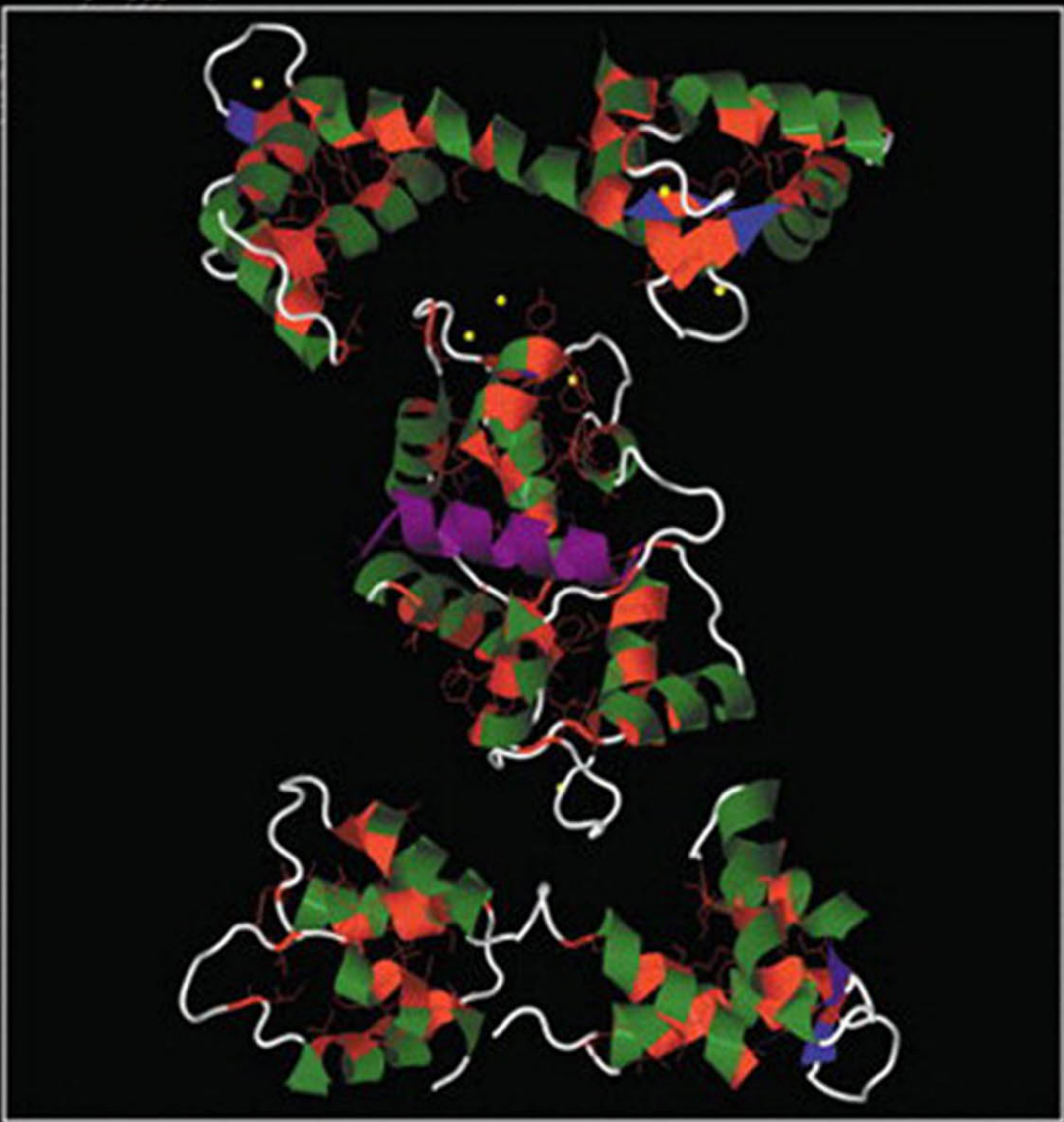
Fifth Edition

Physical Chemistry

Principles and Applications in Biological Sciences



Tinoco • Sauer • Wang
Puglisi • Harbison • Rovnyak

Pearson Advanced Chemistry Series

The need for innovation, adaptability, and discovery is more glaring in our world today than ever. Globally we all look to "thought leaders" for progress, many of whom were, are, or will be students of science. Whether these students were inspired by a book, a teacher, or technology, we at Pearson Education want to do our part to support their studies. The new **Advanced Chemistry Series** supports upper-level course work with cutting-edge content delivered by experienced authors and innovative multimedia. We realize chemistry can be a difficult area of study and we want to do all we can to encourage not just completion of course work, but also the building of the foundations of remarkable scholarly and professional success. Pearson Education is honored to be partnering with chemistry instructors and future STEM majors. To learn more about Pearson's **Advanced Chemistry Series**, explore other titles, or access materials to accompany this text and others in the series, please visit *www.pearsonhighered.com/advchemistry*.

Books available in this series include:

Analytical Chemistry and Quantitative Analysis
by David S. Hage *University of Nebraska Lincoln* and
James R. Carr *University of Nebraska Lincoln*

Forensic Chemistry
by Suzanne Bell *West Virginia University*

Inorganic Chemistry
by Gary Miessler *St. Olaf College,* Paul Fischer *Macalester College,*
Donald Tarr *St. Olaf College*

Medicinal Chemistry: The Modern Drug Discovery Process
by Erland Stevens *Davidson College*

Physical Chemistry: Quantum Chemistry and Molecular Interactions
by Andrew Cooksy *University of California San Diego*

Physical Chemistry: Thermodynamics, Statistical Mechanics, and Kinetics
by Andrew Cooksy *University of California San Diego*

Physical Chemistry
by Thomas Engel *University of Washington* and Philip Reid *University of Washington*

Physical Chemistry: Principles and Applications in Biological Sciences
by Ignacio Tinoco Jr. *University of California Berkeley*, Kenneth Sauer *University of California Berkeley*, James C. Wang *Harvard University*, Joseph D. Puglisi *Stanford University*, Gerard Harbison *University of Nebraska Lincoln*,
David Rovnyak *Bucknell University*

Quantum Chemistry
by Ira N. Levine *Brooklyn College, City College of New York*

FIFTH EDITION

Physical Chemistry

Principles and Applications in Biological Sciences

Ignacio Tinoco, Jr.
University of California, Berkeley

Kenneth Sauer
University of California, Berkeley

James C. Wang
Harvard University

Joseph D. Puglisi
Stanford University

Gerard Harbison
University of Nebraska, Lincoln

David Rovnyak
Bucknell University

PEARSON

Boston Columbus Indianapolis New York San Francisco Upper Saddle River
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montréal Toronto
Delhi Mexico City São Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

