \mu\text{M}$  in the ground state. Using transition state theory and the equation

$$K_d^{\text{Cd}\ddagger} = K_d^{\text{Cd}} \left( \frac{k_2}{k_2^{\text{Cd}}} \right) \quad (6.1)$$

the researchers were able to calculate an apparent dissociation constant of  $K_d^{\text{Cd}\ddagger} = 2.5 \text{nM}$  in the transition state. This much stronger Cd<sup>2+</sup> binding in the transition state was interpreted to indicate at least one additional ligand

attachment. If the added ligands arose from additional interactions with RNA nucleobases, riboses, or phosphate oxygens, this might signal that the metal ion may have participated in a conformation change preceding hammerhead substrate cleavage.

In 1996, the W. G. Scott group published three hammerhead ribozyme structures describing them in a *Science* magazine article (PDB: 299D, 300D, and 301D).<sup>39</sup> While both previously published structures described modified hammerheads, these structural studies were carried out on the unmodified ribozyme–substrate complex. The researchers used time-resolved crystallography to observe structural changes that occur during the ribozyme-catalyzed reaction. (See Section 3.7.2.3.) The reaction is arrested chemically or physically by adjusting the pH, flash-freezing the crystal, or a combination of both methods. In the PDB: 299D structure (containing no divalent metal ions) the crystal was formed in a solution held at pH 6, flash-frozen, and then maintained at 100 K. At pH 6, the cleavage reaction takes place at a slower rate in solution and at a yet slower rate in the crystalline state. The PDB: 299D structure is quite similar to that of the 2'-O-methyl C17 modified hammerhead described above (PDB: 1MME), especially for residues in or interacting with the active site (see Figure 3B of reference 39). The PDB: 300D structure was obtained at pH 5 in the presence of Mn<sup>2+</sup> ions. Six manganese ions appear in the crystal structure; however, none are close enough to oxygen atoms at cleavage site (C<sub>17.0</sub>–A<sub>1.1</sub>) for bonding interactions. In the A<sub>9.0</sub>–G<sub>10.1</sub> region, Mn<sup>2+</sup>–O (Mn<sup>2+</sup>–O distance to a non-bridging phosphate oxygen of A<sub>9.0</sub> = 2.31 Å) and Mn<sup>2+</sup>–N (Mn<sup>2+</sup>–N distance to the N<sub>7</sub> nitrogen atom of G<sub>10.1</sub> = 2.24 Å) interactions are found. Selected bonding distances for these structures are included in Table 6.4.

The overall structure resembles that of PDB: 299D (no divalent metal ions present) and PDB: 1MME (modified C<sub>17.0</sub>). The scissile phosphodiester bond (A<sub>1.1</sub> 5'O–A<sub>1.1</sub> P) and the adjacent ribose (that of C<sub>17.0</sub>) are still in the approximate A-form helical conformation. This places the C<sub>17.0</sub> 2'-OH within 3.5 Å of the A<sub>1.1</sub> phosphate phosphorus atom but on the wrong side for an S<sub>N</sub>2 nucleophilic attack to occur. If crystals of the Mn<sup>2+</sup>-modified ribozyme–substrate complex (PDB: 300D) are dissolved after being soaked in 100 mM MnSO<sub>4</sub> at pH 7 for 30 minutes, a denaturing polyacrylamide gel assay shows residual intact 25-nt substrate as well as 20-nt cleaved product, indicating that some cleavage has taken place at pH 7.

The third crystal (PDB: 301D) was formed by freeze-trapping the RNA four minutes after Mg<sup>2+</sup> was added at pH 8.5. At this pH the hammerhead RNA cleaves itself in the crystalline state, with the cleavage reaction reaching 90% completion within 15 minutes. This structure contained five co-crystallized magnesium ions, and it revealed significant conformational change when compared to both 299D and 300D structures. The change is most evident at the scissile phosphodiester bond located between C<sub>17.0</sub> and A<sub>1.1</sub>. This phosphodiester bond moves approximately 2.9 Å into the catalytic cleft when compared to the same region for the other crystallographic structures. Its change

TABLE 6.4 Selected Bond Distances for X-Ray Crystallographic Structures Discussed in Reference 39 (PDB: 299D, 300D, and 301D)

| Region               |                     | Selected Bond Distances PDB: 299D (Å) pH = 6.0    |  |   |   |  |   |
|----------------------|---------------------|---|--|---|---|--|---|
| A9-G10.1<br>region   | Mn <sup>2+</sup> -O | Mn <sup>2+</sup> -N                               | N <sub>7</sub> G <sub>10.1</sub> -O <sub>1P</sub> A <sub>9</sub> | P G <sub>10.1</sub> -O <sub>1</sub> G <sub>10.1</sub>                                       | P G <sub>10.1</sub> -O <sub>2</sub> G <sub>10.1</sub> | P G <sub>10.1</sub> -O <sub>3'</sub> A <sub>9</sub>  | P G <sub>10.1</sub> -O <sub>5'</sub><br>G <sub>10.1</sub> |
|                      |                     | N(271) <sup>a-</sup><br>O(236)                    | P(257)-O(258)<br>1.49  | P(257)-O(259)<br>1.49   | P(257)-O(243)<br>1.60                                 | P(257)-O(260)<br>1.60                                |   |
|                      |                     | pro-R <sub>p</sub> (O)<br>A <sub>9.0</sub> 3.94   | P A <sub>1.1</sub> -O <sub>1</sub> A <sub>1.1</sub>              | P A <sub>1.1</sub> -O <sub>2</sub><br>A <sub>1.1</sub> -O                                   | P A <sub>1.1</sub> -O <sub>3'</sub> C <sub>7.0</sub>  | P A <sub>1.1</sub> -O <sub>5'</sub> A <sub>1.1</sub> |   |
|                      |                     |   | P(771)-O(772)<br>1.49  | P(771)-O(773)<br>1.49   | P(771)-O(759)<br>1.60                                 | P(771)-O(774)<br>1.59                                |   |
| Region               |                     | Selected Bond Distances PDB: 300D (Å) pH = 5.0    |  |   |   |  |   |
| C17.0-A1.1<br>region | Mn <sup>2+</sup> -O | Mn <sup>2+</sup> -N                               | N <sub>7</sub> G <sub>10.1</sub> -O <sub>1P</sub> A <sub>9</sub> | P G <sub>10.1</sub> -O <sub>1</sub> G <sub>10.1</sub>                                       | P G <sub>10.1</sub> -O <sub>2</sub> G <sub>10.1</sub> | P G <sub>10.1</sub> -O <sub>3'</sub> A <sub>9</sub>  | P G <sub>10.1</sub> -O <sub>5'</sub><br>G <sub>10.1</sub> |
|                      |                     | Mn(875)-<br>O <sub>1P</sub> A <sub>9.0</sub> 2.31 | Mn(875)-<br>N(270) N <sub>7</sub><br>G <sub>10.1</sub> 2.24      | N(270) N <sub>7</sub><br>G <sub>10.1</sub> -O(235)<br>O <sub>1P</sub> A <sub>9.0</sub> 3.32 | P(256)-O(257)<br>1.50                                 | P(256)-O(258)<br>1.49                                | P(256)-O(242)<br>1.60                                     |
|                      |                     | Mn <sup>2+</sup> -N                               | P A <sub>1.1</sub> -O <sub>1</sub> A <sub>1.1</sub>              | P A <sub>1.1</sub> -O <sub>2</sub><br>A <sub>1.1</sub> -O                                   | P A <sub>1.1</sub> -O <sub>3'</sub> C <sub>7.0</sub>  | P A <sub>1.1</sub> -O <sub>5'</sub> A <sub>1.1</sub> |   |
|                      |                     |   | P(770)-O(771)<br>1.49  | P(770)-O(772)<br>1.49   | P(770)-O(758)<br>1.60                                 | P(770)-O(773)<br>1.58                                |   |

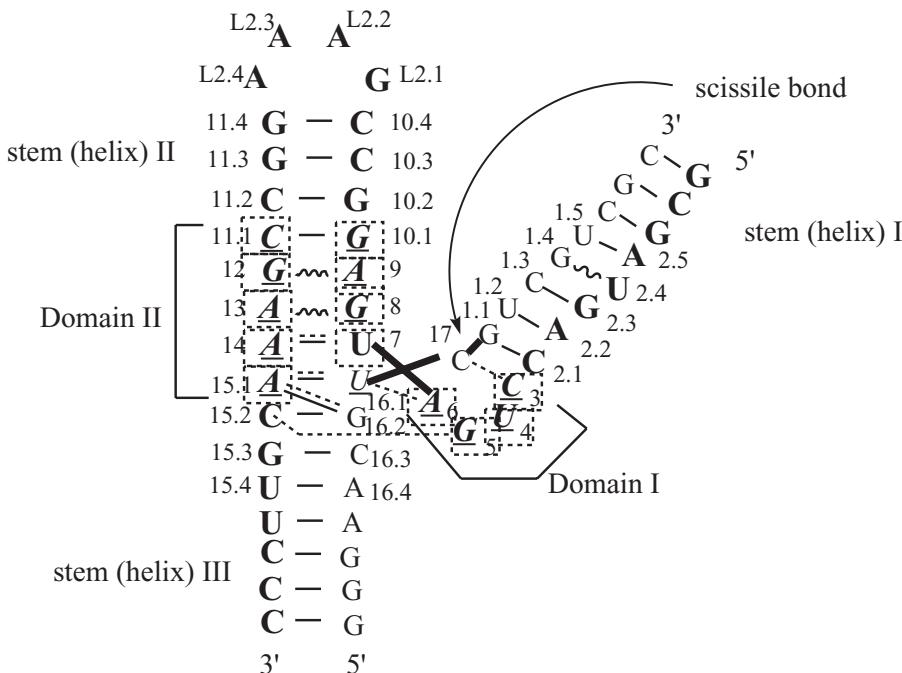
| Selected Bond Distances PDB: 301D (Å) pH = 8.5 |  |   |  |  |  |   |
|--|--|---|--|--|--|---|
|  | Mg <sup>2+</sup> -O  | Mg <sup>2+</sup> -N   | N <sub>7</sub> G <sub>10,1</sub> -O <sub>1P</sub> A <sub>9</sub>             | P G <sub>10,1</sub> -O <sub>1</sub> G <sub>10,1</sub>                        | P G <sub>10,1</sub> -O <sub>3'</sub> A <sub>9</sub>                            | P G <sub>10,1</sub> -O <sub>5'</sub>  |
| A9-G10,1<br>region                             | Mg(877) (Site<br>2)-O(236) O <sub>1P</sub><br>A <sub>9,0</sub> 7.08  | Mg(877) (Site<br>2)-N(271)<br>N <sub>7</sub> G <sub>10,1</sub> 8.27                       | N(271)-O(236)<br>O <sub>1P</sub> A <sub>9,0</sub> 4.42                       | P(257)-O(258)<br>1.50  | P(257)-O(259)<br>1.49  | P(257)-O(243)<br>1.61   |
| C17,0-A1,1<br>region<br>(cleavage<br>site)     | Mg <sup>2+</sup> -O<br>Mg(880) (Site<br>6)-O(773) O <sub>2B</sub><br>pro-R <sub>p</sub> , A <sub>1,1</sub><br>2.43 | Mg <sup>2+</sup> -N<br>Mg(880) (Site<br>6)-N(785)<br>N <sub>7</sub> A <sub>1,1</sub> 4.48 | P A <sub>1,1</sub> -O <sub>1</sub> A <sub>1,1</sub><br>P(771)-O(772)<br>1.50 | P A <sub>1,1</sub> -O <sub>2</sub> A <sub>1,1</sub><br>P(771)-O(773)<br>1.49 | P A <sub>1,1</sub> -O <sub>3'</sub> C <sub>17,0</sub><br>P(771)-O(759)<br>1.63 | P A <sub>1,1</sub> -O <sub>5'</sub> A <sub>1,1</sub><br>P(771)-O(774)<br>1.59 |
| Selected Bond Distances (Å)                    |  |   |  |  |  |   |
| PDB: Accession<br>number                       | Me-Me  | P-P   |  |  |  |   |
| 299D   |  |   | P(771)-P(235)<br>20.6  |  |  |   |
| 300D   | Mn(876) <sup>a</sup> -<br>Mn(875) 13.7   |   | P(234)-P(770)<br>20.7  |  |  |   |
| 301D   | Mg(877)-<br>Mg(878) 10.5   |   | P(771)-P(235)<br>19.6  |  |  |   |
|  | Mg(876)-<br>Mg(880)<br>4.46 Å  |   |  |  |  |   |

<sup>a</sup>Numbers in parentheses are atom numbers from the Protein Data Bank (PDB) files.

in position is accompanied by other nucleobase and ribose movements with concomitant changes in hydrogen bonding patterns in the region. Significant new hydrogen bonds stabilizing the conformational change included one that formed between the furanose oxygen ( $O_4'$ ) of  $C_{17.0}$  and the  $2O'$  of  $U_{16.1}$ 's 2'-hydroxyl group. Importantly, a magnesium ion (site 6) appears to be approximately 2.4 Å from the pro-R<sub>p</sub> phosphate oxygen of residue A<sub>1.1</sub>. Another magnesium ion (site 4) appears near to G<sub>5.0</sub>; this site is known to be required for hammerhead catalytic activity. A third magnesium ion (site 2) appears near the A<sub>9.0</sub>-G<sub>10.1</sub> region but does not interact in the manner seen for the manganese ion in the PDB: 300D structure. The site 6 magnesium ion may be inducing a conformational change leading to transition state in which in-line S<sub>N</sub>2 attack could occur. Although it is not possible to visualize the coordination sphere of the magnesium ion at the resolution of the crystal structures (3.00 Å), the reference 39 authors hypothesized that a coordinated hydroxide could provide the catalyst for the base-catalyzed step of the cleavage reaction. They also favored a mechanism in which the Mg<sup>2+</sup>-phosphate complex rotated relative to the rest of the RNA molecule to reach a transition state that would bring the O $2'$  of  $C_{17.0}$ 's 2'-OH group in line for nucleophilic attack on the phosphate phosphorus atom. (O' indicate ribose oxygen atoms). Herschlag's group, reference 38, commented that the observed PDB: 301D structure might represent an early intermediate on the reaction pathway, but differing considerably from the transition state structure, or might be an "off-pathway" structure. If crystals of the Mg<sup>2+</sup>-modified ribozyme-substrate complex (PDB: 301D) are dissolved after being soaked in 100 mM MgSO<sub>4</sub> at pH 8.5 for 15 minutes, a denaturing polyacrylamide gel assay shows very little residual intact 25-nt substrate and a substantial amount of 20-nt cleaved product (see Figure 2 of reference 39).

In 1997, Scott and co-workers provided evidence gathered in molecular dynamic simulations that a hydroxide ion bridge could exist between two magnesium ions near the hammerhead cleavage site.<sup>40</sup> Starting with data from the X-ray crystallographic structure PDB: 301D described in the previous paragraphs, the  $\mu$ -bridging hydroxide ion was located, in the simulations, between site 6 (that interacts with a nonbridging phosphate oxygen at the cleavage site) and site 1 magnesium ions. Using numbering provided with the Protein Data Bank file for PDB: 301D, one finds that the site 1–site 6 distance—Mg(876)—Mg(880)—is equal to 4.46 Å. Further simulations showed that flipping the ribose at the cleavable phosphate site from the C3'-endo to the 2'-endo conformation brings the relevant 2'-OH nucleophilic group in proximity to the attacked phosphate phosphorus atom and the  $\mu$ -bridging OH<sup>-</sup> ion.

Herschlag's group continued its study of structure–function relationships in the hammerhead ribozyme using a "base-rescue" biochemical method. This method substitutes other atoms or molecules for bases at critical catalytic or structural positions and tests whether catalytic activity is lost. If so, the RNA bases (U, A, G, C) or a modified base (for instance, deazaguanine or 2-aminopurine substituted for guanine) is added to the solution to ascertain

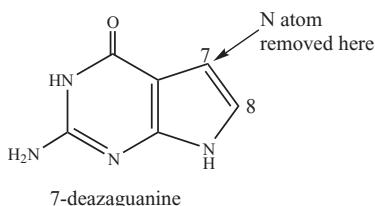


**Figure 6.13** Structure of the hammerhead ribozyme HH16. Dashed boxes surround the 13 substituted positions.

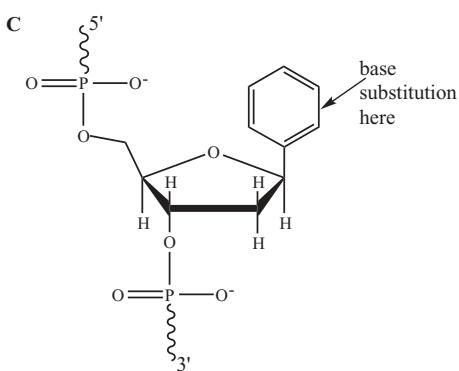
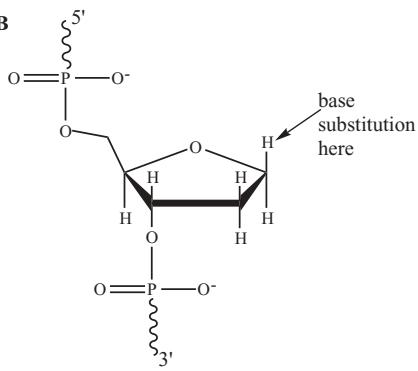
whether catalytic activity can be recovered (rescued). In 1996, the group published a paper in *PNAS* describing the results when abasic nucleotides were substituted at 13 critical positions on the hammerhead HH16 construct.<sup>41</sup> Figure 6.13 shows the 13 substituted positions, each marked with a dashed outline. Figure 6.14 shows the structure of 7-deazaguanine, a reduced abasic nucleotide and a phenyl-substituted nucleotide.

For the 13 hammerhead HH16 mutants prepared by the reference 41 researchers, cleavage activity was significantly reduced in all cases. In four cases, at positions C<sub>3</sub>, A<sub>9</sub>, G<sub>10.1</sub>, and A<sub>13</sub>, addition of the correct base (U, A, G, or C) rescued the activity of the ribozyme. A9X, for instance, is a ribozyme where the adenine base at position 9 has been removed to generate an abasic nucleotide (X = a nucleotide in which H substitutes for the removed base—see Figure 6.14). This altered ribozyme exhibited a  $4 \times 10^{-4}$  times lower catalytic rate than that for the wild-type ribozyme ( $k_{\text{rel}} = k_2^{\text{abasic}} / k_2^{\text{wt}} = 0.0004$ ). When exogenous adenine (3 mM) was added to the reaction mixture, catalysis was rescued:  $k_2^{\text{obsA}} = 0.12 \text{ min}^{-1}$  (or 300-fold) compared to  $1.1 \pm 0.2 \text{ min}^{-1}$  for wild-type ribozyme. For another variant, the G<sub>10.1</sub>X, where X represents the abasic nucleotide mutant, the rescue was to wild-type behavior (from  $0.04 \text{ min}^{-1}$  to  $1.2 \text{ min}^{-1}$ ). The ribozyme positions C3X and A13X also exhibited rescue behavior. The reference 41 authors put forth the following hypotheses: (1) Removal

A



B



**Figure 6.14** Structure of 7-deazaguanine (A), a reduced abasic nucleotide (B), and a phenyl-substituted nucleotide (C).

of the base forming the mutant causes structural deformation in the ribozyme; (2) the perturbed structure may rearrange to make its site accessible to an exogenous (added) base; and (3) binding of the rescuing base may restore the active structure partially or completely and may even enhance catalysis. The researchers found that the G<sub>10.1</sub> N<sub>7</sub> position is important in binding exogenous guanine. Importantly, this is also the site of metal binding in X-ray structures. The following factors for successful rescue were found to be important: (1) the number of interactions formed with neighboring bases by the rescuing base; (2) other local interactions (such as Me<sup>2+</sup> coordination at G<sub>10.1</sub> N<sub>7</sub>); (3) base stacking interactions [purines (A, G) stack better than pyrimidines (U, C)]; (4) structural rigidity at the rescue site versus the ability to adopt alternative structures; and (5) ability to align the RNA into a structural/functional mode through formation of the base–sugar glycosidic bond. The researchers believe that the rescued bases they studied were not necessarily involved in chemical catalysis but rather in repair of conformational defects. They speculated that the technique of base rescue could contribute to drug development and design because RNA ribozymes are being pursued as therapeutic agents having the potential to cleave deleterious RNAs.

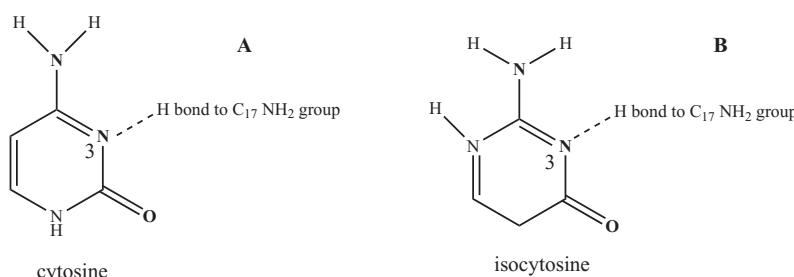
In a 1998 RNA paper, the Herschlag group further examined base rescue at four positions: C<sub>3</sub>, A<sub>9</sub>, G<sub>12</sub>, and A<sub>13</sub> probing transition state interactions at these hammerhead positions.<sup>42</sup> Their model for base rescue, illustrated in Figure 3 of reference 42, shows that removal of a base interferes with the ability of the hammerhead ribozyme core to achieve its catalytic structure but that addition and binding of an exogenous base can allow an active catalytic structure to be generated. This active structure may exhibit enhanced substrate cleavage over that of the wild-type ribozyme. An equation for calculating the observed cleavage rate is given by

$$k_2^{\text{obs}} = \left( \frac{k_2 + k_{\text{rescue}}[\text{Base}]}{1 + [\text{Base}]K_d} \right) \quad (6.2)$$

where  $k_2$  is the cleavage rate constant in the absence of the rescuing base,  $K_d$  is the dissociation constant of the base, and  $k_{\text{rescue}}$  is the apparent second-order rate constant for the rescue =  $k'_2/K_d$ .

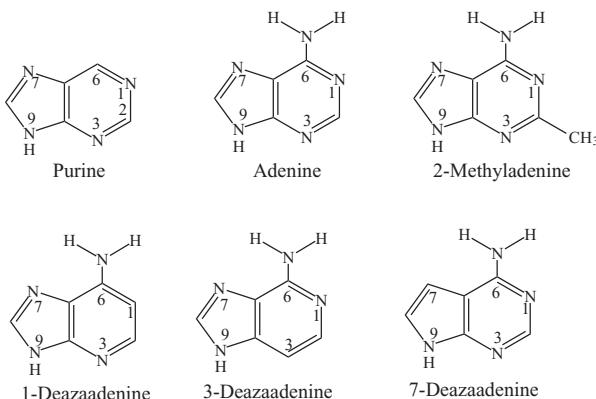
In the model scheme,  $k'_2$  is the cleavage rate constant for the ternary ribozyme–substrate–rescuing base complex. It was found that  $k_{\text{rescue}}$  depended on all the interactions happening in the transition state when the added rescuing base forms the ternary ribozyme–substrate–rescuing base complex from the altered ribozyme–substrate precursor. Many pyrimidine and purine bases were used to test the involvement of specific base functional groups in achieving catalysis.

The cytosine at position 3 was removed to form C3X and the catalytic rate decreased 10<sup>5</sup>-fold. Addition of 50 mM cytosine provides a rescue of 20-fold for the C3X variant while addition of cytosine to the wild-type ribozyme–substrate complex does not affect the catalytic rate. Purine bases and uracil did not provide rescue for the C3X variant whereas isocytosine (see Figure 6.15) does provide efficient rescue. The reference 42 authors concluded that the same hydrogen bonding groups on the pairing face are necessary for rescue because isocytosine and cytosine provide the same needed atoms in the correct alignment (see bold atoms in Figure 6.15).



Bold positions are important for transition state hydrogen bonding in hammerhead ribozymes.

**Figure 6.15** Important hydrogen-bonding positions for (A) cytosine and (B) isocytosine in hammerhead ribozymes.



**Figure 6.16** Bases used in substitution and rescue experiments.

Additional interactions and rearrangements in the transition state with other rescuing bases may take place because it is known from crystal structures that substrate atoms are not in line for nucleophilic attack in the hammerhead ribozyme (at least not in the published crystal structures). Also, a metal ion located  $\sim 20\text{ \AA}$  away from the catalytic site was shown to be crucial for catalysis. This same metal ion appeared likely to take on an additional ligand in the transition state, suggesting that conformational changes had to take place before catalysis.<sup>38</sup>

Herschlag and co-workers continued to study the rescue of A9X using 28 different bases. Only purine bases effected rescue.  $K_{\text{rescue}}$  was  $\sim 30$ -fold higher upon adding adenine than for purine itself (see Figure 6.16), indicating that the 6-amino group of the base is important for rescue. Other experiments indicated that adenine positions 2 and 3 are not important for transition state interactions. Modification of adenine to 1-deazaadenine and 3-deazaadenine indicated that these positions do not engage in critical transition state interactions either.

Changing A9X to A9 $\Phi$  (where  $\Phi$  is a phenyl group, see Figure 6.14) decreased catalysis only a modest amount, indicating that stacking interactions could be important for the A<sub>9</sub> position. Surprisingly, exogenous guanine rescued catalysis better than exogenous adenine, indicating that a hydrogen-bonding pattern is more important than the individual base itself. The authors indicate the possible hydrogen-bonding pattern for guanine to its pairing partner G<sub>12</sub> in Figure 8 of reference 42. Overall, the A<sub>9</sub> position appeared to serve a positioning or structural role in hammerhead catalysis. Similar behavior was seen for the A13X variant hammerhead: pyrimidines do not cause rescue but exogenous guanine does not cause rescue either. One purine, 2-methyladenine (Figure 6.16) does afford modest rescue, and the authors believe that this may be due to fortuitous hydrophobic interactions within the hammerhead central core. In the crystal structure (PDB: 1MME) the base G<sub>12</sub>

seems to be important because it is held in position by four hydrogen bonds to G<sub>8</sub> and A<sub>9</sub> and stacks onto the G<sub>10.1</sub>–C<sub>11.1</sub> pair. However, the G12X variant is not rescued by addition of sparingly soluble guanine and only modestly rescued by a more soluble analog, 7-deazaguanine (see Figure 6.14A).

Authors of reference 42 presented the following conclusions from their study of hammerhead bases using base rescue techniques: (1) Each of the groups on the base-pairing face of C<sub>3</sub> forms interactions important for catalysis; (2) comparison to the ground-state X-ray crystallographic structure (PDB: 1MME) indicates that these interactions develop only in the transition state; (3) bases at A<sub>9</sub>, A<sub>13</sub>, and A<sub>14</sub> provide stacking interactions stabilizing both the ground state (solid-state crystal structure) and transition state (solution base rescue experiments); and (4) enhanced base rescue at position A13 by a methylated base (2-methyladenine) may indicate the importance of hydrophobic interactions for transition state stabilization. The reference 42 authors believe that the base rescue technique can provide efficient, inexpensive testing of structure–function relationships in hammerheads because only one RNA derivative per position needs to be prepared, and commercially available bases and base derivatives can then be tested for rescue. They found in the two studies described here (references 41 and 42) that only five of the 14 tested abasic hammerhead variants could be rescued by exogenous base addition. Additional limitations of the base rescue technique include the difficulty in attaining saturation with poorly soluble exogenous bases and the possibility of observing new non-wild-type binding modes with added bases that would not be found for the base originally removed.

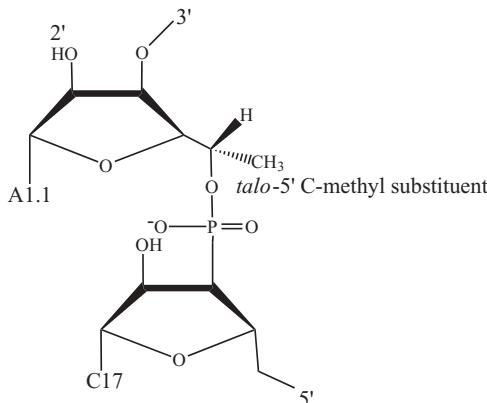
Scott and Uhlenbeck published a hammerhead X-ray crystallographic structure in 1998 (PDB: 359D)<sup>43</sup> that showed a terbium(III), Tb(III), ion competing with a particular Mg<sup>2+</sup> found to bind in the same location in the PDB: 301D crystal described above. The Tb(III) ion interacted with conserved residues guanosine, G<sub>5</sub>, and adenosine, A<sub>6</sub>, in the catalytic core in a similar manner to the magnesium ion in the PDB: 301D structure. The site is approximately 10 Å from the cleavage site in the hammerhead catalytic core, in the so-called uridine turn comprised of residues C<sub>3</sub>, U<sub>4</sub>, G<sub>5</sub>, and A<sub>6</sub>. The Tb(III) ion appeared to interact with the base-pairing face of G<sub>5</sub> and was not within binding distance of any phosphate residues. The closest contact is from N<sub>1</sub> of G<sub>5</sub> to the Tb(III) at 3.8 Å. Herschlag's experiments that removed the G<sub>5</sub> base, described above, confirmed that this residue was essential for catalysis.<sup>41</sup> Other experiments by the Scott group in solution showed that if the Tb(III) ion was added after substrate cleavage had begun in the presence of magnesium ion, cleavage stopped. If higher concentrations of magnesium ion were then added, cleavage resumed. This experiment could be duplicated for the HH16, HH $\alpha$ 1 and HH8 hammerhead constructs. The crystal structure reported as PDB: 359D in reference 43 used the same construct as had been used in previous hammerhead solid-state structures, that of RNA 6 (see Figure 6.10).

The reference 43 researchers put forth two theories to explain their results, both assuming that Tb(III), binding more tightly to the hammerhead construct,

displaces a  $Mg^{2+}$  ion that is essential for catalysis. The first hypothesis asserts that  $Mg^{2+}$  assists in forming a transition state structure, not that observed in the crystal structure, that maintains a particular orientation between  $G_5$  and nearby stem (helix) III. The second asserts that (a)  $Mg^{2+}$  must be released before  $G_5$  can participate in a transition state structure and (b)  $Tb(III)$  with its tighter binding is not released, thereby blocking the required conformational change. The authors tested the hypotheses by placing a bulky group at the 2'-hydroxyl position of  $U_{15.3}$  in stem (helix) III finding that the  $Tb(III)$  inhibition (and presumably  $Tb(III)$  binding) did not take place but catalytic cleavage took place in the presence of  $Mg^{2+}$ . This experiment appears to refute the first hypothesis above— $Mg^{2+}$  assists in forming a transition state structure that maintains a particular orientation between  $G_5$  and nearby stem (helix) III—while leaving the second hypothesis viable and consistent with mechanistic data available in 1998—the  $Mg^{2+}$  binding to  $G_5$  must be released before  $G_5$  can participate in a transition state structure. However, the actual role of  $G_5$  in hammerhead ribozyme catalysis remained open and in need of further experimentation.

Scott's group continued its study of hammerhead X-ray crystallographic structure with the aim of understanding the conformational change required for ribozyme self-cleavage.<sup>44</sup> To create a kinetic bottleneck that would prevent the final chemical cleavage reaction step in the crystal, the researchers introduced a cleavage site modification—a *talo*-5'-C-methyl modified ribose on  $A_{1.1}$  at the C17-A1.1 cleavage site (see Figure 6.17).

Through this modification, they hoped to capture a later conformational intermediate poised to form an inline transition state. The technique used was that of time-resolved crystallography, one that has successfully captured protein intermediates (see Section 3.7.2.3). The data determining the crystal structure has been deposited in the Protein Databank as PDB: 379D.



**Figure 6.17** *Talo*-5' C-methyl substituent at the hammerhead ribozyme active site.

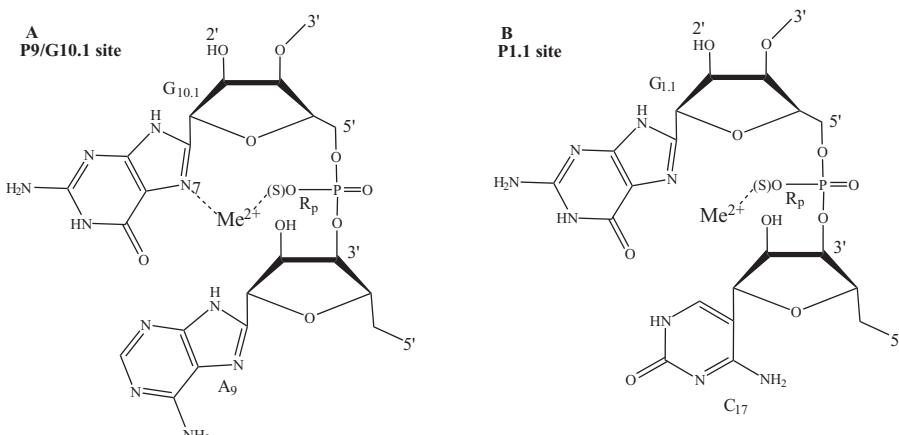
First, the researchers found that the rate of substrate turnover in the *talo*-5'-C-methyl modified RNA was reduced well over 300-fold. The experiments were carried out in the presence of  $\text{Co}^{2+}$  as a divalent ion slightly more catalytically active than  $\text{Mg}^{2+}$ . Further study found that the modified RNA ground-state crystal structure at pH 6 (both in the presence and absence of  $\text{Co}^{2+}$ ) was indistinguishable from the unmodified hammerhead structures reported and discussed previously (PDB: 299D and 300D). The same was found for the structure at pH 8.5 in the presence of  $\text{Co}^{2+}$  in comparison to the structure seen for  $\text{Mg}^{2+}$ -soaked crystals at pH 8.5 (PDB: 301D). The  $\text{Co}^{2+}$ -soaked crystals (100 mM  $\text{CoCl}_2$ , 30 minutes at pH 6 and 50 mM  $\text{CoCl}_2$ , 30 minutes at pH 8.5) served as controls for the current research because they indicated little or no ground-state structure disturbance with inclusion of the *talo*-methyl modification. In these “ground state” structures, the A-form RNA conformation at the cleavage site puts the  $\text{C}_{17}$  O<sub>2'</sub> oxygen maximally out of register (90°) with the  $\text{A}_{1,1}$  phosphorus and O<sub>5'</sub> oxygen for an inline attack mechanism.

Following these findings, the hammerhead ribozyme–substrate complex crystals were soaked for 2.5 hours in 50 mM  $\text{Co}^{2+}$  at pH 8.5. Under these conditions, cleavage in an unmodified RNA would have taken place in the crystalline state even faster than the rate in solution. However, cleavage did not take place in the *talo*-5'-C-methyl modified RNA. Instead, extensive conformational changes were noted in the positions of  $\text{C}_{17}$  and at the scissile phosphate of  $\text{A}_{1,1}$ . The base and ribose of  $\text{C}_{17}$  rotated about 60°, causing the base of  $\text{C}_{17}$  to move over 8.7 Å (relative to its position in the comparable ground-state structure PDB: 301D) and to stack onto the  $\text{A}_6$  nucleobase. These and other changes at the cleavage site pulled the scissile phosphate away from its standard helical geometry and toward a position more relevant to the inline  $S_N2$  mechanism. Essentially the scissile phosphate is rendered more susceptible to nucleophilic attack from the cleavage site  $\text{C}_{17}$  2'-hydroxyl that has also been repositioned in the *talo*-5'-C-methyl-modified RNA crystal structure—the 5' O atom of  $\text{A}_{1,1}$  (leaving group) and the 2' O atom of  $\text{C}_{17}$  (nucleophile) are 3.1 Å apart in the PDB: 379 structure, whereas the same atoms are 4.0 Å apart in the PDB: 301D structure. Modeling different phosphorus coordination spheres at the scissile phosphate site does not unequivocally place the phosphorus atom in the pentacoordination conformation necessary for the  $S_N2$  reaction mechanism. This is due to the inability to visualize the necessary atoms (the crucial phosphate group of  $\text{A}_{1,1}$ , for instance) at the 3.0-Å resolution of the crystal structure. Despite the changes in conformation at the cleavage site, other stacking interactions within the catalytic pocket and in its vicinity are conserved in the activated structure displayed in this work (PDB: 379D). This led the reference 44 authors to conclude that global conformational changes do not take place in the hammerhead ribozyme during cleavage and to describe the PDB: 379D structure as a “late intermediate” in the reaction pathway.

Researchers in the field continued to study the hammerhead ribozyme–substrate complex by biochemical and spectroscopic means. Three papers of the many published between 1999 and 2001 will be used here to illustrate the types of experimental results and to define conclusions based on these. In a *Biochemistry* paper published in 1999, Herschlag and co-workers continued their biochemical study of the hammerhead system and identified a hammerhead ribozyme metal ion binding site that appeared to interact with the hammerhead substrate's cleavage site even though these sites are ~20 Å apart in the X-ray crystallographic ground state structures.<sup>45</sup> The research studied two important hammerhead sites by replacing specific phosphate oxygens with sulfur, determining the effect on catalytic cleavage rate for the replacements and then determining if the thiophilic metal ion Cd<sup>2+</sup> could rescue catalysis. Substitution of specific phosphodiester oxygens by sulfur dramatically lowers Mg<sup>2+</sup> affinity for the site resulting in decreased catalysis. Addition of more thiophilic metals such as Cd<sup>2+</sup> or Mn<sup>2+</sup> is known to “rescue” catalysis.

The two proposed metal binding sites were (1) the ribozyme P9/G10.1 site, located at the junction between stem (helix) II and the conserved catalytic core, and (2) the substrate P1.1, scissile phosphate, site (see Figure 6.18).

For the HH16 hammerhead construct used in this research the specific residues at the first site would be A<sub>9</sub>/G<sub>10.1</sub>, and those at the second, cleavage, site would be G<sub>1.1</sub>/C<sub>17</sub>. The specific phosphate oxygen atoms binding the metal at both sites would be the nonbridging prochiral pro-R<sub>p</sub> phosphate oxygens, the same sites that were occupied by metal ions in the PDB: 301D X-ray crystallographic structure. For instance, looking at Table 6.4 for PDB: 301D, one finds that the bonding distance from Mg<sup>2+</sup> at site 6 to the pro-R<sub>p</sub> phosphate O of A<sub>1.1</sub> is equal to 2.43 Å. It had been shown previously and confirmed in this study that replacement of the pro-R<sub>p</sub> phosphate oxygen with sulfur at either

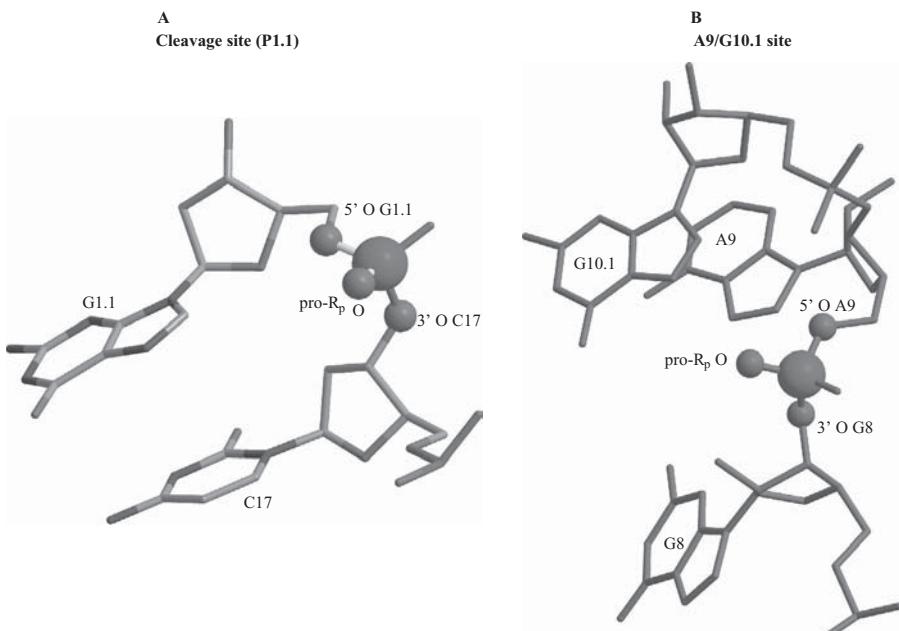


**Figure 6.18** “Rescue” sites in the HH16 hammerhead ribozyme construct.

(or both) site(s), A<sub>9</sub>/G<sub>10,1</sub> or G<sub>1,1</sub>/C<sub>17</sub>, had a deleterious effect on substrate cleavage rate (at least 10<sup>4</sup>-fold decrease in rate). Catalysis could be rescued by addition of Cd<sup>2+</sup> ions. Extensive kinetic studies reported by the reference 45 researchers led to the conclusion that a single rescuing metal ion bound to the P9/G10.1 site acquired a pro-R<sub>p</sub> phosphoryl oxygen at the P1.1 cleavage site as an *additional* ligand in the transition state. The single metal ion was proposed to cause a large conformational change into an “active conformation” and then into a transition state in which cleavage would take place. As shown in Figure 4 of reference 45, helix II, containing the P9/G10.1 site in domain II, is pulled toward the substrate cleavage site at P1.1. In this model, the domain II conformation remains relatively unchanged while domain I of the ribozyme undergoes the most extensive conformational change to form the catalytically active structure. The authors hypothesize that the docking of domain I onto domain II forms a catalytic pocket in which cleavage of the scissile phosphate group could be carried out.

Another paper in this series of three studied the behavior of the same two sites observed spectroscopically using <sup>31</sup>P NMR.<sup>20</sup> The reference 20 researchers used the same site-specific phosphorothioate labeling of the hammerhead RNA at the A9/G10.1 and P1.1 (G1.1/C17) sites as described above. However, they used a DNA substrate or 2'-OMe-substituted nucleotide at the cleavage site in the RNA substrate (RNA-OMe) so that catalytic cleavage would not proceed. The PDB:1HMH X-ray crystallographic structure was used as a basis for visualizing the three-dimensional structure of the two sites. Figure 1B of reference 20 and Figure 6.19 show a close-up view of the two sites.

For an RNA/RNA-OMe hybrid, substitution of the A<sub>9</sub> pro-R<sub>p</sub> phosphoryl oxygen with sulfur showed upfield shifts of 2–3 ppm in the <sup>31</sup>P NMR A<sub>9</sub> resonance with increasing Cd<sup>2+</sup> concentration. Addition of only 0.5 equivalents of Cd<sup>2+</sup> was sufficient to cause the upfield shift indicating relatively high Cd<sup>2+</sup> affinity for this site. The authors also believe that the upfield (rather than the expected downfield shift on metal binding) shift indicated the more covalent nature of the Cd<sup>2+</sup>–S bond. Essentially the same behavior was observed for the A<sub>9</sub> pro-S<sub>p</sub> phosphoryl oxygen. Behavior at the cleavage site differed considerably. For an RNA/DNA hybrid, substitution at the P1.1 phosphate site caused the following behaviors: (1) For the P1.1 pro-R<sub>p</sub> phosphoryl oxygen substitution by sulfur, addition of up to 10 equivalents of Cd<sup>2+</sup> was required to shift the <sup>31</sup>P resonance *upfield* by ≤0.6 ppm; and (2) for the P1.1 pro-S<sub>p</sub> phosphoryl oxygen substitution by sulfur, addition of up to 10 equivalents of Cd<sup>2+</sup> shifted the <sup>31</sup>P resonance *downfield* by ≤0.6 ppm. The reference 20 authors believed that the downfield shift in the latter case indicated Cd<sup>2+</sup> coordination to the pro-R<sub>p</sub> oxygen in the S<sub>p</sub>-substituted phosphorothioate. In other words, Cd<sup>2+</sup> coordinated to the pro-R<sub>p</sub> position at the P1.1 site (G<sub>1,1</sub> in the hammerhead construct used in these experiments) whether that position was taken by a sulfur atom or an oxygen atom. The much larger concentration of Cd<sup>2+</sup> required to obtain shifts at the P1.1 site argued for a weaker metal–ligand interaction at this site. These authors found their NMR results for the substrate



**Figure 6.19** The pro-R<sub>p</sub> oxygen atoms at two sites in the PDB: 1HMH hammerhead ribozyme. (A) The Cleavage site (P1.1) and (B) the A9/G10.1 site. (Adapted with permission from Figure 1 of reference 20. Copyright 2000, American Chemical Society.) Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. (Printed with permission of CambridgeSoft Corporation.)

P1.1 pro-R<sub>p</sub> phosphoryl oxygen–sulfur substitution consistent with Mg<sup>2+</sup> interactions seen in the Scott crystal structure (PDB: 301D) obtained at pH 8.5, because the preceding NMR experiments were conducted at this pH. In the same NMR experiment conducted at pH 5.5, no Cd<sup>2+</sup> shifts of the <sup>31</sup>P NMR spectra were seen for the substrate P1.1 pro-R<sub>p</sub> phosphoryl oxygen–sulfur substitution.

Comparison of the described results for the hammerhead complex to those obtained with several model compounds led to the conclusions that the upfield (coordination to sulfur) or downfield (coordination to oxygen) shifts could be noted for the model compounds as well. Similar NMR results were obtained whether a DNA or an RNA–OMe substrate was used in conjunction with the hammerhead ribozyme RNA, and substitution of sulfur at other phosphate oxygen positions within the ribozyme or substrate did not cause binding of Cd<sup>2+</sup> or <sup>31</sup>P NMR shifts. Comparison of the NMR results with those obtained previously using Mn<sup>2+</sup> in EPR, ENDOR, and ESEEM studies<sup>46</sup> showed that there is at least one high-affinity metal binding site ( $K_d \leq 10$  micromolar, μM) and at least one, and possibly more, low-affinity sites since full hammerhead activity requires >1 millimolar, mM Mn<sup>2+</sup> (or Mg<sup>2+</sup> as determined by other experimentalists). All observations taken together indicate that the A9/G101.1

site is critical for hammerhead catalytic activity and that another divalent metal site is probably required as well. In comparing this statement to the “rearrangement mechanism” of Herschlag described in previous paragraphs, DeRose’s research group postulates that the A9/G101.1 site (higher affinity, larger effect of added Cd<sup>2+</sup>) is required for both ground and transition states while the P1.1-S<sub>Rp</sub> metal binding site (lower affinity, smaller effect of added Cd<sup>2+</sup>) is required for only the transiently populated transition state. The <sup>31</sup>P NMR technique described in reference 20 is useful in that the substituted RNA site is immediately observable with a signal that is sensitive to metal ion binding if it is catalytically important. One limitation may be that phosphorothioate substitutions may also cause conformational changes that interfere with catalysis. This is not suspected in the reference 20 work because Cd<sup>2+</sup> addition supports full activity of the phosphorothioate-substituted ribozyme and substrate complexes.

The third paper in the series compared hammerhead cleavage reactions stimulated by monovalent and divalent cations.<sup>47</sup> In general, the hammerhead cleavage reaction is found to be most efficient in millimolar concentrations of divalent metal cations. This is not a general truth for all ribozymes because some, such as the hairpin ribozyme, can operate in the absence of divalent metal ions. Also, Scott and co-workers described catalytic activity for hammerheads in the absence of divalent metal ions provided very high concentrations of monovalent ions are present.<sup>48</sup> The phosphorothioate–Cd<sup>2+</sup> “rescue” experiments described above suggested that certain phosphoryl groups participate in metal ion interactions important for catalysis. Other experiments have indicated participation by metal ions in catalysis carried out by ribozymes. However, metal ions are also known to be important for the folding and structural properties of DNA and RNA, and it is difficult to separate these structural effects from those only involved in catalysis.

To attempt to differentiate between structural and catalytic effects of metal ions, Herschlag and co-workers<sup>47</sup> use kinetic experiments, with reactions performed under single-turnover conditions, to test the importance of monovalent metal ions for catalysis. They used the HH16 and HH8 ribozyme–substrate complexes in this study. First, they found that for HH8, first-order dependence on mM Mg<sup>2+</sup> concentrations was found up to a saturation rate of 1.3/min at approximately 10 mM Mg<sup>2+</sup>. In contrast, the catalytic cleavage rate in LiCl was very slow below 0.5 M and showed second-order dependence up to a rate of 0.24/min in 4 M LiCl. When abasic (C3X, U7X, G10X, and so on, for HH16) and other mutations (C3U, C17A, and so on, for HH8) were introduced into HH16 and HH8, kinetic experiments like those described previously in the Herschlag group’s work (references 41 and 42) showed that reaction rates in 10 mM Mg<sup>2+</sup> and 4 M Li<sup>+</sup> were mostly analogous. One point of difference was found for the abasic replacements at positions G10.1 and C11.1—that is, G10.1X and C11.1X. While the rate in 10 mM Mg<sup>2+</sup> decreased relative to wild-type HH16 behavior, the rate in 4 M Li<sup>+</sup> increased from the wild-type rate. Since G10.1 is an important divalent metal ion ligand, the researchers concluded that some

hammerhead–metal ion interactions that stimulate catalytic rate in the presence of divalent ions could be absent with monovalent ions.

To test this hypothesis, the Herschlag group conducted the phosphorothioate replacement and rescue experiments as described in the reference 45 article discussed above. The previous result indicated that if the P9 pro-R<sub>p</sub> phosphoryl oxygen and that of P1.1 in the cleavage site were replaced by sulfur, cleavage rates in Mg<sup>2+</sup> were greatly decreased (by 10<sup>3</sup>- to 10<sup>4</sup>-fold). Using the same phosphoryl oxygen–sulfur replacements, the reference 47 authors predicted that if the monovalent Li<sup>+</sup> ion were not used to coordinate the P9 and P1.1 pro-R<sub>p</sub> phosphoryl oxygen sites, the amount of rate decrease would be considerably less. Indeed a rate decrease of approximately 10-fold was found for the reactions in 4M Li<sup>+</sup> substantiating this hypothesis. This and other experimental results led the researchers to conclude that the substantial difference between catalytic stimulation by monovalent and divalent ions is found at the P9/P1.1 site. Overall conclusions from this work and other observations concerning the metal ion's role in structure versus its role in catalysis were as follows: (1) Hammerhead cleavage reactions in the presence of monovalent or divalent metal ions are very similar; (2) the role of the metal or even the ammonium ion (NH<sub>4</sub><sup>+</sup>), also found to increase hammerhead catalytic rate, may be to neutralize phosphate repulsion and facilitate hammerhead core folding into the catalytic conformation; (3) some part of hammerhead rate enhancement does involve at least one divalent metal ion; and (4) the divalent ion may provide its catalytic contribution by electrostatic stabilization of the transition state in addition to its role in facilitating formation of the active hammerhead structure.

The hypothesis that one divalent metal ion coordinated to both the A<sub>9</sub>/G<sub>10.1</sub> and the scissile phosphate at P<sub>1.1</sub> in a hammerhead catalytically active transition state was challenged by Murray and Scott in a *Journal of Molecular Biology* paper published in 2000.<sup>49</sup> The authors initially constructed models using the graphics display program O.<sup>50</sup> The structures were then refined in X-PLOR 3.8.<sup>51</sup> Their goal was to use molecular modeling to define a structure that would obey bonding, RNA stereochemistry, and the inline S<sub>N</sub>2 mechanism for the phosphodiester bond cleavage, as well as incorporate the Herschlag group divalent-ion-coordinating hypothesis. As described above, the Herschlag group proposed that to achieve the catalytically active transition state with one coordinating divalent metal ion, the stem II helix and neighboring domain II with the A<sub>9</sub>/G<sub>10.1</sub> metal ion coordination site remain nearly unchanged and that the domain I (uridine turn) region and its associated cleavage site residue, C<sub>17</sub>, changes conformation from the ground-state crystal structure to accommodate metal ion binding to a P<sub>1.1</sub> pro-R<sub>p</sub> phosphate oxygen.<sup>49</sup> In reference 49, the starting point for all molecular modeling was the authors' best catalytically active hammerhead crystal structure with Mn<sup>2+</sup> bound (PDB: 300D). Murray and Scott found that only one Mg<sup>2+</sup> coordinating position (in its octahedral coordination sphere) could accommodate the single divalent ion hypothesis. This position allowed a stereochemically plausible C<sub>17</sub>–U<sub>16.1</sub> phos-

phate linkage to exist, although the one hydrogen bond seen in the crystal structure pairing U<sub>16,1</sub> and A<sub>15,1</sub> was broken. Minimization of the starting structure provided a solution that required stem I and II helices to become approximately parallel with the stem I helix pushed up into a higher position compared to its crystal structure position (see Figure 4 of reference 49). These authors then attempted to fit the best plausible structure to other experimental data. For example, crosslinking experiments between residues in stems I and II had concluded previously that the distances between 2'-OH groups of residues 2.6 (stem I) and 11.5 (stem II) was set at 16 Å or less. In the crystal structure, this distance was approximately 11 Å. However, if this distance and associated stereochemical constraints were imposed on the energy-minimized structure obtained in this molecular modeling study, the stem I helix would be required to unwind fully—an energetically unlikely event. The reference 49 authors believe a large-scale rearrangement in the hammerhead ribozyme into a transition state as described by Herschlag's group is unnecessary, citing the ability of a hammerhead cleavage reaction product to be formed within a crystalline lattice as they have shown in their previous work (reference 44, PDB: 379D and reference 39, PDB: 301D). They further cite X-ray crystallographic data on a ribozyme–product complex obtained in 2000. This work is discussed in the following paragraphs.

Scott and co-workers next used crystal lattice trapping and X-ray holographic reconstruction to capture and visualize a hammerhead enzyme–product complex.<sup>52</sup> The data have been deposited in the Protein Data Bank as PDB: 488D. First, it should be stated that, upon cleavage, hammerhead ribozyme–complex crystals become disordered and the quality of X-ray diffraction deteriorates to the extent that structures of the ribozyme–product complex cannot be solved. These authors circumvented this problem by mixing a small percentage of modified (or inhibited) RNA substrate into the crystallization solution, resulting in crystalline lattices of sufficient rigidity to trap the ribozyme–product complex in a matrix of unreacted modified ribozyme–substrate complex. The collected data were then manipulated using X-ray holographic methodology, resulting in clearer and more reliable identification of structural elements and conformational changes. Data sets were collected on crystals shown (by HPLC UV absorbance integration) to have undergone 40% and 60% cleavage to product. The data were refined against a previously determined initial state structure (PDB: 299D) and further manipulated using the X-ray holographic procedure (which was itself tested using the initial state structure). The conformational changes were observed in both the 40% and 60% cleavage-to-product crystals.

The final refined structure showed a 2',3'-cyclic phosphate product at the C<sub>17</sub> ribose position and a C<sub>17</sub> nucleotide that had moved dramatically and become almost perpendicular to the Watson–Crick faces of conserved nucleotides G<sub>5</sub> and A<sub>6</sub> in the catalytic pocket. Interactions of functional groups of G<sub>5</sub> and A<sub>6</sub> with the product and C<sub>17</sub> included (1) a hydrogen bond formed between the exocyclic amine (N<sub>6</sub>) of A<sub>6</sub> and the cyclic phosphate nonbridging oxygen

( $A_6N_6-O_{1P} = 2.7 \text{ \AA}$ ) and (2) the close approach of  $G_5$ 's keto oxygen ( $O_6$ ) and the 2'-oxygen of  $C_{17}$  ( $G_5O_6-2'-O = 2.9 \text{ \AA}$ ). These interactions are most interesting because they involve groups on  $G_5$  and  $A_6$  known to be required for hammerhead catalytic cleavage. The authors speculate that the 2'-OH deprotonation necessary to form the 2',3'-cyclic phosphate product might result in a transient  $G_5$  enol form that would produce a  $G_5O_6-2'-O$  hydrogen bond as described in (2) above. Additionally,  $C_{17}$  appears to form a perpendicular stabilizing aromatic interaction with  $G_5$  similar to those found for aromatic substituents in protein structures. The visualized 2',3'-cyclic phosphate resembles a pentacoordinated 2',3',5'-cyclic oxyphosphorane transition state; however, the 5'-O connection appears to be missing in the visualized crystal structure. Therefore, the issue of 5'-oxygen stabilization during cleavage is not addressed in this structure, leaving the authors to hope for a stable transition state analog that mimics a 5-coordinate oxyphosphorane intermediate in future research.

In an earlier X-ray crystallographic study the Scott research group found that terbium(III) inhibited the hammerhead cleavage reaction (reference 43, PDB: 359D) and also found a  $Tb^{3+}$  ion bound to the Watson–Crick face of  $G_5$ . In that work, the authors proposed that this  $Tb^{3+}$  prevented  $G_5$  from making the interactions required for catalysis. This finding agrees with the importance of  $C_{17}-G_5$  and  $A_6$  interactions found in the PDB: 488D structure. Research by other groups has confirmed that the exocyclic groups of both  $A_6$  and  $G_5$  are required for competent catalytic cleavage in the hammerhead ribozyme. The 2000 *Molecular Cell* paper does not address changes in domain II (augmented stem II), especially those discussed previously in the  $A_9/G_{10,1}$  region, and does not provide any insight as to the catalytic role of conserved nucleotides in that region.

In summary, the reference 52 authors propose the following: (1) A conformational rearrangement of the cleavage site nucleotide  $C_{17}$  positions the attacking 2'-O nucleophile in line with the 5'-O leaving group, allowing the  $S_N2$  reaction to proceed (see Figure 6.9); (2) interactions between the scissile phosphate and ribose of  $C_{17}$  and exocyclic functional groups of  $G_5$  and  $A_6$  are involved in establishing the active intermediate; and (3) the RNA itself may possess the necessary catalytic properties and may not require divalent metal ions for the cleavage reaction. Their results do not (1) address how the 5'-oxygen leaving group is stabilized, (2) explain the critical requirement for  $C_3$  in hammerhead catalytic cleavage, or (3) rule out the proposed Herschlag mechanism in which the  $A_9$  phosphate and scissile phosphate coordinate a single metal ion (but see objections in reference 49, discussed above).

In 2001, Scott published a paper that discussed ribozyme catalysis in terms of “orbital steering” and “inline fitness.”<sup>53</sup> In defining the former process, it has been suggested that the catalytic efficiency of enzymes depends on their ability not only to juxtapose the reacting atoms but also to “steer” their orbitals along a path that takes advantage of this strong directional preference.<sup>54</sup> In the latter, an inline fitness parameter has been devised in which a perfectly aligned ( $O2'-P-O5'$  angle for  $S_N2$  nucleophilic attack equals  $180^\circ$ ), and the 2'-oxygen

atom is positioned approximately 3 Å from the attacked phosphorus atom. This situation is assigned an inline fitness value of 1.0. Conversely, for the most poorly aligned situation ( $O_2'-P-O_5'$  angle for  $S_N2$  nucleophilic attack equals 90° in the trigonal bipyramidal transition state), the fitness parameter has a value of 0.0. In published crystal structures the position of the 2'-oxygen atom of the conserved G<sub>8</sub> ribose is located inline with the phosphorus atom and the 5'-oxygen atom of A<sub>9</sub> with the 2'-O-P distance of 2.77 Å. This grouping has an inline fitness parameter of 1.17, yet cleavage does *not* take place at this site. Scott explains this apparent discrepancy by using a hydrogen-bonding network among nucleobases at the G<sub>8</sub>, A<sub>9</sub>, and G<sub>12</sub> positions to orbitally steer the necessary lone pair orbitals away from the phosphorus center, thus preventing nucleophilic attack and bond formation. Scott proposes that the orbital steering mechanism should be invoked in addition to that of proper alignment of attacking and leaving groups in the hammerhead cleavage reaction. In published crystal structures described up to this time, a considerable conformational change would be required to achieve a proper inline fitness parameter at the cleavage site. For instance, the so-called initial state hammerhead ribozyme structure (PDB: 299D) is in a conformation incompatible with the inline attack mechanism and has a fitness parameter of 0.05. In addition to the rate enhancement that would be achieved by conformational changes toward an inline arrangement, Scott believes that orbital steering effects could be responsible for some of the 10<sup>6</sup>-fold (compared to the uncatalyzed reaction) rate enhancement seen for hammerhead cleavage. Other rate enhancement effects—general acid/base catalysis by metal ions or nucleobases or transition state stabilization interactions—also are responsible for some cleavage reaction rate enhancement, although orbital steering effects may be involved with these as well.

Evidence presented in a *Journal of Molecular Biology* paper of 2002 argued that a pH-dependent conformational change, rather than the chemical step of substrate cleavage, could be the rate-limiting step in the hammerhead ribozyme cleavage reaction.<sup>55</sup> First, it is agreed that there is a positive log-linear relationship between pH and hammerhead cleavage rate up to pH 8.0 (using Co<sup>2+</sup> enhancement) or pH 8.5 (using Mg<sup>2+</sup> enhancement). Also, it is known that the hammerhead construct used to determine the Scott group crystal structures (RNA 6 in Figure 6.10) contains a kinetic defect that results in a low concentration of active ribozyme in solution. In solution, the crystal structure RNA 6 construct exhibits a catalytic rate of 0.08/min in 1.8M Li<sub>2</sub>SO<sub>4</sub> at pH 8.5 in the absence or presence of Me<sup>2+</sup>, whereas a kinetically well-behaved hammerhead exhibits rates of 1.0/min at pH 7.5 in 10mM MgCl<sub>2</sub>. Interestingly, the cleavage reaction in the crystal obeys the expected log linear relationship between cleavage rate and pH with a slope of 0.7, the same relationship found for other hammerhead constructs in solution. Additionally, cleavage in the crystal displays a biphasic rate dependence with a slow initial pH-dependent phase and a faster pH-independent phase appearing at pH 6. The faster rate at or above pH 6 is comparable to the fastest rates observed for other

hammerhead RNAs at the same pH 6. The reference 55 researchers also found that substitutions at the 2'-OH position at the cleavage site with either 2'-OCH<sub>3</sub> or 2'-F resulted in crystal structures in which the conformational change leading to the cleavage reaction would not take place. Based on their observations, the Scott group concluded that a pH-dependent conformational change is the rate-determining step for hammerhead self-cleavage reactions and that an ionizable proton at the 2'-position is required to achieve the necessary conformation. They hypothesized then that deprotonation at the cleavage-site ribose 2'-OH drives the needed conformational change.

In 2003, Victoria DeRose published a *Current Opinion in Structural Biology* paper that discussed the known information on metal ion binding to catalytic RNA.<sup>16</sup> Much of the foregoing discussion of the hammerhead ribozyme and the cleavage reaction it catalyzes has been summarized in this paper. DeRose noted that cations provide chemical functionalities not available in the RNA itself and may be necessary for ribozyme catalysis. However, if cations are indeed required, the challenge is to assign their role as necessary for structure (RNA folding) or for catalytic enhancement of the chemical reaction carried out by the ribozyme. DeRose would like to define the metal ion involved in catalysis as one that (1) stabilizes the reaction's transition state and (2) is associated with a group that changes bond order during the catalytic reaction. Spectroscopic analyses—<sup>31</sup>P NMR or EPR studies using paramagnetic Mn<sup>2+</sup>, for instance—have, in rare instances, pinpointed the metal's location but more often have shown the overall influence of metal ions on folding. However, analytic methods have been less successful in identifying metal-dependent changes in ribozyme catalytic activity *not* dependent on folding (conformational change). One biochemical method that comes close to accomplishing this has been the phosphorothioate metal-rescue (PS-rescue) experiment discussed previously. Applying DeRose's criteria, one would say that a metal site identified by PS-rescue to be far away (and stay far away) from the active site is defined as structural. Conversely, if a PS-rescued metal site involves a bond that changes order during the catalyzed chemical reaction then that metal ion may be directly involved in catalysis.

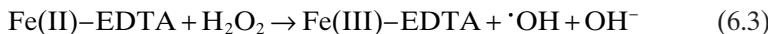
One would like to identify such a site in the hammerhead ribozyme–substrate complex. The A<sub>9</sub>/G<sub>10,1</sub> metal binding site shows possibilities. X-ray crystallography (PDB: 300D) shows an Mn<sup>2+</sup>–A<sub>9</sub> nonbridging phosphate oxygen bond of 2.31 Å in length and an Mn<sup>2+</sup>–G<sub>10,1</sub> N<sub>7</sub> bond 2.24 Å in length (see Table 6.4). If the A<sub>9</sub> prochiral R<sub>p</sub> oxygen is substituted by sulfur in a PS-rescue experiment, Mg<sup>2+</sup> no longer catalyzes the cleavage reaction but catalysis can be rescued by Cd<sup>2+</sup> addition. <sup>31</sup>P NMR shows that Cd<sup>2+</sup> does indeed bind to the A<sub>9</sub> R<sub>p</sub> phosphorothioate. However, no ground-state crystallographic structure shows the A<sub>9</sub>/G<sub>10,1</sub> site as being close to the cleavage site. This led the Herschlag group to present their hypothesis that a single metal ion coordinates to both the A<sub>9</sub>/G<sub>10,1</sub> and the cleavage site in the ribozyme–substrate complex transition state.<sup>45</sup> This hypothesis was in turn challenged by the Scott group showing by molecular dynamics calculations that the proposed single

metal ion binding to both the A<sub>9</sub>/G<sub>10,1</sub> and the cleavage site resulted in an unlikely high-energy state for the ribozyme–substrate complex.<sup>49</sup>

The DeRose group also studied hammerhead RNA constructs using paramagnetic Mn<sup>2+</sup> ions (high-spin *d*<sup>5</sup>, *S* = 5/2) examining these with electron paramagnetic resonance (EPR), electron nuclear double resonance (ENDOR), and electron spin-echo envelope modulation (ESEEM).<sup>46</sup> These instrumental techniques are discussed in Section 3.5.3. The manganese ion is similar to Mg<sup>2+</sup> in ionic radius and hydration enthalpy and yields high ribozyme cleavage rates in 0.1–1.0 M NaCl. The *S* = 5/2 Mn<sup>2+</sup> ion has a distinctive six-line EPR pattern that changes when the ion is bound to RNA, allowing Mn<sup>2+</sup> binding to be quantified. In addition, the unique signal can be probed for metal coordination details.

One high-affinity Mn<sup>2+</sup> site was found at 1:1 concentrations of Mn<sup>2+</sup>:RNA (~100 μM micromolar) that showed subtle line-shape changes indicating metal ion–RNA binding. Given this signal the hammerhead metal ion-binding site could be further examined through hyperfine interactions between nuclei of the RNA ligands and the Mn<sup>2+</sup> electron spin. RNA nuclei with spins that may be examined in this experiment include the <sup>31</sup>P of RNA phosphodiesters, <sup>1</sup>H (exchangeable on aqua ligands and non-exchangeable on sugars and bases), and <sup>14</sup>N on nucleobases. The hyperfine interactions can be observed using ENDOR or ESEEM because both these techniques depend on the influence of nearby nuclear spins on the Mn<sup>2+</sup> EPR signal. Using ENDOR, the researchers identified a specific high-affinity, tightly bound Mn<sup>2+</sup> that was coordinated to a phosphate oxygen (presumably of the A<sub>9</sub> nucleotide) as well as to at least one aqua ligand. Using ESEEM spectroscopy, the same Mn<sup>2+</sup> was seen to be coordinated to the N<sub>7</sub> of G<sub>10,1</sub>. (See Figure 8 reference 46.) The DeRose group also hypothesized that the identified high-affinity site, populated at micromolar concentrations, must be joined by another weaker binding metal ion site since at least millimolar concentrations of Mn<sup>2+</sup> in ~1 M NaCl are necessary to achieve full catalytic activity in the hammerhead. This site is not identified using EPR studies described in reference 46.

The need for both high- and low-affinity metal ion-binding sites was echoed by the hydroxyl radical footprinting experiments carried out by Hampel and Burke.<sup>56</sup> The footprinting method involves rapid hydroxyl-radical generation by Fe(II)–EDTA–hydrogen peroxide solutions to study folding kinetics of RNA complexes. Method details are provided in reference 57, and a brief outline is provided here. The solutions containing the RNA constructs to be studied are incubated with different concentrations of cation, treated with Fe(II)–EDTA–hydrogen peroxide solutions, and then quenched to stop radical generation. Hydroxyl radicals are generated according to the following reaction:



Hydroxyl radicals attack the C4' position of the sugar resulting in sugar decomposition and phosphodiester cleavage. However, if nucleotides are protected

by RNA folding they may not appear on the assay or will appear at lower concentration. Reaction products are assayed through electrophoresis on a 15% (w/v) acrylamide–8M urea gel. The hydroxyl radical footprinting method can be used to accurately define RNA folding rate constants slower than 6–8/min and to probe the structure of the entire hammerhead construct at the individual nucleotide level, because inaccessible nucleotides will not be visualized. Results can then be compared to experiments using base change or functional-group substituted mutants that have diminished catalytic activity.

First, Hampel and Burke described hammerhead secondary structure as determined by fluorescence resonance energy transfer (FRET) experiments. Metal ion titrations using FRET signals demonstrated RNA folding to the typical hammerhead secondary structure Y-shape. (See Figures 6.10, 6.11, and 6.12.) The FRET technique reveals that folding occurs with two discrete transitions: (1) At < mM Mg<sup>2+</sup> the domain II stack forms, and (2) at low mM Mg<sup>2+</sup> the domain I motif assumes a specific structure that orients stem (helix) I close to stem (helix) II. Substitution of the ribose 2'-deoxy nucleotide at the conserved position G<sub>5</sub> alters the geometry of the three-way junction such that stems I and II are not close together while the domain II structure is maintained. From previous studies, it is known that cations are required for folding to the hammerhead global structures described (Figures 6.10, 6.11, and 6.12). The principal metal ion that has been used in folding studies is Mg<sup>2+</sup>. In contrast, many cations have been shown to promote catalysis including divalent cations Mg<sup>2+</sup>, Mn<sup>2+</sup>, Ca<sup>2+</sup>, Sr<sup>2+</sup>, Cd<sup>2+</sup>, and Co<sup>2+</sup> and monovalent cations Na<sup>+</sup>, NH<sub>4</sub><sup>+</sup>, and Li<sup>+</sup>. Although catalytic cleavage rates are enhanced by these cations—some requiring high ionic concentration—catalysis does not have an absolute requirement for divalent metal ions or for inner-sphere coordination by metal ions.

Hampel and Burke observed that protection of hammerhead backbone sites in Mg<sup>2+</sup> solutions required assembly of the full ribozyme–substrate complex. In other words, testing of ribozyme or substrate separately in the hydroxyl footprinting assay showed essentially complete hydrolysis of all nucleotides (Figure 2B of reference 56). In contrast, the fully assembled ribozyme–substrate complex showed protection of nucleotides structurally near the densely packed three-helix junction of hammerhead constructs HH16, HH $\alpha$ 1, and RNA 6. Two of the ribozyme group of protected nucleotides (G<sub>5</sub>, A<sub>6</sub>) are part of the conserved uridine U-turn seen in all known hammerhead constructs. (See Figures 6.10, 6.11, and 6.12.) The footprinting results are collected in Table 6.5.

The three hammerhead constructs display nearly identical solvent-protected (folded) sites except that HH $\alpha$ 1 and RNA 6 show an additional protected site at position 15.3. This is an important site since in the hammerhead crystal structures this site in stem III is approached by domain I, part of the catalytic pocket. The authors believe that a single structure gives rise to the observed pattern of hydroxyl radical protection of nucleotides since protection at individual nucleotides varies in the same manner with Mg<sup>2+</sup> concen-

**TABLE 6.5 Protected Nucleotides in Hydroxyl Footprinting Assay as Adapted from Reference 57**

| Hammerhead Construction | Ribozyme Strand  | Substrate Strand  | Additional Ribozyme Site |
|-------------------------|--|---|--------------------------|
| HH16                    | A <sub>9</sub><br>G <sub>8</sub><br>U <sub>7</sub><br>A <sub>6</sub><br>G <sub>5</sub> | U <sub>1,2</sub><br>G <sub>1,1</sub><br>C <sub>17</sub> |                          |
| HH $\alpha$ 1           | A <sub>9</sub><br>G <sub>8</sub><br>U <sub>7</sub><br>A <sub>6</sub><br>G <sub>5</sub> | A <sub>1,2</sub><br>U <sub>1,1</sub><br>C <sub>17</sub> | G <sub>15,3</sub>        |
| RNA 6                   | A <sub>9</sub><br>G <sub>8</sub><br>U <sub>7</sub><br>A <sub>6</sub><br>G <sub>5</sub> | C <sub>1,2</sub><br>A <sub>1,1</sub><br>C <sub>17</sub> | U <sub>15,3</sub>        |

tration over a range of Mg<sup>2+</sup> or Li<sup>+</sup> concentrations from 0.1 to 100 mM. The importance of the 2'-deoxy substitution at G<sub>5</sub>, a substituted residue that alters hammerhead global folding according to other experimental techniques (it blocks formation of the stem I/stem II catalytic motif in FRET analyses), is also demonstrated in the hydroxyl radical footprinting experiment. In a test of a ribozyme–substrate complex containing a 2'-deoxy G<sub>5</sub>, the nucleotide pattern indicates that the ribozyme–substrate complex does not fold into its solvent-protected form. Looking at PDB: 299D, 300D, and 301D crystal structures, one can find several potential hydrogen bonding partners between G<sub>5</sub>'s 2'-O position and atoms in the nucleotide at position 15.3.

Up to the experimental limit where both folding and catalytic effect could be measured, the researchers always found that folding was faster than catalysis. They also found a log-linear dependence of cleavage rate on pH in agreement with other experimental techniques. In almost all of their experiments, Hampel and Burke<sup>56</sup> used a substrate containing a 2'-deoxy substitution at C<sub>17</sub>, the cleavage site, to prevent substrate cleavage. Only in one experiment with HH16 at pH 6 were the workers able to use an intact substrate. They found that the solvent protection of the ribozyme strand, residues G<sub>5</sub>–A<sub>9</sub>, was unchanged and concluded that the 2'-deoxy C<sub>17</sub> substitution does not affect global folding. The hydroxyl footprinting results showing protection from hydroxyl radical attack near the hammerhead three-stem junction are also in agreement with the crystal structures in which residues G<sub>5</sub>, A<sub>6</sub>, G<sub>8</sub>, and C<sub>17</sub> show reduced solvent accessibility. However, four residues that are protected in the hydroxyl radical attack experiments, U<sub>7</sub> and A<sub>9</sub> in the ribozyme and A<sub>1,1</sub>

and C<sub>1,2</sub> in the substrate, are highly solvent accessible in the X-ray structures. Most of these appear near the edges of the catalytic pocket and could become solvent protected as stems I and II move toward each other.

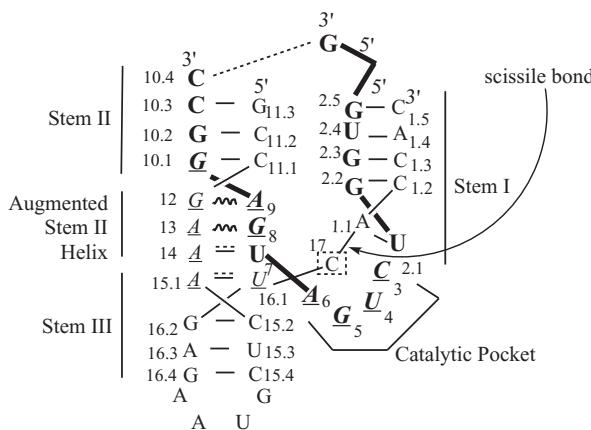
The reference 56 authors conclude the following: (1) Protected nucleotides in the hydroxyl radical footprinting experiments support a global tertiary structure similar or identical to that found in X-ray crystallographic experiments; (2) the protection found for residue 15.3 (U<sub>15,3</sub> in RNA 6 used for crystallography) provides strong evidence that the crystal form is also present in solution; (3) protection at U<sub>7</sub> seen in the hydroxyl radical experiments disagrees with its solvent accessibility in crystal structures indicating a more compact ribozyme–substrate complex form in solution; (4) 2'-deoxy G<sub>5</sub> substitution eliminates solvent protection (and by analogy proper hammerhead folding); (5) since folding is much faster than catalysis, it is difficult to know if the structural form found with hydroxyl radical footprinting is actually on the hammerhead native folding pathway; (6) although specific metal ion binding sites are seen in hammerhead crystal structures and seemingly confirmed by DeRose's NMR methods, the hydroxyl radical footprinting experiments indicate that many different mono- and divalent cations are equally capable of forming a solvent-protected motif; and (7) the data support the hypothesis of a conformational change between ground and transition states; however, the hydroxyl radical footprinting results probably only report on the most abundant species in solution—that is, the ground state.

In 2003, the Scott group published two more crystal structures showing the relationship between the hammerhead ribozyme as a nuclease—catalytic cleavage to produce a 5'-OH terminus and a 2',3'-cyclic phosphate terminus—and as a ligase—essentially the cleavage reaction in reverse to produce a phosphodiester (PDB: 1NYI and 1Q29).<sup>58</sup> The hammerhead ribozyme sequence catalyzes both self-cleavage and self-ligation, depending on which is needed biologically. The most studied hammerhead ribozymes, and all that have been discussed thus far, have greatly favored cleavage over ligation. It is known that engineering a variable length crosslink between two sequentially distant but spatially close 2'-oxygen atoms of residues 11.5 (in stem II) and 2.5 (in stem I) can switch the ribozyme from being a nuclease to being a ligase.<sup>59</sup> The hammerhead RNA begins to favor ligation when the length of the cross-link is increased.<sup>60</sup>

In the first reference 58 structure discussed (PDB: 1NYI), the researchers added an extra guanosine nucleotide at the 5' end of stem I that possessed an unusual 5' to 5' phosphodiester linkage. Concurrently, a corresponding guanosine (G<sub>11,4</sub>) was removed from the 5' end of stem II. The crystallization procedure followed corresponded to that carried out by the Scott group in previous experiments and included freeze-trapping the crystalline state by immersion in liquid nitrogen and maintaining the crystal at 100 K during data collection. It was hoped that the changes in hammerhead construct design would provide more stable crystal packing and hence a higher-resolution X-ray structure than those of 3.0-Å resolution achieved previously. Instead, the unusually linked

5'-5'-connected base on stem I formed a Watson–Crick base pair with C<sub>10.4</sub> creating a tethered hammerhead RNA in the crystal. Although the crystal diffraction did not improve measurably, the new construct did help address other criticisms of previous structures, namely that further movement and conformational change was needed to visualize a possible transition state or that modifications to crystal structure hammerheads prevented achievement of a probable transition state. See discussion of the “later intermediate” structure (PDB: 379D) above in which a *talo*-5'-C-methyl modified RNA (on A<sub>1.1</sub>, at the C17/A1.1 cleavage site) had been used to prevent cleavage in the crystal. In the latter case of modification, criticisms included the prediction that the crystal structures visualized an “off pathway” conformational change not related to achieving the actual hammerhead cleavage site transition state structure. In solution, the PDB: 1NYI-modified hammerhead construct (see Figure 6.20) cleaved in a manner similar to that of the hammerhead initial structures (PDB: 1MME and 299D) (see Figure 6.10), indicating that tethering of stem I to stem II was negligible in solution.

However, tethering of stem I to stem II through C–G hydrogen bonding in the crystal prevented cleavage completely in a construct where, importantly, the RNA leaving group had not been modified. Furthermore, the positions of the C<sub>17</sub> nucleotide in the PDB: 1NYI structure was almost completely superimposable on that found for the *talo*-5'-C-methyl modified RNA of PDB: 379D, allowing the argument that both structures were possible intermediates on the correct pathway to a hammerhead cleavage-site transition state conformation. The reference 58 authors argue that the phosphodiester tether that forms



See key to figures 6.10 and 6.11.

**Figure 6.20** Secondary structure of the tethered hammerhead ribozyme RNA 6 used for X-ray crystallographic structures deposited in the Protein Data Bank as PDB: 1NYI and 1Q29.

during crystallization creates a kinetic bottleneck preventing cleavage but allowing a conformational change (at C<sub>17</sub>) that precedes catalytic cleavage.

Next, the researchers raised the pH in the crystal from 6 to 8.5, the pH found previously to be slightly above the apparent pK<sub>a</sub> of the cleavage reaction. It was hoped that this would drive most of the RNA molecules in the crystal into their catalytically active conformation. Crystals that formed were treated by flash-freezing and diffraction data were collected at 100K as above. These data were deposited as PDB: 1Q29.<sup>58</sup> This crystal structure showed that the C–G hydrogen-bonding tether remained intact, although possibly more strained. At the cleavage site in the PDB: 1Q29 crystal, conformational changes had taken place to achieve an intermediate that had been activated for catalytic cleavage but unable to cleave because of the stem I–stem II tether. The conformation was found to have an “inline fitness” parameter measuring 1.6. This compared to an inline fitness parameter calculated for PDB: 1NYI crystal structure measuring 0.06. As discussed previously, the inline parameter (ideal value 1.0) assesses the ability of the conformation to achieve a cleavage site alignment proper for the S<sub>N</sub>2 mechanism—lining up the O2' (C<sub>17</sub>)–P (A<sub>1,1</sub>)–O5' (A<sub>1,1</sub>) atoms at the correct 180° angle and bringing the O2' (C<sub>17</sub>)–P (A<sub>1,1</sub>) distance to approximately 3.0 Å. Table 6.6 collects the data for PDB: 1NYI and PDB: 1Q29 and compares them to other X-ray crystallographic structures published by the Scott group.

If the reader is using PDB data to visualize these structures, the following nucleotide numbering system will be useful. For PDB: 1NYI, Chain A = ribozyme and Chain B = substrate, C<sub>17</sub> = C120, A<sub>1,1</sub> = A121, and a C–G hydrogen-bonding pair forms between C26 and G500: N<sub>4</sub>–O<sub>6</sub> = 3.30 Å, N<sub>1</sub>–N<sub>3</sub> = 3.12 Å, O<sub>2</sub>–N<sub>2</sub> = 2.91 Å. For PDB: 1Q29, Chain A = ribozyme and Chain B = substrate, C<sub>17</sub> = C120, A<sub>1,1</sub> = A121, and a C–G hydrogen-bonding pair forms between C26 and G10: N<sub>4</sub>–O<sub>6</sub> = 3.10 Å, N<sub>1</sub>–N<sub>3</sub> = 3.11 Å, O<sub>2</sub>–N<sub>2</sub> = 3.14 Å. Fruitful comparison can be made between the PDB: 1Q29-unmodified but tethered RNA “late intermediate” structure (angle = 135°, O2' (C<sub>17</sub>) – P (A<sub>1,1</sub>) = 2.24 Å and fitness = 1.6) and that previously published—PDB: 379D with *talo*-5'-C-methyl modified leaving group at A<sub>1,1</sub> (angle = 111°, O2' (C<sub>17</sub>) – P (A<sub>1,1</sub>) = 2.51 Å and fitness = 0.84). Obviously, the tether is preventing the PDB: 1Q29 structure from achieving the ideal 170–180° angle required for cleavage. Additional differences can be noted in comparison to the hairpin ribozyme ligation site where the 172° angle is almost ideal and the O2'–P distance equal to 2.70 Å indicates atoms approaching for bond formation. Comparison of the tethered late intermediate to the cleaved product PDB: 488D shows that the angle between stem I and stem II widens significantly during cleavage. The reference 58 authors believe that the tether prevents stem I from moving relative to stem II and thus prevents cleavage by restricting the helical motion and unwinding that stem I would undergo during cleavage. They believe that the crystalline lattice also prevents unwinding and that the stem I helical motion and unwinding would be much greater in solution. Finally, they hypothesize that structural interactions between stem I and stem II in hammerhead self-cleaving RNA

**TABLE 6.6** Inline Fitness Data for Hammerhead Ribozyme Crystal Structures<sup>a</sup>

| PDB Accession Number                                 | O2' (C <sub>17</sub> )–P (A <sub>1,1</sub> )–O5' (A <sub>1,1</sub> ) Angle (°) | O2' (C <sub>17</sub> )–P (A <sub>1,1</sub> ) (Å) | O5' (A <sub>1,1</sub> )–P (A <sub>1,1</sub> ) (Å) | O2' (C <sub>17</sub> )–O5' (A <sub>1,1</sub> ) (Å) | Inline Fitness | Reference/Year               |
|--|--|--|---|--|----------------|------------------------------|
| 299D   | 56.9   | 3.46   | 1.59  | 2.91   |                | 39/1996                      |
| 300D   | 60.0   | 3.51   | 1.59  | 3.04   |                | 39/1996                      |
| 301D   | 74.9   | 4.08   | 1.59  | 3.98   |                | 39/1996                      |
| 379D with P modeled into the structure               | 111  | 2.51   | 1.59  | 3.42   | 0.84           | 58/2003 Table 2 <sup>b</sup> |
| 488D uncleaved                                       | 99.4   | 4.37   | 1.59  | 4.89   |                | 52/2000                      |
| 488D cleaved   |  | 1.7  |   |  |                | 52/2000                      |
| 1NYI   | 59.5   | 3.60   | 1.59  | 3.11   | 0.06           | 58/2003 Table 2 <sup>b</sup> |
| 1Q29   | 135  | 2.24   | 1.60  | 3.56   | 1.6            | 58/2003 Table 2 <sup>b</sup> |
| Ideal “Inline Fitness”                               | 180  | 3.00   | 1.59  | 4.59   | 1.0            | 58/2003 Table 2 <sup>b</sup> |
| 299D A <sub>9</sub> –G <sub>8</sub> noncleaving site | 170  | 2.66   | 1.59  | 4.23   | 1.3            | 58/2003 Table 2 <sup>b</sup> |
| Hairpin ribozyme ligation site                       | 172  | 2.70   | 1.59  | 4.28   | 1.3            | 58/2003 Table 2 <sup>b</sup> |

<sup>a</sup>Some data are from Table 2 of reference 58.<sup>b</sup>Note that distance parameter labels in lines 1, *d*(O2'-P), and 2, *d*(O5'-P), are labeled incorrectly in Table 2 of reference 58.

could constitute a biochemical switch that might control whether the RNA functions as a nuclease (cleavage) or a ligase (reversal of the cleavage reaction).

In 2005, Kenneth Blount and Olke Uhlenbeck published an extensive review of biochemical, spectroscopic, and structural research results obtained for the hammerhead ribozyme up to that date.<sup>61</sup> The review is aptly titled “The Structure–Function Dilemma of the Hammerhead Ribozyme.” In the abstract to this article the authors state that the structural and functional data obtained so far for the hammerhead ribozyme do not agree well and that these disagreements have prevented a unified catalytic mechanism from emerging. In the article, they define a “consensus set of functional groups unambiguously required for hammerhead catalysis.” Looking at this set of functional groups in relation to the available crystal structures helped the reference 61 authors define a concise set of disagreements between structural and functional data.

First, these authors describe the natural hammerhead ribozyme as a unimolecular species containing helices I, II, and III. Helices I and II are closed by loops. The constructs used in *in vitro* experiments as described in the above discussion are all bimolecular species. The “substrate” strand contains the scissile phosphate, while the other strand is called the “ribozyme.” A typical hammerhead contains a 15-residue conserved catalytic core domain—adjacent to the cleavage site between residues 17 and 1.1 on the substrate strand—consisting of residues C<sub>3</sub>, U<sub>4</sub>, G<sub>5</sub>, A<sub>6</sub>, and U<sub>7</sub> (not conserved) and G<sub>8</sub>, A<sub>9</sub>, G<sub>10,1</sub>, C<sub>11,1</sub>, G<sub>12</sub>, A<sub>13</sub>, A<sub>14</sub>, A<sub>15</sub>, U<sub>16,1</sub>, and C<sub>17</sub> (not conserved). Hammerhead catalysis is described by three steps in a manner similar to that for protein enzymes: (1) association of ribozyme and substrate; (2) the chemical cleavage step; and (3) dissociation of the ribozyme–products complex. The chemical cleavage step, a transesterification, is uniform for all hammerheads. The reverse reaction, a ligation, is also catalyzed but at a 100-fold lower rate. The chemical cleavage step takes place when the 2’OH of substrate residue 17.0 is deprotonated and attacks the phosphorus of residue 1.1, forming a five-coordinate trigonal bipyramidal transition state. The 3’-product consists of the nucleotide 1.1 ribose’s 5’-oxygen as the leaving group that is expelled and subsequently protonated. The 5’-product of the cleavage is a 2’,3’-cyclic phosphate (see Figure 6.9). Because the chirality of the scissile phosphate is inverted over the course of the reaction, an S<sub>N</sub>2(P) mechanism is implied in which the 2’-oxygen nucleophile must be in line with the scissile phosphorus-5’-oxygen bond for cleavage to occur. Hammerhead cleavage may be catalyzed through (1) general base catalysis—deprotonation of the 2’-OH; (2) general acid catalysis—protonation of the 5’-O leaving group; and (3) transition state stabilization—possibly by negative charge reduction through interaction with cations. Magnesium ions—for example, coordinated near the cleavage site—are proposed to fill this role in catalysis enhancement.

Using an approach that has been successful with protein enzymes and other ribozymes, researchers have used hammerhead X-ray crystallographic structures to identify sites around the catalytic center where mutation might yield

information on function and mechanism of catalysis. In contrast to the success of this method for other systems, analyses of hammerhead structure–function relationships has led to disagreement and controversy.

Hammerhead crystal structures (identified by reference and PDB accession number in Table 6.3) have identified two important domains: (1) domain I consists of the substrate's cleavage site ( $C_{17}$  and  $X_{1,1}$ , where X stands for any nucleotide) and the ribozyme's  $C_3U_4G_5A_6$  loop (called the CUGA turn) and (2) domain II consists of 10 residues connecting helices II and III— $U_{16,1}$ ,  $U_7$ ,  $G_8$ ,  $A_9$ ,  $G_{10,1}$ ,  $C_{11,1}$ ,  $G_{12}$ ,  $A_{13}$ ,  $A_{14}$ ,  $A_{15,1}$ . All of these residues are conserved and essential for all hammerheads, except for  $U_7$ . In the domain II group, there are four adjacent non-Watson–Crick base pairs including two mismatch pairs:  $A_{13}$ – $G_8$ ,  $A_9$ – $G_{12}$  that appear next to each other (as is the case for many ribosomal RNA secondary structures). The two domains are connected at a three-strand junction comprised of nucleotides  $A_6$ ,  $U_7$ ,  $U_{16,1}$ , and  $C_{17}$ . Overall, the hammerhead is Y-shaped with coaxial helices II and III and helices (stems) I and II alongside each other and roughly parallel.

Many hydrogen bonding and interaction characteristics in the hammerhead crystal structures as described by Blount and Uhlenbeck in reference 61 relate to possible functional details. Residues in domain II form non-Watson–Crick base pairs whereas one pair,  $A_{15,1}$ – $U_{16,1}$ , forms only one H-bond, widening the major groove to facilitate formation of the three-strand junction. Cleavage site residue  $C_{17}$  interacts with residues in the CUGA turn, possibly having implications for the catalytic mechanism. In the initial all-RNA hammerhead structure, PDB: 1MME, published in 1995, the 2'-oxygen nucleophile is 90° away from the proper alignment to carry out the  $S_N2(P)$  cleavage mechanism. Since then, as described in the previous discussion, the Scott group has changed crystallization conditions and adjusted the hammerhead construct to increase the attack angle—achieving a  $O2'$  ( $C_{17}$ )–P ( $A_{1,1}$ )– $O5'$  angle of 135°, a  $O2'$  ( $C_{17}$ )–P ( $A_{1,1}$ ) distance of 2.24 Å, and an “inline fitness” parameter of 1.6 in the PDB: 1Q29 structure.

The important presence of metal ions has also been addressed in crystal structure determinations. Since magnesium ions had been found to enhance the cleavage reaction, it was important to find where these ions were placed in the structure. Soaking crystals in  $Mg^{2+}$  ions positioned five  $Mg^{2+}$  ions in the hammerhead ribozyme, as has been described above in the discussion of PDB: 1MME. Importantly, ions were found to be interacting with nucleotides near the catalytic site and in domain II with residues  $A_9$  and  $G_{10,1}$ . Analysis of metal ion placement in the Scott group hammerhead crystal structures has been collected in Table 6.3. Many, but not all, structures where the pH was maintained at 8.5 show close metal ion interactions with important domain II residues  $A_9$  and  $G_{10,1}$ . The points of metal ion attachment at the pro- $R_p$ -oxygen of  $A_9$  and  $N_7$  of  $G_{10,1}$  agree well with the essentiality of these atoms found in biochemical and spectroscopic experiments (see especially Figure 4 of reference 61). Close metal ion interactions are often seen for the  $G_{5,0}$  residue that is part of the CUGA loop of domain I, although few of them connect the metal

ion to the essential N<sub>2</sub>, N<sub>1</sub>, or O<sub>6</sub> positions of the G<sub>5,0</sub> guanosine. Metal ion interactions are almost never seen at the essential positions at the cleavage site—that is, the pro-R<sub>p</sub>-oxygen of A<sub>1,1</sub> or the ribose O2' or O<sub>2</sub> positions of C<sub>17,0</sub>. One interaction in the PDB: 301D structure places a Mg<sup>2+</sup> within 2.4 Å of the pro-S<sub>p</sub>-oxygen of A<sub>1,1</sub>, an atom not found to be essential in biochemical or spectroscopic experiments.

Blount and Uhlenbeck<sup>61</sup> state that the global hammerhead fold in the crystal structures agrees with studies carried out in solution including fluorescence resonance energy transfer (FRET), native gel electrophoresis, transient electric birefringence, and nuclear magnetic resonance (NMR) experiments. Some of these have been discussed previously in this section. However, atomic-level detail of a hammerhead solution structure, especially in the core region, has not been possible. Probably this is due to the highly dynamic nature of the core structure in solution that likely exists as a family of rapidly interconverting different conformations. In fact, a <sup>1</sup>H- and <sup>31</sup>P-NMR study showed that the core region near the essential pro-R<sub>p</sub>-oxygen of A<sub>9</sub> was inherently flexible and that this flexibility could be important for hammerhead cleavage.<sup>62</sup>

In the reference 61 review, the authors establish criteria for evaluation of the many experiments that attempt to define the essential functional groups for hammerhead catalysis. These are as follows: (1) The measured cleavage rate must reflect the chemical step of the reaction,  $k_2$ , and not a conformational change, formation of the substrate–ribozyme complex, or release of the ribozyme–product complex; (2) the hammerhead construct used must have acceptable kinetic properties; and (3) the chemical nature of the functional group modification should not, by itself, cause conformational or mechanistic changes—that is, the pK<sub>a</sub> for the reactions should be the same before and after the modification. To measure  $k_2$  successfully, one must measure the cleavage rate under single-turnover conditions. Kinetically well-behaved hammerheads are hammerheads 6, 8, and 16 (see Figures 6.10, 6.11, and 6.12). In most cases, only one modification to a nucleotide should be undertaken at a time; otherwise, it is difficult to pinpoint the exact nucleotide position causing observed changes. By evaluating available biochemical data according to these criteria, the reference 61 authors established a consensus set of functional groups unambiguously important for hammerhead catalysis. These are collected in Figure 4 and Table 2 of reference 61. If one maps the collected important functional groups—essentially from those residues defined as belonging to domain I and domain II—one finds a set of atoms lining the catalytic pocket starting at G<sub>10,1</sub> in domain II (relatively far from, or distal to, the catalytic site) and ending at the cleavage site, C<sub>17,0</sub>, in domain I. The mapping exercise helps in visualizing why nearby essential functional groups could have important catalytic effects but raises questions as to why far-away functional groups (up to 20 Å away) in domain II should have >100-fold effects on catalytic cleavage. The review authors find 21 essential functional groups in domain II and 13 in domain I. Only four 2'-hydroxyls are essential—those at G<sub>5</sub>, G<sub>8</sub>, C<sub>16,1</sub>, and C<sub>17</sub>. Only four phosphate oxygens appear to be essential—those 5' to X<sub>1,1</sub>, A<sub>9</sub>, A<sub>13</sub>,

and A<sub>14</sub>. As mentioned many times before, X<sub>1,1</sub> and A<sub>9</sub> are important metal ion binding sites. With the catalytic core, only U<sub>7</sub> appears to have no essential functional group.

Next, the reference 61 authors attempted to define some structure–function relationships, based on defined charge–charge, van der Waals stacking, and hydrogen-bonding crystal structure interactions in concert with biochemical and spectroscopic experimental results. They believe that far away, domain II essential functional groups cannot be intimately involved at the active site but rather are important for structural interactions. This means that the hammerhead global fold must be more critical for catalysis than for most other enzymes. One can place hammerhead functional groups in three categories of structure–function correlation: (1) structure and function data that agree; (2) interactions predicted by structural information that cannot be confirmed by functional data; and (3) deleterious functional group modifications that have no apparent basis in structural details. Of course, the last is the most disturbing factor. For instance, G<sub>5,0</sub> in domain I exhibits three essential functional groups at N<sub>2</sub>, N<sub>1</sub>, and O<sub>6</sub>, although it appears to have no interaction partner in the crystal structures. Note, however, that these G<sub>5,0</sub> functional groups do show interactions with metal ions in the crystal structures represented by PDB: 300D, 301D, 1NYI, and 1Q29. In domain II, 10 different essential functional groups on six different nucleotides have no apparent partner in the crystal structures. A consistent metal ion bonding site appearing in most crystal structures involves the pro-R<sub>p</sub>-oxygen of A<sub>9</sub> and the N<sub>7</sub> essential atom of G<sub>10,1</sub>, but little metal ion interaction is seen for residues C<sub>11,1</sub>, G<sub>12</sub>, A<sub>13</sub>, A<sub>14</sub>, or A<sub>15,1</sub>. The PDB: 1NYI structure shows an interaction between Co<sup>2+</sup> and the essential 2'-OH of U<sub>16,1</sub>, and several interactions with G<sub>16,2</sub> just outside domain II are noted in Table 6.3. In category (1), in which structure and function agree, one can site several hydrogen-bonding interactions that appear in the crystal structure between essential functional groups. One of these is between O<sub>2'</sub> of G<sub>8</sub> and the N<sub>2</sub> of G<sub>12</sub> with a bonding distance of 2.4–3.0 Å in most crystal structures. Examples of category (2) include hydrogen-bonding partners found in the crystal structure that have only one essential functional group as part of the hydrogen bond. For instance, the O<sub>2'</sub> of G<sub>5,0</sub> (essential) forms an H-bond with O<sub>2'</sub> of U<sub>15,2</sub> (non-essential) in crystal structures. Lastly, it should be noted that structural–functional agreement for the essential nature of the pro-R<sub>p</sub>-oxygen of A<sub>9</sub> and the N<sub>7</sub> atom of G<sub>10,1</sub> (many crystal structures show metal ion bonding at this location), along with the fact that this site resides approximately 20 Å from the cleavage site in every crystal structure and never gets closer to the cleavage site, raises unanswered questions as to what role is played by this domain II location.

Blount and Uhlenbeck<sup>61</sup> next discuss the progression of crystal structures published by the Scott group that have been extensively discussed previously in this section. They agree that the structures show a progression of conformation changes at the cleavage site, especially in changes in orientation of the C<sub>17</sub> nucleotide and achievement of a proper angle for the inline attack needed

for the S<sub>N</sub>2 mechanism. However, they maintain that it is still unclear whether these changes are on the actual hammerhead reaction pathway. For instance, most of the changes in the crystal structures appear near (proximal to) the cleavage site, whereas few changes are observed for essential functional groups in domain II distal to the cleavage site. The reviewers suggest that the Herschlag group metal ion rescue experiments<sup>45</sup> may be correct in stating that the same Mg<sup>2+</sup> coordinates to the A<sub>9</sub>/G<sub>10.1</sub> and the scissile phosphate in the transition state even though these sites are 20 Å apart in the crystal structures (see Figure 9 of reference 61). Metal ion rescue experiments and the one-metal theory have already been discussed in relation to suggestions by the Scott group that conformational changes to this transition state are energetically too costly.<sup>49</sup> In support of the Herschlag hypothesis, hydroxyl radical footprinting experiments suggest a more compact and less solvent-exposed conformation for the hammerhead in solution compared to conformations found in the crystal structures.<sup>56</sup>

Blount and Uhlenbeck believe that answers may be found to the above-described inconsistencies for the hammerhead ribozyme by studying naturally occurring hammerheads in which the closing loops of stems I and II interact with each other to produce a much more compact and closely packed hammerhead active conformation.<sup>63</sup> The tertiary interaction between the stems appears to lower the Mg<sup>2+</sup> concentration required for full catalytic activity. The reference 61 and 63 authors believe that further study of the naturally occurring hammerheads that exhibit close conformational interactions between domains I and II may lead to a more unifying view of the catalytic cleavage mechanism.

In 2006, the 2.2-Å resolution crystal structure of a full-length *Schistosoma mansoni* hammerhead ribozyme was determined by Martick and Scott.<sup>64</sup> (PDB: 2GOZ). The researchers crystallized a full-length hammerhead ribozyme of 43 nucleotides (nt) in complex with a 20-nt substrate. No metal ions are present in the X-ray crystallographic structure. The full-length hammerhead ribozyme contained the rate-enhancing peripheral domain and had a catalytic core that was very different from the catalytic core present in the structure of the “minimal” hammerhead constructs that have been discussed here—that is, all the previously discussed hammerhead constructs have lacked the peripheral domain present in PDB: 1GOZ.<sup>65</sup> The X-ray structure revealed how tertiary interactions that occur in regions remote from the active site prime the ribozyme for catalysis. For instance, nucleotides G<sub>12</sub> and G<sub>8</sub> are positioned in PDB: 1GOZ to carry out their previously suggested roles as acid–base catalysts. The nucleophile is aligned for an S<sub>N</sub>2 reaction with a scissile phosphate positioned proximal to the A<sub>9</sub> phosphate. Previously unexplained roles of other conserved nucleotides became apparent when observing the distinctly new fold in the PDB: 1GOZ structure. The newly observed ribozyme–substrate interactions allowed the reference 64 authors to explain many of the previously conflicting experimental results. Interested readers should consult reference 64 for more detail.

This section has concentrated on the role of magnesium ions in catalytic RNA. Other important topics related to biological magnesium ion chemistry—in kinases, ATPases, and large systems such as chlorophyll—will not be treated here. In general, cations enable catalysis in the following ways: (1) as general bases perhaps as metal-hydroxo species, activating the nucleophile; (2) as general acids perhaps as metal-aqua species, aiding in the protonation of the leaving group; and (3) as charge neutralizers, stabilizing transition states. The first two criteria could be considered as functional roles whereas the third would be a structural as well as a functional role. For the group I intron, magnesium ions play an important structural role in bringing intron and exon together productively. In a functional manner, magnesium ions are involved in the activation of the guanosine nucleophile in an analogous manner to metal cofactor activation in protein metalloenzymes. Finding out whether a metal ion plays solely a structural role, or also takes part in the actual catalytic activity, is a difficult experimental problem. The phosphorothioate metal rescue (PS-rescue) experiment described in the preceding sections (see Section 6.2.2, for instance) can be successful in predicting sites of catalytic activity, although the situation is complicated by the fact that ribozyme active sites undergo several conformational changes between ground, pre-transition, and transition states. Although the number of magnesium ions playing a functional role in the group I intron may be somewhat uncertain, researchers are in general agreement that  $Mg^{2+}$  has a catalytic role in this ribozyme. At the time of this writing, the role of metal ions in hammerhead ribozyme catalysis is much less certain. Probably metal ions, specifically  $Mg^{2+}$ , play a structural role in assisting rapid and efficient ribozyme folding. Whether or not these ions play a functional/catalytic role remains a question requiring further research.

## 6.3 CALCIUM-DEPENDENT MOLECULES

### 6.3.1 Introduction

Calcium homeostasis has been discussed in Section 5.2.3. The present and following sections will provide some detail on two calcium-containing biomolecules: calmodulin and calcium ATPase. Binding of  $Ca^{2+}$  to calmodulin (Section 6.3.2) activates enzymes such as protein kinases.  $Ca^{2+}$ -ATPases (Section 6.4.2) are cross-membrane calcium pumps that bind and release calcium ions to foster muscle relaxation. Other proteins and enzymes requiring calcium ions are, for example, (1) Troponin C—binding of three to four  $Ca^{2+}$  ions results in muscle contraction; (2) extracellular digestive enzymes such as staphylococcal nuclease (one calcium site), phospholipase A<sub>2</sub> (two calcium sites), and trypsin (one calcium site); (3) structural and storage enzymes that sequester many calcium ions per protein unit including thrombin, phosphodentine (material in teeth), and calsequestrin (calcium storage in the sarcoplasmic reticulum). Calsequestrin is the major calcium-buffering protein localized in the lumen of

adult cardiac SR. It binds calcium with high capacity and low affinity during muscle relaxation. Table 6.6 of reference 1 collects some information on calcium-binding proteins. Chapter 6 of reference 66 discusses calcium-containing biomolecules and its Chapter 7 discusses calcium ions as secondary messengers.

