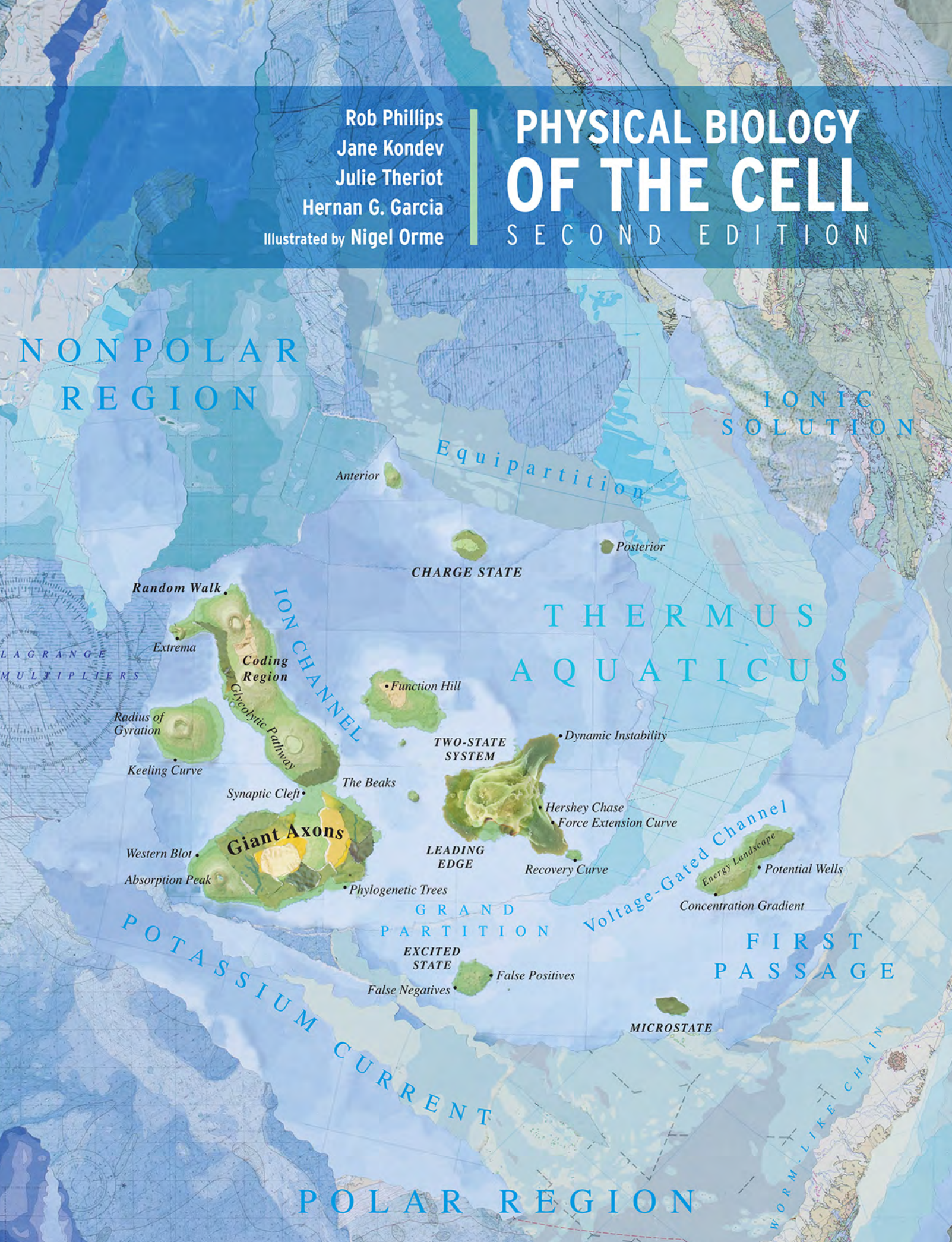
Rob Phillips
Jane Kondev
Julie Theriot
Hernan G. Garcia
Illustrated by Nigel Orme

PHYSICAL BIOLOGY OF THE CELL

SECOND EDITION



Physical Biology of the Cell

This page intentionally left blank
to match pagination of print book

Physical Biology of the Cell

Second Edition

Rob Phillips
Jane Kondev
Julie Theriot
Hernan G. Garcia

Illustrated by
Nigel Orme

Garland Science
Taylor & Francis Group
LONDON AND NEW YORK

Garland Science
Vice President: Denise Schanck
Editor: Summers Scholl
Senior Editorial Assistant: Kelly O'Connor
Cover design and illustrations: Nigel Orme
Production Editor: Natasha Wolfe
Copyeditor: Mac Clarke
Proofreader: Sally Huish
Typesetting: TechSet Composition India (P) Ltd.

Rob Phillips is the Fred and Nancy Morris Professor of Biophysics and Biology at the California Institute of Technology. He received a PhD in Physics from Washington University.

Jane Kondev is a Professor in the Department of Physics and within the Graduate Program in Quantitative Biology at Brandeis University. He attended the Mathematical High School in Belgrade, Serbia, received his Physics BS degree from the University of Belgrade, and his PhD from Cornell University.

Julie Theriot is a Professor in the Department of Biochemistry and the Department of Microbiology and Immunology at the Stanford University School of Medicine. She received concurrent BS degrees in Physics and Biology from the Massachusetts Institute of Technology, and a PhD in Cell Biology from the University of California at San Francisco.

Hernan G. Garcia is a Dicke Fellow in the Department of Physics at Princeton University. He received a BS in Physics from the University of Buenos Aires and a PhD in Physics from the California Institute of Technology.

Excerpt in Chapter 1
"On Exactitude in Science," from COLLECTED FICTIONS by Jorge Luis Borges, translated by Andrew Hurley, © 1998 by Maria Kodama; translation © 1998 by Penguin Putnam Inc. Used by permission of Viking Penguin, a division of Penguin Group (USA) Inc.

© 2013 by Garland Science, Taylor & Francis Group, LLC

This book contains information obtained from authentic and highly regarded sources. Every effort has been made to trace copyright holders and to obtain their permission for the use of copyright material. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems—without permission of the copyright holder.

ISBN 978-0-8153-4450-6

Library of Congress Cataloging-in-Publication Data

Phillips, Rob, 1960-
Physical biology of the cell. – Second edition / Rob Phillips, Jane Kondev, Julie Theriot, Hernan G. Garcia.
pages cm
ISBN 978-0-8153-4450-6 (pbk.)
1. Biophysics. 2. Cytology. I. Title.
QH505.P455 2013
571.6–dc23

2012030733

Published by Garland Science, Taylor & Francis Group, LLC, an informa business, 711 Third Avenue, New York, NY, 10017, USA, and 3 Park Square, Milton Park, Abingdon, OX14 4RN, UK.

Printed in the United States of America
15 14 13 12 11 10 9 8 7 6 5 4 3 2 1



Visit our web site at http://www.garlandscience.com

Dedicated to our friend Jon Widom



This page intentionally left blank
to match pagination of print book

Preface

"The map is not the territory."
Alfred Korzybski

The last 50 years in biology have seen an explosion of both data and understanding that rivals the fertile period between Tycho Brahe's definitive naked-eye investigations of the heavens and Newton's introduction of the "System of the World." One of the consequences of these stunning advances is the danger of becoming overwhelmed by the vast quantities of data coming at us from quarters ranging from next-generation sequencing to quantitative microscopy. For example, at the time of this writing, there are in excess of two million ribosomal RNA sequences deposited on publically accessible databases. But what does it all mean? A central role of scientific textbooks is to attempt to come to terms with broad areas of progress and to organize and distill the vast amounts of available information in a conceptually useful manner. In our view, an effective textbook can act as a map to help curious people discover unfamiliar territories. As with real maps, different purposes are served by different kinds of abstraction. Some maps show roads, some show topography, with each being useful in its own context.

A number of textbook writers have undertaken the formidable task of writing excellent, comprehensive surveys of cell and molecular biology, although each one of these books serves as a slightly different kind of map for the same overlapping territory. Although we cover some of the same material as a typical molecular and cell biology book, our goal in this book is fundamentally different. There is no single, correct way to construct a conceptually simplified map for a huge and complex field such as cell and molecular biology. Most modern biology textbooks organize ideas, facts, and experimental data based on their conceptual proximity for some particular biological function. In contrast, this book examines the same set of information from the distinct perspective of *physical biology*. We have therefore adopted an organization in which the proximity of topics is based on the physical concepts that unite a given set of biological phenomena, instead of the cell biology perspective. By analogy to a map of the United States, a cell biology textbook might describe the plains of Eastern Colorado in the same chapter as the mountains of Western Colorado, whereas our physical biology book would group Eastern Colorado with the rolling fields of Iowa, and Western Colorado with mountainous West Virginia.

This book does not assume extensive prior knowledge on the part of the reader, though a grounding in both calculus and elementary physics is essential. The material covered here is appropriate for a first course in physical biology or biophysics for either undergraduates or graduate students. It is also intended for any scientist interested in learning the basic principles and applications of physical modeling for research in biology, and aims to provide a novel perspective even to scientists who are already familiar with some of the material. Throughout the book, our organization of ideas and data based on proximity in physical biology space juxtaposes topics that are not obviously related in cell biology space. For example, DNA

wrapping around nucleosomes in the eukaryotic nucleus, DNA looping induced by the binding of transcriptional repressors in the context of bacterial gene regulation, and DNA packing into the narrow confines of bacteriophage capsids all appear in the same chapter because they are related by the mechanical rules governing the bending of DNA. Next, the physical and mathematical treatment we derive for DNA bending is directly applied to other kinds of long, thin, biological structures, including the filaments of the cytoskeleton. This organizational principle brings into focus the central thesis of this book, namely, that the appropriate application of a relatively small number of fundamental physical models can serve as the foundation of whole bodies of quantitative biological intuition, broadly useful across a wide range of apparently unrelated biological problems.

During the 12-year journey that led to this book, we benefited immeasurably from the generosity and enthusiasm of hundreds of scientific colleagues who graciously shared their data, ideas, and perspectives. Indeed, in many respects, we view our book as an exercise in *quantitative journalism*, based upon extensive "interviews" with these various scientists in a wide range of disciplines. We offer this book as a report from the front, to share some of the most interesting things that we have learned from our colleagues with any and all inquiring individuals who wish to think both deeply and broadly about the connections between biology and the physical sciences. Our imagined audience spans the range from 18-year-old mechanical engineering undergraduates curious about the application of their discipline to medicine, to 40-year-old string theorists wishing to apply their mathematical and physical talents to living matter, to 70-year-old renowned biologists wondering whether their insights into living systems might be improved by a mathematical treatment.

Although the claim that a handful of simple physical models can shed more than superficial light on complex biological processes might seem naive, the biological research literature is teeming with examples where important quantitative insight into questions of pressing interest has been gained by the application of such models. In every chapter, we have chosen specific examples from classic and current research papers where quantitative measurements on biological systems can be largely understood by recourse to simple, fundamental, physical ideas. In cases where the simplest possible physical models fail to fit the data, the specific quantitative nature of the disparities can often lead to testable new biological hypotheses. For example, a simple calculation estimating the amount of time it would take for a newly synthesized protein to diffuse from the cell body of a motor neuron in the spinal cord to the synapse formed by the same neuron in the foot proves that diffusion is far too slow to get the job done, and an active transport process must occur. Inevitably, researchers performing experiments on biological systems must have physical models explicitly or implicitly in mind, whether imagining how changes in the rate of transcription initiation for a particular gene will lead to changes in the overall amount of the gene product in the cell, or picturing the ways that signaling molecules move through cellular space to encounter their targets, or envisioning how cell movements during embryogenesis lead to the final three-dimensional structures of organs and limbs. In this book, we aim to provide a physical and mathematical toolkit so that people used to thinking deeply about biological problems can make this kind of quantitative intuition explicit; we also hope to provide a perspective on biology that may inspire people from a background more heavily based in physics or

mathematics to seek out new biological problems that are particularly appropriate for this kind of quantitative analysis.

Our general approach follows four steps. First, we introduce a biological phenomenon; second, we perform simple order-of-magnitude estimates to develop a "feeling for the numbers" involved in that process; third, we demonstrate the application of an extremely simple first-pass model; and finally, where possible, we present a refinement of the oversimplified model to better approximate biological reality. Our goal is to share the pleasure in seeing the extent to which simple models can be tailored to reveal the *complexity* of observed phenomena. For our examples, we have chosen particular biological cases that we believe to be worthy illustrations of the concepts at hand and that have captured our imaginations, often because of particularly elegant or clever experiments that were designed to generate intriguing sets of quantitative data. While we have been conscientious in our exploration of these facts and in our construction of simple models, it is inevitable that we will have made errors due to our ignorance and also due to the fact that, in many cases, new discoveries may change the particulars of our case studies. (A list of errors and their corrections will be posted on the book's website as well as the website of one of the authors (R.P.).) Nevertheless, because our goal is to demonstrate the power of applying simple models to complex systems, even when some details are wrong or missing, we hope that any particular lapses will not obscure the overall message. Furthermore, in many cases, we have described phenomena that are still awaiting a satisfying physical model. We hope that many of our readers will seize upon the holes and errors in our exploration of physical biology and take these as challenges and opportunities for launching exciting original work.

Our second edition builds upon the foundations laid in the previous edition, with the addition of two new chapters that focus on central themes of modern biology, namely, light and life and the emergence of patterns in living organisms. The new Chapter 18 focuses on several key ways in which light is central in biology. We begin with an analysis of photosynthesis that illustrates the quantum mechanical underpinnings of both the absorption of light and the transfer of energy and electrons through the photosynthetic apparatus. The second part of our story in that chapter considers the rich and beautiful subject of vision. The new Chapter 20 uses insights garnered throughout the book to ask how it is that organisms ranging from flies to plants can build up such exquisite patterns. Here we explore Turing's famed model of several interacting chemical species undergoing chemical reactions and diffusion and other more recent advances in thinking about problems ranging from somitogenesis to phyllotaxis.

The book is made up of four major parts. Part 1, *The Facts of Life*, largely focuses on introducing biological phenomena. For biology readers already familiar with this material, the hope is that the quantitative spin will be enlightening. For physics readers, the goal is to get a sense of the biological systems themselves. Part 2, *Life at Rest*, explores those problems in biology that can be attacked using quantitative models without any explicit reference to time. Part 3, *Life in Motion*, tackles head-on the enhanced complexity of time-dependent systems exhibiting dynamic behavior. Finally, Part 4, *The Meaning of Life*, addresses various kinds of information processing by biological systems.

Because our hope is that you, our readers, represent a broad diversity of backgrounds and interests, throughout the book we try as much as possible to introduce the origin of the facts and principles

that we exploit. We are reluctant to ever simply assert biological "facts" or physical "results," and would not expect you to blindly accept our assertions if we did. Therefore, we often describe classical observations by biologists over the past centuries as well as the most recent exciting results, and illustrate how current thinking about complex biological problems has been shaped by a progression of observations and insights. Extended discussions of this kind are separated from the main text in sections labeled *Experiments Behind the Facts*. In a complementary way, whenever we find it necessary to derive mathematical equations, we proceed step by step through the derivation and explain how each line leads to the next, so that readers lacking a strong background in mathematics can nevertheless follow every step of the logic and not be forced to take our word for any result. Specific sections labeled *The Math Behind the Models* and *The Tricks Behind the Math* provide summaries for the mathematical techniques that are used repeatedly throughout the book; many readers trained in physics will already be familiar with this material, but biologists may benefit from a brief refresher or introduction. In addition, we include sections labeled *Estimate* that help to develop a "feeling for the numbers" for particularly interesting cases.

Another critical new element in our second edition is a feature called *Computational Exploration*. The idea of these excursions is to show how simple computer analyses can help us attack problems that are otherwise inaccessible. In the first edition, we underemphasized "computation" because we wanted to combat the spurious idea that theory in biology is synonymous with computation. While we made this exaggeration to make a point, we did so at a price, because computation is not only useful, but downright indispensable in some problems. Further, one of the beauties of turning a model into a specific numerical computation is that to get a computer to produce a meaningful number, nothing can be left unspecified. The Computational Explorations are offered as a way for the reader to develop a particular habit of mind, and none of them should be viewed as illustrating the state of the art for making such calculations. Matlab and Mathematica code related to most of these explorations is provided on the book's website.

Although we review the basic information necessary to follow the exposition of each topic, you may also find it useful to have recourse to a textbook or reference book covering the details of scientific areas among biology, physics, chemistry, and mathematics, with which you consider yourself less familiar. Some references that are among our favorites in these fields are suggested at the end of each chapter. More generally, our references to the literature are treated in two distinct ways. Our suggestions for *Further Reading* reflect our own tastes. Often, the choices that appear at a chapter's end are chosen because of uniqueness of viewpoint or presentation. We make no attempt at completeness. The second class of *References* reflect work that has explicitly touched the content of each chapter, either through introducing us to a model, providing a figure, or constructing an argument.

At the end of each chapter, we include a series of problems that expand the material in the chapter or give the opportunity to attempt model-building for other case studies. In the second edition, we have considerably expanded the scope of the end-of-chapter problems. These problems can be used within formal courses or by individual readers. A complete *Solutions Manual*, covering all problems in the book, is available for instructors. There are several different types

of problems. Some, whose goal is to develop a "feeling for the numbers," are arithmetically simple, and primarily intended to develop a sense of order-of-magnitude biology. Others request difficult mathematical derivations that we could not include in the text. Still others, perhaps our favorites, invite the readers to apply quantitative model-building to provocative experimental data from the primary research literature. In each chapter, we have loosely identified the different problems with the aforementioned categories in order to assist the reader in choosing which one to attack depending on particular need. The book's website also includes *Hints for the Reader* for some of the more difficult problems.

Our book relies heavily on original data, both in the figures that appear throughout the book and in the various end-of-chapter problems. To make these data easily accessible to interested readers, the book's website includes the original experimental data used to make all the figures in the book that are based upon published measurements. Similarly, the data associated with the end-of-chapter problems are also provided on the book's website. It is our hope that you will use these data in order to perform your own calculations for fitting the many models introduced throughout the book to the relevant primary data, and perhaps refining the models in your own original work.

Student and Instructor Resources

Figures and PowerPoint® Presentations

The figures from the book are available in two convenient formats: PowerPoint and JPEG. There is one PowerPoint presentation for each chapter, and the JPEGs have been optimized for display on a computer.

Data Sets

The original data used to create both the figures and homework problems are available in Excel® spreadsheets. With this data, the reader can extend the theoretical tools developed in the book to fit experimental data for a wide range of problems. The data files contain explicit statement of all relevant units, and include references to the original sources.

Hints for Problems

This PDF provides both hints and strategies for attacking some of the more difficult end-of-chapter problems. In some cases, the hints provide intuition about how to set up the problem; in other cases, the hints provide explicit mathematical instructions on how to carry through more tricky manipulations.

Matlab® and Mathematica® Code

These files contain code for the Computational Explorations sidebars located throughout the book.

Movies

The movies complement the figures and discussion from the book by illustrating the rich dynamics exhibited by living organisms and the molecules that make them tick.

Solutions Manual

This PDF contains solutions to all problems in the book. It is available only to qualified instructors.

With the exception of the *Solutions Manual*, these resources are available on the *Physical Biology of the Cell*, 2nd Edition, media website:

http://microsite.garlandscience.com/pboc2

Access to the Solutions Manual is available to qualified instructors by emailing science@garland.com.

PowerPoint and Excel are registered trademarks of Microsoft Corporation in the United States and/or other countries.
MATLAB® is a trademark of The MathWorks, Inc.
Mathematica® is a trademark of Wolfram Research, Inc.

Acknowledgments

This book would not have been possible without a wide range of support from both people and institutions. We are grateful for the support of the Aspen Center for Physics, the Kavli Institute for Theoretical Physics at the University of California, Santa Barbara and the ESPCI in Paris, where some of the writing was done. Our funding during the course of this project was provided by the National Science Foundation, the National Institutes of Health, The Research Corporation, the Howard Hughes Medical Institute, and the MacArthur Foundation. We also particularly acknowledge the support of the NIH Director's Pioneer Award and La Fondation Pierre Gilles de Gennes granted to R.P and The Donna and Benjamin M. Rosen Center for Bioengineering at Caltech, all of which provided broad financial support for many facets of this project.

Our book would never have achieved its present incarnation without the close and expert collaboration of our gifted illustrator, Nigel Orme, who is responsible for the clarity and visual appeal of the hundreds of figures found in these pages, as well as the overall design of the book. We also had the pleasure of working with David Goodsell who produced many illustrations throughout the book showing detailed molecular structures. Genya Frenkel also provided assistance on the problems and their solutions. Amy Phillips assisted with editing, responding to reader comments, and obtaining permission for use of many of the previously published images in the figures. Maureen Storey (first edition) and Mac Clarke (second edition) improved our clarity and respectability with their expert copy editing. Our editors Mike Morales (first edition) and Summers Scholl (second edition) have offered great support through the entirety of the project. Simon Hill (first edition) and Natasha Wolfe's (second edition) expert assistance in the production process has been an impressive pleasure.

One of the most pleasurable parts of our experience of writing this book has been our interaction with generous friends and colleagues who have shared their insights, stories, prejudices, and likes and dislikes about biology, physics, chemistry, and their overlap. We are deeply grateful to all our colleagues who have contributed ideas directly or indirectly through these many enjoyable conversations over the past twelve years. Elio Schaechter told us the secret to maintaining a happy collaboration. Lubert Stryer inspired the overall section organization and section titles, and gave us much-needed practical advice on how to actually finish the book. Numerous others have helped us directly or indirectly through inspiration, extended lab visits, teaching us about whole fields, or just by influential interactions along the way. It is very important to note that in some cases these people explicitly disagreed with some of our particular conclusions, and deserve no blame for our mistakes and misjudgments. We specifically wish to thank: Gary Ackers, Bruce Alberts, Olaf Andersen, David Baltimore, Robert Bao, David Bensimon, Seymour Benzer, Howard Berg, Paul Berg, Maja Bialecka, Bill Bialek, Lacra Bintu, Pamela Bjorkman, Steve Block, Seth Blumberg, David Boal, James Boedicker, Rob Brewster, Robijn Bruinsma, Zev Bryant, Steve Burden, Carlos Bustamante, Anders Carlsson, Sherwood Casjens, Yi-Ju Chen, Kristina Dakos, Eric Davidson, Scott Delp, Micah Dembo, Michael Dickinson, Ken Dill, Marileen Dogterom, David Dunlap, Michael Elowitz, Evan Evans, Stan Falkow, Julio Fernandez, Jim Ferrell, Laura Finzi, Daniel Fisher, Dan Fletcher, Henrik Flyvbjerg, Seth Fraden, Scott Fraser, Ben Freund, Andrew J. Galambos, Ethan Garner, Bill Gelbart, Jeff Gelles, Kings Ghosh, Dan Gillespie, Yale Goldman, Bruce Goode, Paul Grayson, Thomas Gregor, Jim Haber, Mike Hagan, Randy Hampton, Lin Han, Pehr Harbury, Dan Herschlag, John Heuser, Joe Howard, KC Huang, Terry Hwa, Grant Jensen, Jack Johnson, Daniel Jones, Jason Kahn, Dale Kaiser, Suzanne Amador Kane, Sarah Keller, Doro Kern, Karla Kirkegaard, Marc Kirschner, Bill Klug, Chuck Knobler, Tolya Kolomeisky, Corinne Ladous, Jared Leadbetter, Heun Jin Lee, Henry Lester, Julian Lewis, Jennifer Lippincott-Schwartz, Sanjoy Mahajan, Jim Maher, Carmen Mannella, William Martin, Bob Meyer, Elliot Meyerowitz, Chris Miller, Ken Miller, Tim Mitchison, Alex Mogilner, Cathy Morris, Dyche Mullins, Richard Murray, Kees Murre, David Nelson, James Nelson, Phil Nelson, Keir Neuman, Dianne Newman, Lene Oddershede, Garry Odell, George Oster, Adrian Parsegian, Iva Perovic, Eduardo Perozo, Eric Peterson, Suzanne

Pfeffer, Tom Pollard, Dan Portnoy, Tom Powers, Ashok Prasad, Mark Ptashne, Prashant Purohit, Steve Quake, Sharad Ramanathan, Samuel Rauhala, Michael Reddy, Doug Rees, Dan Reeves, Joy Rimchala, Ellen Rothenberg, Michael Roukes, Dave Rutledge, Peter Sarnow, Klaus Schulten, Bob Schleif, Darren Segall, Udo Seifert, Paul Selvin, Lucy Shapiro, Boris Shraiman, Steve Small, Doug Smith, Steve Smith, Andy Spakowitz, Jim Spudich, Alasdair Steven, Sergei Sukharev, Christian Sulloway, Joel Swanson, Boo Shan Tseng, Tristan Ursell, Ron Vale, David Van Valen, Elizabeth Villa, Zhen-Gang Wang, Clare Waterman, Annemarie Weber, Jon Widom, Eric Wieschaus, Paul Wiggins, Ned Wingreen, Zeba Wunderlich, Ahmed Zewail, and Kai Zinn.

Finally, we are deeply grateful to the individuals who have given us critical feedback on the manuscript in its various stages, including the many students in our courses offered at Caltech, Brandeis, and Stanford over the last twelve years. They have all done their best to save us from error and any remaining mistakes are entirely our responsibility. We are indebted to all of them for their generosity with their time and expertise. A few hardy individuals read the entire first edition: Laila Ashegian, Andre Brown, Genya Frenkel, Steve Privitera, Alvaro Sanchez, and Sylvain Zorman. We thank them for their many insightful comments and remarkable stamina. For the second edition, we had the great fortune to have Howard Berg read every word of our book always providing pointed and thoughtful commentary. Similarly, Ron Milo has been a constant source of critical commentary, and encouragement throughout the process. Velocity Hughes and Madhav Mani were a tremendous help in reading the entire book in its near final form and providing critical comments at every turn. Justin Bois has also been a source of numerous critical insights. Niles Pierce has also provided his unflagging support throughout this project.

Many more people have given expert commentary on specific chapters, provided specific figures, advised on us end-of- chapter problems, or provided particular insights for either the first or second edition:

Chapter 1

Bill Gelbart (University of California, Los Angeles), Shura Grosberg (New York University), Randy Hampton (University of California, San Diego), Sanjoy Mahajan (Olin College), Ron Milo (Weizmann Institute of Science), Michael Rubinstein (University of North Carolina, Chapel Hill).

Chapter 2

John A. G. Briggs (European Molecular Biology Laboratory), James Boedicker (California Institute of Technology), James Brody (University of California, Irvine), Titus Brown (Michigan State University), Ian Chin-Sang (Queen's University), Avigdor Eldar (Tel Aviv University), Scott Fraser (California Institute of Technology), CT Lim (National University of Singapore), Dianne Newman (California Institute of Technology), Yitzhak Rabin (Bar-Ilan University), Manfred Radmacher (University of Bremen), Michael Rubinstein (University of North Carolina, Chapel Hill), Steve Small (New York University), Linda Song (Harvard University), Dave Tirrell (California Institute of Technology), Jon Widom (Northwestern University).

Chapter 3

Tom Cech (University of Colorado), Andreas Matouschek (Northwestern University), Yitzhak Rabin (Bar-Ilan University), Michael Reddy (University of Wisconsin, Milwaukee), Nitzan Rosenfeld (Rosetta Genomics), Michael Rubinstein (University of North Carolina, Chapel Hill), Antoine van Oijen (Rijksuniversiteit Groningen), Jon Widom (Northwestern University).

Chapter 4

Elaine Bearer (Brown University), Paul Jardine (University of Minnesota, Twin Cities), Michael Reddy (University of Wisconsin, Milwaukee), Michael Rubinstein (University of North Carolina, Chapel Hill).

Chapter 5

James Boedicker (California Institute of Technology), Ken Dill (Stony Brook University), Randy Hampton (University of California, San Diego), Rick James (University of Minnesota, Twin Cities), Heun Jin Lee (California Institute of Technology), Bill Klug (University of California, Los Angeles), Steve Quake (Stanford University), Elio Schaechter (San Diego State University).

Chapter 6

Ken Dill (Stony Brook University), Dan Herschlag (Stanford University), Terry Hwa (University of California, San Diego), Arbel Tadmor (California Institute of Technology).

Chapter 7

Gary Ackers (Washington University in St. Louis), Olaf Andersen (Cornell University), Ken Dill (Stony Brook University), Henry Lester (California Institute of Technology).

Chapter 8

Ken Dill (Stony Brook University), Shura Grosberg (New York University), Michael Rubinstein (University of North Carolina, Chapel Hill), Jeremy Schmit (Kansas State University), Andy Spakowitz (Stanford University), Paul Wiggins (University of Washington).

Chapter 9

Mike Hagan (Brandeis University), Thomas Record (University of Wisconsin, Madison), Bob Schleif (Johns Hopkins University), Pete von Hippel (University of Oregon).

Chapter 10

Zev Bryant (Stanford University), Carlos Bustamante (University of California, Berkeley), Hans-Günther Döbereiner (University of Bremen), Paul Forscher (Yale University), Ben Freund (Brown University), Bill Gelbart (University of California, Los Angeles), Paul Grayson (California Institute of Technology), Mandar Inamdar (Indian Institute of Technology, Bombay), Bill Klug (University of California, Los Angeles), Joy Rimchala (Massachusetts Institute of Technology), Doug Smith (University of California, San Diego), Megan Valentine (University of California, Santa Barbara), Jon Widom (Northwestern University), Paul Wiggins (University of Washington).

Chapter 11

Ashustosh Agrawal (University of Houston), Patricia Bassereau (Institut Curie), Hans-Günther Döbereiner (University of Bremen), Evan Evans (University of British Columbia), Dan Fletcher (University of California, Berkeley), Terry Frey (San Diego State University), Christoph Haselwandter (University of Southern California), KC Huang (Stanford University), Sarah Keller (University of Washington, Seattle), Bill Klug (University of California, Los Angeles), Carmen Mannella (State University of New York, Albany), Eva Schmid (University of California, Berkeley), Pierre Sens (ESPCI, Paris), Sergei Sukharev (University of Maryland, College Park), Tristan Ursell (Stanford University), Paul Wiggins (University of Washington).

Chapter 12

Howard Berg (Harvard University), Justin Bois (University of California, Los Angeles), Zev Bryant (Stanford University), Ray Goldstein (University of Cambridge), Jean-François Joanny (Institut Curie), Sanjoy Mahajan (Olin College), Tom Powers (Brown University), Todd Squires (University of California, Santa Barbara), Howard Stone (Harvard University).

Chapter 13

Howard Berg (Harvard University), Ariane Briegel (California Institute of Technology), Dan Gillespie, Jean-François Joanny (Institut Curie), Martin Linden (Stockholm University), Jennifer Lippincott-Schwartz (National Institutes of Health), Ralf Metzler (Technical University of Munich), Frosso Seitaridou (Emory University), Pierre Sens (ESPCI, Paris), Dave Wu (California Institute of Technology).

Chapter 14

Jean-François Joanny (Institut Curie), Randy Kamien (University of Pennsylvania), Martin Linden (Stockholm University), Ralf Metzler (Technical University of Munich), Pierre Sens (ESPCI, Paris), Arbel Tadmor (California Institute of Technology).

Chapter 15

Anders Carlsson (Washington University in St. Louis), Marileen Dogterom (Institute for Atomic and Molecular Physics), Dan Fletcher (University of California, Berkeley), Dan Herschlag (Stanford University), Jean-François Joanny (Institut Curie), Tom Pollard (Yale University), Dimitrios Vavylonis (Lehigh University).

Chapter 16

Bill Gelbart (University of California, Los Angeles), Jean-François Joanny (Institut Curie), Tolya Kolomeisky (Rice University), Martin Linden (Stockholm University), Jens Michaelis (Ludwig-Maximilians University), George Oster (University of California, Berkeley), Megan Valentine (University of California, Santa Barbara), Jianhua Xing (Virginia Polytechnic Institute and State University).

Chapter 17

Olaf Andersen (Cornell University), Chris Gandhi (California Institute of Technology), Jean-François Joanny (Institut Curie), Stephanie Johnson (California Institute of Technology), Rod MacKinnon (Rockefeller University), Chris Miller (Brandeis University), Paul Miller (Brandeis University), Phil Nelson (University of Pennsylvania).

Chapter 18

Maja Bialecka-Fornal (California Institute of Technology), Bill Bialek (Princeton University), David Chandler (University of California, Berkeley), Anna Damjanovic (Johns Hopkins University), Govindjee (University of Illinois, Urbana-Champaign), Harry Gray (California Institute of Technology), Heun Jin Lee (California Institute of Technology), Rudy Marcus (California Institute of Technology), Tom Miller (California Institute of Technology), Ron Milo (Weizmann Institute of Science), Jose Onuchic (Rice University), Nipam Patel (University of California, Berkeley), Mark Ratner (Northwestern University), Mattias Rydenfelt (California Institute of Technology), Dave Savage (University of California, Berkeley), Klaus Schulten (University of Illinois, Urbana-Champaign), Kurt Warncke (Emory University), Jay Winkler (California Institute of Technology).