Editor in Chief: Adam Jaworski
Executive Editor: Jeanne Zalesky
Senior Marketing Manager: Jonathan Cottrell
Project Editor: Jessica Moro
Assistant Editor: Lisa R. Pierce
Editorial Assistant: Lisa Tarabokjia
Senior Marketing Assistant: Nicola Houston
Associate Media Producer: Erin Fleming
Managing Editor, Chemistry and Geosciences:
Gina M. Cheselka
Project Manager, Production: Wendy Perez
Full Service/Composition: GEX Publishing Services
Design Manager: Mark Ong
Interior and Cover Design: Gary Hespenheide
Manager of Permissions: Timothy Nicholls
Permissions Specialist: Joseph Croscup
Operations Specialist: Jeffrey Sargent

Cover Illustration courtesy of Gerard Harbison. Illustration created using Jmol: an open-source Java viewer for chemical structures in 3D. (*www.jmol.org*)

Data obtained from the RCSB Protein Data Bank (*www.pdb.org*). H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne (2000) The Protein Data Bank. Nucleic Acids Research, 28: 235–242. Data files contained in the PDB archive include:

Top Structure: PDB ID 3CLN

Y. S. Babu, C. E. Bugg, W. J. Cook (1988) Structure of calmodulin refined at 2.2 A resolution. J. Mol. Biol. 204: 191–204.

Middle Structure: PDB ID 1CM1

M.E. Wall, J. B. Clarage, G. N. Phillips Jr. (1997) Motions of calmodulin characterized using both Bragg and diffuse X-ray scattering. Structure 5: 1599–1612.

Bottom Structure: PDB ID 1CFD

H. Kuboniwa, N. Tjandra, S. Grzesiek, H. Ren, C. B. Klee, C.B., A. Bax (1995) Solution structure of calcium-free calmodulin. Nat. Struct. Biol. 2: 768–776.

Credits and acknowledgments borrowed from other sources and reproduced, with permission, in this textbook appear on the appropriate page within the text.

Library of Congress Cataloging-in-Publication Data available upon request from Publisher.

1 2 3 4 5 6 7 8 9 10—**EB**—15 14 13 12

PEARSON

www.pearsonhighered.com

ISBN–10: 0-13-605606-7; ISBN–13: 978-0-13-605606-5

Brief Contents

Contents

Preface

There is a deep sense of pleasure to be experienced when the patterns and symmetry of nature are revealed. Physical chemistry provides the methods to discover and understand these patterns. We think that not only is it important to learn and apply physical chemistry to biological problems, it may even be fun. In this book, we have tried to capture some of the excitement of making new discoveries and finding answers to fundamental questions.

This is not an encyclopedia of physical chemistry. Rather, we have written this text specifically with the life-science student in mind. We present a streamlined treatment that covers the core aspects of biophysical chemistry (thermodynamics and kinetics as well as quantum mechanics, spectroscopy, and X-ray diffraction), which are of great importance to students of biology and biochemistry. Essentially all applications of the concepts are systems of interest to life-science students; nearly all the problems apply to life-science examples.

For the fifth edition, we have extensively revised and updated the treatment of biophysical chemistry, bringing in theoretical approaches earlier and also updating the text to current IUPAC conventions. We have added a new chapter on electrochemistry and expanded our treatment of single molecule methods, quantum mechanics, and magnetic resonance.

Chapter 1 introduces representative areas of active current research in biophysical chemistry and molecular biology: the human genome, the transfer of genetic information from DNA to RNA to protein, ion channels, and cell-to-cell communication. We encourage students to read the current literature to see how the vocabulary and concepts of physical chemistry are used in solving biological problems.

Chapters 2 through 4 cover the laws of thermodynamics and their applications to chemical reactions and physical processes. Essentially all of the examples and problems deal with biochemical and biological systems. For example, after defining work as a force multiplied by the distance moved (the displacement), we discuss the experimental measurement of the work necessary to stretch a single DNA molecule from its random-coiled form to an extended rod, introducing the intuitive and accessible concept of molecular force microscopy. We also include a new and more comprehensive treatment of heat capacities, beginning with the kinetic theory of gases, which is now treated much earlier in chapter 2, and moving systematically to a consideration of what affects the heat capacity of a protein. Molecular interpretations of energies and entropies are emphasized in each of chapters 2 through 4. We also introduce isothermal titration calorimetry in chapter 3. Despite this new content, the length of chapters 2 through 4 as been reduced by over 30 pages, largely by eliminating redundant material.

In chapter 5, we show how the thermodynamic laws discussed in chapters 2 through 4 can be explained by a statistical treatment of molecular motion and interactions, and apply these statistical methods to the conformation of proteins and DNA and the binding of ligands. This section has been combined with the conceptually-related statistical treatment of Maxwell-Boltzmann gases and appears much earlier. In chapter 6, we immediately use these statistical insights to explain physical phenomena such as phase transitions, ligand binding, and surface and membrane effects.

In chapter 7, we present a new and integrated treatment of electrical and electron-transfer phenomena in biophysics, starting with classical electrochemistry, and considering how the chemical processes of electron transfer are linked to the physical processes of ion translocation to explain most of biological energy transduction.

Chapters 8 through 10 cover molecular motion and chemical kinetics. Chapter 8 starts with a discussion of molecular collisions, random walks, and brownian motion. Fluorescence microscopic tracking of single protein molecules diffusing in membranes is shown to beautifully corroborate Einstein's equation relating average distance traveled by a single molecule to its bulk diffusion coefficient. Following this direct experimental demonstration of thermal motion of a molecule, we consider the bulk transport of molecules by diffusion, sedimentation, viscous flow, and electrophoresis. The next two chapters deal with general chemical kinetics and enzyme kinetics, including single-molecule enzyme kinetics.

In the 5th edition, we reflect the rapidly expanding importance of quantum mechanics and diverse powerful spectroscopies in understanding molecular biological phenomena by presenting these subjects in four more focused and augmented chapters. Chapter 11, "Molecular Structures and Interactions: Theory," now focuses solely on the origins and key introductory results of quantum theory, including a review of the postulates. Chapter 12, "Molecular Structures and Interactions: Biomolecules," presents molecular orbital theory, interactions, and an overview of computational methods applied to macromolecules. Similarly, the treatment of spectroscopy is now more focused in two separate chapters on optical (chapter 13) and magnetic (chapter 14) methods, respectively. Chapter 13 increases emphasis on absorption and fluorescence, and includes new material on protein IR spectroscopy. Chapter 14 introduces the classical framework for NMR in more detail and covers new methods in multidimensional and diffusion NMR.

Chapter 15 discusses X-ray diffraction, electron microscopy, and scanning microscopies (such as atomic force microscopy), and emphasizes how structures are determined experimentally. We added a new section on crystal lattices and symmetry, and expanded the discussion of modern methods such as X-ray imaging and free-electron lasers.