### 6.3.2 Calmodulin

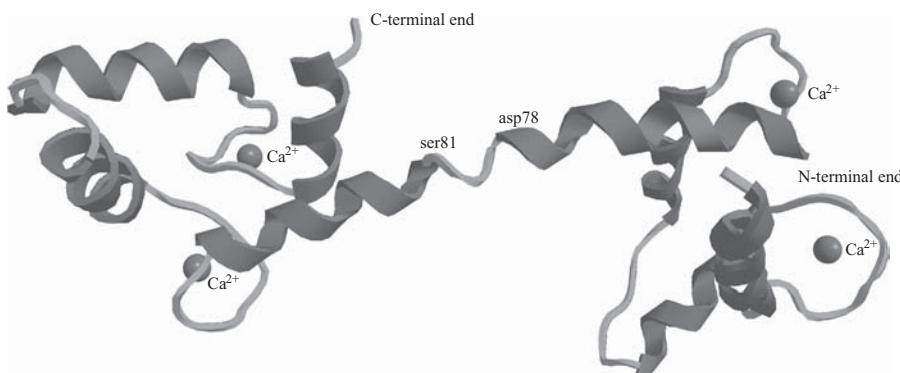
**6.3.2.1 Introduction.** Calmodulin (CaM) is a small acidic protein of 148 amino acid residues weighing approximately 17 kDa. CaM is present in all eukaryotic cells, serves as a primary receptor for intracellular  $\text{Ca}^{2+}$ , and acts as a switch in cellular processes controlled by  $\text{Ca}^{2+}$ -dependent signaling pathways. In the calcium-saturated state,  $\text{Ca}^{2+}\text{-CaM}$  can interact with and modulate many different proteins including adenylate cyclase, nitric oxide synthetases, phosphodiesterases, and calcium pumps (to be discussed in the Section 6.4.2), as well as assorted kinases and phosphatases.<sup>67</sup>  $\text{Ca}^{2+}\text{-CaM}$  activates its target proteins by interacting with peptide sequences within the targets that have a high degree of helicity when they are bound to calmodulin.<sup>68</sup> Because calmodulin binds to so many different proteins, its structure must carry a great deal of flexibility and this factor will become evident in the discussion that follows. The binding of calcium ions causes major changes in calmodulin's tertiary structure, exposing hydrophobic surfaces for interaction with enzyme target substrates. To understand CaM's activity, researchers have studied binding of drug molecules (agonists or antagonists) or small peptides to CaM. An agonist may be defined as a drug that binds to a receptor of a cell and triggers a response by the cell. An agonist often mimics the action of a naturally occurring substance. An antagonist may be defined as a drug that binds to a cell receptor without eliciting a biological response.

In calmodulin, four calcium ions bind to four EF-hand motifs with micromolar affinity. The calcium ion binding lobes have strong resemblance to those of troponin C (TnC). Troponin researchers note that the arrangement of troponin's TnC lobes strongly resembles those of calmodulin, especially after rearrangement of TnC induced by TnC/TnI<sub>1-47</sub> complex formation.<sup>69</sup> Troponin consists of three subunits: (1) Troponin C, the  $\text{Ca}^{2+}$  binding subunit (TnC, 18.4 kDa), Troponin I, the inhibitory subunit (TnI, ~20–25 kDa), and Troponin T, the tropomyosin-binding subunit (TnT, ~31–33 kDa). In muscle, troponin and tropomyosin units bind to the actin-containing muscle thin filament. Polymerized actin forms the backbone of this thin filament. Troponin and tropomyosin units are located at points along polymerized actin at a troponin:tropomyosin:actin ratio of 1:1:7.<sup>70</sup> The two proteins, calmodulin and troponin, have distinct biological functions. Calmodulin functions as an isolated protein and both of its binding lobes serve signaling functions. In contrast, TnC is part of the larger troponin complex and only its N-terminal lobes (sites I and II) play a regulatory role. TnC's C-terminal binding sites (sites III and IV) play a structural role and integrate TnC into the rest of the troponin complex.

Before continuing the discussion of calmodulin's properties and activities *in vivo*, it is important to note that excerpts from the careful reading of primary sources have been used to inform readers on this topic. These sources range in age from the mid- to late 1980s to late 2005 and encompass the work of many different laboratories. Over these 20+ years, methods and terminology have changed substantially. Terminology will create the greatest sources of confusion in the following discussion. For instance, one finds references to calmodulin's helical domains using the letters A through H but also Roman numerals I through VIII. In one 2002 paper, the authors designate the D and E central linker helices as both having the Roman numeral IV.<sup>91</sup> In this system, calmodulin's helical domains have numerals I through VII. Fortunately, the numbering system for calmodulin amino acid residues 1–148 seems to be constant over all sources. Another numbering system creating confusion will arise in the discussions of target enzyme peptide binding domains as they bond to calmodulin. At least two numbering systems are used to identify the peptide residue sequence in the discussions that follow, and these are outlined in Table 6.9. This author has tried to address the confusion in the terminology used, but readers consulting the primary sources referenced here, or newer sources in calmodulin's continuing story, should expect to spend time relating one author's terminology to that of another.

**6.3.2.2 Calmodulin Structure by X-Ray and NMR.** Babu and co-workers published the first X-ray crystallographic structure of  $\text{Ca}^{2+}$ -saturated mammalian calmodulin in 1986 (PDB ascension number 3CLN).<sup>71</sup> As can be seen in Figure 6.21, the molecule is dumbbell-shaped, with an overall length of 65 Å. There are two two-EF-hand domains or lobes—N-terminal, residues 5–73 and C-terminal, residues 74–148—connected by a central seven-turn  $\alpha$ -helix. The center part of the connecting helix is unwound between residues asp78 and

Calmodulin PDB: 3CLN

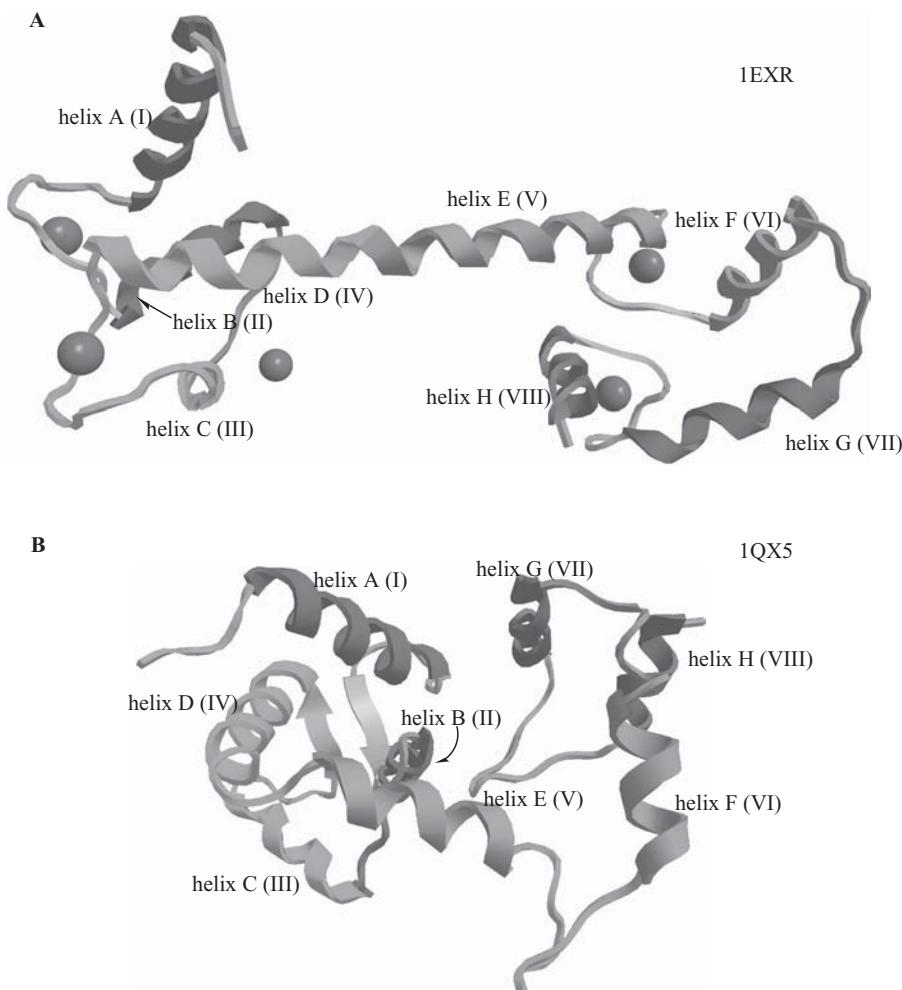


**Figure 6.21** Calmodulin dumbbell shape (PDB: 3CLN). Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. (Printed with permission of CambridgeSoft Corporation.)

ser81. Calmodulin's four  $\text{Ca}^{2+}$ -binding domains have a typical EF hand conformation (helix-loop-helix) and are similar to those described for other  $\text{Ca}^{2+}$ -binding proteins. Each of the four EF-hand domains contains a  $\text{Ca}^{2+}$ -binding loop, and two short, double-stranded antiparallel  $\beta$ -sheets between pairs of adjacent  $\text{Ca}^{2+}$ -binding loops.

The PDB: 3CLN X-ray structure shows a large hydrophobic cleft in each half of the molecule. These hydrophobic regions are believed to represent the sites of interaction with many of the pharmacological agents and target peptides known to bind to calmodulin. Some of these will be described below.

A more recent crystal structure at 1.0 Å resolution of calcium-saturated calmodulin from *Paramecium tetraurelia* has been presented (PDB: 1EXR).<sup>67</sup> The reference 67 authors observe that binding of calmodulin to over 100 different proteins must require a considerable degree of structural plasticity. Indeed, they find at least 36 disordered amino acid residues in their structure—a surprisingly large 24% of the entire protein. The authors state that atomic-resolution structures (resolution of approximately 1.2 Å) usually contain between 6% and 15% of residues in alternative conformations. In this calmodulin structure the central helix contains 10 residues in alternate conformations, including ser70, met72, ser81, glu87, and val85. There are two binding pockets containing mostly hydrophobic amino acid residues, one for each calcium-binding lobe of the protein. These pockets contain at least 16 residues in alternative conformations and others in various disordered modes. The hydrophobic cleft regions in calmodulin's calcium-binding lobes are phenylalanine- and methionine-rich, and these residues provide hydrophobic binding capability for target molecules and peptides. The reference 67 authors state that specific phenylalanine (12, 68, 92, and 141) and methionine (36, 72, 76, and 144) residues present a deformable and polarizable surface that can adapt to bond incoming target molecules. Analyses by these researchers indicate a large number of conformational calmodulin substates in the crystalline environment at the cryogenic temperature of their work (data collected at 100 K), and they also infer that there might be an almost continuously changing set of calmodulin conformations in solution at physiological temperatures. (See the discussion of calmodulin structures that have been determined using NMR methods below.) In addition to large conformation changes observed in going from apo-calmodulin to calcium-saturated states (see Figure 6.22), they note that calcium-saturated calmodulin undergoes additional conformational changes when binding to (wrapping around) its targets. In fact, these additional changes optimize calmodulin's surface complementarity with its targets. The 28 residue central helix (helix D/E) is intimately involved in target recognition and binding because of its exposure to solvent as well as the fact that it can also adopt a large variety of conformations depending on the binding requirements in the calcium-binding lobes. Figure 6.22A illustrates the approximate 90° angles between helices A and B, C and D, E and F, and G and H in holo ( $\text{Ca}^{2+}$ -saturated) calmodulin for the PDB: 1EXR structure. The secondary structure for PDB: 1EXR is identified as follows: (1) helices A (residues 2–18)



**Figure 6.22** X-ray crystallographic structures of: (A) Ca<sup>2+</sup>-saturated calmodulin, PDB: 1EXR; and (B) apo-calmodulin, PDB: 1QX5. Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. (Printed with permission of CambridgeSoft Corporation.) (See color plate)

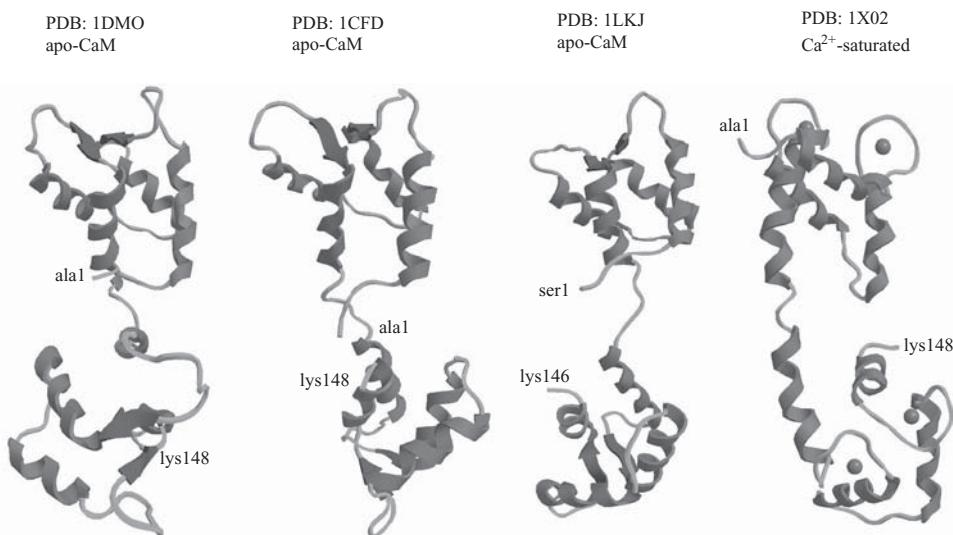
and B (28–39) shown in blue; (2) helix B to helix C connector loop (39–44) shown in orange; (3) helices C (45–55) and D (65–76) shown in green; (4) antiparallel  $\beta$ -sheet connector regions shown in cyan (thr26–thr28, thr62–asp64); (5) helices E (77–92) and F (102–111) shown in magenta; (6) helix F to helix G connector loop (112–117) shown in orange; (7) helices G (118–128) and H (139–146) shown in red; antiparallel  $\beta$ -sheet region shown in cyan (99–101, 135–137). The connector loops between helices in the helix-loop-helix calcium-binding domains are shown in green. See also Table 6.7 for interhelical angles.

The X-ray crystallographic structure of human calmodulin, refining to 1.7-Å resolution, was obtained in 1992 (PDB: 1CLL).<sup>72</sup> The interhelical angle data are included in Table 6.7. The X-ray crystallographic structure of apo-calmodulin—one in which calcium ions are not present—was carried out by Schumacher et al. (PDB: 1QX5).<sup>73</sup> As can be seen from Figure 6.22B, the apo structure consists of more compact coiled globular domains, whereas the calcium-saturated structure (PDB: 1EXR) shows the characteristic calmodulin dumbbell shape. The coloring in Figure 6.22B is the same as that noted above for Figure 6.22A. In the apo-calmodulin structure (PDB: 1QX5), the central E and F helices are not contiguous, helices are shorter, and the loop regions between helices are longer. Note that for the holo-calmodulin structure (PDB: 1EXR), helices A and B (and helices C and D) have opened up relative to each other, forming the hydrophobic cleft where drug molecules or incoming target enzymes can bind.

The solution structure of  $\text{Ca}^{2+}$ -free calmodulin has been determined for several species. In 1995, Zhang, Tanaka, and Ikura published the apo-calmodulin NMR structure for *Xenopus laevis* (PDB: 1DMO).<sup>74</sup> Another apo-calmodulin NMR structure for *Xenopus laevis* was published in the same year by the Ad Bax group (PDB: 1CFC, 25 structures, and 1CFD, averaged structure).<sup>75</sup> These researchers used three- and four-dimensional heteronuclear NMR experiments, including ROE, isotope-filtering combined with reverse labeling, and measurement of more than 700 three-bond J-couplings. (See Sections 3.4.3, 3.4.10, and 3.4.11 for descriptions of these NMR techniques.) As is true for the PDB: 1DMO structure, the PDB: 1CFD protein consists of two globular domains separated by a flexible linker, with no stable, direct contacts between the two domains. In the absence of  $\text{Ca}^{2+}$ , the four helices in each of the two globular domains form a highly twisted bundle, each capped by a short antiparallel  $\beta$ -sheet. (This arrangement is qualitatively similar to that observed in the crystal structure of the  $\text{Ca}^{2+}$ -free N-terminal domain of troponin C.) Other research groups have determined solution structures of apo-calmodulin.<sup>76</sup> In all cases, removal of  $\text{Ca}^{2+}$  ions from the structure changes the conformation of calmodulin to close the hydrophobic pockets that would bind to target enzymes.

More recently, workers in Japan published the solution structure of yeast (*Saccharomyces cerevisiae*) apo-calmodulin (PDB: 1LKJ).<sup>77</sup> Yeast calmodulin is 60% identical in its amino acid sequence with vertebrate CaMs. The 1LKJ N-terminal domain with its two helix-loop-helix calcium-binding domains looks quite similar to those of 1DMO and 1CFD (see Figure 6.23).

The central linker region is more or less unwound and different for each of the three apo-calmodulins discussed here and shown in Figure 6.23. In addition, the C-terminal region takes different conformations in each structure. While the three illustrated C-terminal sections still each consist of two basic helix-loop-helix calcium-binding domains, in (yeast) PDB: 1LKJ the segment corresponding to the fourth  $\text{Ca}^{2+}$ -binding site deviates in its primary structure from a typical EF-hand motif and loses the ability to bind calcium ion. Note



**Figure 6.23** Comparison of three different apo-calmodulin NMR structures (PDB: 1DMO, 1CFD, 1LKJ) and one  $\text{Ca}^{2+}$ -saturated calmodulin (PDB: 1X02). Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. (Printed with permission of CambridgeSoft Corporation.)

that the dumbbell-shaped solution structure of all apo-calmodulins discussed here are quite different from the globular appearance of the X-ray structure apo-calmodulin discussed above (PDB: 1QX5, Figure 6.22B), although the central linker region remains similar as far as its unwound state is concerned. In all cases, removal of  $\text{Ca}^{2+}$  ions from the structure changes the conformation of calmodulin to close the hydrophobic pockets that would bind to target enzymes. Figure 6.23 also shows the secondary structure of a  $\text{Ca}^{2+}$ -saturated calmodulin (PDB: 1X02). This structure, determined in 2006, made use of a new NMR technique called stereo-array isotope labeling (SAIL).<sup>78</sup> The technique uses a stereospecific and regiospecific pattern of stable isotopes that is optimized for the NMR spectra being collected. The SAIL technique offers sharpened lines, spectral simplification without loss of information, and the ability to rapidly collect the structural restraints required to solve a high-quality solution structure for proteins twice as large as commonly solved by NMR. The reference 78 authors state that the SAIL technique makes larger proteins such as the 17-kDa calmodulin molecule accessible to more detailed solution structure determination.

Several other research groups determined solution structures of the N-terminal or C-terminal regions of apo- and  $\text{Ca}^{2+}$ -calmodulin.<sup>79</sup> The Yazawa (reference 79a) group in Japan presented solution structures for  $\text{Ca}^{2+}$ -free and  $\text{Ca}^{2+}$ -saturated N-terminal yeast calmodulin (PDB: 1F54, apo, and 1F55,  $\text{Ca}^{2+}$ -saturated). The binding of two calcium ions caused large rearrangement of the

four N-terminal  $\alpha$ -helices and exposed the hydrophobic surface as observed for vertebrate calmodulin. While the observed overall conformation of yeast and vertebrate peptide backbone are similar, several significant conformational differences were observed between the two proteins. The reference 79a authors believe the differences originate from the 38% disagreement in amino acid sequences between yeast and vertebrate CaM. The  $\beta$ -sheet in apo-CaM, PDB: 1F54 is strongly twisted compared with that in the N-domain of vertebrate CaM, but turns to the normal more stable conformation upon calcium-ion binding. The change in  $\beta$ -sheet conformation may also explain the more cooperative nature of  $\text{Ca}^{2+}$  binding for yeast CaM. There are changes in the amino acid residues found on the hydrophobic surface of yeast calmodulin (leu51 substitutes for the normal vertebrate met51, leu71 for met71, ile55 for val55). These changes are thought to be one of the reasons for the poor activation of target enzymes by yeast CaM. A Swedish group (reference 79b) reported on the  $\text{Ca}^{2+}$ -free and  $\text{Ca}^{2+}$ -saturated C-terminal calmodulin (PDB: 1CMF, apo, and PDB: 1CMG,  $\text{Ca}^{2+}$ -saturated). Structural comparison of apo-CaM with  $\text{Ca}^{2+}$ -saturated CaM showed that  $\text{Ca}^{2+}$  binding caused major rearrangements of the secondary structure elements with changes in inter-residue distances of up to 15 Å and exposure of the hydrophobic interior of the four-helix bundle in the  $\text{Ca}^{2+}$ -saturated form (PDB: 1CMG).

The data in Table 6.7, as found in reference 68, appears to disagree with other interhelical angle data in the literature. For instance, some of the angle data for the calcium-ion free calmodulin disagree about 180° from that in Table 6.7, probably indicating different methods used by researchers to calculate the angles. Table 6.7 is included here to indicate how different methods—NMR and X-ray—plus variable calmodulin sources and constructs, affect the interhelical bond angles. See also Figures 6.22 and 6.23 for structural comparisons.

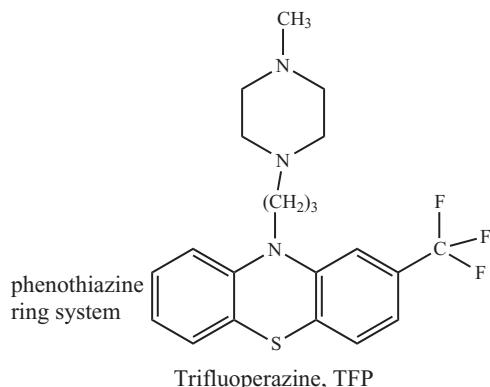
**6.3.2.3 Calmodulin Interactions with Drug Molecules.** Calmodulin interacts with drug molecules containing aromatic rings. These molecules bind in calmodulin's hydrophobic pockets that are exposed upon calcium binding. These drugs must have at least two essential features: (1) a large hydrophobic region consisting of two or three aromatic rings and (2) a positively charged amino group that is at least three atoms away from the ring structure. The aromatic ring region will be held in calmodulin's hydrophobic peptide-binding pocket, and the amino group will be attracted to negatively charged glutamic acid residues present in calmodulin. The glu residues involved in drug binding are usually found in helices A and G. Trifluoperazine (TFP; see Figure 6.24) is an anti-psychotic drug and calmodulin antagonist—that is, TFP binds and inhibits attachment of target enzymes to calmodulin.

TFP has been widely studied in this regard, and a number of structural studies have indicated the binding of one, two, three, or four TFP molecules to calmodulin in the hydrophobic pockets where target enzymes would also bind. Most studies in solution have indicated that CaM has two high-affinity

**TABLE 6.7** Interhelical Angles in Selected Calmodulin Structures

| Calmodulin Structure<br>(Source)                             | Method | Angle (°)<br>A-B (I-II) | Angle (°) C-D<br>(III-IV) | Angle (°)<br>E-F (V-VI) | Angle (°) G-H<br>(VII-VIII) | PDB Entry (Reference Number)                                  |
|--|--------|-------------------------|---------------------------|-------------------------|-----------------------------|---|
| holo $\text{Ca}^{2+}$ -CaM ( <i>Homo sapiens</i> )           | NMR    | 76 ± 2                  | 78 ± 2                    | 70 ± 2                  | 78 ± 2                      | 1J7O (N-terminal, aa 1-76), 1J7P (C-terminal, aa 82-148) (68) |
| apo $\text{Ca}^{2+}$ -free CaM ( <i>Xenopus laevis</i> )     | NMR    | 50 ± 2                  | 48 ± 2                    | 38 ± 2                  | 53 ± 2                      | 1F70 (N-terminal, aa 1-76), 1F71 (C-terminal, aa 82-148) (80) |
| holo $\text{Ca}^{2+}$ -CaM ( <i>Paramecium tetraurelia</i> ) | X-ray  | 91                      | 96                        | 75                      | 89                          | 1EXR (67)   |
| holo $\text{Ca}^{2+}$ -CaM ( <i>Homo sapiens</i> )           | X-ray  | 92                      | 93                        | 75                      | 85                          | 1CLL (72)   |
| CaM-smMLCK ( <i>Homo sapiens</i> )                           | X-ray  | 91                      | 98                        | 81                      | 82                          | 1CDL (86a)  |
| CaM-CaMKII (synthetic construct)                             | X-ray  | 89                      | 87                        | 83                      | 94                          | 1CDM (86b)  |

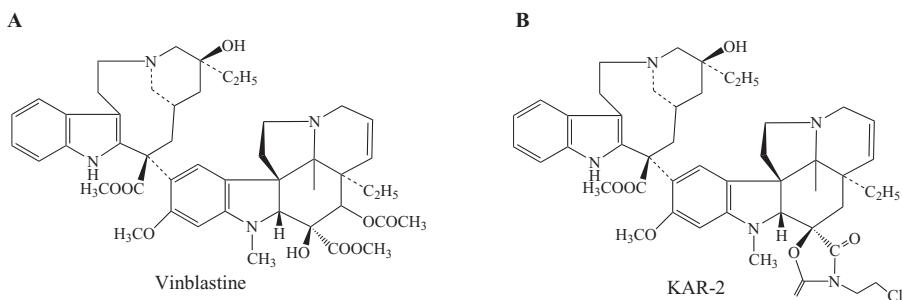
Source: Reference 68.



**Figure 6.24** The calmodulin antagonist trifluoperazine, TFP.

binding sites for TFP.<sup>81</sup> The question arises then as to which calmodulin hydrophobic pockets bind the drug molecules. Possibilities include: (1) both in the N-terminal domain; (2) both in the C-terminal domain; (3) one each in the N- and C-terminal domain; or (4) one in the N- or C-terminal domain and one in an interdomain region. A series of crystal and solution structural studies have answered this question. In a 1994 study by Cook and co-workers the crystal structure of the 1:1 CaM-TFP complex (CaM-1TFP) shows that one TFP molecule binds in the hydrophobic pocket of the C-domain (PDB: 1CTR).<sup>82</sup> In this crystallographic structure, refined to 2.45-Å resolution, residues 75–80 (midpoint of the central helix) and the terminal residue lys148 show almost no electron density and do not appear in the final map. As with many other reported calmodulin structures, the central area of the central helix appears severely bent, disordered, and/or flexible. In spite of the disordered residues, it can be determined that the compact globular structure of the PDB: 1CTR, 1:1 CaM:TFP complex contrasts markedly with the extended dumbbell shape exhibited by the X-ray structures of  $\text{Ca}^{2+}$ -saturated calmodulin such as PDB: 3CLN or PDB: 1EXR (see Figures 6.21 and 6.22A). Thirteen calmodulin amino acid residues form 191 contacts at 5 Å or less between CaM and TFP. The TFP tricyclic phenothiazine ring system lies in the same hydrophobic binding pocket utilized in binding peptides (peptide binding to be discussed below) and makes contact with phe92, ile100, leu105, met109, met124 (25 contacts), ile125, glu127, ala128, val136, phe141, and met144 (20 contacts). The trifluoromethyl group ( $-\text{CF}_3$ ) interacts with glu11 (the only residue from the N-terminal region involved in binding TFP), glu123, met124, and glu127. TFP's aliphatic chain—more interactions are possible if this chain is at least three carbons long—interacts with met124, glu127, and met144 while the piperazine ring, which has a net positive charge, interacts with glu127, ala128, and met144.

In 1998, Ovádi and co-workers studied the structure of calmodulin in complexes with two TFP molecules as well as complexes with TFP and two



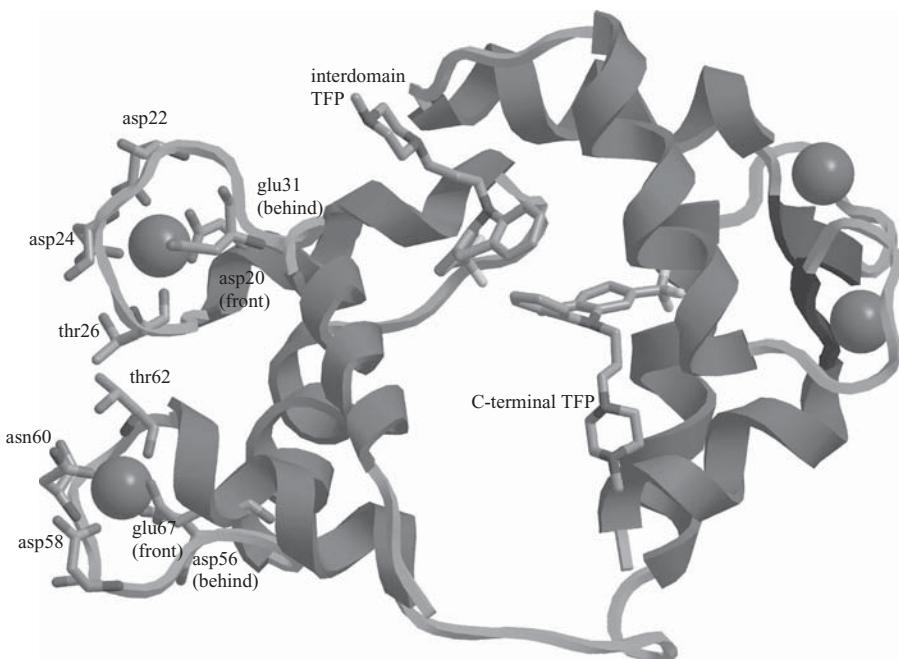
**Figure 6.25** The anticancer drugs (A) vinblastine and (B) KAR-2.

bisindole-based anticancer drug molecules, vinblastine and KAR-2 [ $3''-(\beta\text{-chloroethyl})-2'',4''\text{-dioxo-3,5''-spirooxazolidino-4-deacetoxyvinblastine}$ ] (see Figure 6.25).<sup>83</sup> The X-ray crystallographic structure of the 1:2 CaM:TFP complex has been deposited in the Protein Data Bank as PDB: 1A29.

The X-ray crystallographic structure of the 1:2 CaM:TFP complex (PDB: 1A29) shows binding of one TFP molecule in the C-terminal hydrophobic pocket, with a second TFP in an interdomain site. In the 1:1 structure (PDB: 1CTR), the TFP molecule's tricyclic ring is inverted and rotated 180° from that in the 1:2 CaM:TFP complex (PDB: 1A29). The second TFP molecule residing in the interdomain position shifts toward helix G to make a hydrogen bond with the side chain of glu127 ( $\text{N}_2\text{TFP}_2\text{--O}_{\epsilon 2}\text{ glu127} = 3.21 \text{ \AA}$ ).<sup>83</sup> The reference researchers also studied the 1:2 CaM:TFP complex using near-UV circular dichroism (CD) spectroscopy. In the intact system, with a concentration of 20  $\mu\text{M}$  CaM and varying concentrations of TFP, a negative difference peak characteristic of CaM:TFP complex formation was observed at 263 nm, reaching a saturating value at a concentration of 75  $\mu\text{M}$  TFP. At this concentration, a second positive peak appeared at 270 nm, indicating the binding of the second TFP molecule. In another experiment, calmodulin was cleaved into N- and C-terminal fragments by limited trypsinolysis (cleaving CaM along the central linker at residue lys77). CD spectroscopy of the fragmented CaM was carried out under the same conditions as the previous experiment. The results showed similar behavior (formation of the negative difference peak saturating at 75  $\mu\text{M}$  TFP) but no evidence of a positive difference peak as seen for the intact CaM. The conclusion from this experiment is that the disruption of the CaM central linker region modifies TFP binding and that the C-terminal (and/or N-terminal) intradomain sites become saturated with TFP independently. The position of the second TFP molecule at an interdomain site is confirmed by the PDB: 1A29 X-ray crystal structure (see Figure 6.26).

In Figure 6.26, the calmodulin molecule is shown in cartoon format. The TFP molecules and the calcium ion ligands for calcium sites I and II are shown in stick format. Bond distances for the calcium site I ligands are collected in Table 6.8.

PDB: 1A29



**Figure 6.26** Calmodulin:2trifluoroperazine complex, PDB: 1A29. Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. (Printed with permission of CambridgeSoft Corporation.)

**TABLE 6.8 Bond Distances for the 1:2 CaM:TFP Complex (PDB: 1A29)  $\text{Ca}^{2+}$  at Site I**

| Ligand, Ligand Atom            | Bond Distance to $\text{Ca}^{2+}$ at Site I ( $\text{\AA}$ ) |
|--------------------------------|--|
| asp20, $\text{O}_{\delta 1}$   | 2.29   |
| asp22, $\text{O}_{\delta 1}$   | 2.29   |
| asp24, $\text{O}_{\delta 1}$   | 2.29   |
| thr26, backbone O              | 2.28   |
| glu31, $\text{O}_{\epsilon 1}$ | 2.33   |
| glu31, $\text{O}_{\epsilon 2}$ | 2.29   |

The same authors then studied the formation of ternary complexes with calmodulin, TFP and two bisindole drug molecules: vinblastine and KAR-2 (structures shown in Figure 6.25) using the near-UV CD method. They found that both TFP and KAR-2 are able to bind to calmodulin simultaneously forming a ternary species but that vinblastine is not able to bind to a TFP:CaM complex. If the CD experiment was repeated under conditions in which the

1:2 CaM:TFP complex was expected to form, then KAR-2 was added, results indicated a competition by TFP and KAR-2 for the same interdomain binding site—the second interdomain TFP positive difference peak disappeared and a negative difference peak at 300 nm appeared indicating KAR-2 binding. Using fluorescence experiments, the reference 83 authors also found that KAR-2 and vinblastine can both bind to CaM, forming a ternary complex; however, both bisindoles bind with reduced affinity. The combination of CD and fluorescence experiments led the authors to conclude that TFP and vinblastine compete for the C-terminal CaM binding pocket while KAR-2 competes with a second TFP molecule for an interdomain CaM site. It can also be concluded that TFP and vinblastine are calmodulin antagonists competing with the binding of a target enzyme at the CaM C-terminal hydrophobic pocket. Additionally, it was found that binding of the target peptide melittin to calmodulin prevents the binding of both vinblastine and KAR-2 but not that of TFP. Binding of target peptides to calmodulin is discussed further in Section 6.3.2.4 below.

The crystal structure of a 1:4 CaM:TFP complex was published in 1994 (PDB: 1LIN).<sup>84</sup> This complex shows three TFP molecules binding in or near to the C-domain of CaM, along with one TFP molecule binding to the N-domain. The crystallization mixture contained a 35:1 TFP:CaM ratio in contrast with the 4.5:1 ratio used by the reference 82 authors (PDB: 1CTR), perhaps explaining the higher TFP occupancy in the crystal. A molecular dynamics simulation of the CaM:TFP complex in aqueous solution starting with data from the 1:4 CaM:TFP crystal structure was performed by the PDB: 1LIN researchers. The obtained solution structure is very similar to the 1:2 CaM:TFP crystal structure discussed above. The computer simulation showed that the binding ability of the secondary binding site of the C-domain is higher than that of the primary binding site of the N-domain.

**6.3.2.4 Calmodulin–Peptide Binding.** In order to carry out its purpose, calmodulin (CaM) must first bind calcium ions and then attract and bind peptide segments of its target enzymes in order to activate them (or release the autoinhibitory effect of the target enzyme CaM-binding peptide). Structures of  $\text{Ca}^{2+}$ -saturated calmodulin complexed to various peptides have been studied both in solution by NMR and in the solid state using X-ray crystallography. The structures of binding peptides derived from skeletal (skMLCK, PDB: 2BBM by NMR)<sup>85</sup> and smooth muscle (smMLCK, PDB: 1CDL by X-ray),<sup>86a</sup> as well as from calmodulin kinase II $\alpha$  (CaMKII $\alpha$ , PDB: 1CDM by X-ray),<sup>86b</sup> have been solved. MLCK is the acronym for myosin light chain kinase. The calmodulin-binding peptides assume random coil structures in solution, but in the presence of calmodulin they form amphipathic or amphiphilic (containing both polar and nonpolar residues) helices. All of these peptides have nanomolar (very high) affinities for calmodulin. Table 6.9 shows the primary amino acid sequence of some of the calmodulin-binding peptides, and it is informative to compare them as they are discussed in the following material.

**TABLE 6.9 Numbering System(s) for Calmodulin-Binding Peptides<sup>a</sup>**

|  | -6 | -5 | -4 | -3 | -2 | -1 | 1 | 2 | 3 | 4<br>(1) <sup>b</sup> | 5<br>(2) | 6<br>(3) | 7<br>(4) | 8<br>(5) |
|--|----|----|----|----|----|----|---|---|---|-----------------------|----------|----------|----------|----------|
| skMLCK (M13)<br>(PDB: 2BBM)<br>1–14                  |    |    |    |    |    |    | K | R | R | <b>W</b>              | K        | K        | N        | F        |
| smMLCK (R20)<br>(PDB: 1CDL)<br>1–14                  |    |    |    |    |    |    | A | R | R | <b>W</b>              | Q        | K        | T        | G        |
| CaMKII (PDB:<br>1CDM) 1–10                           | L  | K  | K  | F  | N  | A  | R | R | K | <b>L</b>              | K        | G        | A        | I        |
| CaMKI  |    |    |    |    |    | A  | K | S | K | <b>W</b>              | K        | Q        | A        | F        |
| C20W (PDB:<br>1CFF)                                  |    | L  | R  | R  | G  | Q  | I | L |   | <b>W</b>              | F        | R        | G        | L        |
| C24W   |    |    |    |    |    | Q  | I | L |   | <b>W</b>              | F        | R        | G        | L        |
| C28W   |    | L  | R  | R  | G  | Q  | I | L |   | <b>W</b>              | F        | R        | G        | L        |
| CaMKK $\alpha$ (rat)<br>(PDB: 1CKK)<br>1–16          | V  | K  | L  | I  | P  | S  |   |   |   | <b>W</b>              | T        | T        | V        | I        |
| CaMKK ( <i>C.</i><br><i>elegans</i> ) (PDB:<br>1IQ5) | V  | R  | V  | I  | P  | R  |   |   |   | <b>L</b>              | D        | T        | L        | I        |
| MARCKS (PDB:<br>1IWQ)                                | K  | K  | R  | F  | S  | F  | K | K | S | <b>F</b>              | K        | <b>L</b> | S        | G        |

Source: Reference 85.

<sup>a</sup>Dark-shaded residues interact with calmodulin's C-terminal domain. Light-shaded residues interact with calmodulin's N-terminal domain. For CaMKK peptides, the reverse is true.

<sup>b</sup>Numbers in parentheses correspond to reference 87.

The peptides listed in Table 6.9 comprise the CaM-binding domains of calmodulin target enzymes, many of which are large complex systems whose activity is modulated by calmodulin binding. The names of some of these enzymes appear in the calmodulin introductory section (6.3.2.1). Table 6.9 is adapted from references 85 and 87 and uses the numbering systems found therein. The newer and now more commonly used numbering system found in reference 87 is shown in Table 6.9. The peptide sequences are aligned based on the first hydrophobic anchor residue, shown in bold type and dark shading. Most often this peptide residue is a tryptophan, W (an amino acid with a large aromatic side chain), that forms hydrophobic interactions with the C-terminal domain of CaM. The second anchor residue, many times a phenylalanine, F, also having a large aromatic side chain, is spaced some number of residues further along the peptide chain (shown in bold with light shading) and interacts with hydrophobic residues in CaM's N-terminal domain.

| 9<br>(6) | 10<br>(7) | 11<br>(8) | 12<br>(9) | 13<br>(10) | 14<br>(11) | 15<br>(12) | 16<br>(13) | 17<br>(14) | 18<br>(15) | 19<br>(16) | 20 | 21 | 22 | 23 | 24 |
|----------|-----------|-----------|-----------|------------|------------|------------|------------|------------|------------|------------|----|----|----|----|----|
| I        | A         | V         | S         | A          | A          | N          | R          | F          | K          | K          | I  | S  | S  | S  | G  |
| H        | A         | V         | R         | A          | I          | G          | R          | L          | S          | S          |    |    |    |    |    |
| L        | T         | T         | M         | L          | A          | T          | R          | N          | F          | S          |    |    |    |    |    |
| N        | A         | T         | A         | V          | V          | R          | H          | M          |            |            |    |    |    |    |    |
| N        | R         | I         | Q         | T          | Q          | I          | K          | V          | V          | N          | A  | F  | S  | S  | S  |
| N        | R         | I         | Q         | T          | Q          | I          | K          | V          | V          | N          | A  | F  | S  | S  | S  |
| L        | V         | K         | S         | M          | L          | R          | K          | R          | S          | F          | G  | N  | P  | F  | R  |
| F        | S         | F         | K         | K          |            |            |            |            |            |            |    |    |    |    |    |

Clore, Gronenborn, Bax, and co-workers published the three-dimensional solution structure of the complex between  $\text{Ca}^{2+}$ -saturated calmodulin and a 26-residue synthetic peptide comprised of the calmodulin-binding domain of rabbit skeletal muscle myosin light chain kinase (skMLCK, also referred to as M13) in 1992 (PDB: 2BBM).<sup>85</sup> They noted that the N-terminal (residues 6–73) and C-terminal (residues 83–146) calmodulin domains remained essentially unchanged upon complexation with the peptide. However, the central linker region (residues 65–93), helical in  $\text{Ca}^{2+}$ -saturated calmodulin as determined by published X-ray crystallographic structures (PDB: 3CLN, reference 71) but disrupted near its midpoint in solution structures, also appears disrupted into a long flexible loop in the PDB: 2BBM structure. In fact, the central region (residues 74–82) appears to be a flexible linker between the N- and C-terminal domains changing  $\text{Ca}^{2+}$ -saturated calmodulin from an “extended” dumbbell shape with dimensions of approximately  $65 \times 30 \times 30 \text{ \AA}$  into a “collapsed” globular complex of ellipsoidal shape measuring  $47 \times 32 \times 30 \text{ \AA}$ . The 26-residue

peptide, a random coil in solution, becomes helical in the presence of  $\text{Ca}^{2+}$ -CaM and attaches itself to the  $\text{Ca}^{2+}$ -CaM N- and C-terminal hydrophobic pockets through a large number of hydrophobic and some electrostatic interactions.

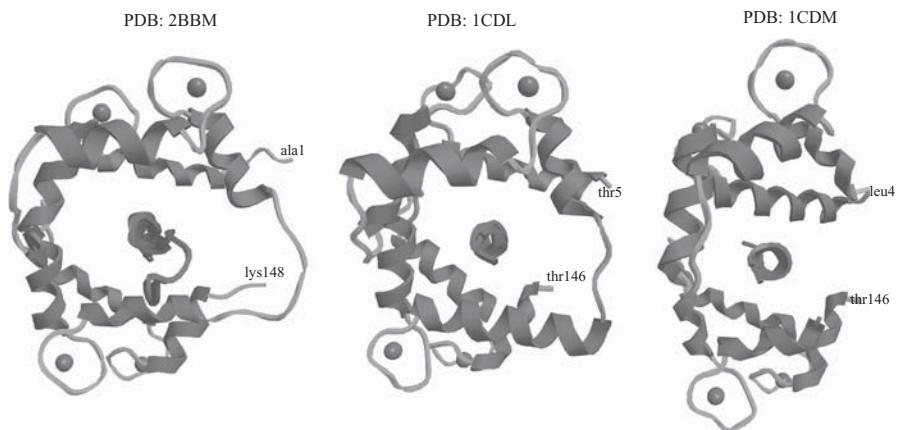
The reference 85 authors found interproton distance restraints using isotope-edited and isotope-filtered 2D, 3D, and 4D nuclear Overhauser effect (NOE) experiments on a 1:1 complex of >95%  $^{15}\text{N}$ - and  $^{13}\text{C}$ -labeled CaM and the 26-residue M13 peptide at natural isotopic abundance. The NOE experiment is described in Sections 3.4.6 and 3.4.10. The one-letter amino acid sequence of the M13 peptide is shown in Table 6.9. The 2D isotope-filtered nuclear Overhauser enhancement spectroscopy (NOESY) experiments yielded interproton distance restraints for the bound M13 peptide, suppressing the signal of  $^{13}\text{C}$ - and  $^{15}\text{N}$ -attached protons. The  $^{13}\text{C}$ -separated,  $^{12}\text{C}$ -filtered 3D NOESY spectra were used to identify NOEs between CaM protons and bound peptide protons. Finally, 3D  $^{13}\text{C}$ - and  $^{15}\text{N}$ -separated and 4D  $^{13}\text{C}$ - $^{13}\text{C}$ -separated NOESY experiments were used to obtain NOEs between protein protons.<sup>85</sup> For details about the NMR methods, see Sections 3.4.6, 3.4.10, and 3.4.11. The 3D experiments showed intense NOE interactions between the M13 tryptophan (trp, W) residue at position 4 (1) and the methyl protons of CaM's met145. (See Table 6.9 for the numbering systems for the peptide positions. The same experiment showed interactions between the aromatic ring protons of phenylalanine (phe, F) at M13's position 8 (5) and the methyl group of CaM's ala88 and the  $\text{CaH}$  proton of ile85. These latter two CaM residues are located near the C-terminal end of the central helix. The 4D experiments yielded information on CaM's shape in the complex. For instance, one result showed that leu39 (N-terminal B helix) and val91 (C-terminal E helix) are  $\leq 5\text{ \AA}$  apart, indicating that the N- and C-terminal domains are in close proximity. This is compared to the  $\text{Ca}^{2+}$ -CaM crystal structure where the methyl groups of these residues are 30–33  $\text{\AA}$  apart. The authors state that the flexibility of the central loop allows the hydrophobic pockets of the two calmodulin domains to approach each other so that the peptide's hydrophobic residues can form the many van der Waals contacts that hold the complex together “rather like two hands capturing a rope.” Figure 2 of reference 85 illustrates interpretations of the NMR data indicating at least 133 intermolecular interactions between CaM and the M13 peptide. Several electrostatic interactions between basic M13 arginine and lysine residues and CaM glutamic acid residues are also proposed by these authors. The most important M13 residues for hydrophobic interaction with calmodulin appear to be those of trp4 (residue 1 the reference 87 numbering system) and phe17 (14), both of which make many van der Waals contacts in the complex. Calmodulin and the M13 peptide appear to be arranged in an antiparallel fashion—that is, trp4 (1) interacts with residues (F89, F92, I100, L105, M109, V121, M124, I125, V136, F141, M144, M145) in CaM's C-terminal domain, whereas phe17 (14) interacts with residues (I27, L32, M36, M51, I52, V55, I63, F68, M71) in CaM's N-terminal domain. As will be seen below, similar arrangements are found for other pep-

tides in binding to calmodulin. It should be noted that calmodulin contains an unusually high number of methionine residues and that all of these are involved in van der Waals interactions with incoming target peptides. It is proposed that one reason calmodulin is able to recognize and bind many different target peptides (of variable length and primary sequence) lies in the flexibility of the unbranched methionine side chain that, through minor conformational changes, adjusts to capture the target.

Later in 1992, Quiocho and co-workers published the 2.4 Å X-ray crystallographic structure of calmodulin complexed with the CaM-binding region of chicken smooth muscle light chain kinase (smMLCK, PDB: 1CDL).<sup>86a</sup> The peptide amino acid sequence is found in Table 6.9. The peptide, known as smMLCK, spans residues 796–815 of smooth muscle light chain kinase. The peptide also serves as an autoinhibitor of smMLCK. The first 12 amino acids appear to fulfill this function, while residues trp4 (1) and 8–17 (5–14) are primarily involved in CaM binding. The CaM–peptide complex exhibits a compact ellipsoidal structure with overall dimensions of approximately 50 × 30 × 25 Å. The NMR structure described in reference 85 (PDB: 2BBM) and the X-ray structure described in reference 86a (PDB: 1CDL) are quite similar, with calmodulin surrounding the peptide helical coil. The extra length of the 2BBM peptide—26 residues versus 18 in the 1CDL structure—is evident in Table 6.9 and Figure 6.27. The flexible central linker loop portion in the Ca<sup>2+</sup>-CaM: smMLCK structure (PDB: 1CDL) spans residues 73–77, in contrast to the longer flexible loop in the NMR Ca<sup>2+</sup>-CaM: skMLCK structure (PDB: 2BBM) that spans residues 74–82. In the X-ray structure (PDB: 1CDL), the calmodulin N- and C-terminal domains are related by a pseudo-twofold axis that runs through the center of the structure relating helices A and E, B and F, C and G, and D and H. The bound peptide, α-helical for almost its entire length, lies in the tunnel running diagonally along the long axis of the calmodulin ellipsoid. N- and C-terminal calmodulin hydrophobic pockets interface with the hydrophobic side of the helical peptide. See Figure 6.27.

In agreement with the NMR structure PDB: 2BBM discussed above, the calmodulin–peptide interactions occur in an antiparallel conformation—that is, N-terminal peptide residues trp4 (1) and thr7 (4) interact with the C-terminal calmodulin pocket while C-terminal peptide residues ala13 (10), ile14 (11), and leu17 (14) reside in the N-domain pocket. The residue numbers in parentheses correspond to the reference 87 numbers in parentheses in Table 6.9 (In reference 86a, there is a third peptide numbering system which this author has ignored.) Of the approximate 185 contacts of <4 Å between calmodulin and peptide, about 80% are van der Waals contacts (hydrophobic) and the remainder electrostatic—salt bridges between peptide basic and calmodulin acidic side chains and hydrogen bonds. As with the NMR structure of reference 85, PDB: 2BBM, all methionine residues of calmodulin are involved in hydrophobic bonding with the peptide. See Figure 6.27.

Arg16 (13) of the peptide chain forms electrostatic (hydrogen bonding and charge-coupling) interactions with the central loop bend and residues in



**Figure 6.27** Three different calmodulin-binding peptides in complexes with calmodulin (PDB: 2BBM, 1CDL, and 1CDM). Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. (Printed with permission of CambridgeSoft Corporation.)

helices D and E of the central linker. As can be seen in Table 6.9, a basic residue (arg, lys, his) occurs at this position on the peptide chain for all calmodulin-binding peptides listed, making this a universally important residue for this group of peptides. Mutagenesis studies on the interaction of CaM with smMLCK have shown the critical importance of peptide residues trp4 (1), arg16 (13), and leu17 (14) for successful CaM-smMLCK binding. The studies also indicate that these residues or close analogs are conserved in other target enzymes (see Table 6.9). Additionally, it has been observed that mutation of three residues in CaM's N-terminal domain converts calmodulin from agonist (assistor of target enzyme activity) to potent antagonist (binder that suppresses target enzyme activity). This observation leads the reference 86a authors to propose that the peptide N-terminal domain first acts with the peptide's C-terminal end, followed by arg16 (13) interaction with the central linker loop-helix domain. Finally, the peptide's C-terminal end would be pulled from its place in the larger target enzyme to interact with CaM's N-terminal domain completing the enzyme activation process.

In late 1993, the Quiocho group reported the 2.00-Å resolution X-ray crystallographic structure of a  $\text{Ca}^{2+}$ -saturated CaM with the CaM-binding domain peptide of bovine brain CaM-dependent protein kinase II $\alpha$  (CaMKII, PDB: 1CDM).<sup>86b</sup> The authors compared the CDM structure to that of  $\text{Ca}^{2+}$ -staurated CaM (PDB: 1CLL, reference 72) and to that of  $\text{Ca}^{2+}$ -staurated CaM complexed with smMLCK (PDB: 1CDL discussed in previous paragraphs). Residues 293–310 of the CaMKII peptide make approximately 135 contacts of  $\leq 4\text{\AA}$  with the  $\text{Ca}^{2+}$ -saturated CaM in a structure very similar to that of PDB: 1CDL discussed above (with approximately 185 contacts). However, in the PDB: 1CDM structure, the CaM central linker loop residues 74–83 are

disordered and do not appear in the final structural renderings (see Figure 6.27). Peptide N-terminal residues arg1, arg2, lys3 form salt bridges with CaM residues glu7, glu14 (N-terminal CaM) and glu120, glu127 (C-terminal CaM) (numbering of peptide residues according to Table 6.9). Major CaM hydrophobic pocket contacts (comprised of nine met and several phe, leu, val, and ala residues) are made between peptide residue leu4 (1) with CaM C-terminal residues and leu13 (10) with CaM N-terminal residues. With the CaMKII peptide, the two major hydrophobic contacts occur at residues (1) and (10) rather than at residues (1) and (14) as seen for skMLCK and smMLCK binding peptides. This residue (1) and (10) behavior is also seen for the CaMKI CaM-binding peptide listed in Table 6.9. It is important to note that the  $\text{Ca}^{2+}$ -binding sites (the EF-hand, helix-loop-helix domains) in all of these structures remain distant from the target enzyme CaM-binding peptides, enabling free access and egress of calcium ions from their binding sites. The reference 86b authors again mention that the CaM central linker helix is more or less disrupted in complexation with different CaM-binding peptides—residues 73–77 are nonhelical in PDB: 1CDL and residues 74–83 in PDB: 1CDM—allowing calmodulin to accommodate itself to a variety of peptide-binding sequences. This flexibility is believed to be a key element of calmodulin's molecular recognition ability and a main contributor in the mechanism of calcium-induced signal transduction.

The research discussed in references 85 and 86 has been summarized by Griesinger, Krebs, and co-workers and extended to a peptide comprised of the N-terminal portion of the calmodulin-binding domain (C20W) of the plasma membrane calcium pump. The experimental method for structure determination was by multidimensional, heteronuclear NMR<sup>87</sup> (PDB: 1CFF). More information about the calcium pump will be found in Section 6.4.2. However, we should say now that the plasma membrane calcium pump is one of the key enzymes controlling  $\text{Ca}^{2+}$  levels within cells. The calmodulin-binding domain of the calcium pump has a length of approximately 30 amino acid residues, along with the capability to form an amphiphilic (polar and non-polar segmented) helix. The domain has an autoinhibitory function binding to two receptor sites near the active center of the calcium pump. The calmodulin-binding domain's N-terminal region interacts with a site near the important phosphorylatable aspartate residue of the pump and its C-terminal region interacts with the cytoplasmic loop between transmembrane domains 2 and 3 of the calcium pump. A unique feature of the plasma membrane calcium pump target enzyme is that it can be activated by the C-terminal but not the N-terminal half of CaM, if these two domains are isolated after trypsinolysis. The reference 87 researchers wanted to find out which part of the CaM-binding domain of the calcium pump is sufficient to interact solely with the C-terminal half of calmodulin and designed three peptides, C20W, C24W, and C28W, to answer this question. As can be seen in Table 6.9, the N-terminal binding domain starting with residue 1 is identical in these three peptides, but the C-terminal domain has been truncated at residue 16 (13) of the C20W peptide.

Small-angle X-ray scattering (SAXS) studies performed on complexes of CaM with C20W and C24W indicated that the CaM–C24W complex had the same globular shape as other CaM–CaM-binding peptide complexes discussed here previously. However, the SAXS results for the CaM–C20W complex indicated a more flexible CaM molecule in solution, and its solution structure (described next) illustrates the differences.

Griesinger, Krebs, and co-workers compared the structures of CaM–C20W complex (PDB: 1CFF) with the CaM–skMLCK (also called M13) complex (PDB: 2BBM) as well as with  $\text{Ca}^{2+}$ -saturated calmodulin (PDB: 3CLN). Comparison of the M13 peptide sequence in Table 6.9 with that of C20W shows that the second CaM-binding hydrophobic residue segment is missing in C20W. This factor leads to significant differences in the C20W–CaM three-dimensional structure in comparison to that seen with the other peptides. Instead of the globular structure noted for the CaM–skMLCK (M13), CaM–smMLCK, and CaM–CaMKII complexes, the CaM–C20W structure resembles that of  $\text{Ca}^{2+}$ -saturated CaM itself.

The C20W peptide, mainly random coil in solution, adopts an  $\alpha$ -helical structure when complexed, and it interacts with calmodulin in an antiparallel manner—that is, the N-terminal half of the peptide binds to calmodulin’s C-terminal domain and vice versa. However, the C20W peptide is missing the second hydrophobic peptide (light shading in Table 6.9) and does not bind to CaM’s N-terminal domain. In a 3D  $^{12}\text{C}$ ,  $^{14}\text{N}$   $\omega_1$ -filtered NOESY  $^1\text{H}$ ,  $^{13}\text{C}$  HSQC experiment, 49 intermolecular NOEs between peptide (unlabeled) and CaM ( $^{13}\text{C}$ - and  $^{15}\text{N}$ -labeled) were defined. In a similar manner as observed with other calmodulin-binding peptides, trp4 (1) showed numerous NOE interactions to the methyl groups of ile100, leu105, met124, ile125, ala128, val136, and met144 of CaM. In addition to trp4 (1), residues leu8 (5) and ile11 (8) on the hydrophobic side of the peptide helix bind in calmodulin’s C-terminal hydrophobic pocket. Figure 8 of reference 87 illustrates the striking difference in the number of peptide–CaM contacts when comparing the CaM–M13 complex (that binds to both the C- and N-terminal domains of CaM) to CaM–C20W (that binds only to the CaM C-terminal domain). All calmodulin C-terminal met residues interact with the peptide, again because met side chains are flexible, the C–S bond is long, and the sulfur atom is more polarizable than carbon, leading to stronger van der Waals interactions. The central linker connecting the two domains of calmodulin (in this case residues 74–84) is ill-defined and does not appear in the structural representation.

Although previous research has indicated that target enzyme CaM-binding domains must interact with both the C- and N-terminal domains of calmodulin to end autoinhibition and activate the target enzyme, the reference 87 authors have shown in their work that the  $\text{Ca}^{2+}$  pump can be activated and autoinhibition by CaM-binding domains negated through binding to calmodulin’s C-terminal domain only. Figure 11 of reference 87 graphically illustrates the exposure of the CaM-binding domain’s tryptophan (trp4 (1)) residue to solvent during the autoinhibition stage when the CaM-binding domain is attached to

the crucial ATP-binding domain of both CaMKI and the  $\text{Ca}^{2+}$  pump. Binding of calmodulin through hydrophobic interaction with trp4 (1) is proposed as the first event in a cascade that removes the autoinhibitory CaM-binding domain from its position on the ATP binding domain of the target enzyme, thus activating the enzyme for its physiological purpose.

More recently, several different calmodulin binding modes to target enzymes have been described. Vetter and Leclerc discuss several of these in their 2003 review article.<sup>88</sup> These authors use the reference 87 numbering system for the target enzyme binding peptides, as seen in Table 6.9. We have discussed the collapsed (compact) structures exhibited by calmodulin when complexed with target peptides skMLCK (reference 85, PDB: 2BBM), smMLCK (reference 86a, PDB: 1CDL), and CaMKII $\alpha$  (reference 86b, PDB: 1CDM) and the extended calmodulin structure when complexed with the C20W peptide (reference 87, PDB: 1CFF). In all of these complexes, the target enzyme binding peptide forms an  $\alpha$ -helix upon calmodulin binding. Vetter and Leclerc review these structures and go on to discuss calmodulin complexes with  $\text{Ca}^{2+}/\text{CaM}$ -dependent kinase kinase,<sup>89</sup> the  $\text{Ca}^{2+}$  activated  $\text{K}^+$  channel,<sup>90</sup> and the anthrax exotoxin (edema factor), a calmodulin-dependent adenylate cyclase.<sup>91</sup>

$\text{Ca}^{2+}/\text{CaM}$ -dependent kinase kinase (CaMKK) is a  $\text{Ca}^{2+}/\text{CaM}$ -dependent serine/threonine kinase that phosphorylates other CaM-dependent kinases. (A kinase is a type of enzyme that transfers phosphate groups from high-energy donor molecules, such as ATP, to specific target molecules—substrates—that are often amino acid residues of the target protein. CaMKK is at the top of a cascade of phosphorylation events. Upon its activation by calmodulin, CaMKK subsequently phosphorylates two serine/threonine kinases, CaM-kinase I (CaMKI) and CaM-kinase IV (CaMKIV). CaMKI and CaMKIV are themselves calmodulin-dependent enzymes. The calmodulin-binding domain of CaMKK is different from those previously discussed in that the CaM binding residues are further apart than those for skMLCK, smMLCK, and CaMKII $\alpha$ . Also a cluster of basic residues (principally arginines (R) and lysines (K)) are located near the C-terminal anchor residue of CaMKK (phe16 in Table 6.9) rather than near the N-terminal anchor residue (trp1 or leu1 in Table 6.9) of the previously discussed CaM-binding peptides.

Many more structure determinations of CaM-binding peptides have been carried out. For instance, the NMR structure determined by Ikura and co-workers for a CaM/CaMKK (from rat) complex (PDB: 1CKK) shows a calmodulin collapsed structure similar to those of CaM/smMLCK, CaM/skMLCK, and CaM/CaMKII $\alpha$  with the rCaMKK peptide cradled between calmodulin's C- and N-terminal domains.<sup>89</sup> However, two features are different for CaM/rCaMKK. The first is that the peptide is bound in an inverted position compared to the others—that is, the N-terminal end of the rCaMKK peptide binds to CaM's N-terminal end and the C-terminal end binds to the C-terminal end rather than vice versa. This factor appears related to the clusters of basic residues on the target enzyme binding peptide—that is, when the

basic residues are near the N-terminal domain, they form salt bridges with glutamate and aspartate residues in calmodulin's C-terminal domain and when the basic residues (as in rCaMKK) are near its C-terminal end, the salt bridges form with the same glu and asp residues in calmodulin's C-terminal domain. The second obvious difference is that the rCaMKK peptide is only partially  $\alpha$ -helical (11 residues from rat CaMKK residues 444–454) and contains an additional well-defined  $\beta$ -hairpin-like loop at its C-terminal end (rat CaMKK residues 455–459). Many of calmodulin's methionine residues are involved in binding the rCaMKK $\alpha$  peptide, as was the case for systems discussed previously. Electrostatic interactions responsible for forming salt bridges and orienting the rCaMKK $\alpha$  peptide's C-terminal end with calmodulin's C-terminal end are the peptide basic arg455, lys456, and arg457 (Table 6.9 positions 12, 13, 14) residues and calmodulin's glu14, glu120, glu123, and glu127 residues. In summary, while the hydrophobic and electrostatic attractions between rCaMKK $\alpha$  peptide and calmodulin appear similar to the binding modes discussed previously, the secondary structure of the rCaMKK $\alpha$  peptide deviates substantially from those of smMLCK and CaMKII $\alpha$ . Significantly and surprisingly, the rCaMKK $\alpha$  peptide's orientation in calmodulin's binding tunnel is reversed from those of smMLCK and CaMKII $\alpha$ .

An X-ray crystallographic structure of a  $\text{Ca}^{2+}$ -CaM/CaMKK complex published by Ikura and co-workers in 2001 confirms the binding and peptide orientation characteristics for this system.<sup>92</sup> In this case, the calmodulin-binding cCaMKK peptide is derived from a nematode Ca(2+)/calmodulin-dependent kinase peptide *Caenorhabditis elegans* (*C. elegans*, PDB: 1IQ5). Although the anchor residues differ from those described for rCaMKK, the structures are quite similar. The peptide residues 337–349 form an  $\alpha$ -helix, while a  $\beta$ -hairpin loop is formed by residues 350–357. Leu337 (Table 6.9, L1) and ile341 (Table 6.9, I5) have hydrophobic interactions with calmodulin's N-terminal domain, while phe352 (Table 6.9, F16), 14 residues away from the primary leu337 N-domain anchor) binds hydrophobically to calmodulin's C-terminal domain. The phe352 residue resides in the  $\beta$ -hairpin loop and probably would not be in a C-domain anchor position if it were part of a continuing  $\alpha$ -helical structure. The loop structure allows the cCaMKK peptide to fold back and places the phe352 residue in position to form the hydrophobic bonds necessary to form the CaM collapsed structure in the same manner discussed above for rCaMKK $\alpha$ . Hydrogen bonding interactions were found between calmodulin and cCaMKK residues for glu87 (CaM)—arg336 (cCaMKK), glu87—thr339, and lys75—thr339 (via a water molecule). Salt bridges form to orient cCaMKK's C-terminal end to CaM's C-terminal domain: glu11—arg349, for instance. These authors compare the CaM/cCaMKK (PDB: 1IQ5) structure to the CaM/smMLCK (PDB: 1CDL) and CaM/CaMKII $\alpha$  (PDB: 1CDM) structures reported previously. They find that differences can be found in comparisons of calmodulin N-terminal (residues 5–73) and C-terminal (residues 84–146) domain orientation when CaM is bound to the different target peptides. When the N-terminal and C-terminal calmodulin domains are superimposed, essen-

tial differences are found in the amount of rotation and translation needed to effect the superimposition. Calmodulin adjusts to these differences through unwinding or expanding its central linker region. The reference 92 authors also find that glu11 is conserved in all complexes compared and forms salt bridges in each: (1) glu 11–arg349 for CaM/cCaMKK; (2) glu 11–arg455 for CaM/rCaMKK $\alpha$ ; (3) glu 11–his805 for CaM/smMLCK; and (4) glu 11–arg297 for CaM/CaMKII $\alpha$ .

The last structure to be discussed in this section is a complex between calmodulin and the calmodulin-binding domain of myristoylated alanine-rich C kinase (MARCKS). MARCKS is a membrane-associated protein predicted to have an extended rod-like shape. It contains a calmodulin-binding domain near the middle of the molecule that interacts not only with CaM, but with other protein effectors. The same MARCKS domain contains phosphorylation sites for several protein kinases. The CaM-binding domain consists of a characteristic amphiphilic sequence with lysine clusters at both its N- and C-terminal ends. The X-ray structure of the 19-residue MARCKS peptide bound to calmodulin was published by Taniguchi and co-workers in 2003 (PDB: 1IWQ).<sup>93</sup> Calmodulin's structure in this complex shows a compact, ellipsoidal form similar to those seen for CaM-smMLCK, CaM-skMLCK, CaM-CaMKII $\alpha$ , and CaM-CaMKK complexes except that the hydrophobic pocket in CaM's N-terminal domain is shallower than for the others. Little electron density is apparent for central linker residues 77–80, indicating the expected flexibility in this region that allows the compact CaM structure. The MARCKS peptide was found to have a short loop (residues 164–166) and a short helical region encompassing residues 157–163 (7 residues) at its C-terminal end and a long loop extending from residues 156 through 148 at its N-terminal end. The peptide's structure contrasts with the helical nature of peptides forming CaM-smMLCK, CaM-skMLCK, and CaM-CaMKII $\alpha$  complexes. The peptide's loop region in the CaM–MARCKS complex is longer than that found for the CaM–CaMKK complexes and its N-terminal loop does not curl back onto the helix as is seen in the CaM–CaMKK complexes. Additionally, the MARCKS peptide adopts the reverse orientation of the CaM-smMLCK, CaM-skMLCK, and CaM-CaMKII $\alpha$  complexes—that is, MARCKS' N-terminal end interacts with CaM's C-terminal domain and its C-terminal end interacts with CaM's N-terminal domain. All regions of the peptide are involved in contacts with CaM even though the N-terminal domain hydrophobic pocket that usually forms in  $\text{Ca}^{2+}$ -saturated CaM is almost absent in the CaM–MARCKS complex, with leu159 of the MARCKS peptide making contact only with the N-terminal domain surface. Leu159 is separated by only one residue from the second anchor point, phe157; however, this residue makes the normal deep hydrophobic pocket interaction in CaM's C-terminal domain. Many ionic contacts, notably those between lys154–glu11 (CaM helix A), lys155–glu123 (CaM helix G), and lys154–glu127 (helix G), stabilize the complex. Lys154 and lys155 reside in the peptide's N-terminal loop region.

Conclusions to be drawn from the discussions of various calmodulin/target enzyme binding peptide complexes include the following: (1) Calmodulin releases the autoinhibition or enhances the catalytic activity of hundreds of enzymes; (2) calmodulin contains four EF hand helix-loop-helix motifs that bind calcium, with the four EF hands being distributed in essentially mirror image N- and C-terminal domains connected by a central linker region; (3) the calcium-ion binding loops remain virtually unchanged and exposed to solvent in all complexes allowing easy ingress and egress for  $\text{Ca}^{2+}$ ; (4) calmodulin's hydrophobic pockets, exposed upon calcium ion binding, are deep enough to accommodate large aromatic, hydrophobic residues such as tryptophan and phenylalanine from the target peptides, or long aliphatic side chains from leucine and isoleucine residues; (5) large areas (2500 to  $>3000 \text{ \AA}^2$ ) of the target enzyme molecule are shielded from solvent through binding to calmodulin; (6) calmodulin methionine residues are plentiful and flexible and contain the polarizable sulfur atom, and met residues are usually involved in hydrophobic binding to target peptides; (7) additional hydrogen bonding and electrostatic interactions connect calmodulin and the target enzyme, and many times these interactions determine the orientation of the target peptide with respect to calmodulin's N- and C-terminal domains; (8) the flexible linker region connecting calmodulin's N- and C-terminal domains adjusts to different target enzyme sequences through  $\alpha$ -helix unwinding and extension of the central linker; and (9) in many cases calmodulin's N-terminal domain, especially helix A, moves more than other parts of the molecule in accommodating differences in target enzyme binding sequence and secondary structure.

Vetter and Leclerc also discuss the X-ray crystallographic structure of calmodulin complexed with the gating domain of a  $\text{Ca}^{2+}$ -activated  $\text{K}^+$  channel (SK channel). This CaM/SK channel complex, studied by Schumacher et al.<sup>94</sup> (PDB: 1G4Y), exhibits a structure that is very different from those seen previously. Calmodulin binds to the SK channel through the channel's CaM-binding domain (CaMBD), located in an intracellular region of the channel. The SK channel calmodulin-binding domain consists of two long  $\alpha$ -helices ( $\alpha 1$  and  $\alpha 2$ ) connected by a loop. The  $\alpha 1$ -helix comprises CaMBD residues 413–439, the loop residues 440–445, and the  $\alpha 2$ -helix residues 446–488. Two CaMBD units form a dimer ( $\alpha 1$ ,  $\alpha 2$  and  $\alpha 1'$ ,  $\alpha 2'$ ) that interacts with two separate calmodulin molecules (CaM1 and CaM2) in an extended conformation. Each calmodulin subunit carries two calcium ions in its N-terminal domain and no calcium ions in its C-terminal domain. Each CaMBD  $\alpha 1$ -helix interacts with the C-terminal domain of one of the calmodulin subunits (CaM1, for instance), whereas the  $\alpha 2$ -helix interacts with the N-terminal domain of the second calmodulin (CaM2) subunit and the central linker region of the first calmodulin (CaM1) subunit. The effect is that each calmodulin subunit wraps around three CaMBD  $\alpha$ -helices. Figure 1 of reference 94 shows the structure clearly. The CaMBD units have an overall positive character (pI 10.5, charge +14.7) and are drawn to the acidic CaM (pI 3.9) units. However, as with other CaM-target enzyme

complexes the specific binding interface between the two proteins is hydrophobic. In the CaM N-lobe, residues phe12, phe19, val35, met36, leu39, phe68, met 71, and met72 create a hydrophobic patch that interacts with CaMBD residues leu480, leu483, and ala484. In this case, the leu480 residue is the anchor residue reaching far into the N-terminal hydrophobic pocket. Interactions of CaMBD units with the calmodulin's C-terminal domains in the CaM–CaMBD complex are different from those seen previously because of the distinct conformation of the calcium-free C-lobe. The calcium-free C-lobe almost adopts the “closed” conformation seen for apo-calmodulin described in Section 6.3.2.2 except that the angle between the G and H helices opens up to create several smaller and discrete hydrophobic binding pockets for CaMBD binding. The CaMBD side-chain residues ala425, leu428, and trp432 bind to calmodulin residues glu127, met124, phe141, and met144. Additionally, leu428 plugs a hydrophobic patch in the CaM C-lobe formed by phe89, phe92, leu105, and met109. Schumacher et al.<sup>94</sup> extend the structural analysis of the CaM/CaMBD system to propose a model for K<sup>+</sup>-channel opening facilitated by the calmodulin/SK-channel calmodulin-binding domain interactions. First, it is necessary to know that the Ca<sup>2+</sup>-activated K<sup>+</sup> channels (SK channels) are voltage-independent and are gated solely by intracellular Ca<sup>2+</sup>. The reference 94 authors propose a chemomechanical model in which calcium ion addition and subsequent CaM–CaMBD complexation changes the orientation of SK-channel S6 helices attached to the CaMBD units, opening them for ion passage. Figure 6 of reference 94 illustrates the chemomechanical model.

Finally in this section, Vetter and Leclerc, reference 88, discuss crystal structures of the C-terminal fragment (residues 291–800) of the *Bacillus anthracis* exotoxin. The anthrax exotoxin (so-called edema factor, EF) is a calmodulin (CaM)-dependent adenylate cyclase. It enters host cells through a transporter produced by the anthrax pathogen. EF acquires calmodulin inside the cell and then converts large amounts of ATP to cyclic AMP (cAMP). The X-ray crystallographic structures of the anthrax exotoxin (edema factor, EF) alone (PDB: 1K8T), EF in complex with calmodulin (PDB: 1K93), and EF in complex with calmodulin and an ATP analog (PDB: 1K90) have been determined by Chester L. Drum and associates.<sup>91</sup> The structures are important because one can compare the exotoxin structure as it changes through complexation with calmodulin (PDB: 1K93) and then with CaM plus 3'dATP, an inactive ATP analog (PDB: 1K90). These are also the first structures of CaM complexed with a catalytically functional CaM-target enzyme. Drum and co-workers found that the structure of EF alone differs substantially from that of CaM-EF, but very few further changes are seen when adding the ATP analog. Therefore, it is assumed that the changes in CaM-EF structure are almost all due to the binding of CaM. CaM, in an extended conformation similar to Ca<sup>2+</sup>-CaM (PDB: 1CLL) and with its central helix unwound between residues 76–78, is inserted deeply between two domains of the exotoxin. The CaM molecule contains two Ca<sup>2+</sup> ions in its C-terminal domain and no Ca<sup>2+</sup> ions in its N-terminal domain. The N-terminal domain is in the “closed” conformation and

is rotated about 80° compared to its fully calcium-saturated structure (PDB: 1CLL). Upon calmodulin binding, large EF movements in its large helical domain (residues 660–800) and conformational changes in its segments designated as switches A, B, and C take place. At least 53 CaM residues interact with 63 EF residues, leading to a large buried surface area of about 5900 Å<sup>2</sup>. This was the largest protein–protein interface between independently stable CaM and EF molecules known at the time of this research. Figure 3 of reference 88 shows a representation of calmodulin binding to the anthrax exotoxin and structural changes that occur upon CaM binding.

Several EF regions are important for describing the changes observed in the EF protein in going from EF alone to the CaM-EF complex. Switch A (EF residues 502–551) includes amino acid side chains that bind CaM. Switch B (residues 578–591), disordered in the EF alone (PDB: 1K8T) structure, becomes clearly visible—ordered—in the CaM-EF (PDB: 1K93) complex. Switch C (residues 630–659) moves 33 Å from its EF alone position to its CaM-EF complex position, and two switch C β-strands and connecting loop (642–652) are converted to a helix in the CaM-EF structure. EF residues 501–540, comprising switch A and a region called C<sub>A</sub>, interact with central linker and C-terminal CaM helices D, E, and F in a fashion similar to that described for other systems described above with both hydrophobic and electrostatic interactions. In particular, EF residue lys525 appears to be a “hot-spot” for CaM binding as its mutation to alanine increases the concentration of CaM necessary for EF activation 200-fold for half-maximum response (EC<sub>50</sub>). The lys525ala mutant has little effect, however, on the enzyme-catalyzed reaction at infinite substrate concentration ( $V_{max}$ ). A second EF region, residues 615–634, 647–672 (EF helix L) and 695–721 (EF helix M), contacts three of the four EF-hand motifs in CaM, one in the Ca<sup>2+</sup>-saturated C-terminal domain and two in the calcium-free N-terminal domain. EF helices L and M contact the first EF-hand motif in calmodulin’s N-terminal domain and appear to stabilize its conformation. The EF residue asp647 forms a salt bridge with calmodulin’s arg90 stabilizing the interaction near EF-hand motif 3 in calmodulin’s C-terminal domain. An asp647ala mutant has a significantly lower  $V_{max}$  but shows little change in EC<sub>50</sub> in contrast to the lys525ala mutant. The authors believe that the Michaelis–Menten behavior (see Section 2.2.4) indicates that CaM binds residues in switch A before those in switch C—that is, binding behavior is more important than catalytic competence for the switch A region mutant and vice versa for the switch C mutation.

**6.3.2.5 Conclusions.** As Vetter and Leclerc state in reference 88 conclusions: “Target recognition and activation by CaM has been studied intensively over the last two decades and today we are further away from a general model than ever before.” Much knowledge has been gained about calmodulin in the work done to date; however, structures of calmodulin in complex with full-length target enzymes, and not just with short peptides, are needed. The structural studies on the calmodulin complex with the anthrax exotoxin are

particularly informative, and readers should look for more studies like this one in the future literature.

## 6.4 PHOSPHORYL TRANSFER: P-TYPE ATPASES

### 6.4.1 Introduction

P-type adenosine triphosphatases (ATPases) are proteins that maintain electrochemical gradients and actively transport ions across biomembranes.<sup>95</sup> The P-type ATPases, also known as cation pumps, use energy from ATP to facilitate ion transport. During active transport, phosphorylation of an aspartic acid residue takes place and a covalent aspartyl-phosphoanhydride intermediate is formed with ATP. A magnesium ion appears to be necessary for the phosphorylation to take place. P-type ATPases include  $\text{Na}^+$ ,  $\text{K}^+$ -ATPases,  $\text{H}^+$ ,  $\text{Na}^+$ -ATPases, and  $\text{Ca}^{2+}$ -ATPases.  $\text{Na}^+$ ,  $\text{K}^+$ -ATPases were discussed in Section 5.4.1.  $\text{Ca}^{2+}$ -ATPases, to be discussed in Section 6.4.2, are found in the sarco(endo)plasmic reticulum of skeletal muscle and are known as SERCA ATPases. This large protein facilitates the re-uptake, into the sarcoplasmic reticulum, of cytosolic  $\text{Ca}^{2+}$  released during muscle contraction. SERCA is also involved in signal pathways using calcium as a secondary messenger.

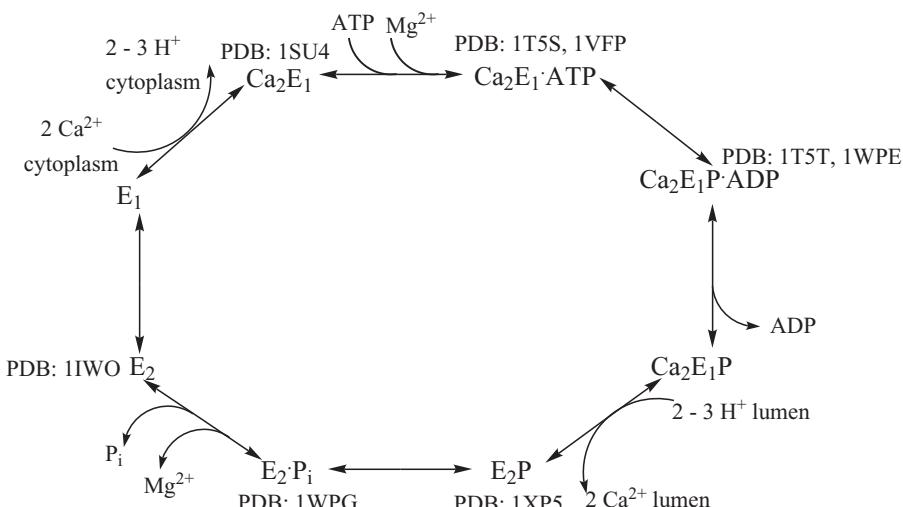
### 6.4.2 Calcium P-Type ATPases

Some sarco(endo)plasmic reticulum (SR)  $\text{Ca}^{2+}$ -ATPases (SERCA) are P-type ATPases that play a major role in muscle contraction-relaxation cycles and are responsible for transporting calcium into the lumen of the sarcoplasmic reticulum. Some definitions are useful in the discussion of  $\text{Ca}^{2+}$ -ATPases:

1. Every eukaryotic cell has three parts: a membrane with channels to let materials into and out of the cell, cytoplasm containing organelles and the cytosol, and a nucleus, containing the cell's chromosomes, or genetic material. The cytoplasm contains everything within the cell, except the nucleus.
2. The endoplasmic reticulum (ER) is responsible for the production of the protein and lipid components of most of the cell's organelles. The ER contains a large number of folds, but the membrane forms a single sheet enclosing a single closed sac. This internal space is called the ER lumen. The smooth endoplasmic reticulum (ER) in muscle cells contains the vesicles and tubules that serve as a store of calcium ions. These are released as one step in the muscle contraction process. Calcium pumps,  $\text{Ca}^{2+}$ -ATPases, serve to move the calcium from the cytoplasm to the ER or SR lumen.
3. The sarcoplasmic reticulum (SR) functions to uptake calcium from the sarcoplasm (the cytoplasm of muscle fiber). The sarcoplasmic reticulum

membrane contains high levels of  $\text{Ca}^{2+}$ -ATPase. With the action of  $\text{Ca}^{2+}$ -ATPase, the SR releases calcium ions into the sarcoplasm to initiate muscle contraction, and it releases the calcium ions into the lumen during relaxation.

The  $\text{Ca}^{2+}$ -ATPase from fast-twitch rabbit skeletal muscle sarcoplasmic reticulum (SERCA1a) is a large membrane protein of 110 kDa that has more than 990 amino acid residues. The enzyme actively transports two  $\text{Ca}^{2+}$  ions per ATP hydrolyzed across the membrane from the cytoplasm to the lumen against a concentration gradient. At the same time, two to three protons are transported in the opposite direction. Activity of  $\text{Ca}^{2+}$ -ATPase is crucial to skeletal muscle function. The reaction cycle, adapted from reference 96, proceeds from the upper left in Figure 6.28 as the high-affinity form of the enzyme,  $E_1$ , receives two moles of calcium ion from the cytoplasm (where  $\text{Ca}^{2+}$  concentration has become unacceptably high) and releases 2–3 protons to the cytoplasm. ATP complexed with  $\text{Mg}^{2+}$  ions enters and is bound to the enzyme at a specific location. The enzyme is then phosphorylated at a specific residue, asp351 (D351) as energy is imparted to the system through an  $\text{ATP} \rightarrow \text{ADP}$  hydrolysis. In the next step, ADP is released. Continuation of the cycle causes calcium ions to be released to the lumen as 2–3 protons enter from the lumen. Phosphate ion is released along with magnesium ions to form the low affinity  $E_2$  state. Transformation of the enzyme's low-affinity state to its high-affinity state restarts the cycle. Figure 6.28 indicates the Protein Data Bank accession numbers for the known X-ray crystallographic structures corresponding to the various cycle intermediates.



**Figure 6.28** The  $\text{Ca}^{2+}$ -ATPase reaction cycle. (Adapted by permission from Macmillan Publishers Ltd from Figure 1 of reference 96, copyright 2004.)

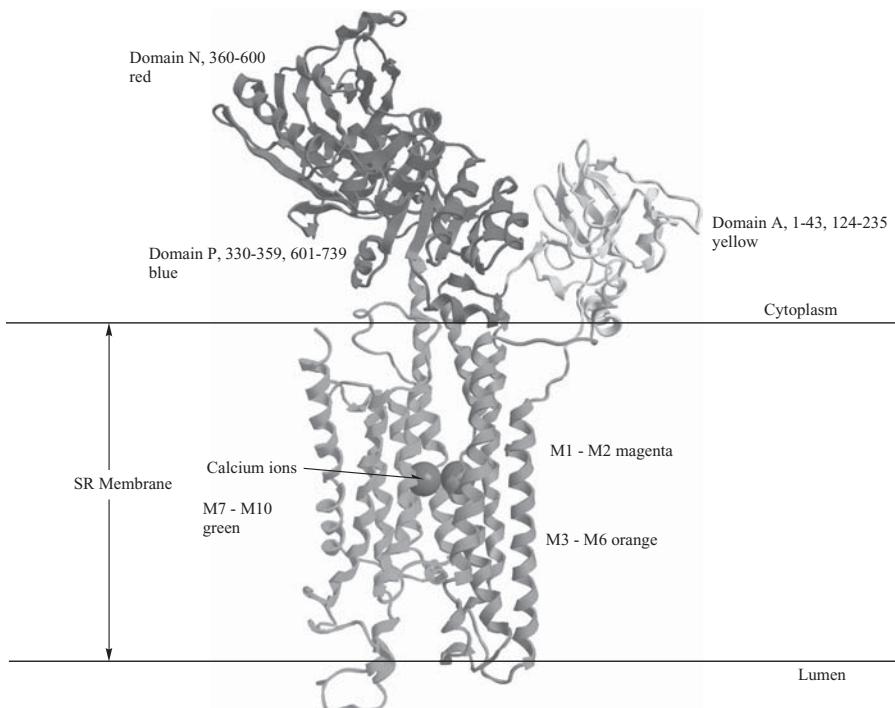
During the cycle, large conformational changes take place within domains of the enzyme to make sure that calcium ions do not “leak” back into the cytoplasm but are “occluded” at binding sites within the SR membrane portion of the enzyme until such time that the enzyme’s “lumenal gate” opens to release the  $\text{Ca}^{2+}$  ions to the lumen.

**6.4.2.1  $\text{Ca}^{2+}$ -ATPase Protein SERCA1a and the  $\text{Ca}^{2+}$ -ATPase Cycle.** In the following paragraphs we will discuss the structural information that has been gathered on crystallizations of the  $\text{Ca}^{2+}$ -ATPase protein from rabbit fast-twitch muscle known as SERCA1a. This will be done through descriptions of the various X-ray crystallographic structures corresponding to stages described in Figure 6.28. The  $\text{Ca}_2\text{E}_1$  state was first described (PDB: 1SU4) by Toyoshima et al.<sup>97</sup> This was the first atomic-scale model for any P-type ATP-dependent pump. The  $\text{Ca}_2\text{E}_1$  state is compared to the  $\text{E}_2$  state (PDB: 1IWO) as described in reference 98. In the PDB: 1IWO X-ray structure, calcium ions are replaced by thapsigargin (TG), a  $\text{Ca}^{2+}$ -ATPase cycle inhibitor.

Very large conformational changes take place between the  $\text{Ca}_2\text{E}_1$  and  $\text{E}_2$  states. (See Figure 6.28.) Moller and co-workers<sup>95</sup> described the structure of a  $\text{Ca}_2\text{E}_1\cdot\text{ATP}$  intermediate in the cycle with  $\text{Ca}_2\text{E}_1\cdot$ adenosine ( $\beta\text{-}\gamma$ -methylene)-triphosphate, AMPPCP (PDB: 1T5S). AMPPCP is an inactive ATP analog that occupies the normal ATP site in the enzyme. Later, the  $\text{Ca}_2\text{E}_1\cdot\text{ATP}$  state is also modeled in the Toyoshima structure deposited as PDB: 1VFP.<sup>99</sup> This structure contains the AMPPCP ATP analog along with calcium and magnesium ions.<sup>99</sup> Moller and co-workers<sup>95</sup> described the  $\text{Ca}_2\text{E}_1\text{P}\cdot\text{ADP}$  intermediate with the structure containing  $\text{AlF}_4^-$  as a phosphate analog (PDB 1T5T). Toyoshima et al.<sup>100</sup> have also described a  $\text{Ca}_2\text{E}_1\text{P}\cdot\text{ADP}$  intermediate containing ADP,  $\text{AlF}_3$ , calcium and magnesium ions (PDB: 1WPE) and an  $\text{E}_2\text{P}$  intermediate state containing ADP,  $\text{MgF}_4^{2-}$ , magnesium and sodium ions, and thapsigargin in the calcium ion sites (PDB: 1WPG). Toyoshima has discussed the PDB: 1WPG structure as an example of the  $\text{E}_2\text{P}$  state, although Moller and co-workers<sup>101</sup> have published the PDB structure PDB: 1XP5 ( $\text{AlF}_4^-$ , magnesium, and potassium ions stabilized by thapsigargin in the calcium ion sites) as an example of the  $\text{E}_2\text{P}$  state. The differences center on the point at which phosphate ions are released to the lumen during the cycle. In this discussion, the Toyoshima group descriptions are followed. Interested readers should consult the more recent literature for updates on the ongoing structural information leading to more accurate mechanistic detail for  $\text{Ca}^{2+}$ -ATPases. One structure eagerly sought by those working in this field would be one in which the asp351 residue is phosphorylated. Toyoshima et al.<sup>100</sup> have modeled a phosphorylated aspartic acid residue from a bacterial response regulator, CheY (PDB: 1QMP), that uses identical residues for stabilization of the phosphoryl group. The modeled aspartylphosphate has been inserted into the  $\text{E}_2\text{P}$  state structure (PDB: 1WPG) containing ADP,  $\text{MgF}_4^{2-}$ , magnesium and sodium ions, and thapsigargin in the calcium ion sites. It is important to remember that all the  $\text{Ca}^{2+}$ -ATPase models discussed are solid-state equilibrium structures of a dynamic system operating

in nature in the solution state. Nevertheless, the models provide an important and fascinating picture of a complex enzyme system. The lessons learned from these studies can be applied to other members of the P-type ATPase superfamily—including  $\text{Na}^+$ ,  $\text{K}^+$ -ATPases (see Section 5.4.1) and  $\text{H}^+$ ,  $\text{Na}^+$ -ATPases—about which less structural information is available.

To introduce details of the SERCA1a protein, we will first look at the large reorganizations that take place between the  $\text{Ca}_2\text{E}_1$  and  $\text{E}_2$  states of the enzyme. Figure 1 of reference 102 shows representations of this  $\text{Ca}^{2+}$ -ATPase with calcium ions (PDB: 1SU4,<sup>97</sup> the  $\text{Ca}_2\text{E}_1$  state) and without calcium ions (PDB: 1IWO, the  $\text{E}_2$  state).<sup>102</sup> The figure includes a representation of the SR membrane as well as a simplified scheme showing the  $\text{Ca}^{2+}$ -ATPase cycle (see Figure 6.28). The SR membrane representation in Figure 1 of reference 102 is an insertion of a bilayer of dioleoylphosphatidylcholine (DOPC) generated through a molecular dynamics calculation by the authors. The PDB: 1SU4 X-ray crystallographic structure (with calcium ions in their binding sites) is shown in this text as Figure 6.29. It is represented as  $\text{E}1\text{Ca}^{2+}$  on the left in Figure 1 of reference 102 (see also Figure 6.28).

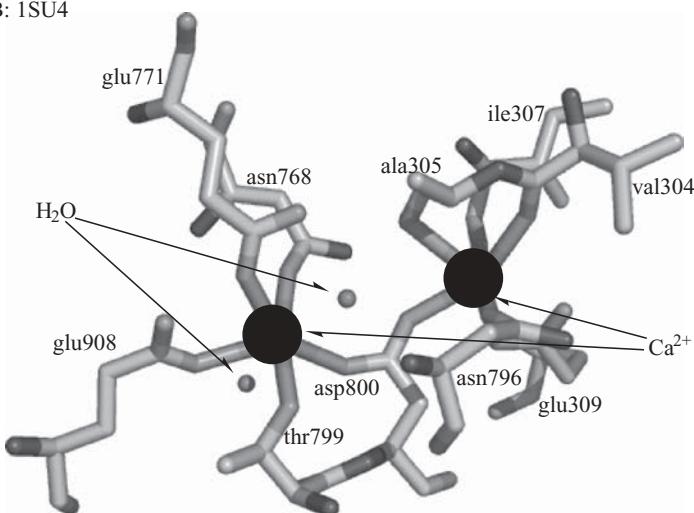


**Figure 6.29**  $\text{Ca}^{2+}$ -ATPase in the  $\text{Ca}_2\text{E}_1$  state as represented by the PDB: 1SU4 X-ray crystallographic structure. Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. (Printed with permission of CambridgeSoft Corporation.) (See color plate)

The right-hand representation in reference 101, Figure 1, without calcium ion and with thapsigargin to stabilize the calcium binding site, is labeled E2(TG). Its X-ray crystallographic structure is deposited as PDB: 1IWO.<sup>98</sup> Figure 1 of reference 98 shows the movement of the N, P, and A domains as the  $\text{Ca}^{2+}$ -ATPase molecule moves from the E1 $\text{Ca}^{2+}$  to the E2(TG) state. The E1 label refers to the high  $\text{Ca}^{2+}$  affinity state in the  $\text{Ca}^{2+}$ -ATPase cycle (Figure 6.28), whereas the E2 label refers to the low  $\text{Ca}^{2+}$  affinity state of the enzyme. In E2(TG), TG refers to thapsigargin, a  $\text{Ca}^{2+}$ -ATPase cycle inhibitor believed to fix the enzyme in an E2 form.  $\text{Ca}^{2+}$ -ATPase denatures in its calcium-free form without the presence of such an inhibitor. Significant shifts in position and orientation of the transmembrane helices take place between the two states, along with large movements of the enzyme domains that protrude into the cytoplasm. First, we will discuss the M1–M10 helical domains of PDB: 1SU4 and 1IWO, followed by discussion of the cytoplasmic domains.

Figure 1 of reference 98 represents a detailed diagram of the E2(TG) (E<sub>2</sub> in Figure 6.28) and E1 $\text{Ca}^{2+}$  (Ca<sub>2</sub>E<sub>1</sub> in Figure 6.28) structures that will be helpful in understanding the differences described below. Figure 6.29 represents the Ca<sub>2</sub>E<sub>1</sub> state. The  $\text{Ca}^{2+}$ -ATPase enzyme portion residing in the membrane consists of 10  $\alpha$ -helices, usually referred to in the literature as helices M1–M10. The M1–M2 helical domain is comprised of enzyme aa residues 44–123. The M1  $\alpha$ -helix contains an amphipathic (containing both polar and nonpolar domains) sector (M1') thought to interface between the cytoplasm and membrane. The orientation of M1' changes substantially between the E2(TG) structure and that of E1 $\text{Ca}^{2+}$  (E<sub>1</sub> in Figure 6.28). In addition, M1 contains a glu58 (E58) and an asp59 (D59), either or both of which are thought to help in guiding calcium ions to their binding location within the membrane portion of  $\text{Ca}^{2+}$ -ATPase. Glu309 (E309) of M4 is another acidic residue that may have access to the cytoplasm and assist calcium ions entering the enzyme. The calcium ions enter the enzyme sequentially and cooperatively; also, the first calcium ion occupies Site I, followed by the second in Site II. M1 and M2 move upwards toward the cytoplasm in moving from E1 $\text{Ca}^{2+}$  (E<sub>1</sub> in Figure 6.28) to other states in the cycle. The M3–M6 helical region, residues 239–329 and 740–821, contains many of the calcium ions ligands. In passing from the E1 $\text{Ca}^{2+}$  state to the E2(TG) state, M4 moves toward the luminal (extracellular) side by one turn of an  $\alpha$ -helix, the upper part of M5 moves toward M4, and the unwound part of M6 rotates by nearly 90°.<sup>102</sup> All of these movements have the effect of decreasing the availability of coordinating oxygen ligands for  $\text{Ca}^{2+}$ .<sup>103</sup> Helices M4–M6 (moving during the E1 $\text{Ca}^{2+}$  to E2(TG) states) and M8 (unmoving) contain the ligands binding  $\text{Ca}^{2+}$  to the enzyme. Membrane helices M7–M10, residues 822–994 are essentially unchanged during the E1 $\text{Ca}^{2+}$  to E2(TG) transition. Site I calcium ions ligands are side-chain oxygens of asn768 (N768) and glu771 (E771) of M5, thr799 (T799) and asp800 (D800) of M6, glu908 (E908) of M8, and two water molecules. Site II ligands include a second side-chain carboxyl oxygen of asp800 plus side-chain oxygens donated by asn796 (N796) of M6 and glu309 (E309) of M4 as well as main-chain carbonyl oxygen

PDB: 1SU4



**Figure 6.30** The calcium ion coordination sphere for the  $\text{Ca}^{2+}$ -ATPase as represented by the PDB: 1SU4 X-ray crystallographic structure. Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft Corporation.) (See color plate)

**TABLE 6.10 Selected Bond Distances for Calcium Ions in  $\text{Ca}^{2+}$ -ATPase, PDB: 1SU4**

| Calcium Ion Site I              |  | Calcium Ion Site II           |                                |
|---------------------------------|--|-------------------------------|--------------------------------|
| Amino Acid Ligand Atom          | Bond Distance to $\text{Ca}^{2+}$ ( $\text{\AA}$ ) | Amino Acid Ligand Atom        | Bond Distance ( $\text{\AA}$ ) |
| Asn768, $\text{O}_{\delta 1}$   | 2.33   | Val304, backbone O            | 2.40                           |
| Glu771, $\text{O}_{\epsilon 2}$ | 2.39   | Ala305, backbone O            | 2.68                           |
| Thr799, $\text{O}_{\gamma 1}$   | 2.36   | Ile307, backbone O            | 2.32                           |
| Asp800, $\text{O}_{\delta 1}$   | 2.38   | Asp800, $\text{O}_{\delta 2}$ | 2.34                           |
| Glu908, $\text{O}_{\epsilon 1}$ | 2.44   | Asn796, $\text{O}_{\delta 1}$ | 2.22                           |
| $\text{H}_2\text{O}_1$          | 2.29   |                               |                                |
| $\text{H}_2\text{O}_2$          | 2.30   |                               |                                |

atoms of residues val304 (V304), ala305 (A305), and ile307 (I307) of M4. Figure 4 of reference 98 shows stereo diagrams of the transmembrane  $\text{Ca}^{2+}$ -binding sites. Figure 6.30 shows the coordination sphere of the two calcium ions of the  $\text{Ca}^{2+}$ -ATPase PDB: 1SU4. Table 6.10 lists selected bond distances for the same  $\text{Ca}^{2+}$ -ATPase PDB: 1SU4.

The cytoplasmic portion of  $\text{Ca}^{2+}$ -ATPase, sometimes called the cytosolic headpiece, contains three important domains. These are (1) the A (actuator or

anchor) domain, (2) the N (nucleotide-binding) domain, and (3) the P (phosphorylation) domain. See Figure 6.29. Domain A is the smallest of these. It contains approximately 110 residues, 1–43, the N-terminal end of the enzyme, and 124–235 falling in sequence between M2 and M3. Domain P is comprised of residues 330–359 and 601–739. The M4 and M5 helices are integrated into the P domain as part of a Rossmann fold, a well-known protein motif containing (usually) a seven-stranded central parallel  $\beta$ -sheet associated with seven surrounding  $\alpha$ -helices. The P domain rotates toward the A domain in passing from the E1Ca<sup>2+</sup> to E2(TG) states. The N domain, residues 360–600, also changes position during the E1Ca<sup>2+</sup> to E2(TG) transition. Domain A rotates and changes position in each of the X-ray and NMR structures known for the Ca<sup>2+</sup>-ATPase cycle and is thought to help control access to the opposite cytosolic and luminal sides of the membrane. In the E1Ca<sup>2+</sup> enzyme (PDB: 1SU4) the domains A, N, and P reside in an “open” position. On going to the E2(TG) structure, the A domain rotates 110°, closing the cytoplasmic gate of the enzyme.

The next discussion will center on changes that take place on moving from the Ca<sub>2</sub>E<sub>1</sub> state to the Ca<sub>2</sub>E<sub>1</sub>·ATP bound state. In the Ca<sub>2</sub>E<sub>1</sub> state, there does not appear to be a relationship between the N domain (nucleotide-binding) and the P domain that contains the phosphorylation site at asp351. Upon addition of nucleotide, in this case the inactive ATP mimic called AMPPCP [adenosine ( $\beta$ - $\gamma$ -methylene)-triphosphate (PDB: 1T5S)<sup>95</sup>], N loop region residues 576–581 form a nucleotide-binding crevice with P loop region residues 720–725. Domain A rotates by approximately 45° to form a lid on one side of this nucleotide-binding pocket. This is illustrated in Figure 2 of reference 95. As the M2 helix continues as a  $\beta$ -strand into the A domain, the rotation after nucleotide binding pulls the M1–M2 helix pair toward the cytoplasm. Helices M1–M2 are lifted by 8–9 Å relative to helices M3–M10. A supplemental movie to reference 95 illustrates the changes (<http://www.sciencemag.org/content/vol304/issue5677/images/data/1672/DC1/1099366s1.mov>). In the movie, domain A appears in yellow, domain N in red, domain P in blue, and membrane helices M1 and M2 in magenta. A well-conserved amino acid sequence in Ca<sup>2+</sup>-ATPases, TGES, starts at thr181 in domain A. This motif helps anchor the A domain and M1–M3 to the P domain, and it is important for carrying out the phosphorylation of asp351 in the P domain. The N domain, residues 360–600 of the enzyme, binds the ATP molecule that enters from the cytoplasm. This ATP will be hydrolyzed to ADP, providing the energy for the release of calcium ions to the lumen. In going from Ca<sub>2</sub>E<sub>1</sub> (PDB: 1SU4) to Ca<sub>2</sub>E<sub>1</sub>·ATP (PDB: 1T5S) the N domain rotates by 90° and closes with the P domain. The ATP analog binds in a highly conserved nucleotide-binding pocket between the N and P domains. Three arginine residues, arg489 and arg560 of the N domain and arg687 of the P domain, stabilize the nucleotide in a bent conformation.<sup>95</sup> The magnesium ion accompanying AMPPCP interacts with its  $\gamma$ -phosphate as well as with side chains of thr353 and thr625 of the P domain and with the main chain amide of gly626 (P domain).

Toyoshima's group has determined the X-ray crystallographic structure of a  $\text{Ca}_2\text{E}_1\cdot\text{ATP}$  state<sup>99</sup> (PDB: 1VFP). The structure shows a rotation of the A domain by about  $30^\circ$  around an axis approximately parallel to the membrane thus making contact with the N domain. The N domain is inclined toward the P domain facilitated by binding of the ATP analogue, AMPPCP. The AMPPCP molecule makes contact with the ring of phe487 and the  $\gamma$ -phosphate of the AMPPCP binds near asp351, the residue to be phosphorylated. Movement of the A domain moves membrane helices M1 and M2. M2 moves one turn of the  $\alpha$ -helix toward the cytoplasm and M1 is drawn up by two turns of its  $\alpha$ -helix. M1 bends at asp59 and the amphipathic (polar–nonpolar) M1' helix lies on the surface of the membrane. Movement of M1 also locks glu309 into place, preventing Site II  $\text{Ca}^{2+}$  ions from escaping back into the cytoplasm—this is called the occluded state. Thus movement of the A domain and its associated M1 helix are of critical importance in locking  $\text{Ca}^{2+}$  ions into place within the membrane. The M3 loop connection to the A domain becomes strained in this process, in preparation for further movement in later parts of the cycle. In addition, the calcium coordinating residues glu309 and asp800 become looser, in preparation for removal of  $\text{Ca}^{2+}$  to the lumen. The AMPPCP molecule is held in place through (1) stacking interaction of the adenine ring with phe487 (N domain), (2) stabilizing the  $\alpha$ -phosphate by hydrogen bonding to arg489 (N domain) and the  $\beta$ -phosphate to arg560 (N domain), and (3) hydrogen bonding between residues in the N and P domains. The distance from the  $\gamma$ -phosphate to the phosphorylation residue, asp351, is  $3.4\text{ \AA}$  in the PDB: 1VFP structure. The P domain Rossmann fold, consisting of a central seven-stranded parallel  $\beta$ -sheet and associated seven  $\alpha$ -helices, moves so that some of its component residues bond to the  $\gamma$ -phosphate and its associated  $\text{Mg}^{2+}$  ions. The movements bend the P domain into two almost orthogonal sectors, and this part of the P domain is brought closer to the N domain.