Chapter 19

James Boedicker (California Institute of Technology), Robert Brewster (California Institute of Technology), Titus Brown (Michigan State University), Nick Buchler (Duke University), Eric Davidson (California Institute of Technology), Avigdor Eldar (Tel Aviv University), Michael Elowitz (California Institute of Technology), Robert Endres (Imperial College London), Daniel Fisher (Stanford University), Scott Fraser (California Institute of Technology), Uli Gerland (Ludwig-Maximilians University), Ido Golding (Baylor College of Medicine), Mikko Haataja (Princeton University), Terry Hwa (University of California, San Diego), Daniel Jones (California Institute of Technology), Tom Kuhlman (University of Illinois, Urbana-Champaign), Wendell Lim (University of California, San Francisco), Chris Myers (Cornell University), Bob Schleif (Johns Hopkins University), Vivek Shenoy (Brown University), Steve Small (New York University), Peter Swain (McGill University), David Van Valen (California Institute of Technology), Ned Wingreen (Princeton University), Sunney Xie (Harvard University).

Chapter 20

Justin Bois (University of California, Los Angeles), Thomas Gregor (Princeton University), KC Huang (Stanford University), Frank Julicher (Max Planck Institute of Complex Systems, Dresden), Karsten Kruse (University of Saarlandes), Andy Oates (Max Planck Institute of Molecular Cell Biology and Genetics, Dresden), Jordi Garcia Ojalvo (Polytechnic University of Catalonia), George Oster (University of California, Berkeley), Andrew Rutenberg (Dalhousie University), David Sprinzak (Tel Aviv University), Carolina Tropini (Stanford University).

Chapter 21

Ralf Bundschuh (The Ohio State University), Uli Gerland (Ludwig-Maximilians University), Daniel Jones (California Institute of Technology), Justin Kinney (Cold Spring Harbor Laboratory), Chris Myers (Cornell University), Eric Peterson (California Institute of Technology), Frank Pugh (Pennsylvania State University), Jody Puglisi (Stanford University), Oliver Rando (University of Massachusetts Medical School), Tony Redondo (Los Alamos National Laboratory), Eran Segal (Weizmann Institute of Science), Boris Shraiman (University of California, Santa Barbara) Peter Swain (University of Edinburgh), Jon Widom (Northwestern University), Chris Wiggins (Columbia University).

Contents

Preface *vii*
Acknowledgments *xiii*
Special Sections *xxix*
Map of the Maps *xxx*

PART 1 THE FACTS OF LIFE

Chapter 1: Why: Biology by the Numbers 3
Chapter 2: What and Where: Construction Plans for Cells and Organisms 35
Chapter 3: When: Stopwatches at Many Scales 87
Chapter 4: Who: "Bless the Little Beasties" 137

PART 2 LIFE AT REST

Chapter 5: Mechanical and Chemical Equilibrium in the Living Cell 187
Chapter 6: Entropy Rules! 237
Chapter 7: Two-State Systems: From Ion Channels to Cooperative Binding 281
Chapter 8: Random Walks and the Structure of Macromolecules 311
Chapter 9: Electrostatics for Salty Solutions 355
Chapter 10: Beam Theory: Architecture for Cells and Skeletons 383
Chapter 11: Biological Membranes: Life in Two Dimensions 427

PART 3 LIFE IN MOTION

Chapter 12: The Mathematics of Water 483
Chapter 13: A Statistical View of Biological Dynamics 509
Chapter 14: Life in Crowded and Disordered Environments 543
Chapter 15: Rate Equations and Dynamics in the Cell 573
Chapter 16: Dynamics of Molecular Motors 623
Chapter 17: Biological Electricity and the Hodgkin–Huxley Model 681
Chapter 18: Light and Life 717

PART 4 THE MEANING OF LIFE

Chapter 19:	Organization of Biological Networks	801
Chapter 20:	Biological Patterns: Order in Space and Time	893
Chapter 21:	Sequences, Specificity, and Evolution	951
Chapter 22:	Whither Physical Biology?	1023
Index		1039

Contents in Detail

Preface *vii*
Acknowledgments *xiii*
Special Sections *xxix*
Map of the Maps *xxx*

PART 1 THE FACTS OF LIFE 1

Chapter 1 Why: Biology by the Numbers 3

1.1 BIOLOGICAL CARTOGRAPHY 3
1.2 PHYSICAL BIOLOGY OF THE CELL 4
Model Building Requires a Substrate of Biological Facts and Physical (or Chemical) Principles 5
1.3 THE STUFF OF LIFE 5
Organisms Are Constructed from Four Great Classes of Macromolecules 6
Nucleic Acids and Proteins Are Polymer Languages with Different Alphabets 7
1.4 MODEL BUILDING IN BIOLOGY 9
1.4.1 Models as Idealizations 9
Biological Stuff Can Be Idealized Using Many Different Physical Models 11
1.4.2 Cartoons and Models 16
Biological Cartoons Select Those Features of the Problem Thought to Be Essential 16
Quantitative Models Can Be Built by Mathematicizing the Cartoons 19
1.5 QUANTITATIVE MODELS AND THE POWER OF IDEALIZATION 20
1.5.1 On the Springiness of Stuff 21
1.5.2 The Toolbox of Fundamental Physical Models 22
1.5.3 The Unifying Ideas of Biology 23
1.5.4 Mathematical Toolkit 25
1.5.5 The Role of Estimates 26
1.5.6 On Being Wrong 29
1.5.7 Rules of Thumb: Biology by the Numbers 30
1.6 SUMMARY AND CONCLUSIONS 32
1.7 FURTHER READING 32
1.8 REFERENCES 33

Chapter 2 What and Where: Construction Plans for Cells and Organisms 35

2.1 AN ODE TO *E. COLI* 35
2.1.1 The Bacterial Standard Ruler 37
The Bacterium *E. coli* Will Serve as Our Standard Ruler 37
2.1.2 Taking the Molecular Census 38
The Cellular Interior Is Highly Crowded, with Mean Spacings Between Molecules That Are Comparable to Molecular Dimensions 48
2.1.3 Looking Inside Cells 49
2.1.4 Where Does *E. coli* Fit? 51
Biological Structures Exist Over a Huge Range of Scales 51
2.2 CELLS AND STRUCTURES WITHIN THEM 52
2.2.1 Cells: A Rogue's Gallery 52
Cells Come in a Wide Variety of Shapes and Sizes and with a Huge Range of Functions 52
Cells from Humans Have a Huge Diversity of Structure and Function 57
2.2.2 The Cellular Interior: Organelles 59
2.2.3 Macromolecular Assemblies: The Whole is Greater than the Sum of the Parts 63
Macromolecules Come Together to Form Assemblies 63
Helical Motifs Are Seen Repeatedly in Molecular Assemblies 64
Macromolecular Assemblies Are Arranged in Superstructures 65
2.2.4 Viruses as Assemblies 66
2.2.5 The Molecular Architecture of Cells: From Protein Data Bank (PDB) Files to Ribbon Diagrams 69
Macromolecular Structure Is Characterized Fundamentally by Atomic Coordinates 69
Chemical Groups Allow Us to Classify Parts of the Structure of Macromolecules 70
2.3 TELESCOPING UP IN SCALE: CELLS DON'T GO IT ALONE 72
2.3.1 Multicellularity as One of Evolution's Great Inventions 73
Bacteria Interact to Form Colonies such as Biofilms 73
Teaming Up in a Crisis: Lifestyle of *Dictyostelium discoideum* 75
Multicellular Organisms Have Many Distinct Communities of Cells 76
2.3.2 Cellular Structures from Tissues to Nerve Networks 77
One Class of Multicellular Structures is the Epithelial Sheets 77
Tissues Are Collections of Cells and Extracellular Matrix 77
Nerve Cells Form Complex, Multicellular Complexes 78
2.3.3 Multicellular Organisms 78
Cells Differentiate During Development Leading to Entire Organisms 78
The Cells of the Nematode Worm, *Caenorhabditis Elegans*, Have Been Charted, Yielding a Cell-by-Cell Picture of the Organism 80
Higher-Level Structures Exist as Colonies of Organisms 82
2.4 SUMMARY AND CONCLUSIONS 83
2.5 PROBLEMS 83
2.6 FURTHER READING 84
2.7 REFERENCES 85

Chapter 3 When: Stopwatches at Many Scales 87

3.1 THE HIERARCHY OF TEMPORAL SCALES 87
3.1.1 The Pageant of Biological Processes 89
Biological Processes Are Characterized by a Huge Diversity of Time Scales 89
3.1.2 The Evolutionary Stopwatch 95

3.1.3 The Cell Cycle and the Standard Clock 99
The *E. coli* Cell Cycle Will Serve as Our Standard Stopwatch 99
3.1.4 Three Views of Time in Biology 105

3.2 PROCEDURAL TIME 106
3.2.1 The Machines (or Processes) of the Central Dogma 107
The Central Dogma Describes the Processes Whereby the Genetic Information Is Expressed Chemically 107
The Processes of the Central Dogma Are Carried Out by Sophisticated Molecular Machines 108
3.2.2 Clocks and Oscillators 110
Developing Embryos Divide on a Regular Schedule Dictated by an Internal Clock 111
Diurnal Clocks Allow Cells and Organisms to Be on Time Everyday 111

3.3 RELATIVE TIME 114
3.3.1 Checkpoints and the Cell Cycle 115
The Eukaryotic Cell Cycle Consists of Four Phases Involving Molecular Synthesis and Organization 115
3.3.2 Measuring Relative Time 117
Genetic Networks Are Collections of Genes Whose Expression Is Interrelated 117
The Formation of the Bacterial Flagellum Is Intricately Organized in Space and Time 119
3.3.3 Killing the Cell: The Life Cycles of Viruses 120
Viral Life Cycles Include a Series of Self-Assembly Processes 121
3.3.4 The Process of Development 122

3.4 MANIPULATED TIME 125
3.4.1 Chemical Kinetics and Enzyme Turnover 125
3.4.2 Beating the Diffusive Speed Limit 126
Diffusion Is the Random Motion of Microscopic Particles in Solution 127
Diffusion Times Depend upon the Length Scale 127
Diffusive Transport at the Synaptic Junction Is the Dynamical Mechanism for Neuronal Communication 128
Molecular Motors Move Cargo over Large Distances in a Directed Way 129
Membrane-Bound Proteins Transport Molecules from One Side of a Membrane to the Other 130
3.4.3 Beating the Replication Limit 131
3.4.4 Eggs and Spores: Planning for the Next Generation 132

3.5 SUMMARY AND CONCLUSIONS 133
3.6 PROBLEMS 133
3.7 FURTHER READING 136
3.8 REFERENCES 136

Chapter 4 Who: "Bless the Little Beasties" 137

4.1 CHOOSING A GRAIN OF SAND 137
Modern Genetics Began with the Use of Peas as a Model System 138
4.1.1 Biochemistry and Genetics 138

4.2 HEMOGLOBIN AS A MODEL PROTEIN 143
4.2.1 Hemoglobin, Receptor–Ligand Binding, and the Other Bohr 143
The Binding of Oxygen to Hemoglobin Has Served as a Model System for Ligand–Receptor Interactions More Generally 143
Quantitative Analysis of Hemoglobin Is Based upon Measuring the Fractional Occupancy of the Oxygen-Binding Sites as a Function of Oxygen Pressure 144
4.2.2 Hemoglobin and the Origins of Structural Biology 144
The Study of the Mass of Hemoglobin Was Central in the Development of Centrifugation 145
Structural Biology Has Its Roots in the Determination of the Structure of Hemoglobin 145
4.2.3 Hemoglobin and Molecular Models of Disease 146
4.2.4 The Rise of Allostery and Cooperativity 146

4.3 BACTERIOPHAGES AND MOLECULAR BIOLOGY 147
4.3.1 Bacteriophages and the Origins of Molecular Biology 148
Bacteriophages Have Sometimes Been Called the "Hydrogen Atoms of Biology" 148
Experiments on Phages and Their Bacterial Hosts Demonstrated That Natural Selection Is Operative in Microscopic Organisms 148
The Hershey–Chase Experiment Both Confirmed the Nature of Genetic Material and Elucidated One of the Mechanisms of Viral DNA Entry into Cells 149
Experiments on Phage T4 Demonstrated the Sequence Hypothesis of Collinearity of DNA and Proteins 150
The Triplet Nature of the Genetic Code and DNA Sequencing Were Carried Out on Phage Systems 150
Phages Were Instrumental in Elucidating the Existence of mRNA 151
General Ideas about Gene Regulation Were Learned from the Study of Viruses as a Model System 152
4.3.2 Bacteriophages and Modern Biophysics 153
Many Single- Molecule Studies of Molecular Motors Have Been Performed on Motors from Bacteriophages 154

4.4 A TALE OF TWO CELLS: *E. COLI* AS A MODEL SYSTEM 154
4.4.1 Bacteria and Molecular Biology 154
4.4.2 *E. coli* and the Central Dogma 156
The Hypothesis of Conservative Replication Has Falsifiable Consequences 156
Extracts from *E. coli* Were Used to Perform *In Vitro* Synthesis of DNA, mRNA, and Proteins 157
4.4.3 The *lac* Operon as the "Hydrogen Atom" of Genetic Circuits 157
Gene Regulation in *E. coli* Serves as a Model for Genetic Circuits in General 157
The *lac* Operon Is a Genetic Network That Controls the Production of the Enzymes Responsible for Digesting the Sugar Lactose 158
4.4.4 Signaling and Motility: The Case of Bacterial Chemotaxis 159
E. coli Has Served as a Model System for the Analysis of Cell Motility 159

4.5 YEAST: FROM BIOCHEMISTRY TO THE CELL CYCLE 161
Yeast Has Served as a Model System Leading to Insights in Contexts Ranging from Vitalism to the Functioning of Enzymes to Eukaryotic Gene Regulation 161
4.5.1 Yeast and the Rise of Biochemistry 162
4.5.2 Dissecting the Cell Cycle 162
4.5.3 Deciding Which Way Is Up: Yeast and Polarity 164
4.5.4 Dissecting Membrane Traffic 166
4.5.5 Genomics and Proteomics 167

4.6 FLIES AND MODERN BIOLOGY 170
4.6.1 Flies and the Rise of Modern Genetics 170
Drosophila melanogaster Has Served as a Model System for Studies Ranging from Genetics to Development to the Functioning of the Brain and Even Behavior 170
4.6.2 How the Fly Got His Stripes 171

4.7 OF MICE AND MEN 173
4.8 THE CASE FOR EXOTICA 174
4.8.1 Specialists and Experts 174
4.8.2 The Squid Giant Axon and Biological Electricity 175
There Is a Steady-State Potential Difference Across the Membrane of Nerve Cells 176
Nerve Cells Propagate Electrical Signals and Use Them to Communicate with Each Other 176
4.8.3 Exotica Toolkit 178

4.9 SUMMARY AND CONCLUSIONS 179
4.10 PROBLEMS 179
4.11 FURTHER READING 181
4.12 REFERENCES 183

PART 2 LIFE AT REST 185

Chapter 5 Mechanical and Chemical Equilibrium in the Living Cell 187

5.1 ENERGY AND THE LIFE OF CELLS 187
5.1.1 The Interplay of Deterministic and Thermal Forces 189
Thermal Jostling of Particles Must Be Accounted for in Biological Systems 189
5.1.2 Constructing the Cell: Managing the Mass and Energy Budget of the Cell 190

5.2 BIOLOGICAL SYSTEMS AS MINIMIZERS 200
5.2.1 Equilibrium Models for Out of Equilibrium Systems 200
Equilibrium Models Can Be Used for Nonequilibrium Problems if Certain Processes Happen Much Faster Than Others 201
5.2.2 Proteins in "Equilibrium" 202
Protein Structures are Free-Energy Minimizers 203
5.2.3 Cells in "Equilibrium" 204
5.2.4 Mechanical Equilibrium from a Minimization Perspective 204
The Mechanical Equilibrium State is Obtained by Minimizing the Potential Energy 204

5.3 THE MATHEMATICS OF SUPERLATIVES 209
5.3.1 The Mathematization of Judgement: Functions and Functionals 209
Functionals Deliver a Number for Every Function They Are Given 210
5.3.2 The Calculus of Superlatives 211
Finding the Maximum and Minimum Values of a Function Requires That We Find Where the Slope of the Function Equals Zero 211

5.4 CONFIGURATIONAL ENERGY 214
In Mechanical Problems, Potential Energy Determines the Equilibrium Structure 214
5.4.1 Hooke's Law: Actin to Lipids 216
There is a Linear Relation Between Force and Extension of a Beam 216
The Energy to Deform an Elastic Material is a Quadratic Function of the Strain 217

5.5 STRUCTURES AS FREE-ENERGY MINIMIZERS 219
The Entropy is a Measure of the Microscopic Degeneracy of a Macroscopic State 219
5.5.1 Entropy and Hydrophobicity 222
Hydrophobicity Results from Depriving Water Molecules of Some of Their Configurational Entropy 222
Amino Acids Can Be Classified According to Their Hydrophobicity 224
When in Water, Hydrocarbon Tails on Lipids Have an Entropy Cost 225
5.5.2 Gibbs and the Calculus of Equilibrium 225
Thermal and Chemical Equilibrium are Obtained by Maximizing the Entropy 225
5.5.3 Departure from Equilibrium and Fluxes 227
5.5.4 Structure as a Competition 228
Free Energy Minimization Can Be Thought of as an Alternative Formulation of Entropy Maximization 228
5.5.5 An Ode to ΔG 230
The Free Energy Reflects a Competition Between Energy and Entropy 230

5.6 SUMMARY AND CONCLUSIONS 231
5.7 APPENDIX: THE EULER–LAGRANGE EQUATIONS, FINDING THE SUPERLATIVE 232
Finding the Extrema of Functionals Is Carried Out Using the Calculus of Variations 232
The Euler–Lagrange Equations Let Us Minimize Functionals by Solving Differential Equations 232

5.8 PROBLEMS 233
5.9 FURTHER READING 235
5.10 REFERENCES 236

Chapter 6 Entropy Rules! 237

6.1 THE ANALYTICAL ENGINE OF STATISTICAL MECHANICS 237
The Probability of Different Microstates Is Determined by Their Energy 240
6.1.1 A First Look at Ligand–Receptor Binding 241
6.1.2 The Statistical Mechanics of Gene Expression: RNA Polymerase and the Promoter 244
A Simple Model of Gene Expression Is to Consider the Probability of RNA Polymerase Binding at the Promoter 245
Most Cellular RNA Polymerase Molecules Are Bound to DNA 245
The Binding Probability of RNA Polymerase to Its Promoter Is a Simple Function of the Number of Polymerase Molecules and the Binding Energy 247
6.1.3 Classic Derivation of the Boltzmann Distribution 248
The Boltzmann Distribution Gives the Probability of Microstates for a System in Contact with a Thermal Reservoir 248
6.1.4 Boltzmann Distribution by Counting 250
Different Ways of Partitioning Energy Among Particles Have Different Degeneracies 250
6.1.5 Boltzmann Distribution by Guessing 253
Maximizing the Entropy Corresponds to Making a Best Guess When Faced with Limited Information 253
Entropy Maximization Can Be Used as a Tool for Statistical Inference 255
The Boltzmann Distribution is the Maximum Entropy Distribution in Which the Average Energy is Prescribed as a Constraint 258

6.2 ON BEING IDEAL 259
6.2.1 Average Energy of a Molecule in a Gas 259
The Ideal Gas Entropy Reflects the Freedom to Rearrange Molecular Positions and Velocities 259
6.2.2 Free Energy of Dilute Solutions 262
The Chemical Potential of a Dilute Solution Is a Simple Logarithmic Function of the Concentration 262
6.2.3 Osmotic Pressure as an Entropic Spring 264
Osmotic Pressure Arises from Entropic Effects 264
Viruses, Membrane-Bound Organelles, and Cells Are Subject to Osmotic Pressure 265
Osmotic Forces Have Been Used to Measure the Interstrand Interactions of DNA 266

6.3 THE CALCULUS OF EQUILIBRIUM APPLIED: LAW OF MASS ACTION 267
6.3.1 Law of Mass Action and Equilibrium Constants 267
Equilibrium Constants are Determined by Entropy Maximization 267

6.4 APPLICATIONS OF THE CALCULUS OF EQUILIBRIUM 270
6.4.1 A Second Look at Ligand–Receptor Binding 270
6.4.2 Measuring Ligand–Receptor Binding 272
6.4.3 Beyond Simple Ligand–Receptor Binding: The Hill Function 273
6.4.4 ATP Power 274
The Energy Released in ATP Hydrolysis Depends Upon the Concentrations of Reactants and Products 275

6.5 SUMMARY AND CONCLUSIONS 276
6.6 PROBLEMS 276
6.7 FURTHER READING 278
6.8 REFERENCES 278

Chapter 7 Two-State Systems: From Ion Channels to Cooperative Binding 281

7.1 MACROMOLECULES WITH MULTIPLE STATES 281
7.1.1 The Internal State Variable Idea 281
The State of a Protein or Nucleic Acid Can Be Characterized Mathematically Using a State Variable 282
7.1.2 Ion Channels as an Example of Internal State Variables 286
The Open Probability $\langle\sigma\rangle$ of an Ion Channel Can Be Computed Using Statistical Mechanics 287

7.2 STATE VARIABLE DESCRIPTION OF BINDING 289
7.2.1 The Gibbs Distribution: Contact with a Particle Reservoir 289
The Gibbs Distribution Gives the Probability of Microstates for a System in Contact with a Thermal and Particle Reservoir 289
7.2.2 Simple Ligand–Receptor Binding Revisited 291
7.2.3 Phosphorylation as an Example of Two Internal State Variables 292
Phosphorylation Can Change the Energy Balance Between Active and Inactive States 293
Two-Component Systems Exemplify the Use of Phosphorylation in Signal Transduction 295
7.2.4 Hemoglobin as a Case Study in Cooperativity 298
The Binding Affinity of Oxygen for Hemoglobin Depends upon Whether or Not Other Oxygens Are Already Bound 298
A Toy Model of a Dimeric Hemoglobin (Dimoglobin) Illustrate the Idea of Cooperativity 298
The Monod–Wyman–Changeux (MWC) Model Provides a Simple Example of Cooperative Binding 300
Statistical Models of the Occupancy of Hemoglobin Can Be Written Using Occupation Variables 301
There is a Logical Progression of Increasingly Complex Binding Models for Hemoglobin 301

7.3 ION CHANNELS REVISITED: LIGAND-GATED CHANNELS AND THE MWC MODEL 305
7.4 SUMMARY AND CONCLUSIONS 308
7.5 PROBLEMS 308
7.6 FURTHER READING 310
7.7 REFERENCES 310

Chapter 8 Random Walks and the Structure of Macromolecules 311

8.1 WHAT IS A STRUCTURE: PDB OR R_G? 311
8.1.1 Deterministic versus Statistical Descriptions of Structure 312
PDB Files Reflect a Deterministic Description of Macromolecular Structure 312
Statistical Descriptions of Structure Emphasize Average Size and Shape Rather Than Atomic Coordinates 312

8.2 MACROMOLECULES AS RANDOM WALKS 312
Random Walk Models of Macromolecules View Them as Rigid Segments Connected by Hinges 312
8.2.1 A Mathematical Stupor 313
In Random Walk Models of Polymers, Every Macromolecular Configuration Is Equally Probable 313
The Mean Size of a Random Walk Macromolecule Scales as the Square Root of the Number of Segments, $\sqrt{N}$ 314
The Probability of a Given Macromolecular State Depends Upon Its Microscopic Degeneracy 315
Entropy Determines the Elastic Properties of Polymer Chains 316
The Persistence Length Is a Measure of the Length Scale Over Which a Polymer Remains Roughly Straight 319
8.2.2 How Big Is a Genome? 321
8.2.3 The Geography of Chromosomes 322
Genetic Maps and Physical Maps of Chromosomes Describe Different Aspects of Chromosome Structure 322
Different Structural Models of Chromatin Are Characterized by the Linear Packing Density of DNA 323
Spatial Organization of Chromosomes Shows Elements of Both Randomness and Order 324
Chromosomes Are Tethered at Different Locations 325
Chromosome Territories Have Been Observed in Bacterial Cells 327
Chromosome Territories in *Vibrio cholerae* Can Be Explored Using Models of Polymer Confinement and Tethering 328
8.2.4 DNA Looping: From Chromosomes to Gene Regulation 333
The Lac Repressor Molecule Acts Mechanistically by Forming a Sequestered Loop in DNA 334
Looping of Large DNA Fragments Is Dictated by the Difficulty of Distant Ends Finding Each Other 334
Chromosome Conformation Capture Reveals the Geometry of Packing of Entire Genomes in Cells 336

8.3 THE NEW WORLD OF SINGLE-MOLECULE MECHANICS 337
Single-Molecule Measurement Techniques Lead to Force Spectroscopy 337
8.3.1 Force–Extension Curves: A New Spectroscopy 339
Different Macromolecules Have Different Force Signatures When Subjected to Loading 339
8.3.2 Random Walk Models for Force–Extension Curves 340
The Low-Force Regime in Force–Extension Curves Can Be Understood Using the Random Walk Model 340

8.4 PROTEINS AS RANDOM WALKS 344
8.4.1 Compact Random Walks and the Size of Proteins 345
The Compact Nature of Proteins Leads to an Estimate of Their Size 345
8.4.2 Hydrophobic and Polar Residues: The HP Model 346
The HP Model Divides Amino Acids into Two Classes: Hydrophobic and Polar 346
8.4.3 HP Models of Protein Folding 348

8.5 SUMMARY AND CONCLUSIONS 351
8.6 PROBLEMS 351
8.7 FURTHER READING 353
8.8 REFERENCES 353

Chapter 9 Electrostatics for Salty Solutions 355

9.1 WATER AS LIFE'S AETHER 355

9.2 THE CHEMISTRY OF WATER 358
9.2.1 pH and the Equilibrium Constant 358
Dissociation of Water Molecules Reflects a Competition Between the Energetics of Binding and the Entropy of Charge Liberation 358
9.2.2 The Charge on DNA and Proteins 359
The Charge State of Biopolymers Depends upon the pH of the Solution 359
Different Amino Acids Have Different Charge States 359
9.2.3 Salt and Binding 360

9.3 ELECTROSTATICS FOR SALTY SOLUTIONS 360
9.3.1 An Electrostatics Primer 361
A Charge Distribution Produces an Electric Field Throughout Space 362
The Flux of the Electric Field Measures the Density of Electric Field Lines 363
The Electrostatic Potential Is an Alternative Basis for Describing the Electrical State of a System 364
There Is an Energy Cost Associated With Assembling a Collection of Charges 367
The Energy to Liberate Ions from Molecules Can Be Comparable to the Thermal Energy 368
9.3.2 The Charged Life of a Protein 369
9.3.3 The Notion of Screening: Electrostatics in Salty Solutions 370
Ions in Solution Are Spatially Arranged to Shield Charged Molecules Such as DNA 370
The Size of the Screening Cloud Is Determined by a Balance of Energy and Entropy of the Surrounding Ions 371
9.3.4 The Poisson–Boltzmann Equation 374
The Distribution of Screening Ions Can Be Found by Minimizing the Free Energy 374
The Screening Charge Decays Exponentially Around Macromolecules in Solution 376
9.3.5 Viruses as Charged Spheres 377

9.4 SUMMARY AND CONCLUSION 379
9.5 PROBLEMS 380
9.6 FURTHER READING 382
9.7 REFERENCES 382

Chapter 10 Beam Theory: Architecture for Cells and Skeletons 383

10.1 BEAMS ARE EVERYWHERE: FROM FLAGELLA TO THE CYTOSKELETON 383
One-Dimensional Structural Elements Are the Basis of Much of Macromolecular and Cellular Architecture 383

10.2 GEOMETRY AND ENERGETICS OF BEAM DEFORMATION 385
10.2.1 Stretch, Bend, and Twist 385
Beam Deformations Result in Stretching, Bending, and Twisting 385
A Bent Beam Can Be Analyzed as a Collection of Stretched Beams 385
The Energy Cost to Deform a Beam Is a Quadratic Function of the Strain 387
10.2.2 Beam Theory and the Persistence Length: Stiffness is Relative 389
Thermal Fluctuations Tend to Randomize the Orientation of Biological Polymers 389
The Persistence Length Is the Length Over Which a Polymer Is Roughly Rigid 390
The Persistence Length Characterizes the Correlations in the Tangent Vectors at Different Positions Along the Polymer 390
The Persistence Length Is Obtained by Averaging Over All Configurations of the Polymer 391
10.2.3 Elasticity and Entropy: The Worm-Like Chain 392
The Worm-Like Chain Model Accounts for Both the Elastic Energy and Entropy of Polymer Chains 392

10.3 THE MECHANICS OF TRANSCRIPTIONAL REGULATION: DNA LOOPING REDUX 394
10.3.1 The *lac* Operon and Other Looping Systems 394
Transcriptional Regulation Can Be Effected by DNA Looping 395
10.3.2 Energetics of DNA Looping 395
10.3.3 Putting It All Together: The J-Factor 396

10.4 DNA PACKING: FROM VIRUSES TO EUKARYOTES 398
The Packing of DNA in Viruses and Cells Requires Enormous Volume Compaction 398
10.4.1 The Problem of Viral DNA Packing 400
Structural Biologists Have Determined the Structure of Many Parts in the Viral Parts List 400
The Packing of DNA in Viruses Results in a Free-Energy Penalty 402
A Simple Model of DNA Packing in Viruses Uses the Elastic Energy of Circular Hoops 403
DNA Self-Interactions Are also Important in Establishing the Free Energy Associated with DNA Packing in Viruses 404
DNA Packing in Viruses Is a Competition Between Elastic and Interaction Energies 406
10.4.2 Constructing the Nucleosome 407
Nucleosome Formation Involves Both Elastic Deformation and Interactions Between Histones and DNA 408
10.4.3 Equilibrium Accessibility of Nucleosomal DNA 409
The Equilibrium Accessibility of Sites within the Nucleosome Depends upon How Far They Are from the Unwrapped Ends 409

10.5 THE CYTOSKELETON AND BEAM THEORY 413
Eukaryotic Cells Are Threaded by Networks of Filaments 413
10.5.1 The Cellular Interior: A Structural Perspective 414
Prokaryotic Cells Have Proteins Analogous to the Eukaryotic Cytoskeleton 416
10.5.2 Stiffness of Cytoskeletal Filaments 416
The Cytoskeleton Can Be Viewed as a Collection of Elastic Beams 416
10.5.3 Cytoskeletal Buckling 419
A Beam Subject to a Large Enough Force Will Buckle 419
10.5.4 Estimate of the Buckling Force 420
Beam Buckling Occurs at Smaller Forces for Longer Beams 420

10.6 SUMMARY AND CONCLUSIONS 421
10.7 APPENDIX: THE MATHEMATICS OF THE WORM-LIKE CHAIN 421
10.8 PROBLEMS 424
10.9 FURTHER READING 426
10.10 REFERENCES 426

Chapter 11 Biological Membranes: Life in Two Dimensions 427

11.1 THE NATURE OF BIOLOGICAL MEMBRANES 427
11.1.1 Cells and Membranes 427
Cells and Their Organelles Are Bound by Complex Membranes 427
Electron Microscopy Provides a Window on Cellular Membrane Structures 429
11.1.2 The Chemistry and Shape of Lipids 431
Membranes Are Built from a Variety of Molecules That Have an Ambivalent Relationship with Water 431
The Shapes of Lipid Molecules Can Induce Spontaneous Curvature on Membranes 436
11.1.3 The Liveliness of Membranes 436
Membrane Proteins Shuttle Mass Across Membranes 437
Membrane Proteins Communicate Information Across Membranes 439
Specialized Membrane Proteins Generate ATP 439
Membrane Proteins Can Be Reconstituted in Vesicles 439

11.2 ON THE SPRINGINESS OF MEMBRANES 440
11.2.1 An Interlude on Membrane Geometry 440
Membrane Stretching Geometry Can Be Described by a Simple Area Function 441
Membrane Bending Geometry Can Be Described by a Simple Height Function, $h(x, y)$ 441

Membrane Compression Geometry Can Be Described by a Simple Thickness Function, $w(x,y)$ 444
Membrane Shearing Can Be Described by an Angle Variable, θ 444
11.2.2 Free Energy of Membrane Deformation 445
There Is a Free-Energy Penalty Associated with Changing the Area of a Lipid Bilayer 445
There Is a Free-Energy Penalty Associated with Bending a Lipid Bilayer 446
There Is a Free-Energy Penalty for Changing the Thickness of a Lipid Bilayer 446
There Is an Energy Cost Associated with the Gaussian Curvature 447

11.3 STRUCTURE, ENERGETICS, AND FUNCTION OF VESICLES 448
11.3.1 Measuring Membrane Stiffness 448
Membrane Elastic Properties Can Be Measured by Stretching Vesicles 448
11.3.2 Membrane Pulling 450
11.3.3 Vesicles in Cells 453
Vesicles Are Used for a Variety of Cellular Transport Processes 453
There Is a Fixed Free-Energy Cost Associated with Spherical Vesicles of All Sizes 455
Vesicle Formation Is Assisted by Budding Proteins 456
There Is an Energy Cost to Disassemble Coated Vesicles 458

11.4 FUSION AND FISSION 458
11.4.1 Pinching Vesicles: The Story of Dynamin 459

11.5 MEMBRANES AND SHAPE 462
11.5.1 The Shapes of Organelles 462
The Surface Area of Membranes Due to Pleating Is So Large That Organelles Can Have Far More Area than the Plasma Membrane 463
11.5.2 The Shapes of Cells 465
The Equilibrium Shapes of Red Blood Cells Can Be Found by Minimizing the Free Energy 466