A new appendix in the fifth edition is an accessible, self-contained, and pragmatic review of the mathematics expectations in this text. We hope the carefully defined scope of the mathematics (a characteristic of previous editions) will be reassuring in preparing to study this text.

We are gratified by the number of faculty who have elected to use this book over the many years since it was first published. We are also grateful for the many students and faculty who have given us their thoughts and impressions. Such feedback has helped improve the book from edition to edition. We are particularly grateful to those of our colleagues who commented on the fifth edition:

Noah W. Allen, III—University of North Carolina, Asheville
Jason Benedict—University of Buffalo
Tim Keiderling—University of Illinois, Chicago
Ruth Ann Murphy—University of Mary Hardin Baylor
Tatyana Smirnova—North Carolina State University
Keith J. Stine—University of Missouri-St. Louis
Gianluigi Veglia—University of Minnesota
Jeff Woodford—Missouri Western State University
Danny Yeager—Texas A&M University
Kazushige Yokoyama—State University of New York, Geneseo

We welcome your comments.

Ignacio Tinoco, Jr.
Kenneth Sauer
James C. Wang
Joseph D. Puglisi
David Rovnyak
Gerard Harbison

New to This Edition

The major theme of this revision is to update and reorder a classic text to reflect changes in scientific knowledge and student learning styles, while retaining the time-tested central core.

- Over 200 new and revised figures to help students visualize and understand the concepts discussed within each chapter.
- At least 20-25% new and revised end-of-chapter problems
- New treatment of the most modern methods, including free-electron laser X-ray imaging, single-molecule microscopy, and isothermal titration calorimetry.
- Quantum mechanics has been split into two chapters covering basic theory and molecular properties; spectroscopy likewise has been split into optical spectroscopy and NMR.
- New chapter on electrical phenomena (chapter 7), which integrates electrobiochemistry, redox biology, and electrophysiology in a single location.
- New focus on single molecule microscopy, dynamics, spectroscopy, and kinetics.
- New molecular based development of heat capacities from ideal gases to proteins, along with new sections on calorimetric measurements.
- Better integration of statistical theories of molecular conformation and binding, which now appear earlier in the text and directly precede experimental treatment of phase transitions, binding, and membranes.
- Addition of **MasteringChemistry** for Physical Chemistry. The **MasteringChemistry** platform is the most widely used and effective online homework, tutorial, and assessment system for the sciences. It delivers self-paced tutorials that focus on your course objectives, provide individualized coaching, and respond to each student's progress. The Mastering system helps instructors maximize class time with easy-to-assign, customizable, and automatically graded assessments that motivate students to learn outside of class and arrive prepared for lecture or lab.

Chapter-by-Chapter Changes

Chapter 1

- new sections on neuroscience and on single-molecule methods
- the Human Genome Initiative section was brought up to date

Chapter 2

- new treatment of molecular-force microscopy
- new development of the concept of the heat capacity at a molecular level, beginning with the kinetic theory of monatomic gases, and extending to proteins
- new and clearer discussion of paths, states and phase transitions

Chapter 3

- expanded introduction and illustrations of Carnot cycle and associated principles
- the thermodynamic square helps students see thermodynamics holistically

Chapter 4

- expanded introduction and illustrations of chemical potential
- isothermal titration calorimetry is included to reflect its broad importance in drug discovery and characterizing substrate interactions

Chapter 5

- reorganized treatment of statistical thermodynamics, beginning with the Maxwell-Boltzmann distribution of gases, and proceeding to statistics of discrete and quantized systems, helix-coil statistics of macromolecules, and ligand binding

Chapter 6

- the Clapeyron equations for phase transitions are now developed from the visual concept of three-dimensional free energy surfaces

Chapter 7

- brand new chapter on electrical phenomena in biophysics
- begins with electrochemical cells and biophysical applications of the Nernst equation
- moves on to transmembrane equilibria, the Donnan effect, ion pumps, and neuroelectrophysiology
- ends with biological redox reactions and a full, up-to-date exploration of oxidative phosphorylation

Chapter 8

- new treatment of molecular collisions and their effect on mean square displacements
- new development of the diffusion equations in one, two and three dimensions using Einstein's original arguments
- new discussion of single molecule microscopic measurements of diffusion in two- and three-dimensions
- new treatment of effect of shape on diffusion and sedimentation coefficients
- improved and rewritten section on analytical ultracentrifugation

Chapter 9

- sections on differential and integrated rate equations rewritten and clarified
- new section on single-molecule kinetics

Chapter 10

- new discussion of Michaelis-Menten kinetics with an emphasis on direct least-squares fitting of data, compared to older, statistically inferior linearization methods

Chapter 11

- now pedagogically focused solely on introducing quantum mechanical origins and key applications
- stronger connection between classical and quantum mechanics is forged
- postulates are now included which can help students build their understanding of the logical structure of quantum mechanics

Chapter 12

- streamlined for focus solely on molecular orbitals, intermolecular and intramolecular interactions
- focus on analyzing coefficients in molecular orbital wavefunctions through introductory material and problem sets

Chapter 13

- introducing optical principles of spectroscopy solely will help students and instructors better organize and build their approach to spectroscopy in biophysics;
- focus on visualizing spectroscopic concepts such as waves, the transition dipole, and the Franck-Condon principle
- protein infra-read spectroscopy gives valuable information on structure, complementary to circular dichroism

Chapter 14

- introduces the vector model for pulsed magnetic resonance
- characterizing molecular rotation and size via relaxation, connecting to Stokes-Einstein principles of chapter 8
- protein 3D-NMR methods and gradient diffusion methods for particle size and microscopy

Chapter 15

- new and highly visual treatment of lattices and symmetry
- new treatment of ‘frontier’ research techniques such as x-ray imaging using free-electron lasers

About the Authors

Ignacio Tinoco was an undergraduate at the University of New Mexico, a graduate student at the University of Wisconsin, Madison, and a postdoctoral fellow at Yale. He then went to the University of California, Berkeley, where he has remained. His research interest has been on the structures of nucleic acids, particularly RNA. He was chairman of the Department of Energy committee that recommended in 1987 a major initiative to sequence the human genome. He is a member of the National Academy of Sciences and of the American Academy of Arts and Sciences. His present research is using single-molecule methods to determine how the ribosome synthesizes proteins.

Kenneth Sauer grew up in Cleveland, Ohio, and received his A.B. in chemistry from Oberlin College. Following his Ph.D. studies in gas-phase physical chemistry at Harvard, he spent three years teaching at the American University of Beirut, Lebanon. A postdoctoral opportunity to learn from Melvin Calvin about photosynthesis in plants led him to the University of California, Berkeley, where he has been since 1960. Teaching general chemistry and biophysical chemistry in the Chemistry Department has complemented research in the Physical Biosciences Division of the Lawrence Berkeley National Lab involving spectroscopic studies of photosynthetic light reactions and their role in water oxidation. His other activities include reading, renaissance and baroque choral music, canoeing, and exploring the Sierra Nevada with his family and friends.