Two X-ray crystallographic structures, PDB: 1T5T<sup>95</sup> and PDB: 1WPE,<sup>100</sup> describe the product of the next part of the cycle:  $\text{Ca}_2\text{E}_1\cdot\text{ATP}$  to  $\text{Ca}_2\text{E}_1\text{P}\cdot\text{ADP}$ . Sorenson's work in reference 95 shows that the  $\text{Ca}_2\text{E}_1\cdot\text{ATP}$  and  $\text{Ca}_2\text{E}_1\text{P}\cdot\text{ADP}$  structures are quite similar. The Sorensen  $\text{Ca}_2\text{E}_1\text{P}\cdot\text{ADP}$  model (PDB: 1T5T) contains the  $\text{ADP}:\text{AlF}_4^-$  molecule, two calcium ions, two magnesium ions, and a potassium ion. The potassium ion may be important for facilitating the dephosphorylation of the enzyme in a later part of the cycle. The PDB: 1T5T structure is intended to mimic the transition state for phosphoryl transfer to asp351 with the tetrafluoroaluminate ion acting as a nonreactive  $\gamma$ -phosphate ion substitute. The differences between the  $\text{Ca}_2\text{E}_1\cdot\text{ATP}$  (PDB: 1T5T) and  $\text{Ca}_2\text{E}_1\text{P}\cdot\text{ADP}$  (PDB: 1T5S) structures lie primarily at the phosphorylation site. The  $\text{Ca}_2\text{E}_1\cdot\text{ATP}$  model shows that the  $\gamma$ -phosphate and the asp351 residue are not close enough for phosphate ion transfer. In this structure, an  $\text{Mg}^{2+}$  ion has coordinated to an asp351 carboxyl ion but not to the  $\gamma$ -phosphate of the AMPPCP. In the  $\text{Ca}_2\text{E}_1\text{P}\cdot\text{ADP}$  model, a second magnesium ion is found between the  $\alpha$ - and  $\beta$ -phosphates, possibly to stabilize the soon-to-be-leaving ADP molecule, and the aluminum atom of the  $\text{AlF}_4^-$  moiety has been pulled

within an approximate 2-Å bonding distance of the asp351. The nearly linear arrangement of the  $\text{AlF}_4^-$ 's aluminum ion and asp351 carboxyl moieties suggest an associative nucleophilic reaction mechanism ( $S_N2$ ) for the phosphoryl transfer involving a pentavalent phosphorus intermediate. Side chains of two aa residues, thr353 and thr625, form hydrogen bonds to the  $\beta$ -phosphate and the  $\text{AlF}_4^-$ , respectively, and the main chain amides of gly626 and thr353 form hydrogen bonds to the  $\text{AlF}_4^-$ , further stabilizing the structure. In the PDB: 1T5T structure, two calcium ions are occluded. The reference 95 authors believe that changes in the A domain have aligned hydrophobic residues in membrane helices M1 (phe57, val62, leu65) and M2 (ile94, leu97, leu98) to close off  $\text{Ca}^{2+}$  ion access to the cytoplasm—the occluded state.

The Toyoshima group has determined the structure of a  $\text{Ca}^{2+}$ -ATPase that they believe mimics the  $\text{Ca}_2\text{E}_1\text{P}\cdot\text{ADP}$  state as well.<sup>100</sup> This structure, PDB: 1WPE, contains an  $\text{AlF}_3\cdot\text{ADP}$  molecule, two calcium ions, and two magnesium ions. In agreement with the reference 95 work, the structure is very similar to the  $\text{Ca}_2\text{E}_1\cdot\text{ATP}$  model (PDB: 1T5S). The P domain is hardly altered from the  $\text{Ca}_2\text{E}_1\cdot\text{ATP}$  structure, and the N and P domains are strongly associated to hold the ADP molecule in place. The important TGES motif in the A domain, beginning at thr181 and continuing through gly182, glu183, and serine 184, participates in hydrogen bonding near the phosphorylation site. Glu183 is an especially important residue, hydrogen bonding through a water molecule to the phosphorylation residue asp 351. The TGES motif may also participate in holding (or releasing) ADP. A movie included as supplemental information with reference 100 illustrates many of the points being discussed here. The movie can be obtained as supplemental information in the reference 100 citing available at [www.nature.com/nature](http://www.nature.com/nature).

The next structures describe the transition from  $\text{Ca}_2\text{E}_1\text{P}\cdot\text{ADP}$  to the  $\text{E}_2\text{P}$  and the  $\text{E}_2\text{P}_i$  states. During this transition, calcium ions have been released to the lumen and ADP has been released as well. Two models have been described. The structure described in reference 101 has been deposited in the PDB as 1XP5 and most closely models the  $\text{E}_2\text{P}$  state, whereas the Toyoshima group structure,<sup>100</sup> PDB: 1WPG, models  $\text{E}_2\text{P}_i$ . The  $\text{E}_2\text{P}$  state is called the ADP-insensitive state. This state is incapable of regenerating ATP, because the enzyme has reconfigured itself once again to allow the release of ADP. In the dephosphorylation transition state depicted by the PDB: 1XP5 structure, calcium ions are no longer occluded and proton counterions have found their way into the enzyme from the lumen. The PDB: 1XP5 structure contains a tetrafluoroaluminate ion,  $\text{AlF}_4^-$ , a potassium ion, a magnesium ion, and a thapsigargin molecule, necessary for stabilizing the now empty calcium ion sites. Release of ADP upon phosphoryl transfer initiates an approximate 50° rotation of the N domain relative to the P domain. This, in turn, enables the A domain to rotate 108° and shift 8 Å around the P domain bringing the TGES motif in apposition to the phosphorylation site. Connections of domain A to the membrane helices M1–M3 effect large changes in their orientation and position. The M1–M3 movements affect the orientation of M4 through M6 relative to M7–M10 that

remain relatively static. In addition, the P domain rotates approximately 30° relative to the M7–M10 segment. All these motions are relative to the Ca<sub>2</sub>E<sub>1</sub>P·ADP model (PDB: 1T5T). The rearrangements and movements destabilize the Ca<sup>2+</sup> binding sites in the membrane but at the same time prevent Ca<sup>2+</sup> movement back to the cytoplasm. The rearrangements have permitted Ca<sup>2+</sup> release into the lumen in exchange for 2–3 protons, most likely now protonating some of the former carboxylate Ca<sup>2+</sup>-binding sites. At the phosphorylation site, the AlF<sub>4</sub><sup>−</sup> group is located between the asp351 residue (P domain) and a water molecule hydrogen-bonded to the main chain carbonyl of thr181 and the side-chain carboxyl of glu183. These residues are part of the important A domain TGES motif as mentioned above. It is believed that the glu183 side chain provides the general base catalysis, abstracting a proton from the water molecule for attack on the phosphorylated residue (here represented by AlF<sub>4</sub><sup>−</sup>). In addition, the TGES motif occupies the former position of ADP between the N and P domains. The associative reaction mechanism is invoked for dephosphorylation (as it was for phosphorylation) because all participants in the reaction are again in a linear arrangement. Reference 101 includes a supplemental movie illustration of the Ca<sup>2+</sup>-ATPase enzyme action available at <http://www.sciencemag.org/content/vol306/issue5705/images/data/2251/DC1/1106289s1.mov>. This movie illustrates the following scenario. Calcium ions enter binding sites within the membrane, are trapped there, and cannot exchange with external calcium ions. ATP, complexed with magnesium, enters and phosphorylates the pertinent amino acid residue in the presence of potassium (or sodium) ion. Phosphoryl transfer from ATP to the enzyme closes the entry hatch for calcium ions. Release of adenosine diphosphate (ADP) opens the exit hatch and allows calcium to escape into the lumen in exchange for protons. Two calcium ions leave in exchange with two or three protons. Finally, release of phosphate and magnesium opens the calcium entry hatch and allows the exchange of protons for calcium. The proposed sequence of events becomes clearer if one steps through the movie after watching it at full speed a few times.

While the Sorensen PDB: 1XP5 structure is thought to be in a E<sub>2</sub>P transition state, the Toyoshima group PDB: 1WPG structure is projected as the E<sub>2</sub>P<sub>i</sub> ground state.<sup>100</sup> This Ca<sup>2+</sup>-ATPase model contains thapsigargin in the Ca<sup>2+</sup> site, the tetrafluoromagnesate ion (MgF<sub>4</sub><sup>2−</sup>) as the bound phosphate analog, Mg<sup>2+</sup>, and Na<sup>+</sup>. Additionally, ADP may bond to the PDB: 1WPG structure, although as a model of the E<sub>2</sub>P<sub>i</sub> state, ADP's site would be occupied by part of a rotated A domain. In the PDB: 1WPG structure, the N and P domains have dissociated somewhat because the ADP molecule holding them together has been lost. Meanwhile the A and P domains have more tightly associated as compared to the Ca<sub>2</sub>E<sub>1</sub>P·ADP modeled by the PDB: 1WPE structure. In fact, the PDB: 1WPG domain A has rotated approximately 110° around an axis perpendicular to the membrane compared to its position in PDB: 1WPE. Rearrangements in the cytoplasmic domains A, N, and P cause marked changes in the transmembrane helices as well. Rotation of the A domain tilts the M1–M2 helices

that in turn interact with M4 and other membrane helices that contain  $\text{Ca}^{2+}$  ligands. The overall effect is to move carboxylate  $\text{Ca}^{2+}$  ligands and decrease affinity for calcium ions, allowing them to be released to the lumen.

An overall scheme for  $\text{Ca}^{2+}$ -ATPase activity, based on structures of five different states in the  $\text{Ca}^{2+}$ -ATPase cycle, has been proposed by Toyoshima et al.<sup>100</sup>:

1.  $\text{Ca}^{2+}$  binds to the  $E_2$  ground state, breaking the closed configuration of the A, N, and P domains and exposing the ATP-binding site.  $\text{Ca}^{2+}$  ions bind to sites in transmembrane helices M4, M5, M6, and M8. The cytoplasmic gate is open and  $\text{Ca}^{2+}$  ions interchange with the cytoplasm.
2. ATP binds, crosslinking the N and P domains. The ATP's  $\gamma$ -phosphate and its bound  $\text{Mg}^{2+}$  ion bend the P domain. The N-domain makes contact with the A domain in a strained conformation. Part of the M1 helix is pulled up and forms a barrier to  $\text{Ca}^{2+}$  ion exchange with the cytoplasm. Calcium ions are occluded.
3. Phosphoryl transfer to asp351 allows dissociation of ADP, and then the N and P domain interface opens. The A domain rotates so that the TGES loop closes into the gap left by the loss of ADP. The A domain rotation, in turn, causes marked rearrangement in transmembrane helices, opening the luminal gate and releasing  $\text{Ca}^{2+}$  ions into the lumen.
4. The TGES loop fixes a particular water molecule that catalyzes the hydrolysis of the aspartylphosphate. Release of inorganic phosphate and  $\text{Mg}^{2+}$  unbends the P domain. Transmembrane helices again rearrange to close the luminal gate.

There are differences between Sorensen et al.<sup>95</sup> and Toyoshima et al.<sup>100</sup> interpretations in the  $E_2\text{P}$  and the  $E_2\text{P}_i$  states concerning the question, When does the luminal gate close? In Sorensen's PDB: 1XP5 structure,  $E_2\text{P}$ , the luminal gate has closed and asp351 is still phosphorylated. The Toyoshima PDB: 1WPG  $E_2\text{P}_i$  structure appears to show that closing of the luminal gate is triggered by phosphate release. Resolution of this difference will come with additional information on the complex mechanism and movements of the  $\text{Ca}^{2+}$ -ATPase enzyme during its cycle. The reader should look for literature updates on  $\text{Ca}^{2+}$ -ATPase intermediates to answer these questions.

## 6.5 CONCLUSIONS

Chapter 6 discussed Group II metal ions in biomolecules, concentrating on magnesium ions in catalytic RNA and on two calcium-containing biomolecules: calmodulin and  $\text{Ca}^{2+}$ -ATPase. Readers interested in the evolutionary aspects of catalytic RNA as a precursor to the DNA-based life forms that exist in the present time could begin by consulting the publications listed in

reference 104. Readers interested in troponin structure and function should consult reference 105, in which Fletterick and co-workers study the structural details of skeletal muscle troponin's subunits TnC, TnI, and TnT using X-ray crystallography. Several researchers working in the field of calcium pumps have written a review on the structural basis for coupling of  $\text{Ca}^{2+}$  transport to ATP hydrolysis by the sarcoplasmic reticulum  $\text{Ca}^{2+}$ -ATPase.<sup>106</sup>

## REFERENCES

1. Cech, T. R. *Biochem. Soc. Trans.*, 2002, **30**(6), 1162–1166.
2. Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R. *Cell*, 1982, **31**, 147–157.
3. Cech, T. R.; Zaug, A. J.; Grabowski, P. J. *Cell*, 1981, **27**, 487–496.
4. (a) Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. *Cell*, 1983, **35**, 849–857. (b) Guerrier-Takada, C.; Altman, S. *Science*, 1984, **223**, 285–286.
5. (a) Michel, F.; Dujon, B. *EMBO J.*, 1983, **2**, 33–38. (b) Waring, R. B.; Scazzocchio, C.; Brown, T. A.; Davies, R. W. *J. Mol. Biol.*, 1983, **167**, 595–605.
6. Gott, J.; Shub, D.; Belfort, M. *Cell*, 1986, **47**, 81–87.
7. (a) Peebles, C. L.; Perlman, P. S.; Mecklenburg, K. L.; Petrillo, M. L.; Tabor, J. H.; Jarrell, K. A.; Cheng, H. L. *Cell*, 1986, **44**, 213–223. (b) van der Veen, R.; Arnberg, A. C.; van der Horst, G.; Bonen, L.; Tabak, H. F.; Grivell, L. A. *Cell*, 1986, **44**, 225–234.
8. (a) Prody, G. A.; Bakos, J. T.; Buzavan, J. M.; Schneider, I. S.; Bruening, G. *Science*, 1986, **231**, 1577–1580. (b) Forster, A. C.; Symons, R. H. *Cell*, 1987, **49**, 211–220. (c) Uhlenbeck, O. C. *Nature*, 1987, **328**, 596–600.
9. Sigler, P. B. *Annu. Rev. Biophys. Bioeng.*, 1975, **4**, 447–527.
10. Pyle, A. M.; Murphy, F. L.; Cech, T. R. *Nature*, 1992, **358**, 123–128.
11. Wang, J-F.; Downs, W. D.; Cech, T. R. *Science*, 1993, **260**, 504–508.
12. Piccirilli, J. A.; Vyle, J. S.; Caruthers, M. H.; Cech, T. R. *Nature*, 1993, **361**, 85–88.
13. Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science*, 1996, **273**, 1678–1685. (PDB: 1GID)
14. Pley, H. W.; Flaherty, K. M.; McKay, D. B. *Nature*, 1994, **372**, 68–74. (PDB: 1HMH)
15. Golden, B. L.; Gooding, A. R.; Podell, E. R.; Cech, T. R. *Science*, 1998, **282**, 259–264. (PDB: 1GRZ)
16. DeRose, V. J. *Curr. Opin. Struct. Biol.*, 2003, **13**, 317–324.
17. Hammann, C.; Lilley, K. M. J. *ChemBioChem.*, 2002, **3**, 690–700.
18. Hampel, K. J.; Burke, J. M. *Biochemistry*, 2003, **42**, 4421–4429.
19. (a) Fang, X.-W.; Thiagarajan, P.; Sosnick, T. R.; Pan, T. *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 8518–8523. (b) Russell, R.; Millett, I. S.; Tate, M. W.; Kwok, L. W.; Nakatani, B.; Gruner, S. M.; Michrie, S. G. J.; Pande, V.; Doniach, S.; Herschlag, D.; Pollack, L. *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 4266–4271.
20. Maderia, M.; Hunsicker, L. M.; DeRose, V. J. *Biochemistry*, 2000, **39**, 12113–12120.

21. Morrissey, S. R.; Horton, T. E.; DeRose, V. J. *J. Am. Chem. Soc.*, 2000, **122**, 3473–3481.
22. Adams, P. L.; Stahley, M. R.; Kosek, A. B.; Wang, J.; Strobel, S. A. *Nature*, 2004, **430**, 45–50. (PDB: 1U6B)
23. Celander, D. W.; Cech, T. R. *Science*, 1991, **251**, 401–407.
24. Guo, F.; Gooding, A. R.; Cech, T. R. *Mol. Cell*, 2004, **16**, 351–362. (PDB: 1X8W)
25. Varani, G.; McClain, W. H. *EMBO Reports*, 2000, **1**(1), 18–23.
26. (a) Christian, E. L.; Yarus, M. *J. Mol. Biol.*, 1992, **228**, 743–758. (b) Christian, E. L.; Yarus, M. *Biochemistry*, 1993, **32**, 4475–4480. (c) Ortoleva-Donnelly, L.; Szewczak, A. A.; Gutell, R. R.; Strobel, S. A. *RNA*, 1998, **4**, 498–519.
27. (a) Shan, S.; Yoshida, A.; Sun, S.; Piccirilli, J. A.; Herschlag, D. *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 12299–12304. (b) Kravchuk, A. V.; Piccirilli, J. A.; Herschlag, D. *Biochemistry*, 2001, **40**, 5161–5171.
28. (a) Michel, F.; Westhof, E. *J. Mol. Biol.*, 1990, **216**, 585–610. (b) Lehnert, V.; Jaeger, L.; Michel, F.; Westhof, E. *Chem. Biol.*, 1996, **3**, 993–1009.
29. Juneau, K.; Podell, E.; Harrington, D. J.; Cech, T. R. *Structure*, 2001, **9**, 221–231. (PDB: 1HR2)
30. Doherty, E. A.; Batey, R. T.; Masquida, B.; Doudna, J. A. *Nat. Struct. Biol.*, 2001, **8**, 339–343.
31. Stahley, M. R.; Strobel, S. A. *Science*, 2005, **309**, 1587–1590. (PDB: 1ZZN)
32. Golden, B. L.; Kim, H.; Chase, E. *Nat. Struct. Mol. Biol.*, 2005, **12**, 82–89. (PDB: 1Y0Q)
33. Hampel, K. J.; Burke, J. M. *Biochemistry*, 2003, **42**, 4421–4429.
34. Hertel, K. J.; Pardi, A.; Uhlenbeck, O. C.; Koizumi, M.; Ohtsuka1, E.; Uesugi, S.; Cedergren, R.; Eckstein, F.; Gerlach, W. L.; Hodgson, R.; Symons, R. H. *Nucleic Acids Res.*, 1992, **20**, 3252.
35. Scott, W. G.; Finch, J. T.; Klug, A. *Cell*, 1995, **81**, 991–1002. (PDB: 1MME)
36. (a) Buzayan, J. M.; van Tol, H.; Feldstein, P. A.; Breuning, G. *Nucleic Acids Res.*, 1990, **18**, 4447–4451. (b) Ruffner, D. E.; Uhlenbeck, O. C. *Nucleic Acids Res.*, 1990, **18**, 6025–6029.
37. (a) Ruffner, D. E.; Stormo, G. D.; Uhlenbeck, O. C. *Biochemistry*, 1990, **29**, 10695–10702. (b) Tuschl, T.; Eckstein, F. *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 6691–6694.
38. Peracchi, A.; Beigelman, L.; Scott, E. C.; Uhlenbeck, O. C.; Herschlag, D. *J. Biol. Chem.*, 1997, **272**(43), 26822–26826.
39. Scott, W. G.; Murray, J. B.; Arnold, J. R. P.; Stoddard, B. L.; Klug, A. *Science*, 1996, **274**, 2065–2069. (PDB: 299D, 300D and 301D)
40. Hermann, T.; Auffinger, P.; Scott, W. G.; Westhof, E. *Nucleic Acids Res.*, 1997, **25**(17), 3421–3427.
41. Peracchi, A.; Beigelman, L.; Usman, N.; Herschlag, D. *Proc. Natl. Acad. Sci. USA*, 1996, **93**(21), 11522–11527.
42. Peracchi, A.; Matulic-Adamic, J.; Wang, S.; Beigelman, L.; Usman, N.; Herschlag, D. *RNA*, 1998, **4**, 1332–1346.
43. Feig, A. L.; Scott, W. G.; Uhlenbeck, O. C. *Science*, 1998, **279**, 81–84. (PDB: 359D)
44. Murray, J. B.; Terwey, D. P.; Maloney, L.; Karpeisky, A.; Usman, N.; Beigelman, L.; Scott, W. G. *Cell*, 1998, **92**, 665–673. (PDB: 379D)

45. Wang, S.; Karbstein, K.; Peracchi, A.; Beigelman, L.; Herschlag, D. *Biochemistry*, 1999, **38**, 14363–14378.
46. Morrissey, S. R.; Horton, T. E.; DeRose, V. J. *J. Am. Chem. Soc.*, 2000, **122**, 3473–3481.
47. O'Rear, J. L.; Wang, S.; Feig, A. L.; Beigelman, L.; Uhlenbeck, O. C.; Herschlag, D. *RNA*, 2001, **7**, 537–545.
48. Murray, J. B.; Seyhan, A. A.; Walter, N. G.; Burke, J. M.; Scott, W. G. *Chem. Biol.*, 1998, **5**, 587–595.
49. Murray, J. B.; Scott, W. G. *J. Mol. Biol.*, 2000, **296**, 33–41.
50. Jones, T. A.; Kjeldgaard, M. *O Version 6.1.0*, Department of Molecular Biology, BMC, Uppsala University, Sweden and Department of Chemistry, Aarhus University, Denmark, 1997.
51. Brünger, A. T. *X-PLOR 3.1: A System for Crystallography and NMR*, Yale University Press, New Haven, CT, 1993.
52. Murray, J. B.; Szoke, H.; Szoke, A.; Scott, W. G. *Mol. Cell.*, 2000, **5**, 279–287. (PDB: 488D)
53. Scott, W. G. *J. Mol. Biol.*, 2001, **311**, 989–999.
54. (a) Storm, D. R.; Koshland, D. E.; Jr. *Proc. Natl. Acad. Sci. USA*, 1970, **66**, 445–452.  
(b) Danforn, A.; Koshland, D. E.; Jr. *Proc. Natl. Acad. Sci. USA*, 1971, **68**, 2463–2467. (c) Mesecar, A. D.; Stoddard, B. L.; Koshland, D. E.; Jr. *Science*, 1997, **277**, 202–206.
55. Murray, J. B.; Dunham, C. M.; Scott, W. G. *J. Mol. Biol.*, 2002, **315**, 121–130.
56. Hampel, K. J.; Burke, J. M. *Biochemistry*, 2003, **42**, 4421–4429.
57. Hampel, K. J.; Burke, J. M. *Methods*, 2001, **23**, 233–239.
58. Dunham, D. M.; Murray, J. B.; Scott, W. G. *J. Mol. Biol.*, 2003, **332**, 327–336. (PDB: 1NYI, 1Q29)
59. Stage-Zimmerman, T. K.; Uhlenbeck, O. C. *Nature Struct. Biol.*, 2001, **8**, 863–867.
60. Blount, K. F.; Uhlenbeck, O. C. *Biochemistry*, 2002, **41**, 6834–6841.
61. Blount, K. F.; Uhlenbeck, O. C. *Annu. Rev. Biophys. Biomol. Struct.*, 2005, **34**, 415–440.
62. Suzumura, K.; Warashina, M.; Yoshinari, K.; Tanaka, Y.; Kuwabara, T.; Orita, M.; Taira, K. *FEBS Lett.*, 2000, **473**, 106–112.
63. Osborne, E. M.; Schaak, J. E.; DeRose, V. J. *RNA*, 2005, **11**, 187–196.
64. Martick, M.; Scott, W. G. *Cell*, 2006, **126**(2), 309–320. (PDB: 2GOZ)
65. Nelson, J. A.; Uhlenbeck, O. C. *Mol. Cell*, 2006, **23**(4), 447–450.
66. Cowan, J. A. *Inorganic Biochemistry, An Introduction*, 2nd ed., Wiley-VCH, New York, 1997. pp. 257–286, 292–303.
67. Wilson, M. A.; Brunger, A. T. *J. Mol. Biol.*, 2000, **301**, 1237–1258. (PDB: 1EXR)
68. Chou, J. J.; Li, S.; Klee, C. B.; Bax, A. *Nat. Struct. Biol.*, 2001, **8**, 990–997. (PDB: 1J7O, 1J7P)
69. Vassylyev, D. G.; Takeda, S.; Wakatsuki, S.; Maeda, K.; Maeda, Y. *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 4847–4852.
70. Takeda, S.; Yamashita, A.; Maeda, K.; Maeda, Y. *Nature*, 2003, **424**, 35–41.

71. Babu, Y. S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.*, 1988, **204**(1), 191–204. (PDB: 3CLN)
72. Chattopadhyaya, R.; Meador, W. E.; Means, A. R.; Quiocho, F. A. *J. Mol. Biol.*, 1992, **228**, 1177–1192. (PDB: 1CLL)
73. Schumacher, M. A.; Crum, M.; Miller, M. C. *Structure (Camb)*, 2004, **12**(5), 849–860. (PDB: 1QX5)
74. Zhang, M.; Tanaka, T.; Ikura, M. *Nat. Struct. Biol.*, 1995, **9**, 758–767. (PDB: 1DMO)
75. Kuboniwa, H.; Tjandra, N.; Grzesiek, S.; Ren, H.; Klee, C. B.; Bax, A. *Nat. Struct. Biol.*, 1995, **2**(9) 768–776. (PDB: 1CFC, 1CFD)
76. Finn, B. E.; Evenas, J.; Drakenberg, T.; Walther, J. P.; Thulin, E.; Forsen, S. *Nat. Struct. Biol.*, 1995, **2**, 777–783. (1CMF, 1CMG).
77. Ishida, H.; Nakashima, K.; Kumaki, Y.; Nakata, M.; Hikichi, K.; Yazawa, M. *Biochemistry*, 2002, **41**(52), 15536–15542. (PDB: 1LKJ)
78. Kainosho, M.; Torizawa, T.; Iwashita, Y.; Terauchi, T.; Ono, A. M.; Guntert, P. *Nature*, 2006, **440**, 52–57. (PDB: 1X02)
79. (a) Ishida, H.; Takahashi, K.; Nakashima, K.; Kumaki, Y.; Nakata, M.; Hikichi, K.; Yazawa, M. *Biochemistry*, 2000, **39**(45), 13660–13668. (PDB: 1F54, 1F55) (b) Finn, B. E.; Evenas, J.; Drakenberg, T.; Walther, J. P.; Thulin, E.; Forsen, S. *Nat. Struct. Biol.*, 1995, 777–783. (PDB: 1CMF, 1CMG).
80. Chou, J. J.; Li, S.; Bax, A. *J. Biomol. NMR*, 2000, **18**, 217–227. (PDB: 1F70, 1F71)
81. Yamaotsu, N.; Suga, M.; Hirono, S. *Biopolymers*, 2001, **58**(4), 410–421.
82. Cook, W. J.; Walter, L. J.; Walter, M. R. *Biochemistry*, 1994, **33**(51), 15259–15265. (PDB: 1CTR)
83. Vertessy, B. G.; Harmat, V.; Bocskei, Z.; Naray-Szabo, G.; Orosz, F.; Ovadi, J. *Biochemistry*, 1998, **37**, 15300–15310. (PDB: 1A29)
84. Vandonselaar, M.; Hickiem R. A.; Quail, J. W.; Delbaere, L. T. *Nat. Struct. Biol.*, 1994, **1**(11), 795–801. (PDB: 1LIN)
85. Ikura, M.; Clore, G. M.; Gronenborn, A. M.; Zhu, G.; Klee, C. B.; Bax, A. *Science*, 1992, **256**, 632–638. (PDB: 2BBM)
86. (a) Meador, W. E.; Means, A. R.; Quiocho, F. A. *Science*, 1992, **257**, 1251–1255. (PDB: 1CDL) (b) Meador, W. E.; Means, A. R.; Quiocho, F. A. *Science*, 1993, **263**, 1718–1721. (PDB: 1CDM)
87. Elshorst, B.; Hennig, M.; Forsterling, H.; Diener, A.; Maurer, M.; Schulte, P.; Schwalbe, H.; Griesinger, C.; Krebs, J.; Schmid, H.; Vorherr, T.; Carafoli, E. *Biochemistry*, 1999, **38**, 12320–12332. (PDB: 1CFF)
88. Vetter, S. W.; Leclerc, E. *Eur. J. Biochem.*, 2003, **270**, 404–414.
89. Osawa, M.; Tokumitsu, H.; Swindells, M. B.; Kurihara, H.; Orita, M.; Shibanuma, T.; Furuya, T.; Ikura, M. *Nat. Struct. Biol.*, 1999, **6**, 404–414. (PDB: 1CKK)
90. Schumacher, M. A.; Rivard, A.; Bachinger, H. P.; Adelman, J. P. *Nature*, 2001, **410**, 1120–1124. (PDB: 1G4Y)
91. Drum, C. L.; Yan, S.-Z.; Bard, J.; Shen, Y.-Q.; Lu, D.; Soelaiman, S.; Grabarek, Z.; Bohm, A.; Tang, W.-J. *Nature*, 2002, **415**, 396–402. (PDB: 1K8T, 1K93, 1K90)
92. Kurokawa, H.; Osawa, M.; Kurihara, H.; Katayama, N.; Tokumitsu, H.; Swindells, M. B.; Kainosho, M.; Ikura, M. *J. Mol. Biol.*, 2001, **312**, 59–68. (PDB: 1IQ5)

93. Yamauchi E.; Nakatsu T.; Matsubara M.; Kato H.; Taniguchi H. *Nat. Struct. Biol.*, 2003, **10**(3), 226–231. (PDB: 1IWQ)
94. Schumacher, M. A.; Rivard, A.; Bachinger, H. P.; Adelman, J. P. *Nature*, 2001, **410**, 1120–1123. (PDB: 1G4Y)
95. Sorensen, T. L-M.; Moller, J. V.; Nissen, P. *Science*, 2004, **304**, 1672–1675. (PDB: 1T5S, 1T5T)
96. Lancaster, C. R. D. *Nature*, 2004, **432**, 286–287.
97. Toyoshima, C.; Nakasako, M.; Nomura, H.; Ogawa, H. *Nature*, 2000 **405**, 647–655. (PDB: 1SU4)
98. Toyoshima, C.; Nomura, H. *Nature*, 2002, **418**, 605–611. (PDB: 1IWO)
99. Toyoshima, C.; Mizutani, T. *Nature*, 2004, **430**, 529–535. (PDB: 1VFP)
100. Toyoshima, C.; Nomura, H.; Tsuda T. *Nature*, 2004, **432**, 361–368. (PDB: 1WPE, 1WPG)
101. Olesen, C.; Sorensen, T. L-M.; Nielsen, R. C.; Moller, J. V.; Nissen, P. *Science*, 2004, **306**, 2251–2255. (PDB: 1XP5)
102. Toyoshima, C.; Nomura, H.; Sugita, Y. *FEBS Lett.*, 2003, **555**, 106–110.
103. Clausen, J. D.; Andersen, J. P. J. *Biol. Chem.*, 2004, **279**(52), 54426–54437.
104. (a) Westhof, E.; Massire, C. *Science*, 2004, **306**, 62–63. (b) Schultes, E. A.; Bartel, D. P. *Science*, 2000, **289**, 448–452. (c) Ikawa, Y.; Tsuda, K.; Matsumura, S.; Atsumi, S.; Inoue, T. *Nucleic Acids Res.*, 2003, **31**(5), 1488–1496.
105. Vinogradova, M. V.; Stone, D. B.; Malanina, G. G.; Karatzafiri, C.; Cooke, R.; Mendelson, R. A.; Fletterick, R. J. *Proc. Natl. Acad. Sci. USA*, 2005, **102**(14), 5038–5043.
106. Moller, J. V.; Olesen, C.; Jensen, A. M.; Nissen, P. J. *Bioenerg. Biomembr.*, 2005, **37**(6), 359–364.

---

# 7

---

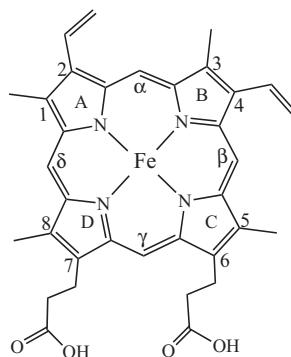
## IRON-CONTAINING PROTEINS AND ENZYMES

### 7.1 INTRODUCTION: IRON-CONTAINING PROTEINS WITH PORPHYRIN LIGAND SYSTEMS

Iron is the most abundant transition metal found in biological systems with a percentage by weight in the human body—for instance,  $5.0 \times 10^{-3}\%$ . Therefore, it is not surprising that iron-containing proteins and enzymes are found in huge numbers and varieties in all biological species. First, one might classify iron-containing species in two categories: those containing a porphyrin ligand system—an iron-bearing heme moiety—and those not containing porphyrin ligands—non-heme iron-containing proteins. This chapter will describe some proteins and enzymes with porphyrin ligand systems with non-heme systems being described briefly in Section 7.9. A section devoted to myoglobin and hemoglobin (7.2) will be followed by six sections (7.3–7.8) discussing the ubiquitous cytochrome metalloenzymes.

### 7.2 MYOGLOBIN AND HEMOGLOBIN

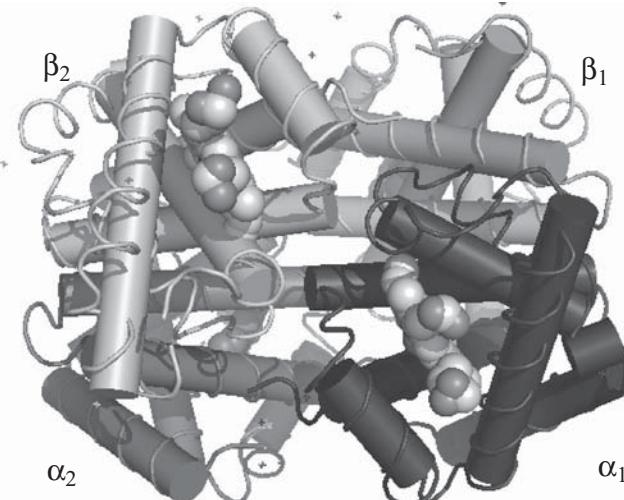
This section outlines the chapter devoted to iron-containing oxygen carriers and their synthetic models in the first edition of this text.<sup>1</sup> Reversible coordination of dioxygen ( $O_2$ ), along with its transport through the bloodstream of vertebrates (and many invertebrates) by the iron-containing metalloproteins



**Figure 7.1** Protoporphyrin IX, heme b, as found in Mb and Hb.

myoglobin (Mb) and hemoglobin (Hb), is critical to the maintenance of biological function. While dioxygen's solubility in water is quite low ( $6.6\text{ cm}^3$  per liter or  $3 \times 10^{-4}\text{ M}$ ), myoglobin and hemoglobin increase O<sub>2</sub>'s solubility in blood approximately 30 times—that is, to  $200\text{ cm}^3$  per liter or  $9 \times 10^{-3}\text{ M}$ .<sup>2</sup> Myoglobin and hemoglobin complex iron through use of a prosthetic heme group, a planar four-coordinate porphyrin ligands such as that shown in Figure 7.1. This heme group is classified as a heme b, the same porphyrin ligand system that ligates iron in cytochrome P450 (Section 7.4), cytochrome b(6)f (Section 7.5), and cytochrome bc<sub>1</sub> (Section 7.6). Protein side-chain ε-nitrogen atoms of histidine complete iron's ligand coordination sphere in myoglobin and hemoglobin.

The structure of hemoglobin and myoglobin were among the first to be solved by X-ray crystallography: for deoxymyoglobin by J. C. Kendrew and H. C. Watson beginning in 1966,<sup>3</sup> and for deoxyhemoglobin by G. Fermi beginning in 1975.<sup>4</sup> These first deoxyhemoglobin structures have been declared obsolete in the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB, <http://www.rcsb.org/pdb/>). In 1984 Fermi and co-workers published the deoxyhemoglobin quaternary structure visualized in Figure 7.2.<sup>5</sup> In Figure 7.2 the four hemoglobin protein chains are visualized in cartoon, cylindrical helical format, and the four hemes are shown in sphere format using the software program Pymol.<sup>TM</sup> The Hb α<sub>1</sub> chain (right, front, lower) is visualized in blue with its heme carbon atoms in gray, oxygens in red, and nitrogens in blue. The Hb α<sub>2</sub> chain (left, rear, lower) is visualized in red with its heme carbon atoms in red. The Hb β<sub>1</sub> chain (right, rear, upper) is visualized in green with its heme carbon atoms in red. The Hb β<sub>2</sub> chain (left, front, upper) is visualized in cyan with its heme carbon atoms in gray. The X-ray crystallographic data, deposited in the Protein Data Bank as PDB: 2HHB, 3HHB and 4HHB, have been updated by many later structures of deoxy- and oxyhemoglobins, many having been modified by site-directed mutagenesis of important aa residues that affect the position of the iron heme



**Figure 7.2** Quaternary structure of deoxyhemoglobin tetramer (PDB: 4HHB). Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft Corporation.) (See color plate)

as well as dioxygen binding and release. (Section 2.3.4 discusses the practice and utility of site-directed mutagenesis.) In late 2006, 381 different data depositions of X-ray and NMR solution studies for deoxy-, oxy-, carbonmonoxy-, nitroso-, and cyano- variants of hemoglobin and its mutant congeners were available in the PDB.

The Figure 7.2 PDB: 4HHB structure of human deoxyhemoglobin was refined at 1.74-Å resolution using data collected from a synchrotron X-ray source. The crystallographic R-factor is 16.0%, or 0.16. The estimated error in atomic positions is 0.1 Å overall, 0.14 Å for main-chain atoms of internal segments, and 0.05 Å for the iron atoms. The geometry of the iron–nitrogen complex closely resembles that of the deoxymyoglobin. In deoxyhemoglobin, the distances of the iron from the mean plane of N(porphyrin) are 0.40(5) Å and 0.36(5) Å, respectively, at the alpha and beta heme centers, in contrast to the corresponding distance of +0.12(8) Å and -0.11(8) Å in oxyhemoglobin (see Tables 7.1 and 7.2 and reference 6). The Fe–N<sub>Im</sub> (proximal histidine N<sub>ε2</sub>) bond length is 2.12(4) Å and the Fe–N<sub>p</sub> (porphyrin nitrogen) bond length is 2.06(2) Å, in good agreement with data presented in Table 7.1 for deoxyhemoglobin.<sup>5</sup>

### 7.2.1 Myoglobin and Hemoglobin Basics

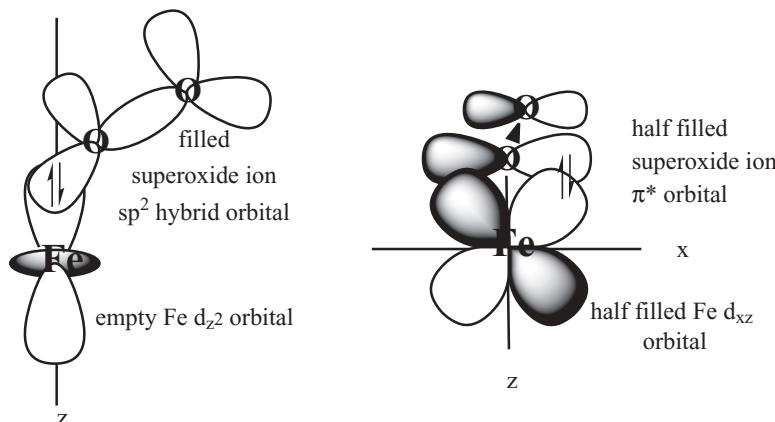
Myoglobin (Mb) is a globular monomeric protein containing a single polypeptide chain of 160 amino acid residues (MW 17.8 kDa) made up of seven

$\alpha$ -helical segments (labeled A–G) and six nonhelical segments. An example of oxymyoglobin's tertiary structure is found in the work of Phillips, who refined X-ray crystallographic data collected for oxymyoglobin at 1.6- $\text{\AA}$  resolution (PDB: 1MBO).<sup>7</sup> Myoglobin's heme prosthetic center contains an iron ion complexed by a porphyrin known as protoporphyrin IX. (See Figure 7.1.) The Fe(II) protoporphyrin cofactor is held in place in the protein principally by noncovalent hydrophobic interactions of some 80 or so residues, principally leucine, isoleucine, valine, and phenylalanine aa residues, and one covalent linkage at the proximal His F8 (his93) residue. The terminology His F8 refers to the eighth residue of the F  $\alpha$ -helical region of the protein's tertiary structure. In newer publications this histidine will usually be referred to as his93, counting aa residues sequentially beginning at the N-terminal end. The so-called distal histidine, described more fully below, is identified as His E7 or his64. Mb stores oxygen in muscle and other cellular tissue binding one oxygen molecule per protein subunit. Hemoglobin (Hb), a tetramer of four globular protein subunits, each of which is nearly identical to a Mb unit (see Figure 7.2), transports oxygen through the blood plasma. Hb's four subunits are comprised of two  $\alpha$  chains of 141 residues and two  $\beta$  chains of 146 residues with a total molecular weight of 64.5 kDa. In hemoglobin,  $\alpha$  and  $\beta$  chains differ slightly, especially in the manner in which the porphyrin is held within the protein. Bond lengths and angles reported in Tables 7.1 and 7.2 illustrate these differences.

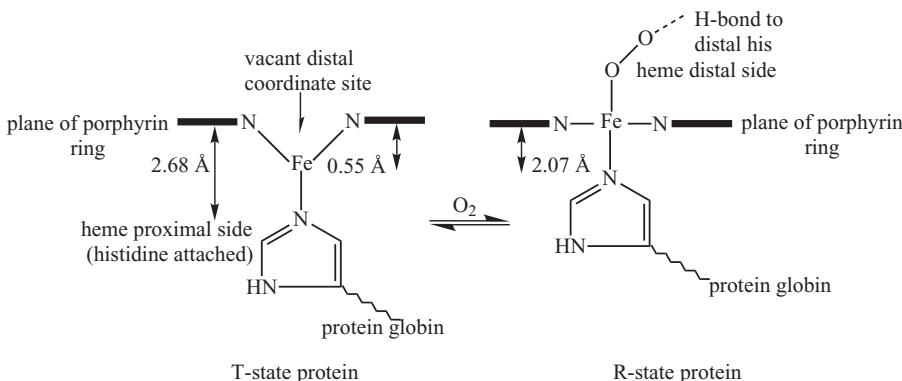
Hemoglobin binds dioxygen in a cooperative manner—that is, once one  $\text{O}_2$  molecule is bound to the enzyme, the second, third, and fourth attach themselves more readily. Both Mb and Hb bind dioxygen only when the iron ion is in its reduced state as iron(II). The terminology *oxy*- and *deoxy*-Mb and Hb refer to the enzyme in its oxygenated or deoxygenated forms, respectively, both with iron(II) metal centers, while *met*- describes oxidized heme proteins containing iron(III) centers. Comparison of reduction potentials in equations 7.1 and 7.2 indicates that dioxygen should oxidize iron(II) under most expected concentration conditions.<sup>8</sup>



Therefore the stability of biological heme– $\text{O}_2$  complexes must arise from kinetic rather than thermodynamic considerations. Some circumstances favoring heme– $\text{O}_2$  stability include: (1) placement of the heme in a hydrophobic pocket within the enzyme that is inaccessible to water molecules and protons; (2) a bent binding mode for dioxygen favored by the prosthetic group's pocket shape that prevents  $\mu$ -oxo dimer formation; (3)  $\sigma$ -bonding donation from an  $sp^2$ -rehybridized superoxide ion to an empty  $d_z^2$  Fe(II) orbital facilitated by a bent orientation of the bound dioxygen (see Figures 7.3, 7.4, and 7.5); and (4) the formation of  $\pi$ -back-bonds through interaction of a half-filled  $d_{xz}$  orbital of Fe(II) with a half-filled  $\pi^*$  orbital of the superoxide ion (see Figure 7.3).<sup>9</sup>



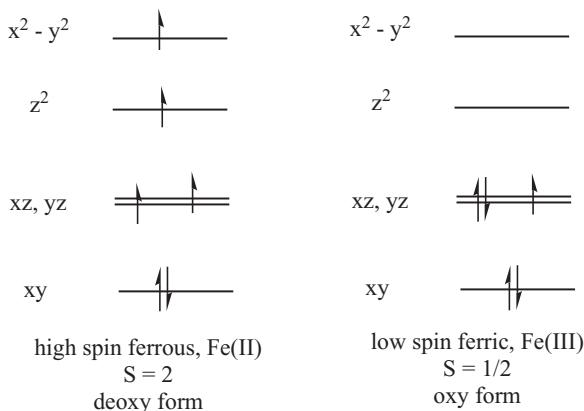
**Figure 7.3** Bonding of the  $\text{O}_2^-$  ion to Fe(III) in Mb or Hb. (Adapted with permission from Figure 3 of reference 9. Copyright 1994 American Chemical Society.)



**Figure 7.4** T and R states for iron hemes. (Adapted with permission from Figure 4 of reference 12. Copyright 1997, American Chemical Society.)

## 7.2.2 Structure of the Heme Prosthetic Group

Deoxymyoglobin and deoxyhemoglobin contain pentacoordinate iron(II) centers in which the metal ion lies out of the plane of the porphyrin's four pyrrole–nitrogen donor ligands.<sup>10</sup> Perutz has called this state the T or tense state.<sup>11</sup> The T state, a term describing the quaternary structure of the Hb tetramer, is one of low oxygen affinity in which the protein is restrained by binding of the so-called proximal histidine. In the T state the Fe(II) center is held approximately 0.55 Å outside of the porphyrin plane and none of the four Hb subunits possess dioxygen ligands. The porphyrin ring also is anchored at the active site by iron's coordination to the proximal histidine's imidazole nitrogen (see Figures 7.4 and 7.5).<sup>12</sup> In the R state or relaxed quaternary state,



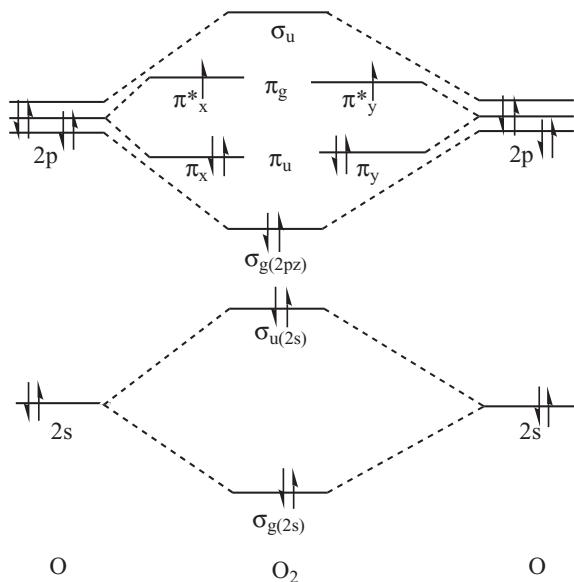
**Figure 7.5** Change from the high, Fe(II), to low, Fe(III), spin states allows iron ion to enter porphyrin plane.

dioxygen is bound to iron on the so-called distal side of the porphyrin ring. Switching from the T state to the R state in the Hb tetramer takes place during or after binding of approximately two dioxygen molecules. Upon binding of dioxygen, constraints within the surrounding protein matrix are relaxed, allowing the iron ion to move toward and nearly into the porphyrin plane.

Binding of the sixth ligand ( $O_2$ ) and consequent movement of the iron atom into the porphyrin plane induces a tertiary structure change as the proximal histidine changes its bond angle with the iron atom. The F helix containing the proximal histidine also changes position. In hemoglobin these factors in turn change the quaternary structure of the Hb tetramer and influence the affinity of the four hemes for dioxygen. The metal ion's movement is accompanied by a change from a high ( $S = 2$ ) to a low-spin state ( $S = 1/2$ ) and from the Fe(II) to Fe(III) oxidation states. In the change from the high- to low-spin states and from Fe(II) to Fe(III), the Fe ion becomes smaller and is thus able to fit better into the porphyrin ring's cavity. (See Figures 7.4 and 7.5.) The dioxygen molecule is guided by the protein pocket surrounding it to attach in a bent structure with an  $Fe-O-O$  bond angle of  $115^\circ$ . As Fe(II) is oxidized to Fe(III), dioxygen becomes the superoxide ion ( $O_2^-$ ). Evidence for this behavior is found in the  $Fe-O$ ,  $Fe-O-O$ , and  $O-O$  distances and bond angles determined by X-ray crystallography and from  $O-O$  bond orders determined by infrared and resonance Raman spectroscopy. Additional evidence for the  $Fe(III)-O_2^-$  moiety is its experimental spin state  $S = 0$  indicating magnetic coupling of the Fe(III) ion's unpaired electron with that of the superoxide ion.

### 7.2.3 Behavior of Dioxygen Bound to Metals

As can be seen from its ground-state molecular orbital diagram in Figure 7.6, dioxygen has a paramagnetic ground state. It is the only stable homonuclear diatomic molecule with this property.



**Figure 7.6** Ground-state electron configuration for the  $\text{O}_2$  molecule.

The triplet ground state ( $0 \text{ kJ/mol}$ ) for the  $\text{O}_2$  molecule is represented by the term symbol  ${}^3\Sigma_g^+$  and has two unpaired electrons, one each in the  $\pi_x^*$  or the  $\pi_y^*$  level. The first excited state ( $92 \text{ kJ/mol}$  above the ground state) is a  ${}^1\Delta_g$  singlet (electrons spin paired with both electrons in either the  $\pi_x^*$  or the  $\pi_y^*$  level). The singlet  ${}^1\Sigma_g^+$  state with paired spin electrons, one each in the  $\pi_x^*$  and  $\pi_y^*$  levels, is the next excited level  $155 \text{ kJ/mol}$  above the ground state. Reduction of  $\text{O}_2$  by one electron yields the superoxide ion ( $\text{O}_2^-$ ) a radical anion. Reduction by two electrons yields the peroxide ion, ( $\text{O}_2^{2-}$ ). As noted in equation 7.1, the reduction potential for  $\text{O}_2$  in the presence of protons is thermodynamically favorable. Therefore reversible binding of  $\text{O}_2$  to a metal can only be achieved if competition with protons and further reduction to superoxide and peroxide are both controlled.<sup>9</sup>

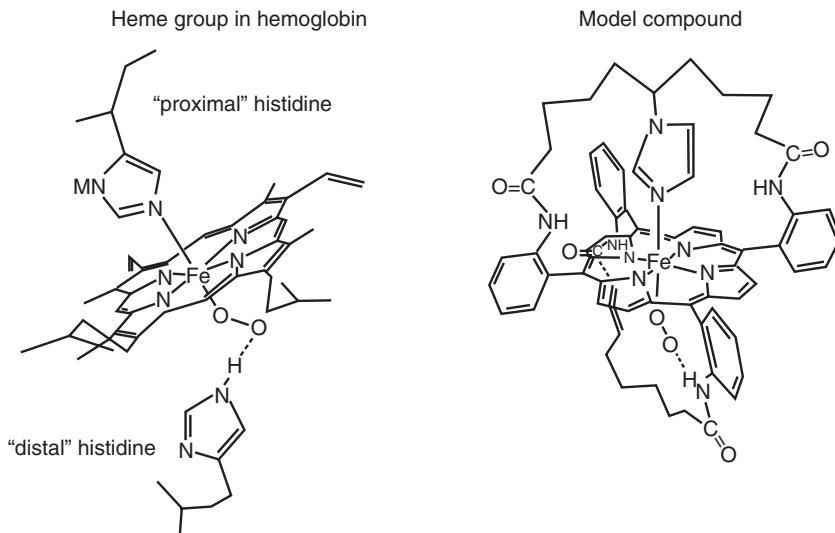
#### 7.2.4 Structure of the Active Site in Myoglobin and Hemoglobin: Comparison to Model Compounds

X-ray crystallographic structures of myoglobin and hemoglobin were first completed in 1966<sup>3</sup> and 1975,<sup>4</sup> respectively. Subsequently, many other X-ray crystallographic studies of deoxy- and oxy- as well as met-myoglobin and hemoglobin have been carried out. Additionally, researchers have studied the carbon monoxide-bound moieties  $\text{MbCO}$  and  $\text{HbCO}$  as well as  $\text{MbNO}$ . Site-directed mutagenesis of residues near the active sites of Mb and Hb have yielded information on the exact nature of  $\text{O}_2$ ,  $\text{CO}$ , and  $\text{NO}$  binding and the small molecule's orientation at the heme site. With information from these

studies confirmed by many other instrumental and analytical techniques, a clear picture of the metalloprotein's active site has emerged.

The active site of hemoglobin and myoglobin consists of an iron(II) protoporphyrin IX (the heme shown in Figure 7.1) encapsulated in a water-resistant pocket and bound to the protein through a single coordinate bond between the imidazole nitrogen of the proximal histidine residue (His93, F8 for myoglobin) and the iron(II) (see Figure 7.4). The proximal Fe–N bond vector has an approximate 10° tilt off the heme normal. Additionally, other protein residues such as the leucine, isoleucine, valine, and phenylalanine interact with the heme, holding it in place through hydrophobic interactions. The five-coordinate Fe(II) can add dioxygen in its sixth, vacant coordination site, and a variety of other small ligands (CO, NO, RCN) may also bind there. Other amino acid residues that control the immediate environment with respect to polar, hydrophobic, or steric interactions surround the distal, vacant coordination site of the deoxy form. When O<sub>2</sub> is bound, it is stabilized by hydrogen-bond interactions through the distal histidine (his64, E7). See Figure 7.4. The H-bond interactions may affect O<sub>2</sub> affinity and inhibit pathways leading to further oxidation and μ-oxo dimer formation (see Section 7.2.6 and Figure 7.8).<sup>9</sup> The distal histidine hydrogen bonding structure for hemoprotein and for a model “amide basket handle” heme is illustrated in Figure 7.7.

Despite the many model compounds that have been prepared, picket fence porphyrin and its many cogeners, first reported by the Collman research group in 1974,<sup>13</sup> remains the only porphyrin-type ligand system yielding



**Figure 7.7** Distal histidine hydrogen bonding structure for hemoglobin (left) and a heme model compound (right). (Reprinted with permission from Figure 12 of reference 9. Copyright 1994, American Chemical Society.)

crystallographic data comparable to Mb oxyheme stereochemistry (see Table 7.1). In fact, the oxygenated picket fence porphyrin model's X-ray structure was known before that of the oxyhemoproteins and anticipated the correct bent geometry of the dioxygen molecule. As described above, the bent geometry facilitates metal  $d\pi$  to  $O_2(\pi^*)$  bonding (see Figure 7.3). Model compounds also include substitutions for the proximal histidine ligand. Figure 7.10 shows substitution by a 1- or 2-methylimidazole (1-MeIm or 2-MeIm). The proximal 2-MeIm ligand of the model compounds is sterically more demanding, causing lengthening of the Fe–O bond, and models the T state (lower dioxygen affinity), whereas the proximal 1-MeIm ligand models the R state (see Figure 7.10).

The information of Tables 7.1 and 7.2 indicates that a water molecule is found in the binding cavity of  $\alpha$  chains of hemoglobin A (HbA) even though this cavity has been called hydrophobic. Indeed, although many hydrophobic groups such as valine, leucine, isoleucine, and phenylalanine are positioned over the porphyrin, the immediate vicinity of the binding site is in fact polar, containing distal histidine, the heme itself, and associated water molecules. Model 1 (deoxy) in Table 7.1 should be a good match for deoxy Hb and be considered as a T-state model compound (see Figures 7.4, left, and 7.10A). Model 2 (oxy) in Table 7.2 should be a good match with  $HbO_2$  and be considered an R-state model compound (see Figures 7.4, right, and 7.10B). The orientation of the proximal, axial base (i.e., the angle  $\phi$  in Table 7.1), the distance of Fe out of the porphyrin plane for deoxy HbA, as well as Mb, and the model compound  $Fe(T_{piv})PP-(2\text{-MeIm})$ , model 1, are similar. In Table 7.2, model 2,

**TABLE 7.1 Deoxy-Heme Stereochemistry for Mb, Hb, and Model Compounds**

|                                     | Mb <sup>14</sup> | HbA ( $\alpha$ H <sub>2</sub> O) <sup>a</sup> ( $\beta$ ) <sup>15</sup> | Fe Model 1 <sup>b</sup> (Deoxy) <sup>6</sup> |
|-------------------------------------|------------------|---|--|
| Fe–N <sub>p</sub> , Å <sup>c</sup>  | 2.03 (10)        | 2.08 (3)<br>2.05 (3)  | 2.072 (5)                                    |
| Fe–N <sub>Im</sub> , Å <sup>c</sup> | 2.22             | 2.16 (6)<br>2.09 (6)  | 2.095 (6)                                    |
| Fe–Porph, Å                         | 0.42             | 0.40 (5)  | 0.43   |
| $\phi$ , deg <sup>d</sup>           | 19               | 18 (1)  | 22.8   |
| Tilt, deg <sup>d</sup>              | 11               | 12 (2)  | 9.6  |

<sup>a</sup>  $\alpha$  and  $\beta$  refer to the particular protein chains in myoglobin or hemoglobin. One notes from the table that the bond distances and angles are slightly different for the two moieties. In each table cell, data are reported in order  $\alpha$  and then  $\beta$ .

<sup>b</sup> Fe model 1 (deoxy form)  $Fe(T_{piv})PP-(2\text{-MeIm})$ . The porphyrin ligand,  $Fe(T_{piv})PP$ , is the picket fence porphyrin, meso-tetrakis( $\alpha, \alpha, \alpha, \alpha$ -o-pivalamidephenyl)porphyrin (Figure 16 of reference 9). In this chapter see Figure 7.9B.

<sup>c</sup> Fe–N<sub>p</sub> is the bond distance to porphyrin nitrogen ligand atoms. Fe–N<sub>Im</sub> is the bond distance to histidine (or imidazole in model compounds) nitrogen ligand atoms.

<sup>d</sup>  $\phi$  is defined as the angle between the plane of the axial base (imidazole or histidine) and the plane defined by the metal, the N ligands of the porphyrin ring and the axial base. Tilt is defined as the angle the axial base moves away from the normal to the metal–porphyrin plane and is affected by methyl substituents such as those on  $Fe(T_{piv})PP-(2\text{-MeIm})$ .

**TABLE 7.2** Oxy-Heme Stereochemistry for Mb, Hb, and Model Compounds

|                                     | MbO <sub>2</sub> <sup>7</sup> | HbO <sub>2</sub> ( $\alpha$ H <sub>2</sub> O) <sup>a</sup><br>( $\beta$ ) <sup>16</sup> | Fe Model 2 <sup>b</sup><br>(Oxy) <sup>17</sup> | Fe Model 3 <sup>c</sup><br>(Oxy) <sup>18</sup> |
|-------------------------------------|-------------------------------|---|--|--|
| Fe–N <sub>p</sub> , Å <sup>d</sup>  | 1.95 (6)                      | 1.99 (5)<br>1.96 (6)  | 1.98 (1)                                       | 1.996 (4)                                      |
| Fe–N <sub>im</sub> , Å <sup>d</sup> | 2.07 (6)                      | 1.94 (9)<br>2.07 (9)  | 2.07 (2)                                       | 2.107 (4)                                      |
| Fe–O, Å                             | 1.83 (6)                      | 1.66 (8)<br>1.87 (3)  | 1.75 (2)                                       | 1.898 (7)                                      |
| O–O, Å                              | 1.22 (6)                      |   | 1.24   | 1.22 (2)                                       |
| Fe–O–O, deg                         | 115 (5)                       | 153 (7)<br>159 (12)   | 131  | 129 (2)  |
| Fe–Porph, Å                         | 0.18 (3)                      | 0.12 (8)<br>−0.11 (8)   | −0.03  | 0.09   |
| ϕ, deg <sup>e</sup>                 | 1                             | 11<br>27  | 20   | 22   |
| Tilt, deg <sup>e</sup>              | 4                             | 3<br>5  | 0  | 7  |

<sup>a</sup>α and β refer to the particular protein chains in myoglobin or hemoglobin. One notes from the table that the bond distances and angles are slightly different for the two moieties. In each table cell, data are reported in order α and then β.

<sup>b</sup>Fe model 2 (oxy form) Fe(T<sub>piv</sub>)PP–(1-MeIm)O<sub>2</sub>. The porphyrin ligand, Fe(T<sub>piv</sub>)PP, is the picket fence porphyrin, meso-tetrakis( $\alpha$ ,  $\alpha$ ,  $\alpha$ ,  $\alpha$ -o-pivalamidephenyl)porphyrin (see Figure 16 of reference 9). In this chapter see Figure 7.9B.

<sup>c</sup>Fe model 3 (oxy form) Fe(T<sub>piv</sub>)PP–(2-MeIm)O<sub>2</sub>. The ligand is (T<sub>piv</sub>)PP–(2-MeIm).

<sup>d</sup>Fe–N<sub>p</sub> is the bond distance to porphyrin nitrogen ligand atoms. Fe–N<sub>im</sub> is the bond distance to histidine (or imidazole in model compounds) nitrogen ligand atoms.

<sup>e</sup>ϕ is defined as the angle between the plane of the axial base (imidazole or histidine) and the plane defined by the metal, the N ligands of the porphyrin ring, and the axial base. Tilt is defined as the angle the axial base moves away from the normal to the metal–porphyrin plane and is affected by methyl substituents such as those on Fe(T<sub>piv</sub>)PP–(2-MeIm).

Fe(T<sub>piv</sub>)PP–(1-MeIm)O<sub>2</sub>, works better as a model for HbO<sub>2</sub> and MbO<sub>2</sub> than does model 3, Fe(T<sub>piv</sub>)PP–(2-MeIm)O<sub>2</sub>, although the ϕ and tilt values are not good matches with MbO<sub>2</sub> in either model.

### 7.2.5 Some Notes on Model Compounds

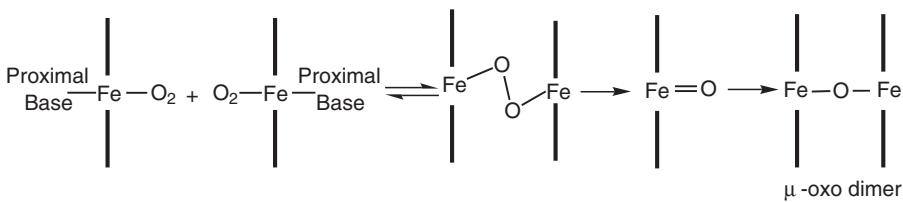
Kenneth Suslick<sup>19</sup> has written several pertinent statements about the study of metalloproteins and their model compounds. This paradigm will guide the study of the varied bioinorganic systems described in this chapter. Additions to Suslick's statement in point 2 stress the importance of modern structural analysis (X-ray crystallography, EXAFS, EPR, and two- and three-dimensional high-field NMR) of the purified metalloproteins and their model compounds. As paraphrased from reference 19, the following statements outline the steps involved in the study of metalloproteins:

1. *Isolation and purification* of the metalloprotein. For instance, Hb was first crystallized in 1849, its physiological purpose of oxygen transport was recognized by 1864, and its molecular weight and primary amino acid sequence was known by 1930.
2. *Measurement* of physical and spectroscopic properties of the active site. X-ray crystallography has been extensively used to characterize metalloproteins, especially with regard to their metal cofactors. Interpretation of the X-ray data must be consistent with spectroscopic data from EXAFS, EPR, and NMR studies. Currently, X-ray crystallographic data are deposited with the Protein Data Bank (PDB) for met-, deoxy-, and oxyhemoglobin and myoglobin from a number of species. More recently, X-ray studies of myoglobin and hemoglobin modified by site-directed mutagenesis of aa residues near and at the active site have led to more detailed information on structure–function relationships.
3. *Characterization* of structural, spectroscopic, and reactivity properties of model compounds—that is, metal cofactor small-molecule analogs.
4. *Comparisons* between the protein and the analogs to reveal new structure–function relationships.

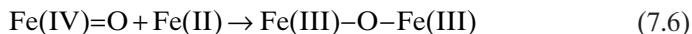
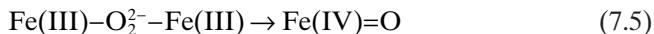
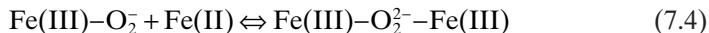
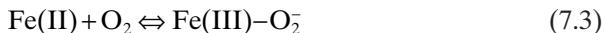
Chemists find many advantages in studying small molecule analogs of metalloproteins. For hemoglobin, as for other systems, one can systematically change one variable at a time. These variables might include modifications of axial ligation, binding site polarity, steric restraint, and solvent effects, among others. The disadvantage for all small-molecule analogs is that the protective environment of the protein itself is lost. In the heme example under discussion, one must synthesize a complex porphyrin ligand system, keep iron in a five-coordinate state until dioxygen is added, prevent irreversible oxidation of the iron center, and produce an environment in which O<sub>2</sub> ligation is favored over CO binding.

### 7.2.6 Iron-Containing Model Compounds

In the case of iron-containing small molecule analogs of Mb and Hb, a rocky road to successful model compounds was encountered. Even though the syntheses of iron porphyrin complexes were carried out, their irreversible oxidation to the  $\mu$ -oxo dimer upon addition of O<sub>2</sub> remained a stumbling block to their study as small molecule analogs of Mb and Hb. Addition of dioxygen to simple, undecorated iron porphyrins led to irreversible oxidation and into a thermodynamic Fe(III) pit with formation of the  $\mu$ -oxo dimer. This behavior is illustrated in the following reactions adapted from reference 20 and the scheme adapted from Suslick's article as shown in Figure 7.8.<sup>19</sup> Initially attempts were made to solve this problem for solution studies through the use of low temperatures and aprotic solvents although the experimental conditions were far from those of physiological systems.



**Figure 7.8** Formation of  $\mu$ -oxo dimer. (Adapted with permission from the reference 19. Copyright 1985, Division of Chemical Education, Inc.)

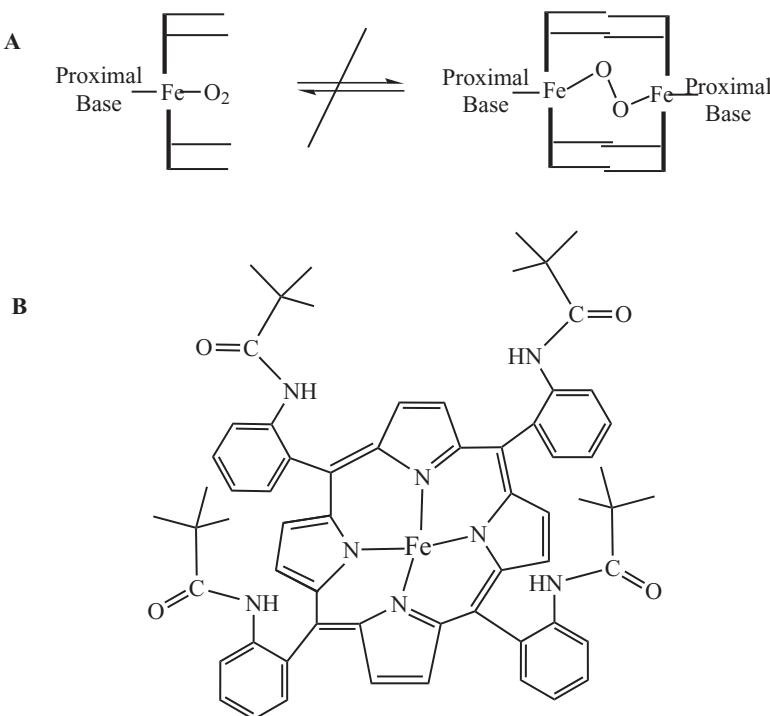


$\mu$ -oxo dimer

Successful iron-containing Mb and Hb models have simulated the protein surroundings of the heme porphyrin through introduction of steric hindrance on the distal side of the porphyrin (Figures 7.9 and 7.10). The porphyrin's decoration is necessary to prevent irreversible oxidation of the Fe(II) by dioxygen and the formation of the  $\mu$ -oxo dimer shown in Figure 7.8. Other strategies used to make Hb and Mb model compound irreversible oxidation slower are to lower the temperature, to attach the Fe(II) complex to a solid support, or to make the observation time faster by using stopped flow kinetics.<sup>19</sup>

In order to prepare and isolate solid-state, crystalline, oxygenated iron–heme model complexes, chemists learned to synthesize (by self-assembly methods) and oxygenate many types of hindered porphyrins. For instance, “capped” porphyrins were synthesized by direct condensation of a suitable tetraaldehyde with four pyrrole molecules.<sup>22</sup> “Picket-fence” porphyrins such as  $[\text{Fe}(\text{TPP})(N\text{-MeIm})]$  (where TPP = *meso*-tetraphenylporphyrin and N-MeIm = *N*-methylimidazole) were synthesized and purified by separation of isomeric mixtures of tetra(*O*-substituted phenyl)porphin.<sup>23</sup> Another picket-fence porphyrin,  $\text{Fe}(\text{T}_{\text{piv}})\text{PP}$ , *meso*-*tetrakis*( $\alpha$ ,  $\alpha$ ,  $\alpha$ ,  $\alpha$ -*o*-pivalamidephenyl) porphyrin, synthesized and studied by Collman's group in the 1970s, is shown in Figure 7.9B.<sup>21</sup> Structural and geometric data on deoxy and oxy forms of this model porphyrin are listed in Tables 7.1 and 7.2. Many other representative examples of hemoglobin model compounds, some of which became known as the “basket” and “pocket” porphyrins, are shown in Figures 8 and 9 of reference 9.

All successful myoglobin and hemoglobin model compounds provide steric bulk on the distal side of the porphyrin ring with a hydrophobic pocket for complexation of dioxygen as well as a bulky alkyl imidazole proximal ligand

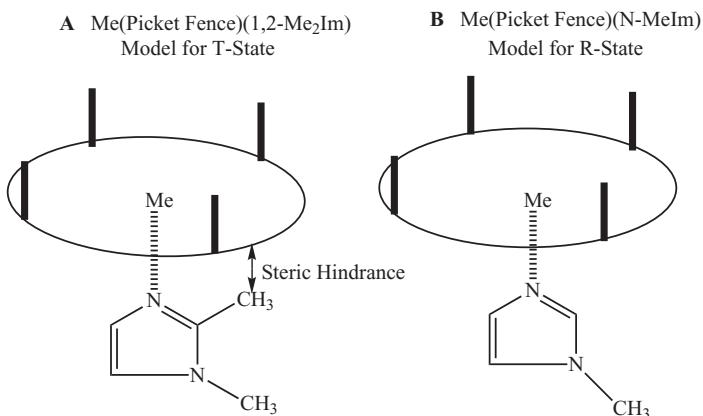


**Figure 7.9** (A) Prevention of  $\mu$ -oxo dimer formation by distal-side porphyrin modifications. (Adapted with permission from the reference 19. Copyright 1985, Division of Chemical Education, Inc.) (B) Picket-fence porphyrin  $\text{Fe}(\text{T}_{\text{piv}})\text{PP}$ , meso-tetrakis( $\alpha$ ,  $\alpha$ ,  $\alpha$ ,  $\alpha$ -*p*-pivalamidophenyl)porphyrin described in references 6, 17, 18, and 21.

(not shown in Figure 7.9B) occupying the other, proximal side of the porphyrin ring. Design of such systems have been many and varied and have led, for example, to models for the T and R states as shown in Figure 7.10.<sup>12</sup> The T-state model (deoxy form in Figure 7.4) uses steric hindrance to prevent the iron ion from entering the plane of the porphyrin ring, while the R-state model (oxy form in Figure 7.4) permits the iron ion into the porphyrin plane. Geometric data for hemoglobin, myoglobin, and a picket-fence model compound, shown in Figure 7.9B, is given in Tables 7.1 and 7.2.

In summary, researchers found a number of methods for avoiding  $\mu$ -oxo dimer formation and preserving a five-coordinate Fe(II) in iron-containing model compounds, through:

1. Modifying the imidazole to preserve the T-state (shown in Figure 7.10A).<sup>22</sup>
2. Modifying the porphyrin to sterically prevent addition of a large sixth ligand. The most well-known version of these is the “picket-fence”



**Figure 7.10** Picket-fence porphyrin models. (A) Model for T-state and (B) model for R-state. (Adapted with permission from Figure 5 of reference 12. Copyright 1997, American Chemical Society.)

porphyrin illustrated in Figure 7.9B although many others have been synthesized.<sup>9,19</sup>

3. Attaching the five-coordinate system to a rigid support—that is, silica gel, reducing its mobility and ability to add a sixth ligand.

### 7.2.7 Binding of CO to Myoglobin, Hemoglobin, and Model Compounds

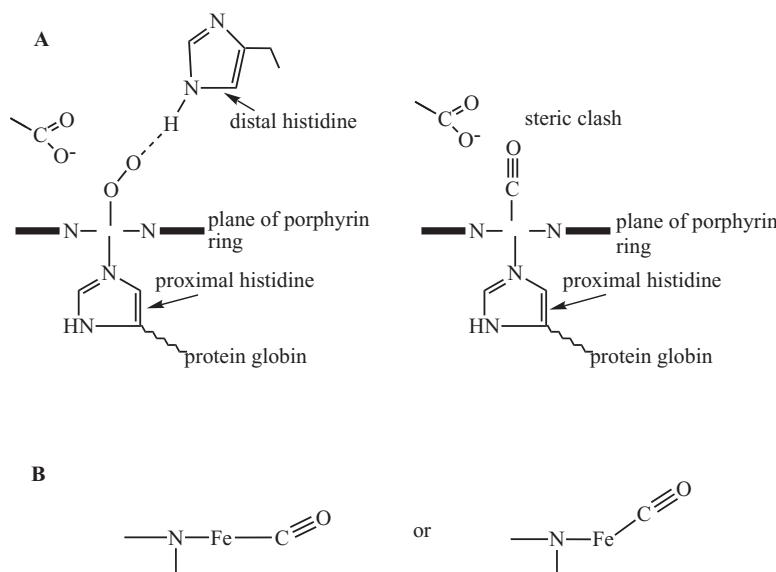
Selective binding of dioxygen rather than carbon monoxide in wild-type biological systems is complicated by the fact that naturally occurring metalloproteins including Mb and Hb produce CO during their degradation processes.<sup>12</sup> Therefore, hemes must be able to carry out their oxygen transport and storage functions in the presence of significant concentrations of CO. In addition, CO binds to myoglobin and hemoglobin with affinities of, respectively, 25 and 200 times those of O<sub>2</sub>. Collman proposes that the discrimination occurs through steric constraints imposed by amino acid residues on the distal side of the porphyrin and by selective hydrogen bonding favoring O<sub>2</sub> over CO coordination. Dioxygen is capable of a bent geometry when bound to the heme facilitating the hydrogen bond to the distal histidine, whereas CO's preferred linear binding mode not only prevents hydrogen bond formation but also results in steric clashes with neighboring amino acids as shown in Figure 7.11.

Not only is carbon monoxide produced during catabolism of hemes, but it is also necessary for maintenance of neuronal function and cell signaling in the vasculature and therefore is present normally in the vicinity of Mb and Hb. For synthetic systems of picket-fence, pocket, or capped porphyrins, CO affinity exceeds that for O<sub>2</sub> sometimes by many orders of magnitude. The in

vitro preference of CO over O<sub>2</sub> binding to Mb and Hb are in direct conflict with the data indicating that under physiological conditions, the dioxygen affinity of Mb and Hb are orders of magnitude greater than that for CO.<sup>20</sup>

For small-molecule, metal–carbon monoxide complexes, the carbon monoxide ligand is almost always in a linear conformation and perpendicular to the metal. If one assumed bonding of CO to Hb or Mb in its normal linear, perpendicular mode, steric conflicts as illustrated in Figure 7.11A would occur and thus one might predict that CO binding would be less favored by physiological systems. Earlier X-ray crystallography of MbCO and Hb(CO)<sub>4</sub> had been less than clear about the carbon monoxide ligand because resolution in protein crystallographic structures was not sufficient to unequivocally distinguish between linear and bent structures. Several possible interpretations of CO binding in a bent conformation may have included, for instance, either of the binding modes illustrated in Figure 7.11B.<sup>20</sup> However, the simple model of linear CO perpendicular to the heme metal ion causing steric crowding versus the bent O<sub>2</sub> molecule relieving steric crowding has proved to be remarkably successful, as J. P. Collman and co-workers reported in a 2004 *Chemical Reviews* article.<sup>24</sup> This reference is the place to look for updates to all the material in Section 7.2.

In attempting to understand how the attachment and release of carbon monoxide, and ultimately dioxygen, happens on a molecular scale, Rodgers



**Figure 7.11** (A) Coordination of O<sub>2</sub> and CO to heme iron in Mb and Hb. (B) Possible Fe-CO bent conformation binding modes.

and Spiro studied the nanosecond dynamics of the R to T transition in hemoglobin.<sup>25</sup> Using pulse-probe Raman spectroscopy, with probe excitation at 230 nanometers, these workers were able to model the R–T interconversion of the hemoglobin molecule as it moved from the R state (HbCO) to the T state (Hb). Under static conditions, laser excitation at 230 nm provided resonance enhancement of vibrational Raman bands of tyrosine ( $\text{tyr}\alpha42$ ) and tryptophan ( $\text{trp}\beta37$ ) side-chain residues that form specific hydrogen bonds across the  $\alpha_1\beta_2$  interface in the T state. These H bonds are broken in the R state. When the researchers collected data on Hb and HbCO under transient conditions, they found a different set of behaviors at much shorter times that reached a maximum intensity at ~50 ns. They believe that these differences reflect tertiary structure changes induced by loss (deligation) of the CO ligand before the R–T interconversion. It is known that CO molecules deligate in less than a picosecond, may recombine with the iron center in geminate fashion before leaving the binding pocket on a ~50-ns time scale, or leave the binding pocket on the same time scale. Therefore the workers believe that the 50-ns differences they detect accompany carbon monoxide ligands leaving the binding pocket rather than CO deligation. In concert with CO leaving the binding pocket, H bonding between residues on the A helix and E helix (the E helix being intimately connected with the heme cofactor because it contains the distal histidine, his58) is detected. Weakening of other H bonds that would allow a shifting of the E helix toward the heme group is also found. In X-ray crystallographic and molecular dynamics studies, it was found that the F helix (proximal to the heme cofactor and containing the proximal histidine ligand his87) moves as well. Taking all this evidence into account, Rodgers and Spiro proposed a model for the R–T reaction coordinate that starts with CO deligation and involves movement of and strain in the E and F helices (on a subpicosecond time scale) along with movement of the Fe ion toward the proximal his residue. The “scissoring” motion of the E and F helices, illustrated in Figure 3 of reference 25, relieves the strain between the helices and allows the CO to leave the binding pocket at the same time.

Moffat and co-workers reported a study of MbCO using nanosecond time-resolved crystallography in 1996.<sup>26</sup> Nanosecond time-resolved crystallography of MbCO is discussed in Section 3.7.2.3, along with a more complete discussion of the reference 26 work. After firing a 10-ns burst of laser light to break the CO–Fe bond, the researchers produced a diffraction image of the crystal through application of a 150-ps X-ray pulse. They showed release of the CO molecule, displacement of the Fe ion toward the proximal histidine, and recombination of the dissociated CO in a time frame of about 100  $\mu\text{s}$ . The reference 26 researchers found that their results using time-resolved crystallography compared well with other spectroscopic studies of HbCO, MbCO, and their models. See Section 3.7.2.3 for more updated examples of CO dissociation from myoglobin and hemoglobin as detected by time-resolved X-ray crystallography. These include references to movie displays illustrating CO dissociation from the heme iron.

### 7.2.8 Conclusions

Myoglobin and hemoglobin have been studied exhaustively since chemists, biochemists, and biologists realized their common and abiding interest in these biomolecules. In Section 7.2, a brief review of the major knowledge categories has been presented. No quantitative information has been presented on kinetic or thermodynamic aspects of O<sub>2</sub> binding, although much is known. Reference 9, for instance, gives an excellent review of quantitative kinetic and thermodynamic information on myoglobin, hemoglobin, and their model compounds.

While much has been learned about myoglobin and hemoglobin, many controversies and uncertainties remain. It is certain that myoglobin stores dioxygen in muscle tissue, whereas hemoglobin carries dioxygen through the bloodstream to all parts of the body. We know the primary, secondary, tertiary, and quaternary structures of the protein matrix with a great deal of certainty and we have much information about the cooperative nature of dioxygen binding to hemoglobin. Most known instrumental and analytical techniques have been used in the study of these important dioxygen carriers. X-ray crystallography, infrared and high-field NMR have yielded much structural information but have not been able to unequivocally resolve a major remaining uncertainty: Why does O<sub>2</sub> preferentially bind to hemes in spite of the known greater affinity of Mb and Hb for CO over O<sub>2</sub>? Consensus will probably arise from a blending of several explanations: hydrogen bonding and electrostatic considerations that favor O<sub>2</sub> coordination and steric considerations that disfavor CO coordination. Certainly the questions and answers will make for interesting future reading.

## 7.3 INTRODUCTION TO CYTOCHROMES

Cytochromes occur in nature in immense variety. These proteins are usually involved in electron transfer reactions and contain a variety of heme cofactors that carry the electrons being transferred and serve as catalytic sites for the redox chemistry. The metalloporphyrin heme cofactors, normally carrying an iron ion that cycles between the Fe(II) and Fe(III) oxidation states, categorize cytochromes according to the type of heme–porphyrin ligand they contain. Some cytochromes listed in the Jena Library of Biological Macromolecules <http://www imb-jena.de/IMAGE.html> are cytochromes a, b, c, f, and P450. The molecules in this database are those that have structural data listed in the Protein Data Bank (PDB) or Nucleic Acid Database (NDB). (As an example, after accessing the website, choose search, Main Search page, click on SWISS-PROT entry description and enter “cytochrome c” in the search box. In mid-2006, this search retrieved 714 cytochrome c, cytochrome c oxidase, cytochrome c peroxidase, cytochrome c<sub>1</sub>, and so on, entries that include references to the structural data available.)

Cytochrome researchers often cite R. P. Ambler's work on classifying cytochromes c (discussed here in Section 7.7).<sup>27</sup> Recently, Ivano Bertini and co-workers published a review of cytochromes c, using a bioinformatics approach to "collect all the sequences of cytochrome c domains in available genome sequences."<sup>28</sup> This work included some helpful notes on cytochrome c nomenclature. The term "cytochrome" may have been first mentioned by Keilin in 1925 to describe a group of heme proteins that could undergo oxidation-reduction reactions.<sup>29</sup> These proteins were characterized by their red color and intense absorption bands in the 510- to 615-nm range. As different cytochromes c were characterized, they were classified in two ways: (1) According to functional class ( $c_1$ ,  $c_2$ ,  $c_3$ , and so on; these criteria have not been applied consistently over time) and (2) according to subscripts assigned based on the experimental wavelength in nanometers of the so-called  $\alpha$ -band in the visible absorption spectrum of the reduced protein ( $c_{550}$ ,  $c_{552}$ , and so on). The most recent attempt on a systematic nomenclature for cytochromes c was published by the International Union of Biochemistry and Molecular Biology in 1989.<sup>30</sup> The reference 28 authors have collected this information in Table 6 of their publication. The table lists the number of hemes in each cytochrome c as well as the iron ion's axial ligands and its Ambler classification.<sup>27</sup> Cytochromes are generally six-coordinate and low-spin in both the Fe(III) and Fe(II) oxidation states, but there are many exceptions to this rule. For instance, cytochrome c' is a high-spin cytochrome c, widely distributed in bacteria with a typical cytochrome c fold—that is, a four  $\alpha$ -helix bundle enclosing a single heme. It is unusual in that it usually exists as a dimer and the heme's iron atom is pentacoordinate with a single histidine axial ligand.

In a 2004 *Chemical Reviews* article, F. A. Walker discussed the geometric and electronic structure of hemes in relation to their spectroscopic footprints and their reduction potentials.<sup>31</sup> As an example of the large number of possible variables, she summarized the factors that affect cytochromes c as (1) the type of axial ligands bound to the heme (his-met for mitochondrial and bacterial cytochromes c, his-his for cytochromes  $c_3$  and other multi-heme cytochromes), (2) the orientation of axial ligands that not only have an  $\sigma$ -donor electron pair but also have a  $\pi$ -donor orbital perpendicular to the  $\sigma$ -donor orbital (examples are his, met, cys), (3) the solvent accessibility of the heme, (4) surface and buried charged groups near the heme plus the overall charge type of the surrounding protein or opposite charges on its partner protein, (5) the dipoles of the protein backbone and side chains, (6) changes in protein conformation and amino acid residue protonation, and (7) solvent pH. In this text's discussion, we will concentrate on eukaryotic mitochondrial cytochrome c.

In mid-2006, a search of the Jena Library for cytochrome b returned 197 entries, including those described in Sections 7.5 (cytochrome  $b_6f$ ) and 7.6 (cytochrome  $bc_1$ ). Cytochrome b contains the heme b subunit (heme protoporphyrin IX), the same metal cofactor contained in hemoglobin and myoglobin (see Figure 7.1). Other cytochrome b subunits occur in the cytochrome  $b_6f$  complex, also known as plasotquinol:plastocyanin reductase (EC 1.10.22) and

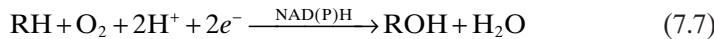
in the cytochrome bc<sub>1</sub> complex, also known as ubiquinol:ferricytochrome c oxidoreductase (EC 1.10.99.1). One can search the <http://www.expasy.org/enzyme/> site to find more information about any enzyme by entering its name. The EC terminology is established by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) at <http://www.chem.qmul.ac.uk/iubmb/nomenclature/>. Cytochrome b binds a high ( $b_H$ ;  $b_{560}$  or  $b_{562}$ ) and a low ( $b_L$ ;  $b_{566}$  or  $b_{565}$ ) potential heme group. These two cytochromes will be discussed in detail in Sections 7.5 and 7.6. Cytochrome P450, discussed in Section 7.4, also contains a cytochrome b.

Cytochrome P450 enzymes are monooxygenases that insert one atom of a dioxygen molecule into a substrate. A wide variety of organic molecules are oxygenated by cytochrome P450. The other oxygen atom of the dioxygen molecule is converted to water during the P450 catalytic cycle (Section 7.4). Cytochrome P450 has its own dedicated website maintained by D. R. Nelson at <http://drnelson.utmem.edu/CytochromeP450.html>.

## 7.4 CYTOCHROME P450: A MONOOXYGENASE

### 7.4.1 Introduction

The cytochrome P450 metalloenzyme is one of a large group of iron-containing oxygenases that utilize atmospheric dioxygen to functionalize molecules using cofactors such as flavins or NAD(P)H as well as non-heme iron or copper and other metalloporphyrin complexes. Cytochromes P450 are known to protect the body from foreign chemicals by oxidizing toxic compounds to metabolically manageable hydrophilic materials. As such, they can be thought of as physiological permanganate ions ( $MnO_4^-$ ). However, P450s can produce toxic or carcinogenic compounds in vivo. The cytochrome P450 metalloenzyme contains an iron–heme cofactor similar to that found in hemoglobin or myoglobin—that is, it is a metalloporphyrin. The porphyrin itself, called protoporphyrin IX, or heme b (see Figure 7.1), contributes four N-donor ligands to the iron ion's coordination sphere in a planar (or nearly planar) arrangement. The metalloenzyme's name resulted from early experimentation: Cytochrome means “cellular pigment” or “P” and 450 refers the wavelength of the absorption maximum found for its ferrous-CO derivative.<sup>32</sup> Cytochrome P450s are ubiquitous in biological systems, and many of them are membrane-bound. The basic reaction catalyzed is



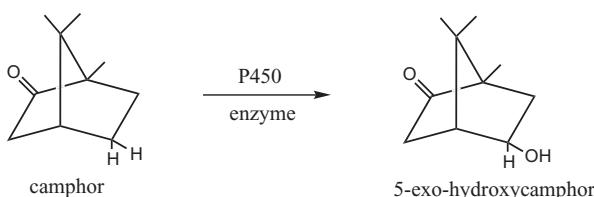
While the uncatalyzed reaction requires extremely high temperatures and proceeds nonspecifically, the enzymatic reaction proceeds regiospecifically and stereospecifically under physiological conditions of temperature and pressure. The P450 enzyme utilizes its iron–porphyrin centers as metal-containing

cofactors (prosthetic groups) that are intimately involved with O<sub>2</sub> activation. Additional coordination sites attach the iron ion in the porphyrin center to the protein through cysteine side chains, or they serve to coordinate dioxygen, NO, oxide anion, water, and other ligands during portions of the catalytic cycle.

Early experiments identified a low-spin ferric, Fe(III), center in an enzyme specific for the substrate camphor, called P450<sub>cam</sub>. The P450<sub>cam</sub> enzyme's metallo-cofactor was identified in the substrate-free resting state by its electron paramagnetic resonance (EPR) spectrum with *g*-values of 2.45, 2.26, and 1.91.<sup>33</sup> When the substrate, camphor, was added, the EPR signal changed to indicate the presence of 60% of the high-spin form with *g*-values of 7.85, 3.97, and 1.78 (experiments carried out at 12 K). Cytochrome P450<sub>cam</sub> continues to be one of the most-studied members of the P450 superfamily and is the central enzyme in a three-component oxidase that converts the 5-methylene group of camphor to the 5-exo alcohol. See Figure 7.12.

In 2005, when several publications discussing cytochrome P450 structure and function appeared in *Chemical Reviews*, there were 4000+ identified P450 genes encoding the cytochrome P450 superfamily.<sup>34</sup> The collection is annotated on a website maintained by David Nelson (<http://drnelson.utmem.edu/CytochromeP450.html>). The Nelson site also details the naming system for this superfamily; for example, cytochrome *Pseudomonas putida* P450<sub>CAM</sub> in the 101 subfamily carries the designation CYP101, while rabbit CYP2B4 is a member of the 2B subfamily. More information about the cytochrome P450 superfamily is found on the <http://metallo.scripps.edu/index.html> web site.

Cytochrome P450 carries out broad functional roles in biological species. One role is the metabolism of drugs and so-called xenobiotics (exogenous compounds not native to the organism). P450's ability to degrade xenobiotics makes it potentially useful in bioremediation. P450's second major role involves the biosynthesis of steroid hormones—for example, cholesterol and other critical signaling molecules. The latter role includes fat-soluble vitamin metabolism and conversion of polyunsaturated fatty acids to biologically active molecules. The cytochrome P450 molecule is also believed to be involved in carcinogenesis.



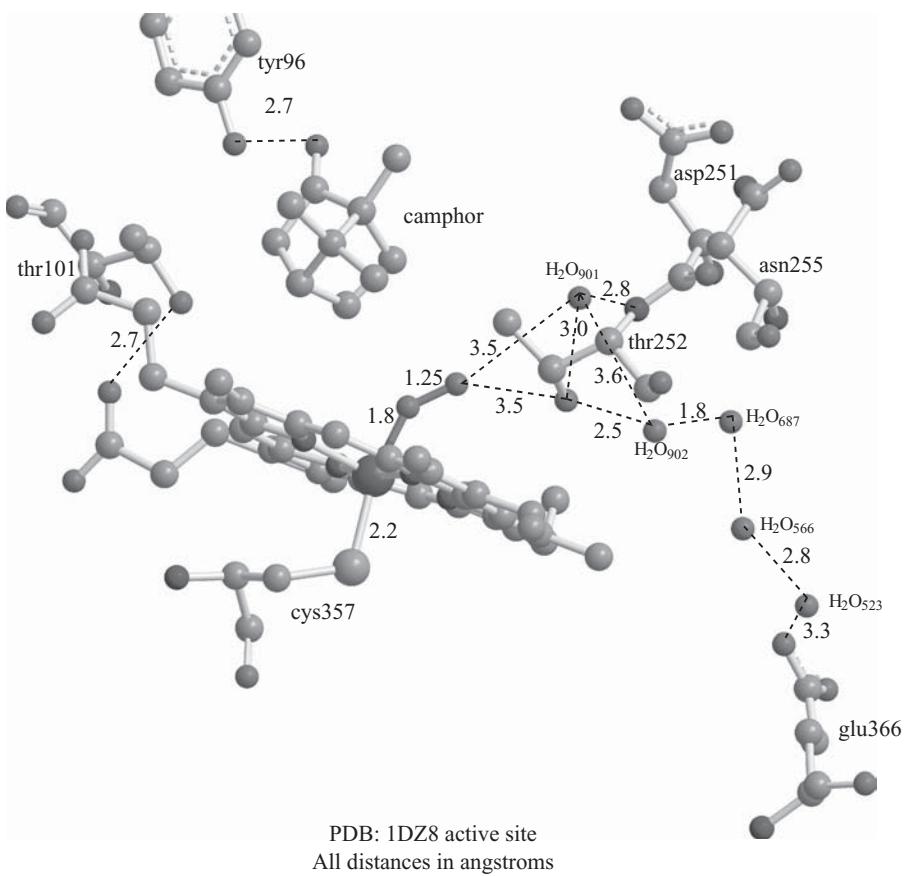
**Figure 7.12** Conversion of camphor to its 5-exo alcohol counterpart using P450 enzyme.

### 7.4.2 Cytochrome P450: Structure and Function

In this discussion, we will concentrate on the *Pseudomonas putida* P450<sub>CAM</sub> (CYP101) enzyme structure and function, especially as concerns the iron-heme center, the enzyme's mechanism of activity as a complex series of individual steps involving protein redox partners and cofactors such as NAD(P)H, and how the interactions of heme iron and dioxygen are controlled by the protein environment. The cytochrome P450 iron heme b porphyrin ligand, protoporphyrin IX, shown in Figure 7.1, is the same as that found in hemoglobin and myoglobin. The heme group is held in the catalytic pocket through a covalent bond between the iron and a cysteine amino acid side chain from the protein (cys357–S–Fe(III) in PDB: 1DZ8 = 2.2 Å) and other hydrogen-bonding interactions with amino acid side chains and specific solvent (water) molecules. There are three important N–H···S hydrogen bonds to leu358, gly359, and gln360 that help retain cysteine's sulfur atom coordination to iron as the metal ion passes through the various ferric, ferrous, and ferryl oxidation states of the catalytic cycle. (See Figures 7.13 and 7.14.)

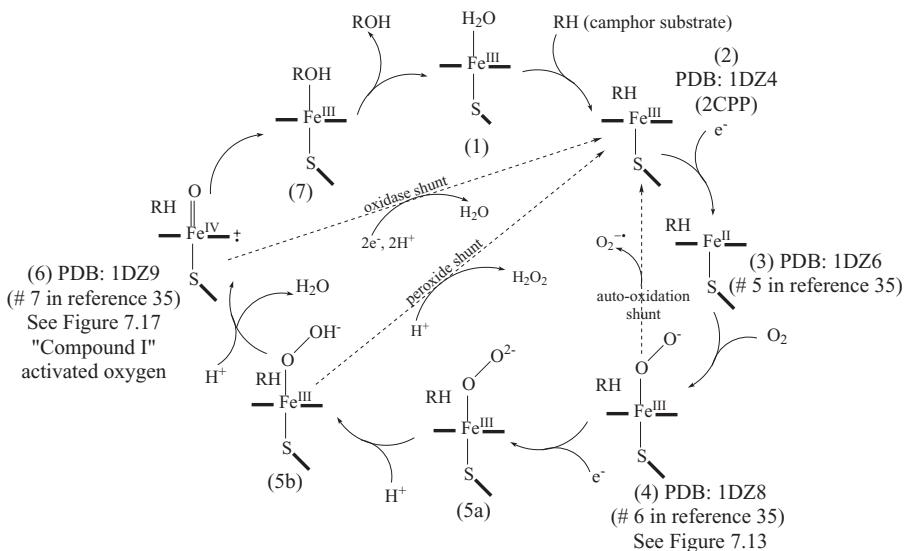
Several CYP101 X-ray crystallographic structures were reported by the Schlichting group in *Science* magazine in 2000<sup>35</sup> (PDB: 1DZ4, 1DZ6, 1DZ8, 1DZ9). References 34 and 35 form the basis for the catalytic cycle found in Figure 7.14.

The resting state of the enzyme, also called met-P450, consists of a six-coordinate, low-spin (l.s.) iron(III) center (**1**—bold numbers in parentheses refer to structures in Figure 7.14). In addition to the four planar porphyrin nitrogen ligands, the Fe(III) is axially coordinated to the sulfur of a conserved cysteine (cys357 in CYP101) and a water molecule. Upon addition of substrate, camphor in this case, the iron(III) center loses its water ligand becoming five-coordinate (**2**). Electron paramagnetic resonance (EPR) spectroscopic results identify the iron(III) center as high-spin (h.s.). X-ray crystallographic structures of this moiety include PDB: 1DZ4, 2CPP,<sup>36</sup> and 1UYU.<sup>37</sup> Next, an electron is added to produce an Fe(II) metal center (**3**). The X-ray crystallographic structure of this intermediate has been published by the Schlichting group as PDB: 1DZ6.<sup>35</sup> This step is followed by the addition of molecular dioxygen (**4**). Questions remain as to the oxidation state of iron in this intermediate, and evidence exists for both the ferrous–O<sub>2</sub> state (Fe<sup>2+</sup>–O–O) and the ferric–super-oxide state (Fe<sup>3+</sup>–O<sub>2</sub><sup>–</sup>). The PDB: 1DZ8 X-ray crystallographic structure represents this intermediate—the active site is modeled in Figure 7.13. This structure is the last relatively stable intermediate in the cycle, although it is known to decompose via the autoxidation shunt producing the physiologically dangerous superoxide radical anion. Addition of an electron to (**4**) generates the ferric–peroxy state, Fe<sup>3+</sup>–OO<sup>2–</sup> (**5a**), and then addition of a proton generates the ferric–hydroperoxy moiety, Fe<sup>3+</sup>–OOH<sup>–</sup>, (**5b**). These non-equilibrating intermediates, found in many heme-containing oxygenase and peroxidase enzymes, have been studied using cryogenic radiolysis.<sup>38</sup> The technique involves radiolysis of the ferric protein in frozen aqueous–organic solutions at 77 K to



**Figure 7.13** Cytochrome P450 catalytic site with camphor—the dioxygen intermediate (PDB: 1DZ8). Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. (Printed with permission of CambridgeSoft Corporation.)

produce the ferrous species. The ferrous product is then annealed at elevated temperatures, and the conformational and chemical relaxation processes are studied by EPR and optical spectroscopic methods. For instance,  $\text{Fe}^{3+}\text{—OOH}^-$  complexes in hemes studied by this method, including CYP101, have shown EPR  $g$  values of 2.3–2.25, 2.2–2.14, and 1.94–1.97. Obtaining such EPR values for the unstable intermediates is then taken as evidence of their presence during the catalytic cycle. Intermediate (**5b**) is known to dissociate to produce hydrogen peroxide,  $\text{H}_2\text{O}_2$ , via the peroxide shunt upon addition of a second proton, resulting in P450 inactivation. The next productive step in the catalytic cycle toward the desired mono-oxygenated substrate involves addition of the second  $\text{H}^+$  to first generate the unstable transient  $\text{Fe}\text{—OOH}_2$  followed by fast heterolytic cleavage of the O–O bond, release of  $\text{H}_2\text{O}$ , and production of the ferryl-oxo-porphyrin  $\pi$ -cation radical,  $\text{Fe(IV)} = \text{O}$  (**6**). This intermediate,



**Figure 7.14** P450CAM catalytic cycle. [Adapted with permission from Figure 2 of reference 34 (copyright 2005, American Chemical Society) and Figure 1 of reference 35 (copyright 2000, AAAS).]

believed to exist in many oxygenase and peroxidase catalytic cycles, is called **Compound I** in many research articles. The O–O heterolytic cleavage reaction to (6) is thought to be promoted by simultaneous electron transfer from the porphyrin to the Fe–O bond as the porphyrin π-cation radical is generated. Two events may happen at this point in the cycle: (1) oxygenation of the substrate or (2) addition of two more electrons and two protons via the oxidase shunt to produce another molecule of water. The latter step is nonproductive and the catalytic cycle has then simply completed the reduction of  $O_2$  to two molecules of water via the reaction:



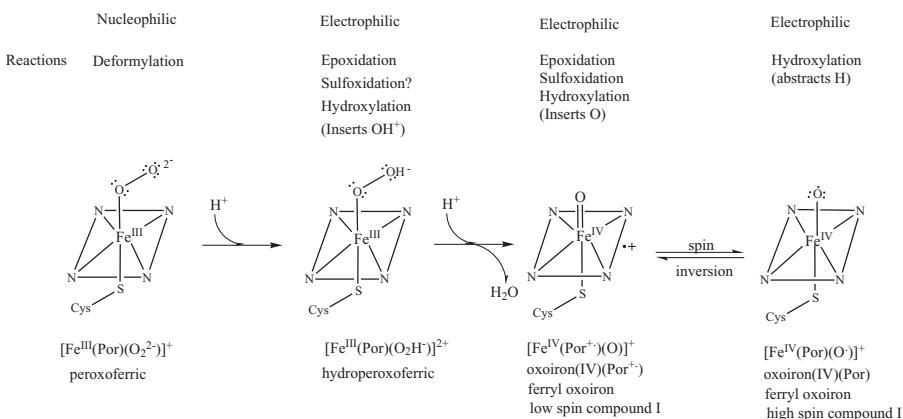
The continuation of the catalytic cycle results in insertion of the ferryl–oxo oxygen regio- and stereospecifically into the camphor C<sub>5</sub> C–H bond (7), followed by release of the 5-exo-hydroxylated camphor and re-generation of the initial member of the cycle.

#### 7.4.3 Cytochrome P450: Mechanism of Activity

As will be discussed further in Section 7.4.5, intermediates (5a), (5b), and (6) (Figure 7.14) are too reactive to build up during catalytic turnover. This has led to the assumption that intermediate (6), or compound I, is the “active oxygen” species responsible for alkane hydroxylation, alkene epoxidations,

and other P450 reactions. The so-called “compound I,” a  $\pi$ -cation radical intermediate that has an oxidation state two electrons above the  $\text{Fe}^{\text{III}}$  resting state, is thought to be analogous to that established for most peroxidases and catalases.<sup>39</sup> Several researchers have observed absorption spectra resembling that of *C. fumago* chloroperoxidase compound I for a transient species formed when substrate-free cytochrome P450 was mixed with *m*-chloroperoxybenzoic acid (*m*-CPBA).<sup>40</sup> Other workers have attempted to obtain Mössbauer and EPR spectra by freeze-quenching substrate-free cytochrome P450 in reactions with peracetic acid as the oxygen donor. These researchers observed a tyrosine radical and an  $\text{Fe}^{\text{IV}}$ –heme derivative that was concluded to be a P450 compound II moiety. The so-called “compound II” species has been identified, again in the catalytic cycles of peroxidases and catalases, as an oxoiron(IV) porphyrin species, one electron above the  $\text{Fe}^{\text{III}}$  resting state of the enzymes.<sup>56a</sup> This intermediate does not appear in Figure 7.14 but will be discussed further in Section 7.4.5. The question then arises: Do the peroxyferric,  $[\text{Fe}^{\text{III}}(\text{Por})(\text{O}_2^-)]^+$ , (5a) or hydroperoxyferric,  $[\text{Fe}^{\text{III}}(\text{Por})(\text{O}_2\text{H}^-)]^{2+}$ , (5b) intermediates function as alternate catalysts to compound I? This possibility is discussed in the literature as the “two-oxidant” or “multiple-oxidant” hypothesis, and again discussion will be deferred to Section 7.4.5. However, note that the peroxyferric (5a) intermediate with its negatively charged peroxy group might be a nucleophilic reagent, whereas all the other proposed intermediates function as electrophilic reagents. (See Figure 7.15.) In addition, Section 7.4.5 will discuss the possibility of side-on peroxide coordination as in the research of J. S. Valentine and co-workers.<sup>60–62</sup>

Most published cytochrome P450 catalytic cycles will show the end-on peroxy coordination as shown in Figure 7.14 (4, 5a, 5b). Shaik and co-workers have used density functional theory methods to propose a “two-state” theory



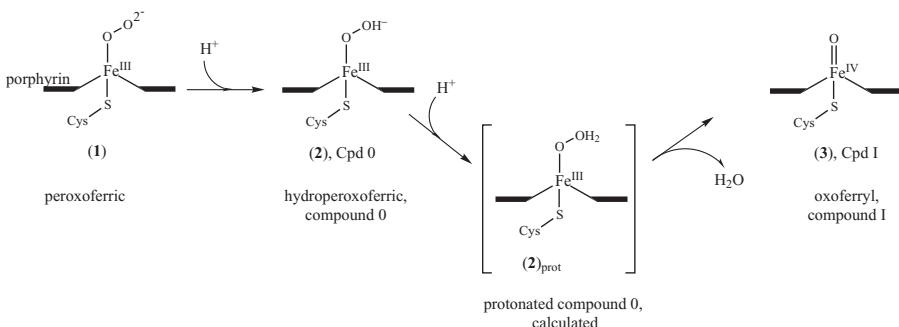
**Figure 7.15** Intermediates in the cytochrome P450 reaction cycle. (Adapted with permission from Scheme 1 of reference 39. Copyright 2004, Society of Biological Inorganic Chemistry.)

that explains most mechanistic questions by invoking low-spin and high-spin states for the compound I intermediate. One recent analysis by the Shaik group of Compound I reactivity, with styrene as substrate, shows that the reaction features multi-state reactivity (MSR) with different spin states (doublet or quartet), different electromeric situations involving radicals, cations, and Fe<sup>III</sup> and Fe<sup>IV</sup> oxidation states.<sup>41</sup> The researchers concluded that low-spin pathways lead to styrene epoxide products via a concerted mechanism that may or may not result in stereochemical scrambling. High-spin pathways have barriers that may prevent ring closure to epoxide and may include intermediates with sufficient lifetimes to allow stereochemical rearrangement and side products such as phenylacetaldehyde or 2-hydroxostyrene.