11.6 THE ACTIVE MEMBRANE 467
11.6.1 Mechanosensitive Ion Channels and Membrane Elasticity 467
Mechanosensitive Ion Channels Respond to Membrane Tension 467
11.6.2 Elastic Deformations of Membranes Produced by Proteins 468
Proteins Induce Elastic Deformations in the Surrounding Membrane 468
Protein-Induced Membrane Bending Has an Associated Free-Energy Cost 469
11.6.3 One-Dimensional Solution for MscL 470
Membrane Deformations Can Be Obtained by Minimizing the Membrane Free Energy 470
The Membrane Surrounding a Channel Protein Produces a Line Tension 472

11.7 SUMMARY AND CONCLUSIONS 475
11.8 PROBLEMS 476
11.9 FURTHER READING 479
11.10 REFERENCES 479

PART 3 LIFE IN MOTION 481

Chapter 12 The Mathematics of Water 483

12.1 PUTTING WATER IN ITS PLACE 483
12.2 HYDRODYNAMICS OF WATER AND OTHER FLUIDS 484
12.2.1 Water as a Continuum 484
Though Fluids Are Composed of Molecules It Is Possible to Treat Them as a Continuous Medium 484
12.2.2 What Can Newton Tell Us? 485
Gradients in Fluid Velocity Lead to Shear Forces 485
12.2.3 $F = ma$ for Fluids 486
12.2.4 The Newtonian Fluid and the Navier–Stokes Equations 490
The Velocity of Fluids at Surfaces Is Zero 491

12.3 THE RIVER WITHIN: FLUID DYNAMICS OF BLOOD 491
12.3.1 Boats in the River: Leukocyte Rolling and Adhesion 493

12.4 THE LOW REYNOLDS NUMBER WORLD 495
12.4.1 Stokes Flow: Consider a Spherical Bacterium 495
12.4.2 Stokes Drag in Single-Molecule Experiments 498
Stokes Drag Is Irrelevant for Optical Tweezers Experiments 498
12.4.3 Dissipative Time Scales and the Reynolds Number 499
12.4.4 Fish Gotta Swim, Birds Gotta Fly, and Bacteria Gotta Swim Too 500
Reciprocal Deformation of the Swimmer's Body Does Not Lead to Net Motion at Low Reynolds Number 502
12.4.5 Centrifugation and Sedimentation: Spin It Down 502

12.5 SUMMARY AND CONCLUSIONS 504
12.6 PROBLEMS 505
12.7 FURTHER READING 507
12.8 REFERENCES 507

Chapter 13 A Statistical View of Biological Dynamics 509

13.1 DIFFUSION IN THE CELL 509
13.1.1 Active versus Passive Transport 510
13.1.2 Biological Distances Measured in Diffusion Times 511
The Time It Takes a Diffusing Molecule to Travel a Distance L Grows as the Square of the Distance 512
Diffusion Is Not Effective Over Large Cellular Distances 512
13.1.3 Random Walk Redux 514

13.2 CONCENTRATION FIELDS AND DIFFUSIVE DYNAMICS 515
Fick's Law Tells Us How Mass Transport Currents Arise as a Result of Concentration Gradients 517
The Diffusion Equation Results from Fick's Law and Conservation of Mass 518
13.2.1 Diffusion by Summing Over Microtrajectories 518
13.2.2 Solutions and Properties of the Diffusion Equation 524
Concentration Profiles Broaden Over Time in a Very Precise Way 524
13.2.3 FRAP and FCS 525
13.2.4 Drunks on a Hill: The Smoluchowski Equation 529
13.2.5 The Einstein Relation 530

13.3 DIFFUSION TO CAPTURE 532
13.3.1 Modeling the Cell Signaling Problem 532
Perfect Receptors Result in a Rate of Uptake $4\pi Dc_0 a$ 533
A Distribution of Receptors Is Almost as Good as a Perfectly Absorbing Sphere 534
Real Receptors Are Not Always Uniformly Distributed 536
13.3.2 A "Universal" Rate for Diffusion-Limited Chemical Reactions 537

13.4 SUMMARY AND CONCLUSIONS 538
13.5 PROBLEMS 539

13.6 FURTHER READING 540
13.7 REFERENCES 540

Chapter 14 Life in Crowded and Disordered Environments 543

14.1 CROWDING, LINKAGE, AND ENTANGLEMENT 543
14.1.1 The Cell Is Crowded 544
14.1.2 Macromolecular Networks: The Cytoskeleton and Beyond 545
14.1.3 Crowding on Membranes 546
14.1.4 Consequences of Crowding 547
Crowding Alters Biochemical Equilibria 548
Crowding Alters the Kinetics within Cells 548

14.2 EQUILIBRIA IN CROWDED ENVIRONMENTS 550
14.2.1 Crowding and Binding 550
Lattice Models of Solution Provide a Simple Picture of the Role of Crowding in Biochemical Equilibria 550
14.2.2 Osmotic Pressures in Crowded Solutions 552
Osmotic Pressure Reveals Crowding Effects 552
14.2.3 Depletion Forces: Order from Disorder 554
The Close Approach of Large Particles Excludes Smaller Particles Between Them, Resulting in an Entropic Force 554
Depletion Forces Can Induce Entropic Ordering! 559
14.2.4 Excluded Volume and Polymers 559
Excluded Volume Leads to an Effective Repulsion Between Molecules 559
Self-avoidance Between the Monomers of a Polymer Leads to Polymer Swelling 561
14.2.5 Case Study in Crowding: How to Make a Helix 563
14.2.6 Crowding at Membranes 565

14.3 CROWDED DYNAMICS 566
14.3.1 Crowding and Reaction Rates 566
Enzymatic Reactions in Cells Can Proceed Faster than the Diffusion Limit Using Substrate Channeling 566
Protein Folding Is Facilitated by Chaperones 567
14.3.2 Diffusion in Crowded Environments 567

14.4 SUMMARY AND CONCLUSIONS 569
14.5 PROBLEMS 569
14.6 FURTHER READING 570
14.7 REFERENCES 571

Chapter 15 Rate Equations and Dynamics in the Cell 573

15.1 BIOLOGICAL STATISTICAL DYNAMICS: A FIRST LOOK 573
15.1.1 Cells as Chemical Factories 574
15.1.2 Dynamics of the Cytoskeleton 575

15.2 A CHEMICAL PICTURE OF BIOLOGICAL DYNAMICS 579
15.2.1 The Rate Equation Paradigm 579
Chemical Concentrations Vary in Both Space and Time 580
Rate Equations Describe the Time Evolution of Concentrations 580
15.2.2 All Good Things Must End 581
Macromolecular Decay Can Be Described by a Simple, First-Order Differential Equation 581
15.2.3 A Single-Molecule View of Degradation: Statistical Mechanics Over Trajectories 582
Molecules Fall Apart with a Characteristic Lifetime 582
Decay Processes Can Be Described with Two-State Trajectories 583
Decay of One Species Corresponds to Growth in the Number of a Second Species 585
15.2.4 Bimolecular Reactions 586
Chemical Reactions Can Increase the Concentration of a Given Species 586
Equilibrium Constants Have a Dynamical Interpretation in Terms of Reaction Rates 588
15.2.5 Dynamics of Ion Channels as a Case Study 589
Rate Equations for Ion Channels Characterize the Time Evolution of the Open and Closed Probability 590
15.2.6 Rapid Equilibrium 591
15.2.7 Michaelis–Menten and Enzyme Kinetics 596

15.3 THE CYTOSKELETON IS ALWAYS UNDER CONSTRUCTION 599
15.3.1 The Eukaryotic Cytoskeleton 599
The Cytoskeleton Is a Dynamical Structure That Is Always Under Construction 599
15.3.2 The Curious Case of the Bacterial Cytoskeleton 600

15.4 SIMPLE MODELS OF CYTOSKELETAL POLYMERIZATION 602
The Dynamics of Polymerization Can Involve Many Distinct Physical and Chemical Effects 603
15.4.1 The Equilibrium Polymer 604
Equilibrium Models of Cytoskeletal Filaments Describe the Distribution of Polymer Lengths for Simple Polymers 604
An Equilibrium Polymer Fluctuates in Time 606
15.4.2 Rate Equation Description of Cytoskeletal Polymerization 609
Polymerization Reactions Can Be Described by Rate Equations 609
The Time Evolution of the Probability Distribution $P_n(t)$ Can Be Written Using a Rate Equation 610
Rates of Addition and Removal of Monomers Are Often Different on the Two Ends of Cytoskeletal Filaments 612
15.4.3 Nucleotide Hydrolysis and Cytoskeletal Polymerization 614
ATP Hydrolysis Sculpts the Molecular Interface, Resulting in Distinct Rates at the Ends of Cytoskeletal Filaments 614
15.4.4 Dynamic Instability: A Toy Model of the Cap 615
A Toy Model of Dynamic Instability Assumes That Catastrophe Occurs When Hydrolyzed Nucleotides Are Present at the Growth Front 616

15.5 SUMMARY AND CONCLUSIONS 618
15.6 PROBLEMS 619
15.7 FURTHER READING 621
15.8 REFERENCES 621

Chapter 16 Dynamics of Molecular Motors 623

16.1 THE DYNAMICS OF MOLECULAR MOTORS: LIFE IN THE NOISY LANE 623
16.1.1 Translational Motors: Beating the Diffusive Speed Limit 625
The Motion of Eukaryotic Cilia and Flagella Is Driven by Translational Motors 628
Muscle Contraction Is Mediated by Myosin Motors 630
16.1.2 Rotary Motors 634
16.1.3 Polymerization Motors: Pushing by Growing 637
16.1.4 Translocation Motors: Pushing by Pulling 638

16.2 RECTIFIED BROWNIAN MOTION AND MOLECULAR MOTORS 639
16.2.1 The Random Walk Yet Again 640
Molecular Motors Can Be Thought of as Random Walkers 640

16.2.2 The One-State Model 641
The Dynamics of a Molecular Motor Can Be Written Using a Master Equation 642
The Driven Diffusion Equation Can Be Transformed into an Ordinary Diffusion Equation 644
16.2.3 Motor Stepping from a Free-Energy Perspective 647
16.2.4 The Two-State Model 651
The Dynamics of a Two-State Motor Is Described by Two Coupled Rate Equations 651
Internal States Reveal Themselves in the Form of the Waiting Time Distribution 654
16.2.5 More General Motor Models 656
16.2.6 Coordination of Motor Protein Activity 658
16.2.7 Rotary Motors 660

16.3 POLYMERIZATION AND TRANSLOCATION AS MOTOR ACTION 663
16.3.1 The Polymerization Ratchet 663
The Polymerization Ratchet Is Based on a Polymerization Reaction That Is Maintained Out of Equilibrium 666
The Polymerization Ratchet Force–Velocity Can Be Obtained by Solving a Driven Diffusion Equation 668
16.3.2 Force Generation by Growth 670
Polymerization Forces Can Be Measured Directly 670
Polymerization Forces Are Used to Center Cellular Structures 672
16.3.3 The Translocation Ratchet 673
Protein Binding Can Speed Up Translocation through a Ratcheting Mechanism 674
The Translocation Time Can Be Estimated by Solving a Driven Diffusion Equation 676

16.4 SUMMARY AND CONCLUSIONS 677
16.5 PROBLEMS 677
16.6 FURTHER READING 679
16.7 REFERENCES 679

Chapter 17 Biological Electricity and the Hodgkin–Huxley Model 681

17.1 THE ROLE OF ELECTRICITY IN CELLS 681
17.2 THE CHARGE STATE OF THE CELL 682
17.2.1 The Electrical Status of Cells and Their Membranes 682
17.2.2 Electrochemical Equilibrium and the Nernst Equation 683
Ion Concentration Differences Across Membranes Lead to Potential Differences 683

17.3 MEMBRANE PERMEABILITY: PUMPS AND CHANNELS 685
A Nonequilibrium Charge Distribution Is Set Up Between the Cell Interior and the External World 685
Signals in Cells Are Often Mediated by the Presence of Electrical Spikes Called Action Potentials 686
17.3.1 Ion Channels and Membrane Permeability 688
Ion Permeability Across Membranes Is Mediated by Ion Channels 688
A Simple Two-State Model Can Describe Many of the Features of Voltage Gating of Ion Channels 689
17.3.2 Maintaining a Nonequilibrium Charge State 691
Ions Are Pumped Across the Cell Membrane Against an Electrochemical Gradient 691

17.4 THE ACTION POTENTIAL 693
17.4.1 Membrane Depolarization: The Membrane as a Bistable Switch 693
Coordinated Muscle Contraction Depends Upon Membrane Depolarization 694
A Patch of Cell Membrane Can Be Modeled as an Electrical Circuit 696
The Difference Between the Membrane Potential and the Nernst Potential Leads to an Ionic Current Across the Cell Membrane 698
Voltage-Gated Channels Result in a Nonlinear Current–Voltage Relation for the Cell Membrane 699
A Patch of Membrane Acts as a Bistable Switch 700
The Dynamics of Voltage Relaxation Can Be Modeled Using an *RC* Circuit 702
17.4.2 The Cable Equation 703
17.4.3 Depolarization Waves 705
Waves of Membrane Depolarization Rely on Sodium Channels Switching into the Open State 705
17.4.4 Spikes 710
17.4.5 Hodgkin–Huxley and Membrane Transport 712
Inactivation of Sodium Channels Leads to Propagating Spikes 712

17.5 SUMMARY AND CONCLUSIONS 714
17.6 PROBLEMS 714
17.7 FURTHER READING 715
17.8 REFERENCES 715

Chapter 18 Light and Life 717

18.1 INTRODUCTION 718
18.2 PHOTOSYNTHESIS 719
Organisms From All Three of the Great Domains of Life Perform Photosynthesis 720
18.2.1 Quantum Mechanics for Biology 724
Quantum Mechanical Kinematics Describes States of the System in Terms of Wave Functions 725
Quantum Mechanical Observables Are Represented by Operators 728
The Time Evolution of Quantum States Can Be Determined Using the Schrödinger Equation 729
18.2.2 The Particle-in-a-Box Model 730
Solutions for the Box of Finite Depth Do Not Vanish at the Box Edges 731
18.2.3 Exciting Electrons With Light 733
Absorption Wavelengths Depend Upon Molecular Size and Shape 735
18.2.4 Moving Electrons From Hither to Yon 737
Excited Electrons Can Suffer Multiple Fates 737
Electron Transfer in Photosynthesis Proceeds by Tunneling 739
Electron Transfer Between Donor and Acceptor Is Gated by Fluctuations of the Environment 745
Resonant Transfer Processes in the Antenna Complex Efficiently Deliver Energy to the Reaction Center 747
18.2.5 Bioenergetics of Photosynthesis 748
Electrons Are Transferred from Donors to Acceptors Within and Around the Cell Membrane 748
Water, Water Everywhere, and Not an Electron to Drink 750
Charge Separation across Membranes Results in a Proton-Motive Force 751
18.2.6 Making Sugar 752
18.2.7 Destroying Sugar 757
18.2.8 Photosynthesis in Perspective 758

18.3 THE VISION THING 759
18.3.1 Bacterial "Vision" 760
18.3.2 Microbial Phototaxis and Manipulating Cells with Light 763
18.3.3 Animal Vision 763
There Is a Simple Relationship between Eye Geometry and Resolution 765
The Resolution of Insect Eyes Is Governed by Both the Number of Ommatidia and Diffraction Effects 768
The Light-Driven Conformational Change of Retinal Underlies Animal Vision 769
Information from Photon Detection Is Amplified by a Signal Transduction Cascade in the Photoreceptor Cell 773

The Vertebrate Visual System Is Capable of Detecting Single Photons 776
18.3.4 Sex, Death, and Quantum Mechanics 781
Let There Be Light: Chemical Reactions Can Be Used to Make Light 784
18.4 SUMMARY AND CONCLUSIONS 785
18.5 APPENDIX: SIMPLE MODEL OF ELECTRON TUNNELING 785
18.6 PROBLEMS 793
18.7 FURTHER READING 795
18.8 REFERENCES 796

PART 4 THE MEANING OF LIFE 799

Chapter 19 Organization of Biological Networks 801

19.1 CHEMICAL AND INFORMATIONAL ORGANIZATION IN THE CELL 801
Many Chemical Reactions in the Cell are Linked in Complex Networks 801
Genetic Networks Describe the Linkages Between Different Genes and Their Products 802
Developmental Decisions Are Made by Regulating Genes 802
Gene Expression Is Measured Quantitatively in Terms of How Much, When, and Where 804

19.2 GENETIC NETWORKS: DOING THE RIGHT THING AT THE RIGHT TIME 807
Promoter Occupancy Is Dictated by the Presence of Regulatory Proteins Called Transcription Factors 808
19.2.1 The Molecular Implementation of Regulation: Promoters, Activators, and Repressors 808
Repressor Molecules Are the Proteins That Implement Negative Control 808
Activators Are the Proteins That Implement Positive Control 809
Genes Can Be Regulated During Processes Other Than Transcription 809
19.2.2 The Mathematics of Recruitment and Rejection 810
Recruitment of Proteins Reflects Cooperativity Between Different DNA-Binding Proteins 810
The Regulation Factor Dictates How the Bare RNA Polymerase Binding Probability Is Altered by Transcription Factors 812
Activator Bypass Experiments Show That Activators Work by Recruitment 813
Repressor Molecules Reduce the Probability Polymerase Will Bind to the Promoter 814
19.2.3 Transcriptional Regulation by the Numbers: Binding Energies and Equilibrium Constants 819
Equilibrium Constants Can Be Used To Determine Regulation Factors 819
19.2.4 A Simple Statistical Mechanical Model of Positive and Negative Regulation 820
19.2.5 The *lac* Operon 822
The *lac* Operon Has Features of Both Negative and Positive Regulation 822
The Free Energy of DNA Looping Affects the Repression of the *lac* Operon 824
Inducers Tune the Level of Regulatory Response 829
19.2.6 Other Regulatory Architectures 829
The Fold-Change for Different Regulatory Motifs Depends Upon Experimentally Accessible Control Parameters 830
Quantitative Analysis of Gene Expression in Eukaryotes Can Also Be Analyzed Using Thermodynamic Models 832

19.3 REGULATORY DYNAMICS 835
19.3.1 The Dynamics of RNA Polymerase and the Promoter 835
The Concentrations of Both RNA and Protein Can Be Described Using Rate Equations 835
19.3.2 Dynamics of mRNA Distributions 838
Unregulated Promoters Can Be Described By a Poisson Distribution 841
19.3.3 Dynamics of Regulated Promoters 843
The Two-State Promoter Has a Fano Factor Greater Than One 844
Different Regulatory Architectures Have Different Fano Factors 849
19.3.4 Dynamics of Protein Translation 854
19.3.5 Genetic Switches: Natural and Synthetic 861
19.3.6 Genetic Networks That Oscillate 870

19.4 CELLULAR FAST RESPONSE: SIGNALING 872
19.4.1 Bacterial Chemotaxis 873
The MWC Model Can Be Used to Describe Bacterial Chemotaxis 878
Precise Adaptation Can Be Described by a Simple Balance Between Methylation and Demethylation 881
19.4.2 Biochemistry on a Leash 883
Tethering Increases the Local Concentration of a Ligand 884
Signaling Networks Help Cells Decide When and Where to Grow Their Actin Filaments for Motility 884
Synthetic Signaling Networks Permit a Dissection of Signaling Pathways 885

19.5 SUMMARY AND CONCLUSIONS 888
19.6 PROBLEMS 889
19.7 FURTHER READING 891
19.8 REFERENCES 892

Chapter 20 Biological Patterns: Order in Space and Time 893

20.1 INTRODUCTION: MAKING PATTERNS 893
20.1.1 Patterns in Space and Time 894
20.1.2 Rules for Pattern-Making 895

20.2 MORPHOGEN GRADIENTS 896
20.2.1 The French Flag Model 896
20.2.2 How the Fly Got His Stripes 898
Bicoid Exhibits an Exponential Concentration Gradient Along the Anterior–Posterior Axis of Fly Embryos 898
A Reaction–Diffusion Mechanism Can Give Rise to an Exponential Concentration Gradient 899
20.2.3 Precision and Scaling 905
20.2.4 Morphogen Patterning with Growth in *Anabaena* 912

20.3 REACTION–DIFFUSION AND SPATIAL PATTERNS 914
20.3.1 Putting Chemistry and Diffusion Together: Turing Patterns 914
20.3.2 How Bacteria Lay Down a Coordinate System 920
20.3.3 Phyllotaxis: The Art of Flower Arrangement 926

20.4 TURNING TIME INTO SPACE: TEMPORAL OSCILLATIONS IN CELL FATE SPECIFICATION 931
20.4.1 Somitogenesis 932
20.4.2 Seashells Forming Patterns in Space and Time 935

20.5 PATTERN FORMATION AS A CONTACT SPORT 939
20.5.1 The Notch–Delta Concept 939
20.5.2 *Drosophila* Eyes 944

20.6 SUMMARY AND CONCLUSIONS 947
20.7 PROBLEMS 948
20.8 FURTHER READING 949
20.9 REFERENCES 950

Chapter 21 Sequences, Specificity, and Evolution **951**

21.1 BIOLOGICAL INFORMATION 952
21.1.1 Why Sequences? 953
21.1.2 Genomes and Sequences by the Numbers 957

21.2 SEQUENCE ALIGNMENT AND HOMOLOGY 960
Sequence Comparison Can Sometimes Reveal Deep Functional and Evolutionary Relationships Between Genes, Proteins, and Organisms 961
21.2.1 The HP Model as a Coarse-Grained Model for Bioinformatics 964
21.2.2 Scoring Success 966
A Score Can Be Assigned to Different Alignments Between Sequences 966
Comparison of Full Amino Acid Sequences Requires a 20–by–20 Scoring Matrix 968
Even Random Sequences Have a Nonzero Score 970
The Extreme Value Distribution Determines the Probability That a Given Alignment Score Would Be Found by Chance 971
False Positives Increase as the Threshold for Acceptable Expect Values (also Called E–Values) Is Made Less Stringent 973
Structural and Functional Similarity Do Not Always Guarantee Sequence Similarity 976

21.3 THE POWER OF SEQUENCE GAZING 976
21.3.1 Binding Probabilities and Sequence 977
Position Weight Matrices Provide a Map Between Sequence and Binding Affinity 978
Frequencies of Nucleotides at Sites Within a Sequence Can Be Used to Construct Position Weight Matrices 979
21.3.2 Using Sequence to Find Binding Sites 983
21.3.3 Do Nucleosomes Care About Their Positions on Genomes? 988
DNA Sequencing Reveals Patterns of Nucleosome Occupancy on Genomes 989
A Simple Model Based Upon Self-Avoidance Leads to a Prediction for Nucleosome Positioning 990

21.4 SEQUENCES AND EVOLUTION 993
21.4.1 Evolution by the Numbers: Hemoglobin and Rhodopsin as Case Studies in Sequence Alignment 994
Sequence Similarity Is Used as a Temporal Yardstick to Determine Evolutionary Distances 994
Modern–Day Sequences Can Be Used to Reconstruct the Past 996
21.4.2 Evolution and Drug Resistance 998
21.4.3 Viruses and Evolution 1000
The Study of Sequence Makes It Possible to Trace the Evolutionary History of HIV 1001
The Luria–Delbrück Experiment Reveals the Mathematics of Resistance 1002
21.4.4 Phylogenetic Trees 1008

21.5 THE MOLECULAR BASIS OF FIDELITY 1010
21.5.1 Keeping It Specific: Beating Thermodynamic Specificity 1011
The Specificity of Biological Recognition Often Far Exceeds the Limit Dictated by Free-Energy Differences 1011
High Specificity Costs Energy 1015

21.6 SUMMARY AND CONCLUSIONS 1016
21.7 PROBLEMS 1017
21.8 FURTHER READING 1020
21.9 REFERENCES 1021

Chapter 22 Whither Physical Biology? **1023**

22.1 DRAWING THE MAP TO SCALE 1023

22.2 NAVIGATING WHEN THE MAP IS WRONG 1027

22.3 INCREASING THE MAP RESOLUTION 1028

22.4 "DIFFICULTIES ON THEORY" 1030
Modeler's Fantasy 1031
Is It Biologically Interesting? 1031
Uses and Abuses of Statistical Mechanics 1032
Out-of-Equilibrium and Dynamic 1032
Uses and Abuses of Continuum Mechanics 1032
Too Many Parameters 1033
Missing Facts 1033
Too Much Stuff 1033
Too Little Stuff 1034
The Myth of "THE" Cell 1034
Not Enough Thinking 1035

22.5 THE RHYME AND REASON OF IT ALL 1035
22.6 FURTHER READING 1036
22.7 REFERENCES 1037

Index *1039*

Special Sections

There are five classes of special sections indicated with icons and colored bars throughout the text. They perform order of magnitude estimates, explore biological problems using computation, examine the experimental underpinnings of topics, and elaborate on mathematical details.

COMPUTATIONAL EXPLORATION

Sizing Up *E. coli* 38
Counting mRNA and Proteins by Dilution 46
Timing *E. coli* 100
Growth Curves and the Logistic Equation 103
Determining the Spring Constant of an Optical Trap 207
Numerical Root Finding 257
Determining Ion Channel Open Probability by Thresholding 284
Numerical Solution of the Cable Equation 707
Electrons in a Well of Finite Depth 733
Extracting Level of Gene Expression from Microscopy Images 817
The Gillespie Algorithm and Stochastic Models of Gene Regulation 849
Scaling of Morphogen Gradients 901
Performing Sequence Alignments Against a Database 974
Searching the *E. coli* Genome for Binding Sites 981

ESTIMATE

Sizing Up *E. coli* 39
Cell-to-Cell Variability in the Cellular Census 44
Sizing Up Yeast 55
Membrane Area of the Endoplasmic Reticulum 60
Sizing Up HIV 68
Sizing Up the Slug and the Fruiting Body 75
Sizing Up Stripes in *Drosophila* Embryos 79
Sizing Up *C. elegans* 82
Timing *E. coli* 101
Timing the Machines of the Central Dogma 109
Timing Development 124
The Thermal Energy Scale 127
Moving Proteins from Here to There 128
Diffusion at the Synaptic Cleft 129
Moving Proteins from Here to There, Part 2 130
Ion Transport Rates in Ion Channels 130
Hemoglobin by the Numbers 143
The Energy Budget Required to Build a Cell 197
Osmotic Pressure in a Cell 266
End-to-End Probability for the *E. coli* Genome 319
The Size of Viral and Bacterial Genomes 322
Chromosome Packing in the Yeast Nucleus 324
Chromosome Organization in *C. crescentus* 327
The Eighth Continent 356
DNA Condensation in Bacteriophage ϕ29? 368
DNA Condensation in Bacteriophage ϕ29 Redux 373
The DNA Packing Compaction Ratio. 399
Sizing Up Nucleosomes 407
Sizing Up Membrane Heterogeneity 436
Vesicle Counts and Energies in Cells 456
Sizing Up Membrane Area in Mitochondria 464
Blood Flow Through Capillaries 493
Mechanics of Leukocyte Rolling 495
Rate of ATP Synthesis in Humans 575
The Rate of Actin Polymerization 577
Equilibrium Polymers? 606
Force Exerted During a Single Motor Step 627
Myosin and Muscle Forces 630
Competition in the ATP Synthase 661
Charge Pumping at Membranes 693
Charge Transfer During Depolarization 697
Solar Energy Fluxes 720
Sizing Up Cyanobacteria 722
Confinement Energies of Electrons 725
Number of Incident Photons Per Pigment Molecule 736
The Tunneling Length Scale 740
Distance Dependence of Tunneling Times 744
Photosynthetic Productivity on Earth 755
Number of Rhodopsin Molecules Per Rod 770
Dynamics of Transcription by the Numbers 837
Bicoid Concentration Difference Between Neighboring Nuclei 908
Genome Size and the Number of Genes 959

EXPERIMENTS

Probing Biological Structure 49
Measurements of Biological Time 92
Genetics 139
Biochemistry 141
Measuring Diffusive Dynamics 513
Taking the Molecular Census 578
Measuring Motor Action 632
Dynamics of Rotary Motors 636
Dynamics of Light and Electrons 742
Measuring Gene Expression 804
Measuring the Process of Chemotaxis 874
Sequencing and Protecting DNA 954

MATH

The Partial Derivative 212
The Beauty of the Taylor Expansion 215
The Stirling Approximation 222
Counting Arrangements of Particles 239
One Person's Macrostate Is Another's Microstate 250
The Method of Lagrange Multipliers 254
The Gaussian Integral 261
Expanding in Sines and Cosines 332
The Gradient Operator and Vector Calculus 366
Fourth Roots of -1 472
Eigenvalues and Eigenvectors 595
The Poisson Distribution 779
Laplace Transforms and Convolutions 858
Linear Stability Analysis for the Genetic Switch 868

TRICKS

Differentiation with Respect to a Parameter 241
Averaging Sums of Random Variables 522
Doing Integrals by Differentiating With Respect to a Parameter 525
Dot Products to Find Amplitudes 792
Phase Portraits and Vector Fields 866

Map of the Maps

Part 1: Map of Alfred Russel Wallace's voyage with the black lines denoting Wallace's travel route and the red lines indicating chains of volcanoes. From *The Malay Archipelago* (1869) by Alfred Russel Wallace.

Chapter 1: Map of the world according to Eratosthenes (220 B.C.E.). Erastosthenes is known for, among many other things, his measurement of the circumference of the Earth, and is considered one of the founders of the subject of geography. From *Report on the Scientific Results of the Voyage of the H.M.S. Challenger During the Years 1872–76*, prepared under the superintendence of C. Wyville Thompson and John Murray (1895).

Chapter 2: Population density in Los Angeles County, as determined in the 2000 census. Darker colors represent denser populations (up to 100,000 people per square mile). From the United States Census Bureau.

Chapter 3: Sedimentary rock layers in the Grand Canyon. Geology and cross section by Peter J. Conley, artwork by Dick Beasley. From the United States National Park Service (1985).

Chapter 4: *Carta marina*, a map of Scandinavia, by Olaus Magnus. A translation of the Latin caption reads: *A Marine map and Description of the Northern Lands and of their Marvels, most carefully drawn up at Venice in the year 1539 through the generous assistance of the Most Honourable Lord Hieronymo Quirino.* This detail shows the sea monsters in the ocean between Norway and Iceland.

Part 2: Tourist map of Père Lachaise cemetery, Paris, France.

Chapter 5: Airplane routes around the nearly spherical Earth. Courtesy of OpenFlights.com.

Chapter 6: Josiah Willard Gibbs articulated the variational principle that shows how to find the equilibrium state of a system by maximizing the entropy. Gibbs spent his entire career in New Haven, Connecticut at Yale University. This 1886 map shows the university buildings during Gibbs' time. Source: Yale University Map Collection. Courtesy of the Yale University Map Collection.

Chapter 7: County map of Virginia and West Virginia, drawn by Samuel Augustus Mitchell Jr. in 1864, after the American Civil War.

Chapter 8: Aerial view of the hedge maze at Longleat Safari and Adventure Park, near Warminster, United Kingdom. Courtesy of Atlaspix/Alamy.

Chapter 9: Topographic map of the Great Salt Lake (Utah, United States) and surrounding region. From the United States Geological Survey (1970).

Chapter 10: Blueprint diagram of the Golden Gate Bridge, San Francisco, California, United States. Courtesy of EngineeringArtwork.com

Chapter 11: Digital elevation map of Mount Cotopaxi in the Andes Mountains, near Quito, Ecuador. Blue and green correspond to the lowest elevations in the image, while beige, orange, red, and white represent increasing elevations. Courtesy of the NASA Earth Observatory (2000).

Part 3: Migration tracks of the sooty shearwater, a small seabird, tracked with geolocating tags from two breeding colonies in New Zealand. Breeding season is shown in blue, northward migration in yellow, and wintering season and southward migration in orange. Over about 260 days, an individual animal travels about 64,000 km in a figure-8 pattern across the entire Pacific Ocean. From S. A. Shaffer et al., "Migratory shearwaters integrate oceanic resources across the Pacific Ocean in an endless summer," *Proceedings of the National Academy of Sciences USA*, **103**: 12799–12802, 2006.

Chapter 12: Worldwide distribution of ocean currents (warm in red, cold in green). Arrows indicate the direction of drift; the number of strokes on the arrow shafts denote the magnitude of the drift per hour. Sea ice is shown in purple. Prepared by the American Geographical Society for the United States Department of State in 1943.

Chapter 13: Temperature map of the sun's corona, recorded by the Extreme Ultraviolet Imaging Telescope at the Solar and Heliospheric Observatory on June 21, 2001. Courtesy of ESA/NASA.

Chapter 14: John Snow's map of the 1854 cholera outbreak in the Soho neighborhood of London. By interviewing residents of the neighborhood where nearly 500 people died of cholera in a ten-day period, Snow found that nearly all of the deaths occurred in homes close to the water pump in Broad Street, which he hypothesized was the source of the epidemic. Reproduced from *On the Mode of Communication of Cholera,* 2nd *Edition*, John Snow (1855).

Chapter 15: Positron emission tomography (PET scan) map of a healthy human brain, showing the rate of glucose utilization in various parts of the right hemisphere. Warmer colors indicate faster glucose uptake. Courtesy of Alzheimer's Disease Education and Referral Center, a service of the National Institute on Aging (United States National Institutes of Health).

Chapter 16: High speed train routes of France, mapped as a transit diagram. Courtesy of Cameron Booth.

Chapter 17: Nile River delta at night, as photographed by the crew in Expedition 25 on the International Space Station on October 28, 2010. Courtesy of Image Science & Analysis Laboratory, Johnson Space Center, Earth Observatory, NASA/GSFC SeaWiFS Project.