James C. Wang was on the faculty of the University of California, Berkeley, from 1966 to 1977. He then joined the faculty of Harvard University, where he is presently Mallinckrodt Professor of Biochemistry and Molecular Biology. His research focuses on DNA and enzymes that act on DNA, especially a class of enzymes known as DNA topoisomerases. He has taught courses in biophysical chemistry and molecular biology and has published over 200 research articles. He is a member of Academia Sinica, the American Academy of Arts and Sciences, and the U.S. National Academy of Sciences.

Joseph Puglisi was born and raised in New Jersey. He received his B.A. in chemistry from The Johns Hopkins University in 1984 and his Ph.D. from the University of California, Berkeley, in 1989. He has studied and taught in Strasbourg, Boston, and Santa Cruz, and is currently professor of structural biology at Stanford University. His research interests are in the structure and mechanism of the ribosome and the use of NMR spectroscopy to study RNA structure. He has been a Dreyfus Scholar, Sloan Scholar, and Packard Fellow.

Gerard Harbison was born in the United Kingdom and raised there and in Ireland. He received his B.A. in biochemistry from Trinity College, Dublin, and his Ph.D. in biophysics from Harvard University. After a brief postdoctoral sojourn at the Max-Planck Institute for Polymer Research in Mainz, Germany, he joined the faculty of Stony Brook University, and then moved to the University of Nebraska Lincoln. He is a Dreyfus Teacher-Scholar, Lilly Foundation Teacher-Scholar and Presidential Young Investigator. His research interests are in nuclear magnetic resonance and electronic structure theory.

David Rovnyak, a native of Charlottesville, Virginia, earned his B.S. in Chemistry at the University of Richmond and Ph.D. in physical chemistry from the Massachusetts Institute of Technology. After performing post-doctoral study at the Harvard Medical School under an NIH-NRSA fellowship, he joined Bucknell University where he has been recognized with the Bucknell Presidential Teaching Award for Excellence. His research focuses on new methods for NMR spectroscopy and physico-chemical behavior of bile acids.

Chapter 1

Introduction

Physical chemistry is everywhere. Physical chemical principles are basic to the methods used to determine the sequence of the human genome, obtain atomic resolution structures of proteins and nucleic acids, and learn how biochemicals react and interact to make a cell function. Once you learn physical chemistry, you will subconsciously apply your knowledge to each scientific paper you read and to each explanation of an experiment you hear or propose yourself. “What provides the energy for the reaction? What about the Second Law? The proposed mechanism seems to violate microscopic reversibility. An intermediate is proposed; I can identify it by fluorescence, or nuclear magnetic resonance (NMR), or Raman scattering.” Most importantly, you will realize that to study any biological process you need to use the methods of physical chemistry. Understanding how the brain works has become among the most active areas of biological research. The neurons in the brain communicate with each other by electrical signals and the transfer of neurotransmitters. The structures of all the players (neurotransmitters, neurotransporters, receptors, ion channels, synapsis proteins, and so forth) need to be determined by X-ray diffraction or by NMR. These molecules move, change conformation, and interact with other molecules in response to action potentials. Their positions, interactions, rates of motion, shapes, and sizes are measured by high resolution optical microscopy, atomic force microscopy, or electron microscopy. Functional magnetic resonance imaging, and infrared absorption and scattering reveal which parts of the brain are most active while you are thinking about whatever you are doing. In the next fifty years new methods and new applications of old methods will reveal new answers and completely new questions to ask about every biological process.

Physical chemistry is a set of principles and experimental methods for exploring chemical and biological systems. The power of physical chemistry lies in its generality. The principles described in this book can be applied to systems as large as the cosmos and as small as an individual atom. Physical chemistry has been especially powerful in understanding fundamental biological processes. In the following chapters, we will present the principles of thermodynamics, transport properties, kinetics, quantum mechanics and molecular interactions, spectroscopy, and scattering and diffraction. We will also discuss various experimentally measurable properties such as enthalpy, electrophoretic mobility, light absorption, and X-ray diffraction. All these experimental and theoretical methods give useful information about whatever problem you want to solve. We emphasize the molecular interpretation of these methods and stress biochemical and biological applications. By learning the principles behind the methods, you will be able to judge the conclusions obtained from them. This is the first step in inventing new methods or discovering new concepts.

First, a quick tour of the book. Chapters 2 through 4 cover the fundamentals of thermodynamics and their applications to chemical reactions and physical processes. Because these chapters review material usually covered in beginning chemistry courses, we emphasize the applications to biological macromolecules. Chapter 5 covers the statistical basis of thermodynamics; it provides a molecular interpretation of thermodynamics. Cooperative binding of ligands to macromolecules, plus helix-coil transitions in nucleic acids and proteins are described. Chapter 6 covers physical equilibria, including osmotic pressure, equilibrium dialysis, and membrane equilibria. Chapter 7 deals with electrochemistry, including Galvanic cells, Donnan equilibria, and transmembrane potentials. The effect of sizes and shapes of molecules on their translational and rotational motions in gases, liquids, and gels are discussed in chapter 8. The driving forces for molecular motion are either random thermal forces that cause diffusion or the directed forces in sedimentation, flow, and electrophoresis. Chapter 9 describes general kinetics, and chapter 10 concentrates on the kinetics of enzyme-catalyzed reactions. Chapter 11 introduces the quantum mechanical principles necessary for understanding bonding and spectroscopy, and chapter 12 describes calculations of protein and nucleic acid conformations using classical force fields (Coulomb's Law, van der Waals' potential). Chapter 13 includes the main spectroscopic methods used for studying molecules in solution: ultraviolet, visible, and infrared absorption; fluorescence emission; circular dichroism; and optical rotatory dispersion. Chapter 14 is devoted to nuclear magnetic resonance (NMR); it discusses the fundamentals of the method for determining structures of proteins and nucleic acids. Chapter 15 starts with the scattering of electromagnetic radiation from one electron and proceeds through the diffraction of X-rays by crystals. Scanning microscope methods are introduced. The appendix contains numerical data used throughout the book, unit conversion tables, and the structures of many of the biological molecules mentioned in the text.

We encourage you to consult other books for background information and greater depth of coverage. Standard physical chemistry texts offer applications of physical chemistry to other areas. Biochemistry and molecular biology texts can provide specific information about such areas as protein and nucleic acid structures, enzyme mechanisms, and metabolic pathways. Finally, a good physics textbook is useful for learning or reviewing the fundamentals of forces, charges, electromagnetic fields, and energy. A list of such books is given at the end of this chapter.