One interesting discussion in the literature involves alkene epoxidations catalyzed by hydroperoxoferric P450, intermediate (**5b**) in Figure 7.14. First, it was found that removing thr252, a highly conserved threonine on P450's distal side near the site of dioxygen binding, impaired the ability of P450<sub>CAM</sub> to hydroxylate camphor.<sup>42</sup> Next, epoxidation of alkenes was investigated in mammalian threonine–alanine mutants, where it was found that substrate alkenes were both epoxidized and hydroxylated but greater amounts of epoxidized product were formed in mutants compared to wild-type enzyme.<sup>43</sup> Further study of the T252A (thr252ala) mutant in the Dawson and Sligar laboratories showed that epoxidized but little or no hydroxylated products were formed and that the peroxide shunt from (**5b**) back to (**2**) in Figure 7.14 was operating.<sup>44</sup> All these results were interpreted to mean that Compound I was not being formed (or being formed in very small concentrations) in these mutants. Rather a second electrophilic oxidant, hydroperoxoferric P450 (**5b**) was responsible for the epoxidation products. This postulation carried the “two-oxidant” name. Dawson and co-workers were able to show that the hydroperoxoferric P450 (**5b**) has sufficient lifetime to be detected using low-temperature EPR/ENDOR methods.<sup>45</sup> Shaik and co-workers, as stated above, postulated the “two states” theory—namely that high- and low-spin states of the compound I intermediate were responsible for differences in cytochrome P450 reactivity.<sup>46</sup> Their calculations also showed that the hydroperoxoferric intermediate should be a poor oxidant with a large energy barrier to overcome.<sup>47</sup> However, their calculations did not explain why the T252A P450<sub>CAM</sub> mutant produced little or no hydroxylated camphor.

More recently, Shaik and co-workers used density functional theory as well as QM(DFT)/MM calculations to show a pathway from the peroxyferric intermediate (**5a**) via the hydroperoxoferric (**5b**) intermediate to compound I (**6**) (see Figure 7.14).<sup>48</sup> In Figure 7.16, using Shaik's terminology, the peroxyferric intermediate is labeled (**1**), the hydroperoxoferric intermediate is labeled (**2**), Cpd 0, and the ferryloxo intermediate is labeled (**3**), Cpd I.

The reference 48 authors used an UB3LYP unrestricted DFT method, Jaguar 5.5 and Gaussian 03 software, beginning calculations with an LACVP\*(Fe)6-31G\*(rest) basis set. See Section 4.4.3. The researchers assumed a direct protonation mechanism involving a framework of amino acid residues



**Figure 7.16** Pathway to cytochrome P450 Compound I. (Adapted with permission from Scheme 1 of reference 48. Copyright 2005 American Chemical Society.)

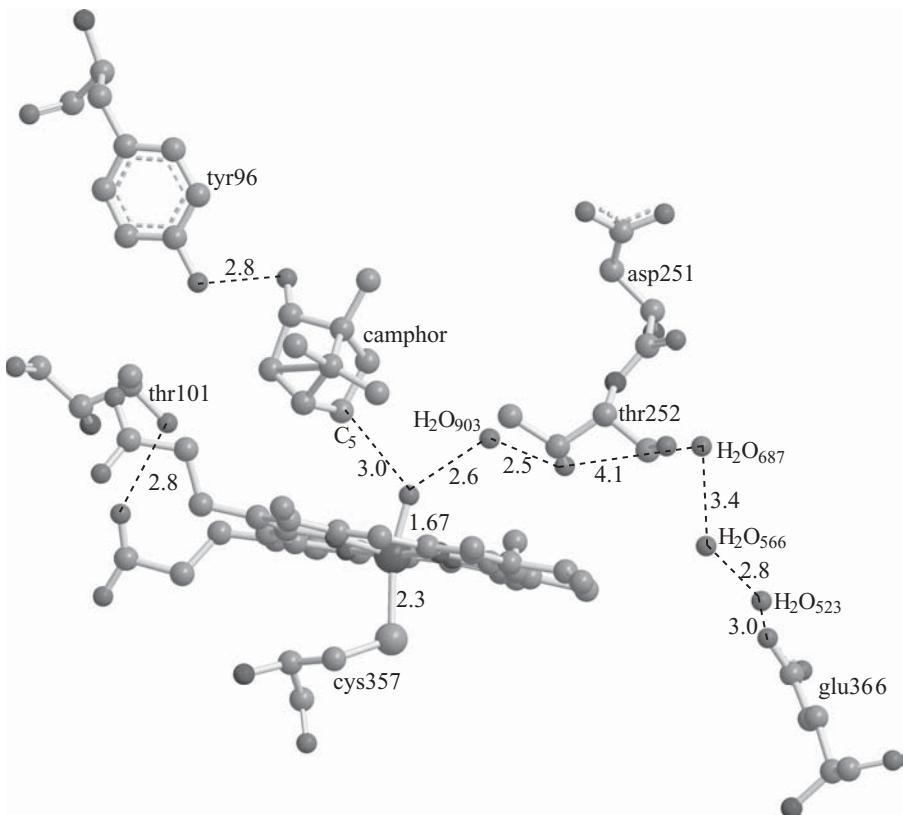
and water molecules in the distal pocket of their starting material, intermediate (**5a**) in Figure 7.14, and intermediate (**1**) in Figure 7.16. Data for the starting material intermediate were taken from the Schlichting PDB: 1DZ8 structure, assumed by those authors to be the  $Fe^{III}-O_2^-$  moiety.<sup>35</sup> Shaik and co-workers believed that the most realistic model for their calculations (96-atom Model E in reference 48) contained the iron–porphyrin ligand system with thiolate axial ligand (proximal) and dioxygen as the distal ligand. The carboxylic acid groups of asp251 and glu366 were used as proton sources (eliminating the hydroxyl group of thr252 as a proton source for any part of the process). The amino acid residues asp251 and thr252 with their connecting peptide bond were important in holding a water molecule, water<sub>901</sub> in PDB: 1DZ8, in place (see Figure 7.13). Acetic acid was used to model glu366. Five water molecules, W<sub>901</sub>, W<sub>902</sub>, W<sub>523</sub>, W<sub>566</sub>, and W<sub>687</sub>, formed a water channel connecting glu366 through thr252 and asp251 to the distal peroxy ligand, completing the model. The starting point of calculation was the structure of (**2**), Cpd 0, obtained by protonation of (**1**) via the asp251–W<sub>901</sub>–thr252 channel as suggested by experimental data<sup>49</sup> and molecular dynamics (MD) simulations.<sup>50</sup> In the Cpd 0, (**2**) model, asp251 has been rotated from its hydrogen-bonded position in the crystal structure (PDB: 1DZ8) as a starting point in investigating O–O bond scission, to get from Cpd 0, (**2**) to Cpd I, (**3**). A glu366–water–thr252 network is established for a subsequent protonation of Cpd 0. However, this does not lead directly to Cpd I but to a “surprise” species, (**2**)<sub>prot</sub>, a ferric complex of water–oxide ( $Fe-O-OH_2$ ) (see Figure 7.16). The authors find that (**2**)<sub>prot</sub> is a genuine minimum energy species in their calculations and also that Cpd 0 and Cpd I are fairly close to each other energetically. Continuing with QM/MM calculations, Shaik and co-workers confirm that Cpd 0 and Cpd I are close in energy whereas (**2**)<sub>prot</sub> is less stable by 24–33 kcal/mol. They believe that more QM(DFT)/MM sampling is needed to confirm these findings.

Conclusions that the reference 48 authors take from their calculations are as follows: (1) Cpd I and Cpd 0 might coexist if the Cpd 0 to Cpd I energy barrier is not too large compared to the barrier for substrate oxidation (highly dependent, it appears, on the substrate itself); (2) energy levels for Cpd 0 and Cpd I both depend on the protein environment (or, for model compounds, on the porphyrin substituents); and (3) thr252 may not be needed as a proton source but is important as a hydrogen-bond donor or acceptor [as an acceptor it assists in the scission of **(2)<sub>prot</sub>** to yield Cpd I]. As evidence for conclusion (3), it has been found experimentally that thrOCH<sub>3</sub> substituted for thrOH forms a competent cytochrome P450 catalyst, indicating that the thr252 proton is not crucial for intermediate protonation, but rather is needed for hydrogen-bonding purposes. The reference 48 authors also speculate further about the **(2)<sub>prot</sub>** intermediate. This species might exist at low concentrations and have a short lifetime but would be a strong oxidant. It may act as the oxidant when Cpd I cannot form as experienced by Dawson and co-workers.<sup>44a</sup> Further study of the **(2)<sub>prot</sub>** intermediate should be forthcoming. Finally the authors state that they have studied the direct protonation mechanism only in relation to the PDB: 1DZ8 structure. Other mechanisms may exist and should be studied further.

#### 7.4.4 Analytical Methods: X-Ray Crystallography

In their *Science* magazine article of 2000,<sup>35</sup> the Schlichting group describes the process involved in obtaining the X-ray crystallographic structures of intermediates in the CYP101 catalytic cycle using cryogenic temperatures and rapid data-collection techniques (PDB: 1DZ4, 1DZ6, 1DZ8 and 1DZ9). Figure 7.14 identifies the intermediates whose structures have been determined. The researchers used the technique of freeze-trapping intermediates and determining their X-ray structures using monochromatic X rays. For instance, to progress from the dioxygen intermediate (Figure 7.13) to the activated oxygen species (Figure 7.17), the second electron in the CYP101 cycle was provided using the X-ray beam itself.

First the ferrous enzyme [PDB: 1DZ6, Figure 7.14 (**3**)] was obtained by diffusion of dithionite reducing agent into the PDB: 1DZ4 crystal (Figure 7.14, **(2)**). The dioxy intermediate [PDB: 1DZ8, Figure 7.14 (**4**)] was then generated by exposure to a high partial pressure of O<sub>2</sub>. Following this structure determination, the PDB: 1DZ8 crystals were irradiated with long-wavelength X rays, producing hydrated electrons within the crystal and presumably driving the reaction toward the activated oxygen species. A third data set was collected after the treated crystal was thawed and refrozen. This data set refined to show a “product complex” PDB: 1DZ9 with a ferryl–oxo porphyrin system in position to hydroxylate the C<sub>5</sub> position of the camphor molecule. [See Figure 7.14 (**6**), active site shown in Figure 7.17.] The reference 35 authors report that the oxidation state of the iron [possibly Fe(IV)] and the electronic state of the



PDB: 1DZ9 active site  
All distances in angstroms

**Figure 7.17** Visualization of cytochrome P450's active site with camphor—the activated oxygen intermediate (PDB: 1DZ9). Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. (Printed with permission of CambridgeSoft Corporation.)

heme (cation radical +•) cannot be unequivocally determined from the X-ray crystallographic structures.

Structural details for the ferric P450–camphor complex (PDB: 1DZ4) were found to be very similar to those found for the PDB: 2CPP structure obtained by Poulos et al.<sup>36</sup> In the PDB: 1DZ4 X-ray structure, the heme iron ion is covalently attached to the thiolate sulfur of cys357. The heme is ruffled and the five-coordinate iron is out of the porphyrin plane by 0.3 Å. The camphor substrate resides in the distal heme pocket held by a single hydrogen bond between its carbonyl oxygen and the side-chain hydroxyl of tyr96. A carboxylate group on the D pyrrole ring of the porphyrin (see Figure 7.1) forms a

hydrogen bond with the hydroxyl group of thr101. Reduction by dithionite to the ferrous P450–camphor complex with freeze-trapping to prevent re-oxidation produces the PDB: 1DZ6 structure—one that is very similar to the ferric analog (PDB: 1DZ4). Next, diffusion of molecular oxygen into the reduced enzyme crystals produced the P450–camphor–O<sub>2</sub> complex (PDB: 1DZ8). This unstable ternary complex's structure was determined with short-wavelength X rays to prevent radiolysis. As can be seen in Figure 7.13, the diatomic oxygen molecule is bound end-on ( $\eta^1$ ) to the heme iron [Fe(II) or Fe(III)–O distance equals 1.8 Å, O–O bonding distance equals 1.25 Å]. The O<sub>2</sub> distal oxygen (furthest from the iron ion) is within 3.5 Å of thr252's hydroxyl group. The camphor molecule has moved slightly away from its previous position but is still within van der Waals contact with the dioxygen (or superoxide) ligand. The porphyrin ring has become flatter and the iron ion is slightly above the porphyrin plane. Obvious changes from PDB: 1DZ4 or 1DZ6 structures upon dioxygen binding include the presence of an ordered water molecule, H<sub>2</sub>O<sub>901</sub>, close to both the dioxygen molecule and the hydroxyl group of thr252. New conformations of conserved residues asp251 and thr252 are seen as the carbonyl oxygen of asp251 has flipped by 90° toward asn255 and the amide nitrogen of thr252 has rotated toward the heme pocket. H<sub>2</sub>O<sub>901</sub> sits in a groove in the distal pocket and interacts with H<sub>2</sub>O<sub>902</sub> that in turn interacts with other water molecules (H<sub>2</sub>O<sub>687</sub>, H<sub>2</sub>O<sub>566</sub>, H<sub>2</sub>O<sub>523</sub>) in a chain reaching toward the highly conserved glu366. (See Figure 7.13.)

A second electron is now added by irradiation of the P450–camphor–O<sub>2</sub> complex (PDB: 1DZ8) with long-wavelength X-rays that results in changes in electron density around the active site. After prolonged exposure to the long-wavelength X rays, thawing and refreezing of these crystals, data collection, and refinement led to a difference map indicating that O–O bond cleavage had taken place (PDB: 1DZ9). The electron density was most consistent with a ferryl–oxo iron (**Compound I**), Fe<sup>IV</sup>=O (P<sup>•+</sup>), complex that had been previously observed in cytochrome c peroxidase and catalase structures.<sup>51</sup> In this structure, shown in Figure 7.17, the Fe(III) or Fe(IV)–O bond distance has shortened to 1.67 Å and the to-be-hydroxylated camphor C<sub>5</sub> atom is within 3 Å of the oxo oxygen atom. Water molecules H<sub>2</sub>O<sub>901</sub> and H<sub>2</sub>O<sub>902</sub> cannot be detected, but electron density for H<sub>2</sub>O<sub>903</sub> appears to establish a similar hydrogen-bonding network reaching toward the highly conserved glu366. The structure is similar to that found by the Poulos research group when P450<sub>CAM</sub> was co-crystallized with 5-exo-hydroxycamphor.<sup>52</sup> In the PDB: 1DZ9 structure, the asp251 carbonyl has flipped back to its original position in ferric P450<sub>CAM</sub> (PDB: 1DZ4).

How are the conserved residues asp251 and thr252 believed to take part in the catalytic mechanism? It is known that D251N (asp251asn) mutants exhibit catalytic rates decreased by two orders of magnitude. Replacement of thr252 leads to a large proportion of uncoupling (moving into the peroxide or oxidase shunts), leading to the conclusion that thr252's role is to provide an important hydrogen bond. The reference 35 authors believe that the P450<sub>CAM</sub>

intermediates that they have described explain how the conserved residues asp251 and thr252 become part of a hydrogen-bonding network (with water molecules) that could deliver protons to the iron-bound dioxygen. Specifically for PDB: 1DZ8, the P450<sub>CAM</sub>–O<sub>2</sub> complex, they believe the carbonyl oxygen of asp251 flips so that it can interact with the amide and amino groups of asn255 upon dioxygen binding. At the same time the amide nitrogen of thr252 rotates toward the active site, providing an additional hydrogen bond to stabilize H<sub>2</sub>O<sub>901</sub>. The thr252 rotation also places its hydroxyl in position to interact with the bound dioxygen molecule as well as both H<sub>2</sub>O<sub>901</sub> and H<sub>2</sub>O<sub>902</sub> (see Figure 7.13). In conclusion the reference 35 authors hope that the structures they presented will provide reference points for future modeling of the P450 reaction mechanism, either computationally or by design and synthesis of model compounds. Computational analyses of cytochrome P450 structures has been discussed previously in this section. This research, by Shaik and co-workers, postulated the “two states” theory—namely that high- and low-spin states of the compound I intermediate were responsible for differences in cytochrome P450 reactivity.<sup>46–48</sup>

#### 7.4.5 Cytochrome P450 Model Compounds

**7.4.5.1 Introduction.** The chemistry and biology of cytochrome P450 model compounds has a rich history. Many intermediates have been identified or proposed to exist in the catalytic cycles proposed by researchers studying cytochrome P450 biomimetic chemistry. These include (1) ferric–peroxy (Fe(III)–O<sub>2</sub><sup>2-</sup>) complexes, (2) oxoiron(IV) porphyrin π-cation radicals (first identified in peroxidase and catalase catalytic cycles and commonly called “compound I”),<sup>53</sup> (3) oxidant-iron(III) porphyrin adducts,<sup>54</sup> (4) oxoiron(V) porphyrins (isoelectronic with oxoiron(IV) porphyrin π-cation radicals),<sup>55</sup> and (5) oxoiron(IV) porphyrins<sup>56</sup> (known as “compounds II” in catalytic cycles of peroxidases and catalases). (See also Figure 7.14.) Note that formation of oxoiron(IV) porphyrin π-cation radicals requires a two-electron oxidation from the Fe(III) porphyrin resting state, whereas the oxoiron(IV) porphyrins require a one electron oxidation. Some of these intermediates are subjects of intense discussion among researchers in the field. For instance, oxoiron(IV) porphyrin π-cation radicals have been proposed as the sole oxidant in P450 chemistry, whereas other researchers have proposed multiple oxidizing species in these oxygen transfer reactions.<sup>57</sup> Evidence in support of the various intermediates proposed for cytochrome P450 model compounds will be presented in Section 7.4.5.3.

**7.4.5.2 A Cytochrome P450 Model Compound: Structural.** The crystal structure of a ferric H93G (his93gly) myoglobin, Mb, cavity mutant at a resolution of 1.7 Å has been published recently (PDB: 1EVP).<sup>58</sup> The H93G mutant protein includes a myoglobin (Mb) heme featuring a high-spin five-coordinate

$\text{Fe}^{\text{III}}$  with the cysteine-like ligand,  $\beta$ -mercaptoethanol (BME) thiolate, an anionic sulfur donor, residing in the proximal cavity. Because of glycine's smaller size, the H93G cavity mutant has a larger proximal pocket that will accommodate foreign natural ligands like BME. This cysteinate-ligated heme complex serves as a structural model for cytochrome P450 as well as other enzymes with similar cofactor cores such as *Calderomyces fumago* chloroperoxidase and nitric oxide synthase. It is known that cysteine must be deprotonated to cysteinate throughout the catalytic cycles of these enzymes. Dawson has postulated that a "push" from the deprotonated cysteine ligand is important for the monooxygenase activity of cytochrome P450.<sup>59</sup> It is known that three amide N–H···S hydrogen bonds from leu358, gly359, and gln360 help retain cysteinate coordination in the proximal pocket throughout the catalytic cycle as iron changes its oxidation from  $\text{Fe}^{\text{II}}$  to  $\text{Fe}^{\text{III}}$  to  $\text{Fe}^{\text{IV}}$  and back again. In the PDB: 1 EVP crystal structure, the BME ligand appears to have an intramolecular hydrogen bond that mimics one of the necessary N–H···S hydrogen bonds in cytochrome P450. The reference 59 authors compare the structure of the model complex, ferric H93G-BME Mb, with the Schlichting X-ray crystallographic study of the analogous five-coordinate, high-spin  $\text{Fe}^{\text{III}}$  P450<sub>CAM</sub> enzyme, PDB: 1DZ4. This structure, which has two symmetry-independent molecules, is discussed in Section 7.4.4 above. Comparison of bond distances and angles are collected in Table 7.3.

Note that the  $S_{\gamma}$ – $\text{Fe}^{\text{III}}$ –C bond angle for the cysteinate ligand suggests that the sulfur atom is  $sp^3$  hybridized and that the  $\text{Fe}^{\text{III}}$ –S bond is covalent. The water molecule at 4.2 Å distance from  $\text{Fe}^{\text{III}}$  in the distal pocket is within hydrogen bonding distance to distal his64 (3.0 Å). The good agreement in the structural data shows that  $\text{Fe}^{\text{III}}$  H93G-BME Mb is an excellent protein-based model system for  $\text{Fe}^{\text{III}}$  P450<sub>CAM</sub>. Spectral data are in good agreement as well. The UV-vis spectrum of  $\text{Fe}^{\text{III}}$  H93G-BME Mb shows a prominent peak at 618 nm that is typical of high-spin ferric heme complexes, and its magnetic circular

**TABLE 7.3 Bond Distances and Angles for  $\text{Fe}^{\text{III}}$  H93G-BME Mb with  $\text{Fe}^{\text{III}}$  P450<sub>CAM</sub>**

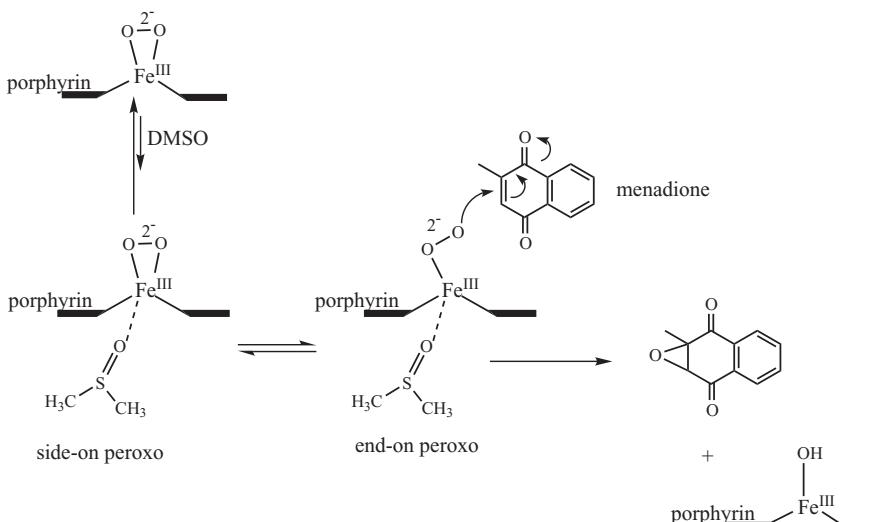
| Bond Distance (Å) or Bond Angle (°)                           | $\text{Fe}^{\text{III}}$ H93G-BME Mb                   | $\text{Fe}^{\text{III}}$ P450 <sub>CAM</sub> <sup>a</sup> |
|---|--|---|
| Proximal ligand, $\text{Fe}^{\text{III}}$ –S                  | 2.41   | 2.37<br>2.36  |
| Distal ligand   | None closer than 4 Å<br>$\text{H}_2\text{O}$ at 4.22 Å | None closer than 4 Å                                      |
| Distance between $\text{Fe}^{\text{III}}$ and porphyrin plane | 0.4  | 0.4   |
| $S_{\gamma}$ – $\text{Fe}^{\text{III}}$ –C bond angle         | 111  | 108<br>110  |

<sup>a</sup>Data given for two symmetry-independent molecules in the unit cell.

dichroism (MCD) spectrum is typical of a five-coordinate thiolate-ligated ferric heme. The question as to why the BME ligand preferentially binds in the proximal pocket in the enzyme model can be answered by noting distal pocket steric hindrance provided by amino acid residues his64 and val68. Large conformational changes would be required to place the BME ligand in the distal pocket (dioxygen-binding site). In conclusion, one should expect future research using the Fe<sup>III</sup> H93G-BME Mb system as a functional model in mechanistic and catalytic studies of the cytochrome P450 enzyme.

**7.4.5.3 Cytochrome P450 Model Compounds: Functional.** Ferric-peroxo species are part of the cytochrome P450 catalytic cycle as discussed previously in Section 7.4.4. For instance, these ferric-peroxo moieties are known to act as nucleophiles attacking aldehydic carbon atoms in oxidative deformylations to produce aromatic species.<sup>60</sup> An example of this work, establishing the nucleophilic nature of [(porphyrin)Fe<sup>III</sup>(O<sub>2</sub>)]<sup>-</sup> complexes, was achieved for alkene epoxidation reactions by J. S. Valentine and co-workers.<sup>61</sup> The electron-deficient compound menadione (see Figure 7.18) yielded menadione epoxide when reacted with a [(porphyrin)Fe<sup>III</sup>(O<sub>2</sub>)]<sup>-</sup> complex.

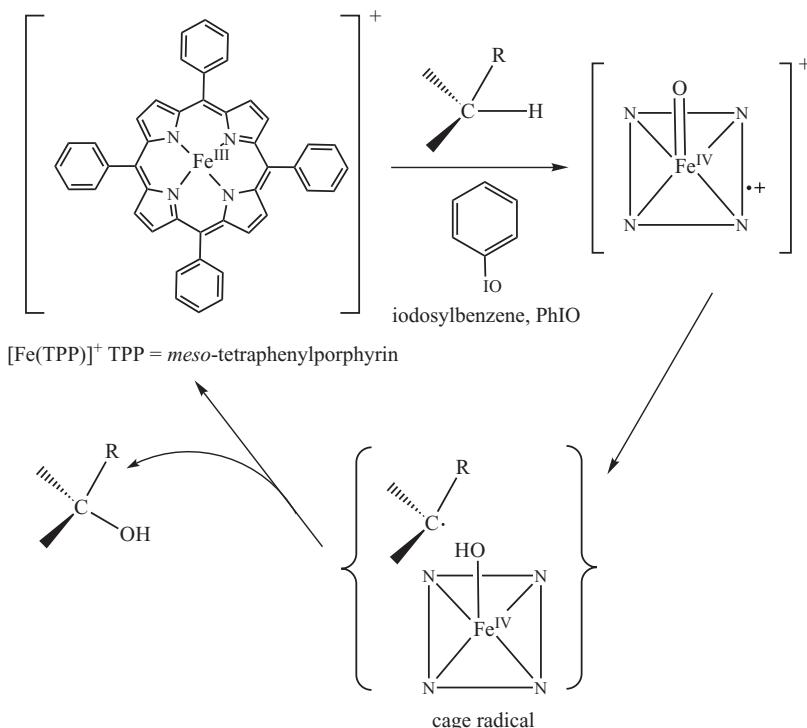
These researchers found that the nucleophilicity of the complex depended on the electronic nature of the porphyrin. While a [(TPFPP)Fe<sup>III</sup>(O<sub>2</sub>)]<sup>-</sup> (TPFPP equals the dianion of *meso*-tetrakis(pentafluorophenyl)porphyrin) complex did not epoxidize alkenes, adding dimethylsulfoxide (DMSO) as an axial ligand restored the complexes' nucleophilicity and ability to epoxidize alkenes. The authors ascribed this restored capability to push the side-on peroxy into a more open, end-on, nucleophilic conformation.<sup>62</sup> Watanabe and co-workers



**Figure 7.18** Epoxidation of menadione by a [(porphyrin)Fe<sup>III</sup>(O<sub>2</sub>)]<sup>-</sup> complex.

found that a  $[(\text{porphyrin})\text{Fe}^{\text{III}}(\text{O}_2)]^-$  complex would deformylate certain aldehydes while a corresponding  $[(\text{porphyrin}^+)^*\text{Fe}^{\text{IV}}-\text{O}]^+$  complex would not.<sup>63</sup>

Publications concerning important early cytochrome P450 model compounds and proposed catalytic cycles by the J. T. Groves group became available starting in the late 1970s.<sup>53a,64</sup> The first article on the use of a synthetic iron(III) porphyrin complex,  $[\text{Fe}(\text{TPP})\text{Cl}]$ , appeared in 1979 (TPP equals the dianion of *meso*-tetraphenylporphyrin). This complex, in the presence of the oxidant iodosylbenzene,  $\text{PhIO}$ , used as a source of oxygen atoms, was found to catalyze alkene epoxidation and alkane hydroxylation reactions. Groves enunciated the “oxygen rebound” mechanism to explain the reactions. In the hydroxylation mechanism, the oxoiron(IV) porphyrin  $\pi$ -cation radical intermediate,  $[\text{Fe}^{\text{IV}}=\text{O}(\text{TPP}^+)]^+$ , forms in the presence of  $\text{PhIO}$  (or other oxidants such as the commonly used 3-chloroperoxybenzoic acid, *m*-CPBA), and it is this intermediate that abstracts a hydrogen atom from the substrate ( $\text{R}-\text{H}$ ) to form the “caged” carbon radical and  $[\text{Fe}^{\text{IV}}=\text{OH}(\text{TPP})]^+$ . The cage ensures that the intermediates do not diffuse away from each other. The two species then recombine in the rebound step to produce the hydroxylated product and the resting state of the  $\text{Fe}^{\text{III}}$ -porphyrin complex (see Figure 7.19).

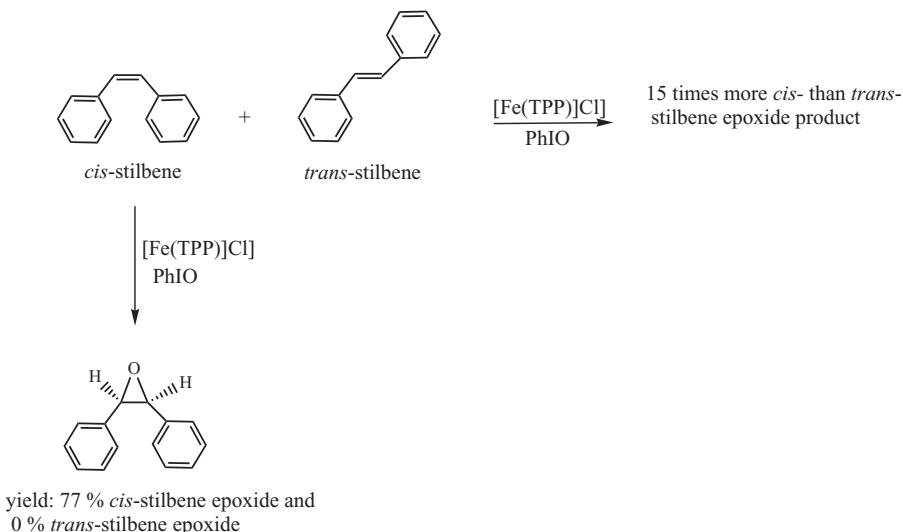


**Figure 7.19** The Groves oxygen rebound mechanism for alkane hydroxylation.

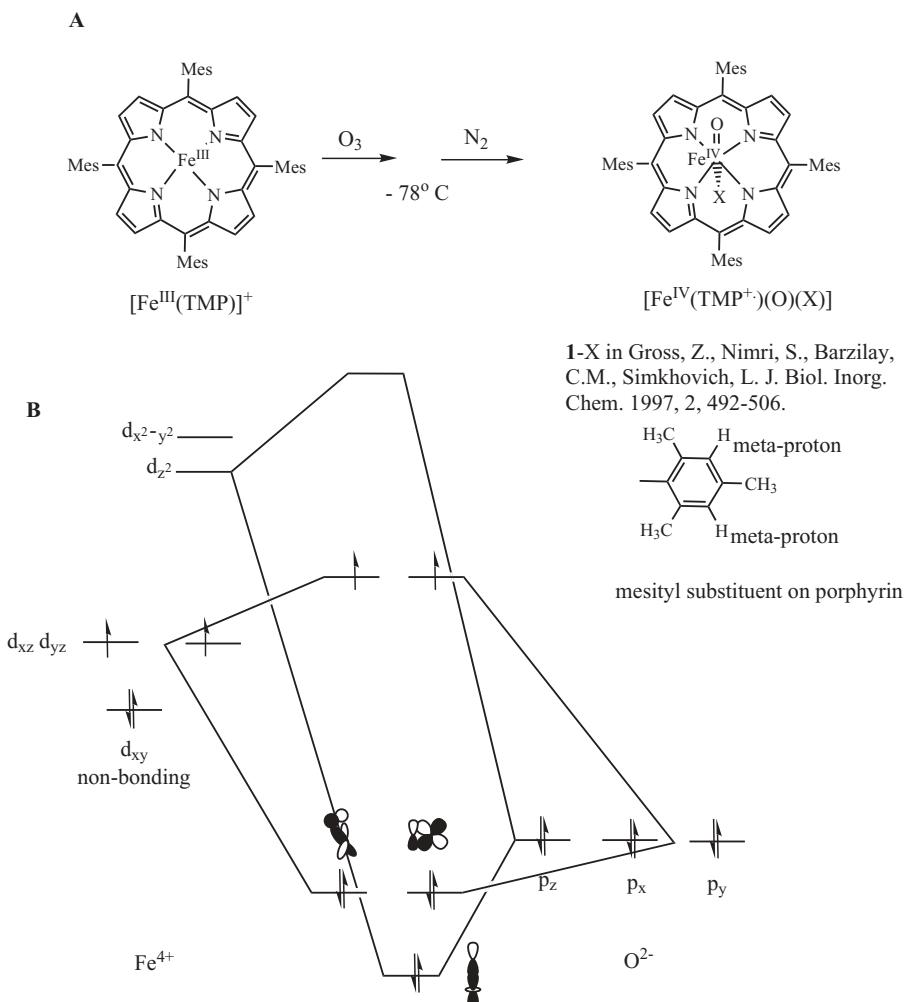
The oxygen rebound mechanism was supported by experimental evidence including (1) high kinetic isotope effects, (2) partial positional or stereochemical scrambling, and (3) allylic rearrangements. For instance, in the presence of  $[\text{Fe}(\text{TPP})\text{Cl}]$  and PhIO, *cis*-stilbene was stereospecifically epoxidized. In addition, it was found that *cis*-stilbene was 15 times more reactive than *trans*-stilbene in competitive epoxidations. (see Figure 7.20).<sup>64b</sup>

The authors ascribed the *cis*-stilbene preference to steric interference between the phenyl groups of *trans*-stilbene and the phenyl groups on the  $[\text{Fe}^{\text{IV}}=\text{O}(\text{TPP}^{+*})]^+$  intermediate. The Groves research group also gathered the first spectroscopic indications of a synthetic, compound I analog.<sup>53a</sup> The  $[\text{Fe}^{\text{IV}}=\text{O}(\text{TMP}^{+*})]^+$  complex (TMP equals the dianion of *meso*-tetramesitylphenylporphyrin) was prepared by the *m*-CPBA (3-chloroperoxybenzoic acid) oxidation of  $[\text{Fe}^{\text{III}}=\text{O}(\text{TMP})\text{Cl}]$ . (See Figure 7.21.)

The  $[\text{Fe}^{\text{IV}}=\text{O}(\text{TMP}^{+*})]^+$  complex exhibited a characteristic bright green color and corresponding visible absorbance in its UV-vis spectrum. In its NMR spectrum, the meta-proton doublet of the porphyrin mesityl groups were shifted more than 70 ppm downfield from tetramethylsilane (TMS) because they were in the presence of the cation radical, while the methyl protons shift between 10 and 20 ppm downfield. In Mössbauer spectroscopy, the isomer shift,  $\delta$  of 0.06 mm/s, and  $\Delta E_Q$  value of 1.62 mm/s were similar to those for other known Fe(IV) complexes. Electron paramagnetic resonance (EPR),<sup>65</sup> resonance Raman (RR),<sup>66</sup> and EXAFS<sup>67</sup> spectroscopies provided additional indications of an  $[\text{Fe}^{\text{IV}}=\text{O}] \pi$ -cation radical intermediate. For instance,



**Figure 7.20** Epoxidation of *cis*- and *trans*-stilbene. (Adapted with permission from Table 1 of reference 64b. Copyright 1983 American Chemical Society.)



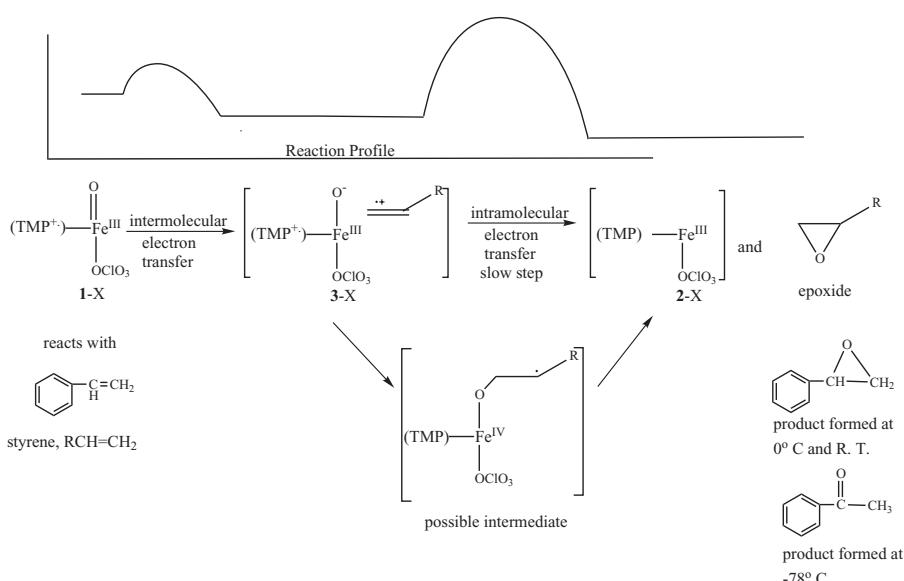
**Figure 7.21** (A) Formation of the  $[\text{Fe}^{\text{IV}}=\text{O}(\text{TMP}^{+*})]^+$  complex using the TMP ligand. (B) Molecular orbital diagram describing bonding in the  $[\text{Fe}^{\text{IV}}=\text{O}(\text{TMP}^{+*})]^+$  complex. (Adapted with permission from Scheme 4 and Figure 2 of reference 68. Copyright 1997, Society of Biological Inorganic Chemistry.)

EXAFS data indicated an  $\text{Fe}=\text{O}$  bond length of  $1.65 \pm 0.05 \text{ \AA}$ , similar to that found for compound I intermediates in several enzymes such as catalases, horseradish peroxidase, cytochrome c peroxidase, and chloroperoxidases.

Studies by other researchers showed that substituents on the porphyrin ligand affected formation of the  $[\text{Fe}^{\text{IV}}=\text{O}(\text{P}^{+*})]^+$  radical. Electron-deficient

iron(III) porphyrin complexes were found to be better catalysts in hydrocarbon oxygenation reactions. For instance,  $[\text{Fe}^{\text{IV}}=\text{O}(\text{TPFPP}^{+*})]^+$  complex (TPFPP equals the dianion of *meso*-tetrakis(pentafluorophenyl)porphyrin), hydroxylates alkanes rapidly and efficiently, whereas  $[\text{Fe}^{\text{IV}}=\text{O}(\text{TMP}^{+*})]^+$ , TMP = dianion of *meso*-tetramesitylphenylporphyrin, does not hydroxylate alkanes under identical conditions.<sup>53c</sup> Gross et al.<sup>68</sup> found that oxygen atom transfer to olefins (epoxidation) is affected by the axial ligand trans to the oxo ligand in  $[\text{Fe}^{\text{IV}}=\text{O}(\text{TMP}^{+*})\text{X}]$ , where  $\text{X} = \text{F}^-$ ,  $\text{CH}_3\text{OH}$ ,  $\text{Cl}^-$ ,  $\text{CH}_3\text{COO}^-$ ,  $\text{CF}_3\text{SO}_3^-$ ,  $\text{ClO}_4^-$ . The rate of styrene epoxidation was in the order  $\text{F}^- >> \text{CH}_3\text{OH} > \text{Cl}^- > \text{CH}_3\text{COO}^- > \text{CF}_3\text{SO}_3^-$ . When  $\text{X} = \text{ClO}_4^-$ , epoxidation of styrene did not take place under the same reaction conditions.<sup>53e</sup> In a continuing study of the effect of the axial ligand on models for cytochrome P450 intermediate (compound I) oxidation reactions, the group showed evidence consistent with a multistep catalytic mechanism.<sup>68</sup> In this proposed mechanism, electron transfer from the olefin is followed by  $\sigma$ -bond formation between oxygen and the olefin radical, rearrangement of the electronic structure, and, finally, release of the products (see Figure 7.22). The study shows that the major kinetic effect of the axial ligands occurs during intramolecular electron transfer from Fe(III) to the porphyrin radical.<sup>68</sup>

Spectroscopic evidence for the  $[\text{Fe}^{\text{IV}}=\text{O}(\text{TMP}^{+*})\text{X}]$  (**1-X**) moiety included that from NMR, EPR, and RR. The EPR spectra of frozen solutions of **1-X**



**Figure 7.22** Cytochrome P450 model compound reaction profile consistent with a multi-step catalytic mechanism. (Adapted with permission from Schemes 7 and 8 of reference 68. Copyright 1997, Society of Biological Inorganic Chemistry.)

complexes at 15 K showed a characteristic  $S = 3/2$  spectrum with  $g_z = 2$ ,  $g_y = 3.5\text{--}3.7$ , and  $g_x = 4.2\text{--}4.4$ . The spin state of 3/2 arises from the two unpaired electrons for a  $\text{Fe}^{4+}\text{--O}^{2-}$  bond plus the unpaired electron from the porphyrin  $\pi$ -cation radical. [See Figure 7.21 for a molecular orbital (MO) diagram.] The NMR data are similar to that found by Groves for  $[\text{Fe}^{\text{IV}}=\text{O}(\text{TMP}^{+\bullet})]^+$  as shown in Table 7.4. As can be noted in this table, the axial ligand, X, does affect the shifts of the mesityl and pyrrole protons.

Resonance Raman (RR) data also show differences for the  $\nu_{\text{Fe}=\text{O}}$  stretching frequency, depending on the axial ligand. For  $\text{X} = \text{F}^-$  and  $\text{Cl}^-$  we have  $\nu_{\text{Fe}=\text{O}} = 801\text{--}806\text{ cm}^{-1}$ , whereas for  $\text{CH}_3\text{OH}$  and  $\text{ClO}_4^-$  we have  $\nu_{\text{Fe}=\text{O}} = 831\text{--}835\text{ cm}^{-1}$ . The reference 68 authors interpret these data as indicating weaker  $\text{Fe}=\text{O}$  bonding when the axial ligands are electron-donating *trans*-donor  $\text{F}^-$  and  $\text{Cl}^-$ . In the molecular orbital diagram of Figure 7.21, one can see that donation of both  $\sigma$ - and  $\pi$ -electron density by the *trans* ligands will have the effect of weakening the  $\text{Fe}=\text{O}$  bond. In this diagram, the low-spin  $\text{Fe}(\text{IV})$  electrons,  $d^4$ , in an approximate octahedral field will have the electron configuration:  $(d_{xy})^2(d_{xz})^1(d_{yz})^1$ . A bond order of 2 is achieved by one two-electron ( $2c\text{--}2e$ )  $d_z^2\text{--}p_z$   $\sigma$ -bond and two three-electron ( $2c\text{--}3e$ )  $\pi$ -bonds,  $d_{xz}\text{--}p_x$  and  $d_{yz}\text{--}p_y$  using the filled  $p$ -shell electrons of  $\text{O}^{2-}$ . The reference 68 researchers studied the kinetic profiles of the various  $[\text{Fe}^{\text{IV}}=\text{O}(\text{TMP}^{+\bullet})(\text{X})]$  species, finding that the **1** $\text{-X}$  compounds underwent rapid intermolecular electron transfer to a **3** $\text{-X}$  intermediate, possibly a  $\text{Fe}(\text{III})$  porphyrin radical with one electron transferred to form a substrate radical. In the following rate-determining step, intramolecular electron transfer would take place to form an  $[\text{Fe}^{\text{IV}}(\text{TMP})(\text{O})(\text{X})]$  intermediate in intimate contact with the substrate radical. Fast transfer of the oxygen atom to substrate completes the reaction to regenerate the  $[\text{Fe}^{\text{III}}(\text{TMP})(\text{X})]$  (**2** $\text{-X}$ ) complex plus epoxide product. The authors present NMR and EPR evidence that points to the **3** $\text{-X}$  intermediate in reactions of  $[\text{Fe}^{\text{IV}}=\text{O}(\text{TMP}^{+\bullet})(\text{ClO}_4^-)]$  with styrene. The reaction with  $\text{X} = \text{ClO}_4^-$  is slow enough so that the **3** $\text{-X}$  intermediate is detectable at low temperatures (see Figure 7.22). In addition to the epoxide product, formed at 0°C and room temperature, the product isomer phenylacetaldehyde is formed at -78°C. The authors postulate, in conclusion, that axial ligands X have most effect on the

**TABLE 7.4** NMR Data for  $[\text{Fe}^{\text{IV}}=\text{O}(\text{TMP}^{+\bullet})]^+$  and  $[\text{Fe}^{\text{IV}}=\text{O}(\text{TMP}^{+\bullet})(\text{X})]$  Complexes<sup>a</sup>

| $[\text{Fe}^{\text{IV}}=\text{O}(\text{TMP}^{+\bullet})(\text{X})]^{(1\text{-X})^{68}}$ | Mesityl meta-H | Ortho- $\text{CH}_3$ | Para- $\text{CH}_3$ | Pyrrole-H of Porphyrin |
|---|----------------|----------------------|---------------------|------------------------|
| $\text{X} = \text{F}^-$   | 71.4           | 26.9, 42.1           | 10.7                | -13.0                  |
| $\text{X} = \text{ClO}_4^-$   | 65.3           | 26.0, 23.5           | 11.1                | -26.9                  |
| $[\text{Fe}^{\text{IV}}=\text{O}(\text{TMP}^{+\bullet})]^{+53\text{a}}$                 | ~70            | ~20                  | ~10                 |                        |

<sup>a</sup> Referenced to tetramethylsilane (TMS).

Sources: References 53a and 68.

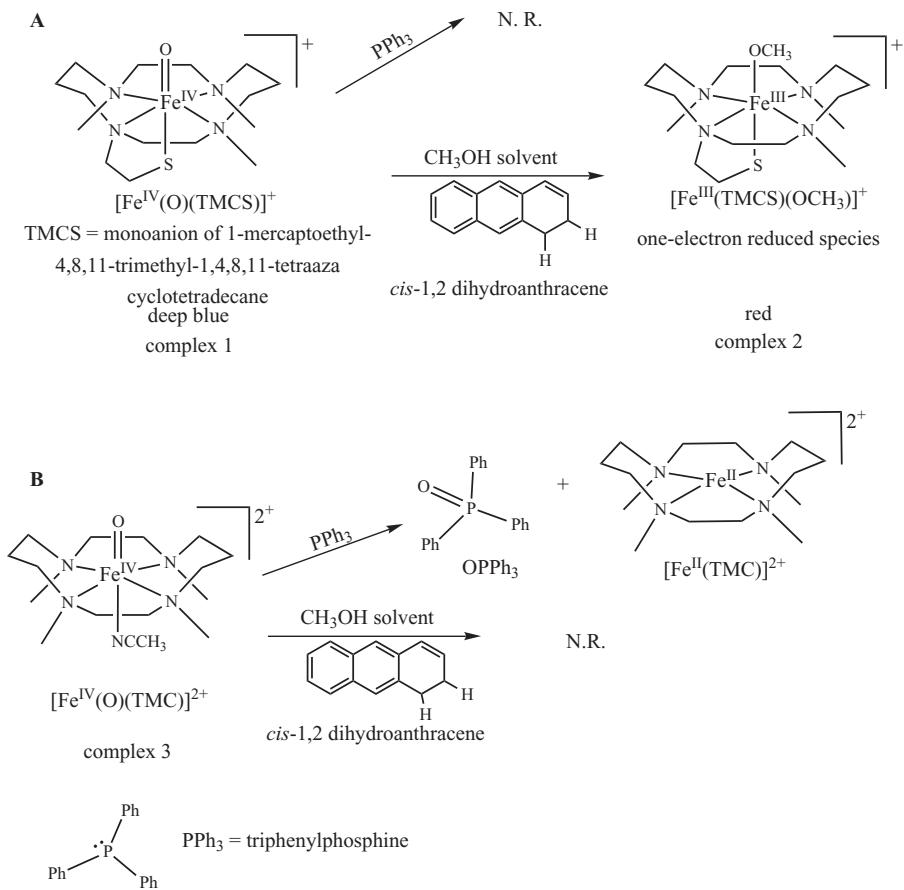
transformation of intermediate  $[(\text{TMP}^{+*})\text{Fe}^{\text{III}}(\text{O}^-)\text{X}]$ , **3-X**, to the possible intermediate  $[(\text{TMP})\text{Fe}^{\text{IV}}(\text{O})\text{X}]$ .

Other examples of oxidant–iron(III) adducts as intermediates in iron porphyrin-catalyzed reactions have been published as listed in references 54a–k. Competitive alkene epoxidation experiments catalyzed by iron porphyrins with peroxy acids,  $\text{RC(O)OOH}$ , or idosylarenes as oxidants have been proposed to have various intermediates such as  $[(\text{porphyrin})\text{Fe}^{\text{III}}(\text{O}-\text{O}-\text{C(O)R})]$  or  $[(\text{porphyrin})\text{Fe}^{\text{III}}(\text{O}-\text{I}-\text{Ar})]$ . Alkane hydroxylation experiments catalyzed by iron porphyrins with oxidant 3-chloroperoxybenzoic acid, *m*-CPBA, have been proposed to operate through the  $[(\text{porphyrin})\text{Fe}^{\text{III}}(\text{O}-\text{O}-\text{C(O)R})]$  intermediate. J. P. Collman and co-workers postulated multiple oxidizing species,  $[(\text{TPFPP}^{+*})\text{Fe}^{\text{IV}}=\text{O}]$  and/or  $[(\text{TPFPP})\text{Fe}^{\text{III}}(\text{O}-\text{I}-\text{Ar})]$  in alkane hydroxylations carried out with various iodosylarenes in the presence of  $\text{Fe}(\text{TPFPP})\text{Cl}$ , where TPFPP is the dianion of *meso*-tetrakis(pentafluorophenyl)porphyrin.<sup>54g</sup>

Oxoiron(V) porphyrins—red in color—an isoelectronic form of oxoiron(IV) porphyrin cation radicals—green—have been proposed (references 55a–c), although the density functional theory calculations have indicated that there are no true oxoiron(V) porphyrins (references 55d–f). Further spectroscopic studies are necessary to confirm the existence of this intermediate.

Oxoiron(IV) porphyrins, one electron above the Fe(III) porphyrin resting state, are known as compounds II in the catalytic cycles of peroxidases and catalases. These intermediates are thought to be poor oxidants, but several research groups have observed very different reactivity for a  $[(\text{TMP})\text{Fe}^{\text{IV}}=\text{O}]$  complex when compared to that for  $[(\text{TMP}^{+*})\text{Fe}^{\text{IV}}=\text{O}]$ . For instance, Groves and co-workers found a mixture of *cis*- and *trans*- $\beta$ -methylstyrene oxide (*cis/trans*~1.2) in reactions involving the  $[(\text{TMP})\text{Fe}^{\text{IV}}=\text{O}]$  complex, whereas  $[(\text{TMP}^{+*})\text{Fe}^{\text{IV}}=\text{O}]$  under the same conditions yielded a *cis/trans* ratio of approximately 11:1.<sup>56c</sup> In general, researchers have found that the oxidizing power of the oxoiron(IV) porphyrin complexes can be tuned by altering the nature of the substituents on the porphyrin ligands. The oxidation mechanism involving these complexes is less well studied than that for other species discussed previously.

Conclusive evidence still does not exist for the presence of thiolate-ligated oxoiron(IV) species in cytochrome P450 catalytic cycles—although see the Schlichting X-ray crystallographic studies in Section 7.4.4. And chemists have attempted, rather unsuccessfully, to synthesize oxoiron(IV) porphyrin model complexes with a thiolate axial ligand. Since the confining enzyme pocket present in the natural enzyme is missing in the model compounds, and the sulfur ligand is as easily oxidized as the iron center itself, efforts to synthesize such complexes have largely failed. However, Eckard Münck and Larry Que have recently published a report on the synthesis and characterization of  $[\text{Fe}^{\text{IV}}(\text{O})(\text{TMCS})]^+$  where TMCS is a pendentate ligand, the monoanion of 1-mercaptoproethyl-4,8,11-trimethyl-1,4,8,11-tetraaza cyclotetradecane.<sup>69</sup> The ligand provides a square pyramidal, non-heme,  $(\text{N}_4\text{SR})_{\text{apical}}$  ligand set that is similar to that in cytochrome P450. (See Figure 7.23.)



**Figure 7.23** Reactions of (A)  $[\text{Fe}^{\text{IV}}(\text{O})(\text{TMCS})]^+$  and (B)  $[\text{Fe}^{\text{IV}}(\text{O})(\text{TMC})]^{2+}$ .

The  $[\text{Fe}^{\text{IV}}(\text{O})(\text{TMCS})]^+$  complex was synthesized at 90% yield starting with  $[\text{Fe}^{\text{II}}(\text{TMCS})]^+$  and using one equivalent of *m*-chloroperbenzoic acid (*m*CPBA) as oxidant in the presence of excess base. The deep blue complex is stable at  $-60^\circ\text{C}$  but decays slowly as the temperature is raised. Mössbauer data ( $\Delta E_Q = 0.22 \text{ mm/s}$  and  $\delta = 0.19 \text{ mm/s}$ ) support the Fe(IV) oxidation state assignment as do X-ray absorption spectroscopy experiments. Extended X-ray absorption fine structure (EXAFS) indicates one oxygen atom at  $1.70(2) \text{ \AA}$ , three nitrogen/oxygen atoms at  $2.09(2) \text{ \AA}$ , one sulfur atom at  $2.33(2) \text{ \AA}$ , and four carbons at  $2.95(2) \text{ \AA}$ . The supposed  $\text{Fe}=\text{O}$  bond distance is longer than that found for other characterized  $\text{Fe}=\text{O}$  species. For instance, the  $\text{Fe}=\text{O}$  bond distance in PDB: 1DZ9 (Section 7.4.4), believed to contain a  $\text{Fe}^{\text{IV}}=\text{O}$  compound I-type moiety is  $1.67 \text{ \AA}$ . The  $\text{Fe}=\text{S}$  distance is significantly shorter than that computed for cytochrome P450 compounds I ( $2.6 \text{ \AA}$ ) and II ( $2.5 \text{ \AA}$ )<sup>70</sup> but are close to those found for PDB: 1DZ9. (See Section 7.4.4 and Figure 7.17.) The researchers found that

the  $[\text{Fe}^{\text{IV}}(\text{O})(\text{TMCS})]^+$  (Figure 7.23, complex 1) did not react with triphenylphosphine,  $\text{PPh}_3$ , to form triphenylphosphine oxide,  $\text{OPPh}_3$  (an oxo-atom two-electron oxidation), while the complex did react with dihydroanthracene at  $-40^\circ\text{C}$  in  $\text{CH}_3\text{OH}$ , undergoing a one-electron reduction to a red species. (See Figure 7.23.) This product exhibits an electrospray ionization mass spectroscopy (ESI-MS) ion at  $m/z = 388.1944$ , corresponding to  $[\text{Fe}(\text{TMCS})(\text{OCH}_3)]^+$  (Figure 7.23, complex 2). The authors postulate that  $[\text{Fe}^{\text{IV}}(\text{O})(\text{TMCS})]^+$  decays by abstracting a hydrogen atom from a dihydroanthracene molecule to form  $[\text{Fe}^{\text{III}}(\text{TMCS})(\text{OH})]^+$  that converts to  $[\text{Fe}^{\text{III}}(\text{TMCS})(\text{OCH}_3)]^+$  in the  $\text{CH}_3\text{OH}$  solvent. Thus, in this model compound, the  $\text{Fe}^{\text{IV}}=\text{O}$  unit is not a two-electron oxo-atom transfer agent but a one-electron oxidant hydrogen abstraction reagent. The conclusion reached is that thiolate ligands in the axial position of  $[\text{Fe}^{\text{IV}}(\text{O})(\text{TMCS})]^+$  P450 model compounds cause more drastic changes in reactivity than found for other axial ligands in similar model compounds. For instance the compound  $[\text{Fe}^{\text{IV}}(\text{O})(\text{TMC})(\text{NCCH}_3)]^+$  (Figure 7.23, complex 3) where TMC = 1,4,8,11-tetramethyl-1,4,8,11-tetraazacyclotetradecane and  $\text{NCCH}_3$  is acetonitrile, does produce the oxo-atom two-electron product,  $\text{OPPh}_3$  when reacted with  $\text{PPh}_3$  (see Figure 7.23).

#### 7.4.6 Cytochrome P450 Conclusions

Cytochrome P450 is a monooxygenase that carries an Fe(II)/Fe(III) heme b cofactor identical to that of hemoglobin and myoglobin. The most-studied cytochrome P450,  $\text{P450}_{\text{CAM}}$ , hydroxylates the camphor molecule. Spectroscopic methods and X-ray crystallography have been used to identify many intermediates in the P450 catalytic cycle, although questions remain as to the exact mechanism of P450-catalyzed hydroxylations, epoxidations, and other monooxygenase-type reactions. The study of P450 model compounds has assisted in identifying possible intermediates in reaction pathways. Some conclusions from study of model compound reactions include the following: (1) Nucleophilic ferric-peroxo porphyrins may epoxidize electron deficient alkenes and deformylate aldehydes; (2) isolation and characterization of oxoiron(IV) porphyrin cation radicals have been very successful in P450 model compounds in contrast to controversy that still surrounds the participation of this intermediate in the P450 enzyme's catalytic cycle; (3) participation of oxidant-iron(III) porphyrin adducts in P450 model compounds may still be questionable; and (4) the mechanism for the activity of oxoiron(IV) porphyrins in oxygenating organic substrates remains largely ill-defined.

### 7.5 CYTOCHROME b(6)f: A GREEN PLANT CYTOCHROME

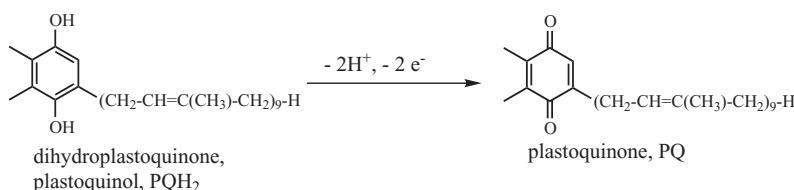
#### 7.5.1 Introduction

Cytochrome b(6)f protein complex, found in green plants, is part of the photosynthetic system—the principal producer of both oxygen and organic

matter on earth. One website devoted to cytochrome b(6)f is found at <http://www.biology.purdue.edu/people/faculty/cramer/Cramer/html/cytb.html>. Four multi-subunit membrane proteins take part in the conversion of the sunlight into chemical energy. These are: photosystem I, photosystem II, cytochrome b(6)f complex, and F-ATPase.<sup>71</sup> Photosystem I generates the most negative redox potential in nature and thus largely determines the global amount of enthalpy in living systems. The cytochrome b(6)f complex interacts with photosystems I and II (PSI and II) by receiving electrons from PSII and passing them to PSI, at the same time pumping protons across the membrane and powering the Q cycle. The Q cycle is the name given to the movement of lipophilic quinone species across membranes coupled to redox reactions and proton uptake and release.<sup>72</sup> Figure 1 of reference 73 shows a schematic diagram of the PSII, cytochrome b(6)f, PSI complex. X-ray crystallographic structures for PSI (PDB: 1JB0)<sup>74</sup> and PSII (PDB: 1FE1, 1IZL)<sup>75</sup> have been published.

The cytochrome b(6)f complex mediates electron transfer between the PSI and PSII reaction centers by oxidizing lipophilic plastoquinol ( $\text{PQH}_2$ ) (see Figure 7.24) and reducing the enzymes plastocyanin or cytochrome  $c_6$ . The electronic connection also generates a transmembrane electrochemical proton gradient that can support adenosine triphosphate (ATP) synthesis instead of electron transport.

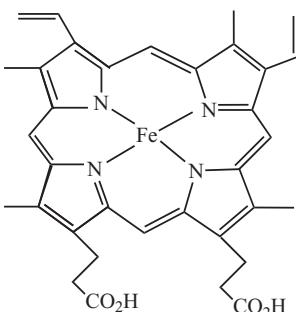
The b(6)f complex consists of a total of four hemes, two belonging to the cytochrome b family, one belonging to the cytochrome f family, a heme “x,” and a [2Fe–2S] Rieske (RISP or ISP) iron–sulfur cluster. The iron–sulfur protein (ISP) component of the cytochrome b(6)f and cytochrome  $bc_1$  (Section 7.6) complexes, now called the “Rieske” ISP, was first discovered and isolated by John S. Rieske and co-workers in 1964 (in the cytochrome  $bc_1$  complex). The homologues of the Rieske proteins include ISP components of cytochrome b(6)f and cytochrome  $bc_1$  complexes, aromatic-ring-hydroxylating dioxygenases (phthalate dioxygenase, benzene, naphthalene and toluene 1,2-dioxygenases), and arsenite oxidase. Comparison of amino acid sequences has revealed the following consensus sequence: cys–X–His–(X)<sub>15–17</sub>–cys–X–X–his, where X stands for any amino acid. The crystal structures of a number of Rieske proteins, including cytochrome b(6)f (discussed in this section) and cytochrome  $bc_1$  (Section 7.6), are known. The overall Rieske protein fold is dominated by an antiparallel  $\beta$ -structure and contains only one  $\alpha$ -helix. The



**Figure 7.24** The plastoquinol/plastoquinone reaction.

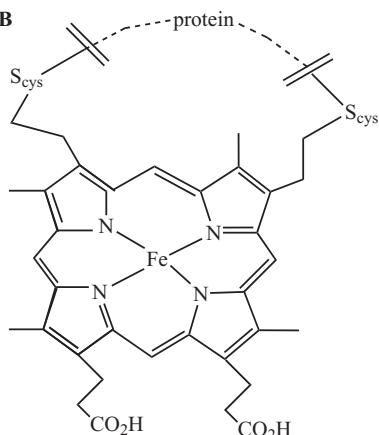
[Fe<sub>2</sub>S<sub>2</sub>] cluster-binding subdomains have the topology of an incomplete anti-parallel β-barrel. One iron atom of the Rieske [Fe<sub>2</sub>S<sub>2</sub>] cluster is coordinated by two cysteine residues through the S<sub>γ</sub> atom and the other is coordinated by two histidine residues through the N<sub>δ</sub> atoms. The ligands coordinating the cluster originate from two loops, with each loop contributing one cysteine and one histidine ligand. See Figure 7.25 for the structures of the cytochrome b, cytochrome f, and heme “x” hemes.

A



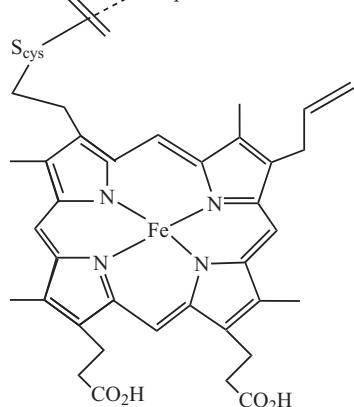
heme b, protoporphyrin IX, Fe(II) and Fe(III) distal and proximal histidine ligands

B



heme c, covalent connection through two cysteine S atoms to protein chain.

C



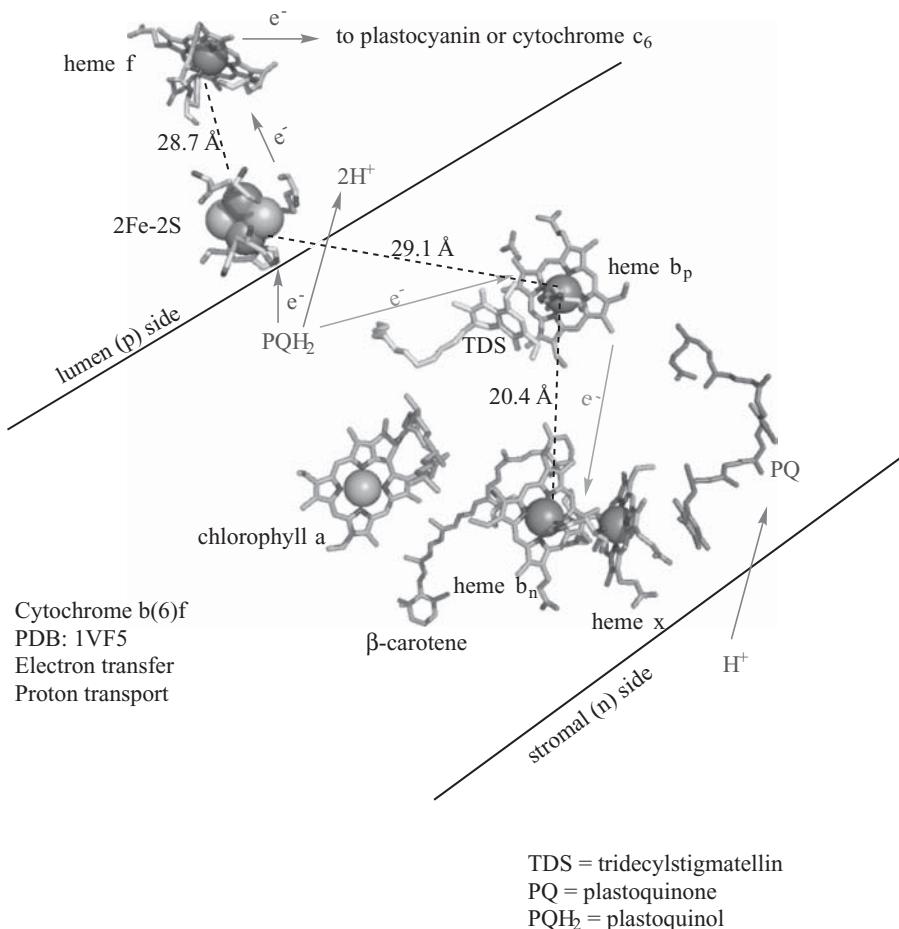
heme ci

heme x, covalent connection through one cysteine S atom (cys35) to protein chain in b(6)f, no aa residue axial ligands, water at 2.53 Å.

heme f, covalent connection through two cysteine S atoms (cys22, cys25) to protein chain in b(6)f, Fe coordination through his26 N<sub>ε2</sub> and backbone amide N of tyr1. Residue numbering from PDB: 1VF5.

**Figure 7.25** Structures of (A) heme b, (B) heme c and heme f, and (C) heme ci or heme x.

The following description of the electron transfer–proton transport scheme is illustrated in Figure 7.26. First, an electron is transferred from doubly reduced dihydroplastoquinone ( $\text{PQH}_2$ ) to a high potential electron transfer chain that consists of the Reiske iron–sulfur protein and the cytochrome protein containing heme f. Rappaport, Lavergne and co-workers have reported a midpoint potential at pH 7.0 of +355 mV for heme f.<sup>76</sup> These two centers reside on the electropositive (lumen or p) side of the membrane, exterior to the membrane. As a result, two protons are transferred to the aqueous lumen phase. A second electron is transferred from  $\text{PQH}_2$  sequentially to heme  $b_p$ ,



**Figure 7.26** Electron transfer and proton transport in cytochrome b(6)f: cofactors (PDB: 1VF5). Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft Corporation.) (See color plate)

heme  $b_n$  and heme x, hemes that are contained within the cytochrome b portion of the complex. Hemes  $b_p$ ,  $b_n$ , and x reside completely within the membrane and within the cytochrome  $b_6$  protein subunit. An uptake of protons from the negative (n or stromal) side of the membrane results in a proton electrochemical gradient across the membrane. Figure 2 of the reference 73 also illustrates the process well.

The cytochrome b(6)f electron transfer and proton translocation functions are similar to those performed by the respiratory cytochrome bc<sub>1</sub> complex described in Section 7.6. In fact, cytochrome f is functionally, but not structurally, analogous to the high-potential cytochrome c<sub>1</sub> in bc<sub>1</sub>; the two b hemes,  $b_p$ , and  $b_n$ , are analogous to the  $b_H$  and  $b_L$  hemes of bc<sub>1</sub>; and the Rieske ISP centers are similar in the two cytochromes. The cytochrome b(6)f PQH<sub>2</sub> → PQ reaction has its counterpart in the dihydroubiquinone → ubiquinone reaction of bc<sub>1</sub>. In both cases the lipophilic quinone species from the bulk bilayer phase move between reduction and proton uptake sites on the negative (n) side of the membrane to sites for oxidation and proton release on the positive (p) side. This cycling is known as the Q (quinone)-cycle mechanism. Overall the core of the b(6)f complex is similar to the analogous respiratory cytochrome bc<sub>1</sub> complex, but the domain arrangement outside the core as well as the arrangement of prosthetic groups (hemes and ISP) differ considerably.

### 7.5.2 Cytochrome b(6)f Metal Cofactor Specifics

Starting with the extrinsic portion, outside the membrane, of the cytochrome b(6)f complex, one first encounters heme f, which is a cytochrome c-type heme. (See Figure 7.25B.) It carries one his26 ligand to its iron center. In keeping with other cytochrome c types, it is attached to the cytochrome f protein via two thioether linkages—cys22 and cys25 in the PDB: 1VF5 X-ray crystallographic structure. This typical cytochrome c-type connector also provides an axial histidine ligand in a common amino acid pattern: CXXXH—that is, C<sub>22</sub>XXC<sub>25</sub>H<sub>26</sub>. Cytochrome f differs from c-type cytochromes in that it has an abnormal second axial ligand, the amide-N backbone nitrogen of the N-terminal tyrosine, tyr1. Next in the electron transfer chain, one finds the Rieske-type [2Fe–2S] cluster, approximately 29 Å distant from heme f toward the membrane lumen (p) side, outside but very close to the membrane surface. The 29-Å distance is believed to be too long for efficient electron transfer between cytochrome f and the [2Fe–2S] cluster. However, the entire iron-sulfur protein (ISP) is mobile, as will be discussed in more detail for cytochrome bc<sub>1</sub> below, so that electron transfer could take place under other conformations. The [2Fe–2S] cluster has bonding contacts with his110, his129, cys108, and cys126 (see Table 7.5). Hemes  $b_p$ ,  $b_n$ , and x reside in the cytochrome  $b_6$  protein within the membrane. Heme  $b_p$ , nearer the lumen (p) side, is held in place by his86 and his187, of the cytochrome  $b_6$  B and D helices. The iron ion in heme  $b_n$  is attached to the protein via the imidazole nitrogens of two histidine ligands, his100 and his202 from cytochrome  $b_6$ 's B and D helices. The

histidine ligations effectively crosslink the B and D helices. The unusually connected heme x (see Figure 7.25C) is covalently linked to the cytochrome b protein via a thioether bond to an invariant (among many cytochrome b(6)f species) cys35, a residue in helix A of cytochrome b<sub>n</sub>. Heme x differs in at least two ways from c-type cytochromes: (1) It features only one rather than the normal two thioether linkages to protein cysteines, and (2) it has no orthogonal histidine ligand to the heme's iron center but rather appears to be linked via a water molecule hydrogen-bonded to a propionate side chain of heme b<sub>n</sub> and to the backbone amide nitrogen of invariant gly38 within cytochrome b<sub>n</sub>. Heme x, having a vacant sixth coordination position, is believed to carry a ferric, high-spin iron ion that should yield an EPR signal. Of the four hemes, heme x is positioned nearest the stromal (n) side of the membrane in a nearly perpendicular orientation to b<sub>n</sub> above it. Selected bond distances are collected in Table 7.5.

**TABLE 7.5 Bond Distances in Cytochrome b(6)f Hemes for PDB: 1VF5**

| Heme                | Complex Subunit/Heme-Protein Connection  | Complex Subunit/Fe Ligands (in Addition to Porphyrin N Atoms)   | Bond Distance (Å)            |
|---------------------|--|---|------------------------------|
| Heme f              | Cyto f subunit/Covalent bonds between S <sub>γ</sub> of cys22 (S <sub>γ</sub> -C = 1.83 Å) and cys25 (S <sub>γ</sub> -C = 1.84 Å) to carbon atoms of the porphyrin's C=CH substituents | Cyto f subunit<br>his26 N <sub>ε2</sub><br>tyr1 backbone amide N  | 2.44<br>2.34                 |
| Rieske [2Fe-2S]     |  | Rieske iron-sulfur protein (RISP or ISP)<br>cys108 S <sub>γ</sub> -Fe <sub>1</sub><br>cys126 S <sub>γ</sub> -Fe <sub>1</sub><br>his110 N <sub>δ1</sub> -Fe <sub>2</sub><br>his129 N <sub>δ1</sub> -Fe <sub>2</sub>                        | 2.32<br>2.32<br>2.27<br>2.25 |
| Heme b <sub>p</sub> | No covalent connection from protein to porphyrin ligand  | Cyto b <sub>6</sub> subunit<br>his86 (helix B) N <sub>ε2</sub> -Fe<br>his187 (helix D) N <sub>ε2</sub> -Fe  | 2.08<br>2.29                 |
| Heme b <sub>n</sub> | No covalent connection from protein to porphyrin ligand  | Cyto b <sub>6</sub> subunit<br>his100 (helix B) N <sub>ε2</sub> -Fe<br>his202 (helix D) N <sub>ε2</sub> -Fe   | 2.51<br>2.30                 |
| Heme x              | Cyto b <sub>6</sub> subunit helix A<br>cys35 S <sub>γ</sub> bonded to carbon atom of the porphyrin's C=CH substituent (S <sub>γ</sub> -C = 1.78 Å)                                     | H <sub>2</sub> O-Fe<br>Fe axial H <sub>2</sub> O ligand<br>hydrogen-bonded to propionate side chain of heme b <sub>n</sub> (OH <sub>2</sub> ⋯O = 3.50 Å) and cyto b <sub>6</sub> gly38 backbone amide N-H (OH <sub>2</sub> ⋯N-H = 2.70 Å) | 2.53                         |

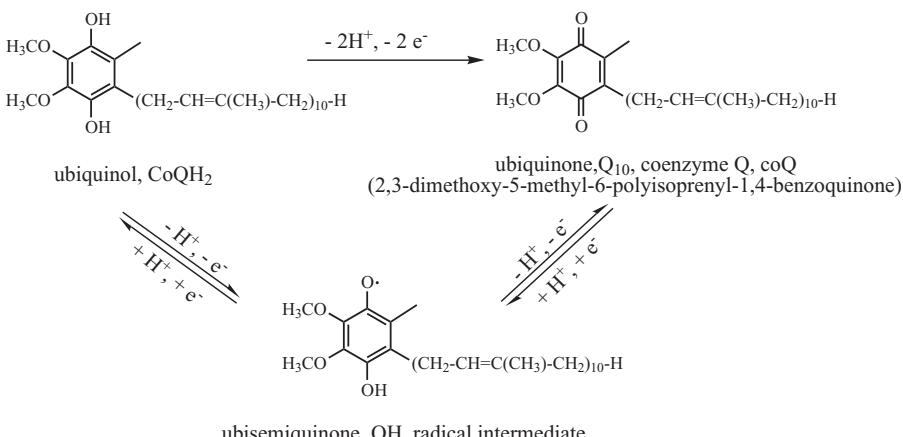
## 7.6 CYTOCHROME bc<sub>1</sub>: A BACTERIAL CYTOCHROME

### 7.6.1 Introduction

Cytochrome bc<sub>1</sub> is a multicomponent enzyme found in the inner mitochondrial membrane of eukaryotes and in the plasma membrane of bacteria.<sup>77</sup> The cytochrome bc<sub>1</sub> complex functions as the middle component of the mitochondrial respiratory chain, coupling electron transfer between ubiquinone/ubiquinol (see Figure 7.27) and cytochrome c.

Other terminology used in referring to cytochrome bc<sub>1</sub> includes complex III of the respiratory chain, or ubiquinol:cytochrome c oxidoreductase. Complex III is one of five essential enzymes of the mitochondrial inner membrane involved in the oxidative phosphorylation process. In this process, carried out in the cellular mitochondria, electrons are passed along through a series of molecules called the electron transport chain. Electrons are generated from NADH (reduced nicotinamide adenine dinucleotide) and ultimately transferred to molecular oxygen. The passage of electrons releases energy stored as a proton gradient across the mitochondrial membrane. The energy released is used by ATP synthase to make ATP (adenosine triphosphate) from ADP. Knowledge about the five respiratory enzyme complexes has been summarized by M. Saratse in a 1999 *Science* magazine publication.<sup>78</sup> Figure 1 of reference 78 is a helpful cartoon diagram illustrating the five enzyme complexes of this respiratory pathway: (1) Complex I, NADH-dehydrogenase (NADH-DH); (2) Complex II, succinate dehydrogenase (SDH); (3) Complex III, cytochrome bc<sub>1</sub> (bc1); (4) Complex IV, cytochrome oxidase (COX); and (5) Complex V, ATP synthase. Cytochrome c and cytochrome c oxidase will be described in Sections 7.7 and in 7.8, respectively.

In cytochrome bc<sub>1</sub>, ubiquinone, also known as coenzymeQ, coQ, or Q<sub>10</sub>, is a lipid-soluble compound that can move within the membrane. It is involved



**Figure 7.27** The ubiquinol/ubiquinone reaction.

as an electron transfer agent, and its reduced alcohol partner, ubiquinol, coQH<sub>2</sub>, is a source of protons for proton transport through respiratory chain components (see Figure 7.27). At the same time as electrons are transferred, a proton gradient is generated across the mitochondrial membrane through a reaction sequence such as



Two active sites within cytochrome bc<sub>1</sub> are required: (1) the Q<sub>o</sub> site where ubiquinol is oxidized and protons are released to the intermembrane space and (2) the Q<sub>i</sub> site where ubiquinone is reduced and protons are taken up from the matrix side of the membrane. The mechanism requires that electrons be transferred from the Q<sub>o</sub> site to the Q<sub>i</sub> site. The Q<sub>o</sub> site is located between the iron–sulfur protein (ISP) and cytochrome b subunits of the cytochrome bc<sub>1</sub> complex, closer to the intermembrane space. The Q<sub>i</sub> site is located wholly within the cytochrome b subunit, closer to the matrix side of the membrane. Both sites communicate with aqueous phases using channels within the protein complex. The ISP and cytochrome b subunits are described in detail below in Sections 7.6.2 and 7.6.3. The translocation of protons across the mitochondrial inner membrane from the matrix space (negative or n side) to the intermembrane space (positive or p side) contributes to the electrochemical proton gradient that drives adenosine triphosphate (ATP) synthesis.<sup>79</sup> The coupling of transmembrane electrochemical proton gradients to ATP synthesis was first proposed by Mitchell in his chemiosmotic hypothesis.<sup>80</sup> More recently, cytochrome bc<sub>1</sub>’s electron transport coupled to proton translocation mechanism has been cited as a version of Mitchell’s “proton-motive Q cycle.”<sup>81</sup> During each iteration of the Q cycle, two electrons are transferred for every four protons translocated. In addition to its respiratory function, multifunctional cytochrome bc<sub>1</sub> is also involved in mitochondrial import protein processing. This major mitochondrial processing activity removes presequences from nuclear encoded precursor proteins.<sup>82</sup>

### 7.6.2 Cytochrome bc<sub>1</sub> Structure

Bovine heart cytochrome bc<sub>1</sub> (PDB: 1BE3 and PDB: 1BGY) as studied by Iwata et al.<sup>83</sup> exists as a dimer in the asymmetric unit cell. Each monomer consists of 11 different polypeptide subunits (SU) with a total of ~2165 amino acid residues and a molecular mass of ~240 kDa. The protein subunits of the complex occupy three separate regions: (1) the intermembrane space (p side) occupied by cytochrome c<sub>1</sub> (subunit 4, SU4), the iron–sulfur protein (ISP, SU5) and subunit 8; (2) the transmembrane region occupied by cytochrome b (SU3), the transmembrane helices of cytochrome c<sub>1</sub> and the ISP, and subunits 7, 10, and 11; and (3) the matrix space (n side) occupied by two large core proteins (subunits 1 and 2) as well as subunits 6 and 9. Subunit 8 is often called the “hinge protein” and is thought to be essential for proper complex formation between cytochrome c (the exit point for some bc<sub>1</sub> complex electrons) and

cytochrome  $c_1$ . On the intermembrane side, cytochrome  $c_1$ , the ISP, and subunit 8 sit on top of cytochrome b. Subunits 10 and 11 maintain contact between cytochrome  $c_1$  and the ISP and may assist in forming the  $bc_1$  complex. Subunit 9 is the mitochondrial targeting presequence of the ISP. This means that in eukaryotes the ISP is nuclear-coded and posttranslationally imported into the mitochondria and then inserted into the inner membrane. In PDB: 1BE3 and PDB: 1BGY, subunit 9 is found between the core 1 and core 2 subunits. These large subunits may be MPPs (mitochondrial processing peptidases), proteins that cleave targeting presequences after they are imported into the mitochondria. In the dimer, as with cytochrome b(6)f, the ISP from one monomer interacts with cytochromes b and  $c_1$  of the other monomer. Electron transfer can only take place when the dimer is assembled; otherwise the electron-carrying hemes and [2Fe–2S] centers are too far apart.

By the end of 1998, five X-ray crystallographic structures of the cytochrome  $bc_1$  complex had been published: (1) PDB: 1QCR, reference 79; (2) PDB: 1BE3, 1BGY, reference 83; and (3) PDB: 1BCC, 3BCC, reference 84. These snapshots along cytochrome  $bc_1$ 's dynamic pathway of electron transfer and proton translocation gave researchers insight not only into the enzyme's structure but also into its mechanism of activity. One of the important findings was the changing position of the iron–sulfur protein (ISP) with its [2Fe–2S] center—it moves as much as 20 Å and rotates as much as 60°. As the data to be presented will show, the ISP appears to change from positions close to cytochrome  $c_1$  to positions close to heme  $b_L$  within cytochrome b while the positions of heme  $b_L$  and  $b_H$  remain fixed with respect to each other and the membrane bilayer. As stated in Section 7.6.1, these structures also indicated that dimer formation was essential for electron transfer and proton translocation to take place in cytochrome  $bc_1$  complex. The enzyme complex contains a cavity between the two monomers of the essential dimer that were thought to allow passage of substrate ubiquinol/ubiquinone to and from the complex and between two active sites for oxidation and/or reduction,  $Q_N$  ( $Q_o$  in most publications) and  $Q_P$  ( $Q_i$  in most publications). An informative cartoon illustrating the complete dimer is found in the perspective by J. L. Smith accompanying the *Science* magazine publication, reference 83.<sup>85</sup>

In 1998, B. K. Jap, S. Iwata, and co-workers solved the complete structure of the 11-subunit mitochondrial cytochrome  $bc_1$  complex using X-ray crystallography of two different crystalline forms, both of which were obtained from oxidized bovine heart mitochondrial  $bc_1$  complex (PDB: 1BE3 equals the P6<sub>5</sub>22 form and 1BGY equals the P6<sub>5</sub> form).<sup>83</sup> The asymmetric unit of the P6<sub>5</sub>22 form, refracting to a resolution of 3.0 Å, contains one monomer of 11 subunits related to the other monomer of the dimer by a crystallographic twofold symmetry axis. The P6<sub>5</sub> form (PDB: 1BGY), refracting to 2.8-Å resolution, contains the whole dimer. The P6<sub>5</sub>22 form's *R* factor (28.5%) is high compared to ordinary water-soluble proteins due to partially ordered co-crystallized lipids not accounted for in the model. In the two different crystal forms, the ISP domains are found in different positions as has been reported for other bovine

heart bc<sub>1</sub> complexes<sup>79</sup> (PDB: 1QCR) and for chicken heart cytochrome bc<sub>1</sub><sup>84</sup> (PDB: 1BCC, 3BCC). In the P6<sub>5</sub>22 form (PDB: 1BE3), the average distances are [2Fe–2S]–heme c<sub>1</sub> = 15.5 Å and [2Fe–2S]–heme b<sub>L</sub> = 35.5 Å, while in the P6<sub>5</sub> structure (PDB: 1BGY) the average distances are 27.5 and 31.0 Å, respectively. In addition, there is a hydrogen bond between the his161 ligand of the ISP and a propionate of heme c<sub>1</sub> in the PDB: 1BE3 crystal. The authors postulate that the 15.5-Å distance and hydrogen bonding facilitate fast electron transfer between the [2Fe–2S] cluster and cytochrome c<sub>1</sub>.

A second research group postulating the movement of the [2Fe–2S] domain is that of Edward A. Berry.<sup>84</sup> This group's X-ray crystallographic structures of chicken heart cytochrome bc<sub>1</sub> as the native enzyme (PDB: 1BCC) and the enzyme in the presence of co-crystallized inhibitors stigmatellin and antimycin A (PDB: 3BCC) showed decidedly different positions for the Rieske iron–sulfur protein (ISP). The native enzyme, PDB: 1BCC, exhibits a shorter distance between the [2Fe–2S] center and heme c<sub>1</sub> than that for the enzyme with co-crystallized inhibitors, PDB: 3BCC. Conversely, the distance between the [2Fe–2S] center and heme b<sub>L</sub> is longer for the native enzyme than that for PDB: 3BCC. The distance data are collected in Table 7.6. Figure 1 of reference 84 illustrates the change graphically.

### 7.6.3 Cytochrome bc<sub>1</sub> Metal Cofactor Specifics

Every cytochrome bc<sub>1</sub> complex contains three subunits with active redox centers: (1) cytochrome c<sub>1</sub>; (2) cytochrome b; and (3) the “Rieske” [2Fe–2S] iron–sulfur protein (ISP). Most researchers studying the structure and function of the complex believe that the ISP is a mobile domain and moves about a pivot point during the catalytic cycle, possibly to facilitate electron transfer. The electron-carrying moieties of cytochrome bc<sub>1</sub> involve the following: (1) the Fe(II)–Fe(III) centers, in order through the bc<sub>1</sub> complex from the inter-membrane space (p side), of (a) heme c<sub>1</sub> in cytochrome c<sub>1</sub>, (b) the high potential [2Fe–2S] center of the Rieske iron–sulfur protein (ISP) of the opposite monomer, and (c) the b<sub>L</sub> (also called b<sub>565</sub>) and the b<sub>H</sub> (also called b<sub>562</sub>) hemes of cytochrome b; and (2) ubiquinone/ubiquinol. Cytochrome c<sub>1</sub>, containing heme c<sub>1</sub>, is an all α-helix protein resembling other members of the Type I cytochrome c family. In PDB: 1BE3, the helices are named: α1, residues 23–38; α1', 57–73; α1'', 97–106; α2, 114–119; α3, 123–132; α5, 178–195; α6, 197–238. These correspond to the numbering system used for cytochrome c; however, the α4 cytochrome c helix appears to be replaced by a distorted double-stranded β-sheet. This and other insertions and deletions are unique to cytochrome c<sub>1</sub> and are believed to relate to its specific functions. One of these functions is to provide a binding surface for cytochrome c to which it transfers electrons out of the cytochrome bc<sub>1</sub> complex. Helix α1 contains at least eight acidic aa residues that help form a large negative docking area for the highly basic cytochrome c. Using the numbering system found for residues in the bovine enzyme (PDB: 1BE3, for instance), heme c<sub>1</sub>'s iron ion ligands are his41

TABLE 7.6 Metal Cofactor Distances in cytochrome bc<sub>1</sub>; [2Fe-2S] Position and Mobility

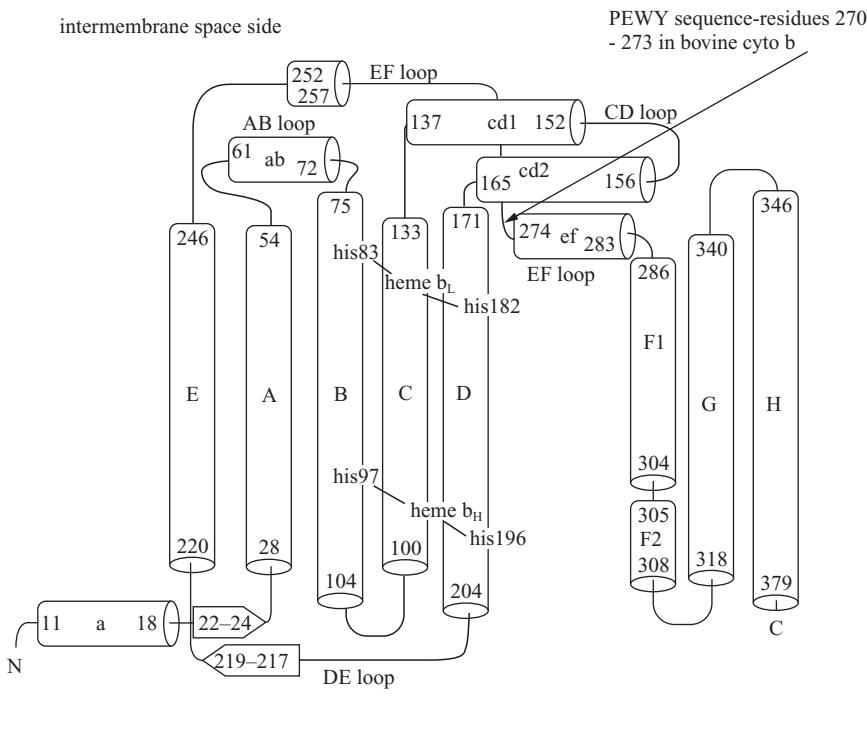
| PDB No.<br>(reference, date) | Species<br>(Crystal Space<br>Group, resolution, Å)                        | Inhibitor or Substrate<br>(PDB abbreviation,<br>binding site)          | [2Fe-2S]<br>(Fe <sub>1</sub> and/or<br>Fe <sub>2</sub> )—heme<br>b <sub>L</sub> (Å) | [2Fe-2S]<br>(Fe <sub>1</sub> and/or<br>Fe <sub>2</sub> )—heme<br>b <sub>H</sub><br>(Å) | heme c <sub>1</sub> —<br>[2Fe-2S]<br>(Fe <sub>1</sub> and/or<br>Fe <sub>2</sub> ) b <sub>L</sub> (Å) | heme b <sub>L</sub> —<br>[2Fe-2S]<br>Position <sup>a</sup> and<br>Mobility <sup>b</sup> |
|------------------------------|---|--|---|--|--|---|
| 1QCR<br>(79, 1997)           | <i>bos Taurus</i> , bovine<br>(14 <sub>1</sub> 22, 2.70)                  | None   | 31  | 27   | 21   | “b” <sup>a</sup>  |
| 1BE3<br>(83, 1998)           | <i>bos Taurus</i><br>(P6 <sub>1</sub> 22, 3.00)                           | None   | 14.9 (Fe <sub>2</sub> )<br>17.1 (Fe <sub>1</sub> )                                  | 35.1 (Fe <sub>2</sub> )<br>35.7 (Fe <sub>1</sub> )                                     | 20.3   | “C <sub>1</sub> ” <sup>a</sup>  |
| 1BCC<br>(84, 1998)           | <i>Gallus gallus</i> , chicken<br>(P2 <sub>1</sub> 2 <sub>1</sub> , 3.16) | None   | 21.3  | 34.3   | “C <sub>1</sub> ” <sup>a</sup>   | “int” <sup>a</sup>  |
| 1BGY<br>(83, 1998)           | <i>bos Taurus</i><br>(P6 <sub>1</sub> , 3.00)                             | None   | 26.1 (Fe <sub>2</sub> )<br>27.0 (Fe <sub>1</sub> )                                  | 31.0 (Fe <sub>2</sub> )<br>31.5 (Fe <sub>1</sub> )                                     | 20.6   | mobile <sup>b</sup>   |
| 1NTM<br>(86, 2004)           | <i>bos Taurus</i><br>(14 <sub>1</sub> 22, 2.40)                           | None   | 32.6 (Fe <sub>2</sub> )<br>31.2 (Fe <sub>1</sub> )                                  | 26.9 (Fe <sub>2</sub> )<br>28.0 (Fe <sub>1</sub> )                                     | 20.6   | mobile <sup>b</sup>   |
| 1NTZ<br>(86, 2004)           | <i>bos Taurus</i><br>(14 <sub>1</sub> 22, 2.60)                           | substrate<br>ubiquinone-2<br>(UQ2, Q <sub>o</sub> and Q <sub>i</sub> ) | 32.5 (Fe <sub>2</sub> )<br>31.0 (Fe <sub>1</sub> )                                  | 26.7 (Fe <sub>2</sub> )<br>28.0 (Fe <sub>1</sub> )                                     | 20.6   | mobile <sup>b</sup>   |
| 1SQP<br>(88, 2004)           | <i>bos Taurus</i><br>(14 <sub>1</sub> 22, 2.70)                           | inhibitor myxothiazol<br>(MYX, Q <sub>o</sub> )                        | 31.0 (Fe <sub>2</sub> )<br>30.4 (Fe <sub>1</sub> )                                  | 29.0 (Fe <sub>2</sub> )<br>29.8 (Fe <sub>1</sub> )                                     | 20.6   | mobile <sup>b</sup>   |

|                    |  |  |  |  |      |   |
|--------------------|--|--|--|--|------|---|
| 1SQX<br>(88, 2004) | <i>bos Taurus</i><br>(I4 <sub>1</sub> 22, 2.60)                  | inhibitor stigmatellin A<br>(SMA, Q <sub>o</sub> ) and substrate<br>ubiquinone-2 (UQ2, Q <sub>i</sub> )<br>inhibitor | 32.3 (Fe <sub>2</sub> )<br>30.8 (Fe <sub>1</sub> ) | 27.4 (Fe <sub>2</sub> )<br>26.1 (Fe <sub>1</sub> ) | 20.6 | “b” <sup>a</sup>  |
| 1PP9<br>(90, 2005) | <i>bos Taurus</i><br>(P2 <sub>1</sub> 2 <sub>21</sub> , 2.10)    | Stigmatellin (SMA, Q <sub>o</sub> )<br>and substrate   | 32.0 (Fe <sub>2</sub> )<br>30.6 (Fe <sub>1</sub> ) | 27.2 (Fe <sub>2</sub> )<br>26.0 (Fe <sub>1</sub> ) | 20.6 | fixed <sup>b</sup> (SMA<br>induced)<br>“b” <sup>a</sup> |
| 1NTK<br>(86, 2004) | <i>bos Taurus</i><br>(I4 <sub>1</sub> 22, 2.60)                  | ubiquinone-10 (UQ, Q <sub>i</sub> )<br>inhibitor   | 32.8 (Fe <sub>2</sub> )<br>31.0 (Fe <sub>1</sub> ) | 26.8 (Fe <sub>2</sub> )<br>27.7 (Fe <sub>1</sub> ) | 20.6 | fixed <sup>b</sup> (SMA<br>induced)<br>“b” <sup>a</sup> |
| 3BCC<br>(84, 1998) | <i>Gallus gallus</i><br>(P2 <sub>1</sub> 2 <sub>21</sub> , 3.70) | Inhibitors   | 31.6   | 26.4   |      | “b” <sup>a</sup>  |
| 1PPJ<br>(90, 2005) | <i>bos Taurus</i><br>(P2 <sub>1</sub> 2 <sub>21</sub> , 2.10)    | Stigmatellin (SIG, Q <sub>o</sub> )<br>and antimycin A<br>(AMY, Q <sub>i</sub> )<br>Inhibitors                       | 32.0 (Fe <sub>2</sub> )<br>30.6 (Fe <sub>1</sub> ) | 27.2 (Fe <sub>2</sub> )<br>26.0 (Fe <sub>1</sub> ) | 20.6 | “b” <sup>a</sup>  |

<sup>a</sup>Position assessed by method of Iwata, et al., reference 83.<sup>b</sup>Mobility assessed by comparisons of anomalous Fe peak heights: data collected at or near the iron K absorption edge. (See references 84 and 88.)

( $N_{\epsilon 2}$ -Fe = 2.03 Å) and met160 ( $S_{\delta}$ -Fe = 2.37 Å). The heme is covalently anchored to the protein through aa residues cys37 and cys40. The cysteine residues occur in a typical cytochrome c CXXCH pattern—in this case cys37-ser38-ser39-cys40-his41. The propionate groups of heme  $c_1$  participate in salt bridges with arg120 of the cytochrome  $c_1$  protein and with his161 of the Rieske iron-sulfur protein (of the opposite monomer). The [2Fe-2S] center of the ISP is held in place by bonds to four ISP aa residues: his141 ( $Fe_2-N_{\delta 1}$  = 2.18 Å), his161 ( $Fe_2-N_{\delta 1}$  = 2.17 Å), cys139 ( $Fe_1-S_{\gamma}$  = 2.21 Å), and cys158 ( $Fe_1-S_{\gamma}$  = 2.25 Å).

The cytochrome b protein contains two hemes, heme  $b_L$  (L = low potential, also called  $b_{565}$  in the literature) and heme  $b_H$  (H = high potential, also called  $b_{562}$  in the literature). Cytochrome b is composed of eight transmembrane helices (A to H). There are seven extra-membrane loops connecting the transmembrane helices. Of these, loops AB (with short helix ab), CD (featuring short helices cd1 and cd2 and a hairpin turn) and EF (with short helix ef) pass through the intermembrane space and are important participants in forming the  $Q_o$  cavity. This cavity extends into the membrane from the intermembrane space, and it sits between the [2Fe-2S] center of the ISP and heme  $b_L$ . The long DE loop is on the matrix side of cytochrome b, helping to form the boundary of the  $Q_i$  pocket that is located near the matrix extending upward into the membrane. (See Figure 7.28 and Figure 1 of reference 86.)



matrix side

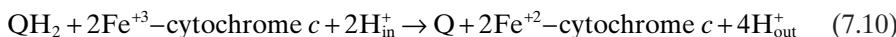
**Figure 7.28** Secondary structure of the cytochrome bc<sub>1</sub> complex.

Both the N- and C-termini are located in the matrix. The two hemes are bound within the four helix bundle of helices A-B-C-D. The iron ion in heme b<sub>L</sub> carries axial his83 (helix B, Fe–N<sub>e2</sub> = 2.08 Å) and his182 (helix D, Fe–N<sub>e2</sub> = 2.06 Å) ligands, while the iron ion in heme b<sub>H</sub> carries his97 (helix B, Fe–N<sub>e2</sub> = 2.07 Å) and his196 (helix D, Fe–N<sub>e2</sub> = 2.05 Å). These histidines are invariant in sequences of cytochrome b as well as in cytochrome b<sub>6</sub> as described for cytochrome b(6)f above.

The Iwata group<sup>83</sup> compared results for their bovine cytochrome bc<sub>1</sub> complex structures (PDB: 1BE3 and 1BGY) with those completed previously for bovine bc<sub>1</sub> (PDB: 1QCR)<sup>79</sup> and for chicken heart bc<sub>1</sub> (PDB: 1BCC and 3BCC).<sup>84</sup> The comparisons led the researchers to propose three positional states for the iron–sulfur protein (ISP): (1) The “c<sub>1</sub>” state in which the ISP is closer to heme c<sub>1</sub> and further away from heme b<sub>L</sub>; (2) the “int” state in which the ISP is in an intermediate position with respect to heme c<sub>1</sub> and heme b<sub>L</sub>; and (3) the “b” state in which the ISP is closer to heme b<sub>L</sub> and further away from heme c<sub>1</sub>. In Table 7.6, X-ray crystallographic structures are classified according to their ISP position or “state.” However, see the qualifying statements in conclusions at the end of Section 7.6.4.

#### 7.6.4 The Cytochrome bc<sub>1</sub> Q Cycle

The proton-motive Q-cycle model, put forward by Mitchell (references 80 and 81) and by Trumper and co-workers,<sup>77</sup> is invoked in the following manner: (1) One electron is transferred from ubiquinol (ubiquinol oxidized to ubisemiquinone; see Figure 7.27) to the Rieske [2Fe–2S] center at the Q<sub>o</sub> site, the site nearest the intermembrane space or p side; (2) this electron can leave the bc<sub>1</sub> complex via an attached cytochrome c or be transferred to cytochrome c<sub>1</sub>; (3) the reactive ubisemiquinone reduces the low-potential heme b<sub>L</sub> located closer to the membrane’s intermembrane (p) side; (4) reduced heme b<sub>L</sub> quickly transfers an electron to high-potential heme b<sub>H</sub> near the membrane’s matrix side; and (5) ubiquinone or ubisemiquinone oxidizes the reduced b<sub>H</sub> at the Q<sub>i</sub> site nearest the matrix or n side. Proton translocation results from the deprotonation of ubiquinol at the Q<sub>o</sub> site and protonation of ubisemiquinone at the Q<sub>i</sub> site. Ubiquinol generated at the Q<sub>i</sub> site is reoxidized at the Q<sub>o</sub> site (see Figure 7.27). Additional protons are transported across the membrane from the matrix (see Figure 7.26 illustrating a similar process for cytochrome b(6)f). The overall reaction can be written



It should be noted that generation of the radical ubisemiquinone or any semiquinone (QH<sup>•</sup>) can have dangerous physiological consequences, such as producing the tissue-damaging superoxide ion, O<sub>2</sub><sup>−</sup>. A. R. Crofts has written a review recently that expands upon known cytochrome bc<sub>1</sub> structure–function relationships.<sup>87</sup> The article reviews knowledge about genetic defects that lead to mutations in respiratory chain proteins including bc<sub>1</sub> complex. These muta-

tions may result in mitochondrial myopathies, causing neuromuscular disease symptoms. In the cytochrome b subunit of bc<sub>1</sub> complex, these myopathies enhance superoxide production at the ubihydroquinone (ubiquinol) oxidation site, thus causing harmful free-radical damage. The bypass reaction producing superoxide appears to be an unavoidable part of the cytochrome bc<sub>1</sub> reaction mechanism and results in cellular aging. Consequently, researchers in the field believe that an understanding of the mechanism of the bc<sub>1</sub> complex may be central to our understanding of the aging process.

Another explanation of the Q-cycle mechanism is given by Iwata et al.<sup>83</sup> The central reaction of the Q cycle is a bifurcation that takes place at the Q<sub>p</sub> site (called Q<sub>o</sub> by other researchers) that is in contact with the positive (intermembrane) side of the membrane. Here, during the hydroquinone (QH<sub>2</sub>) oxidation reaction, one electron is transferred via the high potential [2Fe–2S] center of the ISP and cytochrome c<sub>1</sub> and thus to cytochrome c, while the second electron is transferred to heme b<sub>L</sub>. From there, the second electron is transferred through the membrane dielectric to heme b<sub>H</sub>. Heme b<sub>H</sub> is part of the quinone reduction site Q<sub>N</sub> (called Q<sub>i</sub> by other researchers) that is in contact with the negative (matrix) side of the membrane. The Iwata group proposed a “three-state” model to explain the electron transfer–proton translocation: (1) With fully oxidized complex, before hydroquinone is bound, the ISP is in an intermediate position (Fe<sup>II/III</sup> ions of [2Fe–2S] ~27 Å from cytochrome c<sub>1</sub> and ~31 Å from heme b<sub>L</sub> PDB: 1BGY; Table 7.6, Iwata “Int” state); (2) hydroquinone binds in the Q<sub>p</sub> (Q<sub>o</sub>) site, in a cavity formed near the end of helix C, loop/helix cd1 and loop ef of cytochrome b (see Figure 7.28); (3) hydroquinone is deprotonated to semiquinone status (QH<sup>•</sup>) at this site (the activation barrier of hydroquinone oxidation), causing the ISP to move away to a position in which the Fe<sup>II/III</sup> ions of [2Fe–2S] are ~31–32 Å from cytochrome c<sub>1</sub> and ~26–27 Å from heme b<sub>L</sub> (PDB: 1QCR,<sup>79</sup> 3BCC<sup>84</sup>; Table 7.6, the Iwata “b” state); (4) the semiquinone is tightly bound to the reduced [2Fe–2S] center; a second electron is transferred from the semiquinone to heme b<sub>L</sub>; (5) the interaction between the resulting quinone and the [2Fe–2S] center lessens, and the reduced iron–sulfur center moves back toward heme c<sub>1</sub> for electron transfer in that direction (Fe<sup>II/III</sup> ions of [2Fe–2S] ~15–21 Å from cytochrome c<sub>1</sub> and ~34–35.5 Å from heme b<sub>L</sub> (PDB: 1BE3,<sup>83</sup> 1BCC<sup>84</sup>; Table 7.6, Iwata “c<sub>1</sub>” state); and (6) electrons are transferred from cytochrome c<sub>1</sub> to cytochrome c and from heme b<sub>L</sub> to heme b<sub>H</sub> and the system returns to step (1) of the cycle. The proposed cycle is predicated on the facile movement of the iron–sulfur center in the iron–sulfur protein (ISP) and is supported by X-ray crystallographic structures showing the ISP in differing locations. However, it is important to remember that the X-ray crystallographic studies represent a solid frozen in space and not the dynamic, in solution, physiological state found in nature.