Chapter 18: Single-celled photosynthetic organisms such as the coccolithophore *Emiliana huxleyi* can form gigantic oceanic blooms visible from space. In this April 1998 image, the Aleutian Islands and the state of Alaska are visible next to the Bering Sea that harbors the algal bloom. Courtesy of NASA/GSFC SeaWiFS Project.

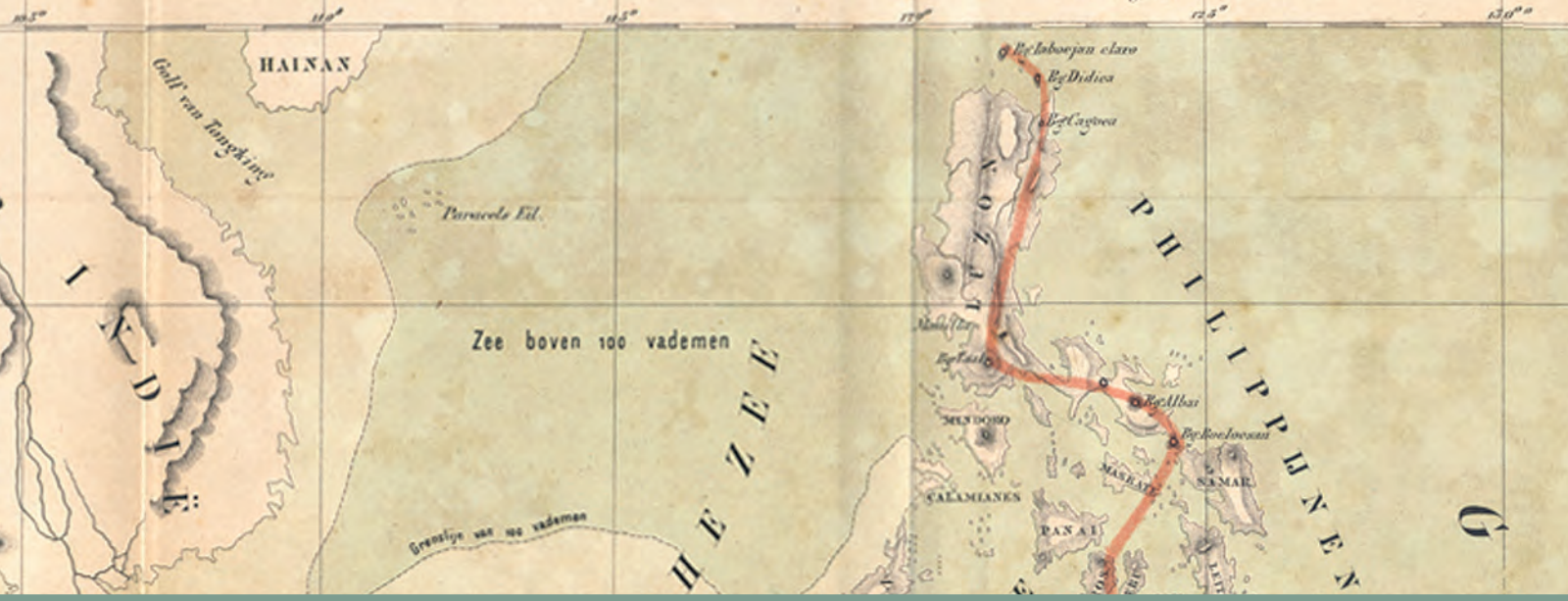
Part 4: A map of the infant universe, revealed by seven years of data from the Wilkinson Microwave Anisotropy Probe (WMAP). The image reveals 13.7 billion year old temperature fluctuations (the range of ±200 microKelvin is shown as color differences) that correspond to the seeds that grew to become the galaxies. Courtesy of NASA/WMAP Science Team.

Chapter 19: Map of the Internet, as of September, 1998, created by Bill Cheswick. *Courtesy of* Lumeta Corporation 2000–2011. Published in Wired Magazine, December 1998 (issue 6.12).

Chapter 20: The Sloan Great Wall measured by J. Richard Gott and Mario Juric shows a wall of galaxies spanning 1.37 billion light years. It stands in the Guinness Book of Records as the largest structure in the universe. Courtesy of Michael Blanton and the Sloan Digital Sky Survey Collaboration, www.sdss.org.

Chapter 21: This map shows the patterns of human migration as inferred from modern geographical distributions of marker sequences in the Y chromosome (blue), indicating patrilineal inheritance, and in the mitochondrial DNA (orange), indicating matrilineal inheritance. Courtesy of National Geographic Maps, Atlas of the Human Journey.

Chapter 22: "The Lands Beyond" drawn by Jules Feiffer for The Phantom Tollbooth (1961) by Norton Juster. Courtesy of Knopf Books for Young Readers, a division of Random House, Inc.

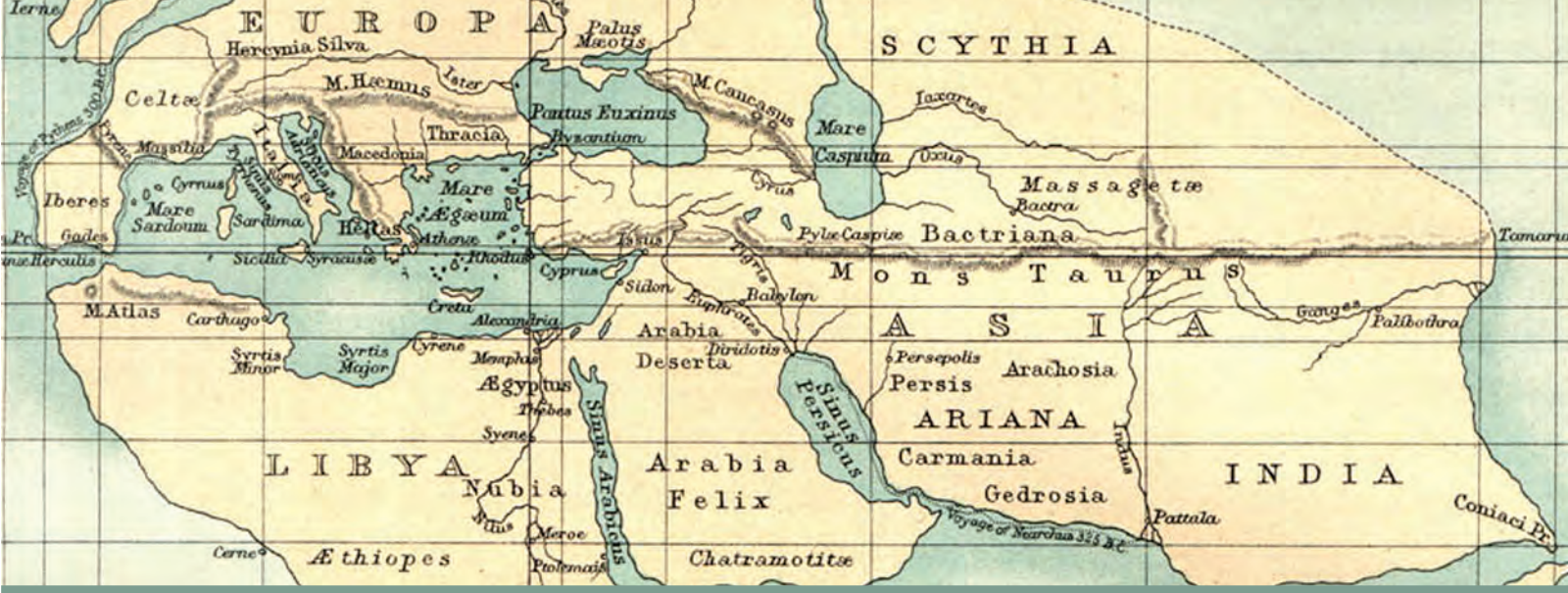


The Facts of Life

PART 1



This page intentionally left blank
to match pagination of print book



Why: Biology by the Numbers

1

Overview: In which the notion of biological numeracy is examined

Charles Darwin once said "All observation should be for or against some view if it is to be of any service," meaning that data are most meaningful against the backdrop of some conceptual or theoretical framework (Darwin, 1903, p. 194). The goal of this chapter is to develop a sense of the kinds of quantitative data that are being obtained in all fields of biology and the models that must be put forth to greet this data. Like with any useful map, we will argue that any good model has to overlook some of the full complexity and detail of a given biological problem in order to generate an abstraction that is simple enough to be easily grasped by the human mind, as an aid to developing intuition and insight. At the same time, it is critical that useful models make meaningful predictions, so they must include at least some of the realistic details of the biological system. The art of model building lies in striking the proper balance between too little detail and too much. Part of our emphasis on model building centers on the role of having a feeling for the numbers: sizes, shapes, times, and energies associated with biological processes. Here we introduce the style of making numerical estimates that will be used throughout the book.

> "To simplify is the greatest form of sophistication."
>
> **Attributed to Leonardo Da Vinci**

1.1 Biological Cartography

In one of the great stories in scientific history, Eratosthenes used the shadow of the noon-day sun at two different positions to work out a first, and surprisingly accurate, measurement of the radius of the Earth. Measurements like those of Eratosthenes made it possible to construct rudimentary maps of the world as it was then known to the Greeks with results such as that decorating the opening to this chapter. Such maps provide a distorted but recognizable view of the world.

In the chapters that follow, our aim is to similarly show how a few fundamental measurements and calculations suffice to provide the rough contours of what is clearly a much more nuanced biological landscape. The idea of such maps is to reveal enough detail to provide a conceptual framework that helps us navigate such biological landscapes, and serve as a starting point to orient ourselves for more in-depth exploration. In constructing such maps, we walk a perilous ridge between descriptions that are so distorted as to be uninformative and those that are so detailed that it is no longer correct to even think of them as maps. These ideas find their literary embodiment in one of Borges' shortest short stories, which serves as the battle cry for the chapters that follow.

> *On Exactitude in Science* . . .
> In that Empire, the Art of Cartography attained such Perfection that the map of a single Province occupied the entirety of a City, and the map of the Empire, the entirety of a Province. In time, those Unconscionable Maps no longer satisfied, and the Cartographers Guilds struck a Map of the Empire whose size was that of the Empire, and which coincided point for point with it. The following Generations, who were not so fond of the Study of Cartography as their Forebears had been, saw that that vast Map was Useless, and not without some Pitilessness was it, that they delivered it up to the Inclemencies of Sun and Winters. In the Deserts of the West, still today, there are Tattered Ruins of that Map, inhabited by Animals and Beggars; in all the Land there is no other Relic of the Disciplines of Geography. Suàrez Miranda: Viajes de varones prudentes, Libro Cuarto, cap. XLV, Lérida, 1658.
>
> Jorge Luis Borges and Adolfo Bioy Casares

1.2 Physical Biology of the Cell

With increasing regularity, the experimental methods used to query living systems and the data they produce are quantitative. For example, measurements of gene expression in living cells report how much of a given gene product there is at a given place at a given time. Measurements on signaling pathways examine the extent of a cellular response such as growth, DNA replication, or actin polymerization, as a function of the concentration of some upstream signaling molecule. Other measurements report the tendency of DNA to wrap into compact structures known as nucleosomes as a function of the underlying sequence. The list of examples goes on and on and is revealed by the many graphs depicting quantitative measurements on biological problems that grace the current research literature.

The overarching theme of the developing field of physical biology of the cell is that quantitative data demand quantitative models and, conversely, that quantitative models need to provide experimentally testable quantitative predictions about biological phenomena. What this means precisely is that in those cases where a biological problem has resulted in a quantitative measurement of how a particular biological output parameter varies as some input is changed (either because it is manipulated by the experimenter or because it varies as part of a natural biological process), the interpretation of that biological data should also be quantitative. In the chapters that follow, our aim is to provide a series of case studies involving some of the key players and processes in molecular and cell biology that demonstrate how physical model building can respond to this new era of quantitative data and sharpen the interplay between theory and experiment.

Model Building Requires a Substrate of Biological Facts and Physical (or Chemical) Principles

One of the first steps in building models of quantitative experiments like those described above is trying to decide which features of the problem are of central importance and which are not. However, even before this, we need to know facts about our system of interest. Throughout this book, we will assert a series of "facts" that fall into several different categories of certainty. We are most comfortable with statements describing observations that you as an individual observer can readily confirm, such as the statement that cells contain protein molecules. A second layer of facts with which we are nearly as comfortable are those arrived at and repeatedly confirmed by decades of experimentation using many different, independent experimental techniques. One fact of this kind is the statement that protein molecules within cells are synthesized by a large piece of machinery called the ribosome. A third category of "facts" that we will take care to mark as speculative are those that stand as compelling explanations for biological observations, but that may rest on unproven or controversial assumptions. In this category, we would include the commonly accepted proposition that modern-day mitochondria are the descendants of what was once a free-living bacterium that entered into a symbiotic relationship with an ancestral eukaryotic cell. In these cases, we will generally explain the observation and the interpretation that underlies our modeling approach. One of the most basic classes of facts that we will need throughout the remainder of the book concerns the nature of the molecules that make up cells and organisms. Because this will set up the common molecular language for the rest of our discussions, we will describe these molecules at some length in the next section.

Complementary to these biological observations, we embrace the proposition that biological entities cannot violate the laws of physics and chemistry. Many fundamental physical and chemical principles can be directly applied to understanding the behavior of biological systems, for example, the physical principle that the average energy of a molecule increases with increasing environmental temperature, and the chemical principle that oil and water do not mix. Along with biological facts described above, such principles will serve as the basis for the models to be described throughout the book.

1.3 The Stuff of Life

Scientists and other curious humans observe the natural world around them and try to make sense of its constituent parts and the ways that they interact. The particular subset of the natural world called life has always held special fascination. Any 3-year-old child knows that a rock is not alive but that a puppy, a tree, and indeed the child himself are alive. Yet the 3-year-old would be hard pressed to produce a rigorous definition of life and, indeed, scientists who have thought about this for decades can only approximate a description of this fascinating property. No single feature can reliably discriminate life from nonlife, but all living systems do share certain central properties. Living systems grow, consume energy from their environment, reproduce to form offspring that are rather similar to themselves, and die.

In the quest to understand the nature of life, scientists over hundreds of years have investigated the properties of living organisms

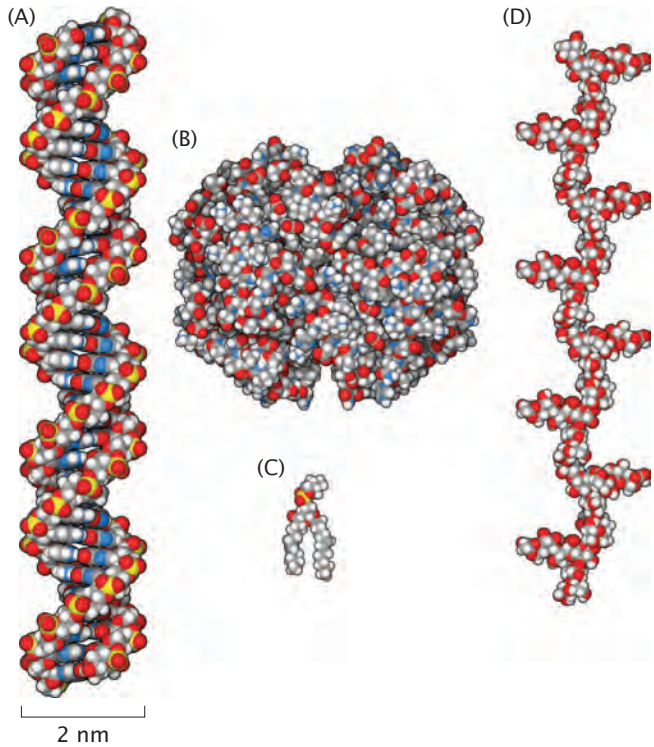


Figure 1.1: Atomic-level structural representation of members of each of the major classes of macromolecules, all drawn at the same scale. Nitrogen is colored in blue, oxygen in red, phosphorus in yellow, carbon in gray, and hydrogen in white. (A) Atomic structure of a small fragment of the nucleic acid DNA in the B form, (B) atomic structure of the oxygen-carrying protein hemoglobin (PDB 1hho), (C) phosphatidylcholine lipid molecule from a cell membrane, and (D) branched complex carbohydrate (M41 capsular polysaccharide) from the surface of the bacterium *Escherichia coli* (PDB 1cap). (Illustrations courtesy of D. Goodsell.)

and their constituent materials with an overarching goal of trying to understand why a puppy is different from a rock, even though they are both made of atoms that obey the same physical laws. Although this quest is still ongoing, some useful generalizations about the material nature of life are now widely accepted. For example, one observed feature of many of the molecules that make up living organisms is that they tend to be relatively large and structurally complex, and are hence called macromolecules. Living organisms also contain a large number of small, simple molecules that are critical to their function, ranging from water and metal ions to sugars such as glucose. Although these small molecules are important to life processes and, indeed, are used by the cell as building blocks to make the macromolecules, a distinction is usually drawn between small molecules that can arise through nonliving chemical processes and characteristic biological macromolecules that are found nowhere but in living organisms. Despite the diversity of macromolecules found in living organisms, the kinds of atoms found within them are surprisingly restricted, consisting mainly of carbon (C), oxygen (O), nitrogen (N), and hydrogen (H), with a smattering of sulfur (S) and phosphorus (P). It is largely the ways in which this humble suite of atoms are connected to each other that create the special properties of the macromolecules found in living systems. Of course, many other kinds of atoms and ions are critical for the biochemical processes of life, but they are generally not covalently incorporated in macromolecular structures.

Organisms Are Constructed from Four Great Classes of Macromolecules

If we take any organism, ranging from a small soil bacterium to a giant redwood tree or a massive whale, and separate out its constituent macromolecules, we will find a small number of distinct classes, chief among these being proteins, carbohydrates, lipids, and nucleic

acids. Figure 1.3 shows representative examples of each of these classes. Broadly speaking, these different classes of macromolecules perform different functions in living organisms. For example, proteins form structural elements within cells and are largely responsible for catalyzing specific chemical reactions necessary to life. Lipids form membrane barriers that separate the cell from the outside world and subdivide the cell into interior compartments called organelles. Carbohydrates are used for energy storage, for creating specific surface properties on the outside of cell membranes, and, sometimes, for building rigid structural units such as cell walls. Nucleic acids perform a critical function as the memory and operating instructions that enable cells to generate all the other kinds of macromolecules and to replicate themselves. Within each of these broad categories there is tremendous diversity. A single organism may have tens of thousands of structurally distinct proteins within it. Similarly, the membranes of cells can comprise hundreds of different lipid species.

Why are these particular classes of macromolecules the stuff of life? A complete understanding of the answer to this question would require knowledge of the specific chemical and physical conditions on the early Earth that gave rise to the first living cells. However, by studying living organisms on Earth today, it is seen that these molecules are wonderfully suited for the perpetuation of life for several reasons. First, each of these classes of molecules can be assembled by the cell from a small number of simpler subunit or precursor molecules. It is the combinatorial assembly of these simple subunits that gives rise to the tremendous structural diversity mentioned above. What this means is that a cell needs only a relatively restricted repertoire of chemical reactions to be able to synthesize these sets of subunits from the food in its environment. The fact that most cells on Earth make a living by consuming other living organisms is probably both a cause and a consequence of the fact that they all share very similar suites of these fundamental building blocks.

Nucleic Acids and Proteins Are Polymer Languages with Different Alphabets

The rest of this book will be largely devoted to exploring the special properties of biological macromolecules and the ways that they are used by cells, with the large-scale goal of understanding how the special properties of life may emerge from fundamental physical and chemical interactions. We will begin by discussing one specific and fundamentally important aspect of macromolecular behavior that is necessary for life as we know it. In particular, a great triumph of the study of molecular biology in the middle part of the twentieth century was the profound realization that the sequence of nucleic acid subunits in the cell's DNA is directly responsible for determining the sequence of amino acids in that same cell's proteins (that is, the discovery of the genetic code). The relationship between a DNA sequence and a protein sequence has been described by Francis Crick as comprising the "two great polymer languages" of cells.

As shown in Figure 1.2, both nucleic acids and proteins are built up from a limited alphabet of units. The nucleic acids are built up from an alphabet of four letters that each represent a chemically distinct "nucleotide" (A – adenine, G – guanine, T – thymine, and C – cytosine, for DNA). The chemical structure of these four nucleotides is shown in Figure 1.3. The protein language is constructed from an alphabet of 20

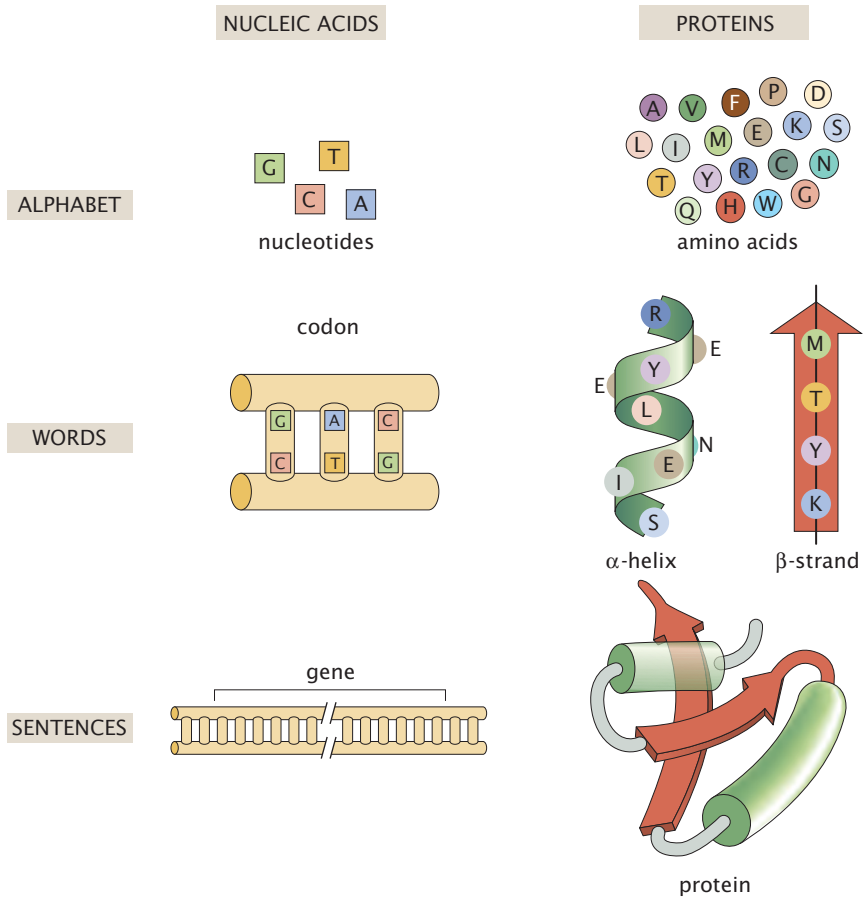


Figure 1.2: Illustration of Crick's "two great polymer languages." The left-hand column shows how nucleic acids can be thought of in terms of letters (nucleotides), words (codons), and sentences (genes). The right-hand column illustrates a similar idea for proteins, where the letters correspond to amino acids, the words to elements of secondary structure such as α helices and β strands, and the sentences to fully folded functional proteins.

distinct amino acids. How can just four nucleotide bases be translated into protein sequences containing 20 different amino acids? The key realization is that the DNA letters are effectively read as "words," called codons, made up of three sequential nucleotides, and each possible combination of three nucleotides encodes one amino acid (with some built-in redundancy in the code and a few exceptions such as the codon for "stop"). In the protein language, groups of amino acid "letters" also form "words," which give rise to the fundamental units of protein secondary structure, namely α helices and β strands, which we will discuss below. Finally, in the DNA language, sentences are formed by collections of words (that is, collections of codons) and correspond to genes, where we note that a given gene codes for a corresponding protein. An example of a complete sentence in the protein language is a fully folded, compact, biologically active enzyme that is capable of catalyzing a specific biochemical reaction within a living cell. Much of the rest of the book will be built around developing models of a host of biological processes in cells that appeal to the macromolecular underpinnings described here.

The sequences associated with nucleic acids and proteins are linked mechanistically through the ribosome, which takes nucleic acid sequences (in the form of messenger RNA (mRNA)) and converts them into amino acid sequences (in the form of proteins). These two polymer languages are linked informationally in the form of the genetic code. In particular, each three-letter collection from the nucleic acid language has a corresponding letter that it codes for in the protein language. The determination of this genetic code is one of the great chapters in the history of molecular biology and the results of that

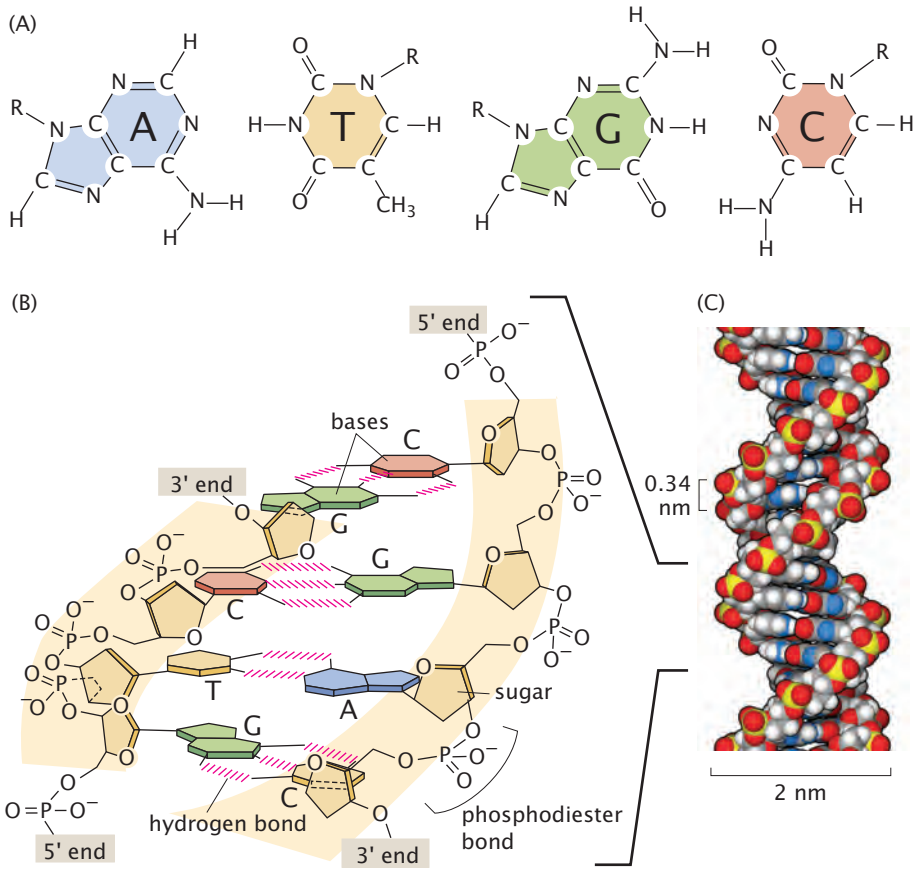


Figure 1.3: The chemical structure of nucleotides and DNA. (A) In DNA, the four distinct subunits or bases are abbreviated: A (adenine), T (thymine), G (guanine), and C (cytosine). In these diagrams, carbon is represented by C, oxygen by O, nitrogen by N and hydrogen by H. The letter R indicates attachment to a larger chemical group (the rest of the molecule); for nucleotides, the R group consists of the pentose sugar deoxyribose attached to phosphate. A single line connecting two atoms indicates a single covalent bond and a double line indicates a double covalent bond. The two large bases, A and G, are called purines and the two small bases, C and T, are called pyrimidines. (B) Illustration of how bases are assembled to form DNA, a double helix with two "backbones" made of the deoxyribose and phosphate groups. The four bases are able to form stable hydrogen bonds uniquely with one partner such that A pairs only with T and G pairs only with C. The structural complementarity of the bases enables the faithful copying of the nucleotide sequence when DNA is replicated or when RNA is transcribed. (C) Space-filling atomic model approximating the structure of DNA. The spacing between neighboring base pairs is 0.34 nm. (Adapted from B. Alberts et al., Molecular Biology of the Cell, 5th ed. Garland Science, 2008.)

quest are shown in Figure 1.4, which shows the information content recognized by the universal translating machine (the ribosome) as it converts the message contained in mRNA into proteins.

1.4 Model Building in Biology

1.4.1 Models as Idealizations

In trying to understand the nature of life, the approach that we take in this book is the same as that followed by scientists for generations. We will approach each system with the goal of gaining some useful insight into its behavior by abstracting and simplifying the highly complex materials of living organisms so that we can apply simple, analytical models that have some predictive value. By definition, we cannot retain a complete atomic description of each macromolecule. Instead, we aim to select only the relevant properties of the macromolecule that speak to the particular aspect of that molecule's behavior that we are trying to address. Therefore, it would be inaccurate to say that we have in mind "a" simple model for a complex macromolecule such as DNA. Instead, we will use a suite of simple models that can be thought of as projections of the complex, multifaceted reality of the DNA molecule into a variety of simpler DNA maps. In the same way, any useful map of a territory emphasizes only the features that will be most relevant to the person who will be using the map. A classic example of the extraction and elegant presentation of only the most useful features is Harry Beck's famous map of the London Underground system, which shows the

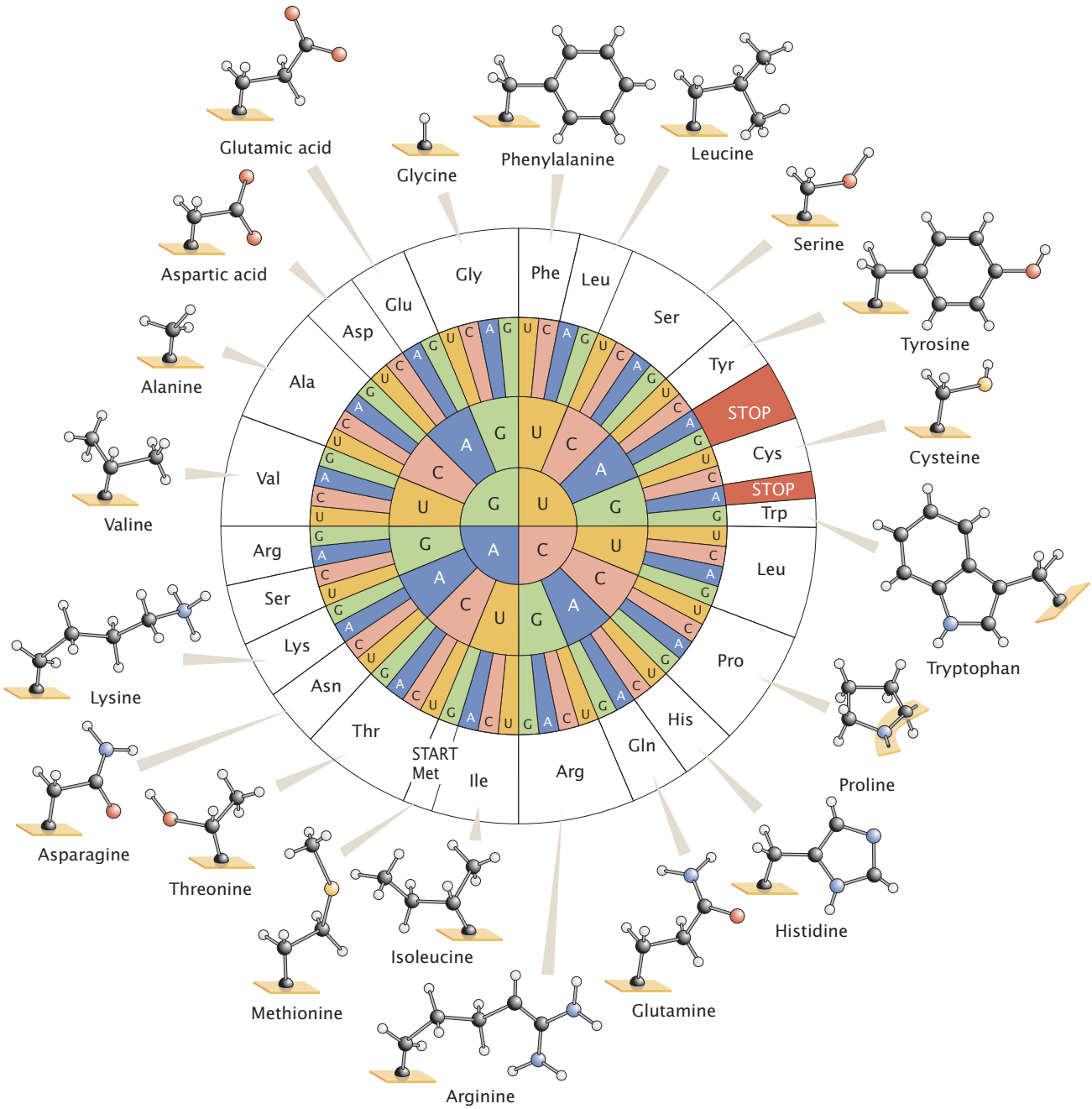


Figure 1.4: Genetic code. In this schematic representation, the first nucleotide in a coding triplet is shown at the center of the ring, the second nucleotide in the middle colored ring, and the third nucleotide in the outer colored ring. In this representation of the genetic code, the four bases are adenine (A), cytosine (C), guanine (G), and uracil (U). Uracil is structurally very similar to thymine (T), and is used instead of thymine in messenger RNA. The amino acids corresponding to each group of triplets are illustrated with their names (outer ring) and atomic structures. Two amino acids, tryptophan and methionine, are encoded by only a single triplet, whereas others, including serine, leucine, and arginine, are encoded by up to six. Three codons do not code for any amino acid and are recognized as stop signals. The unique codon for methionine, AUG, is typically used to initiate protein synthesis.

stations and connecting lines with an idealized geometry that does not accurately represent the complex and rather frustrating topography of London's streets. Although certain features of London's geography are deliberately ignored or misrepresented, such as the actual physical distances between stations, the London Underground map shows the user exactly enough information about the connectivity of the transport system to go anywhere he wishes. In this section, we will

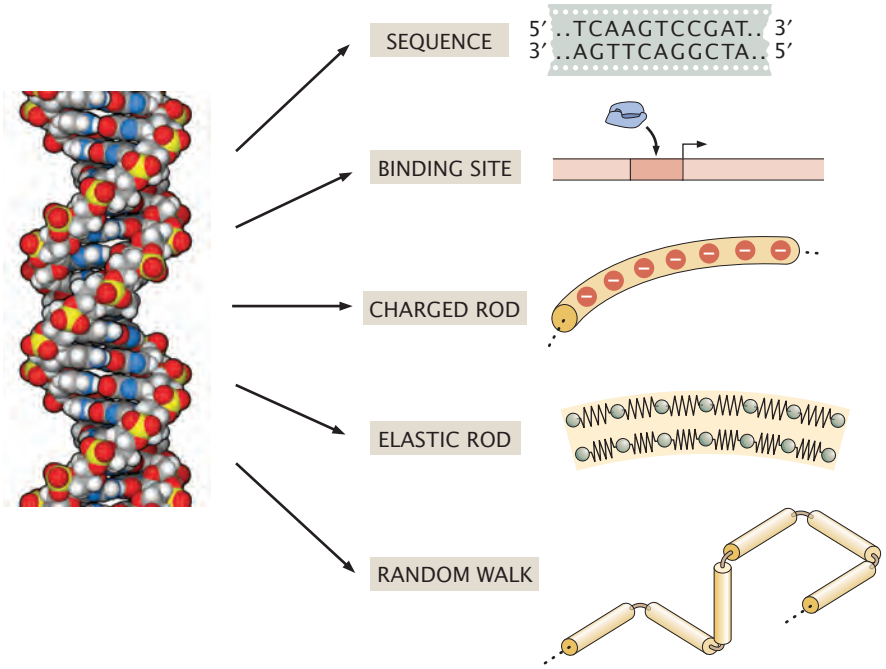


Figure 1.5: Idealizations of DNA. DNA can be thought of as a sequence of base pairs, as a series of binding sites, as a charged rod, as an elastic rod, or as a freely jointed polymer arranged in a random walk, depending upon the problem of interest.

summarize the kinds of physical properties of biological materials that we may wish to consider in isolation. In Section 1.5, we will list the kinds of fundamental physical models that will be applied throughout the book.