In the following sections, we highlight several important biological problems that physical chemistry can address. These examples are meant to give you an overview of how physical chemistry is applied in the biological sciences. Read them for pleasure, without trying to memorize them. Our aim is simply to illustrate some current research from the scientific literature and to point out the principles and methods that are used. We hope to motivate you to learn the material discussed in the following chapters. Many articles in journals such as *Nature* or *Science* apply the methods and concepts described in this book. Read such articles to learn how the book will improve your understanding.

Neuroscience

Eric Kandel, a winner of the 2000 Nobel Prize in Physiology or Medicine, states that the last frontier of the biological sciences is to understand consciousness and the mental processes by which we experience our surroundings. An adult human brain has about 80 billion neurons, and each is connected to about 10,000 other neurons. The interactions and communication among the neurons is who we are. How this all works is left as an exercise to the reader because nobody else has explained it yet. However, in this book you will learn some of the ideas and methods that have been used up to now to begin to find answers, and more importantly, you may learn how to discover the next 'last frontier'.

Magnetic resonance imaging (MRI) uses nuclear magnetic resonance in the presence of a magnetic field gradient (the strength of the magnetic field varies with position) to produce an image. The resonance frequency of a nucleus, such as a proton in water, depends on the magnetic field strength and thus on its position in the sample. Therefore, the intensity at each frequency provides the distribution of water in the sample. For humans the image is analogous to an X-ray picture, but it is much more sensitive to soft tissue. A special application of MRI called functional MRI (fMRI) is used to learn which areas of the brain are most active when you see, hear, feel, smell, or think about different things. Brain activity uses more oxygen than usual, which means more blood flow and a higher concentration of oxyhemoglobin from the lungs. This produces a signal that is different from the slower blood flow with more deoxygenated hemoglobin in the rest of the brain. An fMRI image of a brain of a person listening to music will show activity in a region different from that of a person looking at a picture. These types of images are beginning to teach us which parts of the brain are involved in specific inputs, thoughts, and outputs.

The molecular basis of brain activity depends on the electrical and chemical interaction among the neurons. Neurons are specialized cells with a cell body containing the nucleus with its DNA, an axon that is much longer than the body, and many dendrites for communicating to other neurons. The communication between neurons is done at synapses, where the cells touch. Ion channels open, neurotransmitters are released by one cell and absorbed by the other, and somehow an image is seen, a thought is formed, an idea occurs. Reading this paragraph should trigger hundreds of questions in your mind. I can answer nearly all of them: "Nobody knows". The obvious questions include: What are the structures of all the proteins, nucleic acids, and small molecules involved? How do these molecules get to their sites of action, and how fast do they do it? What are the interactions that trigger and control all the effects? Finally, how do these molecular effects lead us to think, to remember, to dream?

The Human Genome and Beyond

The instructions for making all the molecules that occur in the brain and all the other organs in your body are stored in sequences of base pairs in your DNA. The structure of DNA, determined by X-ray diffraction to be two interwound strands, started the molecular understanding of how genes were stored and replicated (Watson and Crick, 1953). The tenth anniversary of the *Human Genome Project* was celebrated in 2011; it had determined the sequence of a human genome of 3 billion (3×10^9) base pairs (see *Science*, 2011 and Lander, 2011). Genes are sequences of base pairs in double-stranded DNA. In human sperm the DNA is packaged in 23 chromosomes: 22 autosomes plus a male Y chromosome or a female X chromosome. In human eggs the DNA is packaged in 22 autosomes plus a female X chromosome. Thus, each of us acquires 23 pairs of chromosomes; the XX pair makes us female, the XY male. When the human genome project started it was thought that about 100,000 genes coded human proteins. Now the number is estimated to be 20,000 to 25,000, not very different from fruit files. So what is going on? Most of us think we are smarter than a fruit fly.

A DNA sequence of base pairs does not directly code for a sequence of amino acids in a protein; it codes for a sequence of bases in RNA. Some of the RNAs are part of the translation machinery that produces proteins (ribosomal RNAs, transfer RNAs), and some are messenger RNAs that are translated into a sequence of protein amino acids. However, the messenger RNAs in humans are often modified before they are translated; pieces called introns are removed and the remaining exons are spliced together before the messenger RNA is translated. Alternative splicing can thus produce more than one protein from one DNA sequence—one gene. Furthermore, an increasing number of DNA sequences have been found to code for regulatory RNAs, not for proteins; they are RNA genes. These RNAs are not translated into proteins, but they control which proteins are made and when they are made. The 20,000 to 25,000 genes mentioned earlier are *protein-coding* genes; there may be an equal or larger number of *RNA-coding* genes that help define you. Interactions not yet discovered will eventually explain some of the differences between you and a fruit fly. The Human Genome Project is thus typical of nearly all scientific projects. The completion of one project reveals many new projects to investigate.

Determining the precise sequence of 3 billion nucleotides is a heroic task, which has been made possible by the application of many biophysical techniques to separate and characterize molecules. We emphasize the human genome, but the genomes of all kinds of organisms from bacteria to plants to extinct animals are being sequenced, and are revealing new insight into how organisms evolve, differentiate, and exist. The key method used to determine DNA sequence is to measure the fluorescence of fluorescently-labeled DNA bases. A different fluorophore is added to each of the four bases in DNA; this allows the sequence to be read using automated equipment. A newer method of sequencing uses an electric field to pull a single-stranded DNA (or RNA) through a narrow pore in a membrane (Pennisi, 2012). The four bases are different sizes and decrease the current passing through the pore by different amounts. A trace of current vs. time as the nucleic acid strand passes through the pore provides the sequence. The speed of sequencing has increased and the cost has decreased so much that soon we can all be sequenced routinely at birth. Also, DNA sequences have been placed on silicon chips and glass slides. Using the principles of Watson–Crick base pairing, scientists can rapidly identify changes in DNA sequences. These "genes on a chip" are revolutionizing the way that genes and gene expression are analyzed. Many times in science, fundamental advances are allowed by improvements in instrumentation to measure physical properties.

Once the sequence of an organism's genome has been determined, a difficult task begins. What does the string of the four different letters of the DNA alphabet (A, C, G, T) mean? The DNA sequence is first transcribed into the RNA alphabet (A, C, G, U). A messenger RNA is then translated into a protein sequence of 20 amino acids in a three-letter code.[*] As there are 4^3 (64) three-letter words with an alphabet of four letters, the code must be redundant. In the genetic dictionary, most amino acids are coded by two or four different words. Three amino acids have six words each (arginine, leucine, and serine), and two have only one word (tryptophan, methionine). Three of the words do not code for amino acids but instead signal for protein synthesis to stop: UAA, UAG, UGA. One word, AUG, codes for the start of protein synthesis (it also codes for methionine). Sequences before the starting AUG and after the terminating UAA, UAG, or UGA control and regulate the synthesis of the protein. Using the known genetic code, scientists can predict the sequence for a protein that is coded by a given gene. What does this protein do? What does it look like?