Another method used by researchers to indicate the mobility of the ISP protein is to measure the anomalous peak heights of iron atoms belonging to the various hemes and the [2Fe–2S] cluster of the cytochrome bc<sub>1</sub> complex. Considering the heme b<sub>H</sub> iron ion as the most stable position and giving it a

value of 1.0, iron atoms in the other bc<sub>1</sub> subunits can be compared. For instance, the presence of the inhibitor stigmatellin A in crystals increases the stability (decreases the mobility) of the [2Fe–2S] center in the ISP, whereas the inhibitor myxothiazol decreases the stability (increases the mobility of the ISP).<sup>88</sup> See Section 7.6.5 for a discussion of cytochrome bc<sub>1</sub> inhibitors. Other experiments are needed to confirm the movements of protein subunits in the cytochrome bc<sub>1</sub> complex, especially the [2Fe–2S] center, before the catalytic mechanism described here is confirmed.

A third, clearer explanation of the electron transfer, proton translocation cycle is given by Saratse.<sup>78</sup> Each ubiquinol (QH<sub>2</sub>) molecule can donate two electrons. A first QH<sub>2</sub> electron is transferred along a high-potential chain to the [2Fe–2S] center of the ISP and then to cytochrome c<sub>1</sub>. From the cytochrome c<sub>1</sub> site, the electron is delivered to the attached, soluble cytochrome c in the intermembrane space. A second QH<sub>2</sub> electron is transferred to the Q<sub>i</sub> site via the cytochrome b hemes, b<sub>L</sub> and b<sub>H</sub>. This is an electrogenic step driven by the potential difference between the two b hemes. This step creates part of the proton-motive force. After two QH<sub>2</sub> molecules are oxidized at the Q<sub>o</sub> site, two electrons have been transferred to the Q<sub>i</sub> site (where one ubiquinone (Q<sub>10</sub>) can now be reduced, requiring two protons to be translocated from the matrix space). The net effect is a translocation of two protons for each electron transferred to cytochrome c. Each explanation of the cytochrome bc<sub>1</sub> Q cycle has its merits and its proponents. The reader should consult the literature for updates in this ongoing research area.

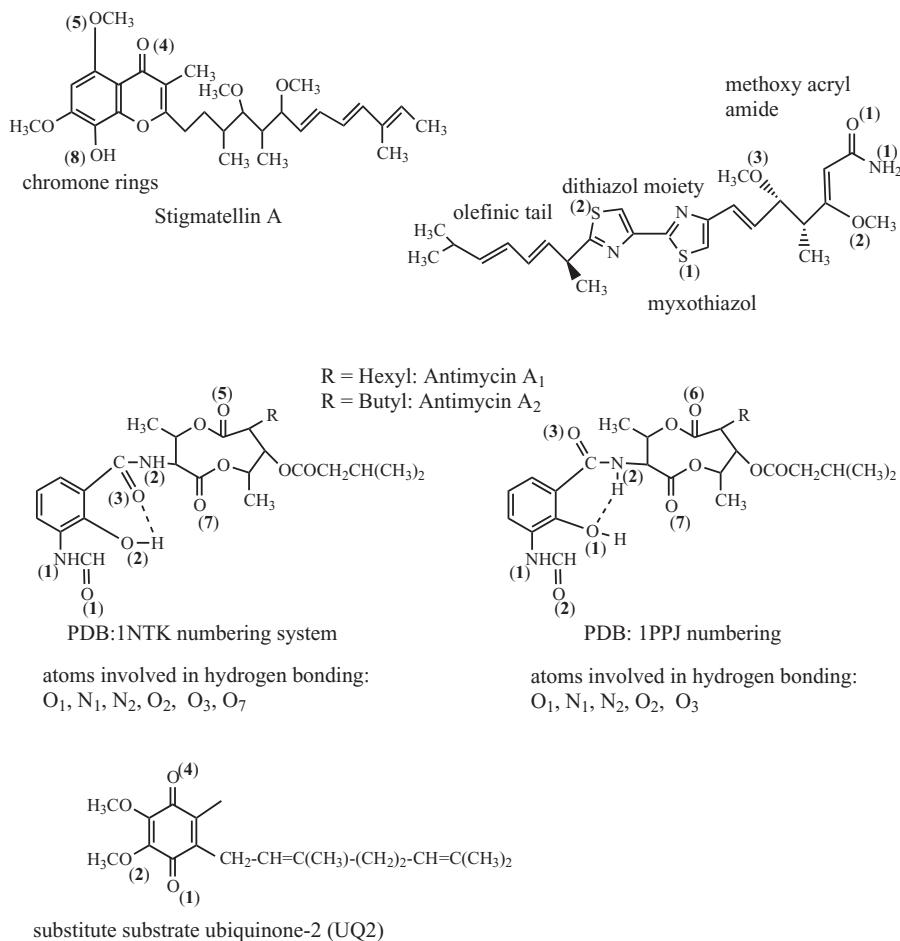
### 7.6.5 Cytochrome bc<sub>1</sub> Inhibitors

Evidence in support of the Q cycle mechanism comes in many forms, including oxidant-induced reduction of cytochrome b, existence of the antibiotic antimycin-sensitive and -insensitive ubisemiquinone radicals, ejection of two protons per electron transferred in bc<sub>1</sub> complexes, and binding of specific inhibitors to sites Q<sub>i</sub> and Q<sub>o</sub>. We will concentrate on the effect of inhibitors in this discussion. The understanding of inhibitor binding to cytochrome bc<sub>1</sub> complex is important to researchers who want to design new antifungal agents as crop protection agents or wish to develop new disease-specific inhibitors for use in animal or human populations. As part of the process, researchers study the enzyme's structure–function relationships or structure–activity relationships (SARs) as a means of understanding its mechanism of activity. The process's end game hopefully leads to design of new and better reagents for disease treatment and cure.

Only a few of the many known cytochrome bc<sub>1</sub> complex inhibitors will be mentioned in this chapter. Cytochrome bc<sub>1</sub> inhibitors that bind at the Q<sub>o</sub> site (p side, near the intermembrane space) include myxothiazol (MYX) and stigmatellin (SMA) (see Figure 7.29). MYX and SMA inhibit electron transfer from ubiquinol (QH<sub>2</sub>) to cytochrome c (outside the bc<sub>1</sub> complex) or to other domains of the bc<sub>1</sub> complex. Antimycin A (see Figure 7.29), a fungicide, binds

to the Q<sub>i</sub> site (n side, near the matrix space) and inhibits the transfer of electrons from heme b<sub>H</sub> to ubiquinone, Q.

Before describing inhibitors, we will discuss one X-ray crystallographic structure that features the substrate substitute ubiquinone-2, UQ2, in both the Q<sub>o</sub> and Q<sub>i</sub> sites (PDB: 1NTZ).<sup>86</sup> Ubiquinone-2, UQ2 (see Figure 7.29), differs from UQ10 (see Figure 7.27) in having two isoprenoid units in its aliphatic tail rather than 10 in UQ10. In the Q<sub>o</sub> pocket, UQ2 is not within hydrogen-bonding distance (<4 Å) of the aa residues normally associated with inhibitor binding—that is, the his161 ligand of the [2Fe–2S] center, glu271 backbone N, or tyr273



(1) numbers in bold in parentheses indicate  
numbering system for the atoms in the molecule

**Figure 7.29** Cytochrome bc<sub>1</sub> inhibitor structures.

hydroxyl groups. The closest approach of a UQ2 atom is that of its methoxy oxygen O<sub>2</sub> to the glu271 backbone N at a distance of 4.1 Å. An intervening water molecule does bring these two atoms within hydrogen-bonding distance (UQ2 O<sub>2</sub>···H<sub>2</sub>O<sub>1069</sub> = 2.8 Å, H<sub>2</sub>O<sub>1069</sub>···E271 N = 3.1 Å in PDB: 1NTZ). The N<sub>ε2</sub> atom of his161 is 4.6 Å distant from the UQ2 carboxyl oxygen O<sub>4</sub>. The isoprenoid tail of UQ2 has hydrophobic contacts with phe274 in cytochrome b's EF loop (short helix ef) as well as met124, ala125, and met129 in the C helix.

In the Q<sub>i</sub> pocket for the PDB: 1NTZ structure, ubiquinone-2 binds in a different manner than the inhibitor antimycin A<sub>1</sub> (see discussion of antimycin A<sub>1</sub> inhibitors below). Different Q<sub>i</sub> amino acid residues interact with the substrate substitute UQ2: (1) leu 200, his201, ser205, tyr224, lys227, and asp228 are important in ubiquinone-2 binding, and (2) trp31, gly38, met190, met194, leu197, and ser35 provide specific interactions with antimycin A<sub>1</sub>. In the Q<sub>i</sub> pocket, hydrogen bonds are formed between a UQ2 carbonyl oxygen and an asp228 side-chain oxygen (UQ2 O<sub>1</sub>···asp228 O<sub>δ2</sub> = 2.5 Å), a second UQ2 carbonyl oxygen and a hydroxyl oxygen of ser205 (UQ2 O<sub>4</sub>···ser205 O<sub>γ</sub> = 4.0 Å), a UQ2 methoxy (OCH<sub>3</sub>) oxygen with the hydroxyl oxygen of ser205 (UQ2 O<sub>3</sub>···ser205 O<sub>γ</sub> = 2.7 Å), and an interaction between the N<sub>ε2</sub> nitrogen and a UQ2 carbonyl group mediated by a water molecule (UQ2 O<sub>4</sub>···H<sub>2</sub>O = 3.4 Å, H<sub>2</sub>O···his201 N<sub>ε2</sub> = 2.6 Å). In addition, an edge-to-face phe220–ubiquinone-2 quinone ring Ar–Ar interaction, similar to that for antimycin A<sub>1</sub>, is evident. The closest interaction distance between ubiquinone-2 in the Q<sub>i</sub> site and the heme b<sub>H</sub> is 4.2 Å.

Effects of the myxothiazol inhibitor on the cytochrome bc<sub>1</sub> complex were studied by the Trumper group beginning in 1984.<sup>89</sup> The researchers found effects identical to those caused by removal of the iron–sulfur protein (ISP) from the complex. Reduction of heme c<sub>1</sub> is blocked but not the reduction of the b hemes, b<sub>L</sub> and b<sub>H</sub>. If the antibiotic antimycin A was present, reduction of all cytochromes hemes was found. Myxothiazol also inhibits reduction of the ISP by ubiquinol. These and other effects were found to be consistent with the proton-motive Q-cycle pathway of electron transfer in which the myxothiazol inhibitor binds to cytochrome b and displaces ubiquinol/ubiquinone from the ISP and Q<sub>o</sub> site. The reference 89 authors postulated that (1) a myxothiazol-induced conformational change in cytochrome b is transmitted to the ubiquinone-binding site on the ISP, or (2) there is a ubiquinone-binding site consisting of peptide domains from both cytochrome b and the ISP. Recent X-ray crystallographic studies have shown postulate (2) to be correct, while postulate (1) structural proofs remain elusive.

The Di Xia research group has completed X-ray crystallographic studies of cytochrome bc<sub>1</sub> with a number of different inhibitors<sup>88</sup> (PDB: 1SQB, 1SQP, 1SQQ, 1SQV, 1SQX). Crystals suitable for crystallography were obtained by co-crystallization of the inhibitor with the native enzyme. Data for two of the inhibitor-modified cytochrome bc<sub>1</sub> complex (PDB: 1SQP and 1SQX) are included in Table 7.6 and Table 7.7. One inhibitor–enzyme complex studied by

**TABLE 7.7 Selected Substrate/Inhibitor Contacts at Q<sub>o</sub> and Q<sub>i</sub> Sites in Cytochrome bc<sub>1</sub>**

| PDB No.<br>(Substrate/<br>Inhibitor)         | Ar-Ar or<br>hydrophobic<br>contacts            | his61 (H161)<br>(from ISP,[2Fe-2S]<br>ligand) (Å)  | glu271 (E271) (from<br>cyt b, EF loop) (Å)  | Q <sub>o</sub> Site<br>tyr273 (Y273)<br>(from cyt b,<br>EF loop) (Å) |
|--|--|--|---|--|
| 1NTZ<br>(substrate<br>ubiquinone-2<br>(UQ2)) | phe274<br>met124<br>ala125<br>met129<br>phe274 | O <sub>4</sub> ...H161 N <sub>ε</sub> = 4.6<br><br>O <sub>2</sub> ...H <sub>2</sub> O <sub>1069</sub> = 2.8,<br>H <sub>2</sub> O <sub>1069</sub> ...E271 N =<br>3.1<br><br>O <sub>1</sub> ...E271 N = 2.84 | O <sub>2</sub> ...E271 N = 4.1<br><br>O <sub>2</sub> ...H <sub>2</sub> O <sub>1069</sub> = 2.8,<br>H <sub>2</sub> O <sub>1069</sub> ...E271 N =<br>3.1<br><br>O <sub>1</sub> ...E271 N = 2.84 | N <sub>i</sub> ...Y273 O<br>(OH) = 3.1                               |
| 1SQP<br>(inhibitor<br>myxothiazol,<br>MYX)   |  |  |   |  |
| 1SQX<br>(inhibitor<br>stigmatellin,<br>SMA)  | phe274<br>met124<br>ala125<br>met129<br>phe274 | O <sub>4</sub> ...H161 N <sub>ε</sub> = 3.0<br>O <sub>5</sub> ...H161 N <sub>ε</sub> = 3.6   | O <sub>8</sub> (OH)...E271 O <sub>ε1</sub> =<br>2.7   |  |
| 1PP9<br>(inhibitor<br>stigmatellin,<br>SMA)  |  |  |   | O <sub>8</sub> (OH)...E271 O <sub>ε1</sub> =<br>2.6                  |
| 1PPJ<br>(inhibitor<br>stigmatellin,<br>SMA)  | phe274   | O <sub>4</sub> ...H161 N <sub>ε</sub> = 2.7<br>O <sub>5</sub> ...H161 N <sub>ε</sub> = 3.5   | O <sub>8</sub> (OH)...E271 O <sub>ε1</sub> =<br>2.6   |  |

|  |   | Q <sub>i</sub> Site  |  |   |  |
|--|---|--|--|---|--|
| PDB<br>(substrate/<br>inhibitor)             | Ar-Ar or<br>hydrophobic<br>contacts   | ser35 S35<br>(from helix A)  | his201 H201<br>(from helix D)  | ser205 S205<br>(from helix D)   | lys227 K227<br>(from helix E)                                |
| 1NTZ<br>(substrate<br>ubiquinone-2<br>(UQ2)) | Ar-Ar with<br>phe220  | O <sub>4</sub> ···H <sub>2</sub> O = 3.4,<br>H <sub>2</sub> O···H201 N <sub>ε2</sub> = 2.6                   | O <sub>4</sub> ···S205 O <sub>γ</sub> =<br>4.0,<br>O <sub>3</sub> ···S205 O <sub>γ</sub> =<br>2.7 Å                                      | O <sub>1</sub> ···H <sub>2</sub> O = 3.6,<br>H <sub>2</sub> O···K227 N <sub>ε</sub> = 3.3 | O <sub>1</sub> ···D228 O <sub>δ2</sub><br>= 2.5 <sup>a</sup> |
| 1SQX<br>(substrate<br>ubiquinone-2,<br>UQ2)  | O <sub>1</sub> ···S35 O <sub>γ</sub> = 4.3  | O <sub>4</sub> ···H <sub>2</sub> O = 2.5,<br>H <sub>2</sub> O···H201 N <sub>ε2</sub> = 2.6                   | O <sub>3</sub> ···S205 O <sub>γ</sub> =<br>3.5   | not involved  | O <sub>1</sub> ···D228 O <sub>δ2</sub><br>= 2.7 <sup>a</sup> |
| 1PP9<br>(substrate<br>ubiquinone-10,<br>UQ)  | Ar-Ar with<br>phe220  | O <sub>1</sub> ···H <sub>2</sub> O = 3.2,<br>H <sub>2</sub> O···S35 O <sub>γ</sub> = 3.0                     | O <sub>4</sub> ···H201 N <sub>ε2</sub> = 2.3   | O <sub>3</sub> ···S205 O <sub>γ</sub> =<br>3.7  | not involved   |
| 1NTK<br>(inhibitor<br>anitmycin A,<br>AY1)   | Ar-Ar with<br>phe220  | O <sub>3</sub> ···S35 O <sub>γ</sub> = 3.0,<br>O <sub>7</sub> ···S35 O <sub>γ</sub> = 3.7                    | N <sub>2</sub> ···H <sub>2</sub> O = 2.5,<br>H <sub>2</sub> O···H <sub>2</sub> O = 2.8,<br>H <sub>2</sub> O···H201 N <sub>ε2</sub> = 2.5 | not involved  | O <sub>1</sub> ···K227 N <sub>ε</sub> =<br>3.7               |
| 1PJ1<br>(inhibitor<br>antimycin A,<br>ANY)   | O <sub>1</sub> ···H <sub>2</sub> O <sub>22</sub> = 2.9,<br>H <sub>2</sub> O <sub>222</sub> ···S35 O <sub>γ</sub> =<br>3.0 | O <sub>3</sub> ···H <sub>2</sub> O = 2.8<br>H <sub>2</sub> O <sub>473</sub> ···H201 N <sub>ε2</sub> =<br>2.7 | not involved   | O <sub>2</sub> ···D228 O <sub>δ2</sub><br>= 2.3 <sup>b</sup>                              |  |
|  |   |  |  | N <sub>1</sub> ···D228 O <sub>δ1</sub><br>= 3.0   | O <sub>1</sub> ···D228 O <sub>δ2</sub><br>= 2.6,             |
|  |   |  |  | N <sub>1</sub> ···K227<br>N <sub>ε</sub> = 2.7  | N <sub>1</sub> ···D228 O <sub>δ1</sub><br>= 2.8              |

<sup>a</sup>The H-donor is questionable in these hydrogen bonds.<sup>b</sup>The H atoms of the salicylamine hydroxyl OH (O<sub>2</sub>) must be the H-donor for both the intramolecular hydrogen bond with the amide linker carbonyl group (O<sub>3</sub>) and the O<sub>ε2</sub> carboxylic acid oxygen of asp228 assuming the carboxylic acid group is deprotonated.

the group included the inhibitor myxothiazol (PDB: 1SQP, resolution 2.7 Å). Myxothiazol, MYX (see Figure 7.29), belongs to class Ia group of inhibitors—that is, it binds in the Q<sub>o</sub> site and has chemical characteristics placing it with others of the “a” subgroup that contain a β-methoxyacrylate or similar entity. Table 1 of reference 88 includes descriptions of a variety of bc<sub>1</sub> inhibitors, classified by the authors into categories based on bc<sub>1</sub>-binding site, induction of ISP mobility by the inhibitor, and inhibitor structural characteristics. Only two of these inhibitor–enzyme complexes, those with inhibitors myxothiazol (Xia class P<sub>m</sub>—inducing mobility of the ISP) and stigmatellin A (Xia class P<sub>f</sub>—inducing a fixed ISP position), will be discussed here. Pharmacologically, myxothiazol is an antifungal agent produced by myxobacterium *M. fulvus*. It is widely used as an inhibitor of cellular respiration. Chemically, it contains a methoxyacrylamide group (similar to the characteristic β-methoxyacrylate group often used in cytochrome bc<sub>1</sub> complex inhibitor studies) and a dithiazole moiety. (See Figure 7.29.) As stated previously, myxothiazol, MYX, increases the mobility of the ISP protein (P<sub>m</sub> group). It prevents electron transfer from ubiquinol to the ISP and causes red shifts in the α- and β-bands of the reduced heme b<sub>L</sub> UV–visible spectrum. In the Q<sub>o</sub> pocket, myxothiazol is held in place by aromatic–aromatic (Ar–Ar) stacking interactions between a thiazole ring and the side chain of phe274. Additionally, there are hydrogen bonds between the O<sub>1</sub> of the methoxyacrylamide moiety and the backbone amide of glu271 (O<sub>1</sub>···N = 2.84 Å) and the N<sub>1</sub> of the methoxyacrylamide moiety and the hydroxyl group of tyr273 (N<sub>1</sub>···O = 3.1 Å). The glu271–MYX interaction is conserved in all class Ia inhibitors. Both glu271 and tyr273 belong to the PEWY sequence, highly conserved in cytochrome b (SU 4 of cytochrome bc<sub>1</sub>) proteins. This sequence is located between helices E and F in the EF loop. (See Figure 7.28.) In bovine bc<sub>1</sub>, the PEWY sequence is composed of pro270–glu271–trp272–tyr273. Mutations in the PEWY sequence are known to block electron transfer and quinol binding at the Q<sub>o</sub> site. Resistance to cytochrome bc<sub>1</sub> inhibitors may also be related to changes in the PEWY sequence. The closest approach of MYX to a heme metal ion occurs between the N<sub>1</sub> position of MYX and the iron ion in heme b<sub>L</sub> (N<sub>1</sub>···Fe = 10.1 Å) and the shortest distance between hem b<sub>L</sub> and any portion of MYX is 5.4 Å.

The Xia group also studied the binding of the inhibitor stigmatellin A (SMA) (PDB: 1SQX, resolution 2.5 Å), a class Ib inhibitor. Class Ib inhibitors have a chromone ring system (see Figure 7.29), inhibit electron transfer from cytochrome c<sub>1</sub> to the ISP, cause an increased redox potential in the ISP, decrease the ISP mobility (P<sub>f</sub> group), and, like class Ia inhibitors, cause a red shift in the reduced heme b<sub>L</sub> UV–vis spectrum. Stigmatellin (SMA) is a natural antifungal agent produced by myxobacteria *S. aurantiaca*. This inhibitor will not bind to cytochrome bc<sub>1</sub> in the absence of the ISP. The protonated N<sub>e2</sub> atom of his161 in the ISP forms two hydrogen bonds with SMA’s methoxy oxygen (O<sub>5</sub>···N<sub>e2</sub> = 3.6 Å) and its carbonyl group (O<sub>4</sub>···N<sub>e2</sub> = 3.0 Å). Part of the chromone ring intercalates between pro270 and ile146. A strong hydrogen bond is

formed between the SMA hydroxyl group and atom O<sub>ε1</sub> of glu271 (O<sub>8</sub> (OH)…O<sub>ε1</sub> = 2.7 Å). Stigmatellin A binding turns the side chain of glu271 about 180° about the χ<sup>1</sup> dihedral angle compared to the native protein. The mostly hydrophobic SMA tail interacts with phe274, met124, ala125, and met129. The closest approach of SMA to a heme metal ion [between the O<sub>8</sub> position of SMA and the iron ion in heme b<sub>L</sub> (O<sub>8</sub>…Fe = 14.6 Å)] is 4 Å longer than the distance for the inhibitor myxothiazol. The reference 88 authors believe that the PDB: 1SQX structure presents stigmatellin A in a stereochemically correct manner, whereas earlier X-ray crystallographic structures (including PDB: 3BCC) may have represented stigmatellin incorrectly or incompletely. Mutations in or near stigmatellin A's binding pocket in cytochrome bc<sub>1</sub> are known to cause resistance to the fungicide, probably because the shape of the Q<sub>o</sub> pocket changes and SMA binding is decreased or eliminated. In the PDB: 1SQX structure, a ubiquinone-2 (UQ2; see Figure 7.29) molecule is located near heme b<sub>H</sub> in the Q<sub>i</sub> pocket. Table 7.7 lists the interactions of this Q<sub>i</sub> site ubiquinone-2 molecule: (1) a carboxyl oxygen of UQ2 interacts with the hydroxyl group of ser35 (O<sub>1</sub>…S35 O<sub>γ</sub> = 4.3 Å); (2) a water molecule intervenes to hydrogen bond the N<sub>ε2</sub> nitrogen of his201 with the UQ2's opposite carboxyl oxygen (O<sub>4</sub>…H<sub>2</sub>O = 2.5 Å, H<sub>2</sub>O…H201 N<sub>ε2</sub> = 2.6 Å); (3) a methoxy oxygen of UQ2 hydrogen bonds to the hydroxyl OH of ser205 (O<sub>3</sub>…S205 O<sub>γ</sub> = 3.5 Å); and (4) a carboxyl oxygen hydrogen bonds to the O<sub>δ2</sub> atom of asp228 (O<sub>1</sub>…D228 O<sub>δ2</sub> = 2.7 Å). In the last case, one must assume that the O<sub>δ2</sub> atom of asp228 is protonated and is the H-donor to the hydrogen bond. At physiological pH, an aspartic acid residue is normally believed to be deprotonated, making the side-chain carboxylic acid oxygens always hydrogen-bond acceptors. One solution is to project an intervening water molecule; however, these may not be present at the needed site, even in crystallographic structures of sufficient resolution to pinpoint their positions. The conclusion must be that interpretation of hydrogen bonding scenarios for inhibitor molecules and their interactions with side-chain or backbone amino acid residue atoms should be carried out with caution. Even if the atoms are close enough to engage in hydrogen bonding, the necessary hydrogen atom may not be present. Also, since the crystals being studied have not necessarily been formed at physiological pH, but instead below this, the aspartic acid residues may indeed be protonated in the crystal with hydrogen's available as H-donors to hydrogen bonds. In Table 7.7, a footnote has been added where questions should be asked about hydrogen bonding partners.

In 2003, the Di Xia group published an X-ray crystallographic study of substrate and inhibitor molecules at the Q<sub>o</sub> and Q<sub>i</sub> site.<sup>86</sup> They intended to compare the structure of the native enzyme (PDB: 1NTM) with the cytochrome bc<sub>1</sub> complex having substrate or inhibitors in the Q<sub>o</sub> and/or Q<sub>i</sub> sites. The complex with substrate ubiquinone UQ2 (PDB: 1NTZ) was discussed previously in this section. The complex with the inhibitor antimycin A<sub>1</sub> (PDB: 1NTK) will be discussed here. A fourth complex with the ubiquinone-model

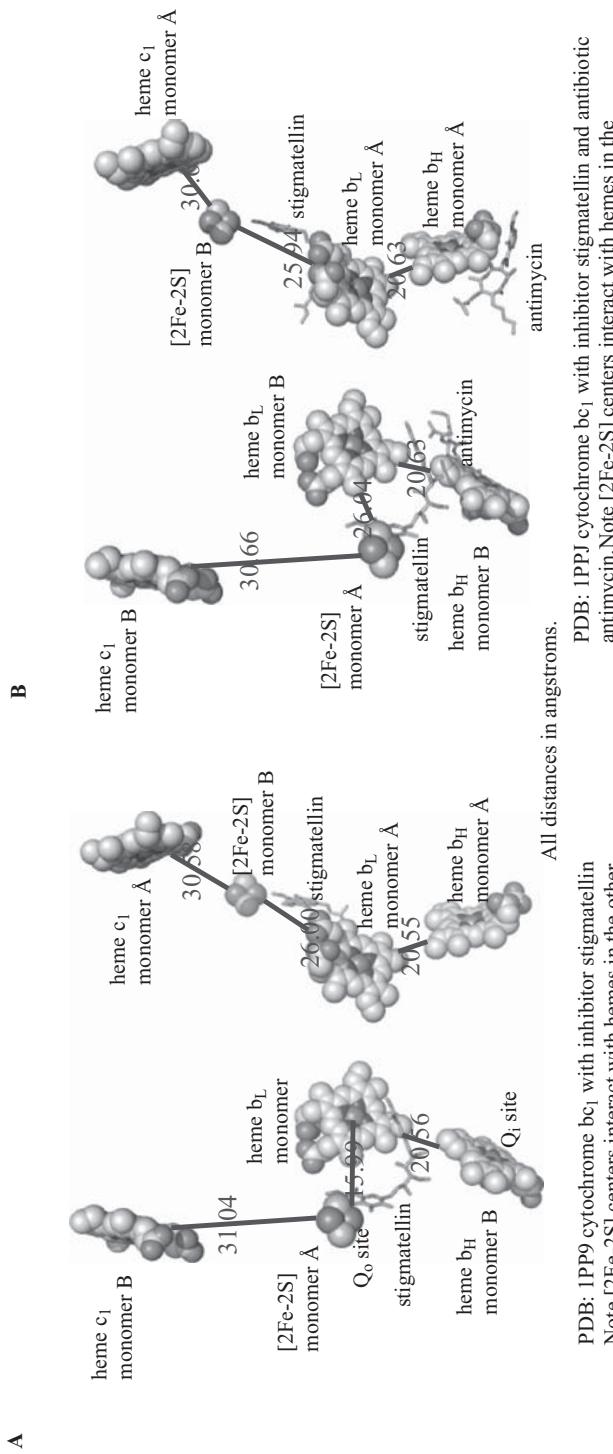
inhibitor NQNO (PDB: 1NUI) competes with ubiquinone for the  $Q_i$  pocket and binds in the  $Q_o$  pocket as well (a Xia classification PN inhibitor). This structure will not be discussed further.

It was known from previous research that the inhibitors antimycin A and NQNO (2-nonyl-4-hydroxyquinoline N-oxide) specifically inhibit activity at the  $Q_i$  site (the quinone reduction site) where a quinone molecule receives two electrons from heme  $b_H$  and two protons from the negative side of the membrane. Antimycin A was one of the first known and is still one of the most potent inhibitors of the mitochondrial respiratory chain. The antimycin  $A_1$  molecule (see Figure 7.29) consists of three parts: (1) a nine-member puckered dilactone ring in the center; (2) a 3-formylamino salicylic acid (3-FASA) group connected to the dilactone through an amide bond; and (3) a hydrophobic tail on the other side of the dilactone ring. The dilactone ring consists of an L-threonine (connected to the salicylate through the amide bond) and a 2-alkyl-3,4-dihydroxyvalerate. It is believed that the 3-FASA portion of the molecule is responsible for the binding specificity. In the antimycin A molecule itself, an intramolecular hydrogen bond exists between the hydroxy group on the 3-FASA aromatic ring and the carbonyl oxygen group of the amide bridge. This hydrogen bond is illustrated in Figure 7.29 for the antimycin A molecule with the PDB: 1NTK label. As we will see later, the numbering system (and the intramolecular hydrogen bond) is different for antimycin A in the PDB: 1PPJ structure to be discussed next. Antimycin  $A_1$  binds in the  $Q_i$  pocket and is therefore called a class II inhibitor or, in the nomenclature of Xia et al., an N inhibitor. Antimycin  $A_1$  binds through a network of six hydrogen bonds and one Ar–Ar interaction. The 3-FASA  $N_1$  atom hydrogen bonds to  $O_{\delta 1}$  of asp228 ( $N_1 \cdots O_{\delta 1} = 3.0 \text{ \AA}$ ), while the  $O_{\delta 2}$  of asp228 binds to the 3-FASA hydroxyl oxygen ( $O_2 \cdots O_{\delta 2} = 2.3 \text{ \AA}$ ). Atom  $O_\gamma(\text{OH})$  of ser35 interacts with dilactone carbonyl group  $O_7$  ( $O_7 \cdots O_\gamma = 3.7$ ) and the carbonyl group with oxygen  $O_3$ , in the amide linker ( $O_3 \cdots O_\gamma = 3.0$ ). The  $N_{\epsilon 2}$  atom of his201 and the carbonyl oxygen of ala17 hydrogen bond through two water molecules (2.8 Å apart) to the antimycin A  $N_2$  salicyamide group. The aromatic group of phe220 has an edge-to-face distance to the 3-FASA ring of 3.1 Å, forming an Ar–Ar interaction. The hydrogen-bonding amino acid residues are part of the  $Q_i$  cavity formed by residues from transmembrane helices A (trp31, asn32, gly34, ser35), D (ala193, met194, leu197, his201), and E (tyr224, lys227, asp228). Residues from the amphipathic (having both negatively and positively charged—hydrophilic—as well as hydrophobic aa residues) surface helix A (phe18), from the A loop (ile27), and from the DE loop (ser205, phe220), and atoms from the high potential heme  $b_H$  (carbon atoms and a propionate side chain) also line the  $Q_i$  cavity.

In 2005, the E. A. Berry group published X-ray crystallographic studies of cytochrome bc<sub>1</sub> with bound inhibitors.<sup>90</sup> The structure of PDB: 1PP9, resolution 2.23 Å, includes stigmatellin A (SMA) in the  $Q_o$  site and substrate ubiquinone (UQ) in the  $Q_i$  site, while PDB: 1PPJ, resolution 2.28 Å, includes stigmatellin A in the  $Q_o$  site and antimycin A<sub>1</sub> (ANY) in the  $Q_i$  site. These

structures are the best resolved of any discussed here so far (PDB: 3BCC, resolution 3.70 Å versus PDB: 1NTK, resolution 2.60 Å). The inhibitor antimycin A was thought to bind near heme b<sub>H</sub> before X-ray crystallography confirmed this fact for two reasons: (1) The heme b<sub>H</sub> UV-visible spectrum shifted to longer wavelengths (red shift) in the presence of antimycin A, and (2) the fluorescence of antimycin A was quenched when it was added to the cytochrome bc<sub>1</sub> complex.<sup>91</sup> The observed EPR signal of the semiquinone (QH<sup>-•</sup>) is eliminated by antimycin A, consistent with predictions of the Q cycle that a Q<sub>i</sub> site inhibitor will inhibit the transfer of electrons from heme b<sub>H</sub> to ubiquinone (Q) or ubisemiquinone (QH<sup>-•</sup>).<sup>92</sup> It has also been postulated that conformational change in several subunits of the bc<sub>1</sub> complex would couple oxidation-reduction events taking place at both the Q<sub>o</sub> and Q<sub>i</sub> sites. For instance, different positions of the ISP protein detected by X-ray crystallographic studies (as discussed in Sections 7.6.2 and 7.6.3) might indicate a link between the bifurcation of electron transfer from ubiquinol to cytochrome b at the Q<sub>o</sub> site and proton-pumping at the Q<sub>i</sub> site where antimycin A binds.<sup>93</sup> These predicted conformational changes involving both the Q<sub>o</sub> and Q<sub>i</sub> sites with and without inhibitors have not been detected in X-ray crystallographic studies (see Table 7.6, for instance). These include those discussed above. Specifically, the PDB: 1NTZ structure with the substrate substitute ubiquinone-2 in both sites has virtually the same overall conformational structure as PDB: 1NTK with antimycin A in the Q<sub>i</sub> site. A conformation comparison between the PDB: 1PP9 structure (stigmatellin in the Q<sub>o</sub> site and ubiquinone-2 in the Q<sub>i</sub> site) and PDB: 1PPJ (stigmatellin in the Q<sub>o</sub> site and antimycin A in the Q<sub>i</sub> site) by Berry and co-workers also indicates little conformational change upon antimycin A binding.<sup>90</sup> (See Figure 7.30.)

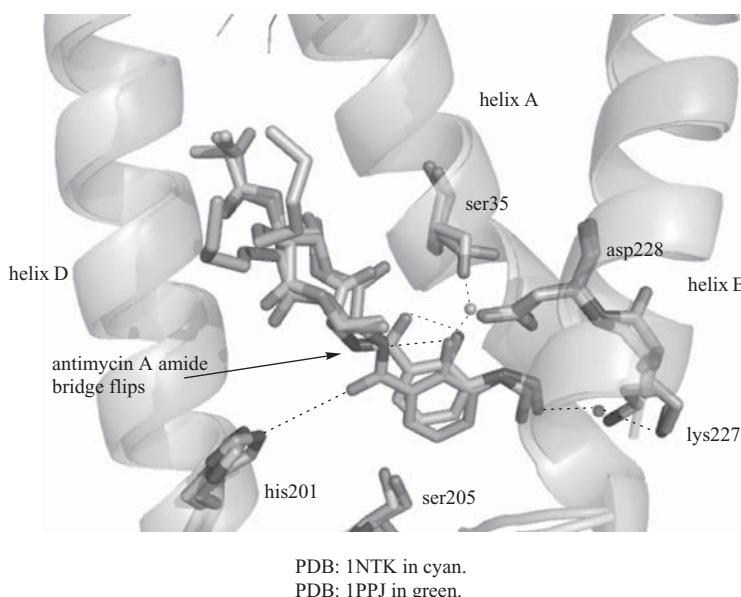
The Berry group researchers provide a detailed analysis of antimycin A binding in the Q<sub>i</sub> site. They identify six well-resolved water molecules that connect antimycin A through a network of hydrogen bonds to aa residues ser28, asn32, ser35, his201, ser205, lys227, asp228, and others. Their data are collected in Tables 4 and 6 of reference 90. Of this group, his201 and ser205 are believed to be most important in binding substrate to the Q<sub>i</sub> site so that inhibitor binding at these residues could prohibit binding of substrate molecules. Many close contacts between antimycin A and heme b<sub>H</sub> can explain antimycin's fluorescence quenching and UV-visible spectral shifts for heme b<sub>H</sub> upon antimycin A binding. There are Ar-Ar contacts between antimycin A's formylaminosalicylate headgroup and phe220 and between the dilactone's methyl group and methyl groups on heme b<sub>H</sub>'s pyrrole rings A and B. The following data use the antimycin A numbering convention of PDB: 1PPJ (reference 90), which, unfortunately, differs from that used for the PDB: 1NTK structure. The two numbering systems are shown in Figure 7.29. Interactions with ser35, his201, lys227, and asp228 are as follows: (1) The hydroxyl group of the formylaminosalicylate ring interacts with the O<sub>γ</sub> of ser35 (O<sub>1</sub>···H<sub>2</sub>O = 2.9, H<sub>2</sub>O···S35 O<sub>γ</sub> = 3.0); (2) the carbonyl oxygen of the amide linker interacts with the N<sub>ε2</sub> atom of his201 through an intervening water molecule (O<sub>3</sub>···H<sub>2</sub>O



PDB: 1PP9 cytochrome  $bc_1$  with inhibitor stigmatellin  
Note [2Fe-2S] centers interact with hemes in the other monomer.

**Figure 7.30** Comparison of cytochrome  $bc_1$  structures: (A) PDB:1PP9 with inhibitor stigmatellin; and (B) PDB:1PPJ with inhibitors stigmatellin and antimycin A. Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft Corporation.) (See color plate)

= 2.8, H<sub>2</sub>O···H201 N<sub>ε2</sub> = 2.7); (3) the N<sub>ζ</sub> atom of lys227's side chain interacts with the formyl carbonyl group through an intervening water molecule (O<sub>2</sub>···H<sub>2</sub>O = 2.6, H<sub>2</sub>O···K227 N<sub>ζ</sub> = 2.7); and (4) both asp228 carboxylic oxygens interact with substituents on the salicylate aromatic ring—the amino nitrogen and the hydroxyl oxygen (N<sub>1</sub>···D228 O<sub>δ1</sub> = 2.8, O<sub>1</sub>···D228 O<sub>δ2</sub> = 2.6). Conformational details of the antimycin A molecule itself in PDB: 1PPJ differ from those found in the small molecule structure<sup>94</sup> and in the PDB:1NTK structure. For instance, an intramolecular hydrogen bond between the phenolic OH group of the salicylamide ring and the carbonyl oxygen in the amide connector to the dilactone ring occurs in the small molecule structure and is shown also in the PDB: 1NTK structure (in PDB: 1NTK, the O<sub>3</sub>···O<sub>2</sub> distance = 2.3 Å). This intramolecular hydrogen bond was believed to be important for antimycin A binding and inhibitory activity, but this is thrown into question by different findings for the PDB: 1PPJ structure. In PDB: 1PPJ, the amide connector has flipped (rotated approximately 180°) with respect to the salicylamide ring so that the hydrogen bond is now formed between the phenolic OH and the NH group of the amide. (O<sub>1</sub>···N<sub>2</sub> = 2.6 Å). See Figure 7.31. Note also that the side chains of ser35 and lys227 differ in conformation between PDB: 1NTK and PDB: 1PPJ. This is most evident for lys227 in Figure 7.31.



**Figure 7.31** Comparison of antimycin A positions in PDB: 1NTK and PDB: 1PPJ. Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft Corporation.) (See color plate)

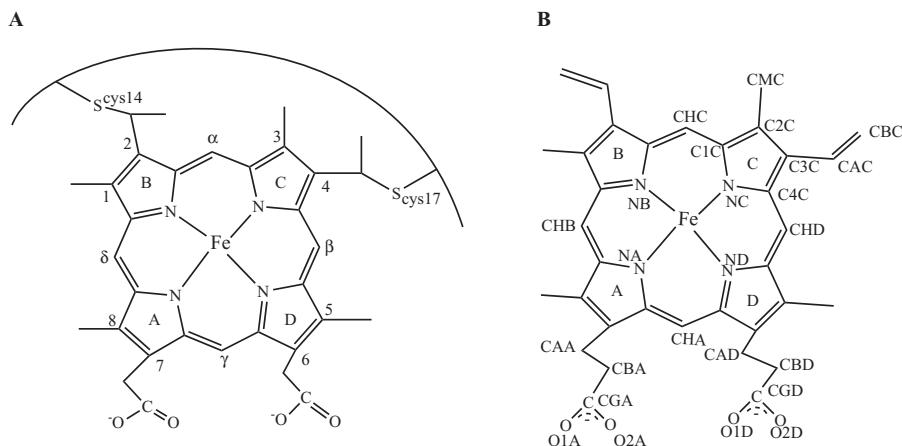
### 7.6.6 Cytochrome bc<sub>1</sub> Conclusions

Overall, the conformation of the cytochrome bc<sub>1</sub> complex does not change appreciably in the absence or presence of substrate or inhibitors except for the movement of the ISP protein and its [2Fe–2S] cluster. Researchers looking for other large changes in tertiary and quaternary structure in the cytochrome bc<sub>1</sub> complex based on the absence or presence of substrates or inhibitors or the oxidation state of iron in the heme and [2Fe–2S] centers must continue their search. Although a detailed discussion on the cytochrome bc<sub>1</sub> complex has preceded these conclusions, much has been left out of these sections as well. For instance, readers may be interested in the large body of knowledge available for yeast cytochrome,<sup>95</sup> including one structure of the cytochrome bc<sub>1</sub> complex with its redox partner cytochrome c attached.<sup>96</sup> An X-ray crystallographic structure of bacterial *Rhodobacter capsulatus* cytochrome bc<sub>1</sub> has been published by the E. A. Berry group.<sup>97</sup> An up-to-date website at <http://sb20.lbl.gov/cytbc1/> lists all the cytochrome bc<sub>1</sub> structures currently in the Protein Data Bank. A. R. Crofts maintains a cytochrome bc<sub>1</sub> webpage at [http://www.life.uiuc.edu/crofts/bc-complex\\_site/](http://www.life.uiuc.edu/crofts/bc-complex_site/) containing detailed information, graphic displays, movies, and links to other useful bc<sub>1</sub> sites. A review focusing solely on the mechanism of ubiquinol oxidation by the cytochrome bc<sub>1</sub> complex has been published recently.<sup>98</sup> Lastly, the reader should remember that X-ray crystallographic structures are frozen moments in time and space. While much of the information gained from these structures cannot be accessed in any other manner at present, confirmation of structure, function, and mechanism should be always be sought through other experimental means.

## 7.7 CYTOCHROMES c

### 7.7.1 Introduction

Cytochromes c (see Figure 7.25B) are ubiquitous in nature and have attributes that make them one of the most studied proteins in bioinorganic research. The protein's relatively small size (~12.5 kDa), thermodynamic stability, high solubility, and high helical content, as well as the presence of a spectroscopically accessible iron-containing heme cofactor, have made it a popular study target for biochemists and biophysicists. Cytochromes c are found outside the cytoplasmic membrane of prokaryotes, in the intermembrane space of eukaryotic mitochondria, or in the lumen of chloroplasts.<sup>99</sup> In eukaryotes, cytochrome c mediates single electron transfer between mitochondrial inner-membrane enzymes cytochrome bc<sub>1</sub> (Section 7.6) and cytochrome c oxidase (Section 7.8), which are complexes III and IV, respectively, of the respiratory chain. Cytochromes c are characterized by the attachment of the heme c cofactor to the protein chain through thioether linkages provided by two cysteinyl residues covalently bonded to vinyl substituents of the heme porphyrin ligand (see Figures 7.25 and 7.32A).



Structure of c-type heme group with Fischer nomenclature.

Thioether linkages to protein chain at positions 2 and 4.

Propionate groups at positions 6 and 7.

**Figure 7.32** (A) Cytochrome c heme Fischer numbering system. (Adapted with permission from Figure 1 of reference 119b.) Copyright 2000 National Academy of Sciences, U.S.A.) (B) The protoheme IX numbering system as outlined in reference 112.

In addition to the four porphyrin nitrogen ligands, cytochrome c's heme iron ion carries one axial histidine ligand. Some variety exists for the iron ion's second axial ligand, although it is often the S<sub>8</sub> atom of a methionine residue. In an early classification study, Ambler identified four classes of cytochromes c depending on the number of hemes, the type and position of the axial iron ligands, and the redox potential.<sup>27</sup> More information about cytochromes c can be found in a 1996 book written by Scott and Mauk.<sup>100</sup> Recently, Bertini et al. have published an extensive study of cytochromes c that broadly corresponds to those of Ambler's class I cytochromes.<sup>101</sup> These cytochromes feature a single-heme domain and possess a "cytochrome c fold." This particular fold has been well described in the literature<sup>100</sup> and can be identified as a superfamily in protein classification tools such as CATH<sup>102</sup> (<http://www.cathdb.info/latest/index.html>) or SCOP<sup>103</sup> (<http://scop.mrc-lmb.cam.ac.uk/scop/>). The CATH database classifies protein structures in the Protein Data Bank (PDB) in a hierarchical manner. Crystal structures resolved to better than 3.0 Å and NMR structures are classified. CATH has four major levels of hierarchy: (1) Class (mainly alpha (α-helical), mainly beta (β-pleated sheet), alpha-beta, few secondary structures); (2) Architecture (overall shape; e.g., alpha four-helical bundle as in cytochrome c); (3) Topology (fold family, determined by overall shape and connectivity of the protein secondary structure); and (4) Homologous superfamily (groups together protein domains thought to share a common ancestor and are therefore homologous). The SCOP (Structural Classification of Proteins) database provides a comprehensive description of structural and

evolutionary relationships between proteins whose structure is known. It surveys protein folds, detailed information about close relatives or any particular protein, and a framework for future research into protein classification systems.

The Bertini group's work<sup>101</sup> begins by describing the well-known "cytochrome c fold," which usually encompasses three structural elements: (1) N- and C-terminal  $\alpha$ -helices ( $\alpha$ 1 and  $\alpha$ 5 in mitochondrial cytochromes); (2) a long  $\alpha$ 3 helix (also called the 60's helix in mitochondrial cytochromes); followed by (3) a short  $\alpha$ -helix and loop that contains the heme iron's second axial ligand—usually a methionine. Alternatively, the heme iron ion's second axial ligand may be another histidine or an asparagine or may be absent altogether. All cytochromes c contain a cys-X-X-cys-his (CXXCH) signature sequence for attaching the heme to the rest of the protein. As stated above, the sulfur atoms of the cysteine residues provide thioether linkages to vinylic heme carbon atoms, while the signature histidine provides one axial ligand for heme iron ion.

Cytochromes c function as electron transfer proteins in aerobic and anaerobic respiration, but have a number of other physiological purposes. For instance, cytochromes c recently have been implicated in apoptosis (cell death) processes.<sup>104</sup> Mitochondrial cytochrome c appears to be necessary for the enzyme cytochrome c oxidase's assembly,<sup>105</sup> although the exact mechanism is unknown. In eukaryotes, a minor role in hydrogen peroxide scavenging has been assigned to some cytochrome c's.<sup>106</sup> The H<sub>2</sub>O<sub>2</sub> scavenging role in mammals takes place through interaction of cytochrome c and cytochrome c peroxidase (CCP).<sup>107</sup> In bacteria, cytochromes c are involved in respiratory processes and in H<sub>2</sub>O<sub>2</sub> scavenging. Often cytochromes c serve as the entry and exit point for electrons in the catalytic cycle of enzymes to which they are attached. This is the case for the cytochrome bc<sub>1</sub> complex as discussed in Section 7.6.

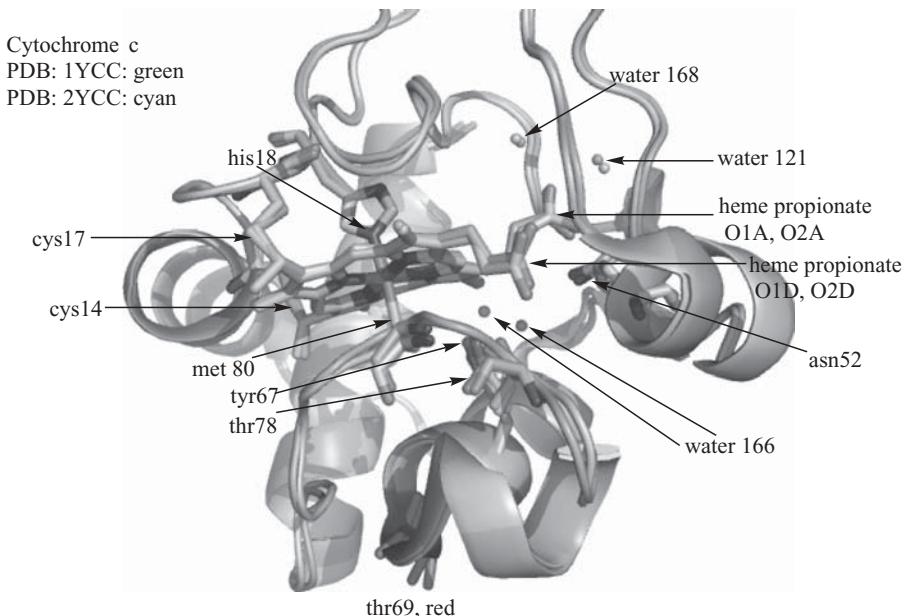
The cytochrome c bioinformatic analyses performed by the Bertini group (reference 101) identified 966 cytochrome c domains in 32 Eukaryota, 59 Gram-negative bacteria, 17 Gram-positive bacteria, and four Archaea. To further reduce the number of cytochromes c described, the authors concentrated on proteins containing a single cytochrome c domain. For instance, a "typical" primate mitochondrial cytochrome c is identified as a single-domain, soluble, 100- to 110-residue protein located in the intermembrane space of the mitochondrion. All mitochondrial cytochromes c are positively charged at physiological pH and are therefore attracted to (and at least sporadically attached to) negative domains on partner enzymes such as cytochrome bc<sub>1</sub> complex, cytochrome c oxidase, or cytochrome c peroxidase. During aerobic respiration, for instance, mitochondrial cytochrome c shuttles electrons from the bc<sub>1</sub> complex (Section 7.6) to cytochrome c oxidase (CCO or CcO, Section 7.8) in the mitochondrion. Figure 2 of reference 101 and Figure 1 of reference 78 show schematic diagrams illustrating the role of cytochrome c in transferring electrons from the bc<sub>1</sub> complex to cytochrome c oxidase as well as the translocation of protons and the connection to ATP (adenosine triphosphate)

synthesis from ADP. The bc<sub>1</sub> complex contains a c<sub>1</sub>-type cytochrome (detailed in Section 7.6) that differs from cytochrome c in several ways. First, the bc<sub>1</sub> complex c-type cytochrome is a much larger protein of about 200 residues that contains at least one transmembrane helix. Normally, cytochrome c proteins reside in the cytoplasm outside the cytoplasmic membrane of prokaryotes, in the intermembrane space of mitochondria, or in the lumen of chloroplasts.<sup>99</sup> The bc<sub>1</sub> complex is also identified in the literature as the ubiquinol/cytochrome c oxidoreductase or the QCR. The interactions between cytochrome c and the bc<sub>1</sub> complex have been extensively studied, and electron transfer partnership between the two has been described.<sup>108</sup> Much less is known about the interactions between cytochrome c oxidase (Section 7.8) and cytochrome c.

### 7.7.2 Mitochondrial Cytochrome c (Yeast)

Mitochondrial cytochrome c has been studied in eukaryotes such as yeast, tuna, and horse. Yeast iso-1 cytochrome c was the subject of X-ray crystallographic studies beginning in the 1990s. The Gary D. Brayer research group published early crystallographic structures of yeast iso-1-ferrocytochrome c (Fe(II)) (PDB: 1YCC)<sup>109</sup> and iso-1-ferricytochrome c (Fe(III)) (PDB: 2YCC).<sup>110</sup> The Fe(II) PDB: 1YCC structure refined to a resolution of 1.23 Å. The molecule exhibited a typical cytochrome c fold with the c-type heme being held in place by covalent linkages between cysteine S<sub>8</sub> thiolates and vinylic carbons of the protoporphyrin IX heme ligand in a CXXH pattern. In this yeast iso-1 cytochrome c, the CXXCH pattern involved cys14-leu15-gln16-cys17-his18. The heme's Fe(II) ligand carries the N<sub>e2</sub> atom of his18 as one axial ligand (N<sub>e2</sub>-Fe = 1.99 Å) and the S<sub>8</sub> atom of met80 as the other (S<sub>8</sub>-Fe = 2.35 Å) (see Figure 7.33 and Table 7.8). The protein's secondary structure organizes into a series of α-helices and reverse turns that envelop the heme prosthetic group in a hydrophobic pocket. In their local environment, the α-helices are sometimes distorted, especially near intrahelical threonine residues. See, for example, the region surrounding thr69, colored red in the lower center of Figure 7.33. In Figure 7.33, the secondary structure is shown in cartoon format with the PDB: 1YCC structure colored green and the PDB: 2YCC structure colored cyan. The heme prosthetic group, the axial ligands his 18 and met80, and selected important hydrogen bonding aa residues (asn52, tyr 67, and thr78) are visualized in stick format. Two heme c-protein-connecting residues, cys17 and cys14, are labeled. Based on refined crystallographic temperature factors, the most rigid parts of the molecule are closely associated with the heme group while surface residues show the greatest flexibility. In the overall molecule, 90% of the polar atoms participate in hydrogen bonding.

The PDB: 1YCC refined model includes 116 water molecules, most of them in hydrogen-bonding networks near the protein's surface. Mostly hydrophobic residues surround the heme crevice; however, at least three important water molecules—wat121, wat166, and wat168—form hydrogen-bonds with the propionate oxygen atoms and polar atoms of nearby aa residues. Wat166 is almost



**Figure 7.33** Two yeast cytochrome c molecules (PDB: 1YCC, Fe(II), green, and PDB: 2YCC, Fe(III), cyan). Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft Corporation.) (See color plate)

within or within hydrogen-bonding distance of three structurally important atoms or groups: (1) the heme O2D propionate oxygen ( $O2D\text{-}wat166 = 4.09 \text{ \AA}$ ); (2) the OH group of tyr67 (tyr67 OH-wat166 =  $2.61 \text{ \AA}$ ); and (3) the  $O_{\gamma 1}$  atom of thr78 (thr78  $O_{\gamma 1}$ -wat166 =  $2.72 \text{ \AA}$ ). The two other water molecules are part of a wider hydrogen-bonding network that includes the OH group of tyr48 and the heme O1A propionate oxygen ( $O1A\text{-OH tyr48} = 2.83 \text{ \AA}$ ,  $O1A\text{-wat121} = 2.81 \text{ \AA}$ ,  $O1A\text{-wat168} = 2.85 \text{ \AA}$ ), the heme O2A propionate oxygen and the backbone N of gly41 ( $O2A\text{-N} = 3.21 \text{ \AA}$ ), the  $N_{\delta 2}$  atom of asn52 ( $O2A\text{-}N_{\delta 2} = 3.34 \text{ \AA}$ ), the  $N_{\epsilon 1}$  atom of trp59 ( $O2A\text{-}N_{\epsilon 1} = 3.09 \text{ \AA}$ ) and wat121 ( $O2A\text{-wat121} = 4.01 \text{ \AA}$ ), the heme O1D propionate oxygen and the  $O_{\gamma 1}$  atom of thr49 ( $O1D\text{-}O_{\gamma 1} = 2.64 \text{ \AA}$ ), the  $O_{\gamma 1}$  atom of thr78 ( $O1D\text{-}O_{\gamma 1} = 2.90 \text{ \AA}$ ) and the backbone N of lys79 ( $O1D\text{-N} = 3.17 \text{ \AA}$ ), and the heme O2D propionate oxygen and the backbone N of thr49 ( $O2D\text{-N} = 2.94 \text{ \AA}$ ). The distance data are taken from the Pymol visualization of the PDB: 1YCC molecule (see Figure 7.33). All of the aa residues involved in the hydrogen-bonding scheme are conserved (at least for yeast and horse heart cytochrome c—see the discussion of the NMR structure of PDB: 1GIW below).

Criteria for hydrogen-bonding geometry in X-ray crystallographic structures have been established as follows: (1) The distance between the acceptor and the hydrogen ( $d_{HA}$ ) must be less than  $2.5 \text{ \AA}$ ; (2) the distance between the

electronegative “heavy” atoms in a hydrogen-bonded pair ( $d_{AD}$ ) must be less than 3.4 Å; and (3) the D–H–A angle must be greater than 90°.<sup>111</sup> Since most X-ray crystallographic structures do not display hydrogen atoms, criteria (2) will be used most frequently in this work. The protoheme IX group numbering system, as outlined in reference 112, is shown in Figure 7.32B.

The PDB: 2YCC X-ray crystallographic structure, with Fe(III) in the heme, refined to a resolution of 1.90 Å.<sup>110</sup> In the  $\alpha$ -helical domains, the PDB: 2YCC structure overlaps almost perfectly with its Fe(II) counterpart (PDB: 1YCC) (see Figure 7.33). Structural adjustments are expressed at the heme and in movements of internally bound water—water molecules bound near the heme. For instance, wat166 moves ~1.6 Å closer to the heme’s iron ion in the oxidized species, PDB: 2YCC, Fe(III)–wat166 = 5.0 Å, versus PDB: 1YCC, Fe(II)–wat166 = 6.6 Å for the reduced protein. Since the heme carries a positive charge in the Fe(III) oxidation state, wat166’s movement toward the iron ion may help compensate if its negative oxygen atom dipole points toward the Fe(III) ion. Other water molecules in the heme’s vicinity in the oxidized structure, wat168 and wat121, form hydrogen-bonding interactions similar to those seen for the reduced species. In PDB: 2YCC, the side chains of amino acid residues asn52 and tyr67 form stabilizing hydrogen bonds, with wat166 making this water molecule a major factor in stabilizing both oxidation states of the heme. Residues 47–79, 65–72, and 81–85, all on the same side of the protein as iron ion ligand met80, have increased mobility in the oxidized form. The ligand met80 itself takes slightly different orientations in the reduced and oxidized forms (see Figure 7.33). It thus appears that oxidation-state-dependent structural changes focus on met80 and the nearby wat166, especially since wat166 has moved to insert itself within hydrogen-bonding distance of the tyr67 OH group and met80’s backbone nitrogen. Distances between the heme’s Fe(III) ion and axial ligands his18 and met80 are very similar to those for the ferrocyanochrome (PDB: 1YCC)—that is, N<sub>e2</sub>–Fe = 2.01 Å and S<sub>8</sub>–Fe = 2.43 Å. The heme ligand itself is more distorted from planarity in the oxidized form (PDB: 2YCC average angular deviation of pyrrole rings from the pyrrole nitrogens’ plane = 12.2°) than in the reduced form (PDB: 1YCC, angular deviation = 9.4°). The normal situation for the heme prosthetic group in these cytochrome c proteins is to take on a shallow saddle-shape distortion from planarity. The propionate group attached to the pyrrole A ring, with heme oxygen atoms O1A and O2A, rotates about the CAA, CBA, and CGB bonds, resulting in a changed hydrogen-bonding network for the oxidized versus the reduced proteins. (See Figure 7.32B for the porphyrin atom identification system.) The differing positions of the heme propionate groups for the Fe(II) versus the Fe(III) structure are apparent in Figure 7.33. The reference 110 researchers conclude that (1) differences in thermal factors for different aa residues in the oxidized and reduced forms (as well as increased mobility for numerous surface residues near axial ligand met80 in the oxidized form) play a role in promoting electron transfer and in stabilizing the protein in its alternate oxidation states; (2) changing hydrogen-bonding networks induce con-

formational changes in the two oxidation states, around the met80 ligand and in the heme (especially around the pyrrole A ring and its attached propionate oxygens).

The Brayer group next began a series of mutational studies on yeast iso-1 ferro- and ferricytochromes c. The first mutant replaced the tyrosine residue at position 67 with phenylalanine—a Y67F mutant. Both oxidation states were studied: (1) the Fe(II)—Y67F mutant deposited as PDB: 1CTZ and (2) the Fe(III)—Y67F mutant as PDB: 1CTY.<sup>112</sup> The reduced state Y67F mutant appears to be very similar to its wild-type (WT) counterpart. Differences are seen for the oxidized mutant (PDB: 1CTY) in comparison to its WT counterpart (PDB: 2YCC). While little change is seen in overall polypeptide chain positioning for either oxidation state, more substantial changes in thermal factor (and therefore chain mobility) are seen, especially comparing the oxidized wild-type and Y67F mutant proteins. In the oxidized Y67F mutant, the heme porphyrin ligand is slightly less distorted from planarity—PDB: 2YCC average angular deviation of pyrrole rings from the pyrrole nitrogens' plane = 12.2°, PDB: 1CTY average angular deviation = 11.6°. Table 2 of reference 112 shows WT versus Y67F mutant comparisons of heme geometry—angular deviations for heme and axial ligands plus heme iron–ligand bond distances. The oxidized Y67F mutant heme met80 ligand changes its orientation from the wild-type; for instance, the C<sub>ε</sub> methyl group moves ~1.8 Å and rotates ~40° about the ligand bond. This movement may be connected to the observation that the met80-conformation-stabilizing wat166 molecule present in the WT protein is missing in the Y67F mutant. Additionally, the stabilizing hydrogen bond between met80 S<sub>8</sub> and tyr67 OH group is not possible for the Y67F mutant. The heme propionate group (O1A and O2A) attached to pyrrole ring A has changed orientation in WT versus Y67F structures as have the orientations of conserved arg38 and asn52 side chains. These reorientations lead to changes in hydrogen-bonding patterns in this region of WT versus Y67F mutant proteins. The differences between both reduced and oxidized wild-type and mutant forms are collected in Table 7 of reference 112. The differences noted include: (1) a drop in the midpoint reduction potential for the Y67F mutant (234 mV) compared to that for the WT (290 mV); (2) changes in thermodynamic properties; and (3) changes in UV–visible absorption spectra that indicate perturbation of the π-conjugated heme system in the mutant protein. The conclusion must be that while the overall fold of the cytochrome c protein does not change, mutations near the heme site will significantly alter the heme environment and thus alter its electron transfer behavior with redox partners.

The Brayer group continued its X-ray crystallographic structural study of yeast iso-1-cytochrome c variants with the following: (1) PDB: 1CRG, Fe(III)—asparagine52 to isoleucine52 (N52I) mutant; (2) PDB: 1CRH, Fe(II)—N52I mutant; (3) PDB: 1CRI, Fe(III)—N52I-Y67F mutant; and (4) PDB: 1CRJ, Fe(II)—N52I-Y67F mutant.<sup>113</sup> In all mutants, the most prominent structural difference from wild-type enzyme is the displacement of the conserved wat166

molecule. As described previously, this same water molecule moves its position substantially, depending on the oxidation state of the enzyme (see discussion of PDB: 1YCC and PDB: 2YCC above). The reference 113 authors suggest that the importance of wat166, along with its associated hydrogen-bonding network, rests on three major attributes: (1) wat166's position modifies the hydrogen-bonding network surrounding met80 in an oxidation-state-dependent manner; (2) wat166 maintains the spatial relationships between nearby amino acid side chains and therefore establishes the hydrogen-bonding network in this region of the protein (especially regarding the tyr67 hydroxyl group's hydrogen bond to S<sub>8</sub> of iron axial ligand met80); and (3) wat166 appears to mediate increases in the mobility of three nearby peptide segment in the enzyme's oxidized state. Regarding point (2) above, the reference 113 authors believe that the absence or presence of the tyr67 OH to met80's S<sub>8</sub> hydrogen bond affects the electron-withdrawing power of the met80 ligand and therefore is a factor in controlling the midpoint reduction potential in this cytochrome c (see discussion of PDB: 1CTY structure above). Regarding point (3) above, their research shows that, in the absence of wat166, mobility changes based on oxidation state do not occur in the N52I or N52I-Y67F mutants. This is an important finding since it is thought that changes in protein side-chain mobility with oxidation state are related to oxidation-state-dependent interactions between cytochrome c and its redox partners. The reference 113 authors also conclude that movement of amino acid side chains (particularly asn52) that participate in the hydrogen-bonding network around the heme pyrrole A propionate are responsible for oxidation-state conformational changes in this region.

In a 1996 *Biochemistry* paper, Brayer, Mauk, and co-workers published a study of cytochrome c electron transfer reactivity in relation to mechanistic and structural contributions by important surface and internal amino acid residues. Measurement of midpoint reduction potentials for wild-type versus mutant proteins showed that wild-type proteins had the highest potential at 290 mV. Mutants that changed the phe82 surface residue, a conserved aa residue thought to be important in interactions between cytochrome c and its redox partners, lowered the reduction potential less. However, a trend toward lower potential tracked with size of the variant amino acid substituted for phe82. For instance, the mutant phe82tyr (280 mV) exhibited less change than the phe82ser (255 mV) or the phe82gly (247 mV) mutants.<sup>114</sup> Mutants that varied conserved internal amino acids thought to be important in hydrogen-bonding networks near the cytochrome c heme—thr78, tyr67, asn52—caused larger changes in the midpoint reduction potential—asn52ala (257 mV), tyr67phe (236 mV), and thr78gly (245 mV). The group also studied reduction kinetics of wild-type and mutant ferricytochromes using [Fe(EDTA)]<sup>2-</sup> (EDTA = ethylenediaminetetraacetate) as a reductant and the oxidation kinetics of WT and mutant ferrocyanochromes using [Co(phen)<sub>3</sub>]<sup>3+</sup> (phen = 1,10-phenanthroline) as an oxidant. They applied Marcus theory (see Section 1.8) to calculate self-exchange rates ( $k_{11}$  or  $k_{11}^{\text{corr}}$ ) for the studied proteins. Electron

transfer proteins will exhibit variations of several orders of magnitude in  $k_{11}$ , depending on their changed electron transfer (eT) reactivity. The eT reactivity can then be related to structural changes (as found in X-ray crystallographic studies) and thus to eT mechanisms for WT versus mutant cytochromes c. In other words, variations in eT reactivity can be interpreted in two ways: (1) in terms of the effects of mutation on native oxidized or reduced cytochromes c (electrochemical effects) or (2) in terms of the mechanism by which proteins change oxidation state (kinetic effects). The results for the surface variants phe82X, where X = tyr, leu, ile, ala, ser, gly, indicated faster eT rates as X decreased in size. All variants exhibited faster rates than did wild-type. Faster relative rates were found for reduction of ferricytochrome c by  $[\text{Fe}(\text{EDTA})]^{2-}$  than for oxidation of ferrocyanide c by  $[\text{Co}(\text{phen})_3]^{3+}$ . The reference 114 authors interpreted the size relationship as favoring faster eT when a closer approach to the redox partner was possible. The researchers also discovered that, at least by this kinetic method, changes in the internal hydrogen-bonding network did not translate to lowered electron transfer reactivity. In fact, for the Y67F mutant discussed above, the relative rate of electron transfer ( $k_{11}^{\text{corr}}$ ) was nearly 11 times faster than that for wild-type ferri- or ferrocyanide c.

The extended study of yeast iso-1-cytochrome c by these researchers leads to several conclusions about the protein: (1) In all studied mutants, the overall cytochrome c fold is unchanged; (2) his18 and met80 are retained as the iron axial ligands in all studied mutants; (3) in the native (wild-type) protein, movement of the wat166 molecule closer to the heme iron in the Fe(III) state and its reorientation to bring its negative dipole closer to the heme iron ion will help relieve positive charge on the heme prosthetic group (with two negative charges on the propionate group the Fe(II) heme is uncharged); (4) breaking of the tyr67 OH–met80 S<sub>δ</sub> hydrogen bond in the Fe(III) state allows the sulfur lone-pair electrons to help balance the heme group's +1 charge; and (5) changes in hydrogen-bonding patterns near the heme are focused near conserved (invariant over different biological species) residues and the heme pyrrole A ring's propionate group.

### 7.7.3 Mitochondrial Cytochrome c (Horse)

One of the first X-ray crystallographic studies of horse heart cytochrome c was published by Takano and Dickerson in 1980.<sup>115</sup> (PDB: 3CYT) These workers first noted the small but significant ferro- and ferricytochromes c conformational differences surrounding the buried water molecule that is hydrogen-bonded to aa residues asn52, tyr67, and thr78 (see Figure 7.33). They saw, in the oxidized state, that the water molecule moves closer to the heme while the heme moves slightly out of its heme crevice, leading to a more polar heme environment. They noted that the cytochrome c heme crevice is bounded by lysines 8, 13, 27, 72, 79, 86, and 87.

In 1990, the Brayer group published the X-ray crystallographic structure of oxidized (Fe(III)) horse heart cytochrome c at a resolution of 1.90 Å (PDB: 1HRC).<sup>116</sup> The secondary structure of PDB: 1HRC overlaps well with its yeast iso-1 counterpart (PDB: 2YCC) discussed above. Conformational differences are noted for (1) residues 22–27, part of a surface β-loop, and (2) residues 41–43 and 56–57, near the heme prosthetic group. The heme groups overlap as well, both displaying shallow saddle-type distortions. Differences can be seen in the propionate group attached to heme pyrrole ring A, and this area seems to exhibit different conformations in most X-ray crystallographic structures studied. The aromatic side chain group of phe82 is positioned in a co-planar manner with the heme in both the horse heart and yeast iso-1 structures. Two internally located water molecules hydrogen-bond conserved amino acid residues in slightly different manners for PDB: 1HRC compared to PDB: 2YCC especially in the area of asn52, the residue that differs most in conformation between the two structures. Amino acid residues close to wat166 (PDB: 2YCC) and its counterpart wat112 (PDB: 1HRC) have very similar interactions. These interactions are collected in Table 7.8 along with axial ligand–iron ion distances.

In total, five water molecules occupy conserved positions in the structures of horse heart, yeast iso-1, tuna, and rice cytochromes c. Three of these are on the surface of the protein, serving to stabilize local polypeptide chain conformations. The remaining two are internally located. One of these mediates a charged interaction between the invariant residue arg38 and a nearby heme propionate. The other is more centrally buried near the heme iron atom and is hydrogen-bonded to the conserved residues asn52, tyr67, and thr78 (see Figure 7.33). It is shown that this latter water molecule shifts in a consistent manner upon change in oxidation state if cytochrome c structures from various sources are compared. The conservation of this structural feature and its close proximity to the heme iron atom strongly implicate this internal water molecule as having a functional role in cytochrome c's mechanism of action.

Recently, an NMR solution structure of horse heart ferrocytochrome c (heme Fe(II) ion) in a solvent mix of 70% water–30% acetonitrile (ACN) was published (PDB: 1LC1 (minimized average nmr structure) and 1LC2 (30 nmr structures)).<sup>117</sup> This ferrocytochrome c features a 104 amino acid residue single chain and an iron-containing heme c prosthetic group. The protein chain has five α-helices: helix I, residues 3–14; helix II, residues 50–55; helix III, residues 60–69; helix IV, residues 70–74; and helix V, residues 88–101. The secondary structure also features two omega loops: turn 1, residues 20–35; and turn 2, residues 35–44. The remainder of the chain consists of random coil segments. The c-type heme is attached covalently to the polypeptide chain through cys14 and cys17. The CXXCH sequence found in cytochromes c displays for PDB: 1LC1 as cys14–ala15–gln16–cys17–his18. His18 and met80 form the heme's axial ligands (his18 N<sub>ε2</sub>–Fe = 1.97 Å, met80 S<sub>δ</sub>–Fe = 2.68 Å). This is the longest met80 S<sub>δ</sub>–Fe distance reported for any structure studied here (see Table 7.8).

**TABLE 7.8** Selected Bonding Distances for Cytochromes c

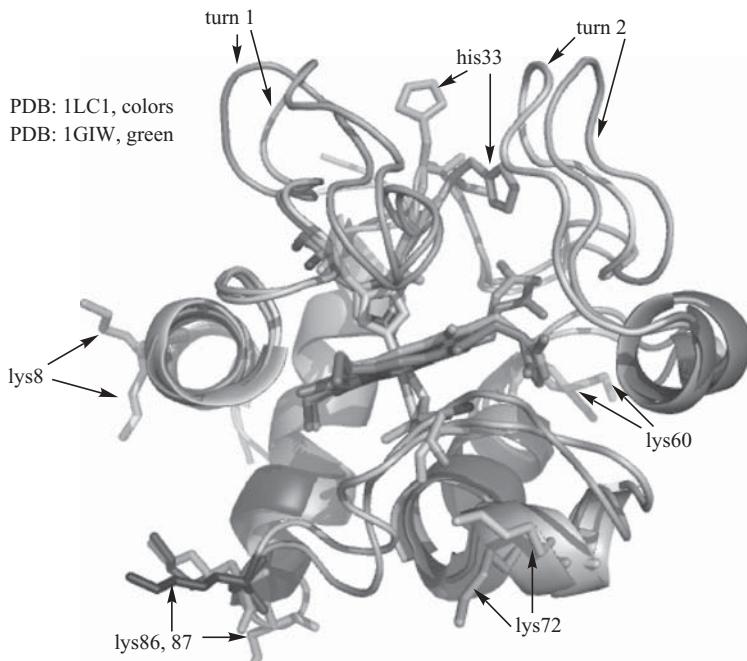
| PDB No.  | Fe (II)/(III)<br>(Species, Crystal Space<br>Group, Resolution Å) | H <sub>2</sub> O ID<br>(watxxx)  | H-Bond Partner—<br>H <sub>2</sub> O Distance (Å)  | Amino Acid to H-Bond Partner<br>Distance (Å) | Iron Ion—Amino<br>Acid Ligand                       | Bond Distance<br>(Å) |
|--|--|--|---|--|---|----------------------|
| 1YCC<br>Fe(II)<br>(yeast <i>iso</i> -1, P4 <sub>3</sub> 2 <sub>1</sub> 2, 1.23)  | wat166   | heme O2D (4.09) <sup>a</sup> ,<br>thr78 O <sub>γl</sub> 2.72,<br>tyr67 OH 2.61,<br>asn52 O <sub>δ1</sub> (5.28) <sup>a</sup> | met80 S <sub>8</sub> —tyr67 OH 3.25,<br>asn52 O <sub>δ1</sub> —heme O2A 3.76,<br>asn52 N <sub>ε2</sub> —heme O2A 3.34 | his18<br>met80                               | N <sub>ε2</sub> —Fe 1.99<br>S <sub>8</sub> —Fe 2.35 |                      |
|  | wat168   | asn52 N <sub>ε2</sub> 3.14<br>arg38 N <sub>ηl</sub> 2.56,<br>heme O1A 2.84   |   |  |   |                      |
| 2YCC<br>Fe(III)<br>(yeast <i>iso</i> -1, P4 <sub>3</sub> 2 <sub>1</sub> 2, 1.90) | wat166   | heme O2D (4.08) <sup>a</sup> ,<br>thr78 O <sub>γl</sub> 2.70,<br>tyr67 OH 2.63,<br>asn52 O <sub>δ1</sub> (6.47) <sup>a</sup> | met80 S <sub>8</sub> —tyr67 OH 3.12,<br>asn52 O <sub>δ1</sub> —heme O2A 3.40,<br>asn52 N <sub>ε2</sub> —heme O2A 3.53 | his18<br>met80                               | N <sub>ε2</sub> —Fe 2.01<br>S <sub>8</sub> —Fe 2.43 |                      |
|  | wat168   | asn52 N <sub>ε2</sub> (4.26) <sup>a</sup><br>arg38 N <sub>ηl</sub> 2.72,<br>heme O1A 2.87                                    |   |  |   |                      |
| 1HRC<br>Fe(III)<br>(horse, P4 <sub>3</sub> , 1.90)                               | wat112   | heme O2D (4.62) <sup>a</sup> ,<br>thr78 O <sub>γl</sub> 2.72,<br>tyr67 OH 2.78,<br>asn52 O <sub>δ1</sub> 3.24                | met80 S <sub>8</sub> —tyr67 OH 3.11,<br>asn52 O <sub>δ1</sub> —heme O1D 3.66,<br>asn52 N <sub>ε2</sub> —heme O2A 3.31 | his18<br>met80                               | N <sub>ε2</sub> —Fe 2.04<br>S <sub>8</sub> —Fe 2.32 |                      |
|  | wat139   | arg38 N <sub>ηl</sub> 2.84,<br>heme O1A 3.53   |   |  |   |                      |

|  |                        |  |                |   |
|--|------------------------|--|----------------|---|
| ILC1<br>Fe(II)<br>(horse, NMR)   | waters not<br>reported | met80 S <sub>8</sub> —tyr67 OH 3.37,<br>thr78 O <sub>δ1</sub> —tyr67 OH (4.14) <sup>a</sup><br>thr78 O <sub>δ1</sub> —O2D 2.57,<br>thr40 O <sub>δ1</sub> —O1A 3.09,<br>asn52 N <sub>ε2</sub> —heme O2D 2.78,<br>asn52 N <sub>ε2</sub> —heme O1A 2.82,<br>asn52 N <sub>ε2</sub> —heme O2A (5.04) <sup>a</sup> ,<br>asn52 O <sub>δ1</sub> —heme O2A (4.90) <sup>a</sup><br>asn52 O <sub>δ1</sub> —heme O1D (6.48) <sup>a</sup> | his18<br>met80 | N <sub>ε2</sub> —Fe 1.97 Å,<br>S <sub>8</sub> —Fe 2.68                            |
| 1GIW<br>Fe(II)<br>(horse, NMR)   | waters not<br>reported | met80 S <sub>8</sub> —tyr67 OH 3.18,<br>thr78 O <sub>δ1</sub> —O2D 2.61,<br>thr78 O <sub>δ1</sub> —O2D 3.27,<br>thr49 O <sub>δ1</sub> —O2D 3.06,<br>asn52 N <sub>ε2</sub> —heme O2D 2.81<br>asn52 N <sub>ε2</sub> —heme O1A (5.27) <sup>a</sup> ,<br>asn52 N <sub>ε2</sub> —heme O2A (6.97) <sup>a</sup> ,<br>asn52 O <sub>δ1</sub> —heme O2A (6.62) <sup>a</sup><br>asn52 O <sub>δ1</sub> —heme O1D (5.76) <sup>a</sup>     | his18<br>met80 | N <sub>ε2</sub> —Fe 1.94 Å,<br>S <sub>8</sub> —Fe 2.36                            |
| 3CYT<br>Fe(III)<br>(tuna, P4 <sub>3</sub> , 1.80)<br>1LFM<br>Co(III)<br>(tuna, P4 <sub>3</sub> , 1.60) |                        | met80 S <sub>8</sub> —tyr67 OH 3.07<br>met80   | his18<br>met80 | S <sub>8</sub> —Fe 2.28<br>N <sub>ε2</sub> —Co 2.05 Å,<br>S <sub>8</sub> —Co 2.38 |

<sup>a</sup>Bond distances too long to be considered as hydrogen bonds

The heme iron atom is 0.16 Å out of the plane formed by the four porphyrin nitrogen ligands and bulges toward the met80 sulfur ligand. Heme pyrrole ring C, which has been proposed to function as part of the physiological electron transfer pathway, exhibits pronounced buckling in the PDB: 1LC1 structure. It is the closest pyrrole ring to the surface residue phe82, also implicated in electron transfer pathways. The reference 117 authors compare their structure in 30% nonaqueous solvent with a horse heart ferrocyanochrome (Fe(II)) solution structure in aqueous solution published by the I. Bertini group in 1999 (PDB: 1GIW (minimized average structure) and 2GIW (40 NMR structures)).<sup>118</sup> In Figure 7.34, the PDB: 1GIW structure is visualized in green, while the PDB: 1LC1 structure is visualized in colors: helix I, yellow; helix II, red; helix III, magenta; helix IV, cyan; and helix V, royal blue. Turn 1 for PDB: 1LC1 is colored orange, while turn 2 is colored gold. Random coil segments for both structures are visualized in green.

The  $\alpha$ -helical portions of the two molecules overlap almost completely, while the loop portions show considerably more variation. Using NMR analytical methods, the reference 117 (PDB: 1LC1, 1LC2) authors indicate higher RMS (root mean square) deviations, and thus higher flexibility, for the backbone atoms in segments 21–24 and 41–45 for the ferrocyanochrome in 30% ACN than for the ferrocyanochrome in aqueous solution. (Compare the orange and gold loop sections for PDB: 1LC1 to the PDB: 1GIW counterparts in green in Figure 7.34.) Polar amino acid residues on the two ferrocyanochromes c surfaces show considerable conformational change, although their backbone atoms remain in similar positions. In Figure 7.34, these residues (lysines 8, 60, 72, 86, 87, leu32, and his33) are labeled. In the ferrocyanochrome in 30% ACN solution, the polar residues are more bent in toward the protein probably because of the unfavorable reaction of the polar side chains with the less polar (than water) solvent medium. The aromatic ring orientation of the important surface residue phe82 is different in the two structures, with the PDB: 1LC1 residue's ring being situated more in parallel with the heme plane. A notable change in the conformation of his33 confirms other studies showing that met80 is removed as a ligand in cytochrome c denaturing solutions and that his33 can become an iron ion ligand in these circumstances.<sup>119</sup> These authors have proposed his33 to be the sixth (and second axial) ligand in the V\* form of ferricytochrome c in 30% ACN solution and an alternative axial ligand replacing met80 in cytochrome c denatured by urea.<sup>120</sup> Similarly, lys72 exhibits a conformational change that may allow it to become a sixth ligand in the IVa form of alkaline ferricytochrome c and the IVa\* form of ferricytochrome c in mixed solvents. The nomenclature for conformationally changed ferricytochromes c includes: (1) state III—the native conformation that prevails at neutral pH; (2) state IV—conversion from state III at  $pK_a$  values between 8.5 and 9.5; and (3) state V—forms at alkaline pH values greater than 9.5. The conformational changes have been studied by a number of different methods, including resonance Raman (RR) spectroscopy by Mauk and co-workers.<sup>121</sup>



**Figure 7.34** Two horse cytochrome c molecules, (PDB: 1LC1, in various colors as described in the text, and PDB: 1GIW in green). Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft, Corporation.) (See color plate)

Conformational changes are found for the ferrocytochrome c hemes as well when the PDB: 1LC1 (30% ACN solution) and PDB: 1GIW (aqueous solution) structures are compared. These are most evident for the propionate moieties, especially those involving the propionate group attached to heme pyrrole ring A (the propionate-7 in terminology of reference 120. (See Figure 7.32 for the nomenclature used for c-type heme groups.) The reference 120 authors believe that this may be due to the loss or movement of water molecule 125. This cannot be confirmed from the PDB: 1LC1 visualizations because the water molecules are not included in the Protein Data Bank data. Sivakolundu and Mabrouk also compare the PDB: 1LC1 ferrocytochrome Fe(II) c X-ray structure with the X-ray crystallographic structure of ferricytochrome Fe(III) c crystallized at low ionic strength (PDB: 1CRC)<sup>122</sup> and with the X-ray crystallographic structure of ferricytochrome c complexed with a physiological binding partner cytochrome c peroxidase, CCP (PDB: 2PCB).<sup>123</sup> They find that the secondary structures are essentially the same and show only minor differences in loop domains. The orientation of the iron ion ligand his18 is quite similar in all structures, whereas all structures show differences in the orientation of the heme propionate-6 group (attached to pyrrole ring D). The

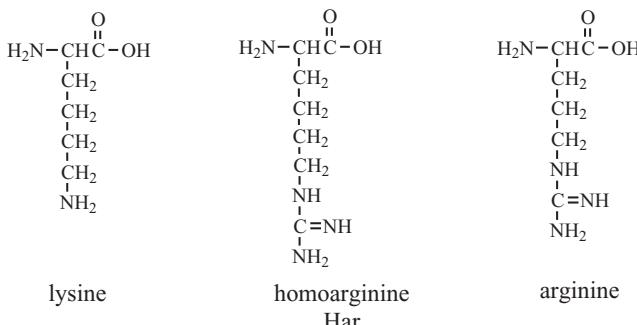
thioether linkages holding the cytochrome c in place appear more strained in the ferrocyanocytocrome c in 30% ACN (PDB: 1LC1) than in any of the other structures. Sivakolundu and Mabrouk discuss the various proposed mechanisms for ferro- and ferricyanocytocrome folding, electron transfer within cytochrome c and its redox partners, and the role of cytochrome c in cell apoptosis in relation to their studies of these cytochromes in nonaqueous solvents. These topics will be discussed in Section 7.7.4.

### 7.7.4 Cytochrome c Folding, Electron Transfer, and Cell Apoptosis

**7.7.4.1 Cytochrome c Folding.** Harry B. Gray and Jay R. Winkler have studied the cytochrome c folding landscape extensively. In a 2002 publication, the researchers used energy landscape theory to interpret experimental investigation of protein folding.<sup>124</sup> Using horse, tuna, and yeast cytochrome c, the workers substituted cobalt(III) ions for iron, finding that this did not change the secondary structure (comparison of tuna cytochrome X-ray crystallographic structure, PDB: 1LFM<sup>124</sup> with PDB: 1CYT<sup>115</sup>) or the absorption spectra. The replacement of the iron ion with cobalt(III) slows down the final step of the cobalt(III)-cytochrome c (Co-cyt c) folding process so that early folding intermediates can be examined by many different physical methods. The Co(III) substitution does not change the situation around the heme, and the X-ray crystallographic structures overlap almost perfectly (see bond distances in Table 7.8). The absorption spectra of horse, tuna and yeast Co-cyt c are virtually identical indicating that their heme environments are the same. For Co-cyt c, the so-called Soret maximum occurs at 427 nm and the Q bands occur at 534 and 568 nm. After addition of the denaturing reagent guanidine hydrochloride (GuHCl), unfolded protein is indicated by blue shifts to 422, 530, and 564 nm. Far-UV circular dichroism (CD) minima for Co-cyt c at 208 and 222 nm are characteristic of  $\alpha$ -helical structures in the folded protein. After addition of the denaturing reagent GuHCl, the far-UV spectra indicate random-coil conformation of the Co-cyt c protein. Trp59 fluorescence, fully quenched by energy transfer to heme in the folded protein, shows considerable fluorescence intensity in the unfolded Co-cyt c protein.

Kinetic experiments by these researchers have shown that horse and tuna Co-cyt c folds several orders of magnitude more slowly than the native proteins.<sup>125</sup> The refolding kinetics are biphasic because misfolding occurs during protein unfolding and folding events. In horse cytochrome c, his26 and his33 are likely “misfolding” ligands, thereby replacing met80, whereas in tuna cytochrome c, which lacks a his33 residue, one of 16 possible lysine residues probably becomes a “misfolding” ligand. To investigate the amino acid lysine’s contribution to misfolding events, the reference 125 researchers replaced all lysines (19 in horse cytochrome c, 16 in tuna) with homoarginine (Har) (see Figure 7.35) and reinvestigated the refolding kinetics.

Homoarginine is not expected to coordinate to the cobalt(III) metal centers. Results show that the tuna Har-Co-cyt c refolded in a single kinetic event



**Figure 7.35** Lysine, homoarginine and arginine.

while the horse Har-Co-cyt c's kinetics remained biphasic (see Figure 2 of reference 124). This result indicates that lysines as well as histidines can coordinate to the cytochrome c metal center during refolding events in denatured cytochrome c at neutral pH. It was also determined that his33 binds more tightly than his26 in unfolded horse Co-cyt c, and therefore his26 mis-ligated protein refolds 3 to 4 times faster than his33 mis-ligated protein. The conclusion is that folding is limited by histidine dissociation from Co(III) in horse Co-cyto c, and this happens faster for his26 because of the smaller loop generated when this aa residue coordinates.

Winkler, Gray, and Lyubovitsky use fluorescence energy transfer (FET) experiments to provide nanosecond-time-scale snapshots of the distance distributions between donor (D) and acceptor (A) labeled domains in rapidly equilibrating populations of folding intermediates.<sup>126</sup> For the yeast Co-cyt c experiments they labeled a C-terminal residue, cys102, with a dansyl fluorophore (5-(((2-iodoacetyl)amino)ethyl)amino)-naphthalene-1-sulfonic acid), DNS. The DNS derivatized-cys102 (donor, D) fluoresces when the protein is unfolded but is significantly quenched by energy transfer to the Co(III)-containing heme (acceptor, A) in the folded protein conformation. First, the DNS(cys102)-Co cyt c is denatured by addition of guanidine hydrochloride (GuHCl). Folding is then triggered by diluting the denatured solution in a stopped-flow mixer. Dilution-mixing is synchronized with excitation by an Nd:YAG laser (355 nm, 50 ps, 0.5 mJ) and the fluorescence decay followed with a picosecond streak camera. The DNS(cys102)-Co cyt c results show (1) a fast collapse of a fraction of the DNS(cys102)-Co cyt c to a DNS-Co(III) average distance of ~30 Å shortly after triggering folding; (2) the observation that most molecules are still in an extended conformation hours after folding initiation; (3) the formation of the Co cyt c native structure on an even longer time scale as both extended and compact conformations collapse to the native conformation; and (4) arrival at a distance distribution characteristic of the native protein (~25–27 Å) after about 18 hours for Co-cyt c. This is in contrast to results for DNS(cys102)-Fe-cyto c that evolves to the native structure within a few hundred milliseconds. The researchers also studied the behavior of these

systems immediately after dilution (concurrent with folding initiation) using NMR and circular dichroism (CD). The CD spectra show that elements of secondary structure, probably compact structures, begin forming shortly after folding initiation and NMR spectra show that these are short-lived (milliseconds). All the experimental evidence suggests that the compact and extended nonnative structures are in rapid equilibrium immediately after denaturant dilution. This conclusion agrees with computational experiments contending that global hydrophobic collapse is not an obligatory step in protein folding. Normally, collapsed intermediates are favored by more hydrophobic and less optimized protein primary sequences, while less hydrophobic and strongly optimized sequences favor extended structures collapsing to native protein conformation. The primary sequence and stability of cytochrome c place it between the two limits, making the extended and collapsed structures degenerate in energy. The relative instability of nonnative collapsed structures prevents misfolded protein structures during self-assembly and reduces the probability that the native protein will transiently adopt an incorrect conformation.

Many other research groups have studied the cytochrome c folding problem. One perspective that addresses some of the controversy surrounding cytochrome c folding pathways is found in a Rousseau group perspective published in 2000.<sup>127a</sup> Readers interested in pursuing this topic should consult reference 127b, in which Akiyama, Morishima, and co-workers studied stepwise formation of  $\alpha$ -helices during cytochrome c folding using CD spectroscopy.

**7.7.4.2 Electron Transfer in Cytochrome c and Its Redox Partners.** Electron transfer (eT) between cytochrome c and its redox partners has been studied by many researchers using many different experimental techniques. For specific recognition between redox partners, complementary protein surfaces are necessary and conformational changes may be required. Two different theories predominate the protein electron transfer discussion: (1) Electron transfer is mediated by the polypeptide backbone and specific amino acid side chains,<sup>128</sup> and (2) rapid electron transfer requires redox cofactors to approach each other at a minimal distance without a requirement for a specific path through the protein.<sup>129</sup> One area that has yielded information is electron transfer between cytochrome c and cytochrome c peroxidase as studied by X-ray crystallography and photochemically induced eT. Cytochrome c peroxidase (CcP) reacts with  $H_2O_2$  and two equivalents of ferrocyanochrome c to produce water, ferricytochrome c, and regenerated CcP. In a multistep process, and before receiving electrons from cytochrome c, CcP forms compound I, then compound II, intermediates of unknown structure. Native, resting state CcP contains a Fe(III) ion in its cytochrome b cofactor. Compound I would nominally contain a Fe(V) ion; however, Mössbauer and other studies have identified a ferryl (Fe(IV)) ion. The other electron lost in forming compound I has been shown to come from a tryptophan side chain, in the species being discussed here, a trp191 cation radical is formed. In mid-2006, no detailed structural information

had been published on the interaction of cytochrome c with its respiratory pathway partner cytochrome c oxidase (Section 7.8), so electron transfer to cytochrome c peroxidase will substitute in this discussion.

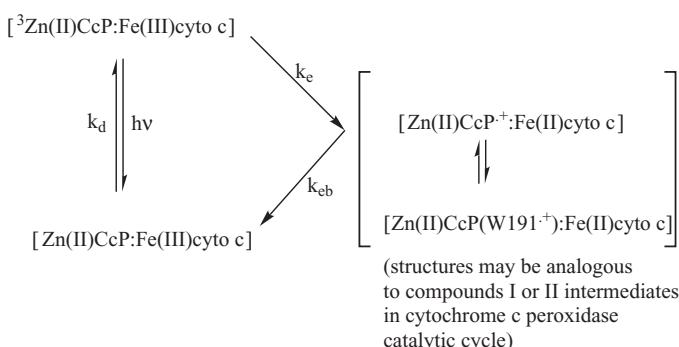
In 1992, Pelletier and Kraut published their study of complexes between yeast cytochrome c peroxidase (CcP) and yeast iso-1-cytochrome c (ccY), PDB: 2PCC, resolution 2.3 Å, and CcP and horse heart cytochrome c (ccH), PDB: 2PCB, resolution 2.8 Å.<sup>130</sup> Both structures show ferricytochrome c bound to CcP through electrostatic, hydrophobic, and van der Waals interactions. The two cytochromes, ccY and ccH, bind somewhat differently to CcP; most notable is a rotation and translation of the cytochromes c relative to CcP. Positively charged ccY surface residues lys73, lys87, and asn70 form longish hydrogen bonds (>3.0 Å) with CcP negatively charged residues asp34 and glu290. The ccY heme pyrrole ring C and especially its vinyl substituent (atoms CAC and CBC, Figure 7.32B) interact with side-chain and backbone atoms of CcP ala193 and ala194. There are a number of van der Waals interactions involving ccY ala81, conserved residue phe82, and gly83 with CcP residues val197 and gln120. Most residues reside on ccY and CcP loops that face each other at the proteins' interface. Lys73 and asn70 are residues in ccY helix IV. The ccY heme iron ion distance to the CcP heme iron ion is 26.4 Å. Two different surface residues from ccH—lys8, lys72—and one identical aa—lys87—form longish hydrogen bonds (>3.0 Å) to two different CcP residues—glu35, asn38—and one identical aa—glu290—at the ccH–CcP interface. Heme atom CBC of ccH interacts with CcP ala 193 and ala194 in similar fashion but at longer distances than for the ccY–CcP pair. Van der Waals interactions also differ for the ccH–CcP pair. The ccH heme iron ion distance to the CcP heme iron ion is 29.4 Å. The authors mention that the ccH–CcP crystals (PDB: 2PCB) were grown in a medium of essentially zero ionic strength, while the ccY–CcP crystals (PDB: 2PCC) were grown in high ionic strength media. Whether these differences are responsible for the different binding arrangements or the differing electron transfer kinetic behavior of the two cannot be determined solely from this work. The reference 130 authors assume that the most stable complex between CcP and the cytochromes c in solution is the one that crystallized. They show that the CcP–ccY structure does represent a true electron transfer complex and that the CcP–ccH structure is not in an optimal geometry for electron transfer (long distance between Fe centers represents one problem). However, they do believe that a highly specific electron transfer complex must be formed between redox partners and that future studies would show this to be the case. These structures did elucidate the important concept that direct electrostatic interactions between the two proteins are not critical in complex formation but that hydrophobic and van der Waals interactions are equally, or perhaps more, important.

Recent structural studies and electron transfer kinetic experiments focus on structures in which a site-specific covalent crosslink between cytochrome c and cytochrome c peroxidase subunits exists. One of these used site-directed mutagenesis to form a disulfide bond between a V197C mutant CcP and an A81C

mutant cytochrome c (cyto c) (PDB: 1S6V, resolution 1.88 Å).<sup>131</sup> The higher resolution of this crystal structure identified the positions of ordered water molecules that bridged the two proteins through hydrogen bonds. As with the PDB:2PCC structure, the direct electrostatic interactions do not appear critical in complex formation. The crosslinked CcP-cyto c complex behaves normally. UV-visible spectroscopic and stopped-flow kinetic studies indicate that the CcP domain of the crosslinked complex reacts with H<sub>2</sub>O<sub>2</sub> to produce the so-called compound I (see description of CcP above). Stopped flow studies also indicate that intramolecular electron transfer between the crosslinked ferrocyanocytocrome c and the CcP trp191 cation radical occurs within two milliseconds. The reference 131 authors believe these results indicate that the ferrocyanocytocrome c resides at or close to its physiological position in the PDB: 1S6V covalent complex (called PCXL in reference 131). Further evidence is shown in kinetic experiments using exogenously added yeast and horse cytochrome c. While wild-type CcP shows high and fast steady-state and single-turnover activity with exogenous cytochromes c, the PCXL covalently bonded CCP-cyto c complex shows much less activity and reacts much more slowly. The authors submit these experiments to indicate that CcP does not have a second independent site for physiologically relevant electron transfer.

Substitution of zinc(II) ions into cytochrome c peroxidase (ZnCcP) has been used to exploit photoactivation of electron transfer (eT) reactions since the mid-1990s.<sup>132</sup> The ZnCcP triplet state (<sup>3</sup>ZnCcP) reduces Fe(III) cytochrome c, and then back electron transfer recombines the charge separation to complete the catalytic cycle (see Figure 7.36).

In the experiment, <sup>3</sup>ZnCcP is produced by laser photolysis, then transient absorption spectroscopy follows the formation ( $k_e$ ) and decay ( $k_{eb}$ ) of ZnCcP<sup>•+</sup> in wild-type and mutant crystals. Kang and Crane have studied the effects of interface mutations on electron transfer rates in single crystals using complexes between a zinc-substituted cytochrome c peroxidase (ZnCcP) and site-directed mutants of yeast cytochrome c (yCc). The mutants replaced the



**Figure 7.36** <sup>3</sup>ZnCcP-Fe(III) reaction cycle. (Adapted with permission from Figure 1 of reference 133. Copyright 2005 National Academy of Sciences, U.S.A.)

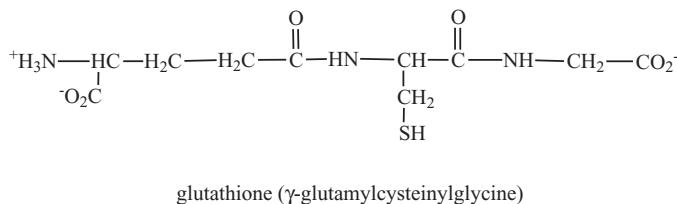
important cytochrome c conserved surface residue phe82 with tryptophan (F82W, PDB: 2B11), tyrosine (F82Y, PDB: 1B12), isoleucine (F82I, PDB: 2B0Z), and serine (F82S, PDB: 2B10).<sup>133</sup> It is known that mutations at phe82 modestly affect yCc stability, structure, and redox potential<sup>134</sup> and reduce rates of back electron transfer ( $k_{eb}$ ) to the resting state in comparison to ZnCcP complexed with WT cytochrome c.<sup>135</sup> (See Figure 7.36 and the discussion on electron transfer in ZnCcP–cytoc complexes that follows.) Substitutions at phe82 resulted in crystal structures with some interesting differences from wild type. For instance, F82W ZnCcP–yCc and F82S ZnCcP–yCc mutants maintain the same contact area between CcP and yCc as wild-type complexes (~1208 Å<sup>2</sup> buried surface area), and the F82Y ZnCcP–yCc and F82I ZnCcP–yCc mutants show CcP to yCc binding at a completely different interface with ~803 Å<sup>2</sup> of buried surface area. Relative to the WT binding, F82Y ZnCcP–yCc and F82I ZnCcP–yCc mutants show a 90° rotation and 8 Å translation of the yCc domain. Figure 2A of reference 133 illustrates the changes graphically. The changed domain-binding mode for the F8Y ZnCcP–yCc mutant is surprising since the difference between phenylalanine (phe,F) and tyrosine (tyr,Y) is only the addition of the hydroxyl group in tyrosine. One might speculate that changed hydrogen-bonding scenarios, altered polarity, or local pH variation could cause the substantial observed structural changes. The F82Y and F82I mutants show a longer donor (D) to acceptor (A) distance than do wild-type or F82W and F82S mutants (~3 Å greater metal to metal, ~7 Å greater heme edge to edge, and ~5 Å greater trp191 to heme). The F82W and F82S mutants show metal-to-metal, heme edge-to heme-edge, and trp191-to-heme geometry very similar to those of the WT yeast cytochrome c–ZnCcP complex. The electron transfer kinetic results are not easy to interpret. First, the back electron transfer rate ( $k_{eb}$ ) is more than 10 times faster than the rate ( $k_e$ ) forming the electron transfer intermediate. Considering only the  $k_{eb}$  rate, the value for the yeast wild-type complex vastly exceeds that for any of the mutants (or for that found for horse heart or tuna wild-type cytochrome c complexes as well). The fastest  $k_{eb}$  rates are found for the F82W and F82Y mutants. While the structure of the F82W ZnCcP–yCc mutant complex is very similar to that of the WT complex (and therefore electron transfer might be expected to approach wild-type rates), the F82Y ZnCcP–yCc complex has a substantially different protein–protein interface (and therefore electron transfer rates might be expected to be much slower than wild-type). Conversely, the F82S ZnCcP–yCc mutant, which does have a structure similar to that of WT, exhibits slower  $k_{eb}$  rates. The reference 133 authors conclude that the insensitivity of  $k_{eb}$  to distance and cofactor (heme) orientation may reflect limiting dynamic processes within the proteins and not just pure electron tunneling.

**7.7.4.3 Apoptosis.** Apoptosis—programmed cell death—involves elimination of damaged cells and maintenance of cell homeostasis. Homeostasis can be defined as the ability of living cells to regulate themselves in a dynamic manner. When apoptosis is deregulated, diseases such as cancers, immune

diseases, and neurodegenerative disorders may occur. Apoptotic events are controlled by cysteine proteases known as caspases. In mammals, caspase activity may be initiated through a cytochrome c pathway. The pathway begins when cytochrome c is released from mitochondria, initiating a series of reactions resulting in caspase activation and eventual cell death. In a recent review, Jiang and Wang focus on regulation of the pathway, its connection to physiology and disease, and potential therapeutics resulting from enhanced understanding of the apoptosis process.<sup>104</sup> Figure 1 of reference 104 shows the pathway controlling cytochrome c's role in cell death. The pathway requires participation of deoxy ATP (dATP) to induce activation of caspase-3, a homolog of caspase-9. Apaf-1, a binding partner of dATP, is also involved in pathway regulation. The oncogene BCL-2 family's role in the process is to prohibit cytochrome c release from the mitochondria. These facts are gathered from much research involving knockout animals—animals lacking the Apaf-1, caspase-9, and caspase-3 genes. The animals exhibited severe defects in central nervous system (CNS) development, indicating that the cytochrome c-mediated caspase activation pathway has an essential role in CNS development. The cytochrome c-mediated caspase activation pathway is also involved in amplification processes in cells known as type II cells. Certain cancers originate in type II cells. It is believed that therapeutics could be developed that target the cytochrome c-mediated caspase activation pathway pathway in synergism with attacks on other cancer-causing pathways. Why does the cytochrome c-mediated caspase activation pathway have its starting point(s) within the mitochondria? The reference 104 authors believe that the pathway can be isolated within the mitochondria and only perform its killing function when released during apoptosis.

The biochemistry of how cytochrome c-mediated caspase activation pathway triggers apoptosis is complex. Selected problems include: (1) formation of the so-called apoptosome with cytochrome c and its other necessary partners (see Figure 3 of reference 104); (2) inhibition of the apoptosome's activity—one example being inhibition with so-called heat shock proteins (HSP); and (3) combination of the apoptosome with other proteins to trigger oncogenes (cancer-causing genes). Questions that remain to be answered include the following: (1) Mechanistically, how do BCL-2 family proteins control release of cytochrome c and other proteins from the mitochondria? (2) How is the cytochrome c-mediated caspase activation pathway activated? and (3) How can the pathway, a promising cancer therapy target, be used to generate clinically effective anti-cancer therapies?