Biological Stuff Can Be Idealized Using Many Different Physical Models

Often, the essence of building a useful and enlightening model is figuring out what features of the problem can be set aside or ignored. Throughout the book, we will come to view molecules and cells from many different angles. An example of this diversity of outlook is shown in Figure 1.5. The nature of the simplified space, and therefore the nature of the projected DNA model, depends upon the specific question that we are asking. Figure 1.5 shows five different projections for the meaning of DNA that we will use repeatedly throughout the book.

DNA is one of the most important molecules, both because it is the carrier of genetic information and because of its iconic role in the history of molecular biology. There is a polarity to each strand of this helical polymeric molecule; that is, the two ends of the strand are structurally and chemically distinct, and are called the 5′ and 3′ ends. In addition, there is heterogeneity along their length, since the chemical identity of the bases from one nucleotide to the next can be different. It is often the informational content of the DNA molecule that garners attention. Indeed, the now-routine sequencing of whole genomes of organisms ranging from bacteria to humans permits us to forget the molecular details of the DNA molecule and to focus just on the sequence of A's, T's, G's, and C's that determines its information content. This representation will be most useful, for example, when we are trying to discern the history of mutational events that have given rise to diversification of species or when we are trying to calculate the information content of genomes. The second representation shown in Figure 1.5 abstracts variations in DNA sequence

in large blocks where one particular stretch of sequence (shown in the figure as the more darkly colored block) represents a portion of sequence that has particularly strong affinity for a protein binding partner, such as the RNA polymerase enzyme that copies the sequence encoded by the DNA into a molecule of mRNA. Both these first two representations draw attention to dissimilarities in detailed chemical structure within the DNA molecule. However, sometimes we will consider DNA simply as a physical entity within the cell where the specific nature of its sequence is not relevant. Instead, other properties of the molecule will be emphasized, such as its net distribution of charge, bending elasticity, or behavior as a long polymer chain subject to thermal undulations. Schematics of models emphasizing these physical properties of DNA are also shown in Figure 1.5.

Proteins are subject to many different idealizations as well. Just as nucleic acids can be thought of as sequences of nucleotides, we can think of proteins in terms of the linear sequence of amino acids of which they are made. The second great polymer language gives rise to proteins, which are considered to provide most of the important biological functions in a cell, including the ability to catalyze specific chemical reactions. The detailed atomic structure of proteins is substantially more complex than that of DNA, both because there are 20 rather than 4 fundamental subunits and because the amino acid chain can form a wider variety of compact three-dimensional structures than can the DNA double helix. Even so, we will extract abstract representations of proteins emphasizing only one feature or another as shown in Figure 1.6 in much the same way we did for DNA.

A protein can be represented as a simple linear sequence of amino acids. In cases where this representation contains too much detail, we will elect to simplify further by grouping together all hydrophobic amino acids with a common designation H and all polar amino acids with a common designation P. In these simple models, the physical differences between hydrophobic and polar amino acids (that is, those that are like oil versus those that are like water) affect the nature of their interactions with water molecules as well as with each other and thereby dictate the ways that the amino acid chain can fold up to make the functional protein. Similarly, we can think of the three-dimensional folded structure of proteins at different levels of abstraction depending on the question under consideration. A common simplification for protein structure is to imagine the protein as a series of cylinders (α-helices) and ribbons (β-strands) that are connected to one another in a defined way. This representation ignores much of the chemical complexity and atomic-level structure of the protein but is useful for considering the basic structural elements. An even more extreme representation that we will find to be useful for thinking about protein folding is a class of lattice models of compact polymers in which the amino acids are only permitted to occupy a regular array of positions in space.

Fundamentally, the amino acid sequence and folded structure of the protein serve a biological purpose expressed in protein function or activity, and we will also develop classes of models that refer only to the activity of the molecule. For example, the vast majority of proteins in the cell are capable of forming specific binding interactions with one another or with other biomolecules; these binding events are usually important to the protein's biological function. When modeling these kinds of interactions, we will envision proteins as "receptors" carrying binding sites that may be occupied or unoccupied by a binding partner, called a "ligand," and our description will make no further

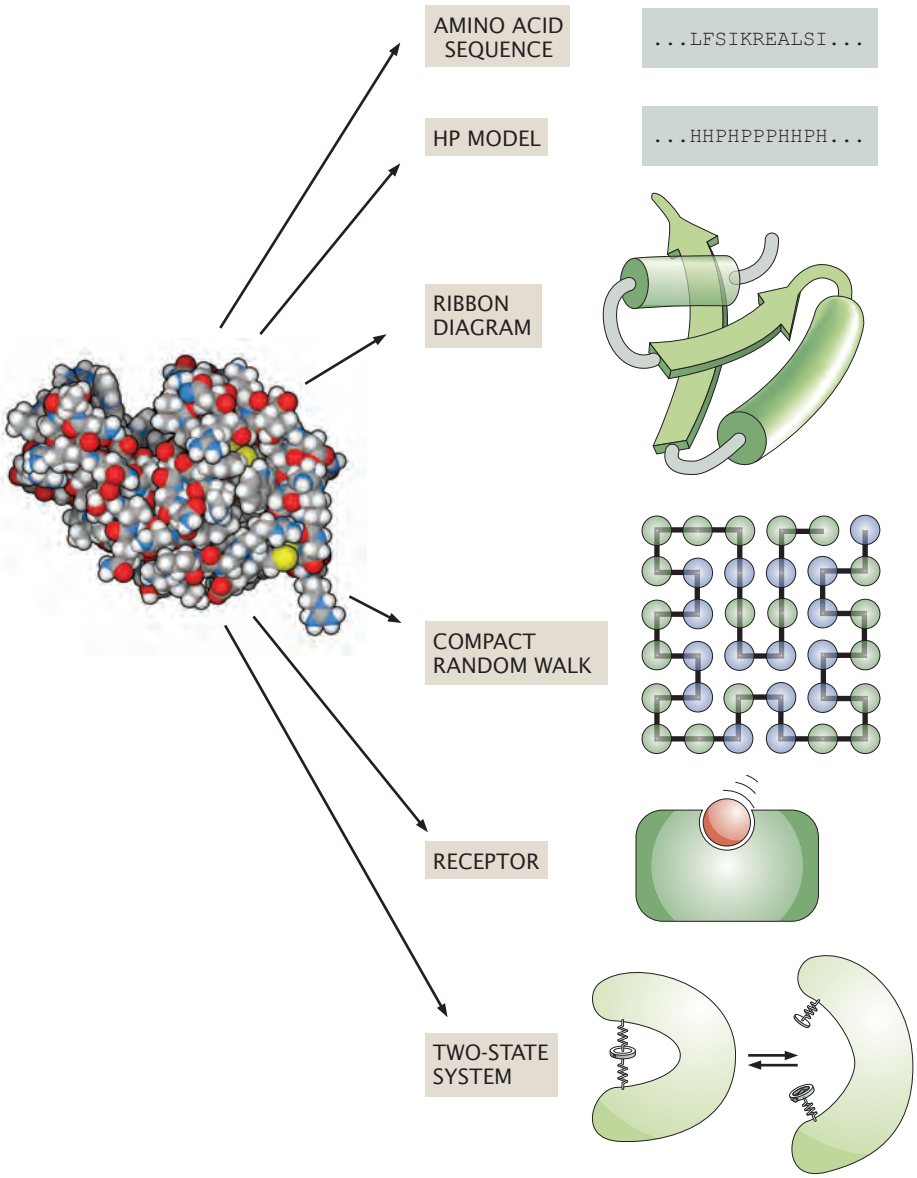


Figure 1.6: Idealizations of protein. Proteins can be thought of as a particular sequence of amino acids, as a simplified sequence reporting only the hydrophobic (H, oil-like) or polar (P, water-like) chemical character of the amino acids, as a collection of connected ribbons and cylinders, as a compact polymer on a lattice, as a binding platform for ligands, or as a two-state system capable of interconverting between different functional forms.

reference to the protein's complex internal structure. For some models, we will focus on changes in activity of proteins, considering them as simple two-state systems that can interconvert between two different functional forms. For example, an enzyme that catalyzes a particular biochemical reaction may exist in an "active" or "inactive" state, depending on the presence of ligands or other modifications to its structure.

By performing these simplifications and abstractions, we are not intending to deny the extraordinary complexity of these molecules and are fully cognizant that amino acid sequence, folding and compaction properties, and conformational changes between states dictate binding affinity, and it is an artificial simplification to consider these properties in isolation from one another. Nonetheless, it is generally useful to consider one single level of description at a time when trying to gain intuition from simple estimates and models, which will serve as our primary emphasis.

We will extend this approach beyond individual macromolecules to large-scale assemblies of macromolecules such as cell membranes

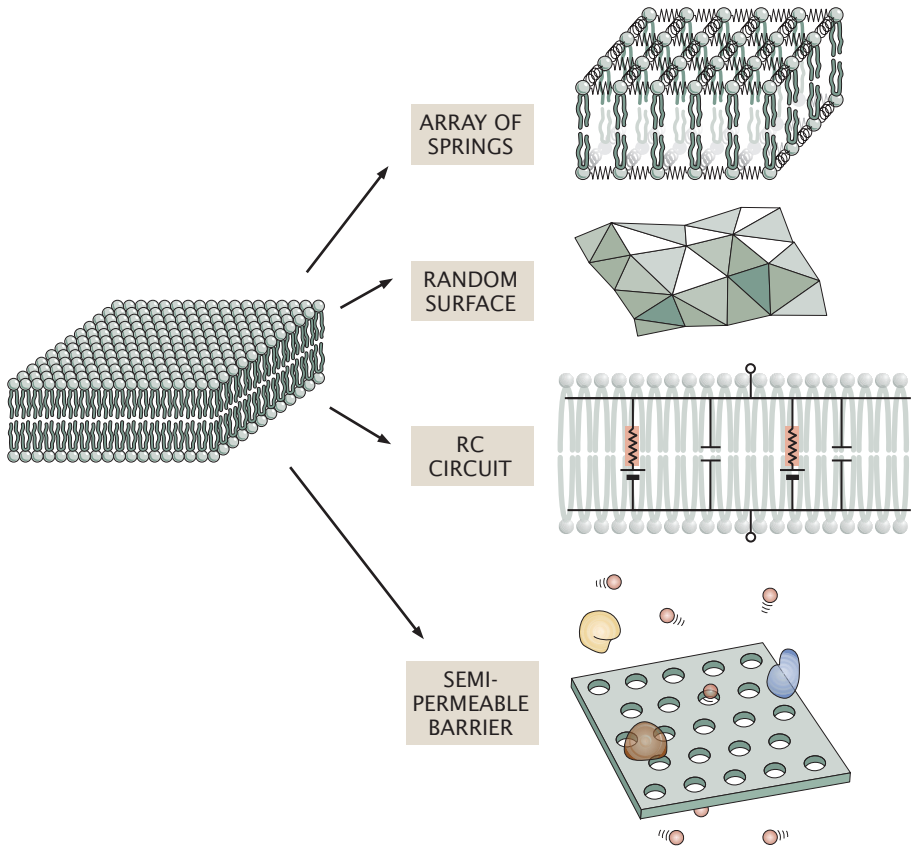


Figure 1.7: Idealization of membranes. A membrane can be modeled as an elastic object that deforms in response to force, as a random surface fluctuating as a result of collisions with the molecules in the surrounding medium, as an electrical circuit element, or as a barrier with selective permeability.

made up of millions of individual lipid and protein molecules. Some of the different ways in which we will think about biological membranes are shown in Figure 1.7. Just as with individual macromolecules, these large-scale assemblies can be usefully characterized by focusing on one aspect of their physical behavior at a time. For example, membranes can be thought of as bendable, springy elastic sheets, as random surfaces, or even as arrays of electrical elements such as resistors and capacitors. In their interactions with other molecules, membranes often serve as selective barriers that will enable some molecular species to cross while blocking others. Depending upon the particular question being asked, we will exploit different idealizations of the membrane.

Audaciously, we will bring these same idealization techniques to bear even on living cells, as shown in Figure 1.8. Although a living cell is exceedingly complex, we will nonetheless choose to extract one physical property of the cell at a time so that we can gain insight from our simple approach. For example, the bacterium *Escherichia coli*, a beloved experimental organism, can for some purposes be thought of as an object carrying an array of protein receptors on its surface. For other purposes, for example when we think of how a bacterium swims through water, we will think of it as an elastohydrodynamic object, that is, a mechanical object that can bend and can interact with the flow of water. The physical properties of the path it follows through water can be further abstracted by considering the cell's large-scale motion in the context of a biased random walk, ignoring the hydrodynamic details that actually enable that motion. Finally, because much biological research focuses on how cells alter the expression of their polymer languages in response to changing conditions, we

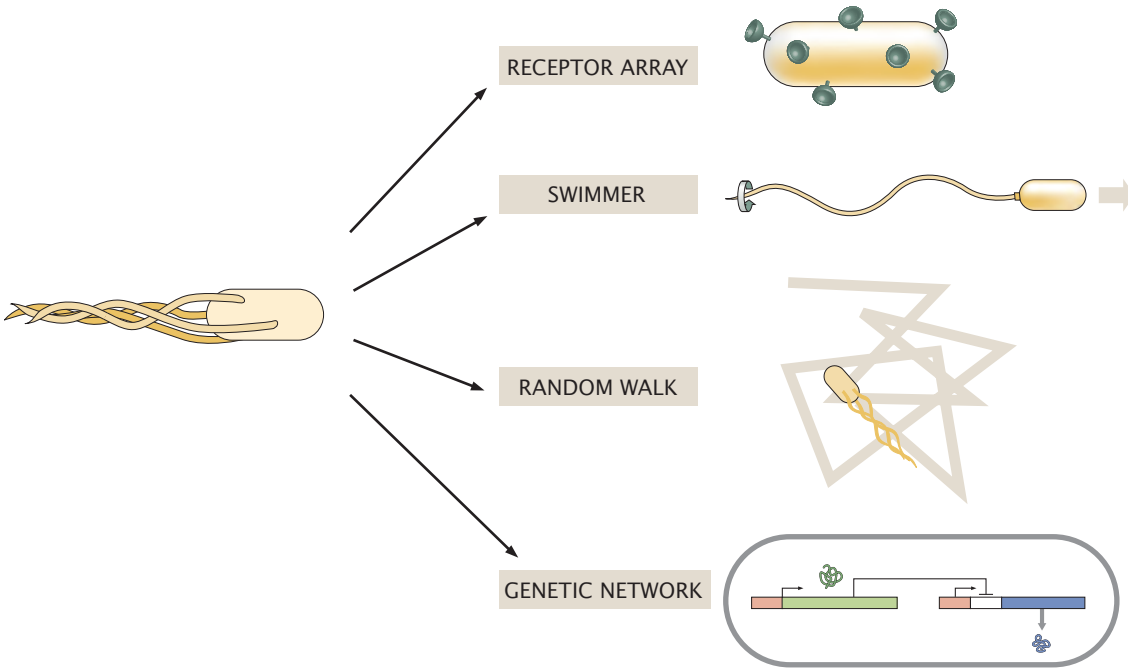


Figure 1.8: *E. coli* idealization. The cell can be modeled as an array of receptors for a ligand of interest, as an elastohydrodynamic object, as a biased random walker moving through water, or as an information processing device characterized by a series of genetic networks.

will frequently find it useful to consider the cell as an information processing device, for example as a network governing the flow of information in the form of gene expression. These descriptions are not mutually exclusive, nor are any of them comprehensive. They are again deliberate abstractions made for practical reasons.

Although so far we have emphasized the utility of abstract projections centering on a single physical property for descriptions of biological entities, the same approach is actually extremely useful in consideration of other systems such as solutions of charged ions in water. Various abstractions of the watery medium of life are shown in Figure 1.9. Sometimes we will pretend that a solution is a regular lattice. For example, if we are interested in ligands in solution and their tendency to bind to a receptor, we adopt a picture of the solution in which the ligands are only permitted to occupy specific discrete positions. Although this is an extreme approximation, it is nonetheless immensely valuable for calculations involving chemical equilibrium and remarkably gives accurate quantitative predictions. As the watery medium interacts with the living creatures it contains, sometimes its hydrodynamic flow properties (or viscous properties) are of most interest. When considering rates of chemical reactions taking place in water, we will find it useful to think of water as the seat of diffusive fluctuations. The macromolecules that exist in the watery medium of cells also interact with specific properties of water. For example, some of the chemical groups in proteins are polar or "hydrophilic" (water-loving), meaning that they are able to form hydrogen bonds with water, while other groups are oil-like and repel water. In the folded protein, the oil-like portions tend to cluster on the inside. In contrast, charged molecules on the surface can interact with the dielectric character of water molecules, which are neutral overall but do exhibit a slight charge separation.

Each of these representations will resurface multiple times throughout the book in different contexts related to different biological questions. Although none of them can give a complete understanding

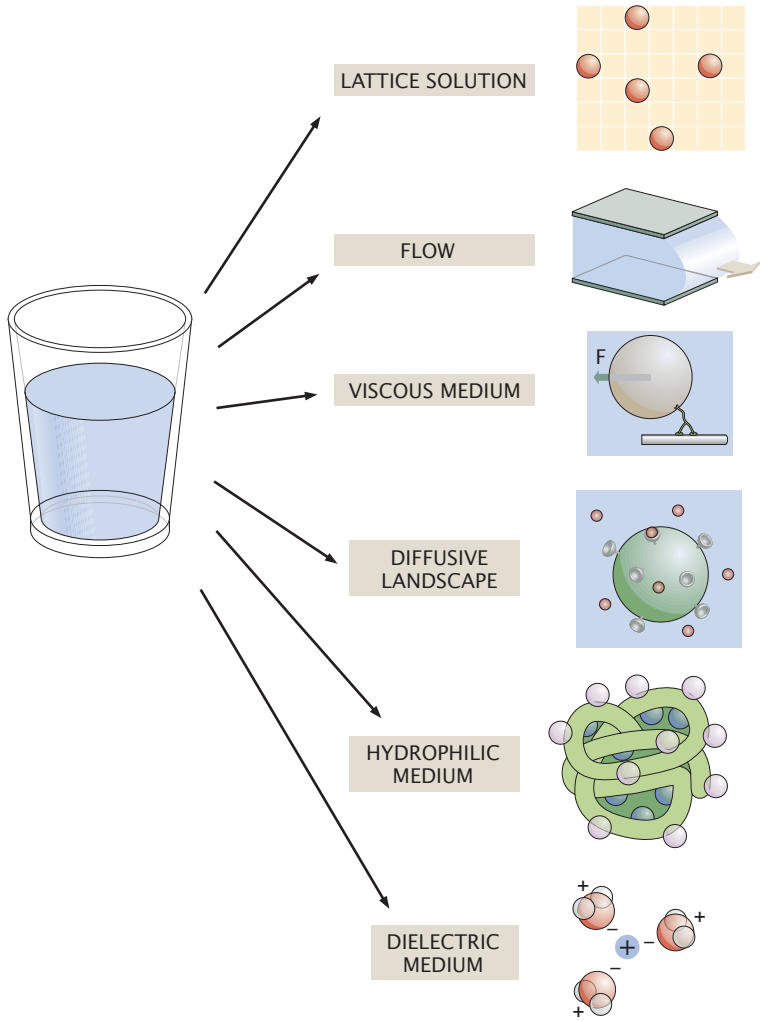


Figure 1.9: Idealization of a solution. A solution can be represented as a lattice of discrete positions where solutes might exist, as a fluid with a mean flow, as a viscous medium that exerts a drag force on objects moving through it, as a fluctuating environment that induces random motion of macromolecules and ligands, as a hydrophilic medium that readily dissolves polar molecules but repels oil-like molecules, or as a dielectric medium that is a poor electrical conductor but can support electrostatic fields.

of the behavior of living cells, each of them serves to provide quantitative insight into some aspect of life and, taken as a whole, they can begin to project a more realistic image of cells than any single abstraction can do alone.

1.4.2 Cartoons and Models

Biological Cartoons Select Those Features of the Problem Thought to Be Essential

We have argued that the art of model building ultimately reflects a tasteful separation of that which is essential for understanding a given phenomenon from that which is not. As is true of all sciences, the history of experimental progress in biology has always involved researchers constructing explicit conceptual models to help them make sense of their data. In some cases, the theoretical framework developed for biological systems is of precisely the mathematical character that will be presented throughout the book, such as that illustrated in the use of thermodynamics and statistical mechanics to understand biochemical reactions or in the development of the theory of the action potential. More commonly, much of the important modeling of biological systems has been in the form of visual schematics

that illustrate the most essential features in a given biological process and how they interact. In fact, the act of drawing the type of cartoon found in molecular biology textbooks or research papers reflects an important type of conceptual model building and either implicitly or explicitly reflects choices about which features of the problem are really important and which can be ignored.

As an example of the model building that takes place in constructing biological cartoons, consider the structure of one of the most beautiful and intriguing eukaryotic organelles, namely, the mitochondrion. Figures 1.10(A) and (B) show a planar section of a mitochondrion as obtained using electron microscopy and part of a three-dimensional structure resulting from cryo-electron microscopy. Figure 1.10(C) shows two cartoons that describe the structure of this important organelle. This pair of drawings emphasizes the central importance of the membranes which bound the organelle and which separate the interior into different, relatively isolated compartments. Though the three-dimensional model represents more detail regarding the precise geometry of the membranes, the essential *conceptual* elements are the same, namely, (a) the mitochondria are closed, membrane-bound organelles and (b) the inner membrane has a complex folded structure. Furthermore, the folds of the inner membrane greatly increase the total amount of surface area of this membrane and therefore increase the area available for the ATP-synthesizing machinery that lies within this membrane, as will be discussed below.

The other cartoons in this figure also represent the mitochondrion, but in very different ways, focusing on specific conceptual elements involved in different aspects of mitochondrial function and behavior. For example, Figure 1.10(D) reveals the surprising existence of mitochondrial DNA, which is a vestige of the microbial origins of this important organelle. Interestingly, these organelles contain their very own DNA genomes, completely distinct from the genome found in the cell's nucleus. As indicated in the diagram, the genome is a circular piece of DNA (like the genome found in bacterial cells, but very different from eukaryotic chromosomes, which are linear pieces of DNA). Each mitochondrion contains several copies of the genome, and they are distributed between the two daughters when the mitochondrion divides.

The diagram provided in Figure 1.10(E) shows a highly magnified segment of the mitochondrial inner membrane, illustrating the proteins that cross the membrane and that are involved in the generation of ATP. A central function of mitochondria in eukaryotic cells is to convert one form of chemical energy, high-energy electrons in the compound NADH, into a different form of chemical energy, a phosphodiester bond in ATP (these different forms of chemical energy and their significance will be discussed in more detail in Chapter 5). The process of performing this conversion requires the cooperative action of several different large protein complexes that span the inner mitochondrial membrane. Sequential transfer of electrons from one protein complex to another results in the transport of hydrogen ions from the mitochondrial matrix across the inner membrane into the intermembrane space, establishing a gradient with an excess of hydrogen ions on the outside. In the final step, the hydrogen ions are allowed to travel back down their concentration gradient into the mitochondrial matrix, by passing through a remarkable protein machine called ATP synthase, which we will revisit several times throughout the rest of the book. ATP synthase catalyzes the construction of a molecule of ATP by combining a molecule of ADP and an

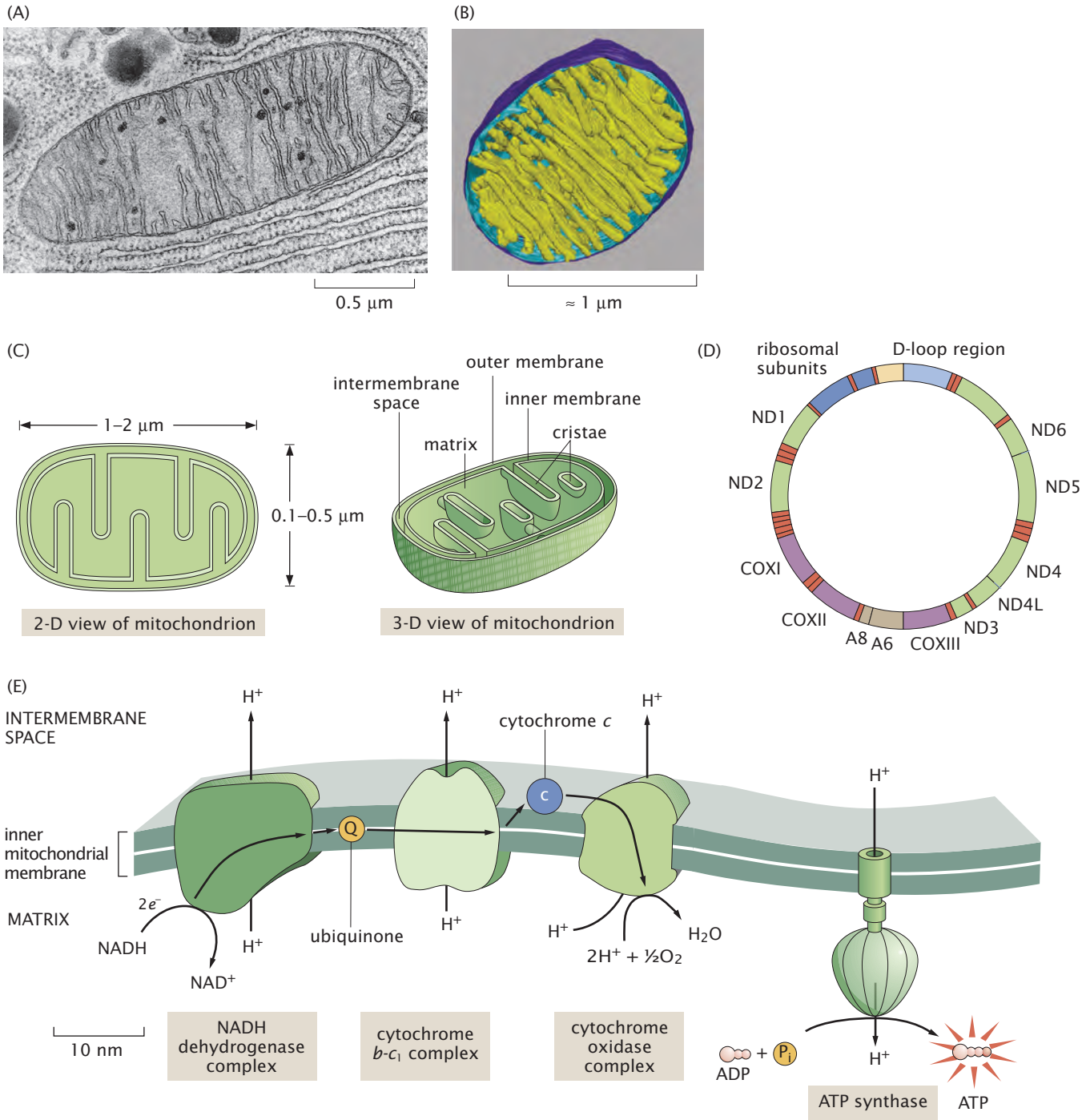


Figure 1.10: Several different ways of illustrating mitochondria. (A) Thin-section electron micrograph showing a mitochondrion found in a cell within the pancreas of a bat. (B) Cryo-electron microscopy reconstruction of the structure of a mitochondrion, shown cut in half to reveal the internal structure. (C) Diagrams showing the arrangement of membranes dividing the mitochondrion into distinct compartments. While the outer membrane forms a smooth capsule, the inner membrane is convoluted to form a series of cristae. Distinct sets of proteins are found in the matrix (inside the inner membrane) and in the intermembrane space (between the inner and outer membranes). (D) Schematic of mitochondrial DNA. (E) Schematic illustration of the major proteins involved in the electron transport chain and in ATP synthesis in the inner membrane of the mitochondrion. The overall purpose of the complex chemical reactions carried out by this series of proteins is to catalyze the creation of the important energy carrier molecule ATP. (A, from D. W. Fawcett, The Cell, Its Organelles and Inclusions: An Atlas of Fine Structure, W. B. Saunders, 1966; B, courtesy of T. Frey; D, adapted from S. DiMauro and E. A. Schon, *N. Engl. J. Med.* 348:2656, 2003; E, adapted from B. Alberts et al., Molecular Biology of the Cell, 5th ed., Garland Science, 2008.)

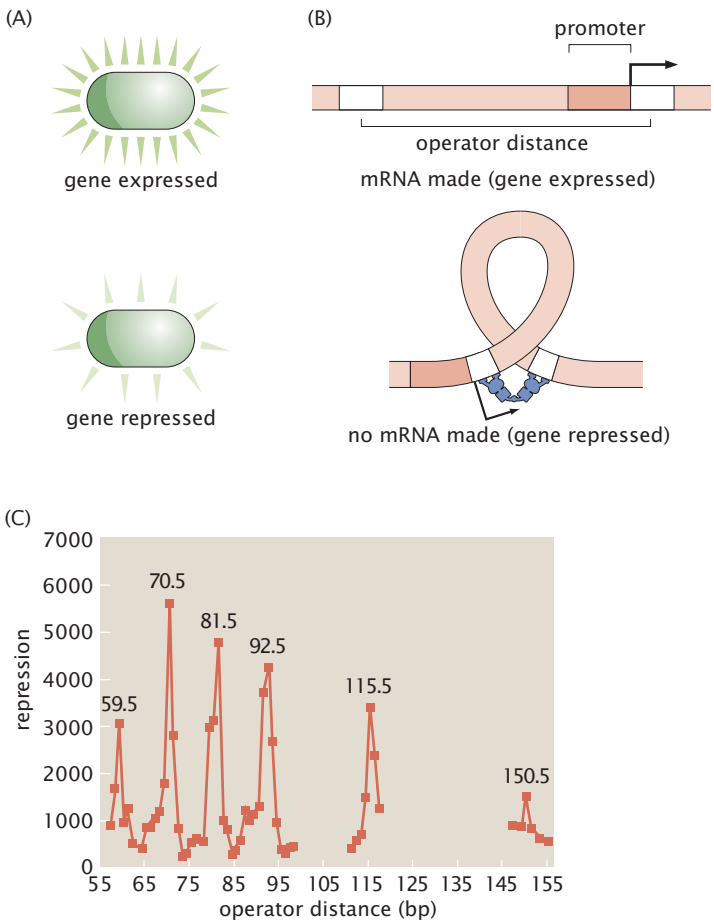


Figure 1.11: Qualitative and quantitative illustrations of gene repression. (A) Bacteria containing a gene whose expression level is regulated by the Lac repressor protein may exist in different states. When repression is low, the gene is expressed, and large amounts of the gene product are present. If the gene being regulated encodes a fluorescent protein, the bacteria will glow brightly when the gene is expressed. In contrast, when the gene is repressed, relatively little fluorescent protein will be produced, and the bacteria will glow dimly. (B) Schematic showing events at the level of the DNA during gene expression and repression. During repression, the Lac repressor protein binds to two distinct operator sites on the DNA (white boxes) and forces the DNA to form a loop. This prevents binding of the protein machine that copies the DNA into mRNA. When no mRNA is produced, the gene is repressed. (C) Quantitative measurements of the strength of gene repression as a function of the exact distance (expressed in base pairs) between the two operator sites. (C, data from J. Müller et al., *J. Mol. Biol.* 257:21, 1996.)

inorganic phosphate ion. The precise molecular details of this process are fascinating and complex; our intention here is merely to point out that this kind of cartoon represents an abstraction of the mitochondrion that emphasizes a completely different set of components and concepts than the membrane-focused cartoons shown above.