Physical chemistry provides the principles that allow bioinformatics scientists to make sense of the vast DNA sequence data of a genome. The protein sequence predicted from a gene is first compared to known protein sequences. If the protein is an essential part of some biochemical process that is common to many or all organisms, related proteins have likely been studied. Computer sequence comparisons establish the relationship between a novel protein and known proteins. The sequences of two proteins of similar biological function from different organisms are almost never identical. However, different protein sequences can adopt similar three-dimensional structures to perform similar functions. Different amino acids can have similar physical properties. For example, both isoleucine and valine have greasy aliphatic side chains and can often be exchanged for each other in a protein with little effect on its activity. Likewise, negatively charged amino acids (aspartic

[*]The structures and names of the nucleic acid bases and the protein amino acids are given in table A.9 in the appendix.

acid or glutamic acid) can often be swapped, and so on. Using this type of logic, computer programs can sometimes predict the function of the unknown gene by its relation to a known protein.

The weakness of this approach is obvious. You require the sequence of a known protein with which to compare the new gene. In addition, what truly determines the function of a protein is not the sequence of amino acids but rather how these amino acids fold into a three-dimensional structure that can perform a specific function—for example, catalysis of a reaction. Biophysical chemists can determine the three-dimensional structures of biological macromolecules, using methods described in this book. Unfortunately, the rate at which structures can be determined lags behind that of sequencing a gene. Nonetheless, comparisons of protein structures often reveal similarities that simple protein-sequence comparisons miss. Triosephosphate isomerase is a protein involved in metabolism, and it has a barrel-like three-dimensional structure. This structure is a rather common motif in proteins, but sequence comparisons by computer can rarely identify its presence.

Determining the three-dimensional structure of a protein would be easy if it could be predicted from its sequence. A protein's amino acid sequence contains the physical characteristics that determine the most stable three-dimensional fold. Biophysical chemists have shown that proteins almost always adopt the most stable three-dimensional structure as determined by the principles of thermodynamics. Thus, physical chemistry provides the framework to predict protein structure. However, predicting the most stable three-dimensional structure of a protein is a very difficult task because a large number of relatively weak interactions stabilize its structure (chapter 12). Precise treatment of these interactions is impossible, so biophysicists and computation biologists use a number of approximations to calculate a protein structure from its sequence (for an example, see Das & Baker, 2008). This is a valid approach to many complex biological problems. How can a scientist know whether a computer program is actually working? Well, she could try it on a sequence of a protein of known structure. But this is of course biased, for our scientist already knows the answer. Scientists in this field in fact resort to friendly competitions. They are asked to predict the structures of proteins whose structures are not known at the beginning of the competition but will be revealed by the end. This provides an unbiased test of various algorithms. This example shows a glimpse of the human side of the scientific process. Although current algorithms cannot predict the structures of protein to the same precision as experimental methods, they are improving. Computer prediction of protein folding and RNA folding is now a highly active area of biophysical research.

In addition to predicting structure and function of a protein from the sequence, you can try to improve the function. For example, will the catalytic activity increase or the specificity change if a crucial aspartic acid is changed to a glutamic acid? Changing one amino acid at a time is slow and tedious; but instead, by randomly changing the RNA sequence that codes for the protein, many mutants of the protein can be made. A selection process is then used to find the one with the desired optimized function. Furthermore, in producing better functions, or new functions, you need not be limited to naturally-occurring amino acids. The translation machinery can be tricked into producing proteins containing amino acids not found in nature (see Brustad and Arnold, 2011).

Transcription and Translation

Genetic information must be faithfully transmitted from DNA to messenger RNA to protein. Copying DNA to RNA is called *transcription;* reading RNA to produce a protein is called *translation*. Two central macromolecular machines are responsible for these processes: RNA polymerases transcribe RNA from DNA, and the ribosome translates RNA into protein. In both systems a series of repetitive tasks must be performed with high fidelity. These machines must be directional, because they copy information in only one direction. The machines are *processive*, in that once they start the process of transcription or translation, they continue through hundreds or even thousands of steps of the process. Finally, these biological processes are highly regulated. Associated factors determine when, where, and how rapidly these processes begin and end. Physical methods have provided important insights into how transcription and translation occur.

The process of transcription was first investigated in simple organisms such as bacteria. The protein that catalyzes transcription consists of only one or a few polypeptide chains. In contrast, in eukaryotic organisms such as humans, the RNA polymerase enzyme consists of ten or more polypeptide chains, reflecting the higher degree of regulation in higher organisms. Transcription begins at specific signals in the DNA called *promoters.* These DNA sequences bind specific *transcription factors* that enhance or prevent transcription. This is an essential feature in the regulation of gene expression. The activity of these transcription factors can be affected by attaching a phosphate group to a protein or by binding of a small molecule cofactor. The classic example is a protein that binds both to small sugars and to DNA, like the lac repressor (Bell and Lewis 2001). These *DNA-binding proteins* recognize specific promoter sequences, which control the expression of genes for sugar metabolism enzymes. When the lactose concentration reaches a certain level, the sugar binds to specific sites on the protein and changes its conformation, such that it binds tighter to its DNA site, thus turning off transcription of genes that would produce more sugars. This is an example of *feedback inhibition.* This example of biological regulation can be explained by the laws of chemical equilibrium and thermodynamics, discussed in chapters 2 through 5.

The high fidelity of transcription is ensured by an elegant kinetic mechanism, determined using the methods of enzyme kinetics described in chapter 10. During a round of polymerization, a nucleoside triphosphate enters the active site of RNA polymerase and pairs with the single-stranded DNA, which has been opened from its double helical form (figure 1.1). The three-dimensional structure of this essential enzyme from bacteria (*E. coli*, see Opalka et al., 2010) and higher organisms (yeast, see Cramer et al., 2000) has been solved. The shape of the active site is such that only the correct Watson–Crick base