Hancock, Desikan, and Neill have studied the redox status of cytochrome c in relation to apoptosis progression.<sup>136</sup> First, they find that the redox status of cytochrome c, as well as the structure of cytochrome c, can change in the presence of reactive oxygen species (ROS, superoxide radical anion  $O_2^-$  or  $H_2O_2$ ) or reduced glutathione (GSH; see Figure 7.37). These authors believe that high levels of cytoplasmic GSH will maintain cytochrome c in its inactive Fe(II) state. In the presence of  $H_2O_2$ , GSH concentrations may drop, cyto-

**Figure 7.37** Glutathione.

chrome c will tend toward the Fe(III) state, and programmed cell death will proceed. The authors propose that the oxidation state of cytochrome c may be a key regulator of apoptosis.

### 7.7.5 Cytochrome c Conclusions

The preceding discussion of cytochromes c provides most detail on eukaryotic, mitochondrial cytochromes c, a small subset of this huge superfamily. Additionally, all the cytochromes c discussed in this section envelop one heme cofactor, although many cytochromes in nature contain more than one heme cofactor. Many other redox proteins contain a cytochrome c domain—a few of these mentioned here include the cytochrome bc<sub>1</sub> complex discussed in Section 7.6, cytochrome c oxidase to be discussed in Section 7.8, and cytochrome c peroxidase, discussed briefly in Section 7.7 (see especially Section 7.7.4.2).

Cytochrome c folding and electron transfer are related topics. An improperly folded protein—engendered through amino acid mutation for instance—will exhibit diminished or nonexistent electron transfer. More reports will be forthcoming from many research groups on the topics of protein folding and electron transfer between proteins. The reader should consult the literature for updates on these related hot topics. The reader is referred to the 2006 *Chemical Reviews* article from the Bertini group (reference 101), which exhaustively surveys the cytochrome c field.

## 7.8 CYTOCHROME c OXIDASE

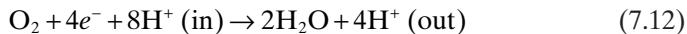
### 7.8.1 Introduction

Cytochrome c and ubiquinol oxidases are part of an enzyme superfamily coupling oxidation of ferrocyanochrome c (in eukaryotes) and ubiquinol (in prokaryotes) to the 4 e<sup>-</sup>/4 H<sup>+</sup> reduction of molecular oxygen to H<sub>2</sub>O. After this introduction, we will concentrate on the cytochrome c oxidase enzyme. The two enzymes, cytochrome c oxidase (CcO) and ubiquinol oxidase, are usually defined by two criteria: (1) The largest protein subunit (subunit I) possesses a high degree of primary sequence similarity across many species; (2) members possess a unique bimetallic center composed of a high-spin Fe(II)/(III) heme in close proximity to a copper ion. Cytochrome c oxidase (CcO) is the terminal

enzyme (Complex IV) in a sequence of membrane-bound electron transfer proteins in the mitochondrial respiratory chain. The respiratory chain, also known collectively as mitochondrial membrane oxidative phosphorylation enzymes (see Figure 1 of reference 78), provides much of the free energy needed for the life processes of aerobic organisms by coupling energy released in electron transport and proton translocation to the synthesis of adenosine triphosphate (ATP). Electron equivalents are provided by cytochrome c enzymes (see Section 7.7) that bind to CcO on the cytosolic side of the membrane (intermembrane space). During the CcO-catalyzed reaction, protons are pumped across the mitochondrial membrane barrier from the matrix side toward the cytosolic side (intermembrane space). An electrochemical potential is thus generated, causing protons to flow back into the mitochondrial matrix. The inward flux of protons is coupled to the synthesis (by ATP synthases) of ATP from ADP and inorganic phosphate ( $\text{PO}_4^{3-}$ ,  $\text{P}_i$ ). Formally, one can write the reaction catalyzed by CcO as



Proton and electron balances are given by the equation

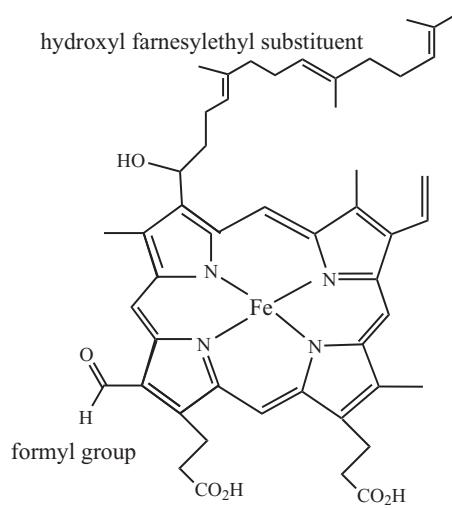


Mammalian and yeast CcO complexes contain 13 subunits with a total molecular mass of approximately 204 kDa. Three subunits are encoded by mitochondrial genes (subunits I–III), and 10 subunits are encoded by nuclear genes (subunits IV–XIII). The discussion here will concentrate on subunits I–III, but all subunits are described in more detail in reference 137b. Figure 3 of reference 138 displays an overall view of the cytochrome c oxidase complex with CcO subunits in different colors. Bacterial CcO complexes consist of two, or at most three, subunits—I, II, III—and these contain the catalytic sites for all CcOs. Subunit I, located mainly in the transmembrane region and consisting of 12 transmembrane helices, contains the ligating amino acid residues for the heme a, heme a<sub>3</sub>, and Cu<sub>B</sub> sites. Subunit II exhibits a large extramembrane domain that resides above the cytosolic surface of subunit I. Subunit II's 10-strand β barrel contains the ligating aa residues for the bimetallic Cu<sub>A</sub> site, in the cytoplasm just above the membrane surface. Subunit II also provides surface contact for cytochrome c binding through acidic amino acid residues glu109, asp119, glu127, asp139, glu157, and asp158. Subunit I also provides cytochrome c contacts via asp50 and asp221, and subunit VIb provides them via asp74 and glu78. The other subunits, IV–XIII, are believed to be important for assembly of the complex and regulation of enzyme activity. One website dedicated to cytochrome c oxidase is found at <http://www-bioc.rice.edu/~graham/CcO.html>.

Cytochrome c oxidase (CcO) has been characterized as a “proton pump” (see, for instance, M. Saratse’s 1999 *Science* magazine article<sup>78</sup>), although CcO

generates a transmembrane proton gradient by a different mechanism than cytochrome bc<sub>1</sub> (see Section 7.6). CcO's substrate, cytochrome c, discussed in Section 7.7, donates electrons to CcO on the cytoplasmic side of the mitochondrial inner membrane (the intermembrane space). Electrons are transferred from cyto c into CcO, first to a bimetallic copper site, Cu<sub>A</sub>, then through a heme iron cofactor, heme a (see Figure 7.38), then to CcO's active site, which contains a heme iron cofactor, heme a<sub>3</sub> (the subscript 3 denotes the site of dioxygen binding in multiple heme enzymes), and a copper ion, Cu<sub>B</sub>. The electrons are used to reduce O<sub>2</sub> to water. In the fully reduced form, all copper ions would be in the Cu(I) oxidation state and both irons in the Fe(II) state. In the fully oxidized form, after having transferred four electrons to dioxygen, the Cu<sub>A</sub> site contains one Cu(II) ion and one Cu(I) ion (or a mixed-valence Cu<sup>1.5+</sup>...Cu<sup>1.5+</sup> moiety; see discussion in Section 7.8.2), heme a and a<sub>3</sub> Fe(III) ions, and a Cu(II) ion at Cu<sub>B</sub>.

Protons needed for production of two water molecules by cytochrome c oxidase are taken from the mitochondrial matrix side through two channels—the so-called D and K channels. These channels—named for the conserved aspartic acid (D) and lysine (K) amino acids in their respective channels—pump one proton per electron across the membrane. In contrast to cytochrome bc<sub>1</sub>—which changes its conformation, depending on redox state and ligand binding—CcO appears to be static in its tertiary structure. The bovine enzyme, for example, shows only a minor conformational change when undergoing reduction and consequent proton pumping—one loop in subunit I facing the cytoplasm is affected. To understand the mechanistic details of the



**Figure 7.38** Heme a structure.

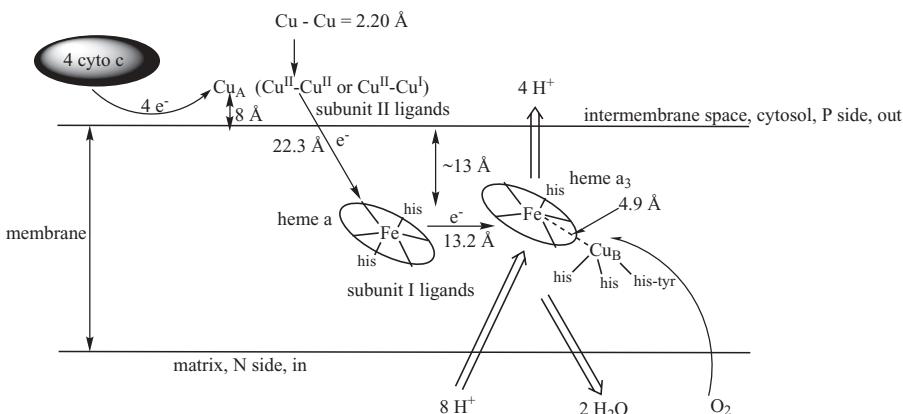
cytochrome c oxidase complex, important questions include (1) the nature of the metal binding sites—metal ions bound, oxidation–reduction behavior, and so on, (2) the metal centers’ amino acid ligands, (3) the mechanism that couples O<sub>2</sub> binding and reduction to proton pumping (translocation), and (4) the electron and proton pathways through the protein and how these are facilitated by the metal centers, channels, and amino acids.

### 7.8.2 Metal-Binding Sites in Cytochrome c Oxidase

X-ray crystallographic structures determined in the mid-1990s established the nature and positions of the metal centers in bovine heart cytochrome c oxidase.<sup>137</sup> Two heme a centers (see Figure 7.38), three copper ions, one magnesium ion, and one zinc ion were located in the crystal structures. Starting from the membrane cytoplasmic side and working toward the matrix side, one first encounters a bimetallic copper site, Cu<sub>A</sub>. The bimetallic Cu<sub>A</sub> site resides within CcO’s subunit II in the region protruding into the mitochondrial membrane’s cytosolic side (intermembrane space) approximately 8 Å above the membrane’s surface. The two copper ions in the Cu<sub>A</sub> site are held in place by six subunit II amino acid residue ligands—two cysteine sulfurs of cys196 and cys200, two imidazole nitrogens of histidine residues, his161 and his204, a methionine sulfur of met207, and the backbone carbonyl of a glutamate, glu198. The sulfur atoms of cys196 and cys200 bridge the two copper ions yielding a short Cu–Cu distance of 2.38 Å in PDB: 1OCC<sup>137b</sup> and 2.20 Å in PDB: 2OCC<sup>137c</sup> (both bovine heart enzyme complexes). The other ligands form an overall tetrahedral coordination sphere for the two Cu<sub>A</sub> ions. While the histidine and cysteine ligands have normal bond lengths and envelop the copper ions in a triangular plane, the methionine and glutamate ligands are loosely coordinated with long bond lengths (2.75 and 2.41 Å, respectively). The Cu<sub>A</sub> and heme a sites (described next) function to transport electrons from external reductant moieties, cytochrome c for one, to the catalytic site. The Cu<sub>A</sub> and heme a sites may also function as electron storage locations within the enzyme. In PDB: 1OCC the distances between the Cu<sub>A</sub> site and Fe ion in heme a is equal to 20.6 Å, while for PDB: 2OCC the Cu<sub>A</sub> site to Fe ion in heme a<sub>3</sub> is equal to 22.3 Å.

The heme a and heme a<sub>3</sub>–Cu<sub>B</sub> centers, located at the same depth within the mitochondrial membrane, are found about 13 Å below the cytosolic membrane surface and ~13 Å apart (see Figure 7.39, or Figure 2 of reference 137a).

The heme units, lying approximately perpendicular to the membrane plane, are ligated to residues within subunit I of CcO: (a) the low-spin Fe<sup>2+</sup> (S = 0) ion within heme a to his61 (subunit I, helix II) and his378 (subunit I, helix X) and (b) the high-spin Fe<sup>2+</sup> pentacoordinate (S = 2) ion within heme a<sub>3</sub> to his376 (subunit I, helix X). Heme a is bridged to heme a<sub>3</sub> through a his378–phe377–his376 linkage. Cu<sub>B</sub>, located 4.70 Å (PDB: 1OCC) or 4.85 Å (PDB: 2OCC) away from the heme a<sub>3</sub> Fe iron ion, is ligated by imidazole nitrogens of three subunit I histidine ligands: his240 (helix VI) and his290 and his291, both in a



**Figure 7.39** Schematic diagram for cytochrome c oxidase. Distances taken from bovine heart X-ray crystallographic structure (PDB: 2OCC). Entry and exit channels for dioxygen, protons, and water are schematic only. (Adapted with permission from Figure 1 of reference 138. Copyright 2004 American Chemical Society.)

nonhelical loop region between helices VII and VIII. In its fully reduced  $Cu^+$  state ( $S = 0$ ) the heme  $a_3$ -Cu<sub>B</sub> center would be EPR silent for low-spin Fe<sup>2+</sup>. In the oxidized state (Fe<sup>3+</sup>,  $S = 5/2$ , Cu<sub>B</sub><sup>2+</sup>,  $S = 1/2$ ) the two metal ions are anti-ferromagnetically coupled and again EPR silent. Heme  $a_3$  and Cu<sub>B</sub> constitute the O<sub>2</sub> binding and reduction to water site, and it was hoped that CcO X-ray crystallographic structures would show dioxygen-binding mode(s). The X-ray crystallographic structure of oxidized CcO (PDB: 1OCC) was not at high enough resolution (2.8 Å) to indicate bridging ligands between the heme  $a_3$  Fe and Cu<sub>B</sub> ions. However, the higher-resolution PDB: 2OCC structure (2.3 Å) does indicate a bridging peroxy or hydroperoxy ligand between the heme  $a_3$  Fe and Cu<sub>B</sub> ions. The positioning of this ligand has been questioned by work on model compounds of the fully oxidized state as will be discussed in Section 7.8.4. (See also Table 7.9.) As of late 2006, no other X-ray crystallographic structures have been published showing a peroxy, hydroperoxy, or superoxo ligand positioned between heme  $a_3$  Fe and Cu<sub>B</sub> ions in CcO. Recently, an X-ray crystallographic structure has been reported that indicates a hydroxide ligand or water molecule bridging between the Fe ion of heme  $a_3$  (Fe-O = 2.13 Å) and the Cu<sub>B</sub> ion (Cu-O = 2.96 Å)<sup>139</sup> (PDB: 2GSM).

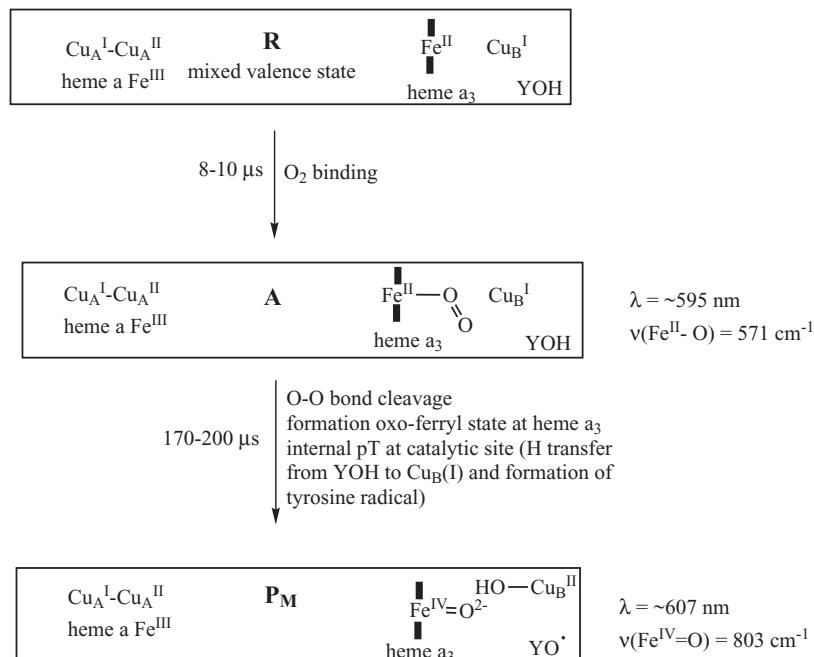
The zinc ion site, residing outside the CcO on the matrix side, was found to have tetrahedral coordination to four cysteine ligands (cys60, cys62, cys82, cys85 from subunit Vb). The polypeptide sequence from cys60 to cys82 exhibits a zinc finger motif (see Section 2.4). The magnesium ion site was less well defined in the reference 137a report, however, asp369 and his368 from subunit I, glu198 in subunit II, and several water molecules are implicated as possible magnesium ligands. In reference 137b, the magnesium ion is postulated to be part of a system/channel for removal of water from the catalytic site. In this

case, a channel of polar atoms from subunits I and II amino acid side chains would guide the water molecules to a cytosolic side (intermembrane space) opening. Overall, cytochrome c oxidase must have a pathway for electron movement and channels to move protons, dioxygen, and water molecules into and out of the enzyme complex.

### 7.8.3 Dioxygen Binding, Proton Translocation, and Electron Transport

In all cytochrome c oxidase enzymes,  $O_2$  is reduced at the bimetallic heme  $a_3$ - $Cu_B$  site. Several theories have been put forward to explain the trajectory of electrons through the enzyme. Two of these, the fully reduced and the mixed valence models, are graphically illustrated in Figure 7.40 and in Figure 11 of reference 140.<sup>140</sup>

In the fully reduced model, four electrons are transferred to dioxygen through sequential one-electron oxidations of heme  $a_3$ 's iron ion, the  $Cu_B$  ion, the heme a iron ion, and one of the bimetallic center's  $Cu_A$  ions. The sequence of electron transferal differs in the mixed valence model, and a tyrosine radical (tyr) is generated. The proposed formation of a tyrosine radical during catalytic turnover arises from the known post-translational modification in most CcO's in which a covalent bond is formed between the his240 ligand of  $Cu_B$



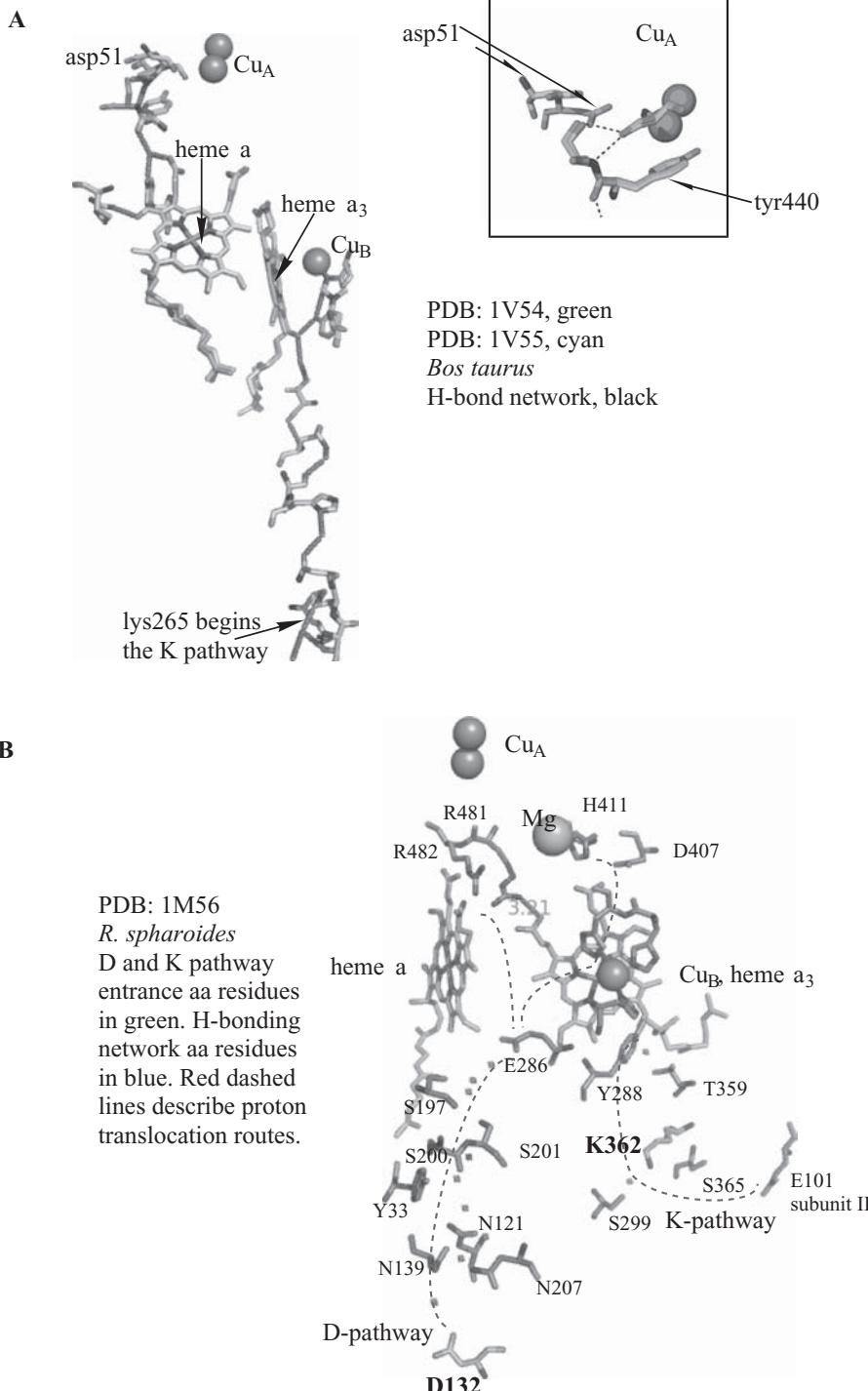
**Figure 7.40** The cytochrome c oxidase reaction cycle starting from the mixed-valence state.

and tyr244 (Y244) (his240 N<sub>ε2</sub>–tyr244 C<sub>ε2</sub> = 1.35 Å). The covalent bond would bring the tyrosine –OH group close to the bound O<sub>2</sub> at the catalytic site.<sup>141</sup> Both proposed catalytic cycles indicate formation of the ferryl–oxo species, Fe(IV)=O in heme a<sub>3</sub>, similar to that found in the myoglobin and hemoglobin cycles. One of the enigmas of CcO's O<sub>2</sub> reducing ability is the enzyme's capability to reduce O<sub>2</sub> to the level of water on a much shorter time scale (<0.2 ms in the mixed-valence case and 0.03 ms for the fully reduced enzyme) than the time it takes CcO to receive an electron from cytochrome c (5–20 ms). This difference is beneficial because it prevents the escape of partially reduced oxygen-containing species such as superoxide (O<sub>2</sub><sup>•</sup>), HO<sup>•</sup> radicals or peroxide anions (O<sub>2</sub><sup>2-</sup>). However, the rapid reduction of O<sub>2</sub> by CcO has also made it difficult to spectroscopically identify any dioxygen intermediates during the catalytic cycle.

Researchers have put forward many proposals concerning proton translocation carried out by cytochrome oxidase. First, two types of protons must be distinguished: (1) “substrate” protons used to produce water molecules and (2) “pumped” protons that reach the intermembrane space and provide the electrochemical energy for the ADP → ATP conversion. In studies of the bovine heart cytochrome c oxidase, the reference 137 researchers determined pathways involving hydrogen-bonding networks through water molecules and amino acid side chains. Figure 10 of reference 137b illustrates possible proton pumping and substrate proton paths traveling from the matrix side through the membrane and into the intermembrane (cytosolic) space. Amino acid side chains, fixed water molecules, heme a and heme a<sub>3</sub> substituents—especially the hydroxyl farnesylethyl group (see Figure 7.38)—and the magnesium ion are involved. For instance, protons to be used at the heme a<sub>3</sub>–Cu<sub>B</sub> dioxygen reduction site—substrate protons—arrive from the matrix in a hydrogen-bonding network beginning at lys265, passing through several water molecules with hydrogen bonding to amino acid side chains thr489, thr490, asn491, his256, ser255, lys319, tyr244, and Cu<sub>B</sub> ligand his240. The hemes, Cu<sub>B</sub>, Cu<sub>A</sub>, and some of the hydrogen-bonding network for PDB: 1V54 and 1V55 are visualized in Figure 7.41A. In this figure, the PDB: 1V54 structure is visualized in green stick format and PDB: 1V55 in cyan.

More recently, Yoshikawa, Tsukihara, and co-workers published a study of fully oxidized (PDB: 1V54) and fully reduced (PDB: 1V55) bovine heart cytochrome c oxidase structures.<sup>142</sup> In this study, they identified an aspartate residue, asp51, which undergoes a substantial change in position between the oxidized and reduced structures (see inset in Figure 7.41A).

An asp51–asn51 mutation of the bovine enzyme abolished its proton pumping capability but showed little change in dioxygen reduction activity. The asp51 residue resides near the enzyme surface at the interface with the cytosolic (intermembrane) side. The oxidized and reduced forms of the enzyme show substantial changes in the hydrogen-bonding interactions near the membrane surface (see Figure 2 of reference 142). The major changes in going from the oxidized to the reduced structures are to deprotonate asp51 and to



**Figure 7.41** (A) Two *B. tauri* cytochrome c oxidase molecules, (PDB: 1V54, in green, and PDB: 1V55 in cyan). Insert shows change in asp51 position. (B) The *R. sphaeroides* cytochrome c oxidase molecule (PDB: 1M56). (Adapted with permission from Figure 1b of reference 143. Copyright 2004, with permission from Elsevier.) Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft Corporation.) (See color plate)

bring the residue's side chain in contact with the intermembrane surface. The hydrogen-bonding network continues through the side-chain hydroxyl groups of ser205 and ser441 and the backbone nitrogen of tyr440, and then it continues through a fixed water molecule ( $H_2O_5$ ) in PDB: 1V54 and 1V55) to the hydroxyl group of tyr371. Water molecule ( $H_2O_5$ ) also hydrogen-bonds to the propionate group attached to pyrrole ring A of heme a. A second fixed water molecule connects tyr371 and arg38, with arg38 also hydrogen-bonded to heme a's formyl substituent. See Figure 7.41A, where some of the hydrogen-bonding network is shown as red lines. The reference 142 authors postulate a new model for the proton pumping process that is driven by a  $pK_a$  change in asp51, involvement of heme a's formyl group, and unidirectional proton transfer through peptide bonds as well as water molecules. The hydrogen-bonding network between arg38 and asp51 includes a peptide bond between tyr440 and ser441. The reference 142 authors summarize the overall process, called the H pathway, in the following manner. For the oxidized state (PDB: 1V54), (1) heme a contains the Fe(III) ion and carries a positive charge, (2) arg38 remains protonated (matrix water molecules are available to arg38 via a water channel), and (3) asp51 is protonated and inside the protein membrane. Upon reduction (PDB: 1V55), the following takes place: (1) The net positive charge on heme a is removed because it now carries a Fe(II) ion, (2) asp51 is exposed to the intermembrane space and its proton is released into this space, and (3) the capacity of the water channel increases so that water molecules are taken up from the matrix space. Upon heme oxidation, (1) asp51 moves back into the protein's interior, (2) the heme a's positive charge decreases the affinity of the heme's formyl group for the proton shared with arg38, (3) the decreased proton affinity promotes proton transfer from arg38 back to asp51, (4) deprotonated arg38 extracts a proton from a water molecule in the water channel, and (5) a hydroxide ion is quickly released, preventing reverse proton transfer (see Figure 2 and 4 of reference 142).

Published reports on electron transfer in cytochrome c oxidase usually couple electron transfer with proton translocation. Several models have been put forward that couple oxidation/reduction of the iron ion in heme a with protonation/deprotonation of a proton pumping site. In a minireview published in *FEBS Letters* in 2004, Namslauer and Brzezinski summarized the structural elements involved in electron transfer as coupled to proton translocation in cytochrome c oxidase.<sup>143</sup> These authors described pathways postulated for the bacterium *Rhodobacter sphaeroides* as determined from the X-ray crystallographic structure of wild-type (PDB: 1M56) as well as an E286Q (glu286gln) mutant (PDB: 1M57).<sup>144</sup> Two proton transfer pathways from the matrix are identified: (1) the D-pathway, named after an aspartic acid residue in subunit I (asp132, D132), and (2) the K-pathway, named after the subunit I residue lys362 (K362). PDB: 1M56, 1M57 numbering system is used in this description. Note that it differs from that of PDB: 1V54, 1V55 discussed above. The K-pathway may be used during reduction at the heme  $a_3\text{-}Cu_B$  catalytic site, perhaps prior to the binding of dioxygen.

In the D-pathway, water molecules and amino acid residues form a hydrogen-bonding network that carries protons forward to a conserved glutamic acid residue in subunit I (glu286, E286) where branching occurs. Some protons are carried from this point toward the O<sub>2</sub> catalytic reduction site—the substrate protons. Other protons, the pumped protons, are transferred toward the cytosolic side (intermembrane space), perhaps via one of the histidines ligated to Cu<sub>B</sub><sup>145</sup> or one of the D-ring propionates of heme a or heme a<sub>3</sub>.<sup>146</sup> Figure 1 of reference 144 and Figure 7.41B illustrate the pathways. In Figure 7.41B, water molecules are shown as red spheres, copper ions are shown as copper colored space-fill spheres, the magnesium ion is shown as a green space-fill sphere, and hemes and heme iron ions in stick format. D- and K-pathway entrance amino acid residues are shown in green stick format. Hydrogen bonding network amino acid residues are visualized as blue sticks. All amino acids are from subunit I, except E101 from subunit II. Red dashed lines describe postulated proton translocation routes from matrix (bottom) to heme a<sub>3</sub>–Cu<sub>B</sub> O<sub>2</sub> reduction site and exit pathways to the intermembrane space (top).

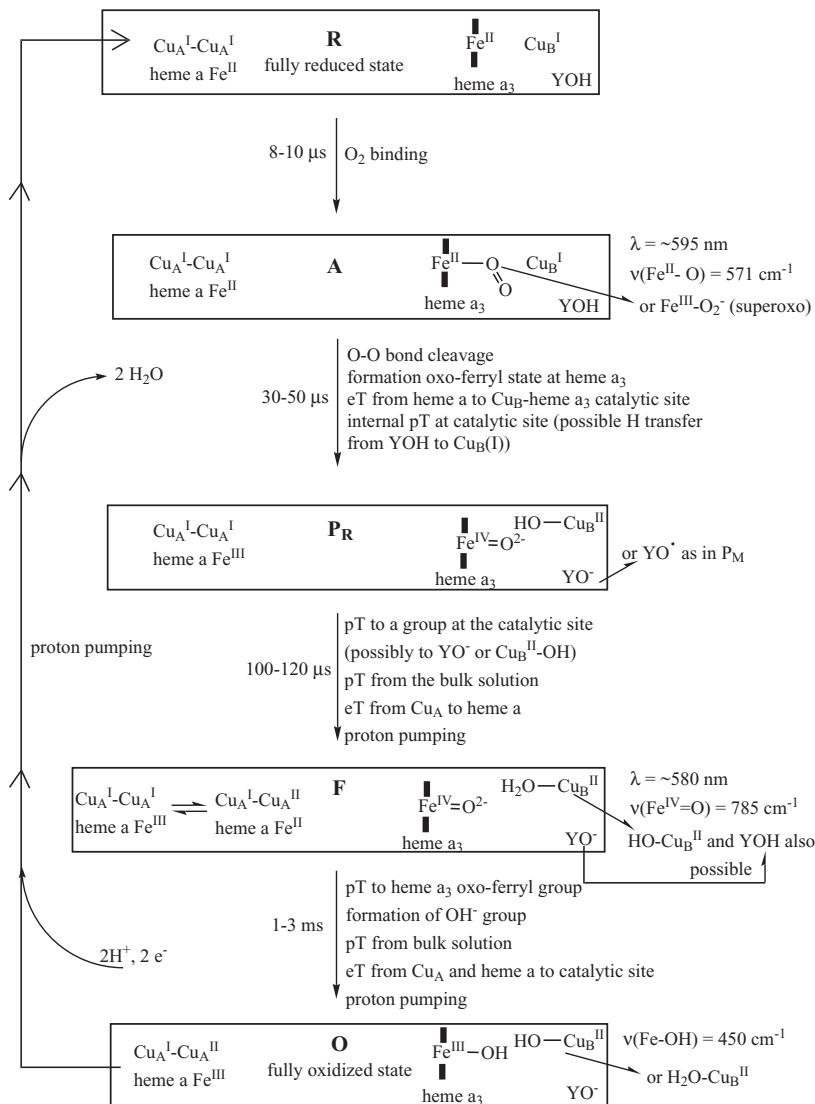
When the wild-type structure (PDB: 1M56) was compared to the E286Q mutant structure (PDB: 1M57), Iwata and co-workers<sup>144</sup> found structural rearrangements around the mutant site that could explain why the E286Q mutant does not transfer protons. First, a hydrogen bond between glu286 and met107 cannot form. Second, structural changes around arg481 and arg482 are observed that alter the interactions of these arg residues with the D-ring propionate of heme a<sub>3</sub> (the arg481–D-ring propionate salt bridge is broken resulting in an increased pK<sub>a</sub> for the propionate). Thus it appears likely that protonation or deprotonation of glu286 is important, and is linked both to turnover at the catalytic site and to the “pumped protons” pathway.

Several other models have been put forward that couple oxidation/reduction of the iron ion in heme a with protonation/deprotonation of a proton pumping site.<sup>147</sup> One model proposed that the redox state of heme a controls the position of water molecules connecting glu286, the D-ring propionate of heme a<sub>3</sub>, and the catalytic site, thus providing a method of proton gating.<sup>148</sup> All models propose that electron transfer (eT) is controlled by proton transfer (pT). For instance, heme a’s role may be to control the timing of eT from cytochrome c/Cu<sub>A</sub> to the catalytic site so that eT is prevented until the key residue, glu286, is reprotonated.

Figures 7.42 and 7.40 illustrate two schemes describing proton translocation and electron transfer within cytochrome c oxidase. Figure 7.42 illustrates the cycle beginning with fully reduced enzyme—that is, Cu<sub>A</sub><sup>I</sup>–Cu<sub>B</sub><sup>I</sup>, with both hemes carrying Fe<sup>II</sup> ions and Cu<sub>B</sub><sup>I</sup>. Figure 7.40 begins the cycle in a so-called mixed valence state—that is, Cu<sub>A</sub><sup>I</sup>–Cu<sub>A</sub><sup>II</sup>, with heme a carrying an Fe<sup>III</sup> ion and heme a<sub>3</sub> carrying an Fe<sup>II</sup> ion and Cu<sub>B</sub><sup>I</sup> (changes from the fully reduced form in bold). Similar schemes are found in references 138, 140, and 143. The reference 140 authors describe both fully reduced and mixed valence possibilities in their Figure 11.

The mixed valence state stores only half the electrons necessary to reduce one molecule of  $O_2$  to water and therefore the reaction stops at the  $\mathbf{P}_M$  state, halfway through the catalytic cycle. Model compounds synthesized in this state provide convenient conditions for enzyme mechanistic studies as will be described in Section 7.8.4. The fully reduced and mixed valence pathways have similar  $\mathbf{A}$  states at the heme  $a_3$ – $Cu_B$  catalytic site after addition of substrate dioxygen. Evidence includes a UV-visible absorption band at 595 nm and a resonance Raman band at  $571\text{ cm}^{-1}$  assigned to  $v(\text{Fe}^{\text{II}}-\text{O})$ . The pathways diverge at that point to produce the  $\mathbf{P}_R$  (fully reduced) or  $\mathbf{P}_M$  (mixed valence) states. The  $\mathbf{P}_M$  state is produced at a ~4–5 times slower rate than  $\mathbf{P}_R$ . Both  $\mathbf{P}_R$  and  $\mathbf{P}_M$  show indications of the ferryl–oxo intermediate at heme  $a_3$ —resonance Raman band at  $803\text{ cm}^{-1}$  assigned to  $v(\text{Fe}^{\text{IV}}=\text{O})$  for  $\mathbf{P}_M$ . In addition, experimental evidence suggests that the  $\mathbf{P}_M$  state contains a tyrosine radical,  $\text{YO}^\bullet$ , and it is postulated that the radical arises from tyr244, which is covalently bonded to  $Cu_B$ 's his240 ligand (bovine enzyme numbering). The fully reduced pathway next undergoes a pH-dependent step to state  $\mathbf{F}$  that exhibits its  $v(\text{Fe}^{\text{IV}}=\text{O})$  resonance Raman band at  $785\text{ cm}^{-1}$ . As shown in Figure 7.42, it is thought that an equilibrium mixture of two redox states at the heme  $a$ – $Cu_A$  sites exists in state  $\mathbf{F}$ . An additional proton has been added to the catalytic site in state  $\mathbf{F}$ ; and although this is shown being added to the  $Cu_B$ –OH moiety of state  $\mathbf{A}$ , the location of this proton has not been established experimentally. The rates of heme  $a \rightarrow Cu_A$  electron transfer and the state  $\mathbf{P}_R \rightarrow$  state  $\mathbf{F}$  transition are pH-dependent, and the step is accompanied by proton pumping (translocation). The fourth electron enters to form the  $\mathbf{O}$  state with its ferric hydroxide intermediate characterized by time-resolved resonance Raman spectroscopy to have a  $v(\text{Fe}-\text{OH})$  stretching mode at  $450\text{ cm}^{-1}$ . The  $\mathbf{F} \rightarrow \mathbf{O}$  transition is pH-dependent and accompanied by proton pumping. What happens next in the cycle must be further protonation and release of  $H_2O$ , but the exact mechanism for these processes is not resolved at the time of this writing.

If one can make overall conclusions about the cytochrome c oxidase cycle, the reference 143 authors do so. They conclude that efficient proton pumping into the cytosolic space (P-side of the membrane) in CcO relies on (1) rapid proton uptake through the D-pathway from the matrix (N-side of the membrane) and (2) finely adjusted  $pK_a$  values of proton shuttling groups—for example, glu286, D-ring propionate, and others within the enzyme complex. Conclusions about the conformation or oxidation state of dioxygen at the heme  $a_3$ – $Cu_B$  catalytic site are more difficult. Probably the presence of the peroxy ligand can be dismissed as a long-lived intermediate, although it appears in many of the cytochrome c oxidase model complexes to be discussed in Section 7.8.4. The dioxygen superoxo ligand intermediate has a greater probability of existence, and the ferryl iron–oxo intermediate has exhibited its spectroscopic signature in many experimental studies. Whether or not the synthesis and investigation of cytochrome c oxidase model compounds can help answer these questions is the topic of the next section.



**Figure 7.42** The cytochrome c oxidase reaction cycle starting from the fully reduced state.

#### 7.8.4 Cytochrome c Oxidase Model Compounds and Associated Analytical Techniques

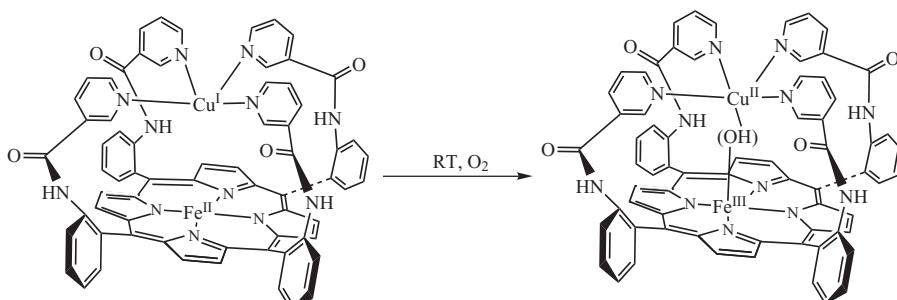
In 2004, Kenneth Karlin and co-workers published an extensive review of cytochrome c oxidase model compounds.<sup>138</sup> Hundreds of CcO models exist, with more being synthesized and studied into the current time. These can be broken down into a number of categories as Karlin's group does in the 2004

*Chemical Reviews* paper.<sup>138</sup> We can only discuss a small number of these here, but some important categories are (1) synthetic Fe(II)–Cu(I) complexes and their reactions with O<sub>2</sub>, (2) oxidized heme–copper models (Fe(III)–X–Cu(II) complexes, where X equals oxo- and hydroxo-bridged complexes, cyanide-bridged complexes, or other X-bridged complexes), (3) crosslinked histidine–tyrosine residues at the heme–copper center, and (4) Cu<sub>A</sub> site synthetic models.

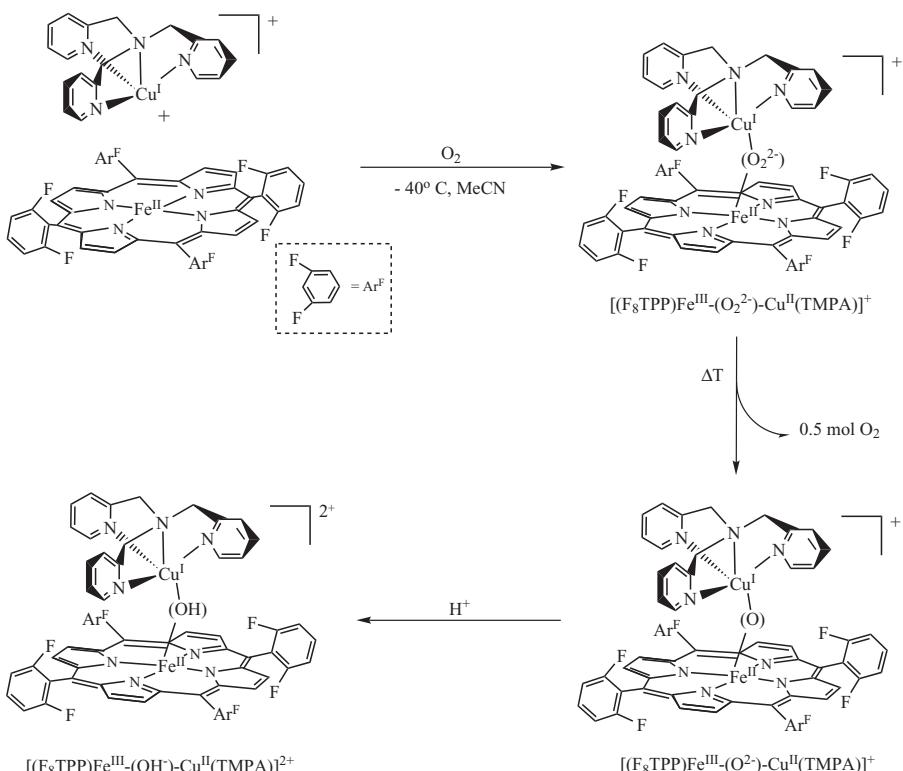
In category (1) above, the earliest synthetic Fe(II)–Cu(I) complexes, prepared beginning in 1981,<sup>149</sup> were adapted from the “picket-fence porphyrin” concept pioneered by the J. P. Collman laboratory.<sup>150</sup> The Gunter and Mander complexes<sup>149</sup> reacted with dioxygen at room temperature to yield oxidized compounds such as [(P)Fe<sup>III</sup>–OH–Cu<sup>II</sup>(N<sub>4</sub>)], P = porphyrin ligand, N<sub>4</sub> = copper ion ligand system. (See Figure 7.43.)

Research on Fe(II)–Cu(I) complexes has continued through the years in research groups headed by J. P. Collman, Y. Naruta, B. Boitrel, and Kenneth Karlin. One system that has been thoroughly characterized by Karlin and co-workers involves the reaction of a [(TMPA)Cu(I)(RCN)]<sup>+</sup> complex with a [(F<sub>8</sub>TPP)Fe(II)] complex. The TMPA ligand is *tris*(2-pyridylmethyl)amine, RCN = MeCN = acetonitrile, and F<sub>8</sub>TPP = *tetrakis*(2,6-difluorophenyl)porphyrinate (–2 charge).<sup>151</sup> These researchers found that reaction with dioxygen in acetonitrile at –40°C resulted in formation of the heterobinuclear peroxy complex [(F<sub>8</sub>TPP)Fe<sup>III</sup>–(O<sub>2</sub><sup>–</sup>)–Cu<sup>II</sup>(TMPA)]<sup>+</sup> rather than homobinuclear heme–peroxy or a binuclear copper–peroxy complexes. The reaction is shown in Figure 7.44.

The product was identified by a number of spectroscopic methods. Dioxygen uptake was measured by spectrophotometric titration. MALDI–TOF–MS (matrix-assisted laser desorption/ionization–time of flight–mass spectrometry), an MS method particularly suited to determining molecular masses of biopolymers and synthetic materials with relative masses up to several hundred kilodaltons, determined that the product contained stoichiometric amounts of the heme starting material, the copper complex, and dioxygen in a 1:1:1 ratio.



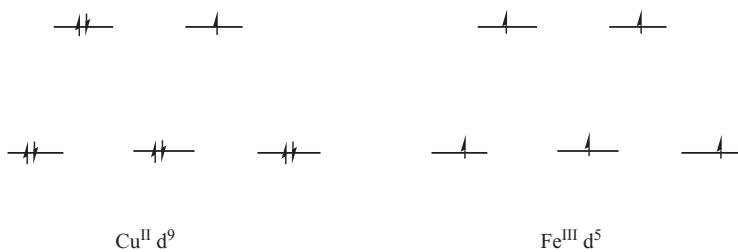
**Figure 7.43** Cytochrome c oxidase model complexes, [(Porphyrin)Fe<sup>III</sup>–OH–Cu<sup>II</sup>(N<sub>4</sub>)], as synthesized in reference 149.



**Figure 7.44** Cytochrome c oxidase model complexes,  $[(\text{F}_8\text{TPP})\text{Fe}^{\text{III}}-(\text{O}_2^{\cdot-})-\text{Cu}^{\text{II}}(\text{TPMA})]^+$ , as synthesized in reference 151.

See 3.7.3 for a discussion of the MALDI-TOF mass spectrometric method. Resonance Raman spectroscopy identified a peroxy stretching vibration  $\nu(\text{O}-\text{O})$  at  $808\text{ cm}^{-1}$ . The resonance was confirmed by testing an  $^{18}\text{O}_2^{\cdot-}$  complex, resulting in a  $-46\text{ cm}^{-1}$  change in the vibration frequency. Solution magnetic moment measurements and  $^1\text{H}$  NMR indicated an  $S = 2$  spin state for the complex that would arise from antiferromagnetic coupling between high-spin Fe(III) and Cu(II) centers (see Figure 7.45).

Mössbauer spectroscopy yielded data consistent with a high-spin Fe(III) ion possessing a peroxide ligand:  $\Delta E_Q = 1.14\text{ mm/s}$  and  $\delta = 0.57\text{ mm/s}$ . Stopped-flow UV-visible spectroscopy indicated that a heme-superoxide intermediate  $[(\text{S})-(\text{F}_8\text{TPP})\text{Fe}^{\text{III}}-(\text{O}_2^{\cdot-})]$  ( $\lambda_{\text{max}} = 537\text{ nm}$ , S = solvent) is generated within  $\sim 1\text{ ms}$  (the mixing time) before the  $[(\text{F}_8\text{TPP})\text{Fe}^{\text{III}}-(\text{O}_2^{\cdot-})-\text{Cu}^{\text{II}}(\text{TPMA})]^+$  complex ( $\lambda_{\text{max}} = 556\text{ nm}$ ) forms. The same  $[(\text{F}_8\text{TPP})\text{Fe}^{\text{III}}-(\text{O}_2^{\cdot-})-\text{Cu}^{\text{II}}(\text{TPMA})]^+$  complex has been generated from alternate starting materials:  $[(\text{F}_8\text{TPP})\text{Fe}^{\text{III}}-(\text{O}_2^{\cdot-})]^{1-}$  and  $[\text{Cu}^{\text{II}}(\text{TPMA})(\text{MeCN})]^{2+}$ .<sup>152</sup> More recently, Karlin, E. I. Solomon, and co-workers have studied the geometric and electronic structure of  $[(\text{F}_8\text{TPP})\text{Fe}^{\text{III}}-(\text{O}_2^{\cdot-})-\text{Cu}^{\text{II}}(\text{TPMA})](\text{ClO}_4)$  using Cu and Fe K-edge EXAFS

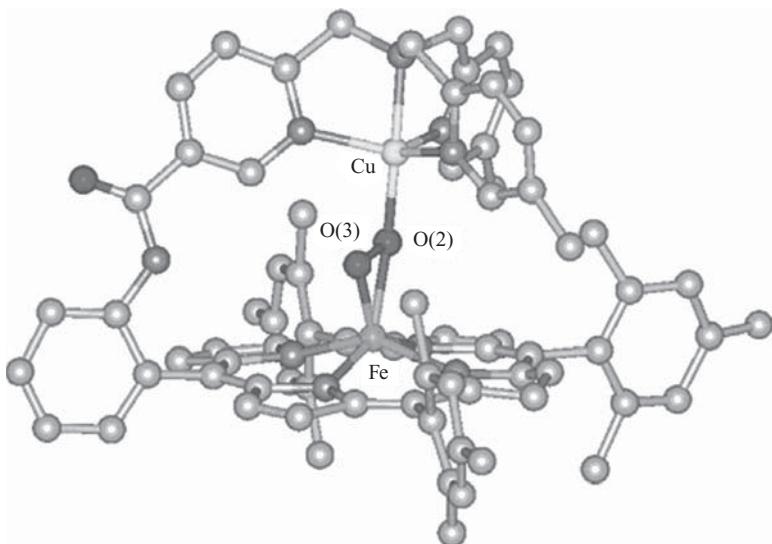


**Figure 7.45** Electron configurations for high-spin Cu(II),  $d^9$ , and Fe(III),  $d^5$ .

spectroscopy and density functional theory.<sup>153</sup> Figure 2 of reference 153 illustrates a consensus  $O_2$  reduction mechanism involving heme  $a_3$  and  $Cu_B$ . The scheme involves the  $Cu_B$  histidine ligand and its covalently attached tyrosine side chain whose hydroxyl group successively becomes a radical (state **P<sub>M</sub>**) and is protonated (state **F**), deprotonated (state **O**), and then protonated (states **R** and **A**) as the cycle resumes. Proton pumping (from matrix to the cytosol) occurs in transitions from states **P<sub>M</sub>** to **F**, **F** to **O**, and **O** to **R**. State **A** has also been formulated as an  $Fe^{III}-O_2^-$  (superoxo) intermediate.<sup>154</sup> (See Figures 7.40 and 7.42.)

In CcO enzyme X-ray crystallographic structures, the heme  $a_3-Cu_B$  centers are separated by 4.4–5.3 Å, depending on the protein derivative and its oxidation state. This distance would preclude strong antiferromagnetic coupling between the high-spin iron and copper ion centers. However, structures that include a peroxide bridge (PDB: 2OCC and the model compounds being discussed here and others) do report strong antiferromagnetic coupling between the two metal centers. One model compound containing a peroxy bridge has had its crystal structure determined.<sup>155</sup> The binuclear heme  $a_3-Cu_B$  peroxy-bridged center of a tethered  $[(TMP)Fe^{III}-(O_2^{2-})-Cu^{II}(5MeTPA)]^+$  shows a  $\mu-\eta^1:\eta^2Fe^{III}-(O_2^{2-})-Cu^{II}$  coordination mode. The all-in-one ligand TMP-5MeTPA (TMP-5MeTPA=6-[bis-(5-methylpyridin-2-ylmethyl)amino]methyl]-N-{2-[tris(2,4,6-trimethylphenyl)porphyrinate]phenyl}nicotinamide) forms the tethered complex with the iron and copper ions. The peroxy anion is bound end-on to the copper ion ( $\eta^1$ ) and side-on to the high-spin iron ion ( $\eta^2$ ). The high-spin heme  $a_3$   $Fe^{III}$  center sits 0.534 Å above the plane formed by the porphyrinate nitrogen ligands. (See Figure 7.46.)

In reference 153, for the untethered complex  $[(F_8TPP)Fe^{III}-(O_2^{2-})-Cu^{II}(TPMA)]^+$ , the authors use density functional theory (DFT) to involve the peroxy ligand in the observed metal center antiferromagnetic coupling—for the  $S_{TOT} = 2$  system,  $-J > 200\text{ cm}^{-1}$ . The spin-unrestricted DFT calculations for the  $S_{TOT} = 2$  spin state, performed on a  $[(P)Fe^{III}-(O_2^{2-})-Cu^{II}(TPMA)]^+$  model, P = porphyrinate, indicate that the peroxy  $\pi_\sigma^*$  orbital donates electron density into the half-occupied  $d_z^2$   $Cu^{II}$  orbital forming an  $\eta^1$  bond. For the  $Fe^{III}$  ion–peroxy  $\eta^2$  bond, there are two major interaction components: (1) The peroxy  $\pi_\sigma^*$  orbital donates electron density into the  $d_{xz}$   $Fe^{III}$



**Figure 7.46** Crystal structure of tethered  $[(\text{TMP})\text{Fe}^{\text{III}}-(\text{O}_2^-)-\text{Cu}^{\text{II}}(\text{5MeTPA})]^+$ . (Reprinted with permission from Figure 7 of reference 138. Copyright 2004 American Chemical Society.)

orbital forming a  $\sigma$  bonding interaction; and (2) the peroxide  $\pi_v^*$  orbital donates electron density into the  $d_{xy}$   $\text{Fe}^{\text{III}}$  orbital, forming a  $\delta$ -bonding interaction. The peroxide  $\pi_o^*$  orbital's interaction with both the half-occupied  $d_z^2$   $\text{Cu}^{\text{II}}$  orbital ( $\eta^1$ ) and the  $d_{xz}$   $\text{Fe}^{\text{III}}$  orbital ( $\eta^2$ ) forms a “superexchange” pathway for strong antiferromagnetic coupling between the metal centers. The two highest-occupied orbitals of the peroxide ion would normally be the degenerate  $\pi^*$  orbitals, but these split under the influence of the metal centers. The  $\pi_o^*$  orbital resides in the  $\text{Fe}-\text{O}_2-\text{Cu}$  plane, while the  $\pi_v^*$  orbital is perpendicular to this plane. The DFT calculations were also used to determine bond distances and angles for fully optimized and constrained ( $\text{Fe}\cdots\text{Cu} = 3.7\text{\AA}$ ) $[(\text{P})\text{Fe}^{\text{III}}-(\text{O}_2^-)-\text{Cu}^{\text{II}}(\text{TPMA})]^+$  model structures. Data are reported in Table 7.9.

The  $[(\text{F}_8\text{TPP})\text{Fe}^{\text{III}}-(\text{O}_2^-)-\text{Cu}^{\text{II}}(\text{TPMA})]^+$  complex was also studied using Cu and Fe K-edge extended X-ray absorption-fine structure (EXAFS) spectroscopy in the solid state and in acetonitrile solution by reference 153 authors. The data were fit with a  $\text{Fe}\cdots\text{Cu}$  distance of  $\sim 3.72\text{\AA}$ . Very little change in bond distances or angles were found between the solid versus solution EXAFS results. In general, there is good agreement between the found X-ray crystallographic bond distances and angles (for Naruta's tethered  $[(\text{TMP})\text{Fe}^{\text{III}}-(\text{O}_2^-)-\text{Cu}^{\text{II}}(\text{5MeTPA})]^+$  complex, reference 155) and those found for the EXAFS experiments as well as the calculated values from DFT calculations. Differences are noted in the  $\text{Fe}-\text{Cu}$  distances, with the fully optimized DFT calculated structure ( $\text{Fe}-\text{Cu} = 4.006\text{\AA}$ ) agreeing better with the

X-ray crystallographic data ( $\text{Fe}-\text{Cu} = 3.916 \text{ \AA}$ ), and both being  $\sim 0.3 \text{ \AA}$  longer than that observed for the  $[(\text{F}_8\text{TPP})\text{Fe}^{\text{III}}-(\text{O}_2^-)-\text{Cu}^{\text{II}}(\text{TMPA})]^+$  (EXAFS data for solid and in solution). The authors believe that a combination of steric and electronic factors is responsible for the differences observed. In comparing their results with those of the CcO enzyme itself, the reference 155 authors note that the peroxy linkage probably does not form in the enzyme because the  $\text{Fe}-\text{Cu}$  distance is too long (4.4–5.3 Å), although one X-ray structure shows a peroxy ligand (PDB: 2OCC). Differences between the model compounds discussed here and CcO protein are (1) the presence of a proximal (opposite the  $\text{O}_2$ -binding site) axial histidine ligand on heme  $a_3$  in CcO, (2) the tyrosine residue covalently bound to a  $\text{Cu}_B$  histidine ligand (which may donate an  $\text{H}^+$  during the catalytic cycle), and (3) the coordination environment of the copper center that may cause electronic structural differences and thus differences in reactivity in the  $\text{Fe}^{\text{III}}-\text{O}_2^--\text{Cu}^{\text{II}}$  center. In spite of these shortcomings, the models do provide a good starting point for understanding the activation and cleavage of the dioxygen bond in cytochrome c oxidase.

Biomimetic efforts have also focused on the oxidized or “resting state” of the heme  $a_3-\text{Cu}_B$  binuclear center in which the porphyrinate iron ion is in the Fe(III) oxidation state and the  $\text{Cu}_B$  ion exists as Cu(II). For instance,  $\text{Fe}^{\text{III}}-\text{X}-\text{Cu}^{\text{II}}$  synthetic model compounds, where  $\text{X} = \text{oxo-}$  or hydroxo- bridged complexes, could represent the CcO “resting state” or catalytic turnover intermediates after dioxygen O–O bond cleavage. Cyanide-bridged compounds could yield information on the toxicity of the  $\text{CN}^-$  ligand. X-ray crystallographic studies on native CcO enzymes over the last 10 years have proposed various connecting ligands for the oxidized enzyme’s heme  $a_3-\text{Cu}_B$  binuclear center. At least one of these—PDB: 2OCC for the oxidized “resting state” bovine heart enzyme—proposed a peroxide ligand connection as has been discussed in Sections 7.8.2 and 7.8.3. The reference 138 authors consider the peroxide bridge unlikely because, for one reason, the metal–O and O–O bond distances in the bovine heart enzyme are considerably longer than that found for similar synthetic complexes (see reference 138, Table 2; also see Table 7.9). Typical bond distances for  $\text{O}_2$ ,  $\text{O}_2^-$  (superoxo), and  $\text{O}_2^{2-}$  (peroxy) are 1.21, 1.28–1.30, and 1.40–1.50 Å, respectively, while the O–O bond distance reported for PDB: 2OCC is equal to 1.63 Å.

Not discussed in the previous section were bovine heart crystal structures featuring heme  $a_3-\text{Cu}_B$  with CO (carbon monoxide)- and  $\text{N}_3^-$  (azide)-bound ligands, although these are common ligands in synthetic models for the oxidized center. (See reference 137c.) Other CcO enzyme X-ray structures of *P. dentrificans* (PDB: 1AR1),<sup>156</sup> *R. sphaeroides* (PDB: 1M56, 1M57),<sup>144</sup> and *T. thermophilus* (PDB: 1EHK)<sup>157</sup> have proposed various connectors:  $\mu$ -oxo and  $\mu$ -hydroxo species, water molecules, or some combination of these for the heme  $a_3-\text{Cu}_B$  binuclear center. Because of their atomic resolution, the X-ray structures are not completely clear as to whether the connector species do, in fact, connect or are simply the nearby ligands of the two metal centers. (See proposed **R**, **A**, **P**, **F** and **O** states as shown in Figures 7.40 and 7.42.) The ref-

TABLE 7.9 Selected Bond Distances and Angles for Fe<sup>III</sup>-(O<sub>2</sub><sup>2-</sup>)<sub>2</sub>-Cu<sup>II</sup> Complexes

| Bond (Å) or Angle (°)        | CcO (PDB: 2OCC) <sup>a</sup>                      | Model Compound Crystal Structure <sup>b</sup> | DFT, Fully Optimized <sup>c</sup> | DFT, Constrained <sup>d</sup> | EXAFS Cu or (Fe) Edge <sup>e</sup> |
|------------------------------|---|---|-----------------------------------|-------------------------------|------------------------------------|
| O <sub>2</sub> -Binding Mode | Trans-μ-1,2                                       | μ-η <sup>1,2</sup>                            | μ-η <sup>1,2</sup>                | μ-η <sup>1,2</sup>            | μ-η <sup>1,2</sup>                 |
| Fe-Cu                        | 4.85  | 3.916   | 4.006                             | 3.700 <sup>d</sup>            | 3.68 (3.71) <sup>e</sup>           |
| O2-O3                        | 1.63 (O1-O2) <sup>a</sup>                         | 1.460 (6)                                     | 1.423                             | 1.446                         |                                    |
| Cu-O2                        | 2.02 (Cu-O2) <sup>a</sup>                         | 1.915 (5)                                     | 1.959                             | 1.854                         | 1.87                               |
| Fe-O2                        | 2.28 (Fe-O1) <sup>a</sup>                         | 2.031 (4)                                     | 2.072                             | 1.900                         |                                    |
| Fe-O3                        | 3.52 (Fe-O2) <sup>a</sup>                         | 1.890 (6)                                     | 1.880                             | 1.891                         | (1.94) <sup>e</sup>                |
| Cu-O-Fe                      | 119.6 (O2) <sup>a</sup> , 157.6 (O1) <sup>a</sup> | 166.3 (3)                                     | 170.1                             | 160.6                         | 149.9                              |
| Cu-O2-O3                     | 93.1 (Cu-O2-O1) <sup>a</sup>                      | 103.4 (4)                                     | 109.0                             | 101.0                         |                                    |
| Fe-O3-O2-Cu                  | 126.8 (Fe-O1-O2-Cu) <sup>a</sup>                  | 178.0   | -175.9                            | -163.8                        |                                    |

<sup>a</sup> Reference 137c. In PDB: 2OCC, the oxygen atom directly bonded to heme  $\alpha_3$  Fe is labeled O1 and the oxygen atom directly bonded to CuB as O2.

<sup>b</sup> Distances and angles for [(TPP)Fe<sup>III</sup>-(O<sub>2</sub><sup>2-</sup>)<sub>2</sub>-Cu<sup>II</sup>(5MeTPA)]<sup>+</sup> from reference 155.

<sup>c</sup> DFT = density function theory calculation for [(P)Fe<sup>III</sup>-(O<sub>2</sub><sup>2-</sup>)<sub>2</sub>-Cu<sup>II</sup>(TPPA)]<sup>+</sup> and EXAFS data for [(F<sub>8</sub>TPP)Fe<sup>III</sup>-(O<sub>2</sub><sup>2-</sup>)<sub>2</sub>-Cu<sup>II</sup>(TPMA)]<sup>+</sup> from reference 153.

<sup>d</sup> Fe-Cu distance constrained at 3.7 Å.

<sup>e</sup> EXAFS Fe edge.

erence 138 authors collect ligand and crystallographic data for the CcO enzyme heme  $a_3$ -Cu $B$  binuclear centers in their Table 5. Table 6 of this reference details bond distance and angle data for numerous Fe $^{III}$ -(X)-Cu $^{II}$  model complexes. Some of these data are collected in Table 7.10.

Chemists have studied a wide variety of Fe $^{III}$ -X-Cu $^{II}$  synthetic model compounds, where X may be an exogenous ligand such as fluoride (F $^-$ ), cyanide (CN $^-$ ), and formate (HCOO $^-$ ) or naturally occurring X groups such as hydroxide (OH $^-$ ), oxide (O $^{2-}$ ), carboxylate (RCOO $^-$ ), chloride (Cl $^-$ ), sulfide (S $^{2-}$ ), cysteinato (S-ligand atom of cysteine), or imidazolate (side chain N-ligand atom of histidine) groups. One goal of the synthetic model research has been to understand the magnetic properties of the CcO enzyme: (1) a strong antiferromagnetic coupling model for the high-spin  $S = 5/2$  ferric heme  $a_3$  with the  $S = 1/2$  Cu $B$  center to produce the  $S = 2$  ground state for the heme  $a_3$ -Cu $B$  center with  $J$  values exceeding 100 cm $^{-1}$ ; (2) evidence in similar systems for a weak coupling model with  $J < 4$  cm $^{-1}$ ; and (3) observation of an active EPR spectrum for some CcO systems.

Space considerations limit the number of Fe $^{III}$ -X-Cu $^{II}$  synthetic model compounds that we can discuss here. We will concentrate on model compounds featuring oxide (O $^{2-}$ ) or hydroxide (OH $^-$ ) bridges. First, the heterobinuclear peroxy complex  $[(F_8TPP)Fe^{III}-(O_2^{2-})-Cu^{II}(TMPA)]^+$ , prepared according to the self-assembly strategy as described above, will lose one-half mole of dioxygen in a slow disproportionation reaction and convert to the  $\mu$ -oxo complex  $[(F_8TPP)Fe^{III}-(O^{2-})-Cu^{II}(TMPA)]^+$  when heated to room temperature (see Figure 7.44).<sup>151</sup> Subsequent addition of one equivalent of acid yields the  $\mu$ -hydroxo complex  $[(F_8TPP)Fe^{III}-(OH^-)-Cu^{II}(TMPA)]^{2+}$ . Karlin and co-workers have used heterobinucleating ligands (ligands for two metal centers connected by a tether) in formation of  $\mu$ -oxo and  $\mu$ -hydroxo complexes.<sup>159</sup> A portion of one such reaction scheme is shown in Figure 7.47. The  $^6L$  (connected to the 6-position of a TMPA pyridine arm) and  $^5L$  (connected to the 5-position) ligands each have a tetradeятate TMPA ligating moiety covalently attached to the edge of a porphyrin. Starting with the  $^6L$  ligand, one can form the di-iron product  $\mu$ -oxo product  $[(^6L)Fe^{III}-O-Fe^{III}-Cl]^+$ , selectively remove one iron to form  $[(^6L)Fe^{III}-OH]$ , and then add copper ion to form  $[(^6L)Fe^{III}-O-Cu^{II}]^+$ . The same procedure can be repeated with the  $^5L$  ligand. The two heterobinuclear products behave differently when acidified. The  $[(^6L)Fe^{III}-O-Cu^{II}]^+$  reaction with H $^+$  produces  $[(^6L)Fe^{III}-OH-Cu^{II}]^{2+}$ , whereas acidification of  $[(^5L)Fe^{III}-O-Cu^{II}]^+$  results in a broken bridge. The reference 159 researchers ascribe the differing reactivity to the Fe-O-Cu bond angles: approximately linear for  $[(^6L)Fe^{III}-O-Cu^{II}]^+$  (and for  $[(F_8TPP)Fe^{III}-(O^{2-})-Cu^{II}(TMPA)]^+$  as well) and substantially bent— $141 \pm 6^\circ$  by EXAFS determination—for  $[(^5L)Fe^{III}-O-Cu^{II}]^+$ . The linear nature of the Fe-O-Cu connection for  $[(^6L)Fe^{III}-O-Cu^{II}]^+$  has been confirmed by an X-ray crystallographic structure determination.<sup>158</sup> The researchers point out that small changes in the structures of  $[(^6L)Fe^{III}-O-Cu^{II}]^+$  versus  $[(^5L)Fe^{III}-O-Cu^{II}]^+$  change the acid-base behavior of these model complexes. The differing behavior of the

**TABLE 7.10 Selected Bond Distances and Angles for Oxidized CcO Enzymes and Fe<sup>III</sup>-(O<sub>2</sub><sup>2-</sup>)-Cu<sup>II</sup> or Fe<sup>III</sup>-(OH<sup>-</sup>)-Cu<sup>II</sup> Model Complexes**

| CcO Crystal Structure or Model Compound   | Bridging Ligand   | Fe...Cu Distance (Å) | Fe...X <sup>a</sup> Distance (Å) | Cu...Y(X) <sup>a</sup> Distance (Å) | Fe-X-Cu Angle (°)                                |
|---|---|----------------------|----------------------------------|-------------------------------------|--|
| PDB: 2OCC (bovine heart)  | O <sub>2</sub> <sup>2-</sup> (possible)                 | 4.85                 | 2.28                             | 2.02                                | 119.6 (O <sub>2</sub> ), 157.6 (O <sub>3</sub> ) |
| <i>R. sphaeroides</i> <sup>b</sup> (PDB: IM56, 1M57)  | OH <sup>-</sup> or H <sub>2</sub> O                     | 4.8                  | 3.6                              | 2.0                                 |  |
| <i>T. thermophilus</i> <sup>c</sup> (PDB: 1EHK)   | O <sup>2-</sup> , OH <sup>-</sup> , or H <sub>2</sub> O | 4.4                  | 2.3                              | 2.3                                 |  |
| [F <sub>8</sub> TPP]Fe <sup>III</sup> -(O <sup>2-</sup> )-Cu <sup>II</sup> (TMPA)] <sup>+</sup> | O <sup>2-</sup>   | 3.60                 | 1.74                             | 1.86                                | 178  |
| [F <sub>8</sub> TPP]Fe <sup>III</sup> -(OH <sup>-</sup> )-Cu <sup>II</sup> (TMPA)] <sup>d</sup> | OH <sup>-</sup>   | 3.66                 | 1.87                             | 1.89                                | 157  |
| [ <sup>6</sup> L]Fe <sup>III</sup> -O-Cu <sup>II</sup> ] <sup>e</sup>                           | O <sup>2-</sup>   | 3.59                 | 1.75                             | 1.85                                | 171  |
| [ <sup>6</sup> L]Fe <sup>III</sup> -O-Cu <sup>II</sup> ] <sup>e,f</sup>                         | O <sup>2-</sup>   | 3.40                 | 1.77                             | 1.84                                | 141  |

<sup>a</sup>X equals the possible bridging ligands O<sub>2</sub><sup>2-</sup>, O<sup>2-</sup>, OH<sup>-</sup>, or H<sub>2</sub>O.

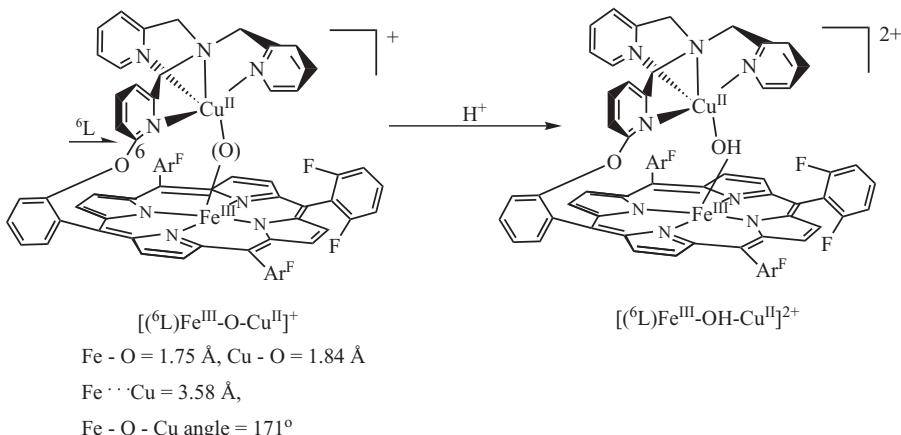
<sup>b</sup>Reference 144.

<sup>c</sup>Reference 157.

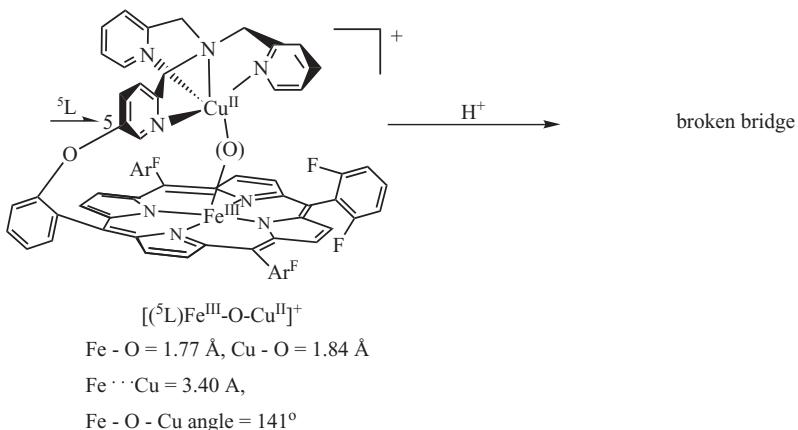
<sup>d</sup>Bond distances and angles determined by X-ray absorption spectroscopy (XAS) methods. See reference 138 and Figures 7.44 and 7.47.

<sup>e</sup>References 138 and 159.

A



B



**Figure 7.47** Acid–base behavior of (A)  $[(^6\text{L})\text{Fe}^{\text{III}}\text{-O-Cu}^{\text{II}}]^+$  versus (B)  $[(^5\text{L})\text{Fe}^{\text{III}}\text{-O-Cu}^{\text{II}}]^+$  as described in reference 159.

models may have analogies to CcO enzyme behavior, because the basicity of  $\text{O}_2$ -derived intermediates may be critical to proton translocation in the enzyme.

Finally, in this section we discuss the model compounds for the Type III binuclear  $\text{Cu}_\text{A}$  center. Electron transfer (eT) to and from  $\text{Cu}_\text{A}$  is remarkably fast, considering the distances to be traveled. Intermolecular eT from cytochrome c  $\rightarrow \text{Cu}_\text{A}$ , over a distance of  $\sim 18\text{--}20\text{ }\text{\AA}$ , has a rate constant,  $k$ , equal to  $6 \times 10^4\text{ s}^{-1}$ , while intramolecular eT from  $\text{Cu}_\text{A} \rightarrow$  heme a, also over an  $\sim 18\text{--}20\text{-}\text{\AA}$  distance, has a rate constant of  $1 \times 10^4\text{ s}^{-1}$ . The  $\text{Cu}_\text{A}$  center exists in two redox states: the reduced form with two cuprous ions ( $\text{Cu}^{\text{I}}\cdots\text{Cu}^{\text{I}}$ ) and the fully delocalized mixed valent center ( $\text{Cu}^{\text{II}}\cdots\text{Cu}^{\text{I}} \leftrightarrow \text{Cu}^{\text{I}}\cdots\text{Cu}^{\text{II}}=\text{Cu}^{1.5}\cdots\text{Cu}^{1.5}$ ). In the

latter delocalized, mixed valence (MV) state, the unpaired electron ( $S = 1/2$ , nuclear spin  $I = 3/2$ ) gives rise to an EPR spectroscopic seven-line hyperfine pattern that persists even at very low temperatures. The delocalized, MV copper ion system— $\text{Cu}^{\text{II}}\cdots\text{Cu}^{\text{I}} \leftrightarrow \text{Cu}^{\text{I}}\cdots\text{Cu}^{\text{II}} \equiv \text{Cu}^{1.5}\cdots\text{Cu}^{1.5}$ —has been designated, in the inorganic literature, as a class III system.<sup>160</sup> As discussed in Section 7.8.2, the two copper ions in  $\text{Cu}_{\text{A}}$  are bridged by cysteinyl sulfur atoms, forming a planar diamond shape with a short 2.58-Å metal-to-metal distance. Two histidine ligands form strong bonds to copper ions in the  $\text{Cu}_{\text{A}}$  center as well. In addition to its distinctive EPR spectrum, the  $\text{Cu}_{\text{A}}$  center exhibits an intense purple color with strong UV-visible bands around 480 and 530 nm. These bands originate from  $\text{S}(\text{cys}) \rightarrow \text{Cu}$  charge-transfer transitions. Resonance Raman spectroscopy identifies  $\text{Cu}_2\text{S}_2$  breathing modes, Cu–S and Cu–N stretching modes, Cu–S twisting modes, and  $\text{Cu}_2\text{S}_2$  accordion bending modes. The unusual core  $\text{Cu}_{\text{A}}$  center structure, with its extremely short Cu–Cu distance and its thiolate bridging, is believed to be largely responsible for its rapid electron transfer capabilities.

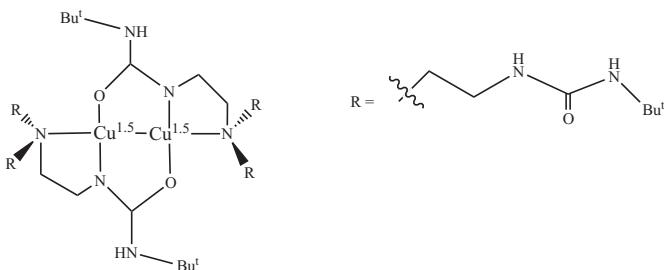
Over the past 10–15 years, several model systems that maintain the copper ion mixed-valence, delocalized electronic state with short Cu–Cu bond distances as in the CcO enzyme— $\text{Cu}^{\text{II}}\cdots\text{Cu}^{\text{I}} \leftrightarrow \text{Cu}^{\text{I}}\cdots\text{Cu}^{\text{II}} \equiv \text{Cu}^{1.5}\cdots\text{Cu}^{1.5}$ —have been synthesized and characterized. One class of ligand is the so-called octaazacryptands, also known as macrobicyclic polyaza ligand systems. These have been studied by the groups of J. Nelson,<sup>161</sup> M. E. Barr,<sup>162</sup> and A. J. Thomson and J. J. McGarvey.<sup>163</sup> The X-ray crystallographic structures of these complexes reveal copper ions of trigonal bipyramidal geometry with Cu–Cu distances between 2.36 and 2.42 Å. The EPR spectra of the complexes show the characteristic seven-line hyperfine pattern arising from spin delocalization over two copper ions ( $I = 3/2$ ) with a Cu–Cu bond arising through direct  $d_{z^2}$  orbital overlap. The hyperfine couplings,  $A_{\perp}$ , in the model compounds are relatively large ( $A_{\perp} \approx 110\text{--}115\text{ G}$ ) compared to the values for CcO's  $\text{Cu}_{\text{A}}$  site ( $A_x = 22.70$ ,  $A_y = 24.50$ ,  $A_z = 38.00$ ). The difference indicates that the copper ions in the model complexes do not possess the extensive spin delocalization onto the ligands as found for the enzyme's  $\text{Cu}_{\text{A}}$  site. Intense UV-visible ( $\lambda = 600\text{--}660\text{ nm}$ ,  $\epsilon = 1500\text{--}3000\text{ M}^{-1}\text{ cm}^{-1}$ ) and near-IR ( $\lambda = 730\text{--}780\text{ nm}$ ,  $\epsilon = 4500\text{--}5000\text{ M}^{-1}\text{ cm}^{-1}$ ) bands indicate metal ion  $d\text{-}d$  and  $\sigma\text{-}\sigma^*$  transitions, respectively. In CcO's  $\text{Cu}_{\text{A}}$  site, UV-visible bands are assigned as  $\text{S} \rightarrow \text{Cu}$  charge transfer bands indicating more involvement of the thiolate ligands that probably play an important role in the enzyme's electron transfer capability. Spectroscopic data on CcO's  $\text{Cu}_{\text{A}}$  site and these model compounds, as well as others to be discussed in the next paragraphs, are collected in Table 14 and 15 of reference 138. The model compounds' structures are collected in Figure 26 of the same publication.

The S. J. Lippard group has studied another set of Type III (class III) mixed-valence (MV) copper complexes containing carboxylate-bridging ligands based on *m*-xylenediamine bis(Kemp's triacid imide), H<sub>2</sub>XDK, or its propyl derivative (H<sub>2</sub>PXDK).<sup>164</sup> The preorganized, negatively charged carboxylate

ligands provide (a) steric hindrance to stabilize the binuclear copper ion MV core and (b) a less positively charged overall complex than the neutral octa-azacryptand ligands discussed above. In spite of their quite different nature, both types of ligands result in model complexes with similar physical properties. X-ray crystallography of the carboxylate-bridged complexes reveals square pyramidal geometry for each copper ion with Cu–Cu distances between 2.40 and 2.42 Å. EPR spectroscopy shows that the complexes exhibit the seven-line signals demonstrating fully delocalized copper ion electronic structure.

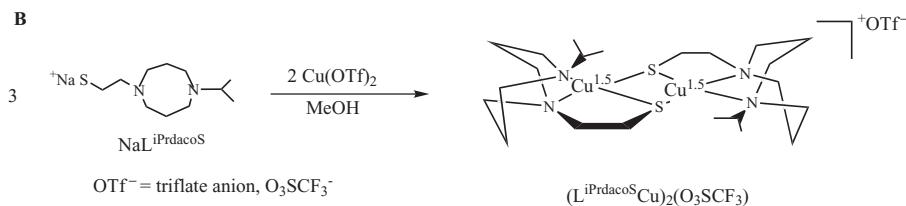
Next in the discussion of Cu<sub>A</sub> site model complexes, Borovik and co-workers have prepared mixed-valence dicopper complexes using the tripodal ligand *tris*[(N'-*tert*-butylureayl)-N-ethyl]amine.<sup>165</sup> The ligand has three N-ethylene-N'-*tert*-butyl urea arms attached to an apical nitrogen atom. (See Figure 7.48A.) This ligand links two copper ions through two mono-deprotonated urea ligands producing a completely delocalized Cu<sup>1.5</sup>...Cu<sup>1.5</sup> core with two μ-1,3-(κN:κO)-ureate bridges and copper centers separated by 2.39 Å. The ligand produces a pseudomacroyclic ligand around the copper ions through intramolecular hydrogen bonds formed from urea groups on adjacent ligands. Spectroscopic

A



*A tris*[(N'-*tert*-butylureayl)-N-ethyl]amine ligand reacted with two Cu ions to form Cu<sup>1.5</sup>...Cu<sup>1.5</sup> complex as described in reference 165.

B



A dithiolate ligand reacted with two Cu ions to form Cu<sup>1.5</sup>...Cu<sup>1.5</sup> complex as described in reference 166.

**Figure 7.48** Two cytochrome c oxidase Cu<sub>A</sub> center model compounds: (A) delocalized Cu<sup>1.5</sup>...Cu<sup>1.5</sup> core with two μ-1,3-(κN:κO)-ureate bridges as reported in reference 165; and (B) a dithiolate-bridged mixed-valence binuclear copper ion complex as reported in reference 166.

properties are similar to those found for (a) the other model complexes discussed and (b) the CcO enzyme's Cu<sub>A</sub> center: (1) UV-visible (568 nm, 1400 M<sup>-1</sup> cm<sup>-1</sup>, 446 nm, 1600 M<sup>-1</sup> cm<sup>-1</sup>) and more intense near-IR (705 nm, 4000 M<sup>-1</sup> cm<sup>-1</sup>) bands and (2) EPR seven-line hyperfine coupling from the fully delocalized mixed-valent copper species.

Tolman and co-workers reported the first example of a dithiolate-bridged mixed-valence binuclear copper ion complex in 1996.<sup>166</sup> (See Figure 7.48B.) X-ray crystallography of this model compound revealed a planar MV [Cu<sub>2</sub>(μ-SR)<sub>2</sub>]<sup>+</sup> core with a Cu–Cu distance of 2.92 Å, considerably longer than that found in the CcO enzyme Cu<sub>A</sub> center (2.43–2.58 Å) or in the other model complexes discussed above (2.39–2.42 Å). The complex involved the use of a newly synthesized ligand, NaL<sup>iPrdacoS</sup>, whose structure is shown in Figure 7.48B. The reaction shows three moles of NaL<sup>iPrdacoS</sup> reacting with two moles of Cu(OTf)<sub>2</sub> to produce the mixed-valent product (L<sup>iPrdacoS</sup>Cu)<sub>2</sub>(O<sub>3</sub>SCF<sub>3</sub>)—one mole of NaL<sup>iPrdacoS</sup> acting as a reducing agent to produce the desired Cu<sup>1.5</sup>...Cu<sup>1.5</sup> product. EPR data for (L<sup>iPrdacoS</sup>Cu)<sub>2</sub>(O<sub>3</sub>SCF<sub>3</sub>) exhibit the expected seven-line hyperfine splitting for the MV product. The UV-visible spectrum of (L<sup>iPrdacoS</sup>Cu)<sub>2</sub>(O<sub>3</sub>SCF<sub>3</sub>) shows absorbances that are red-shifted to longer wavelengths than those for the CcO Cu<sub>A</sub> center; and the near-IR band for (L<sup>iPrdacoS</sup>Cu)<sub>2</sub>(O<sub>3</sub>SCF<sub>3</sub>) occurs at 1466 nm ( $\epsilon = 1200 \text{ M}^{-1} \text{ cm}^{-1}$ ), a much lower energy than that for Cu<sub>A</sub> (800 nm ( $\epsilon = 1600 \text{ M}^{-1} \text{ cm}^{-1}$ )). Further comparisons between CcO's Cu<sub>A</sub> center and the (L<sup>iPrdacoS</sup>Cu)<sub>2</sub>(O<sub>3</sub>SCF<sub>3</sub>) complex have been carried out using absorption, magnetic circular dichroism (MCD), and X-ray absorption (XAS) spectroscopies. Molecular orbital (MO)-splitting diagrams and detailed descriptions of the relationship between these and observed spectroscopic results for CcO's Cu<sub>A</sub> and mixed-valent [Cu<sub>2</sub>(μ-SR)<sub>2</sub>]<sup>+</sup> complexes have been published.<sup>167</sup> For [Cu<sub>2</sub>(μ-SR)<sub>2</sub>]<sup>+</sup> these researchers defined  $\psi$  and  $\psi^*$  molecular orbitals derived from delocalization of electron density between the two copper centers. The doubly occupied  $\psi$  molecular orbital arises from overlap of two Cu  $d_{xy}$  orbitals to form a  $\pi_g^*$  antibonding interaction, while the  $\psi^*$  singly occupied HOMO arises from the same Cu  $d_{xy}$  orbitals to form a  $\pi_u$ -bonding interaction. Figure 6 of reference 167a illustrates the complete molecular orbital diagram for coupling two C<sub>2v</sub> CuS<sub>2</sub> fragments. The  $\psi \rightarrow \psi^*$  transition, occurring at 13,000 cm<sup>-1</sup> for CcO's Cu<sub>A</sub> center is assigned to the class III, mixed valence Cu–Cu overlap because a large change in Cu–Cu separation accompanies the excitation in the near-IR resonance Raman spectrum. Further analysis of the 13,000 cm<sup>-1</sup> band (that converts to an ~800-nm band in UV-visible and MCD spectra) assigns the transition to Cu–S plus Cu–Cu orbital overlaps. In the [Cu<sub>2</sub>(μ-SR)<sub>2</sub>]<sup>+</sup> model, the  $\psi \rightarrow \psi^*$  transition occurs at ~5600 cm<sup>-1</sup> (translating to the UV-visible or MCD absorbances at ~1450–1600 nm), a much lower energy. This observation led to the conclusion that the valence delocalization in the model complex (L<sup>iPrdacoS</sup>Cu)<sub>2</sub>(O<sub>3</sub>SCF<sub>3</sub>)—where the Cu–Cu distance is equal to 2.92 Å—is mediated by the bridging thiolate ligands only, whereas in CcO's Cu<sub>A</sub> center the valence delocalization arises not only through the bridging thiolates but also by direct Cu–Cu overlap, given the Cu–Cu dis-

tance of 2.42–2.58 Å in the enzyme. Research continues on the synthesis of mixed-valent  $[\text{Cu}_2(\mu\text{-SR})_2]^+$  complexes with shorter Cu–Cu distances, and interested readers should search the current literature for updates.

### 7.8.5 Cytochrome c Oxidase Conclusions

Section 7.8 has described the enzyme cytochrome c oxidase, the terminal enzyme (Complex IV) in a sequence of membrane-bound electron transfer proteins in the mitochondrial respiratory chain. CcO receives electrons from cytochrome c (Section 7.7) and transfers them to dioxygen, producing water. In the process, protons are translocated or “pumped” for two purposes: (1) to provide the protons necessary for producing water and (2) to generate an electrochemical potential across the mitochondrial membrane. The electrochemical potential causes a reverse proton flow that is coupled to ATP synthesis. Cytochrome c oxidase contains four diverse metal ion centers that assist in electron transfer through the enzyme: two iron heme a centers, a monometallic  $\text{Cu}_B$  site that, along with heme  $a_3$ , constitutes the dioxygen binding/reduction site, and a bimetallic  $\text{Cu}_A$  center. The metal ion centers and their ligand systems are described in Section 7.8.2.

Cytochrome c oxidase enzyme has been studied by many different analytical methods, often with the goal of identifying intermediates in the CcO catalytic cycle. Kinetic studies, X-ray crystallography of CcO’s 13 subunits in one large complex (PDB: 2OCC and others), and solution NMR studies of CcO segments have been helpful in identifying structural changes that take place when the enzyme’s metal centers undergo oxidation and reduction. (See Figures 7.40 and 7.42.) The nature of dioxygen binding—superoxoxo, peroxo, hydroxo, and  $\mu$ -oxo coordination are some possibilities—and the exact mechanism of its reduction through intermediate stages in the catalytic cycle have proved to be thornier problems, not totally solved at this time. The study of cytochrome c oxidase model compounds has been undertaken by many researchers to assist in determining how dioxygen binds and is reduced, and some aspects of this vast subject are addressed in Section 7.8.4.

As stated in the concluding remarks of reference 138, the overall purpose of heme–copper oxidase synthetic modeling studies is to “elucidate fundamental aspects of the relevant coordination chemistry, metal ion ligation, coordination structures, spectroscopy and spectroscopic structural correlation, and reactivity.” Synthetic approaches to the oxidized form heme  $a_3\text{-Cu}_B$  center ( $\text{Fe}^{\text{III}}\text{-X-Cu}^{\text{II}}$  assemblies) try to identify X, where X is some form of dioxygen, reduced dioxygen, or oxygen in a mono-oxo or hydroxo form, with no clear result as of this writing. Since the heme  $a_3$  Fe-to- $\text{Cu}_B$  distances observed so far are greater than 4.3 Å, possible magnetic coupling between the metal centers is weak at best. The  $\mu$ -oxo and  $\mu$ -hydroxo heme–Cu assemblies of Karlin and co-workers are remarkably stable, but their relevance to the enzyme’s actual intermediates are unclear. Look for further studies involving reactions of the  $\mu$ -oxo and  $\mu$ -hydroxo heme–Cu model complexes with hydrogen peroxide

possibly leading to cytochrome c oxidase-type **P** or **F** transient intermediates (Figures 7.40 and 7.42) relevant to the enzyme's O<sub>2</sub> reduction pathway. Biomimetic studies on Fe<sup>II</sup>-Cu<sup>I</sup> complexes and the reaction products of these species with dioxygen are areas of continuing study. Functional modeling studies of dioxygen-binding to model compounds using electrochemical approaches has been reviewed by Collman et al.<sup>168</sup> The work of Karlin's research group preparing Fe<sup>III</sup>-(O<sub>2</sub><sup>2-</sup>)-Cu<sup>II</sup> products from heterobinuclear Fe<sup>II</sup>-Cu<sup>I</sup> complexes has been discussed in some detail in this section. Naruta's X-ray crystallographic structure (reference 155) showing peroxy side-on ( $\eta^2$ ) binding to the heme's Fe ion and end-on ( $\eta^1$ ) binding to the copper ion confirms earlier conclusions about the peroxy bonding modes garnered from spectroscopic evidence. The question of whether this bonding mode exists in the enzyme itself remains unanswered. Biomimetic studies of the his-tyr covalent cross-link at the heme a<sub>3</sub>-Cu<sub>B</sub> center is a new exciting direction for cytochrome c oxidase modeling not discussed here. A few references are included to start the reader into this research area.<sup>169</sup> Finally, modeling studies of the Cu<sub>A</sub> dicopper center, which serves as a one-electron transfer conduit for many copper oxidases (and nitrous oxide reductase), continue into the present time. The mixed-valence delocalized Cu<sup>1.5</sup>...Cu<sup>1.5</sup> center with cysteine bridging ligands has provided challenges for chemists synthesizing workable models that mimic the enzyme's structure and function. Better Cu<sub>A</sub> models are needed, especially those that mimic the sulfur and nitrogen ligand coordination sphere of the CcO enzyme like those of Tolman (reference 166) but that also have the short Cu-Cu distance found in the native cytochrome c oxidase.

## 7.9 NON-HEME IRON-CONTAINING PROTEINS

### 7.9.1 Introduction

The preceding sections of Chapter 7 have discussed iron-containing proteins and enzymes having a porphyrin ring system. Section 7.9 presents a short introduction to the many non-heme iron-containing proteins and enzymes. Two of these are iron-sulfur proteins (Section 7.9.2) and iron-oxo proteins (Section 7.9.3).

### 7.9.2 Proteins with Iron-Sulfur Clusters

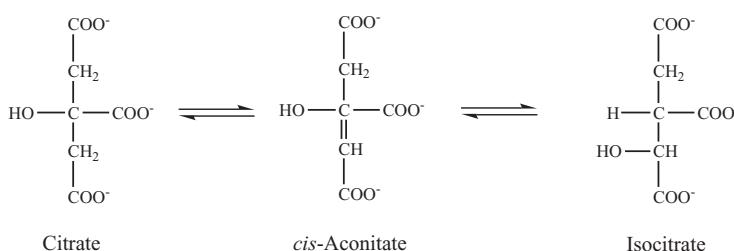
One large class of non-heme iron-containing biomolecules involves proteins and enzymes containing iron-sulfur clusters. Iron-sulfur clusters are described in Sections 1.7 (Bioorganometallic Chemistry) and 1.8 (Electron Transfer) as well as in Section 3.6 (Mössbauer Spectroscopy). See especially Table 3.2 and the descriptive examples discussed in Section 3.6.4. Iron-sulfur proteins include rubredoxins, ferredoxins, and the enzymes aconitase and nitrogenase. The nitrogenase enzyme was the subject of Chapter 6 in the first edition of this text—see especially Section 6.3 for a discussion of iron-sulfur clusters.<sup>1</sup> In this

second edition, information on iron–sulfur clusters in nitrogenase is found in Section 3.6.4. Rubredoxin and ferredoxin iron–sulfur proteins are electron transferases, as are the previously discussed cytochromes. Rubredoxins are found in anaerobic bacteria and contain iron ligated to four cysteine–sulfur ligands. Ferredoxins are found in plant chloroplasts and mammalian tissue and contain spin-coupled [2Fe–2S] clusters.

**7.9.2.1 The Enzyme Aconitase.** The enzyme aconitase catalyzes the elimination or addition of water in the second step of the citric acid (Krebs) cycle, catalyzing the interconversion of citrate and isocitrate via *cis*-aconitate. See reference 8, pages 190–196, Figure 7.49, and equation 7.13.



The aconitase superfamily contains five phylogenetic groups: (1) mitochondrial aconitases (mAcn); (2) cytoplasmic aconitases (cAcn) and iron regulatory proteins (IRP1 and IRP2) of higher organisms and bacterial aconitase As (AcnA); (3) homoaconitases; (4) fungal and bacterial isopropylmalate isomerases (IPMI); and (5) bacterial aconitase Bs (AcnB). Aconitase was one of the first iron–sulfur cluster-containing proteins to be studied in detail. Early in its investigation, it was found that activation of aconitase required conversion of a [3Fe–4S] cluster within the protein into a [4Fe–4S] cluster.<sup>170</sup> The Fe(II) ion inserted into the [3Fe–4S] cluster in the presence of a reductant, termed Fe<sub>a</sub> or Fe<sub>4</sub>, remains in its Fe<sup>2+</sup>, high-spin, six-coordinate state. Fe<sub>4</sub> becomes directly involved in binding substrate via carboxyl and hydroxyl ligands for citrate ( $\beta$ -carboxyl) and isocitrate ( $\alpha$ -carboxyl) and via  $\alpha$ - or  $\beta$ -carboxyl ligands for *cis*-aconitate.<sup>171</sup> (See reaction 7.13, Figures 7.49 and 7.50.) Mössbauer spectroscopy indicates that the [Fe<sub>3</sub>S<sub>4</sub>]<sup>+</sup> cluster is first reduced. This is followed by the slower uptake step of Fe(II) during enzyme rearrangement with the formation of the active enzyme containing a [Fe<sub>4</sub>S<sub>4</sub>]<sup>n+</sup> cluster. The [Fe<sub>4</sub>S<sub>4</sub>]<sup>n+</sup> cluster can exist with the iron ions (as tetrahedrally coordinated high-spin  $d^6$  or  $d^5$  species) in a series of oxidation and overall spin states, [Fe<sub>4</sub>S<sub>4</sub>]<sup>3+</sup>



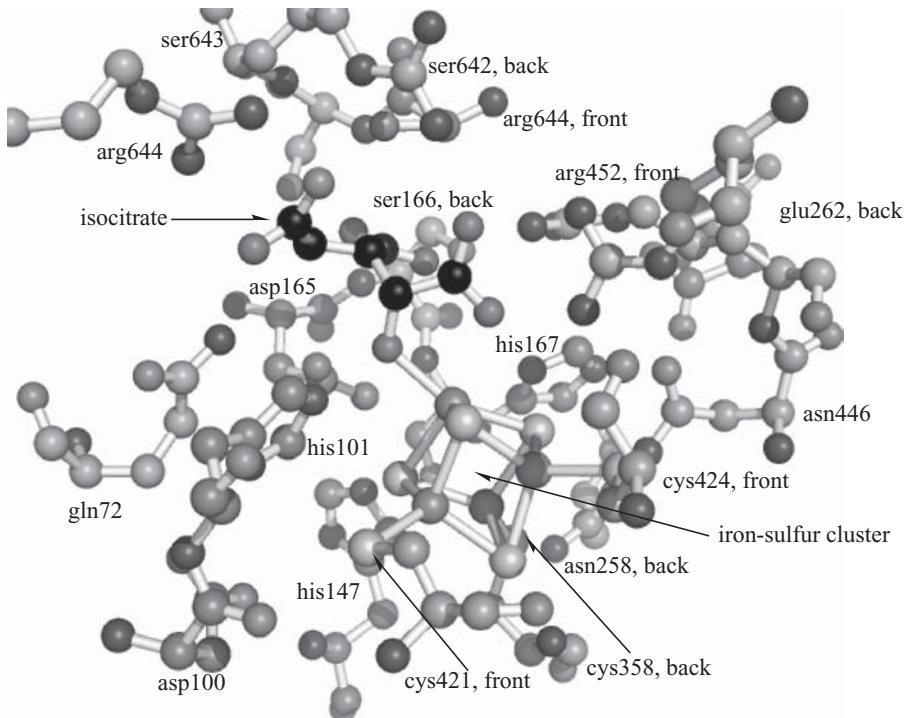
**Figure 7.49** Formation of isocitrate from citrate as catalyzed by the enzyme aconitase.

( $S = \frac{1}{2}$ ),  $[\text{Fe}_4\text{S}_4]^{2+}$  ( $S = 0$ ), and  $[\text{Fe}_4\text{S}_4]^+$  ( $S = \frac{1}{2}$ ), containing iron ions in +2, +3, and +2.5 states.

The enzyme aconitase contains four domains: Domains 1, 2, and 3 are closely associated with the [4Fe–4S] iron–sulfur cluster that resides in the active site, while domain 4 is attached by a peptide linker to create a large cleft that leads toward the active site. In 1992, the C. D. Stout group solved the X-ray crystallographic structure of porcine heart mitochondrial aconitase with bound substrate isocitrate (PDB: 7ACN).<sup>172</sup> The 82.8-kDa protein contains 754 amino acid residues in addition to an isocitrate group and an  $[\text{Fe}_4\text{S}_4]^{2+}$  iron–sulfur cluster. The same research group solved the structure of bovine heart enzyme in complex with the reaction intermediate analog and inhibitor nitro-isocitrate bound (PDB: 8ACN). Both substrate compounds bind to the unique iron atom in a [4Fe–4S] iron–sulfur cluster—the Fe not coordinated by a cysteine ligand—via hydroxyl and carbonyl oxygens. The substrate-coordinating  $\text{Fe}_a$  atom coordinates an  $\alpha$ -carboxyl oxygen of isocitrate, a  $\beta$ -carboxyl oxygen of citrate, and either the  $\alpha$ - or  $\beta$ -carboxyl oxygen of *cis*-aconitate as well as a substrate hydroxyl and  $\text{H}_2\text{O}$  from solvent. Twenty-three amino acid residues from all four aconitase domains contribute interactions within the [4Fe–4S] cluster-containing active site.

Figure 7.50 shows some of the important amino acid residues surrounding the isocitrate substrate (represented in ball-and-stick format with black carbon atoms). The iron–sulfur cluster, also in ball-and-stick format, shows sulfur atoms colored yellow and shows iron atoms in orange-brown. Residues cys358, cys421, cys424, asn258, and asn446 contribute ligation and other interactions with the iron–sulfur cluster—carbon atoms colored orange in Figure 7.50. Residues arg447, arg452, arg580, arg644, gln72, ser166, and ser643 participate in substrate recognition—carbon atoms colored green in Figure 7.50. Residues ala74, asp568, ser571, and thr567 contribute hydrogen bonds to active site residue side chains. Ser642 and three acidic side-chain histidine pairs (asp100–his101, asp165–his147 and glu262–his167) contribute specifically to the catalytic reaction shown in equation 7.13 and Figure 7.49. (The carbon atoms of these residues are colored magenta in Figure 7.50. For clarity, not every aa residue is shown.) For example, ser642 appears to be stabilized as an alkoxide to assist in proton abstraction from  $\text{C}_\beta$  of the isocitrate moiety. In addition to the catalytic group abstracting the proton, another catalytic group must protonate the substrate hydroxyl to remove this group as a water molecule. An extensive hydrogen-bonding network is involved in the protonation of substrate hydroxyl and Fe-bound hydroxyls. Some typical PDB: 7ACN bond lengths are displayed in Table 7.11.

The X-ray structures of other aconitases have appeared in the literature. Recently, the crystal structure of human iron regulatory protein, IRP1, in its aconitase form, has been published.<sup>173</sup> Iron regulatory proteins (IRPs) control the translation of proteins involved in iron uptake, storage, and utilization by binding to specific noncoding sequences of the corresponding mRNAs known



**Figure 7.50** Important amino acid residues surrounding the isocitrate substrate (black carbon atoms) and the iron–sulfur cluster of aconitase (PDB: 7ACN). Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft Corporation.) (See color plate)

**TABLE 7.11 Selected 7ACN Bond Distances (PDB: 7ACN)**

| Bonding Atoms                        | Bond Length (Å) | Hydrogen-Bonding Atoms                  | Bond Length (Å) | Hydrogen-Bonding Atoms                         | Bond Length (Å) |
|--------------------------------------|-----------------|---|-----------------|--|-----------------|
| Fe <sub>1</sub> –S <sub>cys358</sub> | 2.34            | Ser642 O <sub>γ</sub> –O <sub>ict</sub> | 3.60            | Asp100 O <sub>δ1</sub> –his101 N <sub>δ1</sub> | 2.76            |
| Fe <sub>2</sub> –S <sub>cys424</sub> | 2.26            | Ser642 O <sub>γ</sub> –O <sub>ict</sub> | 3.41            | Asp165 O <sub>δ2</sub> –his147 N <sub>ε2</sub> | 2.84            |
| Fe <sub>3</sub> –S <sub>cys421</sub> | 2.35            |   |                 | Glu262 O <sub>ε1</sub> –his167 N <sub>ε2</sub> | 2.80            |
| Fe <sub>4</sub> –O <sub>2ict</sub>   | 2.49            |   |                 |  |                 |
| Fe <sub>4</sub> –O <sub>7ict</sub>   | 2.15            |   |                 |  |                 |

ict, isocitrate.

as iron-responsive elements (IREs). Strong interaction between IRPs and IREs assures proper iron homeostasis in animal cells under iron shortage. Conversely, under iron-replete conditions, IRP1 binds a [4Fe–4S] cluster and functions as cytosolic aconitase. Comparison with known structures of other aconitase enzymes reveals conserved folds and active site environments throughout the aconitase superfamily.

### 7.9.3 Iron-Oxo Proteins

A second large class of non-heme iron proteins contains iron-oxo clusters. Two of these are (1) transferrins for metal transport and (2) ferritins for metal storage. Transferrins are eukaryotic iron-binding glycoproteins that control the level of free iron in biological fluids. Transferrin proteins, loaded with Fe(III) ions, encounter transferrin receptors on cell surfaces, bind to the receptor, and are transported into cells via a vesicle. Inside the cell, the vesicle is acidified, causing transferrin to release its iron ions. The receptor is then transported through the endocytic cycle back to the cell surface, ready for another round of iron uptake. Each transferrin molecule has the ability to carry two Fe(III) ions. The Wellcome Trust Sanger Institute website for transferrin is found at <http://www.sanger.ac.uk//cgi-bin/Pfam/getacc?PF00405>.