Each of these different kinds of cartoon is a valid and informative model of certain aspects of the mitochondrion, but they serve essentially nonoverlapping conceptual purposes. Furthermore, each cartoon represents the culmination of thousands of separate experiments covering decades of hard-won, detailed knowledge; experts have carefully sifted the raw data and made difficult decisions about which details are important and which are dispensable in considering each of these aspects of mitochondrial reality. Throughout the rest of this book, we will take advantage of the hard work of conceptualization that has already gone into constructing these kinds of diagrams. Our usual goal will be to take these existing biological models one step further, by rendering them into a mathematical form.

Quantitative Models Can Be Built by Mathematicizing the Cartoons

One of the key mantras of the book is that the emergence of quantitative data in biology generates a demand that the cartoons of molecular biology be mathematicized. In particular, in some cases only by constructing a mathematical model of a given biological problem can our

theoretical description of the problem be put on the same footing as the data itself, that is, quantitatively. A concrete example that will arise in Chapter 19 is the measurement of gene expression as a function of the distance between two binding sites on DNA as shown in Figure 1.11. The mechanistic basis of these data is that a certain protein (Lac repressor) binds simultaneously at two sites on the DNA and forms a loop of the DNA between the two binding sites. When the Lac repressor protein is bound to the DNA, it prevents the binding of the protein machine that copies the DNA into RNA. The strength of this effect (repression) can be measured quantitatively. The experimental data show that the amount of repression depends in a complex way on the precise distance between the two binding sites.

Our argument is that a cartoon-level model of this process, while instructive, provides no quantitative basis for responding to these data. For example, why are there clear periodic peaks in the data? Why does the height of the peaks first increase and then decrease? In Chapter 19, using the framework of statistical mechanics, we will derive a mathematical model for this process that attempts to predict the shapes and sizes of these features in the data while at the same time highlighting some features of DNA mechanics that remain unclear. Further, by constructing a physical model in mathematical terms of this process, new experiments are suggested that sharpen our understanding of transcriptional regulation. In nearly every chapter, we will repeat this same basic motif: some intriguing quantitative observation on a biological system is described first in terms of a cartoon, which is then recast in mathematical form. Once the mathematical version of the model is in hand, we use the model to examine previous data and to suggest new experiments.

As a result of examples like this (and many others to be seen later), one of the central arguments of the present chapter is that there are many cartoons in molecular biology and biochemistry that have served as conceptual models of a wide variety of phenomena and that reflect detailed understanding of these systems. To make these models quantitatively predictive, they need to be cast in mathematical form; this in keeping with our battle cry that quantitative data demand quantitative models. As a result, one way of viewing the chapters that follow is as an attempt to show how the cartoons of molecular biology and biochemistry have been and can be "translated into a mathematical form."

1.5 Quantitative Models and the Power of Idealization

Quantitative models of the natural world are at the very heart of physics and in the remainder of this book we will illustrate how pervasive these ideas have become in biology as well. This approach is characterized by a rich interplay between theory and experiment in which the consequences of a given model are explored experimentally. In return, novel experimental results beckon for the development of new theories. To illustrate the way in which particular key models are recycled again and again in surprisingly varied contexts, we consider one of the most fundamental physical models in all of science, namely, the simple spring (also known as the harmonic oscillator).

1.5.1 On the Springiness of Stuff

The concept of springiness arises from the confluence of the very important mathematical idea of a Taylor series (explained in detail in "The Math Behind the Models" on p. 215) and an allied physical idea known as Hooke's law (the same Hooke who ushered in the use of microscopy in biology and coined the term "cell"). These ideas will be developed in detail in Chapter 5 and for now we content ourselves with the conceptual framework. The fundamental mathematical idea shared by all "springs" is that the potential energy for almost any system subjected to small displacements from equilibrium is well approximated by a quadratic function of the displacement. Mathematically, we can write this as

$$\text{energy} = \tfrac{1}{2}kx^2. \tag{1.1}$$

What this equation states is that the potential energy increases as the square of the displacement x away from the equilibrium position. The "stiffness" k is a measure of how costly it is to move away from equilibrium and reflects the material properties of the spring itself.

Another way of thinking about springs is that they are characterized by a restoring force that is proportional to how far the spring has been displaced from its equilibrium position. Mathematically, this idea is embodied in the equation

$$F = -kx, \tag{1.2}$$

where F is the restoring force, x is the displacement of the spring from its equilibrium position, and k is the so-called spring constant (the stiffness). The minus sign signals that the force is a restoring one towards $x = 0$, which designates the equilibrium position. This result is known as Hooke's law. When stated in this way, this kind of physical model gives the impression of an abstract example of masses on frictionless tables with pulleys and springs. But empirically, scientists have found that precisely this mathematical model arises naturally in many different practical contexts, as shown in Figure 1.12.

In biology, the simple spring comes cloaked in many different disguises, as we will see throughout the book. From a technological perspective, many of the key single-molecule techniques used to study macromolecules and their assemblies invoke this description. Both optical tweezers and the atomic force microscope can be mapped precisely and unequivocally onto spring problems. However, as shown in Figure 1.12, there are many other unexpected examples in which the problem can be recast in a way that is mathematically equivalent to the simple spring problem introduced above. Simple spring models have been invoked when considering the bending of DNA, the beating of the flagellum on a swimming sperm, and even the fluctuations of membranes at the cell surface. As these are all mechanical processes, it is fairly straightforward to see how the idea of a spring can apply to them. But the real power of the simple spring model will be revealed when we see how it can be directly applied to nonmechanical problems. For example, when we discuss biochemical reactions in a cell, we will imagine molecules moving on an "energy landscape" where they accumulate in "potential wells." The mathematics of simple springs will help us understand the rates that govern these biochemical reactions. The ideas of harmonic oscillators and potential wells will also illuminate our exploration of protein conformational changes. Even more abstractly, we will see that changes in

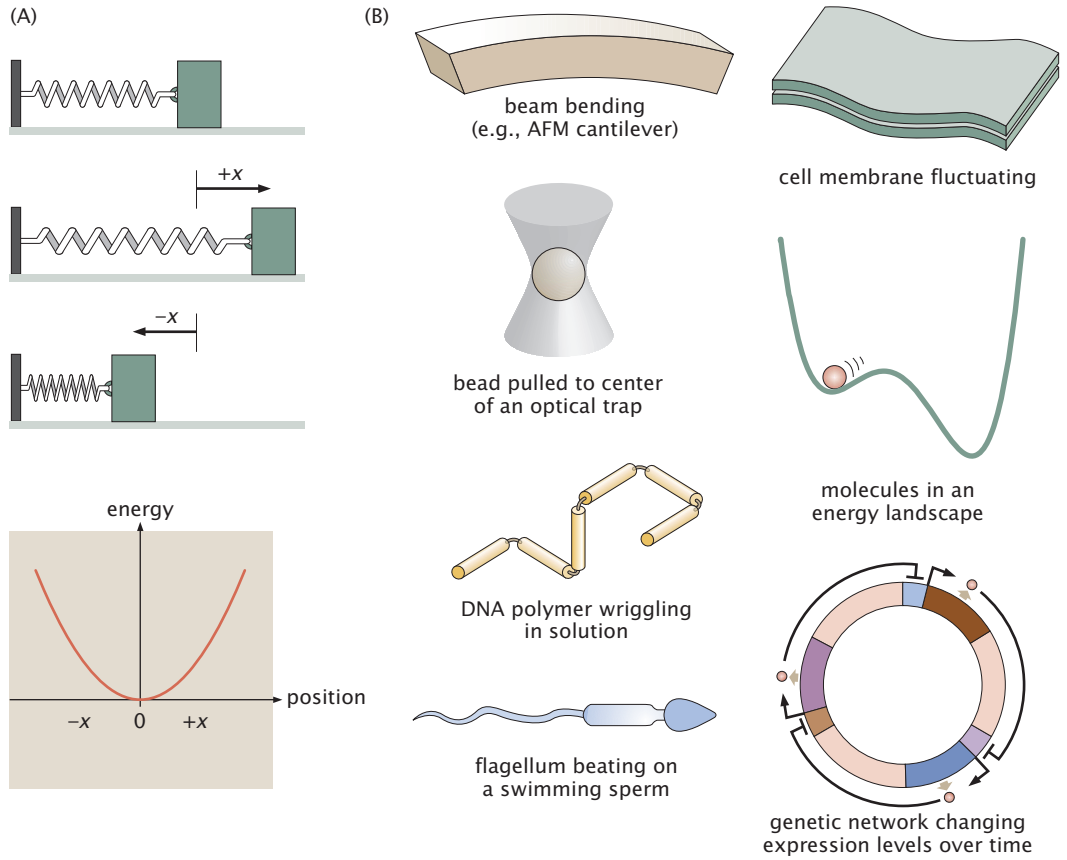


Figure 1.12: Gallery showing some of the different ways in which the idea of a spring is used in biological model building. (A) Classical physics representation of the harmonic oscillator. A simple spring is attached at one end to an immovable wall, and at the other end to a block-shaped mass that can slide on a frictionless table. The position of the mass changes as the spring extends and compresses. The graph at the bottom shows the energy of the spring as a function of the mass position. When the spring is relaxed at its resting length, the energy is at a minimum. (B) Multiple manifestations of the simple harmonic oscillator relevant to biological systems.

gene expression over time for certain interesting kinds of genetic networks can also be surprisingly well described using the mathematics of simple harmonic oscillators.

1.5.2 The Toolbox of Fundamental Physical Models

The overall strategy of this book will be to apply this kind of fundamental physical thinking to the widest possible range of biological problems. Through careful analysis of the quantitative assumptions that go into the analysis of biological data, it has become clear that a mere handful of fundamental physical models are often sufficient to provide a rigorous framework for interpretation of many kinds of quantitative biological data.

Beyond the harmonic oscillator introduced above, we believe that there are fewer than 10 other key physical concepts that require mastery. A scientist who understands and appreciates the applications of these concepts in biology is afforded a powerful framework for zeroing in on the most important, unexplained open questions. It is not our contention that this family of broad reaching physical models is sufficient to explain biological phenomena, far from it. Rather, we contend that any observations that can be completely quantitatively accounted for within the framework of one of these models require no further speculation of unidentified mechanism, while data that cannot be embraced within this physical framework cry out for more investigation and exploration.

Our choice of this set of concepts is purely utilitarian. We are not attempting to write a book surveying all of physics, or surveying all of

biology. Modern, cutting-edge research at the interdisciplinary interface between biology and physics demands that scientists trained in one discipline develop a working knowledge of the other. Students currently training in these fields have the opportunity to explore both in detail and to understand their complementarity and interconnectedness. In this book, we have attempted to address all three audiences: physicists curious about biology, biologists curious about physics, and broad-minded learners ready to delve into both. Our hope is that these ideas will provide a functional foundation of some of the key physical concepts most directly relevant to biological inquiry. At the same time, we hope to provide enough examples of specific biological phenomena to allow a physical scientist to identify further important unsolved problems in biology.

The bulk of this book comprises detailed dissections and case studies of these fundamental concepts and concludes with chapters showing applications of several of these concepts simultaneously. As noted above, many individual chapters each focus on a single class of physical model. The key broad physical foundations are as follows:

- simple harmonic oscillator (Chapter 5)
- ideal gas and ideal solution models (Chapter 6)
- two-level systems and the Ising model (Chapter 7)
- random walks, entropy, and macromolecular structure (Chapter 8)
- Poisson–Boltzmann model of charges in solution (Chapter 9)
- elastic theory of one-dimensional rods and two-dimensional sheets (Chapters 10 and 11)
- Newtonian fluid model and the Navier–Stokes equation (Chapter 12)
- diffusion and random walks (Chapter 13)
- rate equation models of chemical kinetics (Chapter 15)

1.5.3 The Unifying Ideas of Biology

One of the most pleasing and powerful aspects of science is its ability to link together seemingly unrelated phenomena. Few examples from the physical sciences are more potent at making this point than the surprising unification of the age-old phenomenon of magnetism known from lodestones and the early days of compasses and navigation to the diverse manifestations of electricity such as the lightning that attends summer thunderstorms to the many and varied appearances of light such as rainbows or the apparent bending of a spoon in a glass of water. All of these phenomena and many more were brought under the same roof in the form of the modern theory of electromagnetism.

The toolbox of fundamental physical ideas introduced above focused on the way in which certain key ideas from physics have served to buckle together a huge number of observations in a way that is not just descriptive, but, more importantly, is predictive. But such unifying concepts are certainly not unique to the physical sciences. Biology has far-reaching ideas of its own. Just as it is true that a mere handful of fundamental physical models are often sufficient to provide a rigorous framework for interpretation of many kinds of quantitative biological data, there is a similar handful of fundamental biological ideas that serve as an overarching foundation for all that we think about in biology. Indeed, many earlier authors

(see the references at the end of the chapter) have provided various interesting perspectives on what one might call the "great ideas of biology."

On the second page of the nearly 1400 pages of the *Feynman Lectures on Physics*, Feynman says "If, in some cataclysm, all of scientific knowledge were to be destroyed, and only one sentence passed on to the next generation of creatures, what statement would contain the most information in the fewest words?" Feynman chooses the idea that "all things are made of atoms". However, a survey of scientists would not be unanimous on this point, and some fraction of them would come down in favor of something along the lines of "all living organisms are related through descent from a common ancestor through the process of evolution by natural selection." Though perhaps it has become hackneyed through overuse, the words of Dobzhansky remain both timely and true: "Nothing in biology makes sense except in the light of evolution." Indeed, there is a strong argument that evolution is biology's greatest idea and will leave its indelible stamp on nearly everything we say in the chapters that follow.

While Darwin and Wallace labored to find an understanding of the process of evolution, Mendel was hard at work in establishing the early ideas on the underlying mechanisms of inheritance, resulting in discoveries that founded another one of biology's uniquely great ideas, namely, genetics. One of the most remarkable adventures of the human mind has been the quest to relate Mendel's abstract particles of inheritance to the molecules of the cell, with the successful insight being the realization that DNA is the carrier of genetic information. History is often characterized by various ages and it is no exaggeration to say that we are now living in the DNA Age.

Another of the key ideas that is now largely taken for granted, but is a unifying concept all the same and was hard won in the making, is the cell theory, the insight that all living organisms are made up of cells and that individual cells are themselves the quantum unit of living matter. The biochemical processes that take place within cells also reveal a remarkable unity, whether we think of the common metabolic pathways shared by so many organisms or the fact that the molecular machinery that implements our shared genetic heritage is based upon a few common macromolecular assemblies such as polymerases and the ribosome.

The key broad biological foundations that will serve as a point of departure for much of what follows in the remainder of the book are as follows:

- the theory of evolution (Chapters 3, 4, 18, 20, and 21, and many others less explicitly)
- genetics and the nature of inheritance (Chapters 2–5 and 21)
- the cell theory (Chapters 2, 3, 11, and 14)
- the unity of biochemistry (Chapters 2, 3, 5, 6, and 15)

One of the features we find especially intriguing about physical biology is that this perspective yields a different view of the proximity of different biological phenomena. As illustrated in Figure 1.13, the way we think about problems depends greatly upon both the kinds of questions we are asking and on how we go about answering them. The development of a frog and a fly both raise questions that a developmental biologist would recognize as related: what determines the timing and synchronization of the developmental process,

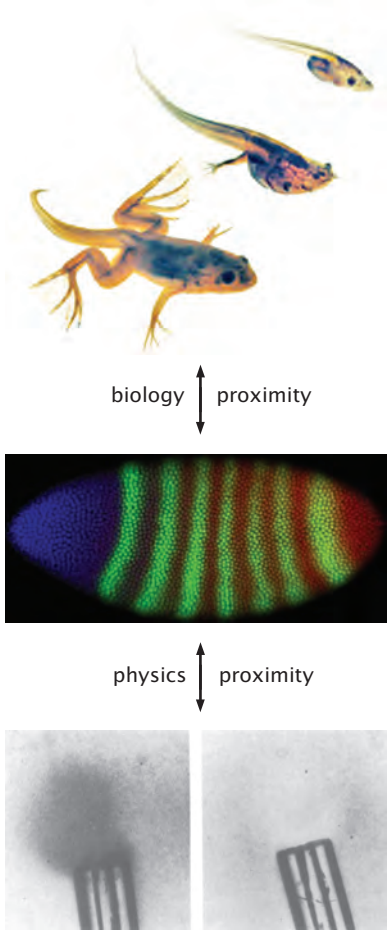


Figure 1.13: Proximity of ideas in cell biology and in physical biology. During the early development of the fruit fly *Drosophila* (middle), the eventual fate of the cells is determined by the patterned expression of a group of regulatory proteins (colored stripes). In a traditional organization of biological topics, the process of cell fate determination in *Drosophila* development might be discussed in the context of development of other animals such as the frog (top). However, the critical physical concepts that enable cell specification in the fly embryo involve processes such as diffusion and precise measurement of chemical concentrations processes that are also central to very different biological phenomena that have nothing to do with animal development. One of these phenomena is illustrated at the bottom, where on the left a group of bacteria is seen swarming toward the tip of a glass micropipette that is leaking a delicious amino acid (serine) into the medium. When the serine is switched for a toxin (phenol), as shown on the right, the bacteria quickly reverse course and swim away. In both the developing embryo and the swimming bacterium, cells measure the concentrations in their environment and make decisions as a result of similar physical measurements. (Top, courtesy of R. J. Denver, photo by David Bay; middle, adapted from E. Poustelnikova et al., *Bioinformatics* 20:2212, 2004; bottom, adapted from B. A. Rubik et al., *Proc. Natl Acad. Sci. USA* 75:2820, 1978.)

how is polarity established, etc.? On the other hand, pattern formation in the fly embryo and bacterial chemotaxis are conceptually linked when viewed as questions about the limits of chemical gradient detection. In the chapters that follow, we seek these kinds of unconventional statements of proximity with the ambition of using proximity in the physical biology sense to shed light on a host of different problems.

1.5.4 Mathematical Toolkit

One of the most surprising aspects of science is the way in which the same ideas show up again and again. For example, the random walk model does duty not only as a way to think about the diffusive motions of molecules in solution such as described in Figure 1.13 but also as a tool for describing the conformations of polymers. Thus far, we have argued for such unity and pervasiveness in the context of physics and biology. But as will be seen throughout the book, certain overarching mathematical ideas reveal themselves again and again as well. Indeed, this observation has been canonized in much the same way as Dhobzhansky's remark about evolution, in the title of another famous essay, this time by Eugene Wigner and entitled "The Unreasonable Effectiveness of Mathematics in the Natural Sciences."

One example of this unreasonable effectiveness has to do with our ability to assign numbers to our degree of belief when faced with

questions where there are many outcomes, such as whether a receptor will be bound by a ligand or not. Probability theory will serve as one of the mathematical cornerstones of what we have to say in coming chapters. In particular, we will repeatedly and often appeal to four key distributions: the binomial distribution, the Gaussian distribution, the Poisson distribution, and the exponential distribution, each of which will help us sharpen the way we think about many different biological questions.

However, we will need more than just the powerful calculus of uncertainty provided by the modern theory of probability and statistics. Ideas from the differential and integral calculus are also foundational in much of what we will say, especially through our use of differential equations. Often, we will want to know how the quantity of some particular species, such as the number of microtubules, varies in time. With our calculus toolkit, one of our favorite tools will concern the mathematics of superlatives, how do we find the largest or smallest values of some function or functional? These ideas will be explored in detail in Section 5.3.1 (p. 209).

Sometimes, the push for a strictly analytical treatment of the mathematics that arises in thinking about our models will be either too tedious or demanding. In such cases, we will turn to the amplification in personal mathematical power that is available to us all now as a result of computers. Indeed, the exploitation of such computation will serve as one of the centerpieces through the remainder of the book in the form of "Computational Explorations" in which we use Matlab to numerically explore our models.

1.5.5 The Role of Estimates

In this era of high-speed computation, the temptation to confront new problems in their full mathematical complexity is seductive. Nevertheless, the core philosophy of the present book advocates a spirit that is embodied in a story, perhaps apocryphal, of the demise of Archimedes. Plutarch tells us that Archimedes was deeply engaged in a calculation, which he was performing with a stick in the sand, at the time when the Roman army invaded Syracuse. When confronted by a Roman soldier and asked to identify himself, Archimedes told the soldier to wait until he was done with his calculation. The soldier replied by impaling Archimedes, bringing the life of one of the great scientists of history to an early and unjust end. Our reason for recounting this story is as a reminder of the mathematical character of many of the models and estimates that we will report on in this book (but hopefully with less dire personal consequences). In particular, we like to imagine that a model is of the proper level of sophistication to help build intuition and insight if, indeed, its consequences can be queried by calculations carried out with a stick in the sand, without recourse to sophisticated computers or detailed references.

Throughout the book, we will repeatedly make use of estimates to develop a feeling for the numbers associated with biological structures and biological processes. This philosophy follows directly from the biological tradition articulated by the pioneering geneticist Barbara McClintock, who emphasized the necessity of developing a "feeling for the organism," in her case the crop plant maize. We believe that a quantitative intuition is a critical component of having a feeling for the organism. Indeed, we think of model building as proceeding through a series of increasingly sophisticated steps starting with

simple estimates to develop a feel for the relevant magnitudes, followed by the development of toy models that can (usually) be solved analytically, followed in turn by models with an even higher degree of realism that might require numerical resolution.

The estimates that we advocate are all meant to be sufficiently simple that they can be done on the back of an envelope or written with a stick in the sand on a desert island. In keeping with this purpose, we will rarely use more than one significant digit in our estimates. One useful rule of thumb that we will adopt is the geometric mean rule (see *Guesstimation* by Weinstein and Adam for a more complete description). The idea is that when making an estimate where we are uncertain of the magnitudes, begin by attempting to guess a lower and upper bound for the magnitude of interest. With these two guesses in hand, multiply them together and take the square root of the result as your best guess for that poorly known quantity. This strategy is illustrated in Figure 1.14. What this scheme really amounts to is averaging the exponents rather than the quantities themselves.

Naturally, this introduces somewhat low precision, for example, an estimate may yield an answer of 100 molecules in a system when the precisely measured value is actually 42, but has the benefit of greater simplicity and clarity. In general we will consider an estimate to be successful if the result is within an order of magnitude of the measured value. Indeed, as shown in Figure 1.15, we will often resort to simple arithmetic rules in which we are intentionally numerically imprecise in the name of rapidly getting to an order-of-magnitude result. In cases where our estimates differ by more than an order of magnitude, it will be clear that we have made an incorrect assumption and we will use this as an opportunity to learn something more about the problem of interest.

The precept that animates much of the book is that we are closest to understanding a phenomenon, not when we unleash the full weight of a powerful numerical simulation of the problem at hand, but rather when we are engaged in developing a feel for the scaling of the quantities of interest and their associated numerical values.

In many cases, one of the most powerful tools that can be unleashed in order to understand the relevant numerical magnitudes in a given problem is known as *dimensional analysis*. Indeed, in many examples throughout the book, whether when talking about fluid flow using the Navier–Stokes equations or the nature of the genetic switch, we will write our equations in dimensionless form. One of the reasons that these dimensionless forms are so powerful is that they permit us to write the equations describing the phenomenon of interest in the "natural variables" of the problem. For example, when thinking about a circuit with a resistor and a capacitor, the natural time unit is given by $\tau = RC$, the product of the resistance and the capacitance. In thinking about the establishment of morphogen gradients during anterior–posterior patterning in flies, the fundamental length scale is given by $\lambda = \sqrt{D\tau}$, where D is the diffusion constant of the morphogen and τ is its degradation time. Further, when we are trying to estimate some quantity of interest, a powerful trick is to simply use the relevant constants describing that problem to construct a quantity with the dimensions of the quantity we are trying to estimate.

In the morphogen gradient example, if we assume that the length scale λ over which the pattern changes significantly is determined by the diffusion constant of the morphogen and its mean lifetime, then the dimensions of these two quantities determine the formula for the length scale λ up to a numerical constant. To be specific, the diffusion

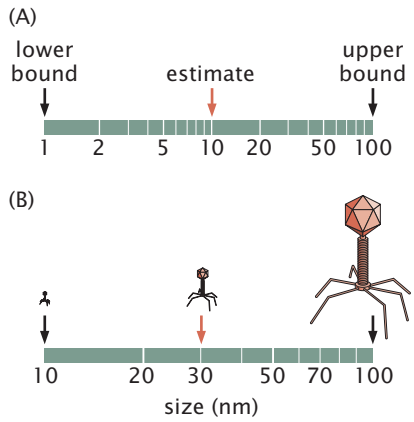


Figure 1.14: Geometric mean rule for estimates. (A) For cases in which we don't know the relevant magnitude, a powerful scheme is to guess an upper and lower bound and to take their geometric mean $\sqrt{x_{\text{lower}} x_{\text{upper}}}$ as our best guess. (B) Application of the geometric mean rule to guess the size of bacteriophages that are larger than 10 nm and smaller than 100 nm. The geometric mean guess is for a typical size of roughly 30 nm.

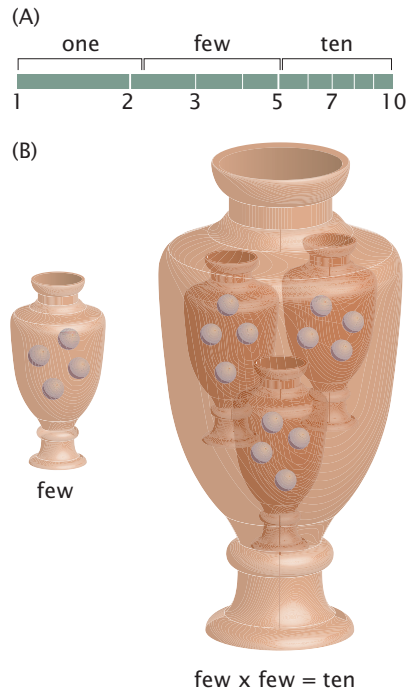


Figure 1.15: Few times one rule for multiplication. (A) Scheme for assigning numbers either to "one," "few," or "ten (concept from "Street Fighting Mathematics" by Sanjoy Mahajan)." (B) Demonstration that few $\times$ few $\approx$ 10.

constant has dimensions of Length^2/Time, while the mean lifetime has dimensions of Time. The only way these two quantities can be combined into one that has dimensions of Length is by multiplying them and then taking a square root, leading to the formula for λ given above. The particular power of dimensional analysis comes from comparing the estimate it yields with the measured value of the relevant quantity. In the morphogen example, a large discrepancy of the measured λ and the one estimated from the formula obtained by dimensional analysis would indicate that our determination of the morphogen diffusion constant or its lifetime is wrong. If more careful measurements of these two parameters still lead to a discrepancy, then we are led to an even more exciting conclusion, namely, that there is something fundamentally missing in our understanding of how the morphogen gradient is established, something that would introduce yet another parameter that determines λ.

Taylor and the Cover of Life *Magazine* One of the greatest stories of the power and subtlety of estimation and, particularly, of dimensional analysis is associated with G. I. Taylor. During the early days of the atomic bomb, the energetic yield of the first explosion was classified as a military secret. This did not keep the people who make such decisions from releasing several time lapse photos of the explosion that showed the radius of the fireball as a function of time. On this meager information, shown on the cover of *Life* magazine, Taylor was able to use the principles of fluid mechanics to *estimate* the yield of the explosion, an estimate that was exceedingly close to the secret value of the actual yield.

Interestingly, this famous episode from the cultural history of the physical sciences could have happened just as well in biology. Pictures like that shown in Figure 1.16 were the thrilling outcomes of the early days of the use of electron microscopy in biology. Here too, by observing the physical region occupied by the DNA molecule from the ruptured bacterial virus (bacteriophage), it is possible to make a highly simplified estimate of the *genomic* size (that is, the length in base pairs) of the bacteriophage's DNA.

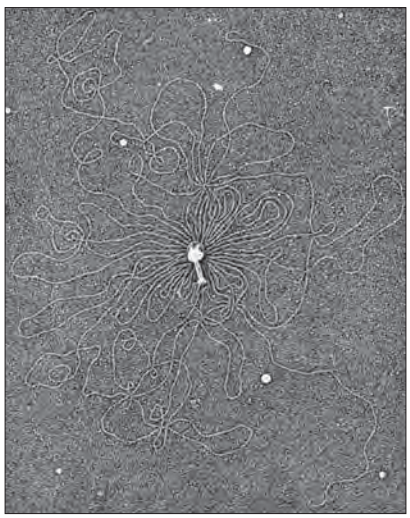


Figure 1.16: Electron microscopy image of a bacteriophage genome that has escaped its capsid. Simple arguments from polymer physics can be used to estimate the genomic size of the DNA by examining the physical size of the randomly spread DNA. We will perform these kinds of calculations in Chapter 8. (Adapted from G. Stent, Molecular Biology of Bacterial Viruses. W. H. Freeman, 1963.)

Why do numbers matter? Why is it useful to know the number of copies of a given molecule that is controlling the expression of a gene of interest within an *E. coli* cell? What merit is there in estimating the rate of actin polymerization at the leading edge of a motile cell? What is the value in knowing the time it takes a given protein to diffuse a millimeter? First, simple estimates serve as a reality check to help us see whether or not our impressions of how a system works are reasonable at all. Second, a sense of the numbers can tell us what kinds of physical constraints are in play. For example, by knowing the size of a bacterium and its swimming speed, we can construct a characteristic number (the Reynolds number) that tells us about the importance of acceleration in the dynamical problem of swimming. Similarly, a knowledge of the number of the specific DNA-binding proteins known as transcription factors that control the expression of a gene of interest can tell us about the importance of fluctuations in the gene expression process.

As will be seen in Chapters 3 and 19, biological oscillators are one of the centerpieces of the regulatory architecture of almost all living organisms. Oscillatory behavior is seen in examples ranging from the clocks that control the cell cycle in developing embryos to the circadian rhythms that make jet lag the unwanted but constant companion

of the weary traveler. As we explore mathematical models of regulatory biology, we will see that minimal oscillators can be constructed from genetic networks involving just two proteins, namely, an activator and a repressor. But what we will also see in contemplating these important networks is that for many of the possible parameters we might choose for the protein production and degradation rates, they don't yield oscillatory behavior at all, but rather stable and boring solutions in which the protein concentrations relax after some transient period into stationary values. Hence, even a clear picture of the wiring of some network cannot tell us what that network does in the normal life of a cell and calls for a quantitative examination of how that network behaves as a function of its governing parameters.

1.5.6 On Being Wrong

When performing simple quantitative estimates, or when attempting to write simple mathematical models to describe some aspect of a biological process, we are easily tempted to worry about whether our estimates and models are "right" or "wrong." There are at least four different ways in which a model might justly be called "wrong," but we believe that committing each of these kinds of wrongness can be a useful and worthwhile experience.

One category of wrong models might fail because they use an inappropriate type of fundamental physical model to abstract an aspect of a particular biological problem. This kind of wrongness might be illustrated by the question: how many hours can fit in the back of a pickup truck? The investigator will quickly learn that the predictions of such a model will simply have no relevance to the actual experiment or biological phenomenon being considered. It is inevitable that any person learning to conceptualize the kinds of complicated issues we explore in biology will make some of these kinds of mistakes at first. With time and experience, these errors will become less frequent.

A second category of wrong models might fail because they do not include enough detail about the system. For example, when we try to predict how fast *E. coli* bacterial cells can swim through water, we might be able to write an equation that correctly predicts the order of magnitude of the average speed of the cells, but fails to provide any basis for predicting how the speed might fluctuate over time. This kind of error, we believe, is not so much wrongness as incompleteness. As we have argued, it is actually extremely important that physical models be incomplete in order to be simple enough to help in building intuition. A model that is nearly as complex as the system itself may be able to make very accurate predictions, but will not be useful for promoting understanding.

A third category of wrong models includes those where inappropriate assumptions are made about the relative importance of different physical considerations. Although these models may be mathematically correct, they will fail to make accurate predictions for the behavior of cells in the real world. For example, if the swimming *E. coli* cell stops rotating its flagellum, then its future trajectory will be governed by a combination of inertia, viscous drag, and Brownian motion. A modeler whose primary experience lies in hydrodynamics for large objects such as submarines or fish may be inclined to ignore the contribution of Brownian motion, but for the bacterium this is the primary determinant of its behavior. Familiarity with the biological system and the opportunity to directly compare model predictions

with experimental data for iteratively refined generations of models will usually allow this problem to be overcome.