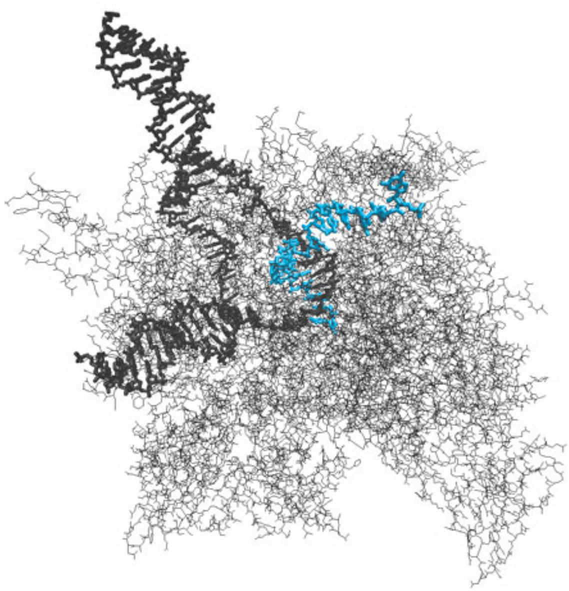


FIGURE 1.1 Three-dimensional structure of the RNA polymerase from *E. coli*, the enzyme responsible for transcribing the RNAs. The enzyme breaks base pairs in the double-stranded DNA (in black) to produce an open loop. The new RNA strand (in blue) is synthesized complementary to one of the DNA strands in the loop. This structure is a combination of X-ray diffraction and cryo-electron-microscopy data plus computer modeling (Opalka et al., 2010). The coordinates were obtained from the Protein Data Base. (Courtesy of Troy Lionberger, University of California Berkeley.)

pair is tolerated; the wrong nucleoside triphosphate does not make a good fit into the active site and is more rapidly ejected. For DNA polymerase, the enzyme that copies DNA during cell division, the push for fidelity is so strong that the enzyme contains an editing function. If a wrong nucleotide is incorporated into the DNA, it is snipped out, and the correct nucleotide is incorporated. This drive for fidelity is understandable, considering the drastic effects mutations can have on protein function. On the other hand, the polymerases have to perform their functions rapidly, so they have evolved a trade-off between high fidelity and reasonable rates of polymerization. Such trade-offs are a hallmark of biological chemistry.

The regulation of transcription is a central process in biology; the requirement for a complex macromolecular assembly to perform RNA transcription in higher organisms derives from the need for regulation. Cells must sense outside stimuli and respond, usually by rapidly synthesizing or degrading a protein or proteins. Recent biochemical experiments have revealed elaborate *signal transduction pathways.* A protein on the surface of a cell, called a *receptor,* will bind to an external signal, which may be a specific hormone or another extracellular-signaling molecule. The receptor molecule spans the cell membrane, and the binding of the hormone causes a change in its three-dimensional structure, activating an enzymatic activity (a *kinase*) that adds a phosphate group to a protein. When a phosphate group modifies a protein, the protein's shape and activity can change. Often, a cascade of kinase events occurs, where protein 1 phosphorylates protein 2, which, in turn, phosphorylates protein 3, and so on. The final targets of these cascades are often transcription factors which can turn transcription on or off depending on the desired result of a signaling event. Certain human cancers occur when these signaling pathways—and thus the ability of a cell to respond to external stimulus—are disrupted. Signaling pathways are very complex and biologists are still identifying their many components. Physical methods and reasoning, however, will be required to unravel the mechanisms of these signaling pathways.

The *ribosome* (figure 1.2), where translation occurs, is more complex than RNA polymerases. The ribosome in bacteria consists of two subunits which weigh 0.80×10^6 and 1.4×10^6 daltons. These enormous subunits each consist of at least one RNA chain and 20 to 30 proteins. The adaptors between the genetic code of RNA and the protein amino acid, first proposed by Crick, are called *transfer RNAs* (tRNAs). A single loop of the tRNA contains three nucleotides—the *anticodon*—that can form Watson–Crick base pairs with a given codon; the amino acid that corresponds to that codon is attached at the 3′-end of the tRNA. The three-dimensional structure of tRNA shows that these two parts are located 7.5 nm apart. The ribosome is able to select the correct tRNA that binds to the appropriate codon. The messenger RNA (mRNA) runs through a cleft between the subunits, and the anticodon portion of the tRNA interacts with the smaller (30S) subunit. Once the correct tRNA is selected at the A-site, the 3′-end of the tRNA sits within the larger subunit, where peptide bond formation is catalyzed between the amino acid and a peptide-chain containing tRNA (which is bound at the adjacent codon at the P-site). The ribosome then must move by three nucleotides in the mRNA to the next codon; this precise directional movement is called *translocation.*

The basic mechanism of translation was delineated over 40 years ago, but molecular details of how the ribosome performs the task of protein synthesis have been revealed only recently (reviewed in Schmeing and Ramakrishnan, 2009, and Moore, 2012). The structure of the ribosome (Ban et al., 2000; Schuwirth et al., 2005) showed that the biological functions of the ribosome are dominated by the RNA components; RNA catalyzes the formation of the peptide bond, making it an RNA enzyme—a ribozyme. Kinetic studies, similar to those done on polymerase enzymes, have revealed the origins of translational fidelity. The strategy used by the ribosome is somewhat similar to that used by polymerases. In the case of the ribosome, the base pairing between codon and anticodon occurs about