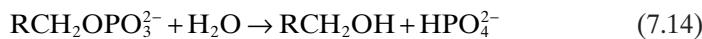
Ferritins are globular protein complexes, usually consisting of 24 protein subunits, forming a hollow, roughly spherical three-dimensional shape. They are the main intracellular iron storage proteins in both prokaryotes and eukaryotes, keeping iron ions in a soluble and nontoxic form. Each ferritin 24-mer can hold up to 4500 iron atoms or ions. Some ferritins are heavy (H)-chain rich; the H-chains catalyze oxidation of Fe(II). Some ferritins are light (L)-chain rich; the L-chains promote nucleation and storage of Fe(III) ions. Recent X-ray crystallographic studies of L-chain human ferritin have been carried out<sup>174</sup> (PDB: 2FFX, 2FG4, 2FG8). Human H-chain ferritins have been studied as well and compared to horse H-chain ferritins<sup>175</sup> (PDB: 1AEW (horse), 2FHA (human)).

A website-based ferritin tutorial, last updated in 1998, can be found at <http://www.chemistry.wustl.edu/~edudev/LabTutorials/Ferritin/FerritinTutorial.html>. The website is accompanied by a *Journal of Chemical Education* publication describing ferritin and a general chemistry laboratory experiment for analysis of the iron content of ferritin.<sup>176</sup> The Wellcome Trust Sanger Institute ferritin website is at <http://www.sanger.ac.uk//cgi-bin/Pfam/getacc?PF00210>. M. A. Carrondo has written a review of bacterioferritins—ferritins isolated from bacteria, some of which contain heme iron centers.<sup>177</sup> Stephen J. Lippard's group has synthesized and characterized mixed-valent dodecairon oxo complexes that are models for the biominerization of ferritin. The iron-oxo clusters exhibited properties of both discrete molecules and extended solids and were deemed to be representative of a new class of nanometer-sized compounds bridging the molecular solid-state boundary.<sup>178</sup> Information on the X-ray crystallographic structure and electronic properties of

the ferritin model,  $[Fe_{12}(O)_2(OMe)_{18}(OAc)_6(MeOH)_{4,67}]$ , was published by the Lippard group in 1994.<sup>179</sup>

Hemerythrin is an oligomeric protein responsible for dioxygen ( $O_2$ ) transportation in the marine invertebrates. Myohemerythrin is a monomeric  $O_2$ -binding protein, found in the muscles of marine invertebrates. Hemerythrin and myohemerythrin are essentially colorless when deoxygenated, but turn a violet-pink in the oxygenated state. Neither enzyme contains a heme group. In both hemerythrin and monohemerythrin, the dioxygen-binding site is a binuclear iron center. The iron ions are coordinated to the protein through the carboxylate side chains of one glutamate, one aspartate, and five histidine amino acid residues. Hemerythrin and myohemerythrin are often described according to oxidation and ligation states of the iron center: (1) Fe(II)–OH–Fe(II), reduced, deoxy; (2) Fe(II)–OH–Fe(III), semi-met; (3) Fe(III)–O–Fe(III)–OOH<sup>–</sup>, oxidized, oxy; and (4) Fe(III)–OH–Fe(III)–other ligands, oxidized, met. The X-ray crystallographic structures of oxy (PDB: 1HMO) and deoxy (PDB: 1HMD) were determined at 2.00 Å resolution in the early 1990s.<sup>180</sup> More information on hemerythrin and associated proteins can be found at the website <http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF01814>.

Purple acid phosphatases (PAPs) occur widely in nature and are responsible for hydrolysis of orthophosphate monoethers to alcohols under acidic conditions according to the reaction



In the plant form, PAPs contain a binuclear metal center containing hexacoordinate Zn(II) and Fe(III) ions, while in the mammalian form, PAPs contain binuclear iron centers having Fe(II)–Fe(II) or Fe(II)–Fe(III) oxidation states. The Fe(III)–M(II) center constitutes the active form of the enzyme. Ligands include histidine N<sub>δ</sub> or N<sub>ε</sub> atoms, mono- or bidentate O atoms from aspartate (O<sub>δ</sub>) or glutamate (O<sub>ε</sub>) residues, the hydroxy group of a tyrosine residue, a bridging O(H) ligand, and/or water or hydroxide ions. The purple color of PAPs is due to a tyr → Fe(III) charge transfer (CT) band around 560 nm. Recent X-ray crystallographic structures have mostly characterized the catalytically inactive Fe(III)–Fe(III) form of the mammalian enzyme.<sup>181</sup>

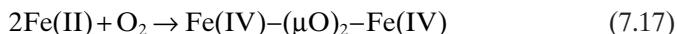
**7.9.3.1 Methane Monooxygenases.** Methane monooxygenase (MMO) and ribonucleotide reductases (RNR) belong to a class of enzymes that contain a nonheme diiron(II) center with two bridging carboxylate ligands as well as terminal carboxylate and histidine ligands. Only the MMO enzyme will be discussed in any detail here. The enzyme methane monooxygenase (MMO) is utilized by methanotrophic bacteria (bacteria capable of using methane as both a carbon and an energy source) to oxidize methane ( $CH_4$ ) to methanol ( $CH_3OH$ ) with incorporation of one oxygen atom from  $O_2$ .



Methane is the most difficult hydrocarbon substrate to oxidize because it has a high C–H bond energy (104 kcal/mol), no dipole moment, and no functionality that would assist its binding to a catalytic site. Researchers interested in the methane monooxygenase enzyme believe that it might be harnessed to convert natural gas, mostly methane, to more useful methanolic fuels and to oxidize other alkanes to their more useful alcohol counterparts. Wide varieties of organic molecules are oxidized by this enzyme; for instance, halogenated hydrocarbons can be oxidized by MMO in bioremediation efforts.<sup>182</sup> The MMO enzyme of methanotrophic bacteria can be utilized to remove the harmful greenhouse gas methane from the atmosphere, thus playing an environmental role here as well.

Two types of methane monooxygenases have been studied: (1) soluble methane monooxygenase (sMMO) and (2) particulate (membrane-bound) methane monooxygenase (pMMO). The well-studied sMMO is produced by methanotrophs under copper-limiting conditions. All methanotrophs produce pMMO—found in intracytoplasmic membranes—but it is the less well-studied enzyme.

Soluble methane monooxygenase (sMMO) is a multicomponent enzyme consisting of the following subunits: (1) a 251-kDa hydroxylase protein (MMOH) containing a non-heme dinuclear iron center; (2) a 39-kDa NADH-dependent, [2Fe–2S]- and FAD-containing reductase (MMOR); (3) a 16-kDa regulatory protein (MMOB); and (4) a protein of unknown function that binds to MMOH and inhibits sMMO activity (MMOD). The enzyme activates dioxygen for further oxidation chemistry, incorporating one of its oxygens into the product—hence the name monooxygenase. The sMMO's MMOH component catalyzes the reaction with methane (converting it to methanol) at an active site containing a non-heme di-iron moiety according to the following reactions adapted from reference 183:

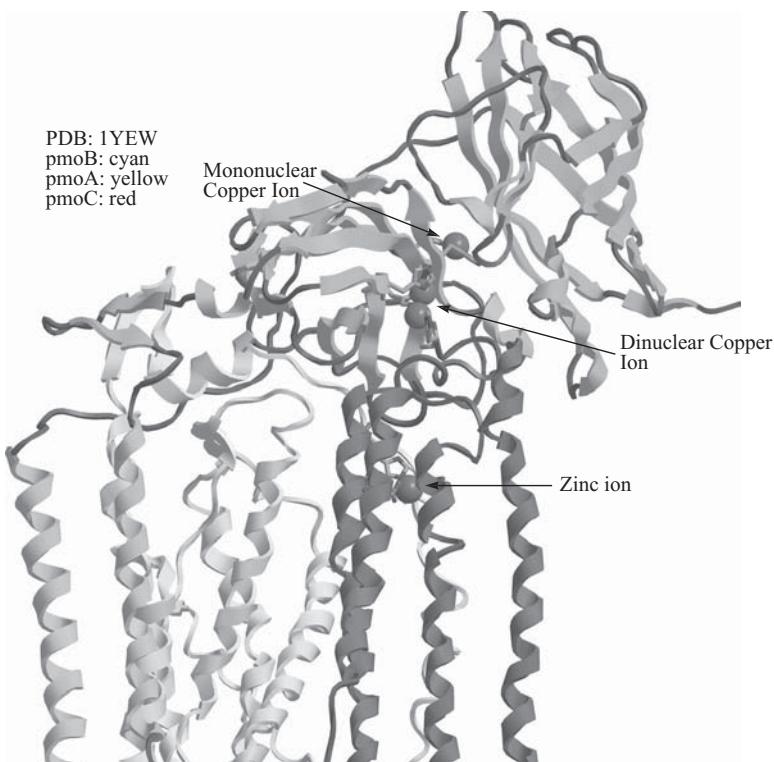


Three enzymes capable of hydroxylating methane are known: soluble methane monooxygenase (sMMO), particulate methane monooxygenase (pMMO), and a related enzyme ammonia monooxygenase (AMO). All methanotrophs produce pMMO—found in intracytoplasmic membranes—but until now it has been the less well-studied enzyme. Particulate MMO (pMMO) is composed of three subunits: (1) pmoB ( $\alpha$ , ~47 kDa), (2) pmoA ( $\beta$ , ~24 kDa), and (3) pmoC ( $\gamma$ , ~22 kDa). Each of these subunits contains membrane-spanning helices. A 100-kDa  $\alpha\beta\gamma$  polypeptide arrangement has been proposed, but a definitive molecular mass for the enzyme is not known at this time. The number of metal centers in pMMO remains controversial, with reported values of 2–15 copper ions and 0–2 iron ions per ~100 kDa of purified pMMO.<sup>184</sup> Five trinuclear copper centers were proposed by Chan on the basis of the hyperfine splitting in the electron paramagnetic resonance (EPR) spectrum ( $g = 2.06$ ),

and the reference 185 authors hypothesized that the centers could be characterized as two catalytic (C) centers and three electron transfer (E) functional units. R. L. Lieberman in the Rosenzweig group at Northwestern observed a type 2 mononuclear Cu(II) EPR signal only.<sup>184b,c,e,186</sup> These workers also proposed a mononuclear copper site plus a copper-containing cluster based on extended X-ray absorption fine structure (EXAFS) best fitted with a 2.57-Å Cu–Cu interaction.<sup>184b,187</sup> In 2005, the researchers were able to crystallize *Methyllococcus capsulatus* (Bath) pMMO and reported its 2.8-Å resolution X-ray crystallographic structure<sup>188</sup> (PDB: 1YEW).

X-ray crystallography of the particulate MMO revealed a trimer composed of three copies (protomers) each of pmoA, pmoB, and pmoC to form a  $\alpha_3\beta_3\gamma_3$  cylinder ~105 Å long and ~90 Å in diameter. A soluble region, composed mainly of six  $\beta$ -barrel structures, extends ~45 Å away from the membrane. The soluble region rests on 42 transmembrane (TM) helices, 14 from each protomer. In the soluble region, a ~11-Å hole, lined with hydrophilic residues—glutamic acid, aspartic acid, and arginines—is located in the center of the trimer. The hole extends into the TM region, widening eventually to ~22 Å. The reference 188 authors believe that the trimer is physiologically relevant.

One protomer in the trimer can be described in the following manner using PDB: 1YEW data as visualized in Figure 7.51. The chain identified as pmoB includes residues 33–414, and it is colored cyan in Figure 7.51. The pmoB secondary structure includes two antiparallel  $\beta$ -barrel structures, one at the N-terminal end and one at the C-terminus. The N-terminal  $\beta$ -barrel includes seven strands and is oriented ~90° in relation to the C-terminal  $\beta$ -barrel, which contains eight strands. The two  $\beta$ -barrels are separated by a  $\beta$ -hairpin followed by the two TM (transmembrane) helices. The pmoB secondary structural motif is similar to that of subunit II of cytochrome c oxidase (Section 7.8). Both pMMO and cytochrome c oxidase contain monomeric copper(II) and dinuclear Cu–Cu centers within the  $\beta$ -barrel motif. In cytochrome c oxidase, these are identified as the dinuclear Cu<sub>A</sub> and monomeric Cu<sub>B</sub> sites. In pMMO, the mononuclear copper ion, colored blue in Figure 7.51, is located ~25 Å above the membrane and near the surface of the N-terminal  $\beta$ -barrel. This copper ion is coordinated by the N<sub>δ</sub> atoms of his48 and his72, shown in stick figure format in Figure 7.51, with a nearly linear geometry. Gln404 resides within 3 Å of this site as well. This mononuclear copper center resembles the type 2 site found in multicopper oxidases such as ascorbate oxidase and ceruloplasmin, and it probably is responsible for the type 2 Cu(II) EPR signal seen for purified pMMO. The second pmoB copper ion site, blue in Figure 7.51, is also within its N-terminal  $\beta$ -barrel, ~21 Å from the mononuclear site. The site's location, about 10 Å above the lipid bilayer interface, is similar to that of cytochrome c oxidase's Cu<sub>A</sub> site. The Cu–Cu distance in PDB: 1YEW refines to ~2.6 Å, similar to the 2.57 Å value found from EXAFS studies. The coordination sphere for one copper ion includes his33, the N-terminal amino acid found for the pmoB protomers of PDB: 1YEW. Both the N-terminal amino nitrogen and the side-chain N<sub>δ</sub> of his33 are within bonding distance to Cu<sub>1</sub>.



**Figure 7.51** Particulate MMO, pMMO (PDB: 1YEW). Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. (Printed with permission of CambridgeSoft Corporation.) (See color plate)

The second copper ion, Cu<sub>2</sub>, is ligated to the N<sub>δ</sub> of his137 and the N<sub>ε</sub> of his139. The dinuclear copper center ligands are shown as stick figures in Figure 7.51. His137 and his139 are highly conserved in pMMO and AMO from a number of organisms. His33 has hydrogen bonds to the side chain of glu35 as has his139 to the carbonyl oxygen of gly152. Both glu35 and gly152 are also highly conserved.

The 1PDB: 1YEW pmoA and pmoC subunits reside mainly in the membrane. Subunit pmoA includes residues 7–244 is colored yellow in Figure 7.51. The pmoA subunit contains seven TM α-helices, some of which pack against the TM helices from pmoB. Two pmoA helices face the trimer center hole and interact with counterparts from the other two protomers (not shown in Figure 7.51).

The subunit pmoC, colored red in Figure 7.51, is comprised of five TM α-helices that are oriented approximately parallel to the membrane normal and to each other. A single zinc metal ion, in a distorted tetrahedral coordination sphere, has been modeled within the membrane. Conserved amino acid resi-

dues from pmoA (glu195) and pmoC (asp156, his160, his173) hold this zinc ion in place. These are shown in stick format in Figure 7.51.

Since the nuclearity of biological copper centers does not correlate with reactivity,<sup>189</sup> any of the pMMO sites could be catalytic. The authors believe that the dinuclear site looks promising, because it is unusual. The dinuclear site has an adjacent cavity lined by several conserved hydrophobic residues—pro94 from pmoB and leu78, ile163, val164 from pmoC—that could be a binding site for methane. Similarites between the dinuclear copper and the cytochrome c oxidase (CcO) subunit II sites could indicate that this site is involved in electron transfer. However, the CcO dinuclear copper site has an entirely different coordination sphere that includes bridging cysteine sulfur atoms and terminal coordination to two histidines, a methionine, and a carbonyl oxygen. Other arguments can be put forward for the mononuclear copper center as well as the zinc center as catalytic sites in pMMO. For instance, the zinc site is rich in amino acids with carboxylic acid side chains and could accommodate a diiron center similar to that described for sMMO. The physiological source of electrons for pMMO is unknown, but a possible docking site is found in a negatively charged patch on the outer surface of the soluble pmoB domains close to the dinuclear copper site—conserved pmoB residues glu35 and asp368 as well as pmoC residue asp45 are found here.

Conclusions from the PDB: 1YEW X-ray crystallographic study of pMMO are as follows: (1) The pMMO structure reveals an unexpected trimeric arrangement; (2) three metal ions, a mononuclear copper(II) center, and a dinuclear Cu–Cu site reside within the soluble regions of the pmoB subunit; (3) the third metal center, occupied by a zinc ion, lies within the membrane and is coordinated by aa residues from pmoA and pmoC; (4) metal centers lie close enough together to possibly foster electron transfer; and (5) neither the site of methane oxidation nor the pathway for methane ingress or egress can be determined from these results.

In 2005, the Rosenzweig group published an article in *Inorganic Chemistry* entitled: “X-ray Crystallography and Biological Metal Centers: Is Seeing Believing?”.<sup>190</sup> This article serves as an excellent guide to bioinorganic chemists. First, the article lays out restrictions inherent in X-ray crystallographic techniques. Some of these are as follows: (1) Very few protein crystal structures are determined to atomic resolution, and therefore bond distances and angles between small atoms such as carbon, nitrogen, oxygen, and definitely hydrogen cannot be known with certainty; (2) the number and type of metal ion centers in biological molecules cannot be known with certainty from X-ray crystallography only, because metal ions may be adventitious, the actual biological metal ion may be replaced by other metal ions during crystallization, or the metal ion may be lost altogether; (3) the metal ion’s oxidation state cannot be known with certainty from X-ray crystallographic results alone; (4) the metal ion can be reduced by the X-ray beam, and the true oxidation state and resulting coordination sphere in a studied “catalytic intermediate” may not be the one observed; and (5) exogenous ligands, such as O<sub>2</sub>, must

be discovered in the crystallographic electron density map—not a trivial endeavor for structures at medium or low resolution ( $>2.8\text{ \AA}$ )—or discovered by other means.

Fortunately, there are corresponding advantages to the study of biological systems using X-ray crystallography. Questions about metal ion location, the coordination sphere of amino acid residues and exogenous ligands surrounding the metal, and the geometry about the metal center are routinely answered through crystallography. Often, X-ray crystallography is a starting point for protein and enzyme mechanistic and computational studies. Questions arising from the protein's X-ray structure often inspire researchers to synthesize model compounds that may further clarify a catalytic cycle, or result in a small molecule catalyst mimicking the biomolecule's reaction.

In reference 190, the authors describe the spectroscopic and X-ray crystallographic techniques they used to determine the pMMO structure. First, EPR and EXAFS experiments indicated a mononuclear, type 2 Cu(II) center ligated by histidine residues and a copper-containing cluster characterized by a  $2.57\text{ \AA}$  Cu–Cu interaction. A functional iron center was also indicated by Inductively Coupled Plasma–Atomic Emission Spectroscopy (ICP–AES).<sup>184b,187</sup> ICP–AES uses inductively coupled plasma to produce excited atoms that emit electromagnetic radiation at a wavelength characteristic of a particular element. The intensity of this emission is indicative of the concentration of the element (iron in this case) within the sample.

In X-ray crystallographic techniques, the reference 190 authors used copper single-wavelength anomalous dispersion (SAD) to obtain phases, and then they improved the phases through noncrystallographic symmetry averaging. As described in Section 3.3.3, the anomalous dispersion technique chooses a wavelength (available using tunable synchrotron X-ray sources) that will cause a transition of electrons of heavy atoms in the crystal. The wavelength chosen is near the metal's absorption edge. In this case, anomalous data collected at an energy of  $9400\text{ eV}$  located six copper sites arranged in pairs around a three-fold axis. This was the first indication that pMMO existed as a trimer. Additional data collection at  $9686\text{ eV}$  identified zinc ions, as well as confirming the presence of the copper ions. Multiple data sets at the iron edge did not reveal the presence of iron ions in the pMMO crystal. The putative dinuclear copper site cannot be identified as such by the X-ray crystallographic analysis because the EXAFS distance reported ( $\text{Cu–Cu} = 2.57\text{ \AA}$ ) is shorter than the  $2.8\text{-\AA}$  maximum resolution of the PDB: 1YEW data.

In general, the collection and analysis of data at different wavelengths is an important strategy for identifying metal ions and for solving the phase problem for protein X-ray crystallographic structures. Tunable synchrotron radiation sources allow data collection near the absorption edge of a specific metal ion. Absorption of radiation by the metal ion leads to anomalous scattering, a useful breakdown in Friedel's law (which states that  $I(h\ k\ l) = I((-h\ -k\ -l))$ ). This effect, detected in anomalous difference Fourier and Patterson maps, is exploited in the multiwavelength anomalous dispersion (MAD) and single-

wavelength anomalous dispersion (SAD) phasing techniques. The SAD technique collects one dataset at one wavelength. This yields two possible phasing solutions. The mean is taken of these to calculate an electron density map sufficiently detailed to identify the metal ion's location. The MAD technique collects multiple data sets at different wavelengths. This is useful to identify different metal ions because the anomalous dispersion value varies strongly with wavelength.

In the PDB: 1YEW structure, the zinc ion site in pmoC is not identified in pMMO by other spectroscopic techniques such as ICP-AES. The reference 190 authors believe that this zinc ion is either adventitious (perhaps depositing from the crystallization buffer) or in a location usually occupied by another metal ion, possibly another copper ion or an iron ion *in vivo*.

The reference 190 authors come to the following conclusions for the PDB: 1YEW pMMO structure: (1) At 2.80-Å resolution and using SAD and MAD data collection techniques, the metal ions and their coordinating amino acid ligands can be identified; however, determining meaningful bond distances are not possible; and (2) at 2.80-Å resolution, small exogenous ligands and water (solvent or ligating) cannot be identified or placed with accuracy, although the dinuclear copper site in pMMO probably has bridging ligands. In general, X-ray crystallography is an excellent technique for determining metal ion stoichiometry, locating metal ions within a protein fold, and identifying amino acid ligands. Using anomalous data collection techniques, SAD and especially MAD, researchers can identify specific metal ions in the crystal. Determining metal ion oxidation states in crystals is problematic, and any determination should be confirmed by single-crystal spectroscopy or other experimental techniques. Bond distances and angles (especially for light atoms), as well as detection of exogenous ligands, depend critically on the resolution of the crystal structure. Very few crystal structures are determined at the 1.8- to 2.0-Å resolution needed to achieve these measurements accurately. The authors end on a hopeful note, stating that the technological advances in synchrotron radiation and in crystallographic hardware and software, plus the use of robotics for protein expression and crystallization, will result in more and better X-ray crystallographic structures of biomolecules.

## 7.10 CONCLUSIONS

Chapter 7 has reported on the importance of iron in biological species. Because iron is the most abundant transition metal found in biological species, one would expect a wide variety of iron-containing proteins and metalloenzymes. Only a few of these have been treated in any detail in this chapter. Little or no mention has been made of how or why iron ions evolved to be the most biologically abundant transition metal ions; probably their usefulness in redox situations and for electron transport has something to do with their popularity. Iron homeostasis in biological species has not been discussed, although this

important topic is a subject of much research, especially into the disease states caused by faulty iron storage and/or distribution. Several red blood cell diseases—anemia, iron overload (hemochromatosis), and sickle cell anemia—occur because of faults in iron homeostasis. Hematological malignancies such as lymphoma (Hodgkin's disease and others) sicken and kill many people each year. Readers interested in these topics will find whole websites and many research publications devoted to them.

## REFERENCES

1. Roat-Malone, R. M. *Bioinorganic Chemistry: A Short Course*, John Wiley & Sons, Hoboken, NJ, 2002.
2. Hay, R. W. *Bio-Inorganic Chemistry*, Ellis Horwood Limited, Halsted Press, New York, 1984.
3. Nobbs, C. L.; Watson, H. C.; Kendrew, J. C. *Nature*, 1966, **209**, 339–341.
4. Fermi, G. *J. Mol. Biol.*, 1975, **97**(2), 237–256.
5. Fermi, G.; Perutz, M. F.; Shaanan, B.; Fourme, R. *J. Mol. Biol.*, 1984, **175**, 159–174. (PDB: 2HHB, 3HHB, 4HHB)
6. Jameson, G. B.; Molinaro, F. S.; Ibers, J. A.; Collman, J. P.; Brauman, J. I.; Rose, E.; Suslick, K. S. *J. Am. Chem. Soc.*, 1980, **102**, 3224–3237.
7. Phillips, S. E. *J. Mol. Biol.*, 1980, **142**, 531–554. (PDB: 1MBO)
8. Cowan, J. A. *Inorganic Biochemistry, An Introduction*, 2nd ed., Wiley-VCH, New York, 1997.
9. Momenteau, M.; Reed, C. A. *Chem. Rev.*, 1994, **94**, 659–698.
10. Lippard, S. J.; Berg, J. M. *Principles of Bioinorganic Chemistry*, University Science Books, Mill Valley, CA, 1994.
11. Perutz, M. *Mechanisms of Cooperativity and Allosteric Regulation in Proteins*, Cambridge University Press, New York, 1990.
12. Collman, J. P. *Inorg. Chem.*, 1997, **36**, 5145–5154.
13. Collman, J. P.; Gagne, R. R.; Reed, C. A.; Robinson, W. T.; Rodley, C. A. *Proc. Natl. Acad. Sci. USA*, 1974, **71**, 1326–1329.
14. Takano, T. *J. Mol. Biol.*, 1977, **110**, 569–584. (PDB: 3MBN, 5MBN)
15. Perutz, M. F.; Fermi, G.; Luisi, B.; Shaanan, B.; Liddington, R. C. *Acc. Chem. Res.*, 1987, **20**, 309–321.
16. Shaanan, B. *J. Mol. Biol.*, 1983, **171**, 31–59. (PDB: 1HHO)
17. Jameson, G. B.; Rodley, G. A.; Robinson, W. T.; Gagne, R. R.; Reed, C. A.; Collman, J. P. *Inorg. Chem.*, 1978, **17**, 850–857.
18. Jameson, G. B.; Molinaro, F. S.; Ibers, J. A.; Collman, J. P.; Brauman, J. I.; Rose, E.; Suslick, K. S. *J. Am. Chem. Soc.*, 1978, **100**, 6769–6770.
19. Suslick, K. S.; Reinert, T. *J. J. Chem. Ed.*, 1985, **62**(11), 974–983.
20. Jameson, G. B.; Ibers, J. A. in Bertini, I.; Gray, H.; Lippard, S. J.; Valentine, J. S. *Bioinorganic Chemistry*, University Science Books, Mill Valley, CA 1994.

21. Collman, J. P.; Gagne, R. R.; Reed, C. A.; Halbert, T. R.; Lang, G.; Robinson, W. T. *J. Am. Chem. Soc.*, 1975, **97**, 1427–1439.
22. Collman, J. P.; Reed, C. A. *J. Am. Chem. Soc.*, 1973, **95**, 2048–2049.
23. Collman, J. P.; Gagne, R. R.; Halbert, T. R.; Marchon, J. C.; Reed, C. A.; Halbert, T. R.; Lang, G.; Robinson, W. T. *J. Am. Chem. Soc.*, 1973, **95**, 7868–7870.
24. Collman, J. P.; Boulatov, R.; Sunderland, C. J.; Fu, L. *Chem. Rev.*, 2004, **104**, 561–588.
25. Rodgers, K. R.; Spiro, T. G. *Science*, 1994, **265**, 1697–1699.
26. Srajer, V.; Teng, T.-Y.; Ursby, T.; Pradervand, D.; Ren, Z.; Adachi, S.-I.; Schildkamp, W.; Bourgeois, D.; Wulff, M.; Moffat, K. *Science*, 1996, **274**, 1726–1729.
27. Ambler, R. P. *Biochim. Biophys. Acta*, 1991, **1058**, 42–47.
28. Bertini, I.; Cavallaro, G.; Rosato, A. *Chem. Rev.*, 2006, **106**, 90–115.
29. Keilin, D. *Proc. R. Soc. London, Ser. B: Biol. Sci.*, 1925, **98**, 312.
30. Palmer, G.; Reedijk, J. *Biochim. Biophys. Acta*, 1991, **1060**, 599.
31. Walker, F. A. *Chem. Rev.*, 2004, **104**, 589–615.
32. Lippard, S. J.; Berg, J. M. *Principles of Bioinorganic Chemistry*, University Science Books, Mill Valley, CA 1994, pp. 303–311.
33. Lipscomb, J. D. *Biochemistry*, 1980, **19**, 3590–3599.
34. Denisov, I. G.; Makris, T. M.; Sligar, S. G.; Schlichting, I. *Chem. Rev.*, 2005, **105**, 2253–2277.
35. Schlichting, I.; Berendzen, J.; Chu, K.; Stock, A. M.; Maves, S. A.; Benson, D. E.; Sweet, R. M.; Ringe, D.; Petsko, G. A.; Sligar, S. G. *Science*, 2000, **287**, 1615–1622. (PDB: 1DZ4, 1DZ6, 1DZ8, 1DZ9)
36. Poulos, T. L.; Finzel, B. C.; Howard, A. J. *J. Mol. Biol.*, 1987, **195**, 687–700. (PDB: 2CPP)
37. Wade, R. C.; Winn, P. J.; Schlichting, I. *J. Inorg. Biochem.*, 2004, **98**, 1175–1182. (PDB: 1UYU)
38. (a) Denisov, I. G.; Makris, T. M.; Sligar, S. G. *Methods Enzymol.*, 2002, **357**, 103. (b) Ibrahim, M.; Denisov, I. G.; Makris, T. M.; Kincaid, J. R.; Sligar, S. G. *J. Am. Chem. Soc.*, 2003, **125**, 13714. (c) Denisov, I. G.; Makris, T. M.; Sligar, S. G. *J. Biol. Chem.*, 2001, **276**, 11648. (d) Davydov, R.; Ledbetter-Rogers, A.; Martasek, P.; Larukhin, M.; Sono, M.; Dawson, J. H.; Masters, B. S. S.; Hoffman, B. M. *Biochemistry*, 2002, **41**, 10375. (e) Davydov, R.; Macdonald, I. D. G.; Makris, T. M.; Sligar, S. G.; Hoffman, B. M. *J. Am. Chem. Soc.*, 1999, **121**, 10654. (f) Davydov, R. M.; Yoshida, T.; Ikeda-Saito, M.; Hoffman, B. M. *J. Am. Chem. Soc.*, 1999, **121**, 10656. (g) Davydov, R.; Matsui, T.; Fujii, H.; Ikeda-Saito, M.; Hoffman, B. M. *J. Am. Chem. Soc.*, 2003, **125**, 16208. (h) Davydov, R.; Satterlee, J. D.; Fujii, H.; Sauer-Masarwa, A.; Busch, D. H.; Hoffman, B. M. *J. Am. Chem. Soc.*, 2003, **125**, 16340. (i) Denisov, I. G.; Ikeda-Saito, M.; Yoshida, T.; Sligar, S. G. *FEBS Lett.*, 2002, **532**, 203. (j) Makris, T. M.; Davydov, R.; Denisov, I. G.; Hoffman, B. M.; Sligar, S. G. *Drug Metab. Rev.*, 2002, **34**, 691. (k) Ibrahim, M.; Kincaid, J. R. *J. Porphyrins Phthalocyanins*, 2004, **8**, 215.
39. Jin, S.; Bryson, T. A.; Dawson, J. H. *J. Biol. Inorg. Chem.*, 2004, **9**, 644–653.
40. (a) Egawa, T.; Shimada, H.; Ishimura, Y. *Biochem. Biophys. Res. Commun.*, 1994, **201**, 1464–1468. (b) Kellner, D. G.; Hung, S. C.; Weiss, K. E.; Sligar, S. G. *J. Biol. Chem.*, 2002, **277**, 9641–9644.

41. Kumar, D.; de Visser, S. P.; Shaik, S. *Chem. Eur. J.*, 2005, **11**, 2825–2835.
42. (a) Martinis, S. A.; Atkins, W. M.; Stayton, P. S.; Sligar, S. G. *J. Am. Chem. Soc.*, 1989, **111**, 9252–9253. (b) Imai, M.; Shimada, H.; Watanabe, Y.; Matsushima-Hibiya, Y.; Makino, R.; Koga, H.; Horiuchi, T.; Ishimura, Y. *Proc. Natl. Acad. Sci. USA*, 1989, **86**, 7823–7827.
43. Vaz, A. D. N.; McGinnity, D. F.; Coon, M. J. *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 3555–3560.
44. (a) Jin, S.; Makris, T. M.; Bryson, T. A.; Sligar, S. G.; Dawson, J. H. *J. Am. Chem. Soc.*, 2003, **125**, 3406–3407. (b) Gelb, M. H.; Malkonen, P.; Sligar, S. G. *Biochem. Biophys. Res. Commun.*, 1982, **104**, 853–858. (c) Maryniak, D. M.; Kadkhodayan, S.; Crull, G. B.; Bryson, T. A.; Dawson, J. H. *Tetrahedron*, 1993, **49**, 9373–9384.
45. Davydov, R.; Perera, R.; Jin, S.; Yang, T.-C.; Bryson, T. A.; Sono, M.; Dawson, J. H.; Hoffmann, B. M. *J. Am. Chem. Soc.*, 2005, **127**, 1403–1413.
46. (a) de Visser, S. P.; Ogliaro, F.; Harris, N.; Shaik, S. *J. Am. Chem. Soc.*, 2001, **123**, 3037–3047. (b) de Visser, S. P.; Ogliaro, F.; Sharma, P. K.; Shaik, S. *J. Am. Chem. Soc.*, 2002, **124**, 11809–11826. (c) Ogliaro, F.; Cohen, S.; de Visser, S. P.; Shaik, S. *J. Am. Chem. Soc.*, 2000, **122**, 12892–12893. (d) de Visser, S. P.; Ogliaro, F.; Sharma, P. K.; Shaik, S. *Angew. Chem. Int. Ed.*, 2002, **41**, 1947–1951.
47. Ogliaro, F.; de Visser, S. P.; Cohen, S.; Sharma, P. K.; Shaik, S. *J. Am. Chem. Soc.*, 2002, **124**, 2806–2817.
48. Kumar, D.; Hirao, H.; de Visser, S. P.; Zheng, J.; Wang, D.; Thiel, W.; Shaik, S. *J. Phys. Chem. B*, 2005, **109**(42), 19946–19951.
49. (a) Vidakovic, M.; Sligar, S. G.; Li, H.; Poulos, T. L. *Biochemistry*, 1998, **37**, 9211–9214. (b) Gerber, N. C.; Sligar, S. G. *J. Am. Chem. Soc.*, 1992, **114**, 8742–8743. (c) Aikens, J.; Sligar, S. G. *J. Am. Chem. Soc.*, 1994, **116**, 1143–1144. (d) Gerber, N. C.; Sligar, S. G. *J. Biol. Chem.*, 1994, **269**, 4260–4266.
50. Taraphder, S.; Hummer, G. *J. Am. Chem. Soc.*, 2003, **125**, 3931–3940.
51. (a) Fulop V.; Phizackerley, R. P.; Soltis, S. M.; Clifton, I. J.; Wakatsuki, S.; Erman, J.; Hajdu, J.; Edwards, S. L. *Structure*, 1994, **2**(3), 201–208. (b) Gouet, P.; Jouve, H. M.; Williams, P. A.; Andersson, I.; Andreoletti, P.; Nussaume, L.; Hajdu, J. *Nat. Struct. Biol.*, 1996, **3**(11), 951–956.
52. Li, H.; Narasimhulu, S.; Havran, L. M.; Winkler, J. D.; Poulos, T. L. *J. Am. Chem. Soc.*, 1995, **117**, 6297–6299.
53. (a) Groves, J. T.; Haushalter, R. C.; Nakamura, M.; Nemo, T. E.; Evans, B. J. *J. Am. Chem. Soc.*, 1981, **103**, 2884–2886. (b) Fujii, H. *Coord. Chem. Rev.*, 2002, **226**, 51–60. (c) Dolphin, D.; Traylor, T. G.; Xie, L. Y. *Acc. Chem. Res.*, 1997, **30**, 251–259. (c) Goh, Y. M.; Nam, W. *Inorg. Chem.*, 1999, **38**, 914–920. (d) Dawson, J. H. *Science*, 1998, **240**, 433–439. (e) Gross, Z.; Nimri, S. *Inorg. Chem.*, 1994, **33**, 1731–1732. (f) Gross, Z.; Nimri, S.; Barzilay, C. M.; Simkovich, L. *J. Biol. Inorg. Chem.*, 1997, **2**, 492–506.
54. (a) Groves, J. T.; Watanabe, Y. *Inorg. Chem.*, 1987, **26**, 785–786. (b) Watanabe, Y.; Yamaguchi, K.; Morishima, I.; Takehira, K.; Shimizu, M.; Hayakawa, T.; Orita, H. *Inorg. Chem.*, 1991, **30**, 2581–2582. (c) Machii, K.; Watanabe, Y.; Morishima, I. *J. Am. Chem. Soc.*, 1995, **117**, 6691–6697. (d) Nam, W.; Lim, M. H.; Lee, H. J.; Kim, C. *J. Am. Chem. Soc.*, 2000, **122**, 6641–6647. (e) Nam, W.; Jin, S. W.; Lim, M. H.; Ryu, J. Y.; Kim, C. *Inorg. Chem.*, 2002, **41**, 3647–3652. (f) Suzuki, N.; Higuchi, T.; Nagano, T. *J. Am. Chem. Soc.*, 2002, **124**, 9622–9628. (g) Collman, J. P.; Chi  n, A. S.;

- Eberspacher, T. A.; Brauman, J. I. *J. Am. Chem. Soc.*, 2000, **122**, 11098–11100. (h) Nam, W.; Lim, M. H.; Moon, S. K.; Kim, C. *J. Am. Chem. Soc.*, 2000, **122**, 10805–10809. (i) Nam, W.; Lee, H. J.; Oh, S. Y.; Kim, C.; Jang, H. G. *J. Inorg. Biochem.*, 2000, **80**, 219–225. (j) Wadhwani, P.; Mukherjee, M.; Bandyopadhyay, D. *J. Am. Chem. Soc.*, 2001, **123**, 12430–12431. (k) Khavasi, H. R.; Davarani, S. S. H.; Safari, N. *J. Mol. Catal. A*, 2002, **188**, 115–122.
55. (a) Nanthakumar, A.; Goff, H. M. *J. Am. Chem. Soc.*, 1990, **112**, 4047–4049. (b) Yamaguchi, K.; Watanabe, Y.; Morishima I. *J. Chem. Soc. Chem. Commun.*, 1992, 1721–1723. (c) Murakami T.; Yamaguchi, K.; Watanabe, Y.; Morishima, I. *Bull. Chem. Soc. Jpn.*, 1998, **71**, 1343–1353. (d) Ogliaro, F.; de Visser, S. P.; Groves, J. T.; Shaik, S. *Angew. Chem. Int. Ed.*, 2001, **40**, 2874–2878. (e) Dey, A.; Ghosh, A. *J. Am. Chem. Soc.*, 2002, **124**, 3206–3207. (f) Nam, W.; Choi, S. K.; Lim, M. H.; Rohde, J.-U.; Kim, I.; Kim, J.; Kim, C.; Que, L., Jr. *Angew. Chem. Int. Ed.*, 2003, **42**, 109–111.
56. (a) Weiss, R.; Gold, A.; Trautwein, A. X.; Terner, J. In: Kadish, K. M.; Smith, K. M.; Guilard, R., eds., *The Porphyrin Handbook*, Vol. 4, Academic Press, New York, 2000, pp. 65–96. (b) Chin, D.-H.; La Mar, G. N.; Balch, A. L. *J. Am. Chem. Soc.*, 1980, **102**, 5945–5947. (c) Groves, J. T.; Gross, Z.; Stern, M. K. *Inorg. Chem.*, 1994, **33**, 5065–5072. (d) Groves, J. T.; Stern, M. K. *J. Am. Chem. Soc.*, 1988, **110**, 8628–8638. (e) Nam, W.; Park, S.-E.; Lim, I. K.; Lim, M. H.; Hong, J.; Kim, J. *J. Am. Chem. Soc.*, 2003, **125**, 14674–14675.
57. Nam, W.; Ryu, Y. O.; Song, W. *J. J. Biol. Inorg. Chem.*, 2004, **9**, 654–660.
58. Qin, J.; Perera, R.; Lovelace, L. L.; Dawson, J. H.; Lebioda, L. *Biochemistry*, 2006, **45**, 3170–3177. (PDB: 1EV)
59. (a) Sono, M.; Roach, M. P.; Coulter, E. D.; Dawson, J. H. *Chem. Rev.*, 1996, **96**, 2841–2887. (b) Dawson, J. H. *Science*, 1988, **240**, 433–439.
60. Wertz, D. L.; Valentine, J. S. *Struct. Bonding*, 2000, **97**, 37–60.
61. Sisemore, M. F.; Burstyn, J. N.; Valentine, J. S. *Angew. Chem. Int. Ed. Engl.*, 1996, **35**, 206–208.
62. Selke, M.; Valentine, J. S. *J. Am. Chem. Soc.*, 1998, **120**, 2652–2653.
63. Goto, Y.; Wada, S.; Morishima, I.; Watanabe, Y. *J. Inorg. Biochem.*, 1998, **69**, 241–247.
64. (a) Groves, J. T.; Nemo, T. E.; Myers, R. S. *J. Am. Chem. Soc.*, 1979, **101**, 1032–1033. (b) Groves, J. T.; Nemo, T. E. *J. Am. Chem. Soc.*, 1983, **105**, 5786–5791. (c) Groves, J. T.; McClusky, G. A.; White, R. E.; Coon, M. J. *Biochem. Biophys. Res. Commun.*, 1978, **81**, 154–160. (d) Groves, J. T. *J. Chem. Ed.*, 1985, **62**, 928–931.
65. Bill, E.; Ding, X. Q.; Bominaar, E. L.; Trautwein, A. X.; Winkler, H.; Mandon, D.; Weiss, R.; Gold, A.; Jayaraj, K.; Hatfield, W. E.; Kirk, M. L. *Eur. J. Biochem.*, 1990, **188**, 665–672.
66. (a) Hashimoto, S.; Tatsuno, Y.; Kitagawa, T. *J. Am. Chem. Soc.*, 1987, **109**, 8096–8097. (b) Kincaid, J. R.; Schneider, A. J.; Paeng, K. J. *J. Am. Chem. Soc.*, 1989, **111**, 735–737.
67. Penner-Hahn, J. E.; McMurry, T. J.; Renner, M. W.; Latos-Grazynski, L.; Eble, K. S.; Davis, I. M.; Balch, A. L.; Groves, J. T.; Dawson, J. H.; Hodgson, K. O. *J. Biol. Chem.*, 1983, **258**, 12761–12764. (b) Groves, J. T.; McMurry, T. J. *Rev. Port. Quim.*, 1985, **27**, 102–103.
68. Gross, Z.; Nimri, S.; Barzilay, C. M.; Simkhovich, L. *J. Biol. Inorg. Chem.*, 1997, **2**, 492–506.

69. Bukowski, M. R.; Koehntop, K. D.; Stubna, A.; Bominaar, E. L.; Halfen, J. A.; Münck, E.; Nam, W.; Que, L., Jr. *Science*, 2005, **310**, 1000–1002.
70. (a) Schoneboom, J. C.; Lin, H.; Reuter, N.; Thiel, W.; Cohen, S.; Ogliaro, F.; Shaik, S. *J. Am. Chem. Soc.*, 2002, **124**(27), 8142–8151. (b) Ogliaro, F.; de Visser, S. P.; Shaik, S. *J. Inorg. Biochem.*, 2002, **91**, 554–567.
71. Nelson, N.; Ben-Shem, A. *Bioessays*, 2005, **27**(9), 914–922.
72. Stroebel, D.; Choquet, Y.; Popot, J.-L.; Picot, D. *Nature*, 2003, **426**, 413–418. (PDB: 1Q90)
73. Kurisu, G.; Zhang, H.; Smith, J. L.; Cramer, W. A. *Science*, 2003, **302**, 1009–1014. (PDB: 1VF5)
74. Jordan, P.; Fromme, P.; Witt, H. T.; Klukas, O.; Saenger, W.; Krauss, N. *Nature*, 2001, **411**, 909–917. (1JB0)
75. (a) Zouni, A.; Witt, H. T.; Kern, J.; Fromme, P.; Krauss, N.; Saenger, W.; Orth, P. *Nature*, 2001, **409**, 739–743. (PDB: 1FE1) (b) Kamiya, N.; Shen, J.-R. *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 98–103. (PDB: 1IZL)
76. Alric, J.; Pierre, Y.; Picot, D.; Lavergne, J.; Rappaport, F. *Proc. Natl. Acad. Sci. USA*, 2005, **102**(44), 15860–15865.
77. (a) Brandt, U.; Trumppower, B. *Crit. Rev. Biochem. Mol. Biol.*, 1994, **29**, 165–197. (b) Trumppower, B. L.; Gennis, R. B. *Annu. Rev. Biochem.*, 1994, **63**, 675–716. (c) Trumppower, B. L. *J. Biol. Chem.*, 1990, **265**, 11409–11412.
78. Saratse, M. *Science*, 1999, **283**, 1488–1493.
79. Xia, D.; Yu, C.-A.; Kim, H.; Xia, J.-Z.; Kachurin, A. M.; Zhang, L.; Yu, L.; Deisenhofer, J. *Science*, 1997, **277**, 60–66. (PDB: 1QCR)
80. Mitchell, P. J. *Nature*, 1961, **191**, 144–148.
81. (a) Mitchell, P. J. *Theor. Biol.*, 1976, **62**, 327–367. (b) Crofts, A. R. In Martonosi, A. N., ed., *The Enzymes of Biological Membranes*, Vol. 4, Plenum, NY, 1985, pp. 347–382.
82. Braun, H. P.; Emmermann, M.; Kruft, V.; Schmitz, U. K. *EMBO J.*, 1992, **11**, 3219–3227.
83. Iwata, S.; Lee, J. W.; Okada, K.; Lee, J. K.; Iwata, M.; Rasmussen, B.; Link, T. A.; Ramaswamy, S.; Jap, B. K. *Science*, 1998, **281**, 64–71. (PDB: 1BE3, 1BGY)
84. Zhang, Z.; Huang, L.; Shulmeister, V. M.; Chi, Y.-I.; Kim, K. K.; Hung, L.-W.; Crofts, A. R.; Berry, E. A.; Kim, S.-H. *Nature*, 1998, **392**, 677–684. (PDB: 1BCC, 3BCC)
85. Smith, J. L. *Science*, 1998, **281**, 58–59.
86. Gao, X.; Wen, X.; Esser, L.; Yu, L.; Yu, C. A.; Xia, D. *Biochemistry*, 2003, **42**, 9067–9080. (PDB: 1NTK, 1NTM, 1NTZ, 1NU1)
87. Crofts, A. R. *Annu. Rev. Physiol.*, 2004, **66**, 689–733.
88. Esser, L.; Quinn, B.; Li, Y. F.; Zhang, M.; Elberry, M.; Yu, L.; Yu, C. A.; Xia, D. *J. Mol. Biol.*, 2004, **341**, 281–302. (PDB: 1SQB, 1SQP, 1SQQ, 1SQV, 1SQX)
89. von Jagow, G.; Ljungdahl, P. O.; Graf, P.; Ohnishi, T.; Trumppower, B. L. *J. Biol. Chem.*, 1984, **259**(10), 6318–6326.
90. Huang, L.-S.; Cobelli, D.; Tung, E. Y.; Berry, E. A. *J. Mol. Biol.*, 2005, **351**, 573–597. (PDB: 1PP9, 1PPJ)
91. Berden, J. A.; Slater, E. C. *Biochim. Biophys. Acta*, 1972, **256**, 199–215.

92. Yu, C. A.; Nagaoka, S.; Yu, L.; King, T. E. *Biochem. Biophys. Res. Commun.*, 1978, **82**, 1070–1078.
93. (a) Brandt, U.; von Jagow, G. *Eur. J. Biochem.*, 1991, **195**, 163–170. (b) Zhang, Z.; Huang, L.; Shulmeister, V. M.; Chi, Y.-I.; Kim, K. K.; Hung, L.-W.; Crofts, A. R.; Berry, E. A.; Kim, S.-H. *Nature*, 1998, **392**, 677–684. (c) Brandt, U. *Biochim. Biophys. Acta*, 1998, **1365**, 261–268. (d) Brandt, U. *J. Bioenerg. Biomembr.*, 1999, **31**, 243–250.
94. Kim, H.; Esser, L.; Hossain, M. B.; Xia, D.; Yu, C.-A.; Rizo, J.; van der Helm, D.; Deisenhofer, J. *J. Am. Chem. Soc.*, 1999, **121**, 4902–4903.
95. (a) Lange, C.; Nett, J. H.; Trumppower, B. L.; Hunte, C. *EMBO J.*, 2001, **20**, 6591–6600. (PDB: 1KB9) (b) Palsdottir, H.; Lojero, C. G.; Trumppower, B. L.; Hunte, C. *J. Biol. Chem.*, 2003, **278**, 31303–31311. (PDB: 1P84)
96. Lange, C.; Hunte, C. *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 2800–2805. (PDB: 1KYO)
97. Berry, E. A.; Huang, L. S.; Saechao, L. K.; Pon, N. G.; Valkova-Valchanova, M.; Daldal, F. *Photosynth. Res.*, 2001, **81**, 251–275. (PDB: 1ZRT)
98. Mulkidjianian, A. Y. *Biochim. Biophys. Acta*, 2005, **1709**(1), 5–34.
99. Kuchar, J.; Hausinger, R. P. *Chem. Rev.*, 2004, **104**, 509–525.
100. Scott, R. A.; Mauk, A. G. *Cytochrome c—A Multidisciplinary Approach*, University Science Books, Sausalito, CA, 1996.
101. Bertini, I.; Cavallaro, G.; Rosato, A. *Chem. Rev.*, 2006, **106**, 90–115.
102. Orengo, C. A.; Michie, A. D.; Jones, S.; Jones, D. T.; Swindells, M. B.; Thornton, J. M. *Structure*, 1997, **5**, 1093–1108.
103. Lo Conte, L.; Ailey, B.; Hubbard, T. J.; Brenner, S. E.; Murzin, A. G.; Chothia, C. *Nucleic Acids Res.*, 2000, **28**, 257–259.
104. Jiang, X.; Wang, X. *Annu. Rev. Biochem.*, 2004, **73**, 87–106.
105. Pearce, D. A.; Sherman, F. *J. Biol. Chem.*, 1995, **270**, 20879–20882.
106. Pettigrew, G. W.; Moore, G. R. *Cytochromes c; Biological Aspects*, Springer-Verlag, Berlin, 1987.
107. Dudkina, N. V.; Eubel, H.; Keegstra, W.; Boekema, E. J.; Braun, H. P. *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 3225–3229.
108. (a) Maneg, O.; Ludwig, B.; Malatesta, F. *J. Biol. Chem.*, 2003, **278**, 46734–46740. (b) Witt, H.; Malatesta, F.; Nicoletti, F.; Brunori, M.; Ludwig, B. *J. Biol. Chem.*, 1998, **273**, 5132–5136.
109. Louie, G. V.; Brayer, G. D. *J. Mol. Biol.*, 1990, **214**, 527–555. (PDB: 1YCC)
110. Berghuis, A. M.; Brayer, G. D. *J. Mol. Biol.*, 1992, **223**, 959–976. (PDB: 2YCC)
111. Stickle, D. F.; Presta, L. G.; Dill, K. A.; Rose, G. D. *J. Mol. Biol.*, 1992, **220**, 1143–1159.
112. Berghuis, A. M.; Guillemette, J. G.; Smith, M.; Brayer, G. D. *J. Mol. Biol.*, 1994, **235**, 1326–1341. (PDB: 1CTY, 1CTZ)
113. Berghuis, A. M.; Guillemette, J. G.; McLendon, G.; Sherman, F.; Smith, M.; Brayer, G. D. *J. Mol. Biol.*, 1994, **236**, 786–799. (PDB: 1CRG, 1CRH, 1CRI, 1CRJ)
114. Rafferty, S. P.; Guillemette, J. G.; Berghuis, A. M.; Smith, M.; Brayer, G. D.; Mauk, A. G. *Biochemistry*, 1996, **35**, 10784–10792.

115. Takano, T.; Dickerson, R. E. *Proc. Natl. Acad. Sci. USA*, 1980, **77**, 6371–6375. (PDB: 3CYT)
116. Bushnell, G. W.; Louie, G. V.; Brayer, G. D. *J. Mol. Biol.*, 1990, **214**, 585–595. (PDB: 1HRC)
117. Sivakolundu, S. G.; Mabrouk, P. A. *J. Biol. Inorg. Chem.*, 2003, **8**, 527–539. (PDB: 1LC1, 1LC2)
118. Banci, L.; Bertini, I.; Huber, J. G.; Spyroulias, G. A.; Turano, P. *J. Biol. Inorg. Chem.*, 1999, **4**, 21–31. (PDB: 1GIW)
119. (a) Colon, W.; Wakem, L. P.; Sherman, F.; Roder, H. *Biochemistry*, 1997, **36**, 12535–12541. (b) Russell, B. S.; Melenkivitz, R.; Bren, K. L. *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 8312–8317.
120. Sivakolundu, S. G.; Mabrouk, P. A. *J. Inorg. Biochem.*, 2003, **94**(4), 381–385.
121. Döpner, S.; Hildebrandt, P.; Rosell, F. I.; Mauk, A. G. *J. Am. Chem. Soc.*, 1998, **120**, 11246–11255.
122. Sanishvili, R.; Volz, K. W.; Westbrook, E. M.; Margoliash, E. *Structure*, 1995, **3**, 707–716. (PDB: 1CRC)
123. Pelletier, H.; Kraut, J. *Science*, 1992, **258**, 1748–1755. (PDB: 2PCB)
124. Tezcan, F. A.; Findley, W. M.; Crane, B. R.; Ross, S. A.; Lyubovitsky, J. G.; Gray, H. B.; Winkler, J. R. *Proc. Natl. Acad. Sci. USA*, 2002, **99**(13), 8626–8630. (PDB: 1LFM)
125. Tezcan, F. A.; Winkler, J. R.; Gray, H. B. *J. Am. Chem. Soc.*, 1999, **121**, 11918–11919.
126. Lyubovitsky, J. G.; Gray, H. B.; Winkler, J. R. *J. Am. Chem. Soc.*, 2002, **124**, 5481–5485.
127. (a) Yeh, S.-R.; Rousseau, D. L. *Nat. Struct. Biol.*, 2000, **7**(6), 443–445. (b) Akiyama, S.; Takahashi, S.; Ishimori, K.; Morishima, I. *Nat. Struct. Biol.*, 2000, **7**, 514–520.
128. Winkler, J. R.; Gray, H. B. *Chem Rev.*, 1992, **92**, 369–379.
129. Moser, C. C.; Keske, J. M.; Wanrcke, K.; Farid, R. S.; Dutton, P. P. *Nature*, 1991, **355**, 796–802.
130. Pelletier, H.; Kraut, J. *Science*, 1992, **258**, 1748–1755. (PDB: 2PCB, 2PCC)
131. Guo, M.; Bhaskar, B.; Li, H.; Barrows, T. P.; Poulos, T. L. *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 5940–5945. (PDB: 1S6V)
132. Nocek, J. M.; Zhou, J. S.; DeForest, S.; Priyadarshy, S.; Beratan, D. N.; Onuchic, J. N.; Hoffman, B. M. *Chem. Rev.*, 1996, **96**, 2459–2489.
133. Kang, S. A.; Crane, B. R. *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 15465–15470. (PDB: 2B10, 2B11, 2B12, 2B0Z)
134. (a) Rafferty, S. P.; Pearce, L. L.; Barker, P. D.; Guillemette, J. G.; Kay, C. R.; Smith, M.; Mauk, A. G. *Biochemistry*, 1990, **29**, 9365–9369. (b) Lo, T. P.; Guillemette, G.; Louie, G. V.; Smith, M.; Brayer, G. D. *Biochemistry*, 1995, **34**, 163–171. (c) Louie, G. V.; Pielak, G. J.; Smith, M.; Brayer, G. D. *Biochemistry*, 1998, **27**, 7870–7876.
135. Everest, A. M.; Wallin, S. A.; Stemp, E. D. A.; Nocek, J. M.; Mauk, A. G.; Hoffman, B. M. *J. Am. Chem. Soc.*, 1991, **113**, 4337–4338.
136. Hancock, J. T.; Desikan, R.; Neill, S. J. *Free Radic. Biol. Med.*, 2001, **31**(5), 697–703.
137. (a) Tsukihara, T.; Aoyama, H.; Yamashita, E.; Tomizaki, T.; Yamaguchi, H.; Shinzawa-Itoh, K.; Nakashima, R.; Yaono, R.; Yoshikawa, S. *Science*, 1995, **269**,

- 1069–1074. (b) Tsukihara, T.; Aoyama, H.; Yamashita, E.; Tomizaki, T.; Yamaguchi, H.; Shinzawa-Itoh, K.; Nakashima, R.; Yaono, R.; Yoshikawa, S. *Science*, 1996, **272**, 1136–1144. (PDB: 1OCC). (c) Yoshikawa, S.; Shinzawa-Itoh, K.; Nakashima, R.; Yaono, R.; Yamashita, E.; Inoue, N.; Yao, M.; Fei, M. J.; Libeu, C. P.; Mizushima, T.; Yarnaguchi, H.; Tomizaki, T.; Tsukihara, T. *Science*, 1998, **280**, 1723–1729. (PDB: 1OCR, 2OCC, 1OCO, 1OCZ)
138. Kim, E.; Chufán, E. E.; Kamaraj, K.; Karlin, K. D. *Chem. Rev.*, 2004, **104**, 1077–1133.
139. Qin, L.; Hiser, C.; Mulichak, A.; Garavito, R. M.; Ferguson-Miller, S. *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 16117–16122. (PDB: 2GSM)
140. Collman, J. P.; Boulatov, R.; Sunderland, C. J.; Fu, L. *Chem. Rev.*, 2004, **104**(2), 561–588.
141. (a) Pinakoulaki, E.; Pfitzner, U.; Ludwig, B.; Constantinos, V. C. *J. Biol. Chem.*, 2002, **277**, 13563–13568. (b) Proshlyakov, D. A.; Pressler, M. A.; DeMaso, C.; Leykam, J. F.; DeWitt, D. L.; Babcock, G. T. *Science*, 2000, **290**, 1588–1591. (c) Morgan, J. E.; Verkhovsky, M. I.; Palmer, G.; Wikström, M. *Biochemistry*, 2001, **40**, 6882–6892.
142. Tsukihara, T.; Shimokata, K.; Katayama, Y.; Shimada, H.; Muramoto, K.; Aoyama, H.; Mochizuki, M.; Shinzawa-Itoh, K.; Yamashita, E.; Yao, M.; Ishimura, Y.; Yoshikawa, S. *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 15304–15309. (PDB: 1V54, 1V55)
143. Namslauer, A.; Bzezinski, P. *FEBS Lett.*, 2004, **567**, 103–110.
144. Svensson-Ek, M.; Abramson, J.; Larsson, G.; Tornroth, S.; Brzezinski, P.; Iwata, S. *J. Mol. Biol.*, 2002, **321**, 329–339. (PDB: 1M56, 1M57)
145. (a) Iwata, S.; Ostermeier, C.; Ludwig, B.; Michel, H. *Nature*, 1995, **376**, 660–669. (b) Wikström, M.; Bogachev, A.; Finel, M.; Morgan, J. E.; Puustinen, A.; Raitio, M.; Verkhovskaya, M.; Verkhovsky, M. I. *Biochim. Biophys. Acta Bioenerg.*, 1994, **1187**, 106–111.
146. (a) Puustinen, A.; Wikström, M. *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 35–37. (b) Wikström, M.; Verkhovsky, M. I.; Hummer, G. *Biochim. Biophys. Acta*, 2003, **1604**, 61–65. (c) Mills, D. A.; Ferguson-Miller, S. *Biochim. Biophys. Acta*, 2002, **1555**, 96–100. (d) Behr, J.; Michel, H.; Mäntele, W.; Hellwig, P. *Biochemistry*, 2000, **39**, 1356–1363.
147. (a) Ruitenberg, M.; Kannt, A.; Bamberg, E.; Ludwig, B.; Michel, H.; Fendler, K. *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 4632–4636. (b) Ruitenberg, M.; Kannt, A.; Bamberg, E.; Fendler, K.; Michel, H. *Nature*, 2002, **417**, 99–102. (c) Verkhovsky, M. I.; Belevich, N.; Morgan, J. E.; Wikström, M. *Biochim. Biophys. Acta*, 1999, **1412**, 184–189. (d) Forte, E.; Barone, M. C.; Brunori, M.; Sarti, P.; Giuffre, A. *Biochemistry*, 2002, **41**, 13046–13052. (e) Forte, E.; Scandurra, F. M.; Richter, O. M.; D’Itri, E.; Sarti, P.; Brunori, M.; Ludwig, B.; Giuffre, A. *Biochemistry*, 2004, **43**, 2957–2963.
148. Wikström, M.; Verkhovsky, M. I.; Hummer, G. *Biochim. Biophys. Acta*, 2003, **1604**, 61–65.
149. Gunter, M. J.; Mander, L. N.; Murray, K. S. *J. Chem. Soc., Chem. Commun.*, 1981, 799–801.
150. Collman, J. P. *Acc. Chem. Res.*, 1977, **10**, 265–272.
151. Ghiladi, R. A.; Hatwell, K. R.; Karlin, K. D.; Huang, H.-W.; Moeenne-Loccoz, P.; Krebs, C.; Huynh, B. H.; Marzilli, L. A.; Cotter, R. J.; Kaderli, S.; Zuberbuehler, A. D. *J. Am. Chem. Soc.*, 2001, **123**, 6183–6184.

152. Chufán, E. E.; Karlin, K. D. *J. Am. Chem. Soc.*, 2003, **125**, 16160–16161.
153. Del Rio, D.; Sarangi, R.; Chufán, E. E.; Karlin, K. D.; Hedman, B.; Hodgson, K. O.; Solomon, E. I. *J. Am. Chem. Soc.*, 2005, **127**(34), 11969–11978.
154. (a) Collman, J. P.; Sunderland, C. J.; Berg, K. E.; Vance, M. A.; Solomon, E. I. *J. Am. Chem. Soc.*, 2003, **125**, 6648–6649. (b) Liu, J. G.; Naruta, Y.; Tani, F. *Angew. Chem., Int. Ed.*, 2005, **44**, 1836–1840.
155. Chishiro, T.; Shimazaki, Y.; Tani, F.; Tachi, Y.; Naruta, Y.; Karasawa, S.; Hayami, S.; Maeda, Y. *Angew. Chem., Int. Ed.*, 2003, **42**, 2788–2791.
156. Ostermeier, C.; Harrenga, A.; Ermler, U.; Michel, H. *Proc. Natl. Acad. Sci. USA*, 1997, **94**, 10547–10553. (PDB: 1AR1)
157. Soulimane, T.; Buse, G.; Bourenkov, G. P.; Bartunik, H. D.; Huber, R.; Than, M. E. *EMBO J.*, 2000, **19**, 1766–1776. (PDB: 1EHK)
158. Ju, T. D.; Ghiladi, R. A.; Lee, D.-H.; van Strijdonck, G. P. F.; Woods, A. S.; Cotter, R. J.; Young, V. G., Jr.; Karlin, K. D. *Inorg. Chem.*, 1999, **38**, 2244–2245.
159. Obias, H. V.; van Strijdonck, G. P. F.; Lee, D.-H.; Ralle, M.; Blackburn, N. J.; Karlin, K. D. *J. Am. Chem. Soc.*, 1998, **120**, 9696–9697.
160. Robin, M. B.; Day, P. *Adv. Inorg. Chem. Radiochem.*, 1967, **10**, 247–422.
161. (a) Harding, C. J.; McKee, V.; Nelson, J. *J. Am. Chem. Soc.*, 1991, **113**, 9684–9685. (b) Harding, C. J.; Nelson, J.; Symons, M. C. R.; Wyatt, J. *J. Chem. Soc. Chem. Commun.*, 1994, 2499–2500.
162. Barr, M. E.; Smith, P. H.; Antholine, W. E.; Spencer, B. *J. Chem. Soc. Chem. Commun.*, 1993, 1649–1652.
163. Farrar, J. A.; McKee, V.; Al-Obaidi, A. H. R.; McGarvey, J. J.; Nelson, J.; Thomson, A. *J. Inorg. Chem.*, 1995, **34**, 1302–1303.
164. LeCloux, D. D.; Davydov, R.; Lippard, S. J. *Inorg. Chem.*, 1998, **37**, 6814–6826.
165. Gupta, R.; Zhang, Z. H.; Powell, D.; Hendrich, M. P.; Borovik, A. S. *Inorg. Chem.*, 2002, **41**, 5100–5106.
166. Houser, R. P.; Young, V. G., Jr.; Tolman, W. B. *J. Am. Chem. Soc.*, 1996, **118**, 2101–2102.
167. (a) Gamelin, D. R.; Randall, D. W.; Hay, M. T.; Houser, R. P.; Mulder, T. C.; Canters, G. W.; de Vries, S.; Tolman, W. B.; Lu, Y.; Solomon, E. I. *J. Am. Chem. Soc.*, 1998, **120**, 5246–5263. (b) Williams, K. R.; Gamelin, D. R.; LaCroix, L. B.; Houser, R. P.; Tolman, W. B.; Mulder, T. C.; de Vries, S.; Hedman, B.; Hodgson, K. O.; Solomon, E. I. *J. Am. Chem. Soc.*, 1997, **119**, 613–614. (c) George, S. D.; Metz, M.; Szilagyi, R. K.; Wang, H.; Cramer, S. P.; Lu, Y.; Tolman, W. B.; Hedman, B.; Hodgson, K. O.; Solomon, E. I. *J. Am. Chem. Soc.*, 2001, **123**, 5757–5767.
168. Collman, J. P.; Boulatov, R.; Sunderland, C. J.; Fu, L. *Chem. Rev.*, 2004, **104**, 561–588.
169. (a) Pesavento, R. P.; Pratt, D. A.; Jeffers, J.; van der Donk, W. A. *Dalton Trans.*, 2006, 3326–3337. (b) Pratt, D. A.; Pesavento, R. P.; van der Donk, W. A. *Org. Lett.*, 2005, **7**, 2735–2738. (c) Kim, E.; Kamaraj, K.; Galliker, B.; Rubie, N. D.; Moenne-Loccoz, P.; Kaderli, S.; Zuberbuhler, A. D.; Karlin, K. D. *Inorg. Chem.*, 2005, **44**(5), 1238–1247. (d) McCauley, K. M.; Vrtis, J. M.; Dupont, J.; Van der Donk, W. A. *J. Am. Chem. Soc.*, 2000, **122**, 2403–2404. (e) Collman, J. P.; Wang, Z.; Zhong, M.; Zeng, L. *J. Chem. Soc., Perkin Trans. 1*, 2000, 1217–1222. (f) Kamaraj, K.; Kim,

- E.; Gallikler, B.; Zakharov, L. N.; Rheingold, A. L.; Zuberbuehler, A. D.; Karlin, K. D. *J. Am. Chem. Soc.*, 2003, **125**, 6028–6029. (g) He, Z.; Colbran, S. B.; Craig, D. C. *Chem. Eur. J.*, 2003, **9**, 116–129.
170. (a) Beinert, H.; Thomson, A. J. *Arch. Biochem. Biophys.*, 1983, **222**, 333–361. (b) Beinert, H.; Emptage, M. H.; Dreyer, J.-L.; Scott, R. A.; Hahn, J. E.; Hodgson, K. O.; Thomson, A. J. *Proc. Natl. Acad. Sci. USA*, 1983, **80**, 393–396. (c) Kennedy, M. C.; Emptage, M. H.; Dreyer, J. L.; Beinert, H. *J. Biol. Chem.*, 1983, **258**, 11098–11105. (d) Kent, T. A.; Dreyer, J.-L.; Kennedy, M. C.; Huynh, B. H.; Emptage, M. H.; Beinert, H.; Münck, E. *Proc. Natl. Acad. Sci. USA*, 1982, **79**, 1096–1100.
171. (a) Kennedy, M. C.; Werst, M.; Tesler, J.; Emptage, M. H.; Beinert, H. M.; Hoffman, B. M. *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 8854–8858. (b) Tesler, J.; Emptage, M. H.; Kennedy, M. C.; Beinert, H.; Hoffman, B. M. *J. Biol. Chem.*, 1986, **261**, 4840–4846.
172. Lauble, H.; Kennedy, M. C.; Beinert, H.; Stout, C. D. *Biochemistry*, 1992, **31**, 2735–2748. (PDB: 7ACN, 8ACN)
173. Dupuy, J.; Volbeda, A.; Carpentier, P.; Darnault, C.; Moulis, J. M.; Fontecilla-Camps, J. C. *Structure*, 2006, **14**, 129–139. (PDB: 2B3X, 2B3Y)
174. Wang, Z.; Li, C.; Ellenburg, M.; Soistman, E.; Ruble, J.; Wright, B.; Ho, J. X.; Carter, D. C. *Acta Cryst., D*, 2006, **62**, 800–806. (PDB: 2FFX, 2FG4, 2FG8)
175. Hempstead, P. D.; Yewdall, S. J.; Fernie, A. R.; Lawson, D. M.; Artymiuk, P. J.; Rice, D. W.; Ford, G. C.; Harrison, P. M. *J. Mol. Biol.*, 1997, **268**, 424–448. (PDB: 1AEW (horse), 2FHA (human))
176. Donlin, M. J.; Frey, R. F.; Putnam, C.; Proctor, J. K.; Bashkin, J. K. *J. Chem. Ed.*, 1998, **75**(4) 437–441.
177. Carrondo, M. A. *EMBO J.*, 2003, **22**(9), 1959–1968.
178. Taft, K. L.; Papaefthymiou, G. C.; Lippard, S. J. *Science*, 1993, **259**, 1302–1305.
179. Taft, K. L.; Papaefthymiou, G. C.; Lippard, S. J. *Inorg. Chem.*, 1994, **33**, 1510–1520.
180. Holmes, M. A.; Le Trong, I.; Turley, S.; Sieker, L. C.; Stenkamp, R. E. *J. Mol. Biol.*, 1991, **218**, 583–593. (PDB: 1HMO, 1HMD)
181. (a) Uppenberg, J.; Lindqvist, F.; Svensson, C.; Ek-Rylander, B.; Andersson, G. *J. Mol. Biol.*, 1999, **290**, 201–211. (PDB: 1QFC) (b) Strater, N.; Jasper, B.; Scholte, M.; Krebs, B.; Duff, A. P.; Langley, D. B.; Han, R.; Averill, B. A.; Freeman, H. C.; Guss, J. M. *J. Mol. Biol.*, 2005, **351**, 233–246. (PDB: 1WAR, 2BQ8)
182. (a) Hanson, R. S.; Hanson, T. E. *Microbiol. Rev.*, 1996, **60**, 439–471. (b) Sullivan, J. P.; Dickinson, D.; Chase, C. A. *Crit. Rev. Microbiol.*, 1998, **24**, 335–373.
183. Noodleman, L.; Lovell, T.; Han, W.-G.; Li, J.; Himo, F. *Chem. Rev.*, 2004, **104**, 459–508.
184. (a) Yu, S. S.; Chen, K. H.; Tseng, M. Y.; Wang, Y. S.; Tseng, C. F.; Chen, Y. J.; Huang, D. S.; Chan, S. I. *J. Bacteriol.*, 2003, **185**, 5915–5924. (b) Lieberman, R. L.; Shrestha, D. B.; Doan, P. E.; Hoffman, B. M.; Stemmler, T. L.; Rosenzweig, A. C. *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 3820–3825. (c) Choi, D. W.; Kunz, R. C.; Boyd, E. S.; Semrau, J. D.; Antholine, W. E.; Han, J. I.; Zahn, J. A.; Boyd, J. M.; de la Mora, A. M.; DiSpirito, A. A. *J. Bacteriol.*, 2003, **185**, 5755–5764. (d) Nguyen, H. H.; Elliott, S. J.; Yip, J. H.; Chan, S. I. *J. Biol. Chem.*, 1998, **273**, 7957–7966. (e) Basu, P.; Katterle, B.; Andersson, K. K.; Dalton, H. *Biochem. J.*, 2003, **369**, 417–427.

185. Chan, S. I.; Chen, K. H.-C.; Yu, S. S.-F.; Chen, C.-L.; Kuo, S. S.-J. *Biochemistry*, 2004, **43**, 4421–4430.
186. Lemos, S. S.; Yuan, H.; Perille-Collins, M. L. *Curr. Top. Biophys.*, 2002, **26**, 43–48.
187. Liberman, R. L.; Rosenzweig, A. C. *Crit. Rev. Biochem. Mol. Biol.*, 2004, **39**, 147–164.
188. Liberman, R. L.; Rosenzweig, A. C. *Nature*, 2005, **434**, 177–182. (PDB: 1YEW)
189. Mirica, L. M.; Ottenwalder, S.; Stack, R. D. P. *Chem. Rev.*, 2004, **104**, 1013–1045.
190. Sommerhalter, M.; Lieberman, R. L.; Rosenzweig, A. C. *Inorg. Chem.*, 2005, **44**, 770–778.

# INDEX

Following the page number, the letter f indicates the entry is a figure; t, a table

- Ab initio calculations, 119, 161, 170–174, 180
- Absorption, electron paramagnetic resonance, EPR, 123–126
- Absorption, X-ray absorption spectroscopy, XAS, EXAFS, 78–83
- ABX system, 106
- Acid-base, see Hard-soft acid-base theory
- Aconitase, 21, 44, 132, 454–456 see also Iron-sulfur clusters
- Active transport, 193, 197–200, 327
- Adenine (A), 47–48, 48t, 50f, 54, 61, 141, 263, 273, 276, 276f, 334  
deazaadenine, 276, 276f  
in hammerhead ribozyme, 263, 273, 280–283, 289  
nicotinamide adenine dinucleotide (NADH, NAD<sup>+</sup>), 24, 24f, 44, 44f, 212, 388
- metal binding, 47, 48t, 141, 201–203, 280–283, 289
- methyl adenine, 276–277, 276f
- Adenosine nucleotide:
  - in group I intron, 252
  - in hammerhead ribozyme, 277
- Adenosine monophosphate (AMP), 193–194, 203
- Adenosine diphosphate (ADP), 45, 192, 336, 388, 410
- Adenosine triphosphate (ATP), 26, 45, 48t, 95, 192, 193, 383, 388–389, 410, 430
- Adenosine triphosphate analogs, see ATP analogs
- Adenosine triphosphatases, see also P-type ATPase, 327
- A DNA, 50, 52f, 53t
- Agarose gel, 59
- Alcohol dehydrogenase, 45, 66
- Aldehyde, 45
  - deformylation, 375, 382
  - phenylacetaldehyde (2-hydroxystyrene), 367, 379
- preparation of porphyrins, 354

- Alkali metals ( $\text{Na}^+$ ,  $\text{K}^+$ ): see also  $\text{Na}^+/\text{K}^+$   
 ATPase and potassium ( $\text{K}^+$ ) ion channels  
 biological functions, 1, 2t, 4t, 6, 7t, 9, 9t, 44, 190, 192, 197t  
 $\text{Ca}^{2+}$ -ATPase and, 329, 336  
 coordination geometry, 2t, 4t, 9, 9t  
 facilitated diffusion and, 197–199, 198f  
 homeostasis, 189–190, 192, 193–200, 198f  
 kinetics, 11t  
 passive diffusion and, 195–196  
 thermodynamics, 9t, 193–199
- Alkaline earth metals ( $\text{Mg}^{2+}$ ,  $\text{Ca}^{2+}$ ): see also  $\text{Ca}^{2+}$ -ATPase, calmodulin, troponin, ribozymes  
 biological functions, 2t, 4t, 44, 191, 197, 197t  
 $\text{Ca}^{2+}$ -ATPase and, 44, 95–97, 327–329  
 calcium-dependent molecules and, 301–302  
 calmodulin and, 302–326  
 coordination geometry, 4t  
 homeostasis, 193, 194–195, 199  
 ion channels and, 199, 204, 215  
 matrix metalloproteinase (MMP) and, 120, 121f  
 ribozymes and, 238–301, 254f, 256f, 260f  
 troponin and, 3, 44, 194, 301–302, 306, 338
- Allosteric, 44
- Allosteric and nonallosteric enzymes, 44–46, 46f
- Allosteric Interactions See cooperativity, hemoglobin, hemocyanin, R state/T state
- Alpha helix, 34–35, 36f, 36t, 37f  
 $\text{Ca}^{2+}$ -ATPase and, 333–334, 337  
 calmodulin and, 303–308, 305f, 308f, 310–311, 316, 318–319, 318f, 320–326  
 cytochrome bc<sub>1</sub> iron sulfur protein and, 168, 393–394, 394f, 395–396, 402, 406f  
 cytochrome b(6)f and, 387, 387t  
 cytochrome c 360 and, 408, 411, 412f, 421
- cytochrome c oxidase and, 429  
 gated channels and, 197  
 hemoglobin and myoglobin and, 348, 358  
 lysozyme and, 37–38  
 matrix metalloproteinase and, 120–121, 121f  
 $\text{Na}^+/\text{K}^+$ -ATPase and, 201  
 potassium ion channels and, 209, 212, 215–217, 217f, 218–234, 219f  
 Rieske protein and, 383  
 zinc fingers and, 63–73, 68f, 70f, 71f
- American Chemical Society, 183
- AMBER force field, 119, 161, 166, 175, 179–180
- Amino Acids, 30–33, 31t, 32t  
 abbreviations, 31t, 32t  
 D/L configuration, 332  
 essential, 30  
 structures, 31t, 32t  
 interactions with metal ions, 32–33, 33t
- Amphipathic (amphiphilic), 37, 233, 313, 331, 334, 404
- Amplification, 69, 428
- Anaerobic bacteria, 24, 455
- Anions, 7  
 $\beta$ -mercaptoethanol(BME) thiolate, 373  
 biological phosphorus and, 193  
 cytochrome P450 models and, 380, 381f  
 peroxyo, 82, 349, 363, 366, 368, 372, 374f, 382, 433, 433–439, 440f, 441–442, 443, 454  
 porphyrin(s) and, 374, 376, 378, 380  
 superoxide radical anion, 82, 346, 347f, 349, 363, 364, 371, 396, 428, 435, 440f, 442–443, 440f, 443, 453  
 triflate, 451
- Annual Reviews, 183
- Antibodies, 69, 215, 218, 225, 228–229
- Anticodon, 55
- Antiferromagnetism:  
 antiferromagnetic coupling, 17  
 cytochrome c oxidase and, 461  
 cytochrome c oxidase models and, 437, 438, 440, 441, 449  
 hemoglobin and myoglobin and, 348

- Apoenzyme, 44  
Acquired drug resistance (ADR), 43, 55  
Area detectors, 95–96  
Arginine, 32f, 57f, 151  
  Ca<sup>2+</sup>-ATPase and, 333  
  calmodulin and, 316, 321  
  cytochrome c and, 415, 421f  
  methane monooxygenase and, 459  
  Na<sup>+</sup>/K<sup>+</sup>-ATPase and, 203  
  potassium ion channels and, 214–218, 221–222, 226–227, 229, 232–234  
  zinc fingers and, 68–69  
Arsenic (As), 2, 2t  
Ascorbate oxidase, 461  
Asparagine (Asn, N) 31f, 56  
  cytochrome c and, 410, 411  
Aspartic Acid, aspartate (Asp, D), 32–33, 32f, 33f  
aconitase and, 454, 455f, 457t, 456–458  
Ca<sup>2+</sup>-ATPase and, 328–329, 331–332, 332f, 332t, 333–337  
calmodulin and, 303, 303f, 305, 312f, 312t, 319, 322, 326  
cytochrome bc<sub>1</sub> and, 399, 402–406, 406f  
cytochrome c and, 388  
cytochrome c oxidase and, 425, 429, 429–432, 433f  
cytochrome P450 and, 364f, 368, 370f, 371–372  
matrix metalloproteinase and, 120, 121f  
methane monooxygenase and, 459  
Na<sup>+</sup>/K<sup>+</sup>-ATPase and, 200–202  
potassium ion channels and, 214, 216, 217f, 221, 223, 227, 230–231  
P-type ATPases and, 327–329  
zinc fingers and, 68, 71, 71f, 72t  
Associative mechanism (A, S<sub>N</sub>2), 10  
  Ca<sup>2+</sup>-ATPase and, 335–337  
ribozymes and, 258–262, 264, 268, 272, 279, 283–284, 286–287, 294, 296–297, 300  
Atomic Force Microscopy, 139–143  
A-T pairs, 47, 49, 50f, 141, 254f  
ATP, see adenosine triphosphate  
ATP analogs, 329, 333–334  
ATPases, see also Ca<sup>2+</sup>-ATPase, Na<sup>+</sup>/K<sup>+</sup>-ATPase, 44–45, 95–97, 383  
A-U pairs, 49, 249, 253, 254f, 261, 263, 285, 297  
AU/TG codon, 56  
Automated sequencing technology, 61–62  
Avogadro's number, 13  
Azurin, 4, 57–58  
  
B<sub>12</sub> see also coenzyme B<sub>12</sub>, cobalamin, vitamin B<sub>12</sub>, 6t  
Back-bonding, 346, 347f  
BamHI, restriction enzyme, 59  
Barium (Ba), 191t, 193, 195  
Base-pairing, nucleic acids  
  Hoogsteen, 49, 244, 257  
  Watson-Crick, 47–50, 50f, 240f, 244, 253, 263, 285–286, 297  
Bases, see nucleobases  
B DNA, 47, 50, 52f, 53t, 66  
Beryllium (Be), 11t, 190t  
Beta pleated sheet, 34–35, 36f, 37f  
  cytochrome c and, 409  
  in lysozyme and, 37f, 38  
  in zinc fingers and, 64–65, 65f, 67, 70f, 72  
  parallel, antiparallel, 36f, 37f  
β-mercaptoproethanol(BME) thiolate, 373  
Bioavailability of metal ions, see homeostasis  
Bioorganometallic chemistry, 19–22  
  16–18e rule, 19–20, 19t, 20f  
  iron-sulfur clusters, 21–22  
Bohr effect, see hemoglobin  
Bohr magneton, 122  
Boltzmann distribution, 100, 124  
Boltzmann equilibrium, 109  
Bone, 44, 140, 191t, 192, 194  
Bragg's law, 79–80, 88–89, 92  
Brookhaven protein data bank, see Protein Data Bank, PDB  
  
Ca<sup>2+</sup>-ATPase, 191t, 200, 201  
  aspartylphosphate, 329, 337  
  Ca<sup>2+</sup>-ATPase cycle, 328f, 329–337  
  calcium binding, 329–333, 336–337  
  calcium pump, 96, 190t, 319–321, 329

- comparison to  $\text{Na}^+/\text{K}^+$ -ATPase, 201–203  
 EF hand, 191t, 302–304, 306, 319, 324, 326  
 homeostasis and, 194–195  
 introduction to, 301, 327–329  
 thapsigargin and, 329, 331, 335, 336  
 X-ray crystallography, 96–97, 330–337, 330f, 332f, 332t
- Cadmium (Cd), 2, 7t, 174  
 calcium homeostasis and, 195  
 hammerhead ribozyme folding and, 290  
 $\text{Mg}^{2+}$  substitute in ribozymes, 267t  
 PS rescue experiment and, 242, 268, 280–283, 288
- Calcium ion, see also  $\text{Ca}^{2+}$ -ATPase, calmodulin, troponin  
 biological distribution, 3, 4t, 44, 191t, 193, 197, 204, 301  
 calcium pump, 96, 190t, 319–321, 329  
 coordination chemistry, 4t  
 homeostasis, 194–195, 199  
 in  $\text{Ca}^{2+}$ -ATPase, 95–97, 301, 327–337, 330f, 332f, 332t  
 in calmodulin, 301–327  
 in matrix metalloproteinase, 120, 121f  
 in troponin, 301, 302  
 ion channels and, 204, 215  
 ligand preferences, 4t  
 phosphate minerals in bone, 191t, 192–193, 194, 199
- Calmodulin:  
 apo-calmodulin, 304–309, 305f, 307f, 309t  
 $\text{Ca}^{2+}$  binding, 194, 195, 301–304, 306–308, 306f  
 drug interactions, 308–313, 312f  
 introduction to, 302–303  
 KAR-2 and, 311–313, 311f  
 NMR solution structures, 306–308, 319  
 peptide binding to, 313–326, 314–315t, 318f  
 trifluoperazine and, 308–313, 310f, 312f  
 vinblastine and, 311–313, 311f  
 X-ray crystallography, 303–308, 305f, 323–325
- Calsequestrin, 301–302  
 Capped porphyrin, see porphyrin  
 Carbon Dioxide, 43, 45  
 Bohr effect and, 192  
 Carbon Monoxide:  
 as a ligand, 7t, 8, 16, 19, 20  
 binding to iron porphyrins, 353, 357–358, 357f  
 binding to hemoglobin and myoglobin, 349–350, 356–359  
 $^{13}\text{C}$  NMR, 105f  
 cytochrome P450 and, 361  
 cytochrome c oxidase and, 440, 445  
 distal histidine interactions, 357, 357f  
 flash photolysis from  $\text{CcO}$ , 145  
 organometallics and, 20–21  
 spectrochemical series and, 16  
 time-resolved crystallography, 146–148
- (Carbonmonoxy)hemoglobin, 349, 356–359  
(Carbonmonoxy)myoglobin, 349, 356–359
- Carboxypeptidase, 40
- Carboxyl group, carboxylate, see also aspartic acid, glutamic acid, 4t, 5t, 6t, 9t  
 biotin and, 220f  
 citrate, isocitrate and *cis*-aconitate, 451, 451f  
 peptides, 30, 32, 34, 34f, 40, 42
- Catalase, 82, 366, 371, 372, 378, 380
- Catalysis:  
 catalyst, 22, 29–30, 162, 262, 272, 300, 366, 369, 378, 378f, 464  
 enzyme catalysis, 4, 6t, 29–30, 172, 191t  
 organometallic, 22  
 ribozymes and, 240, 241, 243–244, 247, 249–250, 262–263, 268, 272–278, 280–281, 283–284, 286–292, 296–301, 336  
 transition state, 171, 227, 242–244, 258, 260f, 261–262, 268, 272, 275–278, 275f, 281–288, 292–293, 296, 300–301, 334–336
- Catalytic RNA, see ribozyme(s)
- CD, see circular dichroism

- Charge carriers, metals in biological systems, 4t
- cDNA, 59–60, 60f
- Celera Genomics, 62–63
- Cesium (Cs), 190t, 193, 211
- CHARMM force field, 167–169, 177–178, 180
- Ceruloplasmin, Cp, 461
- Chemical Abstracts, 183–184
- Chemical degradation method, 59, 152
- Chloride ion ( $\text{Cl}^-$ ), 7t, 19, 19t, 20f  
bridge in electron transfer, 23, 23f  
intra- and extracellular concentration, 193–194, 196, 199, 204
- Chloroethane, 106
- 1-chloropropane, 103
- Chromosome mapping, 141
- Chromatography: see also HPLC, GC-MS, 39–40, 149, 158  
affinity, 40  
gel filtration, 40  
ion exchange, 40
- Chromium (Cr):  
biological significance, 2, 7t, 10–11, 11t  
chromium(III), inert, 11, 11t  
chromium(II), labile, 10, 11t  
electron transfer and, 23, 23f  
Jahn-Teller effect and, 10–11, 18, 19f
- Circular dichroism (CD), see also magnetic circular dichroism (MCD) 40, 130, 311–313, 422–424
- CisDDP, cisPlatin, *cis*-dichlorodiammin eplatinum(II), 16, 20
- Cloning, see also gene cloning, 29, 57, 58–61, 60f, 62
- Cobalamins, see also coenzyme B<sub>12</sub>, vitamin B<sub>12</sub>, 6t, 22
- Cobalt (Co):  
biological significance, 2, 2t, 6t, 7t, 11, 11t, 44  
B- to Z-DNA transformation, 141  
cofactor and, 44  
cobalt NMR, 105f  
 $\text{Co}(\text{phen})_3]^{3+}$  as cytochrome c oxidant, 415–416  
cytochrome c folding and, 422–429  
electron transfer, 23, 23t
- hammerhead ribozyme and, 267t, 279, 287, 290
- inert, 11
- labile, 23
- ligand field stabilization energy, 18
- Mössbauer source, 133
- vitamin B<sub>12</sub>, coenzyme B<sub>12</sub>, 22, 44
- Codons  
genetic mutations, 56–57, 57f  
in genetic transmission, 55–56
- Coenzyme Q, coQ, see ubiquinone
- Coenzymes, 44
- Cofactors, see also prosthetic groups, 13, 24, 44, 117, 133  
nitrogenases and, 137–138, 137f, 139  
hemes, as, 117, 146
- Competitive Inhibitors, 46–47, 46f
- Compound I, see cytochrome c, cytochrome P450, see also ferryl-oxo intermediates
- Compound II, see cytochrome c, cytochrome P450
- Computer hardware:  
high performance cluster (HPC), 160  
Macintosh, 158, 175, 177, 178–181  
parallel computers, 159–160  
PC, 158–160, 175, 177, 180–181  
workstations, 158–159, 172, 174, 177–179, 181
- Computer software, graphics and calculation:  
AMBER, 119, 161, 166, 175, 179, 180  
AMPAC<sup>TM</sup>, 179  
CAChe<sup>TM</sup>, 176–177  
CambridgeSoft<sup>TM</sup>, 176, 179  
CHARMM, 168–169, 177–178–180  
Gaussian, 119, 168–170, 173, 176–177, 179, 182  
GROMOS, 174, 180  
HyperChem<sup>TM</sup>, 162, 164, 177–178  
Hypercube, 177  
MacroModel<sup>TM</sup>, 175, 178, 179  
Maestro<sup>TM</sup>, 178  
MOBY, 180  
MOMEC, 162, 164, 178  
MOPAC<sup>TM</sup>, 176, 179, 182  
PCMODEL<sup>TM</sup>, 164, 179  
PyMOL, 176  
Schrödinger, Inc., 178, 179

- SemiChem, 179  
 Serena Software, 164, 179  
 Spartan<sup>TM</sup>, 170, 175, 180  
 Tripos, 159, 175, 179  
 Wavefunction, Inc., 180  
 Windows, 158, 175–182
- Computer software, mathematical:  
 Mathematica, 180–181  
 MathType, 180  
 SigmaPlot, Deltagraph, 180
- Computer software, nomenclature:  
 ACD/Name, 182  
 AutoNom Standard, 181
- Computer software, visualization resources:  
 Chime, 35, 181–182  
 ISIS/Draw, 181–182  
 projected stereo images, 183  
 Protein Explorer, 182  
 PyMOL, 176  
 RasMol, 182
- Conformation  
 A-, B-, Z-DNA, 47, 50, 52f, 53, 53t, 66  
 Sugars in DNA and RNA, 33, 47, 48t, 240, 259, 289
- Cooperativity  
 calcium in calmodulin, 308  
 calcium in Ca<sup>2+</sup>-ATPase, 331  
 dioxygen binding in hemoglobin, 346, 359
- Coordination geometry  
 common transition metal, 14–16, 14f  
 ion size influence, 8  
 octahedral, 4t, 5t, 6t, 14f, 15f, 17–18, 17f, 18t, 21, 79, 81, 109, 126, 163, 167, 252, 258, 284, 379  
 square planar, 6t, 10, 14f, 15–17, 15f, 18t, 19, 81, 127, 163  
 square pyramid, 4t, 5t, 14f, 18t, 21, 82, 380, 451  
 tetrahedral, 4t, 5t, 6t, 14f, 15–18, 15f, 18t, 21, 63, 65–66, 79, 81, 120, 135t  
 trigonal bipyramidal, 16, 21, 257, 287, 296, 450  
 trigonal planar, 5t  
 trigonal prism, 21
- Coordination number, 4t, 5t, 6t, 10, 12, 18, 79, 81, 103, 162–163
- Copper, see also copper proteins:  
 antiferromagnetic coupling and, 17, 442  
 azurin and, 57–58  
 biological significance, 2t, 4, 5t, 6t, 361  
 cytochrome c oxidase and, 17, 129–131, 411–461  
 electron transfer and, 5t, 24, 26, 129, 132, 143–146  
 EXAFS studies, 81–82  
 model complexes, 439, 447, 450–453, 441f, 442f  
 X-ray source and, 94
- Copper proteins, see also azurin,  
 cytochrome c oxidase, methane monooxygenase:  
 copper-zinc superoxide dismutase, 143  
 electron transfer and, 5t, 24, 26, 57–58, 117, 129  
 ENDOR studies, 131–132  
 EPR studies and properties, 17, 129–131, 130f, 443, 445–447, 451–452, 456, 458, 460–461  
 EXAFS studies, 82–83, 82–83, 461  
 ligands and geometries, 5t, 6t, 7t, 10–11, 11t, 18  
 model complexes, 439, 441f–442f, 447, 450–453  
 monooxygenase activity, 459–460  
 NMR studies, 442, 453  
 Type I, II, III, 5t, 6t, 17, 166, 449–450  
 UV-visible spectroscopy, 144, 439, 442–444, 448t
- Correlation spectroscopy, see also nuclear magnetic resonance (NMR)/(COSY), 112–115
- Cross reactions, electron-transfer, 25, 25f
- Crown ethers, 6–7, 7f, 7t, 9t
- Crystallization techniques, 84–88, 96, 215, 228, 231, 241, 252, 258, 285, 292, 294, 297, 313, 399, 463, 465
- Crystallography, see X-ray crystallography
- Cysteine (Cys, C), 20, 22, 31f, 33f, 34f, 43  
 as a ligand, 24, 131–132, 137f, 138, 362–363, 373, 384, 450–451

- biotinylation and, 220–221, 220f  
cadmium binding, metallothionein, 174  
cystine dimer, 23–24  
cytochromes and, 362–363, 373, 384, 384f, 399, 408–411, 427–429, 455  
disulfide bond, 37–38, 37f  
fluorescence resonance energy transfer (FRET) and, 224–225  
iron binding in iron sulfur clusters, 137f, 138, 168–169, 454–456, 457f  
modes of interaction with metal ions, 33f  
zinc binding in zinc fingers, 63–65, 68, 68f, 70f, 72, 72t  
C-terminal, 34–35, 35f, 37f, 38, 40, 65, 151, 201–202, 208f, 214, 216, 410, 423, 459  
calmodulin and, 303, 303f, 306–326, 309t, 312f, 314–315t  
potassium ion channels and, 208f, 213–214, 216, 218, 220–233  
Curie temperature, 124, 136  
Cyclic adenosine monophosphate, (cAMP), 194, 325  
Cytochrome b(6)f  
electron transfer, 385–386, 385f  
proton translocation, 385f, 386  
Q-cycle mechanism, 386, 389–390  
Rieske iron sulfur protein (ISP), 22, 383–384, 385f, 386–387, 387t  
X-ray crystallography, 383, 385f, 386  
Cytochrome bc<sub>1</sub>  
electron transfer, 388–391, 396–399, 402, 405  
inhibitors, 396–404  
proton translocation, 389–391, 395–397  
Q-cycle mechanism, 396–397  
Rieske iron sulfur protein (ISP), 22, 389–391, 394–397, 399, 405  
stigmatellin and, 167–170, 385f, 391, 397, 398f, 400t, 402–405, 406f  
X-ray crystallography, 399, 405  
Cytochrome c  
Compound I, 424  
Compound II, 424  
electron transfer, 405, 408, 410–413, 415–416, 420–427  
ferryl-oxo iron species, see also compound I, 424  
mechanism of activity, 408, 417  
proton translocation, 410  
X-ray crystallography, 424  
Cytochrome c oxidase  
bimetallic copper site (Cu<sub>A</sub>), 129–131, 431–432, 433f–434f, 436f, 438  
electron transfer and, 5t, 26, 129, 132, 143–146, 426–427, 429–430, 434, 437–439, 449–450, 461  
EPR studies, 129–131, 433, 443, 445, 444f  
EXAFS studies, 82–83, 442, 444–445  
ferryl-oxo iron species, 435, 439  
introduction to, 429  
mechanism of activity, 439, 443, 453  
model compounds, 439–440, 443, 445, 447  
proton translocation and, 430, 434  
X-ray crystallography, 451–453, 461  
Cytochrome c peroxidase, 166, 359, 365–366, 369, 371–372, 377, 410, 421, 424–426, 426f  
Cytochrome P450  
compound I, 82, 172, 363, 365–368, 365f, 366f, 368f, 371–372, 376–378, 381  
compound II, 366  
concerted mechanism, 367–369  
electron transfer, 365, 378–379, 378f, 383  
ferryl-oxo iron species, see also compound I, 82, 172, 363–371  
model compounds, 372–382  
multistep mechanism, 378, 378f  
“oxygen rebound” mechanism, 375–376, 375f  
 $\pi$ -cation radical, 364–366, 370, 372, 375–376, 379, 382  
X-ray crystallography, 363, 369–372, 370f, 382  
Cytoplasm, 55, 96, 200, 408, 428  
Ca<sup>2+</sup>-ATPase and, 327–337  
definition, 327–328  
potassium ion channels and, 204, 213, 216, 221, 224–226, 229–233

- Cytosine (C), 47, 48t, 50f, 54, 61, 71, 71f, 141  
 hammerhead ribozyme and, 266, 275, 275f  
 Cytosol, 54, 193, 327, 428f, 438
- Databases, online, 42, 62, 84, 97, 117, 119, 139, 152, 159, 175, 181, 182, 184–185, 200, 205, 359
- d-d* transitions, 15, 78t, 450, 452  
 forbidden, 15
- d* electrons, 4t, 5t, 6t, 11, 11t, 14–18, 15f, 17f, 18f, 18t
- Degenerate energy states, 11, 14–15, 18
- Dehydrogenases, 4t, 45, 66, 388
- Deletions, chromosomal, 56–57, 57f, 190t
- Denaturation, 39, 269, 272, 331, 420, 422–424
- Density functional theory (DFT), 83, 161, 170, 171–173, 174–178, 180, 366–367, 380, 443
- Deoxyhemoglobin, 90–91, 344–346, 347, 349–350, 353  
 stereochemistry, 351, 351t  
 quaternary structure, 344–346, 345f
- Deoxymyoglobin, 147–148, 344–346, 347, 349–350, 351, 351t, 353
- Deoxynucleotide triphosphates (dNTPs), 55
- Deoxyribonucleic acid, see DNA
- Diamagnetism, 17, 102, 136, 138
- Dideoxy method, 59
- Diffraction:  
 atomic force microscopy, 140  
 multiwavelength anomalous diffraction (MAD), 92–93, 95, 241, 247, 283, 294–295, 358, 464–465  
 X-ray crystallography, 78–79, 83–98, 147–148, 211
- Dioxygen:  
 activation, 362  
 binding proteins, see also  
 hemoglobin, cytochrome P450, cytochrome c oxidase, methane monooxygenase, 30, 135t, 143, 146–148, 172, 192–193, 344–345  
 formation of superoxide anion, 346, 347f, 348  
 ground state, 348–349, 349f  
 monooxygenases and, 22  
 orientation in hemoglobin, 346–347, 347f  
 reactions, 62  
 singlet state, 162, 349  
 solubility, 344  
 superoxide ion radical, 62, 346–349, 347f, 363, 371, 395, 424  
 toxic radicals from, 363, 397  
 transport, 5t, 458–459  
 triplet state, 124
- Dioxygen binding:  
 binding mode to hemoglobin, 346, 347f, 348, 351, 356  
 cooperative, 91, 346, 348, 359  
 cytochrome P450 and, 172, 361–364, 364t, 367–372  
 cytochrome c oxidase and, 425, 429f, 429–443  
 distal effects  
 cytochrome P450 and, 172, 367–374, 373t  
 hemoglobin and, 346–348, 347f, 350–351, 350f, 354, 355f, 356–358, 357f, 384  
 hemerythrin and, 459  
 hemoglobin and, 135t, 192–193, 344–345, 348  
 kinetics, 346  
 methane monooxygenase and, 459–460  
 model compounds and, 90–91, 352–356, 374, 439–454  
 non-cooperative, 90–91  
 proximal effects:  
 cytochrome c oxidase and, 440–443  
 cytochrome P450 and, 368, 373–374, 373t  
 hemoglobin and, 346, 347, 348, 350–351, 350f, 354–355, 354f, 355f, 357f, 358, 384  
 thermodynamics, 346
- Dioxygenases, 383
- Diphosphoglycerate, (DPG), 192–193, 193f
- Dissociative mechanism (D, S<sub>N</sub>1), 10, 12

- Distal effects, see also dioxygen, 147, 172, 298–300
- DNA
- A-form, see A DNA
  - B-form, see B DNA
  - backbone structure, 50, 50f, 51f
  - base-pairing, 47, 49–50, 50f, 57
  - bending and unwinding, 54, 65
  - cleavage reagents, 59, 62
  - cloning, 29, 57, 58–61, 60f, 62
  - codons 55–57, 57f
  - conformation, 49–53, 53t
  - conformational recognition by zinc finger proteins, 65, 67
  - double helix, 30, 47, 49–51, 51f, 54–55, 60f, 61, 67, 70–72, 71f
  - exons, see also RNA, ribozyme, 58–59, 63
  - genetic transmission and, 53–56
  - grooves, major and minor, 47, 50, 53, 53t, 64–68, 71
  - helical structures, 30, 50–53, 52f, 53t
  - insertion into a plasmid, 57, 59–60, 60f
  - interactions with proteins, 63–73
  - nucleobases, 48–49t, 52f
  - polymorphs, 50, 63
  - primary strand, 65, 69, 71–73, 70f
  - repair, 26, 55, 57
  - replication, 53–55, 54f, 58, 140
  - sugar pucker, 50–51, 52f, 53t
  - synthesis, 55
  - transcription, 53–55, 54f, 57, 59, 63–66, 70, 72
  - translation, 53–57, 54f, 141, 148, 153
  - Z-form, see Z DNA
- DNA-binding proteins, see zinc finger proteins
- DNA-polymerase, 54–55, 61, 64, 241, 260
- Donnan equilibria, 195f
- Dose-response curve, essential elements, 2, 3f
- Double-stranded DNA (dsDNA), 30, 47, 49–50, 51f, 54–55, 61, 67, 70, 71f, 141
- Dysprosium (Dy), 103
- EDTA, see ethylenediaminetetraacetic acid
- 18-electron rule, 19–21, 19t
- Electromagnetic spectrum, 77–78, 78f
- Electron carriers, 4, 24
- Electron microscope, 142
- Electron-nuclear double resonance spectroscopy, ENDOR, 76, 129, 131–132, 137–139, 243, 282, 289, 367
- Electron paramagnetic resonance (EPR) spectroscopy, 122–132
- active spin states, 122–124, 123f
  - anisotropy, 125
  - applications, 122
  - axial spectra, 126–128, 126f, 131, 134, 136
  - characteristics of, 122–128
  - copper ion(s) and, 129–131
  - derivative spectrum, 124–126, 124f, 126f
  - descriptive examples, 17, 129–132, 137–139, 234, 243, 282, 288–289, 362–367, 376–379, 387, 454
  - Fe-S clusters, 137–139
  - g-value(s), 122–127, 130f
  - hemoglobin and, 17, 127
  - hyperfine coupling, 127–129, 131, 243, 289, 450–452, 460
  - iron-sulfur clusters, see Fe-S clusters
  - isotropic spectra, 126, 126f, 131
  - rhombic spectra, 126–127, 126f
  - Mössbauer and, 137
  - spin-orbit coupling, 125–127
  - superhyperfine coupling, 127–129
  - Zeeman splitting, 122, 125
  - zero-field splitting (zfs), 125, 127
- Electron-spin echo envelope modulation, ESEEM, 129, 243, 282, 289
- Electron transfer, see also Marcus theory, 22–26
- cofactors, 24, 24f
  - cross reaction, 25, 25f
  - cytochrome b(6)f and, see cytochrome b(6)f
  - cytochrome bc<sub>1</sub> and, see cytochrome bc<sub>1</sub>
  - cytochrome c and, see cytochrome c