The fourth category of wrong models is by far our favorite. These are the wrong models that drive breakthrough experiments. When a person who is familiar with a particular biological system writes down a quantitative model for that system incorporating the correct basic physical models and uses the correct order-of-magnitude estimates for relevant parameters nevertheless finds that it makes dramatically incorrect predictions, this is almost always a golden opportunity to learn something fundamentally new about the system. Science advances when people notice that something does not quite make sense, that something is puzzling, that something seems to be missing. It is our hope that readers of this book will develop sufficient confidence in their ability to create appropriate models that when they find a model that seems to be wrong in this way they will seize the opportunity to make a new scientific discovery.

1.5.7 Rules of Thumb: Biology by the Numbers

A mastery of any field requires knowledge of a few key facts. For the physicist who uses quantum mechanics, this means knowing the ground state energy of the hydrogen atom, the charge on an electron, Planck's constant, and several other key physical parameters. In physical biology, there are similarly a small set of numbers that are worth carrying around as the basis of simple estimates of the kind one might carry out with a stick in the sand. In that spirit, the key numbers given in Table 1.1 are those that we find useful when thinking about physical biology of the cell. In the chapters that follow, these key numbers will permit us to form the habit of making quantitative estimates as a preliminary step when confronting new problems.

Throughout the remainder of the book, we will repeatedly make reference to certain numerical rules of thumb such as that the volume of an *E. coli* cell is roughly $1\,\mu\text{m}^3$ or that the natural energy scale of physical biology is k_BT (that is, Boltzmann's constant multiplied by room temperature in kelvins ≈ 4 pN nm (piconewton nanometers)). It is convenient to express this natural energy scale in terms of piconewtons and nanometers when we are considering the forces generated by protein molecular motors moving along DNA, for example, but this natural energy scale can also be expressed in different units depending on the type of calculation we are doing. When considering processes involving charge and electricity, it is convenient to recall that k_BT is approximately equal to 25 meV (millielectron volts). When considering biochemical reactions, it is more useful to express k_BT in biochemical units, as 0.6 kcal/mol (kilocalories per mole) or 2.5 kJ/mol (kilojoules per mole).

The quantitative rules of thumb in Table 1.1 will serve as the basis of our rough numerical estimates that could be carried out using a stick in the sand without reference to books, papers, or tables of data. Where do these numbers come from? Each comes from the results of a long series of experimental measurements of many different kinds. Right now we will simply assert their values, but later in the book we will provide numerous examples showing how these kinds of measurements have been made and also noting that in many cases, there remains uncertainty about the precise values of the parameters that emerge from such measurements. How precisely must we know these numbers? As stated earlier, our approach is to be generally satisfied

Table 1.1: Rules of thumb for biological estimates.

	Quantity of interest	Symbol	Rule of thumb
E. coli			
	Cell volume	$V_{E.\,coli}$	$\approx 1\,\mu m^3$
	Cell mass	$m_{E.\,coli}$	≈ 1 pg
	Cell cycle time	$t_{E.\,coli}$	≈ 3000 s
	Cell surface area	$A_{E.\,coli}$	$\approx 6\,\mu m^2$
	Macromolecule concentration in cytoplasm	$c_{E.\,coli}^{\text{macromol}}$	≈ 300 mg/mL
	Genome length	$N_{\text{bp}}^{E.\,coli}$	$\approx 5 \times 10^6$ bp
	Swimming speed	$v_{E.\,coli}$	$\approx 20\,\mu$m/s
Yeast			
	Volume of cell	V_{yeast}	$\approx 60\,\mu m^3$
	Mass of cell	m_{yeast}	≈ 60 pg
	Diameter of cell	d_{yeast}	$\approx 5\,\mu$m
	Cell cycle time	t_{yeast}	≈ 200 min
	Genome length	$N_{\text{bp}}^{\text{yeast}}$	$\approx 10^7$ bp
Organelles			
	Diameter of nucleus	d_{nucleus}	$\approx 5\,\mu$m
	Length of mitochondrion	l_{mito}	$\approx 2\,\mu$m
	Diameter of transport vesicles	d_{vesicle}	≈ 50 nm
Water			
	Volume of molecule	V_{H_2O}	$\approx 10^{-2}$ nm^3
	Density of water	ρ	1 g/cm^3
	Viscosity of water	η	≈ 1 centipoise (10^{-2} g/(cm s))
	Hydrophobic embedding energy	$\approx E_{\text{hydr}}$	2500 cal/(mol nm^2)
DNA			
	Length per base pair	l_{bp}	$\approx 1/3$ nm
	Volume per base pair	V_{bp}	≈ 1 nm^3
	Charge density	λ_{DNA}	$2\,e/0.34$ nm
	Persistence length	ξ_{p}	50 nm
Amino acids and proteins			
	Radius of "average" protein	r_{protein}	≈ 2 nm
	Volume of "average" protein	V_{protein}	≈ 25 nm^3
	Mass of "average" amino acid	M_{aa}	≈ 100 Da
	Mass of "average" protein	M_{protein}	$\approx 30{,}000$ Da
	Protein concentration in cytoplasm	c_{protein}	≈ 150 mg/mL
	Characteristic force of protein motor	F_{motor}	≈ 5 pN
	Characteristic speed of protein motor	v_{motor}	≈ 200 nm/s
	Diffusion constant of "average" protein in cytoplasm	D_{protein}	$\approx 10\,\mu m^2/s$
Lipid bilayers			
	Thickness of lipid bilayer	d	≈ 5 nm
	Area per molecule	A_{lipid}	$\approx \frac{1}{2}$ nm^2
	Mass of lipid molecule	m_{lipid}	≈ 800 Da

in those cases where our numerical estimates are of the right order of magnitude. To that end, we make certain gross simplifications. For example, when thinking about proteins and the amino acids that make them up, we will often adopt a minimalistic perspective in which the mass of amino acids is *approximated* to be 100 Da (daltons, or grams per mole). In actuality, real amino acids have masses that range over a wider set of values from 75 to 204 Da. Likewise, the rest of the numbers in Table 1.1 represent only rough average values to use as a starting point for quantitative estimates. As we progress through the examples in the rest of the book, these values will become increasingly familiar. However, our book does not make a systematic attempt at the important task of collecting values for a wide range of biologically important numbers. For this, the interested reader is encouraged to visit the "BioNumbers" website to search out what is known about the quantitative status of a given biological parameter and to be on the lookout for those cases in which we cite such BioNumbers by their BioNumber ID (BNID).

1.6 Summary and Conclusions

When applying physical and quantitative thinking to biological problems, two different skills are critically important. First, the biological system must be described in terms of fundamental physical models that can help to shed light on the behavior of the system. Fewer than 10 truly basic physical models are able to provide broad explanatory frameworks for many problems of interest in biology. Second, appropriate quantitative estimates to give order-of-magnitude predictions of sizes, numbers, time scales, and energies involved in a biological process can serve as a reality check that our conception of the system is appropriate and can help predict what kinds of experimental measurements are likely to provide useful information. Armed with these skills, an investigator can begin to systematize the large amounts of detailed quantitative data generated by biological experimentation and determine where new scientific breakthroughs might be sought.

1.7 Further Reading

Alberts, B, Johnson, A, Lewis, J, et al. (2008) Molecular Biology of the Cell, 5th ed., Garland Science. This book provides a discussion of much of the biological backdrop for our book.

Schaechter, M, Ingraham, JL, & Neidhardt, FC (2006) Microbe, ASM Press. This book is full of insights into the workings of microscopic organisms.

Stryer, L (1995) Biochemistry, 5th ed., W. H. Freeman. We like to imagine our readers with this book at their side. Stryer is full of interesting insights presented in a clear fashion.

Schleif, R (1993) Genetics and Molecular Biology, 2nd ed., The Johns Hopkins University Press. Schleif's book discusses many of the same topics found throughout our book from a deeper biological perspective, but with a useful and interesting quantitative spin.

Vogel, S (2003) Comparative Biomechanics—Life's Physical World, Princeton University Press. Though largely aimed at describing problems at a different scale than those described in our book, Vogel's book (as well as his many others) presents an amazing merger of biology and physical reasoning.

Nurse, P (2003) The Great Ideas of Biology, *Clinical Medicine* **3**, 560. This article describes Nurse's views on some of the greatest ideas in biology.

Bonner, JT (2002) The Ideas of Biology, Dover Publications. This book espouses Bonner's view on the overarching themes of modern biology.

The spirit of physical biology we are interested in developing here has already been described in a number of other books. The following are some of our favorites:

Dill, K, & Bromberg, S (2011) Molecular Driving Forces, 2nd ed., Garland Science. Despite a deceptive absence of difficult mathematics, their book is full of insights, subtlety, and tasteful modeling.

Boal, D (2012) Mechanics of the Cell, Cambridge University Press. Boal's book illustrates the way in which physical

reasoning can be brought to bear on problems of biological significance. We have found that many of the exercises at the back of each chapter are the jumping off point for the analysis of fascinating biological problems.

Howard, J (2001) Mechanics of Motor Proteins and the Cytoskeleton, Sinauer Associates. Howard's book is a treasure trove of information and concepts.

Nelson, P (2004) Biological Physics: Energy, Information, Life, W. H. Freeman. This book discusses a wide range of topics at the interface between physics and biology.

Benedek, GB, & Villars, FMH (2000) Physics with Illustrative Examples from Medicine and Biology, 2nd ed., Springer-Verlag. This three-volume series should be more widely known. It is full of useful insights into everything ranging from the analysis of the Luria–Delbrück experiment to the action potential.

Sneppen, K, & Zocchi, G (2005) Physics in Molecular Biology, Cambridge University Press. This book presents a series of vignettes that illustrate how physical and mathematical reasoning can provide insights into biological problems.

Jackson, MB (2006) Molecular and Cellular Biophysics, Cambridge University Press. Jackson's book covers many themes in common with our own book and is especially strong on ion channels and electrical properties of cells.

Waigh, TA (2007) Applied Biophysics, Wiley. This interesting book covers many of the same topics covered here.

Harte, J (1988) Consider a Spherical Cow, University Science Books. This book does a beautiful job of showing how order of magnitude estimates can be used to examine complex problems. Those that are intrigued by this book might consider dipping into his Consider a Cylindrical Cow (University Science Books, 2001) as well.

Weinstein, L, & Adam, JA (2008) Guesstimation: Solving the World's Problems on the Back of a Cocktail Napkin, Princeton University Press. This wonderful little book shows by example how to do the kind of estimates advocated here and also is the basis for the rule captured in Figure 1.14.

Goldreich, P, Mahajan, S, & Phinney, S, Order-of-Magnitude Physics: Understanding the World with Dimensional Analysis, Educated Guesswork, and White Lies, online book, available at www.inference.phy.cam.ac.uk/sanjoy/oom/. This excellent work embodies the science of estimation so helpful to biology as well as physics.

Cohen, J (2004) Mathematics is biology's next microscope, only better; Biology is Mathematics' next physics, only better, *PLoS Biology* **2**, e439. This article argues for the way that quantitative reasoning will enrich biology and for the ways that biology will enrich mathematics.

Mahajan, S (2010) Street-Fighting Mathematics, The MIT Press. Much of our mathematics will be of the street-fighting kind and there is nowhere better to get a feeling for this than from Mahajan's book.

1.8 References

Bray, D (2001) Reasoning for results, *Nature* **412**, 863. Source of Darwin's quote at the beginning of the chapter.

Darwin, F (ed.) (1903) More Letters of Charles Darwin: A Record of His Work in a Series of Hitherto Unpublished Letters, Volume I, John Murray.

DiMauro, S, & Schon, EA (2003) Mitochondrial respiratory-chain diseases, *N. Engl. J. Med.* **348**, 2656.

Fawcett, DW (1996) The Cell, Its Organelles and Inclusions: An Atlas of Fine Structure, W. B. Saunders. Fawcett's atlas is full of wonderful electron microscopy images of cells.

Milo, R, Jorgensen, P, Moran, U, et al. (2010) BioNumbers—the database of key numbers in molecular and cell biology, *Nucl. Acids Res*. **38**, D750.

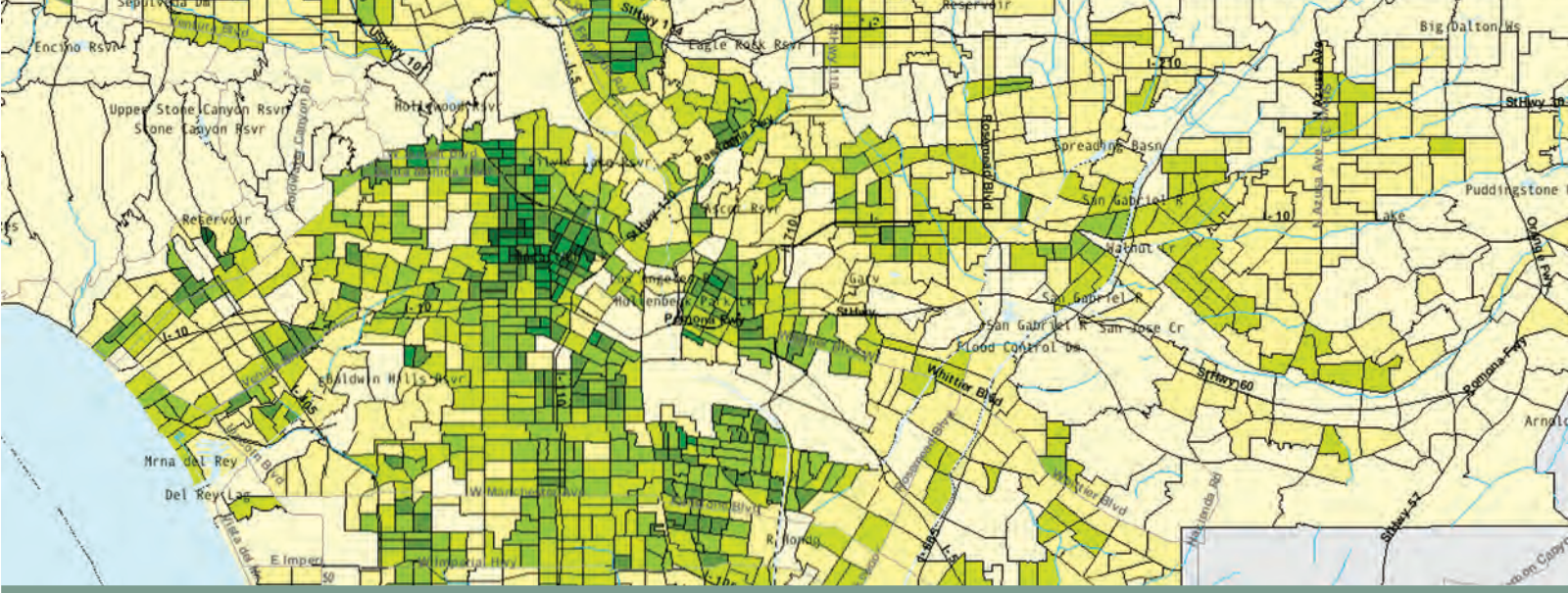
Müller, J, Oehler, S, & Müller-Hill, B (1996) Repression of *lac* promoter as a function of distance, phase and quality of an auxiliary *lac* operator, *J. Mol. Biol*. **257**, 21.

Poustelnikova, E, Pisarev, A, Blagov, M, et al. (2004) A database for management of gene expression data *in situ, Bioinformatics* **20**, 2212.

Rubik, BA, & Koshland, DE, Jr. (1978) Potentiation, desensitization, and inversion of response in bacterial sensing of chemical stimuli, *Proc. Natl. Acad. Sci*., **75**, 2820.

Stent, G (1963) Molecular Biology of Bacterial Viruses, W. H. Freeman.

This page intentionally left blank
to match pagination of print book



What and Where: Construction Plans for Cells and Organisms

2

Overview: In which we consider the size of cells and the nature of their contents

Cells come in a dazzling variety of shapes and sizes. Even so, their molecular inventories share many common features, reflecting the underlying biochemical unity of life. In this chapter, we introduce the bacterium *Escherichia coli* (commonly abbreviated as *E. coli*) as our biological standard ruler. This cell serves as the basis for a first examination of the inventory of cells and will permit us to get a sense of the size of cells and the nature of their contents. Using simple estimates, we will take stock of the genome size, numbers of lipids and proteins, and the ribosome content of bacteria. With the understanding revealed by *E. coli* in hand, we then take a powers-of-10 journey down and up from the scale of individual cells. Our downward journey will examine organelles within cells, macromolecular assemblies ranging from ribosomes to viruses, and then the macromolecules that are the engines of cellular life. Our upward journey from the scale of individual cells will examine the biological structures resulting from multicellularity, this time with an emphasis on how cells act together in contexts ranging from bacterial biofilms to the networks of neurons in the brain.

> “By the help of Microscopes there is nothing so small as to escape our inquiry; hence there is a new visible World discovered.”
>
> **Robert Hooke**

2.1 An Ode to *E. coli*

Scientific observers of the natural world have been intrigued by the processes of life for many thousands of years, as evidenced by early written records from Aristotle, for example. Early thinkers wondered about the nature of life and its “indivisible” units in much the same way that they mused about the fundamental units of matter. Just as physical scientists arrived at a consensus that the fundamental unit of matter is the atom (at least for chemical transactions), likewise, observers of living organisms have agreed that the indivisible unit of life is the cell. Nothing smaller than a cell can be shown to

be alive in a sense that is generally agreed upon. At the same time, there are no currently known reasons to attribute some higher "living" status to multicellular organisms, in contrast to single-celled organisms.

Cells are able to consume energy from their environments and use that energy to create ordered structures. They can also harness energy and materials from the environment to create new cells. A standard definition of life merges the features of metabolism (that is, consumption and use of energy from the environment) and replication (generating offspring that resemble the original organism). Stated simply, the cell is the smallest unit of replication (though viruses are also replicative units, but depend upon their infected host cell to provide much of the machinery making this replication possible).

The recognition that the cell is the fundamental unit of biological organization originated in the seventeenth century with the microscopic observations of Hooke and van Leeuwenhoek. This idea was put forth as the modern cell theory by Schwann, Schleiden, and Virchow in the mid-nineteenth century and was confirmed unequivocally by Pasteur shortly thereafter and repeatedly in the time since. Biologists agree that all forms of life share cells as the basis of their organization. It is also generally agreed that all living organisms on Earth shared a common ancestor several billion years ago (or, more accurately, a community of common ancestors) that would be recognized as cellular by any modern biologist.

In terms of understanding the basic rules governing metabolism, replication, and life more generally, one cell type should be as good as any other as the basis of experimental investigations of these mechanisms. For practical reasons, however, biologists have focused on a few particular types of cell to try to illuminate these general issues. Among these, the human intestinal inhabitant *E. coli* stands unchallenged as the most useful and important representative of the living world in the biologist's laboratory.

Several properties of *E. coli* have contributed to its great utility and have made it a source of repeated discoveries. First, it is easy to isolate because it is present in great abundance in human fecal matter. Unlike most other bacteria that populate the human colon, *E. coli* is able to grow well in the presence of oxygen. In the laboratory, it replicates rapidly and can easily adjust to changes in its environment, including changes in nutrients. Second, it is nearly routine to deliberately manipulate these cells (using insults such as radiation, chemicals, or genetic tricks, for example) to produce mutants. Mutant organisms are those that differ from their parents and from other members of their species found in the wild because of specific changes in DNA sequence that give rise to biologically significant differences. For example, *E. coli* is normally able to synthesize purines to make DNA and RNA from sugar as a nutrient source. However, particular mutants of *E. coli* with enzymatic deficiencies in these pathways have lost the ability to make their own purines and become reliant on being fed precursors for these molecules. A more familiar and frightening example of the consequence of mutations is the way in which bacteria acquire antibiotic resistance. Throughout the book, we will be using specific examples of biological phenomena to illustrate general physical principles that are relevant to life. Often, we will have recourse to *E. coli* because of particular experiments that have been performed on this organism. Further, even when we speak of experiments on other cells or organisms, often *E. coli* will be behind the scenes coloring our thinking.

2.1.1 The Bacterial Standard Ruler

The Bacterium *E. coli* Will Serve as Our Standard Ruler

Throughout this book, we will discuss many different cells, which all share with *E. coli* the fundamental biological directive to convert energy from the environment into structural order and to perpetuate their species. On Earth, it is observed that there are certain minimal requirements for the perpetuation of cellular life. These are not necessarily absolute physical requirements, but in the competitive environment of our planet, all surviving cells share these features in common. These include a DNA-based genome, mechanisms to transcribe DNA into RNA, and, subsequently, translation mechanisms using ribosomes to convert information in RNA sequences into protein sequence and protein structure. Within those individual cells, there are many substructures with interesting functions. Larger than the cell, there are also structures of biological interest that arise because of cooperative interactions among many cells or even among different species. In this chapter, we will begin with the cell as the fundamental unit of biological organization, using *E. coli* as the standard reference and standard ruler. We will then look at smaller structures within cells and, finally, larger multicellular structures, zooming in and out from our fundamental cell reference frame.

Figure 2.1 shows several experimental pictures of an *E. coli* cell and its schematization into our standard ruler. The AFM image and the electron micrograph in Figures 2.1(A) and (B) show that these bacteria have a rod-like morphology with a typical length between 1 and 2 μm and a diameter between 0.5 and 1 μm. The reader is invited to use light microscopy images of these bacteria and to determine the size of an *E. coli* cell for him or herself in the "Computational Exploration" at the end of this section.

The length unit of 1 μm, or micrometer, is so useful for the discussion of cell biology that it has a nickname, the "micron." To put this standard ruler in perspective, we note that, with its characteristic length scale of 1 μm, it would take roughly 50 such cells lined up end to end in order to measure out the width of a human hair. On the other hand, we would need to divide the cell into roughly 500 slices of equal width in order to measure out the diameter of a DNA molecule. Using these insights, the question of how many bacteria can dance on the head of a pin is answered unequivocally in Figure 2.1(D).

Note that the average size of these cells depends on the nutrients with which they are provided, with those growing in richer media having a larger mass. An extremely elegant experiment that explores the connection between growth rate and mass is shown in Figure 2.2. Our reference growth condition throughout the book will be a chemically defined solution referred to by microbiologists as "minimal medium" that is a mixture of salts along with glucose as the sole carbon source. In the laboratory, bacteria are often grown in "rich media," which are poorly defined but nutrient-rich mixtures of extracts from organic materials such as yeast cultures or cow brains. Although microorganisms can grow very rapidly in rich media, they are rarely used for biochemical studies because their exact contents are not known. For consistency, we will therefore refer primarily to experimental results for bacterial growth in minimal media.

Because of its central role as the quantitative standard in the remainder of the book, it is useful to further characterize the geometry of *E. coli*. One example in which we will need a better sense of the

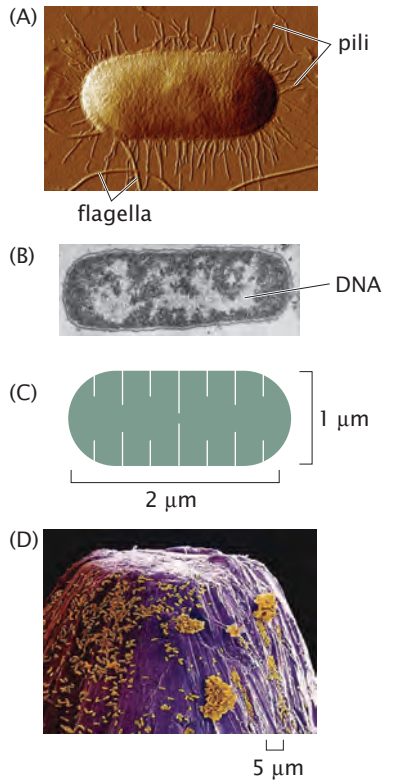


Figure 2.1: *E. coli* as a standard ruler for characterizing spatial scales. (A) Atomic-force microscopy (AFM) image of an *E. coli* cell, (B) electron micrograph of a sectioned *E. coli* bacterium, and (C) the *E. coli* ruler. (D) Bacteria on the head of a pin giving an impression of how our standard ruler compares with the dimensions of everyday objects. (A, courtesy of Ang Li; D, courtesy of Tony Brain.)

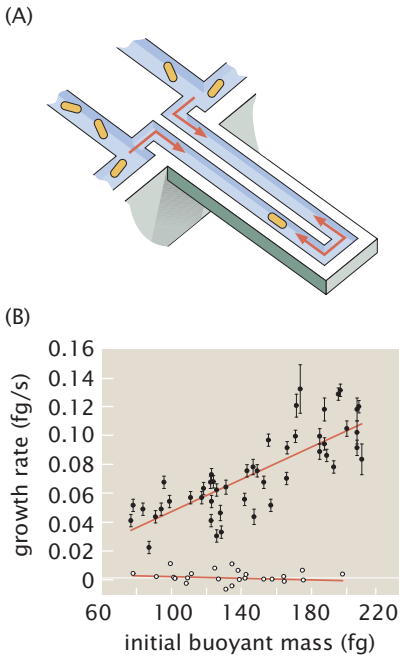


Figure 2.2: *E. coli* mass and growth rate as measured using a cantilever assay. (A) Schematic of the geometry of the device used to characterize the buoyant mass of bacterial cells by dynamically trapping them within the hollow cantilever. The hollowed out cantilever oscillates with a slightly different frequency when there is a cell present in its interior. (B) Relation between the buoyant mass at the start of the experiment and the growth rate. *E. coli* K12 cells grown at 37 °C. Filled circles correspond to cells that are growing normally and open circles correspond to fixed cells. (Figures adapted from M. Godin et al., *Nat. Meth.* 7: 387, 2010.)

geometry of cells and their internal compartments is in the context of reconciling experiments performed *in vitro* (that is, in test tubes) and *in vivo* (that is, in living cells). Results from experiments done *in vitro* are based upon the free concentrations of different molecular species. On the other hand, in *in vivo* situations, we might know the number of copies of a given molecule such as a transcription factor inside the cell (transcription factors are proteins that regulate expression of genes by binding to DNA). To reconcile these two pictures, we will need the cellular volume to make the translation between molecular counts and concentrations. Similarly, when examining the distribution of membrane proteins on the cell surface, we will need a sense of the cell surface area to estimate the mean spacing between these proteins, which will tell us about the extent of interaction between them. For most cases of interest in this book, it suffices to attribute a volume $V_{E.\ coli} \approx 1\,\mu\text{m}^3 = 1\,\text{fL}$ to *E. coli* and an area of roughly $A_{E.\ coli} \approx 6\,\mu\text{m}^2$ (see the problems for examples of how to work out these numbers from known cellular dimensions).

Computational Exploration: Sizing Up *E. coli* As already revealed in Figure 2.1, the simplest way for us to figure out how big bacteria are is to stick them under a microscope and to look at them. Of course, to do this in reality requires some way to relate the cells we see magnified in our microscope and their actual size. In modern terms, what this really means is figuring out the relationship between the size of the pixels produced by our camera and real world length units such as microns.

In this Computational Exploration, the reader is invited to take the images shown in Figures 2.3(A) and (C) and to use them to determine how many nanometers are in a pixel. To that end, the idea is to read the image file from Figure 2.3(A) into Matlab and to produce a graph of the intensity for one row of the image as shown in Figure 2.3(B). Each one of the lines in the graticule shown in Figure 2.3(A) is separated by 10 μm. With this information in hand, the reader can work out the size of our pixels and hence install the missing scale bar on Figure 2.3(C), which is the ultimate objective of this Computational Exploration.

COMPUTATIONAL EXPLORATION

2.1.2 Taking the Molecular Census

In the remainder of this section, we will proceed through a variety of estimates to try to get a grip on the number of molecules of different kinds that are in an *E. coli* cell. Why should we care about these numbers? First, a realistic physical picture of any biological phenomenon demands a quantitative understanding of the individual particles involved (for biological phenomena, this usually means molecules) and the spatial dimensions over which they have the freedom to act. One of the most immediate outcomes of our cellular census will be the realization of just how crowded the cellular interior really is, a subject explored in detail in Chapter 14. Our census will paint a very different picture of the cellular interior as the seat of biochemical reactions than is suggested by the dilute and homogeneous environment of the biochemical test tube. Indeed, we will see that the mean spacing between protein molecules within a typical cell is less than 10 nm. As we will see below this distance is comparable to the size of the proteins themselves!

Taking the molecular census is also important because we will use our molecular counts in Chapter 3 to estimate the rates of macromolecular synthesis during the cell cycle. How fast is a genome replicated? What is the average rate of protein synthesis during the cell cycle and, given what we know about ribosomes, how do they maintain this rate of synthesis? A prerequisite to beginning to answer these questions is the macromolecular census itself.

Ultimately, to understand many experiments in biology, it is important to realize that most experimentation is comparative. That is, we compare "normal" behavior to perturbed behavior to see if some measurable property has increased or decreased. To make these statements meaningful, we need first to understand the quantitative baseline relative to which such increases and decreases are compared. There is another sense in which numbers of molecules are particularly meaningful which will be explored in detail in subsequent chapters that has to do with whether we can describe a cell as having "a lot" or "a few" copies of some specific molecule. If a cell has a lot of some particular molecule, then it is appropriate to describe the concentration of that molecule as the basis for predicting cellular function. However, when a cell has only a few copies of a particular molecule, then we need to consider the influence of random chance (or stochasticity) on its function. In many cases, cells have an interesting intermediate number of molecules where it is not immediately clear which perspective is appropriate. However, knowing the absolute numbers always gives us a reality check for subsequent assumptions and approximations for modeling biological processes.

Because of these considerations, much effort among biological scientists has been focused on the development of quantitative techniques for measuring the molecular census of living cells (both bacteria and eukaryotes). In this chapter, we will rely primarily on order-of-magnitude estimates based on simple assumptions. These estimates are validated by comparison with measurements. In subsequent chapters, these estimates will be refined through explicit model building and direct comparison with quantitative experiments.

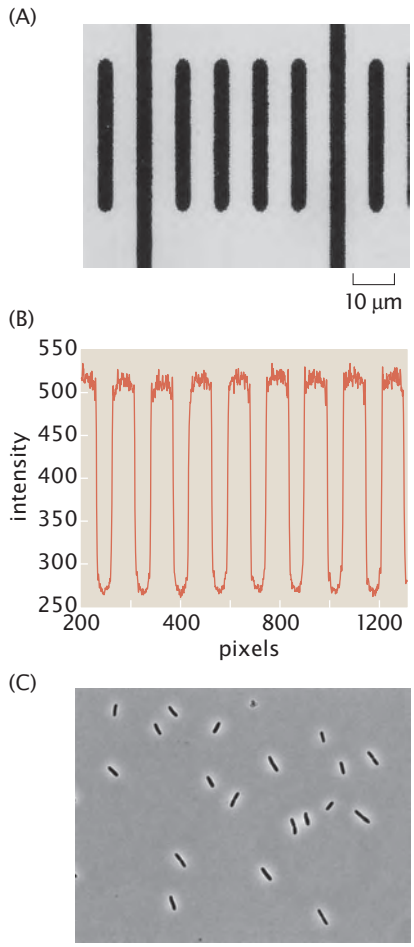


Figure 2.3: Sizing up *E. coli*. (A) Image of a graticule at 100× magnification. (B) Matlab plot of the intensity for a horizontal cut through the image. (C) Phase contrast image of a field of view with several *E. coli* cells taken at the same magnification as in (A).

ESTIMATE

Estimate: Sizing Up *E. coli* As already noted in the previous chapter, cells are made up of an array of different macromolecules as well as small molecules and ions. To estimate the number of proteins in an *E. coli* cell, we begin by noting that, with its 1 fL volume, the mass of such a cell is roughly 1 pg, where we have assumed that the density of the cell is that of water, which is 1 g/mL, though clearly Figure 2.2 shows that this assumption is not true. Measurements reveal that the dry weight of the cell is roughly 30% of its total and half of that mass is protein. As a result, the total protein mass within the cell is roughly 0.15 pg. We can also estimate the number of carbon atoms in a bacterium by considering the chemical composition of the macromolecules of the cell. This implies that roughly half the dry mass comes from the carbon content of these cells, a figure that reveals of the order of 10^{10} carbon atoms per cell. Two of the key sources that have served as a jumping off point for these estimates are Pedersen et al. (1978) and Zimmerman and Trach (1991), who describe the result of a molecular census of a bacterium.

As a first step toward revealing the extent of crowding within a bacterium, we can estimate the number of proteins

by assuming an average protein of 300 amino acids with each amino acid having a characteristic mass of 100 Da. These assumptions are further examined in the problems at the end of the chapter. Using these rules of thumb, we find that the mean protein has a mass of 30,000 Da. Using the conversion factor that $1\,\text{Da} \approx 1.6 \times 10^{-24}\,\text{g}$, we have that our typical protein has a mass of $5 \times 10^{-20}\,\text{g}$. The number of proteins per *E. coli* cell is estimated as

$$N_{\text{protein}} = \frac{\text{total protein mass}}{\text{mass per protein}} \approx \frac{15 \times 10^{-14}\,\text{g}}{5 \times 10^{-20}\,\text{g}} \approx 3 \times 10^{6}. \quad (2.1)$$

If we invoke the rough estimate that one-third of the proteins coded for in a typical genome correspond to membrane proteins, this implies that the number of cytoplasmic proteins is of the order of 2×10^6 and the number of membrane proteins is 10^6, although we note that not all of these membrane-associated proteins are strictly transmembrane proteins.