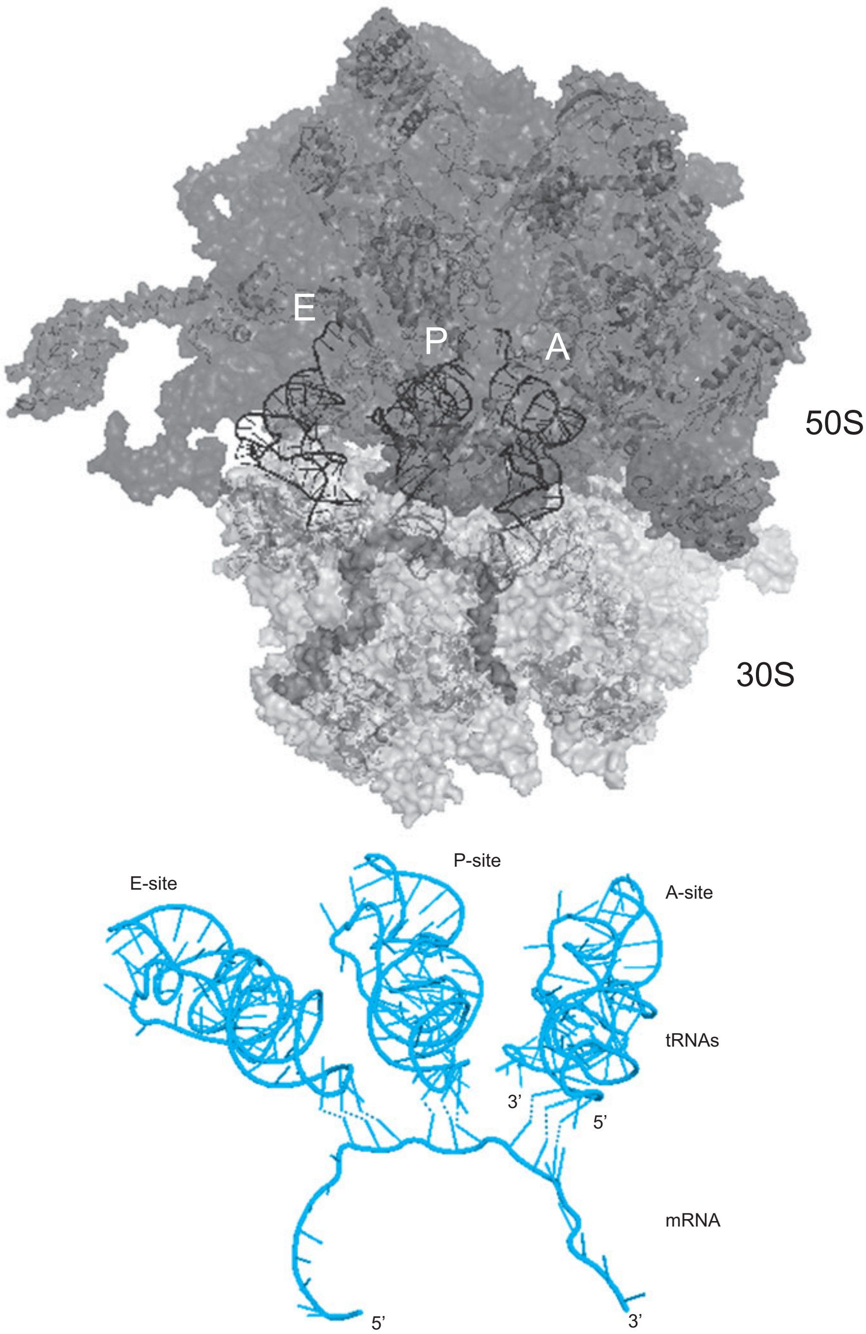


FIGURE 1.2 (top) The architecture of the ribosome; the large, 50S, subunit is on top and the small, 30S, subunit is on the bottom. Three transfer RNAs are shown reading the sequence in the messenger RNA. The structure was solved by X-ray crystallographic methods (Zhang et al., 2009). (bottom) A close-up view of the Watson-Crick base pairing between each codon on the messenger RNA with the anticodon of each transfer RNA as it occurs on the amino acid site (A-site), peptide site (P-site), and exit site (E-site) of the ribosome. The coordinates are from the Protein Data Base. (Courtesy of Shannon Yan, University of California Berkeley.)

7.5 nm from the site of peptide bond formation; the ribosome couples this base pairing to another chemical reaction: hydrolysis of guanosine triphosphate, GTP, which is bound to a protein factor that escorts the tRNA to the ribosome. Rate constants for tRNA dissociation and ribosomal conformational changes are modulated by whether the correct or the incorrect tRNA is present. Structural biologists have obtained detailed views of the ribosomal particles. The two subunits of the ribosome interact through an interface that is entirely RNA. Adjustments of this interface allow the ribosome to translocate down the mRNA. As biochemical experiments have predicted, the structures show that RNA forms the critical active sites for tRNA binding and peptidyl transfer. The RNA folds into a complex three-dimensional structure, which the protein components of the ribosome (many of which bind to ribosomal RNA) stabilize. The molecular rationale for how the ribosome performs translation will only be revealed by physical chemical investigations.

Ion Channels

Cells perform spectacular feats of chemistry. *Ion channels* are proteins that span the lipid membrane of a cell and specifically allow one ion type to traverse the channel. Ion channels are critical for many biological processes, including signaling by neurons. Ion channels can be remarkably selective. Potassium ion (K^+) channels are about 10,000-fold more selective for K^+ than for Na^+ (sodium) even though their ionic radii are 1.33 and 0.95 Å, respectively. Also, the ion channels must allow a large number of ions to pass across a membrane in a directional manner in a short time period. Finally, many ion channels are controlled by external conditions. They are opened or closed to ion passage by factors such as the voltage difference across the membrane. The methods of physical chemistry have been invaluable in determining how ion channels work (Doyle et al. 1998). When ion channels do not function properly, the results can be disastrous. Many human diseases are linked to impairment in these molecular highways. For example, cystic fibrosis, one of the most common genetic diseases, is caused by mutations in a Cl^- (chloride ion) channel. The disrupted function for this channel leads to a buildup of thick, fibrous mucus in the lungs, which impairs breathing.

Determining the three-dimensional structure of a K^+ channel to atomic resolution was a significant breakthrough in understanding how ion channels work. The protein is a tetramer of identical subunits. Long rods of alpha helix span the membrane. The protein is not merely a tube through which potassium flows. The overall shape of the protein is like a flower, with the petals opening toward the outside of the membrane and narrowing at the inside of the membrane (figure 1.3). The ions pass through a channel in the center of the tetrameric protein. How are potassium ions specifically selected and transported?

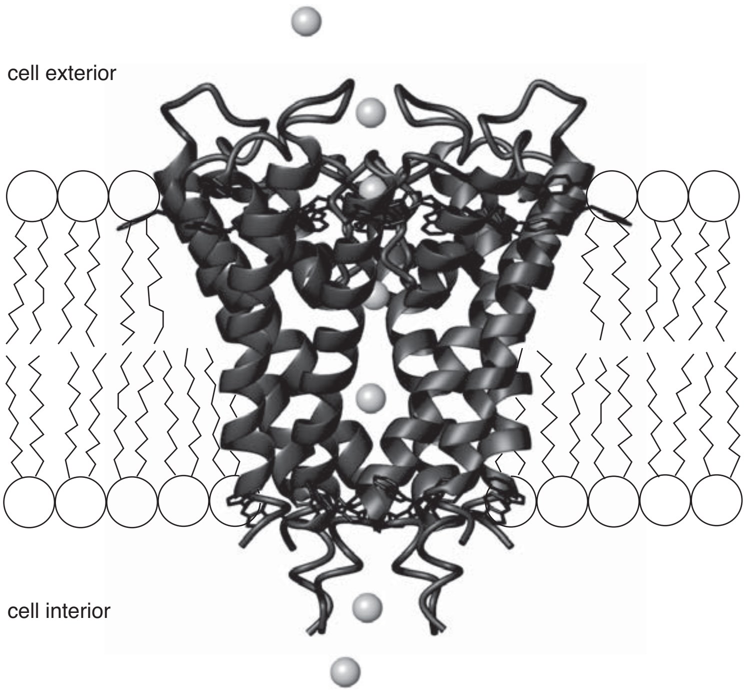


FIGURE 1.3 The three-dimensional structure of a K^+ channel, showing schematically the position of the channel within a cellular membrane. Intracellular (in the cell interior) and intercellular (on the cell surface) domains are indicated. Potassium ions are shown as spheres and are transported directionally from the exterior to the interior of the cell. (By permission of Roderick Mackinnon, MD; Professor, The Rockefeller University; Investigator, Howard Hughes Medical Institute.)