- cytochrome c oxidase and, see cytochrome c oxidase  
 electron tunneling, 26, 427  
 Frank-Condon principle, 25  
 inert and labile ions, 23  
 inner-sphere mechanism, 23, 23f  
 metalloenzymes and, 24, 26  
 outer-sphere mechanism, 23, 25  
 self-exchange reaction, 25f
- Electron tunneling, 26, 427
- Electronic spectroscopy, 437  
 charge transfer transitions, 81, 450, 459  
*d-d* transitions, 15, 78t, 450, 452
- Electrophiles, 366–357, 366f
- Electrophoresis, 40–42, 59, 62, 143, 148, 290, 298
- Elements:  
 bulk, 1–2, 2t  
 composition of the human body, 2t  
 essential, 1–3, 4t, 5t, 6t  
 overview of biochemical roles, 4t, 5t, 6t  
 trace, 1–2  
 ultratrace, 2
- Endoplasmic reticulum (ER), 95–96, 194, 212, 327–328
- ENDOR, see Electron-nuclear double resonance spectroscopy
- Energy storage, see adenosine triphosphate (ATP)
- Enzyme kinetics:  
 competitive inhibitors, 46–47  
 initial rate, 45  
 Lineweaver-Burk plot, 45–47, 46f  
 Michaelis-Menten equation, 45  
 noncompetitive inhibition, 46–47  
 rate laws, 45
- EPR, see Electron paramagnetic resonance spectroscopy
- Equilibrium constants, 8, 12–13, 25, 27  
 passive diffusion and, 195–196, 196f, 198
- ESEEM, see Electron-spin echo envelope modulation
- Essential elements, 1–3, 4t, 5t, 6t
- Ethylenediaminetetraacetic acid (EDTA), 244, 248, 289, 415–416
- Eukaryotic, 53–54, 58–59, 66  
 $\text{Ca}^{2+}$ -ATPases and, 327  
 calmodulin and, 97, 302  
 cytochromes and, 82, 117, 129, 360, 388, 390, 408–429, 455  
 potassium ion channels and, 209, 212, 228–229  
 ribozymes and, 239, 262
- Europium (Eu), 103, 154
- Evolution, biological, 61, 249, 337, 410
- Ewald sphere, 85, 88–89, 85f, 89f
- Extended X-ray absorption fine structure (EXAFS); 76, 78–83  
 characteristics of, 78–81  
 cytochrome c and, 82–83  
 cytochrome c oxidase and, 131  
 cytochrome c oxidase models and, 439, 440, 440f–442f, 453–454  
 cytochrome P450 and, 82, 173  
 cytochrome P450 models and, 376–377, 381  
 descriptive examples, 81–83  
 methane monooxygenase and, 459–460  
 model compounds and, 352–353  
 nitrogenase and, 139
- Facilitated diffusion, 197–199  
 gated channels, see also voltage-gated channels, 197
- FAD see flavin adenine dinucleotide
- Fe-Fe bonds, 20f, 22
- Fe(TPP)(1-MeIm), 82, 135, 354, 356f
- Fe(TPP)(1,2-Me<sub>2</sub>Im), 365f
- Fe<sub>2</sub>S<sub>2</sub> centers, see iron-sulfur clusters, Rieske iron sulfur protein (ISP)
- Fe<sub>3</sub>S<sub>4</sub> centers, see iron-sulfur clusters
- Fe<sub>4</sub>S<sub>4</sub> centers, see iron-sulfur clusters
- Ferredoxins, 21, 24, 455
- Ferritin, 45, 458–459
- Ferryl-oxo intermediates, see hemoglobin, myoglobin, cytochrome c, cytochrome c oxidase, cytochrome P450
- First-order kinetics, see kinetics
- Flash photolysis, 144–145
- Flavin adenine dinucleotide (FAD), 460
- Flavin mononucleotide (FMN), 24
- Fletcher-Reeves method, 165

- Fluorescent resonance energy transfer (FRET), 153, 222, 224, 232, 242, 290–291, 298
- Footprinting, 242, 289–292, 291t, 300
- Force fields, 119, 161–169, 174–180
- Free induction decay (FID), 110–118, 112f, 113f, 115f
- Franck-Condon principle, 25
- Free radicals, 124, 248
- Fuoss-Eigen equation, 12
- g-value, see Electron paramagnetic resonance spectroscopy (EPR)
- Gallium (Ga), 7t, 11t, 173
- Gas chromatography, see GC-MS
- G-C pairs, 47, 49, 50f, 141, 247, 249, 251, 253, 253f, 263, 277, 294
- GC-MS, 149, 158
- Gel electrophoresis, 41–43, 59, 62, 269, 272, 290, 298
- Gel filtration chromatography, 40
- Gene expression, 56–58
- Gene regulation
- zinc as a factor in, 4t
- Genes
- cloning, 29, 57, 58–61, 60f, 62
  - dsDNA, 55, 72–73, 141
  - exons, 58–59, 63, 239, 240, 240f, 244–246, 248, 253–260, 259t, 301
  - functions of, 58, 63–65
  - introns, 58–59, 63, 239–261, 240f, 248t, 259t, 301
  - ssDNA
- Genetic engineering, 58, 62, 69
- Genetic information, transmission of, 53–56
- Genomics, 29–30, 61–63, 69
- Geometric structures, metals in biological systems, 4t, 5t, 6t, 6, 13–19
- Glutamic acid, glutamate (Glu, E), 32f, 33f
- as a ligand, 33f
  - calmodulin and, 308, 316, 322
  - cytochrome c oxidase and, 131, 429–451, 433f–434f, 436f, 440f–442f, 451f, 454, 461
- interaction with metal ions, 33f, 131, 231
- potassium ion channels and, 231
- Glutamine (Gln, Q), 31f, 58, 131
- Glutathione (GSH), 429f
- Glycine (Gly,G), 31f, 32, 56, 203, 215–216, 230–231, 373
- Glycosidic linkage, 47
- GROMOS force field, 174, 180, 188
- Group I intron, 246f, 256f
- base triple and, 244, 247, 249, 251, 252–253, 253f, 254f, 257, 259t, 261
  - cadmium (Cd) and, 242
  - exogenous guanosine, 240, 240f, 242, 244–245, 249, 259
  - exon-splicing mechanism, 240–241, 240f, 245, 245f
  - G-dependent mechanism, 245, 245f
  - internal guide sequence (IGS), 240–241, 245, 246, 255, 256, 259
  - magnesium and, 239, 241–244, 248, 248f, 249, 254, 252–257, 256f, 258–259, 259t, 260f, 261
  - manganese and, 243, 250, 261
  - phosphorothioate metal rescue (PS-rescue), 242–243, 250
- <sup>31</sup>P NMR and, 231
- prochiral phosphate (pro-R, pro-S), 243–244, 244f, 248t, 250, 257–258, 259t
- role of metal, 241–244
- scissile phosphate, 250, 255, 257–261, 259t, 260f
- S<sub>N</sub>2 mechanism, 258, 260, 260f
- Tetrahymena* pre-rRNA, 238–242
- transition state, 260f
- wobble base, 240f, 246, 249, 256
- X-ray crystallography, 241, 249, 252–256f, 260
- GSH, see glutathione
- Guanine (G), 47, 48t, 50f, 54, 55, 61, 67–89, 73
- metal binding to, 48t
- Guanosine, 48t, 52f, 239
- group I intron and, 240f, 244–261, 246f, 248t, 253f, 254f, 256f, 259t, 260f
- hammerhead ribozyme and, 263, 267t, 270–271t, 277, 282f, 292, 298, 301

- metal binding to, 48t  
*Tetrahymena* pre-rRNA ribozyme and, 239–241
- Guanosine monophosphate, 48t  
 cyclic guanosine monophosphate (cGMP), 194
- Guanosine triphosphate (GTP), 239
- Hamiltonian equation, 123, 127
- Hammerhead ribozyme:  
 base rescue 274f, 276f  
 cadmium (Cd) and, 257t, 268, 280–283, 288, 290  
 constructs, 264f, 265f, 266f  
 exogenous base addition and, 273–277  
 hammerhead fold and, 263, 292, 298  
 hydroxyl radical footprinting and, 289–292, 291t, 300  
 inline fitness and, 286–287, 294, 295t, 297  
 magnesium and, 267t, 268–269, 270–271t, 272, 277–284, 287–301  
 manganese and, 267t, 269, 270–271t, 272, 280, 282, 284, 288–290  
 orbital steering and, 286–287  
<sup>31</sup>P NMR, 281–283, 288–289, 298  
 pH and, 267t, 269, 270–271t, 272, 279, 282, 287–288, 291, 294, 297  
 phosphorothioate metal rescue (PS-rescue), 268, 281–284, 288, 301  
 prochiral phosphate (pro-R, pro-S), 262, 265, 267t, 268, 271t, 272, 280–282, 282f, 284, 288, 297–299  
 role of metal ion, 241–244  
 scissile phosphate, 264f, 265f, 266f, 266, 268–269, 273f, 279–281, 284, 286, 293f, 296, 300  
 S<sub>N</sub>2 mechanism and, 262, 262f, 264, 268, 272, 279, 284, 287, 294, 296–297, 300  
 terbium (Tb) and, 267t, 277–278, 286  
 time-resolved crystallography, 269, 278  
 X-ray crystallography, 262–264, 266, 267t, 268, 270–271t, 272, 274, 277–278, 280–291, 285–286, 288, 292–294, 296, 300
- Hanging drop method, 84
- Hard anions, 6–8, 7t, 40, 242
- Hard-soft acid-base theory (HSAB), 6–8, 7t
- Harmonic oscillators, electron exchange, 25
- Hartree-Fock (HF) theory, 161, 168–169, 171–173, 175, 178, 180
- Helical rise, defined, 51, 53t
- Heme(s), see also porphyrin, 117–118, 129, 166–168, 343–345, 354, 357f, 359, 384f, 387t, 406f, 409f, 441f
- Hemerythrin, 459
- Hemocyanin, (Hc), 4–5, 5t
- Hemoglobin (Hb):  
 Bohr effect, 192  
 carbon monoxide binding, 349–350, 356–359  
 cooperativity, 90–91, 346, 359  
 deoxy-, 90–91, 344–348, 345f, 348f, 349–351, 351t  
 dioxygen binding, 345–349, 357f, 359  
 diphosphoglycerate, DPG, 192–193, 193f  
 distal histidine, 346, 350–351, 350f, 356–358, 357f  
 EPR, 17  
 ferryl-oxo iron intermediate, 354, 364, 371, 376  
 ligands, 344, 344f, 347, 350–352, 351t  
 met-346, 349, 363  
 model compounds, 90–91, 349–359  
 Mössbauer spectroscopy, 135t  
 μ-oxo iron species, 346, 350, 353–355, 354f, 355f  
 oxy-, 90–91, 135t, 344–346, 348f, 350–352, 350f, 351t, 352t, 353–355  
 proximal histidine, 345–351, 347f, 350f  
 quaternary structure, 344–348, 345f, 359  
 R- and T-states, 91, 146, 347–348, 347f, 351, 355–358, 356f  
 resonance Raman, 146, 348, 358  
 subunits, 346–347  
 tetramer, 345f, 346–348  
 time-resolved X-ray crystallography, 358  
 vibrational spectroscopy, 348, 358

- X-ray crystallography, 344–345, 349, 351, 352t, 357–359
- Histidine (His, H), 7–8, 32, 32f, 33f, 56–57, 57f
- aconitase and, 452
- as a hemoglobin ligand, see hemoglobin
- as a myoglobin ligand, see myoglobin
- cytochrome b(6)f and, 386–387, 387t
- cytochrome bc<sub>1</sub> and, 168–169, 397
- cytochrome c and, 117, 410–411, 415, 422
- cytochrome c oxidase and, 131–132, 429–430, 435
- cytochrome c oxidase model compounds and, 439, 440, 441, 442, 453
- hemoglobin model compounds and, 91
- methane monooxygenase and, 459–460
- modes of interaction with metal ligands, 33f
- myoglobin and, 147
- nitrogenase and, 138
- zinc fingers and, 63–72, 64f, 65f, 68f, 72t
- Histone proteins, 53
- Holoenzymes, 44, 54, 55
- Homeostasis
- alkali metals and, 189–190, 192, 193–200, 198f
  - apoptosis and, 427
  - calcium and, 194–195
  - iron and, 466
  - iron regulatory proteins (IRP or IRE) and, 455
  - phosphorus and, 192–193
- Homocitrate, 137f
- Hoogsteen base pairs, 49, 244, 257
- HPLC, 42, 149, 158
- Human genome, 30, 41, 54, 57, 61–64, 66
  - genetic transmission and, 53–56
- Human Genome Project, 61–63
- Hybridization
- DNA-RNA hybrids, 50, 72–73, 266, 281
  - molecular orbitals and, 164–165, 171–173, 346–347, 347f, 373
- Hydrogen bonds:
- aconitase and, 454, 457
  - Ca<sup>2+</sup>-ATPase and, 334–336
  - calmodulin and, 311, 317, 322, 324
  - characteristics of, 34, 422–423
  - cytochrome bc<sub>1</sub> and, 391, 399, 401, 403, 405, 406f, 407
  - cytochrome b(6)f and, 387, 387t
  - cytochrome c and, 422, 424–426
  - cytochrome c oxidase, 410–411, 415, 430–435, 449
  - cytochrome P450 and, 363, 368–382
  - force fields and, 167
  - hemoglobin and, 350, 350f, 356, 358, 359
  - iron-sulfur clusters and, 169–170, 391
  - methane monooxygenase and, 459
  - Na<sup>+</sup>/K<sup>+</sup>-ATPase and, 201–203
  - nucleic acid base pairs and, 47, 49, 50f, 54, 55
  - potassium ion channels and, 211, 223, 228
  - protein secondary structure and, 34–35, 36f, 37–38
  - ribozymes and, 240, 244–249, 250–256, 253f, 264f, 265f, 272, 275–277, 275f, 285–287, 291–294, 297–299
  - zinc finger domains and, 65, 67–68, 72t
- Hydrogen peroxide, 166, 289, 364, 410, 453
- Hydrogenase, 4t, 6t, 22, 45, 66, 172, 388
- Hydrolysis, 39, 45, 198–200, 198f, 290, 328, 337–338, 459
- Hydroxyl ion, 4t
- Hydroxyl radical, see footprinting, hammerhead ribozyme
- Hydroxylase, 5t
- Imidazole, 16, 90–91, 119, 135t, 169, 354, 355, 432
- Imidazole nitrogen, 3t, 5t, 6t, 7–8, 16, 347, 350, 351–354, 351t, 352t, 386, 432
- Immunoglobulins, 45
- Inertness, 10–11
- Infrared Raman spectroscopy, 78, 348

- Infrared spectroscopy, 158, 178, 203, 450–451, 453
- Inner-sphere:  
electron transfer, 23, 23f  
reaction rates, 11–12
- Interchange mechanism, 10, 12
- Introns, 58–59, 63, 238–261, 240f, 246f, 248t, 256f, 259t, 301
- Inverted region, see electron transfer and Marcus theory
- Iodide ions, 7
- Ion channels, see potassium ( $K^+$ ) ion channels
- Ion exchange chromatography, 40
- Ionization chambers, X-ray absorption spectrometry, 80
- Ionophore(s), 190–191t
- Ion pump(s), see active transport, potassium ( $K^+$ ) ion channels, 198f
- Iridium (Ir), 11t, 16, 254
- Iron (Fe):  
abundance, 2–3, 2t, 343  
characteristics of, 5t, 343  
chelating agents for, see ethylenediaminetetraacetate (EDTA)  
cytochromes and, 4, 5t, 117–118, 166–167, 173, 360, 384f  
cytochrome bc<sub>1</sub> and, 389–407, 393f, 406f  
cytochrome b(6)f and, 382–388, 385f, 387t  
cytochrome c and, 408–428, 412f, 421f  
cytochrome c oxidase and, 129, 428–454, 436f  
cytochrome P450 and, 361–372, 364f, 370f  
electron transfer, 5t, 23–24, 26, 58, 117, 129, 143–146, 350, 365, 378–379, 378f, 383, 385–386, 385f, 388–391, 396–397, 399, 402, 405, 406f, 408, 410–411, 413–416, 420, 422, 424–427, 429–430, 437–439, 440f, 444f, 449–450, 453–454  
enzymes and, 82  
EPR studies, 17, 129, 137–139, 362–364, 366–367, 376, 378–379, 387, 433, 447
- essentiality, 2–3, 2t
- ferritin and, 45, 458–459
- [ $Fe_xS_y$ ] clusters, see also iron sulfur clusters 4, 21–22, 21f, 24, 132, 136, 457f
- heme(s) and, 117–118, 129, 166–168, 343–345, 354, 357f, 359, 384f, 387t, 409f, 431f, 433f
- hemerythrin, 459
- hemoglobin/myoglobin and, 4, 5t, 17, 82, 343–359
- high-spin, 17
- hydrogenase and, 22, 172
- low-spin, 17
- MbCO and, 147–148, 358
- methane monooxygenase and, 459–463
- model compounds and, 90, 350–359, 350f, 351t, 352t, 372–382, 439–440, 442f, 446t, 448t
- Mössbauer spectroscopy and, 22, 76, 132–139, 153
- $\mu$ -oxo dimer, 346, 350, 353–355, 354f, 355f
- nitrogenase and, 4, 5t, 22, 137–139, 188
- organometallics, 20f
- oxidation states, 14, 82, 134, 139, 168, 346, 359
- porphyrins and, 17, 344f, 346, 353–356, 355f, 356f, 357f, 359, 375f, 377f, 384f, 409f
- redox potentials, 346
- Rieske iron sulfur protein (ISP), 22, 167–168, 383–384, 385f, 386–387, 387t, 389–391, 394–396, 399, 406f, 408
- storage and transport, 45, 189, 458, 465
- Iron-oxo proteins, 458–465
- ferritin, 45, 458–459
- hemerythrin, 459
- methane monooxygenase, 459–463
- transferrin, 458
- Iron-sulfur clusters, 21–22, 136, 454–455, 458
- aconitase and, 21, 44, 132, 454–458, 455f
- characteristics of, 135t, 136

- cytochrome b(6)f and, 132  
cytochrome bc<sub>1</sub> and, 132, 168–169  
ferredoxin and, 21, 24, 455  
FeS<sub>4</sub> centers, 135t  
Fe<sub>2</sub>S<sub>2</sub> centers, 135t, 168  
  cytochrome b(6)f and, 22, 383–384,  
    385f, 386–387, 387t  
  cytochrome bc<sub>1</sub> and, 22, 389–391,  
    393, 395, 397, 399, 401, 403, 406f,  
    405–407  
Fe<sub>3</sub>S<sub>4</sub> centers, 135t, 137–139  
Fe<sub>4</sub>S<sub>4</sub> centers, 135t  
Mössbauer spectroscopy, 137–139,  
  454–455  
nitrogenase and, 137–139  
  rubredoxin and, 22, 24, 135t, 455  
Iron-sulfur proteins (ISP), see also  
  Rieske iron sulfur protein  
aconitase, 21, 44, 132, 454–458  
cytochrome b(6)f and, 22, 132, 383–  
  384, 385f, 386–387, 387t  
  cytochrome bc<sub>1</sub>, 168–169, 389–391,  
    393, 395, 397, 399, 401, 403, 406f,  
    405–407  
ferredoxin and, 21, 24, 455  
nitrogenase, and, 137–139  
  rubredoxin and, 22, 24, 135t, 455  
Isoelectric focusing (IEF), 40–41  
Isoleucine (Ile, I), 31f, 34, 34f, 324, 346,  
  350–351, 416, 427  
Isomer shift, see Mössbauer  
  spectroscopy  
Isozymes, 44  
Jahn-Teller effects, 10–11, 18, 126, 163,  
  178  
*J* values, see nuclear magnetic  
  resonance  
KAR-2, 311–313, 311f  
Kinetics, 8–13, 143–146  
  associative, 10  
  dissociative, 12  
  first-order, 9–11  
  flash photolysis, 144–146  
  interchange mechanism, 12  
  Marcus theory, 25–26  
  pseudo-first-order conditions, 9–10  
  rate laws and rate constants, 11–13  
  second order, 11–12  
  stopped flow, 143–144  
Kinetics, enzyme, 43–47  
  Michaelis-Menten, 13, 45–47, 326  
K<sub>M</sub>, 45–46, 46f  
K<sub>ML</sub>, 8–9, 9t  
Lability, 10, 23f  
Lanthanide ions, 14  
  EPR and, 125  
  lanthanide-based energy transfer  
    (LRET), 223–224  
  NMR probes, 103  
Lead (Pb), 2, 2t, 7, 7t  
Least-squares method, 90, 165  
Leucine (Leu, L), 31f, 57f, 58, 64, 324,  
  346, 350–351  
Lewis bases, 19  
Library of Congress, as information  
  resource, 183  
Ligand exchange rates, 10–11, 11t  
  water exchange rates, 11t  
Ligand field stabilization energies  
  (LFSE), 11, 15, 18  
Ligand field theory, 11, 18  
Ligand-metal charge transfer (LMCT),  
  81, 450, 459  
Ligand preference, 3, 6, 4t, 5t, 6t, 8,  
  9–10, 9t, 47, 55, 347f, 356–357,  
  359  
Lineweaver-Burk plot, 46, 46f  
Lithium (Li), 2, 171, 190t, 193, 199, 211,  
  283–284, 287, 290–291  
L-stereoisomers, 32  
Lumen, 96, 301  
Ca<sup>2+</sup>-ATPase and, 327–329, 328f,  
  330f, 333–337  
  cytochrome b(6)f and, 385–386, 385f  
  cytochrome c and, 410–411, 415  
Lyases, 45, 116  
Magnesium (Mg), 1, 2t, 7t, 9–10, 9t, 11t,  
  47, 138, 141, 189, 197, 197t  
ATP complexes, 44, 47  
catalytic RNA and, see ribozymes  
cytochrome c oxidase and, 436f  
homeostasis and, 191t, 192–193, 195,  
  197t  
structure and triggers, 4t

- Magnetic circular dichroism (MCD) spectroscopy, 40, 77, 130, 374, 452
- Magnetic measurements, see also electron paramagnetic resonance (EPR), 17
- Magnetic receptivity, 99
- Manganese (Mn), 2, 2t, 4t, 7t, 11t, 18, 361  
electron transfer and, 23  
group I introns and, 243, 250, 261, 272  
hammerhead ribozyme and, 267t, 269, 270–271t, 280–290  
structure and triggers, 4t
- Marcus theory, 25–26, 415
- Mass spectrometry (MS), 76, 142, 148–153  
cytochrome c oxidase models and, 440  
electrospray ionization-MS (ES-MS), 382  
instrumentation, 149  
MALDI-TOF MS, 43, 149–150, 152, 441  
peptide and protein sequencing and, 150–153  
protein folding and, 149  
proteomics and, 40, 42–43, 150–153
- Matrix metalloproteinase, 119–120, 121f
- Maxam-Gilbert sequencing method, 59, 62
- M center (nitrogenase), 137–138, 137f
- Mechanisms, see outer sphere, inner sphere, interchange, associative, dissociative, and enzyme of interest
- Medicinal chemistry, 14, 73, 190t  
polymerase chain reaction (PCR) and, 61
- MEDLINE, 184–185
- Messenger RNA (mRNA), 54–56, 58–59, 239, 458
- Metal chelators, 8, 32, 47, 81f
- Metal ion complexation:  
kinetics, 9–13  
thermodynamics, 8–9
- Metal-ligand interactions, 6, 25
- Metal-metal bonds, 21–22
- Metallocenes, 19–20
- Metalloenzyme systems, see also individual proteins, 22, 27, 30, 44, 73
- Metalloproteins: see also individual proteins  
characteristics of, 1, 14, 27, 30, 60, 73  
electron exchange, 24  
isolation, 353  
model compounds and, 352–353  
modeling and, 166  
purification, 84, 353
- Metallothionein, 174
- Metals, in biological systems, 3–6
- Metal toxicity, 2
- Methane monooxygenase, 22, 459–463  
electron transfer and, 461–463  
EPR studies, 460–464  
EXAFS studies, 461–464
- Particulate methane monooxygenase (pMMO), 460–465  
soluble methane monooxygenase (sMMO), 460–463
- X-ray crystallography, 461–465, 463f
- Methionine (Met, M), 22, 31f, 33f, 40, 56  
calmodulin and, 304, 317, 322, 324  
cytochrome c iron ligand and, 117, 408, 410–411  
cytochrome c oxidase copper ligand and, 131, 429, 430, 463
- Michaelis constant, 45
- Michaelis-Menten enzyme kinetics, 13, 45–47, 326  
competitive inhibition, 46–47, 46f  
noncompetitive inhibition, 46–47, 46f  
rate equation, 45  
stopped-flow kinetics, 143–144, 354, 419, 421, 437
- MM2/MM3 force field, 161, 164, 166, 178
- MNDO, 161, 174–176, 179
- MNDO/d, 174, 179
- Model compounds, see individual proteins
- Molecular modeling (MM), 158–159, 160–170, 178–180, 182  
biomolecule modeling, 166–167  
characteristics of, 160–164  
cytochrome bc<sub>1</sub> and, 167–170

- descriptive examples, 167–170  
hammerhead ribozyme and, 284–285  
 $\text{Na}^+/\text{K}^+$  ATPase and, 202–203  
potassium ion channels and, 225
- Molybdenum (Mo):  
characteristics of, 6t  
in nitrogenase, 4, 6t, 14, 19–20
- Monooxygenases, see also cytochrome P450, methane monooxygenase, 22, 172, 372–382, 455–461
- Monte Carlo method, 165–166, 166f
- MOPAC, 176, 179, 182
- Mössbauer spectroscopy:  
applications, generally, 132  
characteristics of, 132–134, 133f  
descriptive examples, 137–139  
iron-sulfur clusters, 137–139  
isomer shift ( $\delta$ ), 134–135, 135t  
magnetic hyperfine interactions, 136–137  
nitrogenase, 137–139  
quadrupole splitting ( $\Delta E_0$ ), 134–135, 135t
- Multiple-wavelength anomalous dispersion (MAD), 92–93, 95, 241, 247, 464–465
- $\mu$ -oxo iron species, 346, 350, 353–355, 354f, 355f
- Mutagenesis, site-directed, 56–58, 83, 205, 220, 222, 224–225, 249, 344–345, 349, 353, 425
- Mutations, genetic, 56–58
- Myoglobin, 3, 5t, 8, 17, 30, 43, 82, 83, 166–167  
active site, 350–351  
carbon monoxide binding, 146–148, 167, 356–358  
basics, 344–347  
cytochrome P450 model and, 363, 372, 382  
deoxy-, 147–148, 344  
dioxygen binding, 146–147, 344, 346f, 349  
EPR and, 127  
ferryl-oxo iron intermediate, 354, 364, 371, 376  
flash photolysis, 146–147  
ligands and, 344, 344f, 347, 350–352, 351t
- met-, 346, 349, 353  
model compounds, 354–355  
Mössbauer spectroscopy and 135t  
 $\mu$ -oxo iron species, 353–355, 354f, 355f  
oxy-, 346, 349, 350–352, 351t, 352t, 353–355
- time-resolved crystallography and, 146–148
- X-ray crystallography, 344, 346, 349, 356–359
- NAD, see nicotinamide adenine dinucleotide
- NADP, see nicotinamide adenine dinucleotide phosphate
- $\text{Na}^+/\text{K}^+$ -ATPase, 44, 45, 190t, 191t, 199–203  
active transport and, 197–199  
facilitated diffusion and, 197  
X-ray crystallography and, 200–203
- National Cancer Institute (NCI), 184–185
- National Human Genome Research Institute, 61–62
- National Institutes of Health (NIH), 62, 97, 182
- National Library of Medicine, 184–185
- nephelauxetic effect, 103
- Neural network, 39, 165
- Newton-Raphson method, 165
- Nickel (Ni):  
characteristics of, 2t, 16  
effect of Ni(II) on dsDNA, 141  
ligand geometry, 6t, 16, 81
- Nicotinamide adenine dinucleotide (NAD), 44, 44f, 388
- Nicotinamide adenine dinucleotide phosphate (NADP), 24, 24f, 212
- Nitrogen ( $\text{N}_2$ ), 2t, 5t
- Nitrogenase, 4, 5t, 6t, 22, 33, 137–139, 166, 450  
iron-sulfur clusters, 137–139  
ENDOR studies, 137–139  
M center (FeMoco), 137f  
Mössbauer spectroscopy, 137–139  
X-ray crystallography, 139
- Nonallosteric enzymes, 44–46, 46f
- Noncompetitive inhibitors, 46–47

- N7 sites, see also adenine, guanine, 47, 141, 248t, 268
- N-terminal, 34–35, 353f, 37f, 38, 40, 42, 57, 68–69, 151, 200–202, 333, 346, 386, 410
- calmodulin and, 303, 303f, 306–326, 309t, 312f, 314–315t
- methane monooxygenase and, 459–460
- potassium ion channels and, 208f, 213–214, 216, 218, 220–233
- Nuclear magnetic relaxation, 100, 107–108, 109–110
- Nuclear magnetic resonance (NMR) spectroscopy:
- applications, generally, 98–101
  - basic information, 99–101
  - Boltzmann distribution, 100, 109
  - calmodulin and, 303–308, 303f, 305f, 307f, 309t, 311–312, 312f, 313–326, 318f
  - characteristics of, 98–101
  - chemical shift, 99, 101–104, 105–106, 105f, 109–117, 172, 177, 281–282, 377–379
  - contact shifts, 103
  - COSY, 112–116, 113f, 118
  - cytochrome c structure and, 408, 411, 415–417, 416f, 421
  - descriptive examples, 117–122
  - Fourier transform (FT), 80–81, 92, 94, 110–112, 111f, 149
  - free induction decay (FID), 110–118, 112f, 113f, 115f
  - instrumentation 101–102, 110–111
  - J* values, see also spin-spin coupling, 104–106, 117, 106f
  - lanthanides and, 103, 125
  - magnetic properties of nuclei, 98–100
  - magnetogyric ratio, 98–100
  - model compounds, 282, 352–353, 376–378, 439–410
  - multi-dimensional, 116–117
  - NOESY, 112, 115–116, 118, 316, 320
  - nuclear magnetic relaxation, 100, 107–108, 109–110
  - nuclear overhauser effect (NOE), 107, 108–110, 112, 122, 202, 316, 320
  - nuclear screening, 101–104, 105f
  - nuclear spin, 77–78, 99, 99f, 104, 106–108
  - receptivity, 98, 109
  - relaxation times  $T_1$  and  $T_2$ , 107–109, 110, 116
  - spectral integration, 106–107, 111
  - spectrum, obtaining, 110–111
  - spin-spin coupling, 104–106, 106f, 107, 108, 112–118, 122
  - spin-spin decoupling, 106–107, 113
  - TOCSY, 112–116, 118
  - two-dimensional (2D), 111–116
- Nuclear overhauser effect (NOE), 107, 108–110, 112, 122, 202, 316, 320
- Nuclear overhauser effect spectroscopy (NOESY), 112, 115–116, 118, 316, 320
- Nucleic acids: see also DNA, RNA
- DNA: see also DNA
- building blocks, 47, 48–49t
  - molecular structures, 48–49t
  - genes and cloning, 58–61
  - genetic information transmission, 53–56
  - genetic mutations, 56–58
  - genomics, 61–63
  - human genome, 30, 39, 41, 50, 54, 57, 61–63
- RNA: see also ribozymes
- building blocks, 47, 48–49t
  - molecular structures, 48–49t
  - site-directed mutagenesis, 56–58, 249
- Nucleophiles:
- ribozymes and, 241–244, 251, 254, 257–260, 259t, 260f, 262f, 279, 286, 296–297, 300–301
- Nucleophilic attack, 245, 248f, 255, 257–258, 262, 269, 272, 276, 279, 286–287, 335, 366f, 374, 382
- Nucleosides, defined, 47, 48–49t
- group I intron ribozyme and, 244
- Nucleotide(s):
- defined, 47, 48–49t
  - $\text{Na}^+/\text{K}^+$ -ATPase and, 200, 202
  - polymerization, 30
  - properties in DNA, 53t

- ribozymes and, 241–242, 247, 247–249, 248t, 253–263, 259t, 264f, 265f, 268, 273, 274f, 281, 285–286, 289–294, 291t, 296–299, 333  
sequencing and, 58–59, 149  
transcription and, 55
- <sup>18</sup>O labeling, proteomic studies, 42–43
- Oligonucleotides:  
catalytic RNA (ribozymes) and, 240f, 245  
mass spectrometry, sequencing and, 148–150  
zinc finger protein and, 66
- Organometallic chemistry, see Bioorganometallic chemistry
- Orbitals:  
highest occupied molecular (HOMO)  
lowest unoccupied molecular (LUMO)  
Slater-type (STO)
- Outer-sphere (OS):  
electron transfer, 10, 23–25  
reaction rates, 11–12
- Oxidation, metals in biological systems, 2–5
- Oxidation-reduction, 14, 23  
electron transfer and, 22–26
- Oxidation state:  
copper ions in cytochromes, 443  
copper ions in methane  
monooxygenase, 463–465  
iron ions in hemoglobin, 349  
iron ions in cytochromes, 359, 360, 363, 366–367, 369, 373, 412–414, 426, 428, 445  
iron ions in iron-sulfur clusters, 4, 24  
metals in biological systems, 2–5, 4t, 5t, 6t, 14, 19  
nuclear magnetic resonance and, 103  
Mössbauer isomer shift ( $\delta$ ) and, 134, 135t  
nitrogenase Mössbauer isomer shift ( $\delta$ ) and, 138–139
- X-ray absorption energy and, 79, 81–82
- Oxidoreductases, 45, 167, 212, 214, 229, 361, 388, 411
- Oxygen, O<sub>2</sub>:  
insertion enzymes, see also cytochrome P450, methane  
monooxygenase, 22, 82, 172  
molecular orbital diagram, 349f  
percent composition of human body, 2t  
reduction enzymes, see also specific cytochromes, 24, 82, 143  
transport proteins, see also hemoglobin, myoglobin, hemocyanin, 4, 5t, 30, 43, 83, 90–91, 135t, 146–147, 192–193, 343–349  
solubility, 344
- Oxygenation: see also methane  
monooxygenase, cytochrome P450  
cytochrome P450, 365, 365f, 378  
methane monooxygenase, 22, 459–460
- Oxy-heme stereochemistry, 351–352, 351t, 352t
- Oxyhemoglobins, see hemoglobin, oxy-
- Oxymyoglobin, see myoglobin, oxy-
- Palladium (Pd), 7t, 16
- Paramagnetism, 17, 102–104, 108, 122, 125–129, 136–138, 153, 243, 288–289, 348, 362–263, 376, 460
- Parkinson's disease, 160
- PASCAL, 185
- Passive diffusion, 193, 195–197
- Peptide bonds, formation of, 33–34, 34f
- Peroxidase, 82, 366, 372–373, 377, 380  
cytochrome c peroxidase, 166, 359, 363, 365–366, 371–372, 377, 410, 421, 424–426, 426f, 429
- Peroxide ion, see anions
- Peroxo ligand(s), see also cytochrome c oxidase, cytochrome c peroxidase, cytochrome P450, 82, 166, 363, 365f, 366f, 368, 368f, 372, 374, 374f, 382, 433, 439, 441f, 443–445, 451
- pH, 190t  
cytochromes and, 360, 385, 402, 408, 418, 420, 424, 437–439

- electron transfer kinetics and, 144  
 hammerhead crystal structures and, 267t, 269–272, 270–271t, 279, 282, 287–288, 291, 294, 297  
 homeostasis and, 197, 199  
 isoelectric for amino acids, 40–41  
 oxygen carriers  
   passive diffusion and, 195  
   physiological, 31f, 32, 32f  
   protein crystallization and, 84  
   proton channels and, 194  
 Phage display, 69  
 Phenylalanine (Phe, F), 31f  
   calmodulin and, 304, 314–316, 315t, 324  
   cytochrome c and, 415, 424  
   hemoglobin and myoglobin and, 346, 350–351  
   Na<sup>+</sup>/K<sup>+</sup> ATPase and, 201  
   potassium ion channels, 209  
   zinc fingers and, 64  
 Phosphates, 4t, 7, 7t  
   nucleotides and, 24, 47–50, 48t, 50f, 55, 65, 69, 201–202, 212  
   group I intron and, 244–261, 254f  
   hammerhead ribozyme and, 261–301  
   homeostasis and, 192–194  
   prochiral, 243, 244f  
   ribozymes and, 240–244  
   scissile phosphate, 250, 255, 257–261, 264f, 265f, 266f, 268–269, 279–281, 284, 286, 293f, 296, 300  
 Phosphodiester groups, 47, 55, 68, 239, 242, 244, 263, 265, 269, 280, 284, 289, 292–293  
 Phosphorus (P), 1–2, 7t, 107, 192–194, 243  
   group I intron and, 257–262, 360f  
   hammerhead ribozyme and, 269, 272, 279, 287, 296  
   prochiral, 243, 244f  
   P type-ATPases and, 335  
 Phosphorylation, 56, 193, 200, 244, 260, 319, 321, 323, 327, 329, 333–337, 388, 430  
 Phosphatases, 191t, 302, 327, 459  
 Plasmids, 59–61, 60f  
 Platinum (Pt), 14, 16, 19, 96  
   spin-spin coupling and, 106–107, 106f  
 Platinum-containing anticancer agents:  
   *cis*-dichlorodiammineplatinum(II),  
   *cis*DDP, cisplatin, 16, 16f, 20, 20f, 47, 53, 55, 141  
   drug resistance, 55  
 Po1ak-Ribiere minimizer, 165  
 Polyacrylamide gel electrophoresis:  
   Two-dimensional (2-DE or 2-D gels), 41–42, 269, 272  
 Polyethyleneglycol (PEG), 84  
 Polymerase chain reaction (PCR), 60–61, 141  
 Polymerization, 30  
 Polypeptide chains, formation of, 30, 33–35, 34f  
 Porphyrin: see also heme(s)  
   capped, 354, 356  
   π-cation radical, 364–366, 370, 372, 375–376, 379, 382  
   picket-fence, 90–91, 135t, 350–352, 351t, 352t, 354–356, 355f, 356f, 441  
 Potassium (K), 1, 2t, 4t, 6, 7t, 9, 9t, see also alkali metals, Na<sup>+</sup>/K<sup>+</sup>-ATPase, potassium ion channels  
 Potassium (K<sup>+</sup>) ion channels, 199, 203–235  
   basic information, 203–205  
   biotinylation and, 220–221, 220f, 224, 233–234  
   inactivation peptides and, 204, 212–214, 229–231  
   ion conduction (pore), 209, 212–215, 217f, 222–227, 230–234  
   KvAP and, 204, 206–207t, 209f, 214–215, 218, 223–234  
   monoclonal antibodies (Fabs) and, 204, 206–207t, 215, 225, 229  
   pore(s) and, 194, 204–205, 206–207t, 208–209, 208f, 210f, 211–216, 217f, 218–235  
   Shaker channels and, 204–205, 208–209, 208f, 223–232  
   tetraethylammonium (TEA) ion and, 209, 212  
   voltage-dependent (voltage-gated), 197, 199, 204–205, 206–207t, 208–209, 212–218, 213f, 219f, 224–225, 229

- voltage-independent and, 233  
voltage sensors (voltage sensor paddles) and, 204, 215–218, 219f, 223–232  
X-ray crystallography and, 200, 201, 204, 206–207t, 208–209, 208f, 212, 213f, 215–216, 217f, 219f, 223–235  
Praseodymium (Pr), 104  
Prokaryotic, 58–59  
  cytochromes and, 82, 405, 408–409, 426, 455  
  potassium ion channels and, 208–209, 228–231  
Prosthetic group or cofactor, 24, 44  
Protein Data Bank (PDB), 83, 94, 176, 179–182  
Proteins:  
  amino acid building blocks, 30–33  
  enzyme kinetics, 45–47  
  enzymes, 43–47  
  folding, 38–39, 63, 65–66, 149  
  helical structures, 34–36, 36f, 36t, 37f, 38  
  pleated sheets,  $\beta$ -pleated sheets, 34–36, 36f, 37f, 38  
  profiling, 41–42  
  proteomics, 30, 40–43, 150–153  
  quaternary structure, 38  
  secondary structure, 34–38, 36f, 37f  
  sequencing, 39–43, 150–153  
  tertiary structure, 37–39, 37f  
Proteomics, 30, 39–43, 150–153  
Proton translocation, see cytochrome  $bc_1$ , cytochrome b(6)f, and cytochrome c oxidase  
P type ATPases, 327–338, see also  $\text{Ca}^{2+}$ -ATPase  
PubMed, 184  
Purines, 47, 48–49t  
Pyrimidines, 47, 48–49t  
  
Quadrupole splitting, see Mössbauer spectroscopy  
Quantitative structure activity relationship (QSAR), 175  
Quantum mechanics:  
  ab initio calculations, 119, 161, 170–174, 180  
density function theory, 83, 161, 170–180  
semiempirical methods, 161, 170, 173–175, 177, 179–180  
Quench-flow, 143  
Quinones, see also ubiquinone, 24  
  
Radium (Ra), 191t  
Ramachandran map, 166  
Raman spectroscopy, 77–78, 348, 358, 376, 379, 420, 439, 442–443, 450, 454  
  pulse-probe, 176  
Rapid quench, 143  
Recombinant DNA, 57–61, 60f  
Replication, 53–55, 54f  
Research Collaboratory for Structural Bioinformatics' Protein Data Bank (RCSB-PDB), see protein data bank  
Resonance Raman (RR) spectroscopy, see Raman spectroscopy  
Restriction endonucleases, 58–60, 60f  
Retroviruses, genetic transmission, 54  
R-factor, 93–94, 97, 345, 390  
R groups:  
  in amino acids, 30, 31f, 32f  
Ribonucleic acid (RNA): see also RNA, ribozyme  
  as building blocks, 47–48, 48t  
  molecular structures, 48t  
  synthesis, 55  
Ribonucleotide reductase, 459  
Ribosomal RNA (rRNA), 54  
Ribozyme(s)  
  group I intron, 244–261, see also group I intron  
  hammerhead, 261–301, see also hammerhead ribozyme  
  *Tetrahymena* pre-rRNA, 238–242, see also group I intron  
Rieske iron sulfur protein (ISP), 22, 167–168, 383–387, 387t, 391, 394–396  
RNA, see also ribozyme  
  exon, 239–240, 240f, 244–248, 246f, 251–260, 259t, 260f

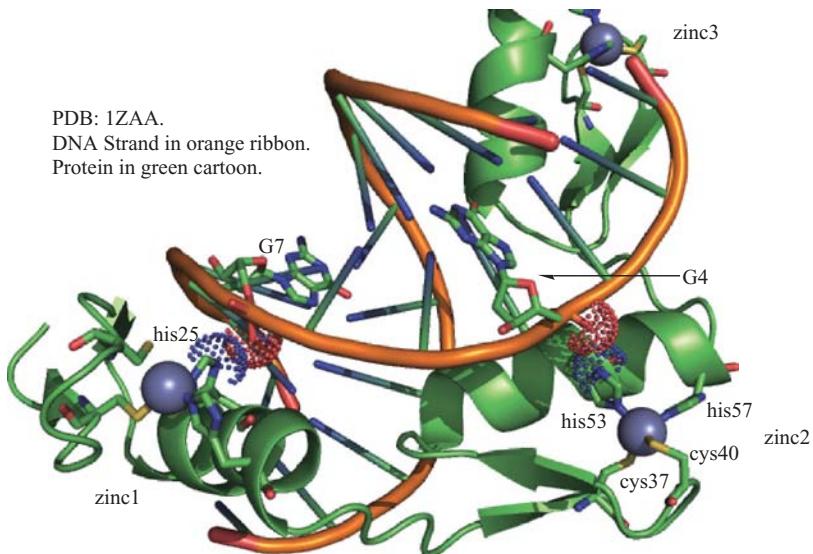
- intron, 238–261, 240f, 246f, 256f, 259t, 301  
repair, 274
- RNA polymerase, 55
- RNAi, 54
- rRNA, 54, 238–242, 245
- Rubidium (Rb), 190t, 193, 211
- Rubredoxin (Rd), 22, 24, 135t, 455
- Sanger sequencing method, 58, 62
- Sarcoplasmic reticulum (SR), 95–96, 195, 201, 301, 327–328, 338
- Scanning electron microscope (SEM), 142
- Scanning tunneling microscope (STM), 142
- Scattering:  
X-ray absorption spectrometry, 79–80, 83  
X-ray crystallography, 87, 89, 89f, 92–93, 465
- Scientific World, 185
- Second-order rate law, 9–12
- Self-assembly  
porphyrins and, 354  
cytochrome c folding and, 422
- Self-consistent reaction field (SCRF), 161, 178
- Self-exchange, electron transfer process, 25, 57–58  
cytochrome c and, 415
- Semiempirical methods, 161, 170, 173–175, 177, 179–180
- Single-nucleotide polymorphisms (SNPs), 63
- Single-stranded DNA (ssDNA), 59
- Site-directed mutagenesis, 56–58
- Slater-type orbitals (STOs), 170–171, 173
- Sodium (Na), see also see  $\text{Na}^+/\text{K}^+$ -ATPase, 1, 2t, 4t, 7t, 9, 9t, 141, 150, 197  
active transport and, 197–199, 198f  
homeostasis and, 190–192, 190t, 192–193, 197t  
passive diffusion and, 197
- Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), 41
- Sodium-potassium ATPase, see  $\text{Na}^+/\text{K}^+$ -ATPase
- Sodium pump, see see  $\text{Na}^+/\text{K}^+$ -ATPase
- Soft cations, 6–7, 7t, 8, 242, 252, 268
- Southern blot technique, 59
- Spectrochemical series, 16
- Spectroscopy, characteristics of, see also specific types of spectroscopy, 77–78
- Spin-echo method, 108, 129, 243, 289
- Spin-orbit coupling, 125–127, 125f
- Spin-spin coupling, see nuclear magnetic resonance (NMR)
- Spin-spin decoupling, see nuclear magnetic resonance (NMR)
- Square-planar complexes, 6t, 10, 14–17, 14f, 15f, 18t, 19, 20f
- Stability constants, 8–9
- Stereoviewing, 58, 182
- Stigmatellin, 167–170, 385f, 391, 393, 397, 398f, 400, 402, 403
- STO-3G, 161, 171
- Stochastics, 165
- Stop codon, 56
- Stopped-flow kinetics, 143–144, 354, 423, 426, 442
- Strain energy, 162–164, 178
- Strontium (Sr), 191t, 193, 195, 290
- Structure-activity relationships (SAR), 20, 175  
cytochrome bc<sub>1</sub> inhibitors and, 397
- Sulfur ligands, 6–8, 7t
- Superoxide dismutase (SOD), 3–4, 32, 143
- Synchrotron radiation, 79, 93–95, 97, 146, 209, 345, 464–465
- Tapping Mode™ scanning, 140
- Technetium (Tc), 14
- Tetrahymena* pre-rRNA, see Group I intron
- Thapsigargin, 329, 331, 335, 336
- Thermodynamics, 8–9  
cytochrome c and, 408, 410  
facilitated diffusion and, 197–199  
heme-O<sub>2</sub> complexes and, 346, 349, 359  
molecular modeling and, 164  
 $\mu$ -oxo dimer and, 353–354  
thermochemistry and, 171, 173, 180
- Thiol(s), 20, 141, 216

- Thiolates, 4t, 5t, 6t, 82, 173  
cytochrome c oxidase and, 445,  
453–454  
cytochrome P450 and, 368, 370,  
373–374, 380–382  
Thymine (T), 47, 49t, 50f, 54, 61, 141  
Tin (Sn), 2, 7t  
Topoisomerase, 54  
Total correlation spectroscopy  
(TOCSY), 112–116, 118  
Toxicity, elements, 2, 3f, 189, 190–191t,  
361  
Trace metals, essential, 1–2  
Transcription, 53–57, 54f, 59, 141, 238  
zinc finger proteins and, 64–72  
Transferases, 22, 24, 45, 455  
Transfer RNA (tRNA), 54–56, 239, 241,  
263  
Transition metals, 3, 4, 11, 13–18, 19, 23,  
138, 141  
biological functions, 3–6, 4t, 5t, 6t  
coordination geometry, 14–16, 14f,  
15f, 18t  
electron paramagnetic resonance  
(EPR), 125, 129  
high spin, low spin *d* electron  
configuration, 17–18, 17f, 18f  
Jahn-Teller effects, 10–11, 18, 126,  
163, 178  
ligand field stabilization energy,  
LFSE, 11, 15, 18  
molecular modeling, molecular  
mechanics and, 161–162, 164, 174,  
176, 178–179  
number of *d* electrons, 14  
spectrochemical series, 16  
Translation, 53–57, 54f, 77, 141, 148, 153,  
390  
Transmission electron microscopes,  
142  
Transverse relaxation time, T<sub>2</sub>, see  
nuclear magnetic resonance,  
relaxation time  
Trifluoperazine, 308–313, 310f, 312f  
Troponin, 3, 44, 194, 301–302, 306, 338  
Tryptophan (trp, W), 30, 31f, 40, 209,  
211, 222, 358, 424, 427  
calmodulin and, 314, 314–215t, 316–  
321, 324–325  
cytochrome bc<sub>1</sub> and, 397, 398f, 399,  
400t, 402–405  
cytochrome c and, 408, 410–417,  
420–422  
potassium ion channels and, 222–223  
Trypsin, 40, 42, 151, 195, 301  
T-state, *see* Hemoglobin  
Two-dimensional correlation  
spectroscopy (COSY), 112–116,  
113f, 118  
Tyrosine (tyr, Y), 32f, 33f, 38, 40, 56,  
203, 358, 427, 459  
cytochrome b(6)f and, 386, 387t  
cytochrome bc<sub>1</sub> and, 397, 399, 400t,  
402–405  
cytochrome c and, 408–409, 409f, 410,  
412f  
cytochrome c oxidase and, 425, 429–  
431, 434f, 434f, 435, 436f,  
437–440, 440f, 442f, 445  
cytochrome P450 and, 364f, 366, 370,  
370f  
potassium ion channels and, 209, 211,  
217f  
zinc fingers and, 64, 64f, 67  
Ubiquinone, 170, 386, 388–392, 388f,  
395, 397–399, 398f, 400t, 401t,  
403–405  
U.S. Department of Energy, Human  
Genome Program, 61  
Uptake factor, 8–9  
Uracil (U), 47, 49t, 246, 275  
UV-visible spectroscopy, 79, 142, 144,  
176  
cytochrome bc<sub>1</sub> and, 395–397, 398f,  
399, 400t, 402–405  
cytochrome c and, 411, 417, 421  
cytochrome c oxidase and, 434, 437,  
445  
cytochrome P450 and, 373, 376  
Valence electrons, in metal clusters, 21  
Valine (Val, V), 30, 31f, 38, 40, 211, 230,  
319, 346, 350, 351, 374, 420  
Ca<sup>2+</sup>-ATPase and, 332, 332f, 332t, 335  
calmodulin and, 304, 308, 310, 316,  
319, 320, 325  
Na<sup>+</sup>/K<sup>+</sup>-ATPase and, 201

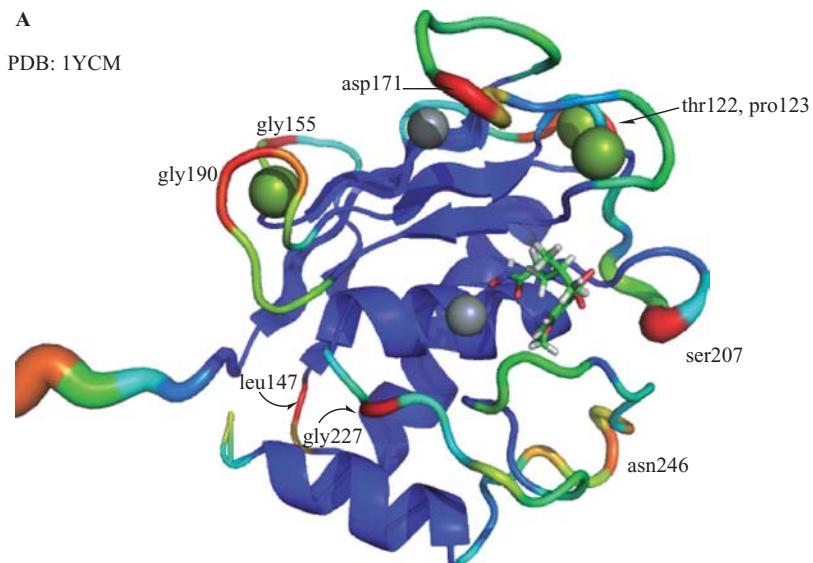
- potassium ion channels and, 211, 217f, 221, 223, 224, 230  
 van der Waals forces, 37–38, 163, 167, 211, 299, 316–317, 320, 371, 425  
 Vanadium (V), 2, 11t, 14, 19  
 Vector quantity, 85, 88–89, 98–101, 99f, 101f, 113, 122  
 Vibrational spectroscopy, 163, 172, 177, 179  
 cytochrome c oxidase, 433f, 435, 436f, 437  
 Vinblastine, 311–313, 311f  
 Visualization software:  
     Chime, 35, 181–182  
     ISIS/Draw, 181–182  
     projected images, 183  
     Protein Explorer, 182  
     RasMol, 182  
 Vitamin B<sub>12</sub>, 6t, 22, 44  
 $V_{\max}$ , 45–47, 46f  
 Voltage-gated channels, see also  
     potassium ion channels, 199, 205, 206–207t, 209, 215, 235
- Water:  
     aconitase and, 454–456  
     Ca<sup>2+</sup>-ATPase crystal structure and, 97, 331, 332f, 335–337  
     cytochrome bc<sub>1</sub> crystal structure and, 169–170, 396–397, 399, 402–405, 408–410  
     cytochrome b(6)f and, 387  
     cytochrome c structure and, 408–412, 409f, 413–417  
     cytochrome c oxidase and, 425, 429–432, 433f, 442f, 445  
     cytochrome P450 and, 361, 262, 363–365, 365f, 368, 371–373  
     exchange, aqua ions, 10–12, 11t  
     group I intron ribozyme structure and, 248, 251–252  
     hemoglobin/myoglobin structure and, 346, 350, 351  
     methane monooxygenase and, 460  
     potassium ion channel structure and, 209–212, 222, 225–228, 234  
     stopped-flow kinetics and, 143–144  
 Watson-Crick base pairing, 47, 49, 50f  
     ribozymes and, 240f, 244, 254, 263, 285–286, 293, 297  
 Wild-type genes, 43, 220, 224  
     ribozymes and, 251, 273, 275, 277, 283  
 World Wide Web, as information resource, 181–185  
     computer software downloads, 160, 175, 176, 181–182, 185  
     literature, 183–185  
     materials, 183–184  
     online societies, 183–184
- X-ray absorption near-edge structure (XANES), 79  
 X-ray absorption spectroscopy, (XAS, EXAFS), 76, 78–83, 446t, 448t  
     applications, generally, 78–79  
     characteristics of, 78–81  
     descriptive examples, 81–83  
 X-ray crystallography, 83–95  
     aconitase structure and, 451–456, 455f, 457f  
     applications, generally, 83–84  
     Ca<sup>2+</sup>-ATPase structure and, 96–97, 328f, 329–337  
     calmodulin structure and, 97, 303–308, 303f, 305f, 307f, 309t, 311–312, 312f, 313–326, 318f  
     characteristics of, 88–95  
     crystal habits, 84–88  
     cytochrome bc<sub>1</sub> structure and, 390–393, 391t, 395, 406f  
     cytochrome b(6)f structure and, 383, 385f, 386–387  
     cytochrome c structure and, 402, 408–412, 409, 413–417, 420–421, 421f  
     cytochrome c oxidase structure and, 427–429, 432, 432f, 434–435, 437–439, 440f, 441f, 445, 449  
     cytochrome P450 structure and, 363, 364f, 369–372, 370f  
     descriptive examples, 95–98  
     electron density, 91–94, 96–97  
     Ewald sphere, 85, 85f, 88, 89f  
     hemoglobin structure and, 83, 344–345, 351t, 352t, 353

- methane monooxygenase structure and, 459–460
- myoglobin structure and, 83, 146–148, 344–345, 351t, 352t, 353, 358, 372–374
- $\text{Na}^+/\text{K}^+$  ATPase, 200–203
- ORTEP diagrams, 90–91, 90f
- phase angle, 83, 92–93, 96
- potassium ion channel and, 205–235
- protein crystal disorder, 91–92, 94
- R-factor, 93–94, 97
- seven crystal system, 87, 87t
- static disorder, 91–92
- time-resolved, 146–148, 269, 278, 358
- unit cells, 85–88, 86f, 90, 91–92, 96, 146, 373t, 389
- zinc finger proteins, 67–73
- YAG laser, 419
- Z DNA, 50, 53, 53t
- Zeeman effect/interaction, 122, 125
- Zero-field splitting (zfs), 125, 127
- zero-order reaction, 9, 45, 46f
- ZIF268, 65–70, 68f, 70f
- Zinc (Zn), 2, 4t, 7t, 9, 9t, 11t
  - superoxide dismutase and, 3, 33
  - structure and triggers, 4t
- Zinc-finger proteins, 63–73, 65f, 71f, 72t
  - characteristics of, 63–66
  - descriptive examples, 67–73
  - structure, 64f, 65f, 65–66, 68f, 71f
- ZINDO, 161, 176
- Zwitterions, amino acids, 31t, 32t, 32, 34, 40





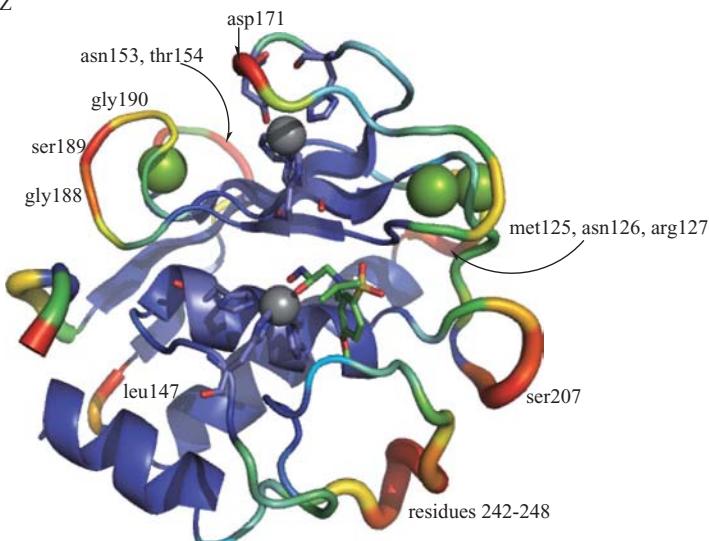
**Figure 2.22** Zinc-finger–DNA interactions for PDB: 1ZAA as described in reference 31 and the text. Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft Corporation.)



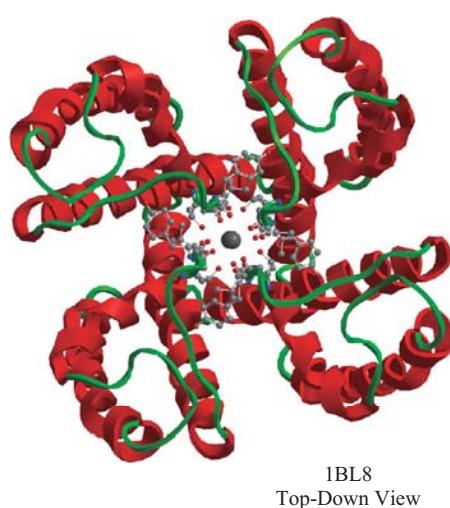
**Figure 3.18** Secondary structure of matrix metalloproteinase 12 (MMP12): (A) PDB: 1YCM (structure determined using solution NMR).

**B**

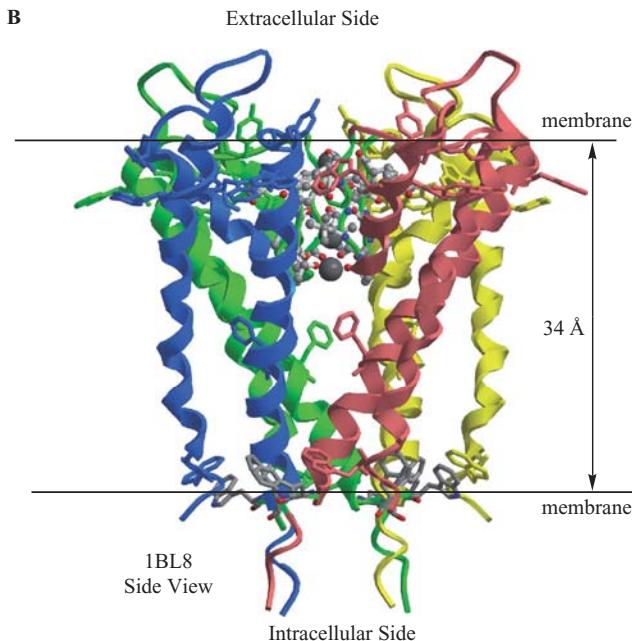
PDB: 1RMZ



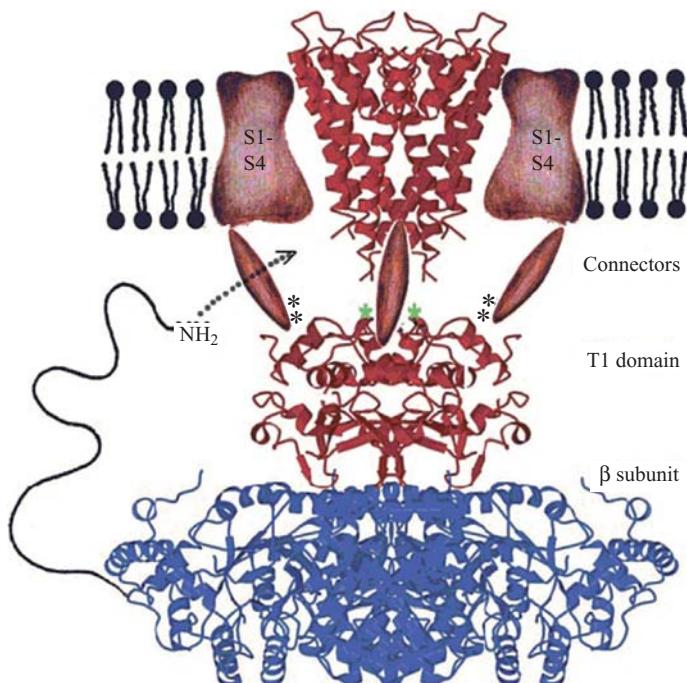
**Figure 3.18 (Continued) (B)** PDB: 1RMZ (structure determined using X-ray crystallography). Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft Corporation.)

**A**

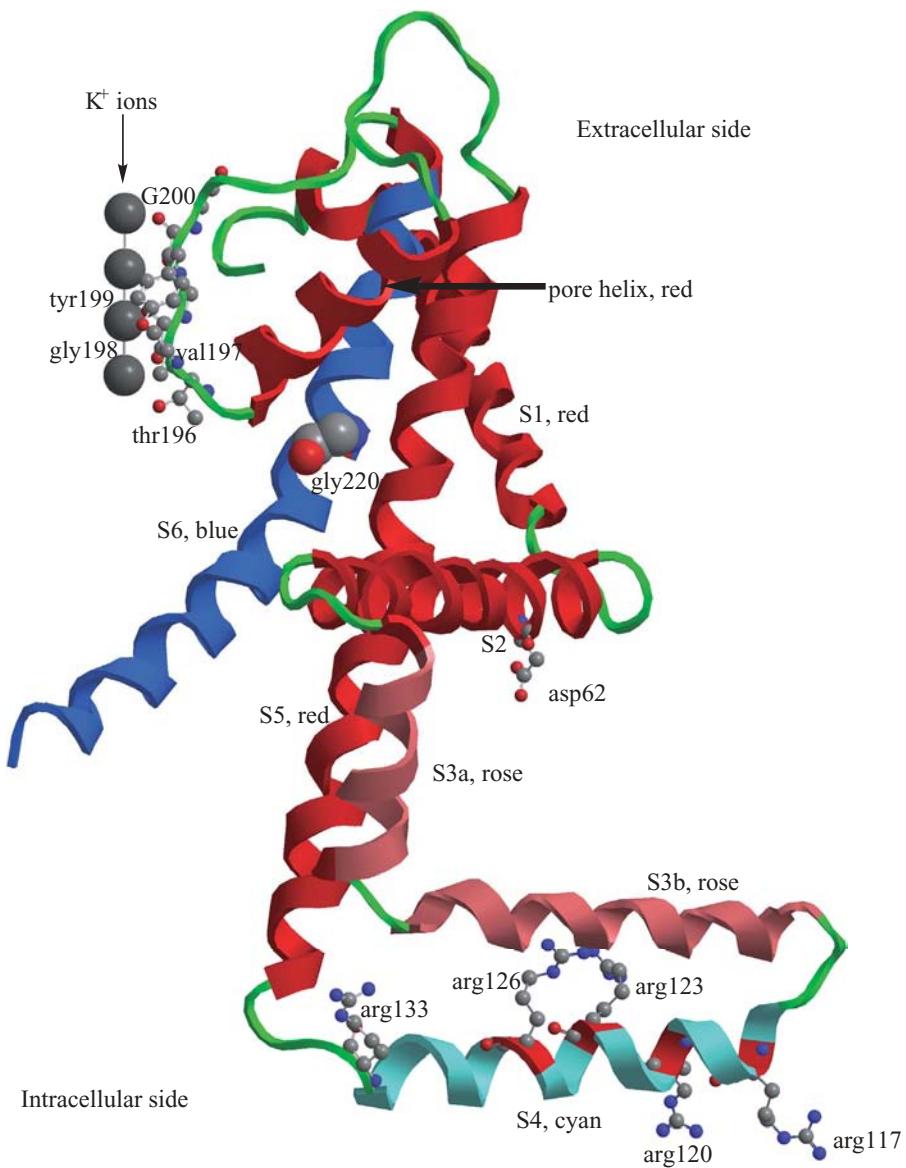
**Figure 5.5 (A)** Top-down view of the  $\text{K}^+$  ion channel, PDB: 1BL8.



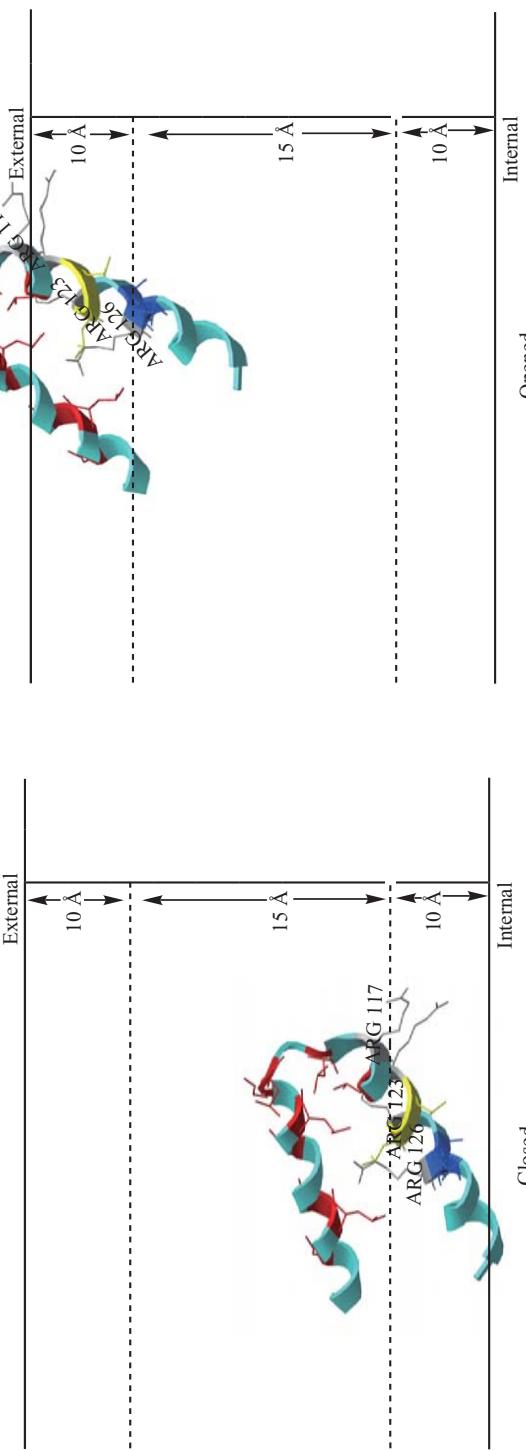
**Figure 5.5 (Continued)** (B) Side view of the  $\text{K}^+$  ion channel, PDB: 1BL8. Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. (Printed with permission of CambridgeSoft Corporation.)



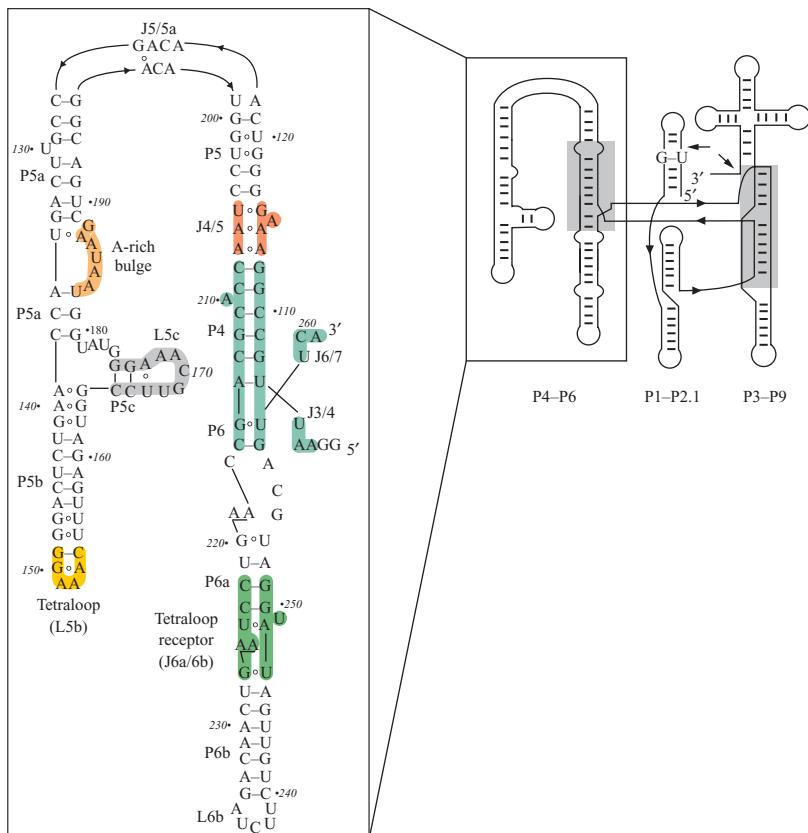
**Figure 5.6** Composite model of a voltage-dependent  $\text{K}^+$  channel. (From Figure 5 of reference 16. Reprinted with permission of AAAS.)



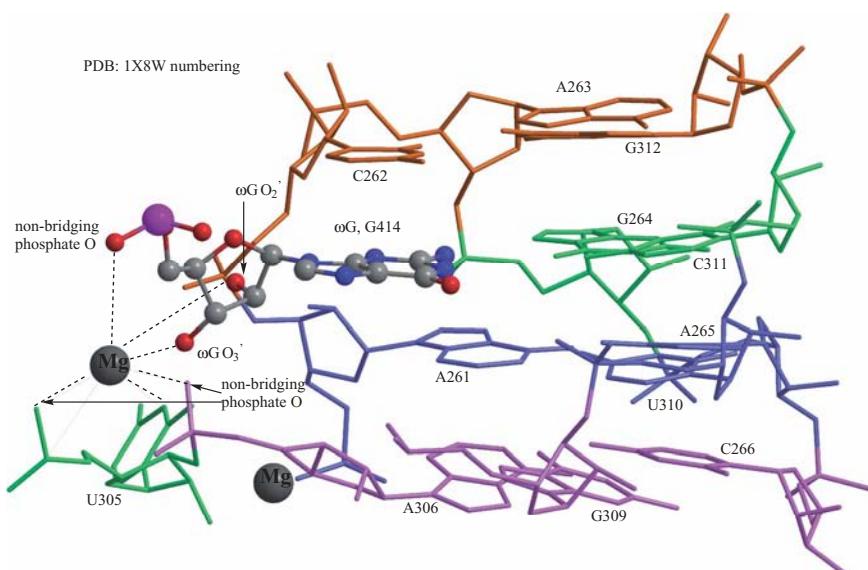
**Figure 5.7** View of the K<sup>+</sup> ion channel, the pore helix, and helices S1–S6, PDB: 1ORQ. Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. (Printed with permission of CambridgeSoft Corporation.)



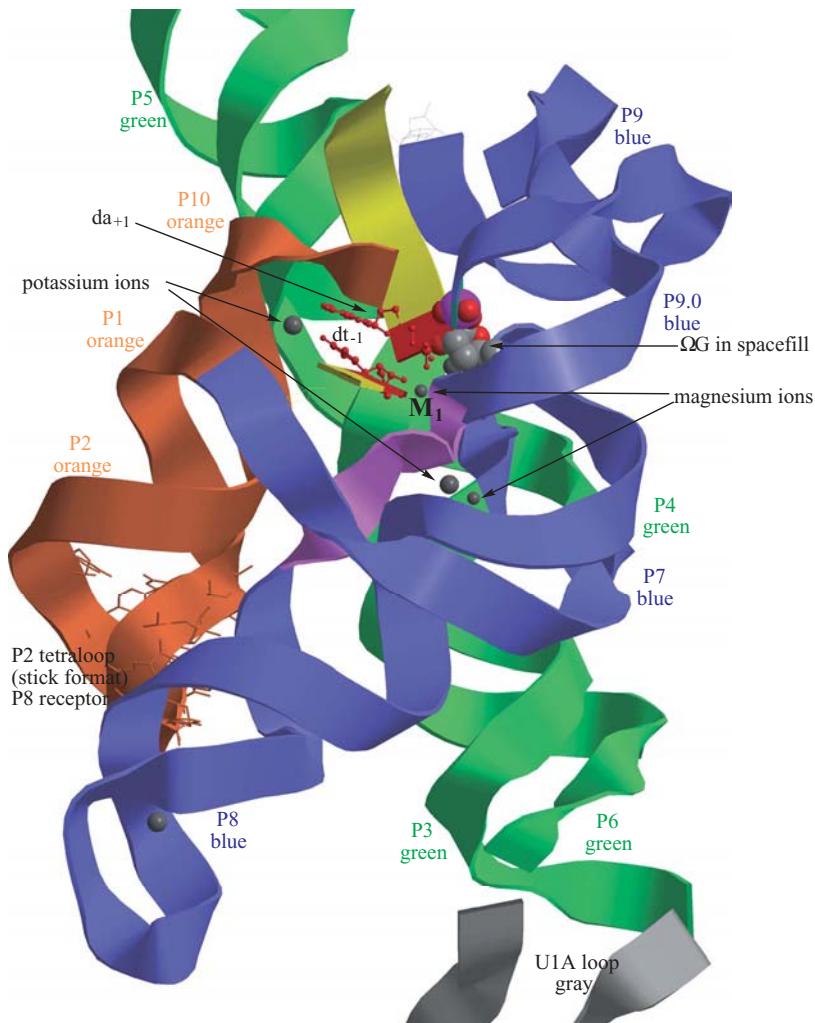
**Figure 5.8** Closed and opened views of the potassium ion channel pore according to reference 25. S4 helix of PDB: 1ORQ. Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. (Printed with permission of CambridgeSoft Corporation.)



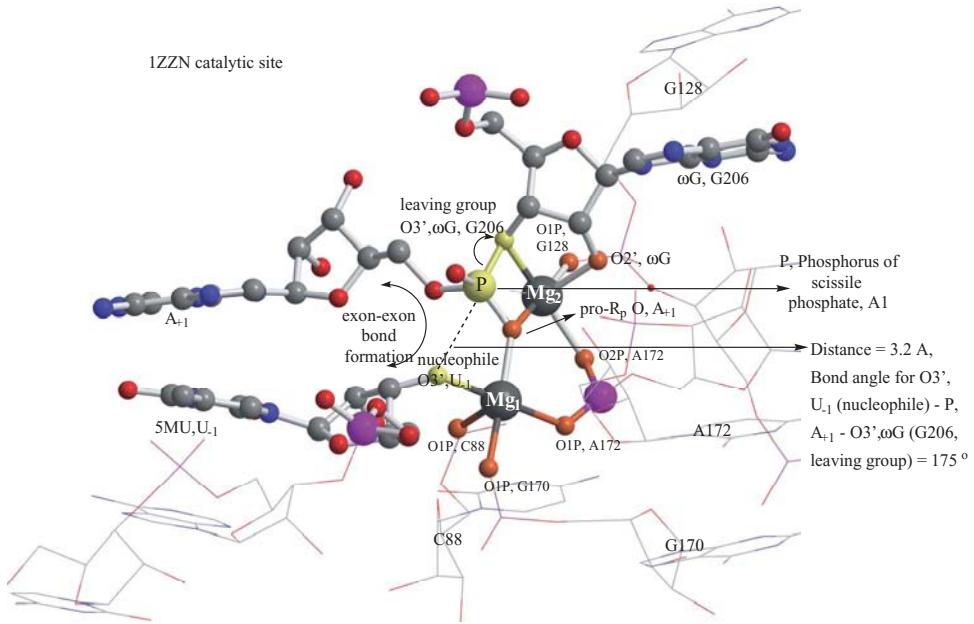
**Figure 6.4** Secondary structure of a group I intron ribozyme. (From Figure 1 of reference 13. Reprinted with permission of AAAS.)



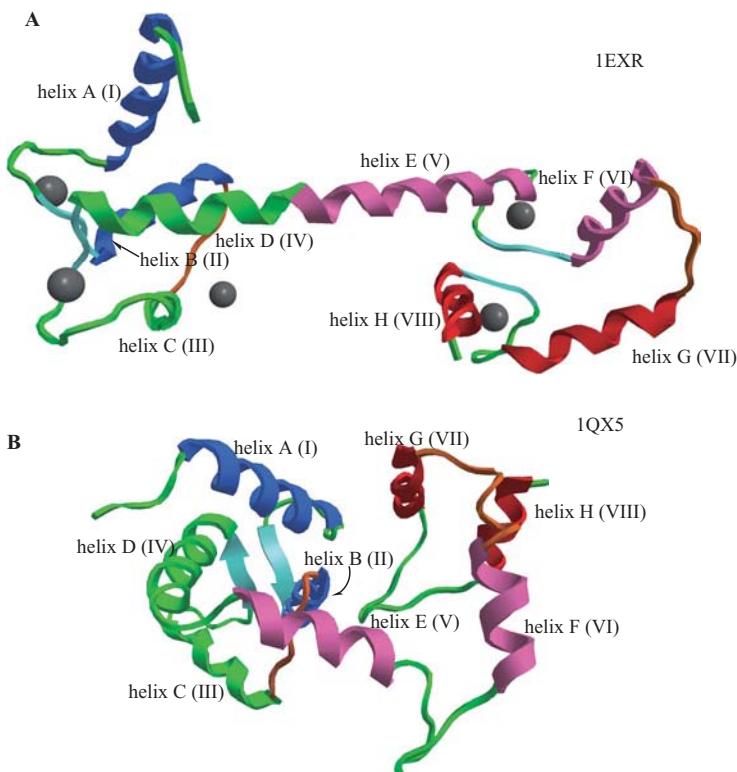
**Figure 6.6** Base triples surrounding  $\Omega G$  in PDB: 1X8W. Magnesium ions in position to facilitate nucleophilic attack. Visualized using Chem3D Ultra and ChemDraw Ultra, version 10.0. (Printed with permission of CambridgeSoft Corporation.)



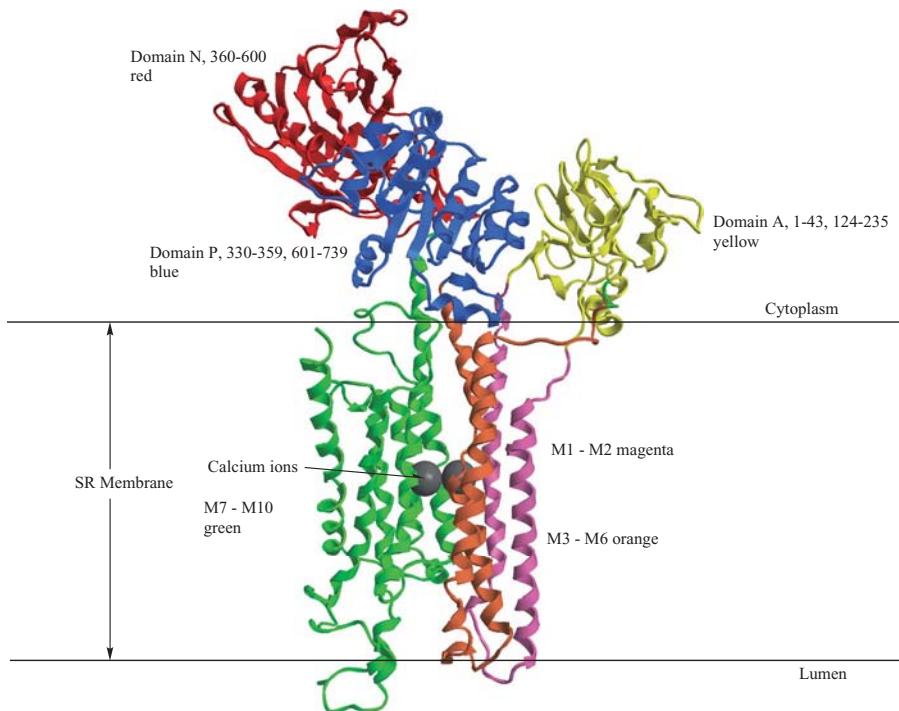
**Figure 6.7** A bacterial self-splicing group I intron with both exons (PDB: 1U6B). Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0.(Printed with permission of CambridgeSoft Corporation.)



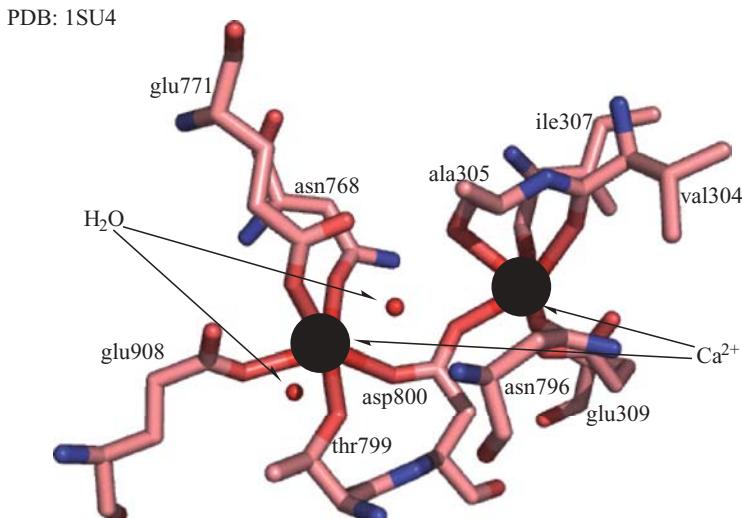
**Figure 6.8** A group I intron transition state stabilized by two magnesium ions (PDB: 1ZZN). Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. (Printed with permission of CambridgeSoft Corporation.)



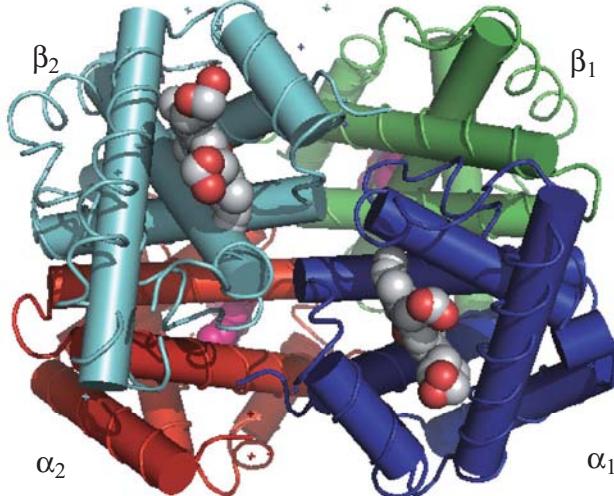
**Figure 6.22** X-ray crystallographic structures of: (A)  $\text{Ca}^{2+}$ -saturated calmodulin, PDB: 1EXR; and (B) apo-calmodulin, PDB: 1QX5. Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. (Printed with permission of CambridgeSoft Corporation.)



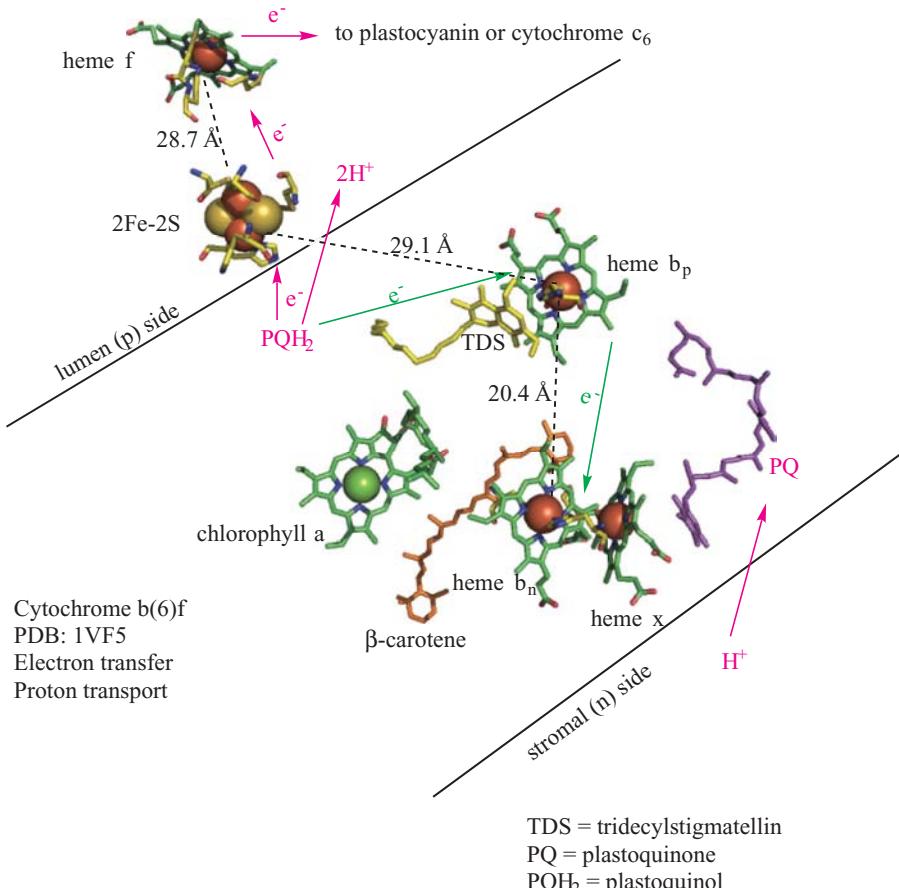
**Figure 6.29**  $\text{Ca}^{2+}$ -ATPase in the  $\text{Ca}_2\text{E}_1$  state as represented by the PDB: 1SU4 X-ray crystallographic structure. Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. (Printed with permission of CambridgeSoft Corporation.)



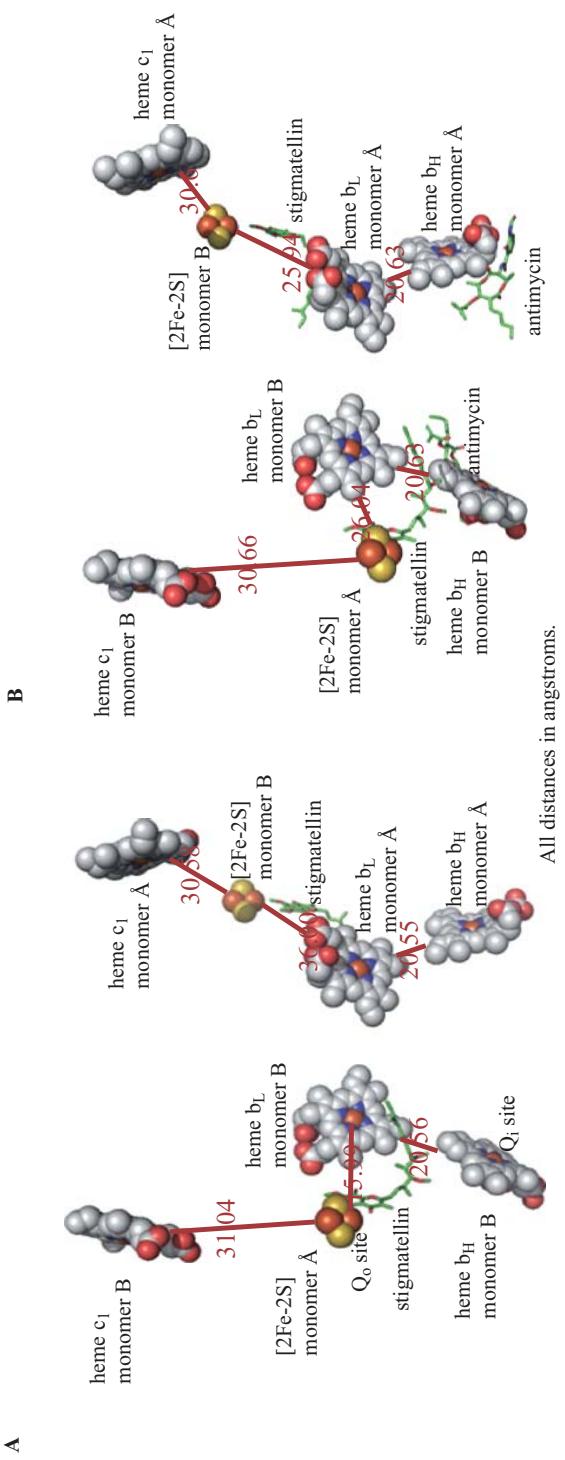
**Figure 6.30** The calcium ion coordination sphere for the  $\text{Ca}^{2+}$ -ATPase as represented by the PDB: 1SU4 X-ray crystallographic structure. Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft Corporation.)



**Figure 7.2** Quaternary structure of deoxyhemoglobin tetramer (PDB: 4HHB). Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft Corporation.)

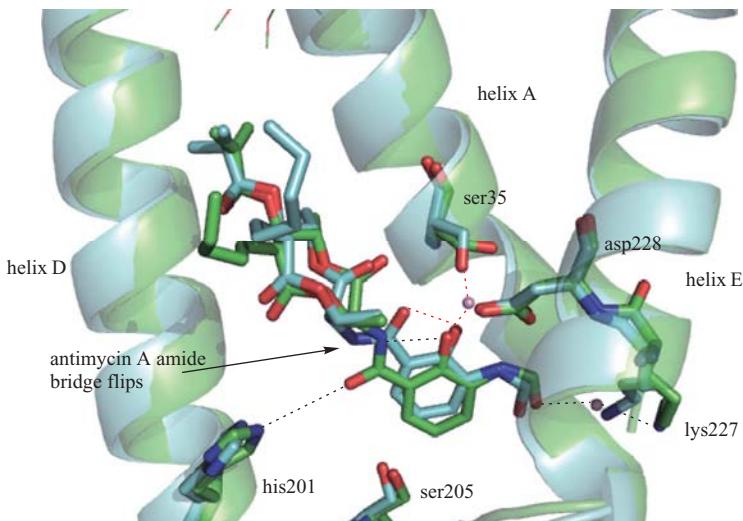


**Figure 7.26** Electron transfer and proton transport in cytochrome b(6)f: cofactors (PDB: 1VF5). Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft Corporation.)

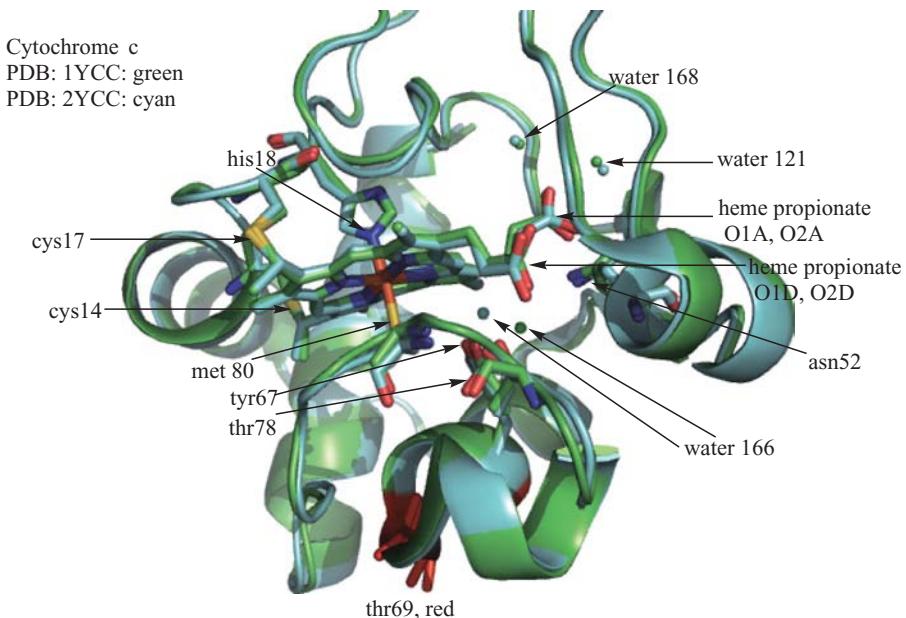


PDB: 1PP9 cytochrome  $bc_1$  with inhibitor stigmatellin  
Note [2Fe-2S] centers interact with hemes in the other monomer.

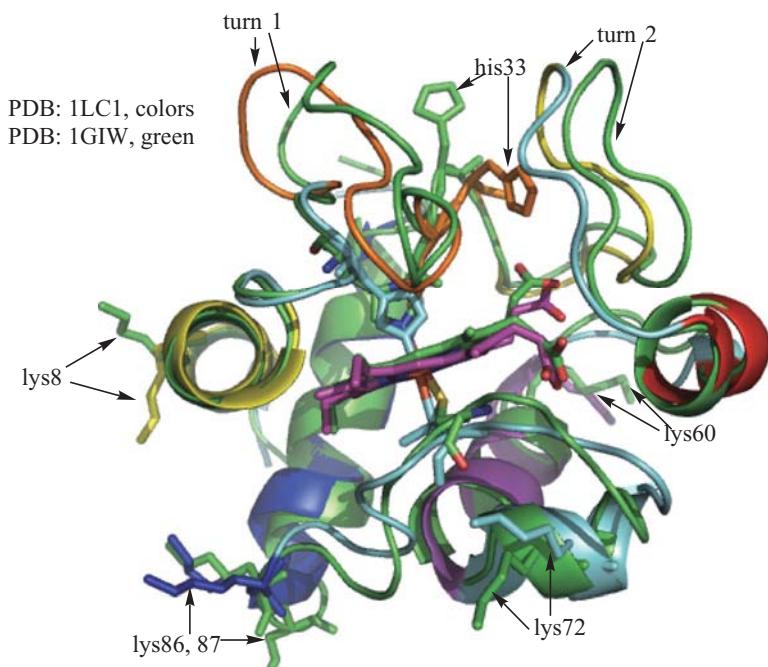
**Figure 7.30** Comparison of cytochrome  $bc_1$  structures: (A) PDB: 1PP9 with inhibitor stigmatellin; and (B) PDB: 1PPJ with inhibitors stigmatellin and antimycin A. Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft Corporation.)



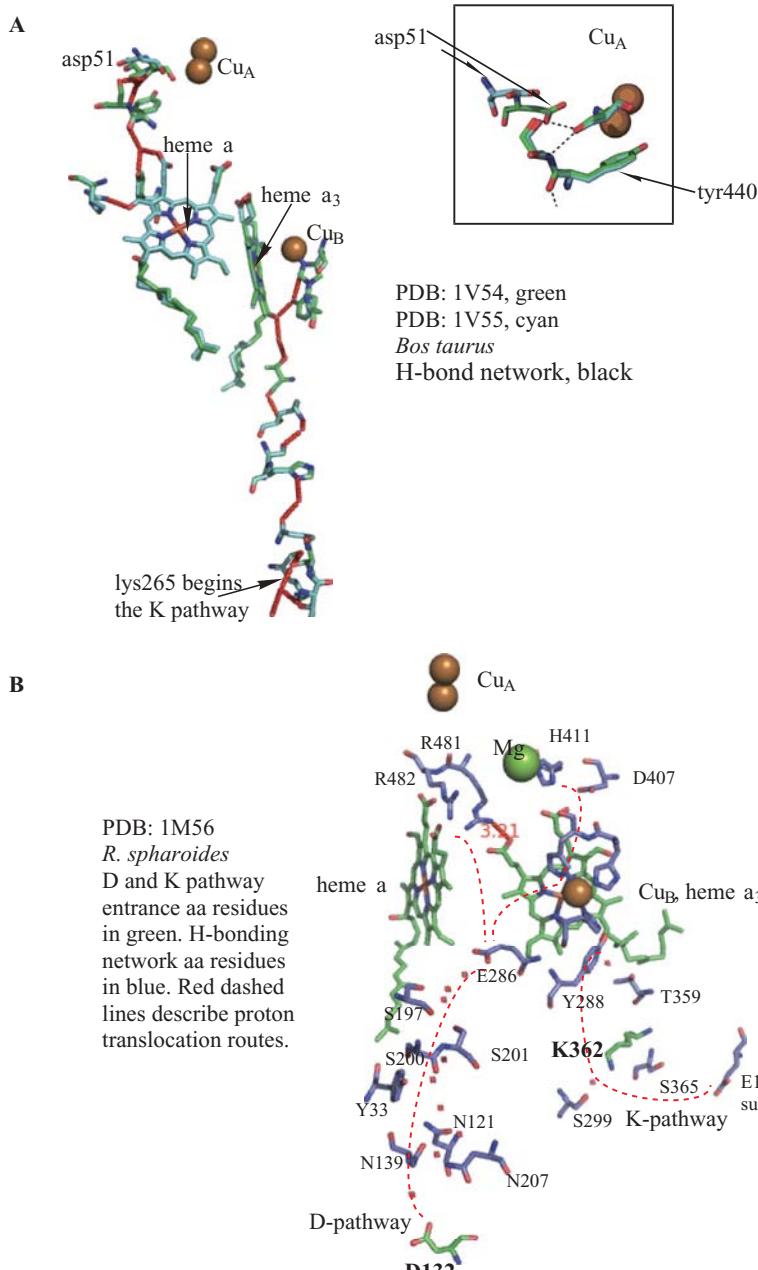
**Figure 7.31** Comparison of antimycin A positions in PDB: 1NTK and PDB: 1PPJ. Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft Corporation.)



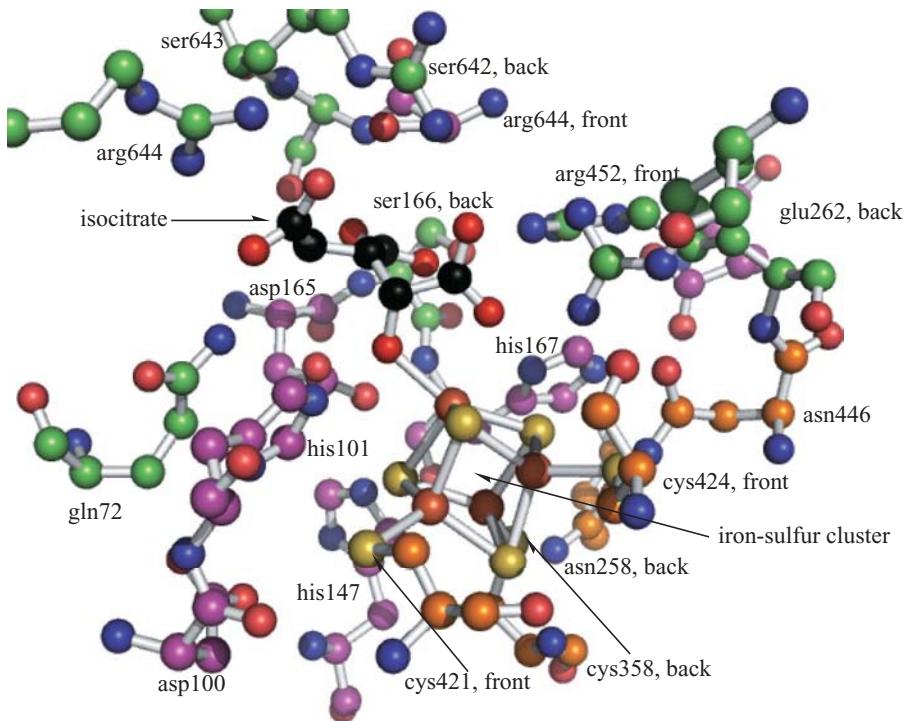
**Figure 7.33** Two yeast cytochrome c molecules (PDB: 1YCC, Fe(II), green, and PDB: 2YCC, Fe(III), cyan). Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft Corporation.)



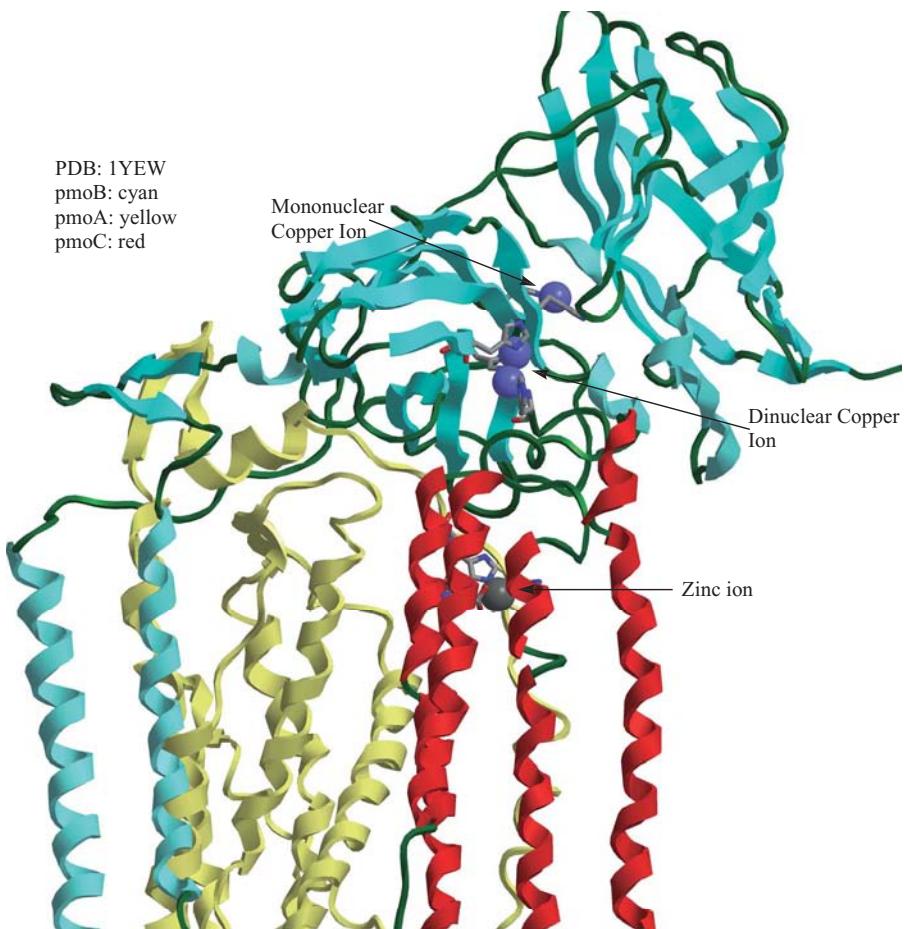
**Figure 7.34** Two horse cytochrome c molecules, (PDB: 1LC1, in various colors as described in the text, and PDB: 1GIW in green). Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft, Corporation.)



**Figure 7.41** (A) Two *B. taurus* cytochrome c oxidase molecules, (PDB: 1V54, in green, and PDB: 1V55 in cyan). Insert shows change in asp51 position. (B) The *R. sphaeroides* cytochrome c oxidase molecule (PDB: 1M56). (Adapted with permission from Figure 1b of reference 143. Copyright 2004, with permission from Elsevier.) Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft Corporation.)



**Figure 7.50** Important amino acid residues surrounding the isocitrate substrate (black carbon atoms) and the iron–sulfur cluster of aconitase (PDB: 7ACN). Visualized using The PyMOL Molecular Graphics System and ChemDraw Ultra, version 10.0. (Printed with permission of Delano Scientific, LLC and CambridgeSoft Corporation.)



**Figure 7.51** Particulate MMO, pMMO (PDB: 1YEW). Visualized using CambridgeSoft Chem3D Ultra 10.0 with notations in ChemDraw Ultra 10.0. (Printed with permission of CambridgeSoft Corporation.)