Another interesting use of this estimate is to get a rough impression of the number of ribosomes—the cellular machines that synthesize proteins. We can estimate the total number of ribosomes by first estimating the total mass of the ribosomes in the cell and then dividing by the mass per ribosome. To be concrete, we need one other fact, which is that roughly 20% of the protein complement of a cell is ribosomal protein. If we assume that all of this protein is tied up in assembled ribosomes, then we can estimate the number of ribosomes by noting that (a) the mass of an individual ribosome is roughly 2.5 MDa and (b) an individual ribosome is roughly one-third by mass protein and two-thirds by mass RNA, facts that can be directly confirmed by the reader by inspecting the structural biology of ribosomes. As a result, we have

$$\begin{aligned} N_{\text{ribosome}} &= \frac{0.2 \times 0.15 \times 10^{-12}\,\text{g}}{830{,}000\,\text{Da}} \times \frac{1\,\text{Da}}{1.6 \times 10^{-24}\,\text{g}} \\ &\approx 20{,}000 \text{ ribosomes.} \end{aligned} \quad (2.2)$$

The numerator of the first fraction has 0.2 as the fraction of protein that is ribosomal, 0.15 as the fraction of the total cell mass that is protein, and 1 pg as the cell mass. Our estimate for that part of the ribosomal mass that is protein is 830,000 Da. The size of a ribosome is roughly 20 nm (in "diameter") and hence the total volume taken up by these 20,000 ribosomes is roughly $10^8\,\text{nm}^3$. This is 10% of the total cell volume.

Idealizing an *E. coli* cell as a cube, sphere, or spherocylinder yields (see the problems) that the surface area of such cells is $A_{E.\,coli} \approx 6\,\mu\text{m}^2$. This number may be used in turn to estimate the number of lipid molecules associated with the inner and outer membranes of these cells as

$$N_{\text{lipid}} \approx \frac{4 \times 0.5 \times A_{E.\,coli}}{A_{\text{lipid}}} \approx \frac{4 \times 0.5 \times (6 \times 10^6\,\text{nm}^2)}{0.5\,\text{nm}^2} \approx 2 \times 10^7, \quad (2.3)$$

where the factor of 4 comes from the fact that the inner and outer membranes are each *bilayers*, implying that the lipids effectively cover the cell surface area four times. A lipid bilayer consists of two sheets of lipids with their tails pointing toward

each other. The factor of 0.5 is based on the crude estimate that roughly half of the surface area is covered by membrane proteins rather than lipids themselves. We have made the similarly crude estimate that the area per lipid is $0.5\,\mathrm{nm}^2$. The measured number of lipids is of the order of 2×10^7 as well.

In terms of sheer numbers, water molecules are by far the majority constituent of the cellular interior. One of the reasons this fact is intriguing is that during the process of cell division, a bacterium such as *E. coli* has to take on a very large number of new water molecules each second. The estimate we obtain here will be used to examine this transport problem in the next chapter. To estimate the number of water molecules, we exploit the fact that roughly 70% of the cellular mass (or volume) is water. As a result, the total mass of water is 0.7 pg. We can find the approximate number of water molecules as

$$\begin{aligned} N_{\mathrm{H_2O}} &\approx \frac{0.7 \times 10^{-12}\,\mathrm{g}}{18\ \mathrm{g/mol}} \times 6 \times 10^{23}\ \text{molecules/mol} \\ &\approx 2 \times 10^{10}\ \text{water molecules.} \end{aligned} \tag{2.4}$$

It is also of interest to gain an impression of the content of inorganic ions in a typical bacterial cell. To that end, we assume that a typical concentration of positively charged ions such as K^+ is 100 mM, resulting in the estimate

$$\begin{aligned} N_{\mathrm{ions}} &\approx \frac{(100 \times 10^{-3}\ \mathrm{mol}) \times (6 \times 10^{23}\ \text{molecules/mol})}{10^{15}\,\mu\mathrm{m}^3} \times 1\,\mu\mathrm{m}^3 \\ &= 6 \times 10^7. \end{aligned} \tag{2.5}$$

Here we use the fact that $1\,\mathrm{L} = 10^{15}\,\mu\mathrm{m}^3$. This result could have been obtained even more easily by noting yet another simple rule of thumb, namely, that one molecule per *E. coli* cell corresponds roughly to a concentration of 2 nM.

The outcome of our attempt to size up *E. coli* is illustrated schematically in summary form in Figure 2.4. A more complete census of an *E. coli* bacterium can be found in Neidhardt et al. (1990). The outcome of experimental investigations of the molecular census of an *E. coli* cell is summarized (for the purposes of comparing with our estimates) in Table 2.1.

How is the census of a cell taken experimentally? This is a question we will return to a number of different times, but will give a first answer here. For the case of *E. coli*, one important tool has been the use of gels like that shown in Figure 2.5. Such experiments work by breaking open cells and keeping only their protein components. The complex protein mixture is then separated into individual molecular species using a polyacrylamide gel matrix. First, the protein mixture is distributed through a tube-shaped polyacrylamide gel that has been polymerized to contain a stable pH gradient, and then an electric field is applied across the gel. The net charge on each protein depends on the pH and on the number and type of charged (protonatable) amino acid side chains that it contains. For example, the carboxylic acid group on aspartate will be negatively charged at high pH, but

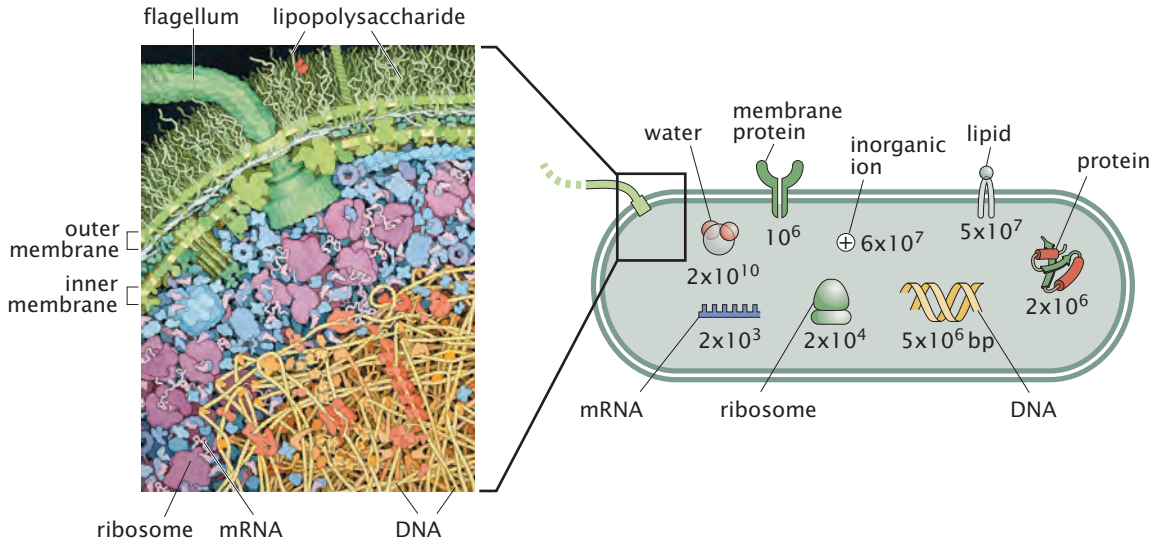


Figure 2.4: Molecular contents of the bacterium *E. coli*. The illustration on the left shows the crowded cytoplasm of the bacterial cell. The cartoon on the right shows an order-of-magnitude molecular census of the *E. coli* bacterium with the approximate number of different molecules in *E. coli*. (Illustration of the cellular interior courtesy of D. Goodsell.)

Table 2.1: Observed macromolecular census of an *E. coli* cell. (Data from F. C. Neidhardt et al., Physiology of the Bacterial Cell, Sinauer Associates, 1990 and M. Schaechter et al., Microbe, ASM Press, 2006.)

Substance	% of total dry weight	Number of molecules
Macromolecules		
Protein	55.0	2.4×10^6
RNA	20.4	
23S RNA	10.6	19,000
16S RNA	5.5	19,000
5S RNA	0.4	19,000
Transfer RNA (4S)	2.9	200,000
Messenger RNA	0.8	1,400
Phospholipid	9.1	22×10^6
Lipopolysaccharide (outer membrane)	3.4	1.2×10^6
DNA	3.1	2
Murein (cell wall)	2.5	1
Glycogen (sugar storage)	2.5	4,360
Total macromolecules	**96.1**	
Small molecules		
Metabolites, building blocks, etc.	2.9	
Inorganic ions	1.0	
Total small molecules	**3.9**	

will pick up a hydrogen ion and will be neutral at low pH. Conversely, the amine group on lysine will be neutral at high pH, but will pick up a hydrogen ion and will be positively charged at low pH. The pH where a protein's charge is net neutral is called its "isoelectric point."

When a protein finds itself in a region of the gel where the pH is below its isoelectric point, there will be an excess of positive charge associated with the protein, and it will move toward the cathode when the electric field is applied. When a protein finds itself in a region of the gel where the pH is above its isoelectric point, it will have a net negative charge, and will therefore move toward the anode. When

all the proteins in the mixture have moved to the location of the pH where each is neutral, then all movement stops. At this point, "isoelectric focusing" is complete, and all of the proteins are arrayed along the tube-shaped gel in positions according to their net charge. Next, a charged detergent is added that binds to all proteins so the total number of detergent molecules associated with an individual protein is roughly proportional to the protein's overall size. The detergent-soaked isolectric focusing gel tube is placed at one end of a flat, square gel that also contains detergent, and an electric field is applied in a direction perpendicular to the first field. Because the net charge on the detergent molecules is much larger than the original net charge of the protein, the rate of migration in the second direction through the gel is determined by the protein size.

After the procedure described above, the individual protein species in the original mixture have been resolved into a series of spots on the gel, with large, negatively charged proteins at the upper left-hand corner and small, positively charged proteins at the lower right-hand corner for the gel shown in Figure 2.5. The proteins can then be stained with a nonspecific dye so that their locations within the gel can be directly observed. The intensity of the spots on such a gel can then be used as a basis for quantifying each species. The identity of the protein that congregates in each spot can be determined by physically cutting each spot out of the gel, eluting the protein, and determining its size and amino acid content using mass spectrometry. Similar tricks are used to characterize the amount of RNA and lipids, for example, resulting in a total census like that shown in Table 2.1.

More recently, several other methods have been brought to bear on the molecular census of cells. As shown in Figure 2.6, two of these methods are the use of mass spectrometry and fluorescence microscopy. In mass spectrometry, fragments of the many proteins contained within the cell are run through the mass spectrometer and their absolute abundances determined. The results of this technique applied to *E. coli* are shown in Figure 2.6(A). Alternatively, in fluorescence microscopy, a library of cells is created where each strain in this library expresses a particular protein from the *E. coli* proteome fused to a fluorescent protein. By calibrating the fluorescence corresponding to an individual fluorescent protein, one can measure the amount

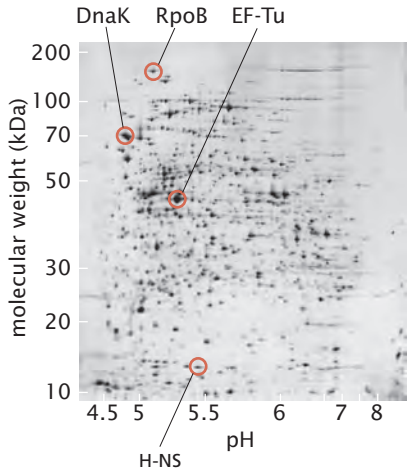


Figure 2.5: Protein census of the cell. Measurement of protein census of *E. coli* using two-dimensional polyacrylamide gel electrophoresis. Each spot represents an individual protein with a unique size and charge distribution. The spots arising from several well-known bacterial proteins are labeled. (Copyright Swiss Institute of Bioinformatics, Geneva, Switzerland.)

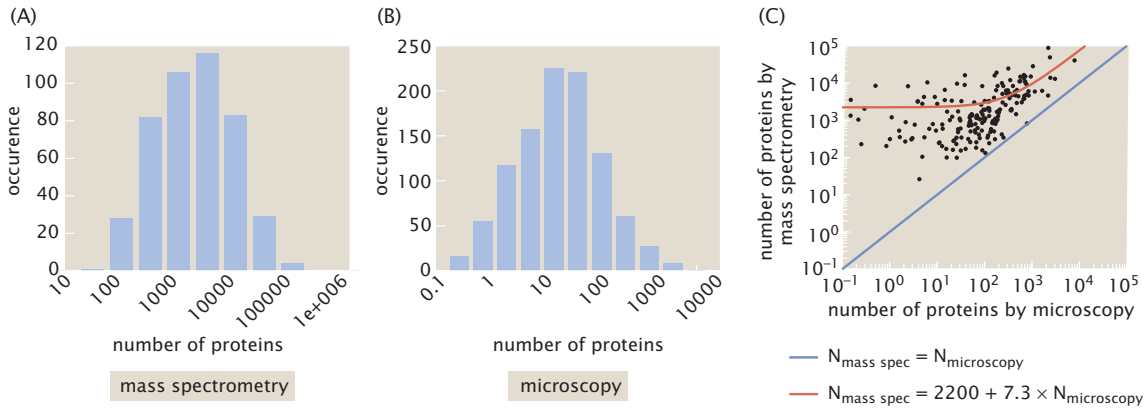


Figure 2.6: Protein census of *E. coli* using several techniques. A histogram of protein number in *E. coli* is shown from results using (A) mass spectrometry and (B) fluorescence microscopy with protein fusions to fluorescence proteins. (C) Comparison of mass spectrometry and fluorescence methods for the same collection of proteins showing a discrepancy between the two techniques as revealed by the fact that the best fit is not given by a line of slope one. (A, adapted from P. Lu et al., *Nat. Biotechnol.* 25:117, 2007; B, adapted from Y. Taniguchi et al., *Science* 329:533, 2010.)

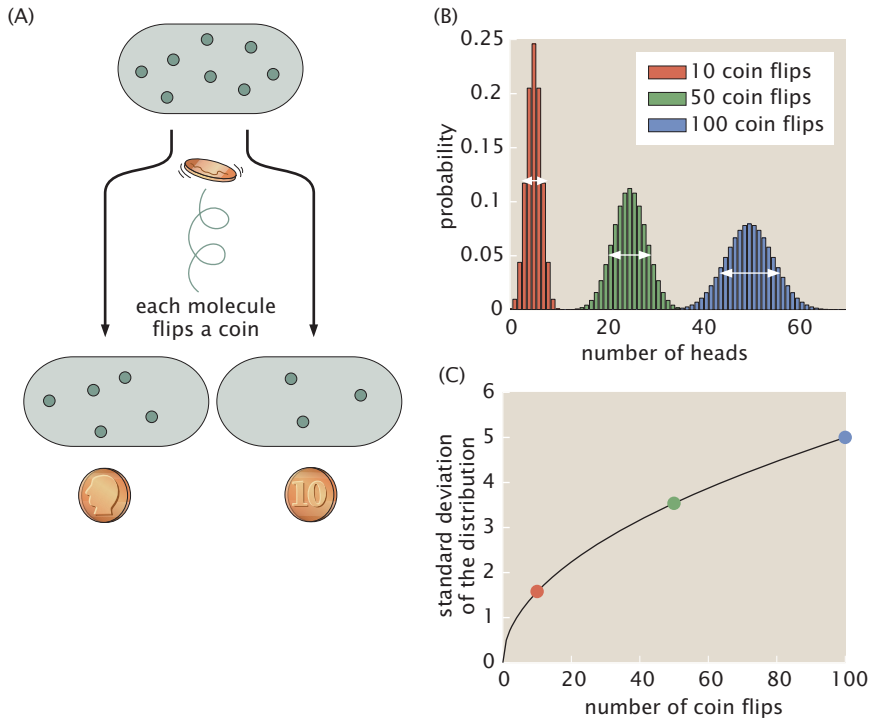


Figure 2.7: Schematic of the random partitioning of molecules during the process of cell division. (A) When the cell divides, each of the molecules chooses a daughter cell via a coin flip. (B) Probability distribution for the number of heads for different choices of the total number of coin flips. (C) The width of the distribution as a function of the number of coin flips.

of protein fusion in each strain of the library by comparing the total fluorescence in the cell with the single-molecule standard. Such a measurement for *E. coli* results in the histogram showed in Figure 2.6(B). It is important to point out, however, that these two methods are not in agreement, as shown in Figure 2.6(C). It is seen that, for a given protein, fluorescence microscopy tends to undercount proteins with respect to mass spectrometry (or mass spectrometry tends to overcount with respect to fluorescence microscopy). There are various possible sources for this discrepancy, ranging from systematic errors in the two experimental techniques to the fact that the experiments were done under different growth conditions leading to very different cell cycle times and cell sizes. One of the key features revealed by fluorescence measurements is the importance of cell-to-cell variability, which we can estimate using simple ideas from probability theory.

Estimate: Cell-To-Cell Variability in the Cellular Census

ESTIMATE

When cells divide, how much variability should we expect in the partitioning of the mRNA and protein contents of the different daughter cells? Conceptually, we can think of the passive molecular partitioning process as a series of coin flips in which each molecule of interest flips a coin to decide which of the two daughters it will go to. This idea is illustrated schematically in Figure 2.7.

To make a simple estimate of the variability, we turn to one of the most important probability distributions in all of science, namely, the binomial distribution. Despite the apparently contrived nature of coin flips, they are precisely the statistical experiment we need when thinking about cellular concentrations of molecules that are passively partitioned during cell division. The idea of the binomial distribution is that we carry out N trials of our coin flip process with the outcomes being heads and tails. However, a more biologically relevant

way to think of these outcomes is that with each cell division, the macromolecule of interest partitions either to daughter 1 or to daughter 2, with the probability of going to daughter 1 given by p and the probability of going to daughter 2 given by $q = 1 - p$. If we are dealing with a fair coin, or if the sizes of the two daughters are equal and there are no active segregation mechanisms in play, then both outcomes are equally likely and have probability $p = q = 1/2$.

To be precise, the probability of having n_1 of our N molecules partition to daughter cell 1 is given by

$$p(n_1, N) = \frac{N!}{n_1!(N - n_1)!} p^{n_1} q^{N-n_1}. \tag{2.6}$$

The factor $N!/[n_1!(N - n_1)!]$ reflects the fact that there are many different ways of flipping n_1 heads (h) and $N - n_1$ tails (t) and is given by the famed binomial coefficients (see Figure 2.8). For example, if $N = 3$ and $n_1 = 2$, then our three trials could turn out in three ways as *thh*, *hth*, and *hht*, exactly what we would have found by constructing $3!/(2!1!)$. Note also that throughout this book we will see that, in certain limits, the binomial distribution can be approximated by the Gaussian distribution (large-N case) or the Poisson distribution ($p \ll 1$ case).

If the number of copies of our mRNA or protein of interest is 10, then on division each daughter will get 5 of these molecules on average. More formally, this can be stated as

$$\langle n_1 \rangle = Np. \tag{2.7}$$

For the case of a fair coin (i.e., $p = 1/2$), this reduces to precisely the result we expect. This intuitive result can be appreciated quantitatively in the distributions shown in Figure 2.7(B). Here, the average number of heads obtained for a number of coin flips is given by the center of the corresponding distribution. The width of these distributions then gives a sense for how often there will be a result that deviates from the mean. As a result, to assess the cell-to-cell variability, we need to find a way to measure the width of the distribution. One way to characterize this width is by the standard deviation, which is computed from the function given in Equation 2.6 as follows:

$$\langle n_1^2 \rangle - \langle n_1 \rangle^2 = Npq, \tag{2.8}$$

a result the reader is invited to work out in the problems at the end of the chapter.

With these results in hand, we can now turn to quantifying the fluctuations that arise during partitioning. There is no one way to characterize the fluctuations, but a very convenient one is to construct the ratio

$$\frac{\sqrt{\langle n_1^2 \rangle - \langle n_1 \rangle^2}}{\langle n_1 \rangle} = \frac{1}{\sqrt{N}} \tag{2.9}$$

This result tells us that when the number of molecules is small (i.e., <100), the cell-to-cell variation is itself a significant fraction of the mean itself.

Such claims are not a mere academic curiosity and have been the subject of careful experimental investigation. On multiple occasions in the remainder of this book, we will examine partitioning data (for mRNA, for proteins, for carboxysomes,

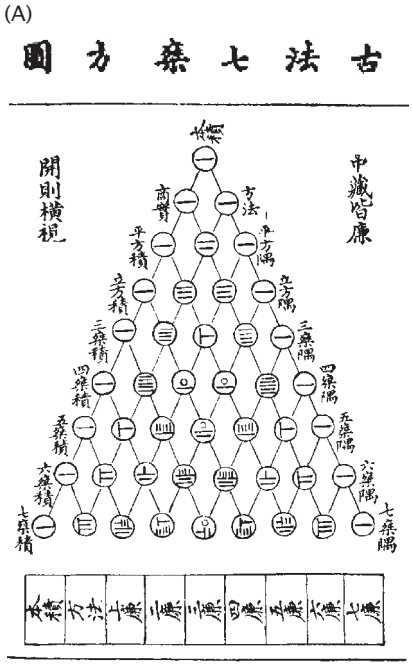


Figure 2.8: Binomial coefficients. (A) Though commonly known as Pascal's triangle, Chinese mathematicians had also earlier appreciated the symmetry of this collection of numbers. (B) The binomial coefficients that arise when raising $p + q$ to the power n, where n characterizes which row we are talking about, with the top "row" denoted by $n = 0$.

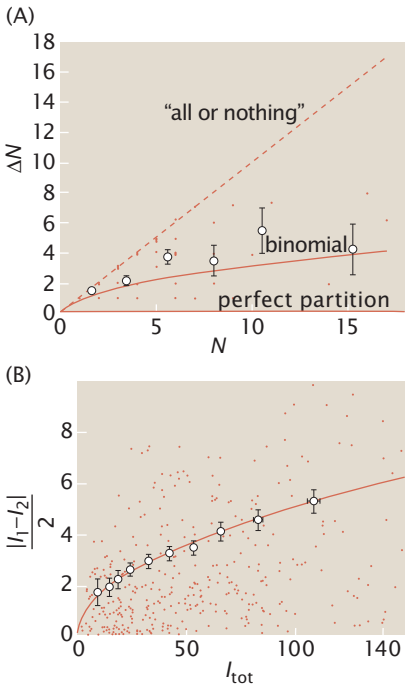


Figure 2.9: Binomial partitioning of mRNA and proteins in *E. coli* during cell division. (A) Difference ΔN in the number of mRNAs between the two daughters given that the mother has N mRNAs. The curves show three possible partitioning mechanisms involving "all or nothing" in which one daughter takes all of the mRNAs, binomial partitioning, and "perfect partitioning" in which each daughter gets exactly half of the proteins from the mother cell. (B) Difference in the fluorescence intensity of the two daughter cells for a fluorescent fusion to a repressor protein as a function of the fluorescence intensity of the mother cell. The line corresponds to binomial partitioning model. (A, adapted from I. Golding et al., *Cell* 123:1025, 2005; B, adapted from N. Rosenfeld et al., *Science* 307:1962, 2005.)

etc.) in which cell lineages are followed and the partitioning of the macromolecules or complexes of interest are measured directly. An example of such data for the case of mRNA and proteins is shown in Figure 2.9. As a result of the ability to fluorescently label individual mRNAs as they are produced, it is possible to characterize the number of such mRNAs in the mother cell before division and then to measure how they partition during the division process. Using fusions of fluorescent proteins to repressor molecules, similar measurements were taken in the context of proteins as shown in Figure 2.9(B). In both of these cases, the essential idea is to count the number of molecules in the mother cell and then to count the number in the two daughters and to see how different they are. Similar data is shown in Figure 18.7 (p. 724) for the case of the organelles in cyanobacteria known as carboxysomes, which contain the all-important enzyme Rubisco so critical to carbon fixation during the process of photosynthesis.

Computational Exploration: Counting mRNA and Proteins by Dilution As will be highlighted throughout this book, many of our key results depend critically upon the number of molecular actors involved in a given process. As shown in Figure 18.7 (p. 724), for example, fluorescence microscopy has made it possible to track the dynamics and localization of key cellular structures such as the carboxysomes. One of the main conclusions of that study was that the partitioning of the carboxysomes is not a random process since the *statistics* of the partitioning of these organelles between the two daughter cells does not follow the binomial distribution, which would be characteristic of random partitioning. The goal of this Computational Exploration is to explore the statistics of random partitioning from the perspective of simulations, with reference not only to the carboxysome example, but also to clever new ways of taking the molecular census. Using these simulations, we will show how the fluctuations in the partitioning of a fluorescent protein between daughter cells is related to the cells' intensity as schematized in Figure 2.10. Data from this type of experiment are shown in Figure 19.15 (p. 816).

The advent of fluorescent protein fusions now makes it possible to assess the relative quantities of a given protein of interest on the basis of the fluorescence of its fusion partner. However, for the purposes of the strict comparison between theory and experiment that we are often interested in, it is necessary to have a calibration that permits us to convert between the arbitrary fluorescence units as measured in a microscope and the more interesting molecular counts necessary for plugging into our various formulae or that we use in dissecting some molecular mechanism. To effect this calibration, we exploit the idea of a strict linear relation between the observed intensity and the number of fusion proteins as indicated by the relation

$$I_{tot} = \alpha N_{tot}, \tag{2.10}$$

where α is the unknown calibration factor linking fluorescence and the number of fluorescent molecules (N_{tot}).

As indicated in Figure 2.10, a series of recent experiments all using the same fluctuation method make it possible to

COMPUTATIONAL EXPLORATION

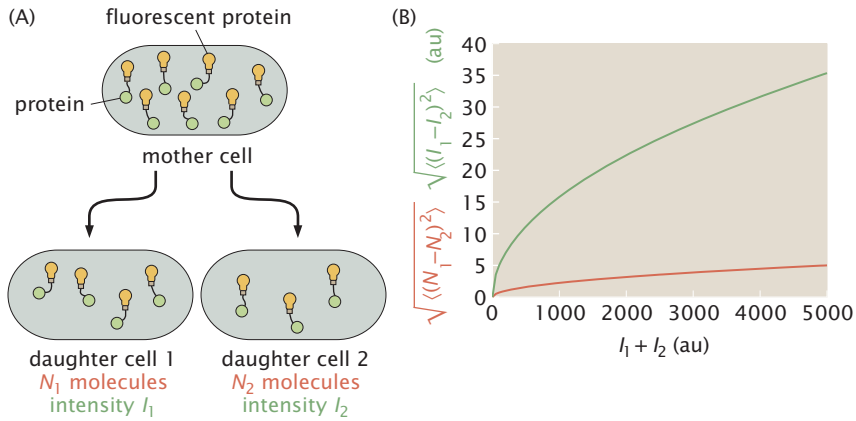


Figure 2.10: Intensity of the daughter cells as a result of binomial partitioning. (A) The mother cell starts out with a certain number of fluorescent proteins (represented by light bulbs) and upon division, these proteins are partitioned according to the binomial distribution. The number of fluorescent proteins and their corresponding total fluorescent intensity are related by the calibration factor α. (B) Plot of the intensity difference between the two daughters as a function of the intensity of the mother cell and the corresponding difference in partitioning of fluorescent molecules as a function of the total number of molecules in the mother cell.

actually determine N_{tot} by determining the unknown calibration factor. The idea of the experiment is to exploit the natural fluctuations resulting from imperfect partitioning of the fluorescent proteins between the two daughter cells. The key point is that for the case in which the partitioning between the two daughters is random (i.e., analogous to getting heads or tails in a coin flip), the size of those fluctuations depends in turn upon the total number of proteins being segregated between the two daughters. The larger the number of proteins, the larger the fluctuations. In particular, as shown in the problems at the end of the chapter, we can derive the simple and elegant result that the average difference in intensity between the two daughter cells is given by

$$\langle (I_1 - I_2)^2 \rangle = \alpha I_{tot}, \tag{2.11}$$

where I_1 and I_2 are the intensities of daughters 1 and 2, respectively, and I_{tot} is the total fluorescence intensity of the mother cell ($I_{tot} = I_1 + I_2$).

The simple analytic result of Equation 2.11 can perhaps be best appreciated through an appeal to a stochastic simulation. The idea is to "simulate" the dilution of fluorescent proteins during repeated cell divisions where all the mother cells are presumed to start with N copies of the fluorescent protein and the partitioning of those proteins to the two daughters is determined strictly by a simple binary random process (i.e., heads or tails of a coin flip of an honest coin). Conceptually, the algorithm for performing this simulation can be summarized as follows:

1. Choose the number of fluorescent molecules N in the mother cell and compute the intensity $I_{tot} = \alpha N$.
2. Generate N random numbers between 0 and 1. If a given random number is between 0 and 0.5, assign the protein to daughter 1 and otherwise assign the protein to daughter 2.
3. Compute the "intensity" of the daughters by multiplying N_1 and N_2 by α.
4. Perform the same operation for 100 cells (for example).
5. Choose a new N and repeat the process.

By considering the same division process again and again, for different choices of N, we can generate a curve that captures

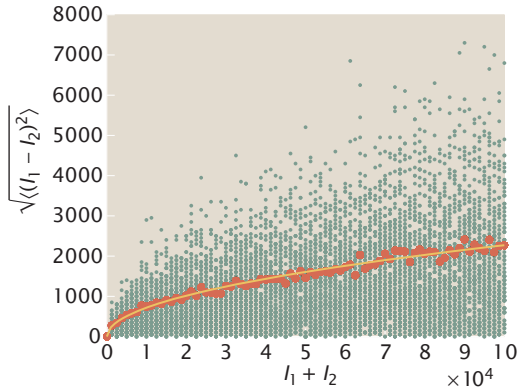


Figure 2.11: Simulation results of fluorescent protein dilution and counting experiment. For a given total number of proteins, N, in the mother cell, we simulate 100 different outcomes of partitioning by cellular division and plot their corresponding partitioning error (green dots). For each value of N, we can calculate the root mean square in partitioning (red dots) and fit them to the expected functional form given by the square root of Equation 2.11.

the statistics set forth above. In particular, for each and every N, 100 separate "divisions" were simulated and the results of each such division process are shown as the data points. For a given value of N, we can calculate the root mean square error in partitioning and its corresponding intensity, given by $\sqrt{\langle (I_1 - I_2)^2 \rangle}$. To see an example of the outcome of such a simulation, examine Figure 2.11.

Examples of Matlab code that could be used to perform this 'Computational Exploration' can be found on the book's, website.

The Cellular Interior Is Highly Crowded, with Mean Spacings Between Molecules That Are Comparable to Molecular Dimensions

One of the most intriguing implications of our census of the molecular parts of a bacterium is the extent to which the cellular interior is crowded. Because of experiments and associated estimates on the contents of *E. coli*, it is now possible to construct illustrations to depict the cellular interior in a way that respects the molecular census. The crowded environs of the interior of such a cell are shown in Figure 2.4. This figure gives a view of the crowding associated with any *in vivo* process. In Chapter 14, we will see that this crowding effect will force us to call in question our simplest models of chemical potentials, the properties of water, and the nature of diffusion. The generic conclusion is that the mean spacing of proteins and their assemblies is comparable to the dimensions of these macromolecules themselves. The cell is a very crowded place!

The quantitative significance of Figure 2.4 can be further appreciated by converting these numbers into concentrations. To do so, we recall that the volume of an *E. coli* cell is roughly 1 fL. The rule of thumb that emerges from this analysis is that a concentration of 2 nM implies roughly one molecule per bacterium. A concentration of 2 μM implies roughly 1000 copies of that molecule per cell. Concentration in terms of our standard ruler is shown in Figure 2.12. This figure shows the number of copies of the molecule of interest in such a cell as a function of the concentration.

We can use these concentrations directly to compute the mean spacing between molecules. That is, given a certain concentration, there is a corresponding average distance between the molecules. Having a sense of this distance can serve as a guide to thinking about the likelihood of diffusive encounters and reactions between various molecular constituents. If we imagine the molecules at a given concentration arranged on a cubic lattice of points, then the mean spacing between

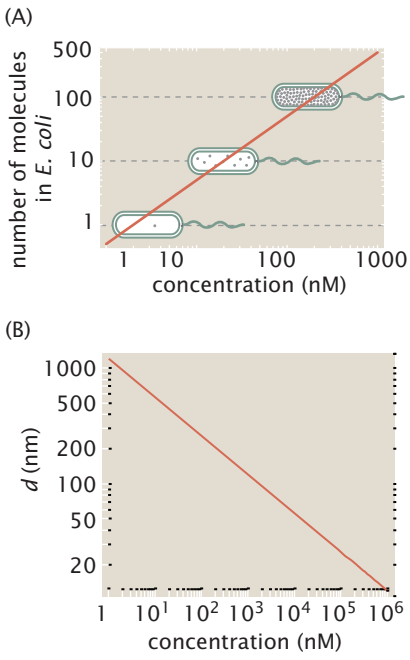


Figure 2.12: Physical interpretation of concentration. (A) Concentration in *E. coli* units: number of copies of a given molecule in a volume the size of an *E. coli* cell as a function of the concentration. (B) Concentration expressed in units of typical distance d between neighboring molecules measured in nanometers.

those points is given by

$$d = c^{-1/3}, \tag{2.12}$$

where c is the concentration of interest (measured in units of number of molecules per unit volume). Larger concentrations imply smaller intermolecular spacings. This idea is formalized in Figure 2.12(B), which shows the relation between the mean spacing measured in nanometers and the concentration in nanomolar units.

2.1.3 Looking Inside Cells

With our reference bacterium in mind, the remainder of the chapter focuses on the various structures that make up cells and organisms. To talk about these structures, it is helpful to have a sense of how we know what we know about them. Further, model building requires facts. To that end, we periodically take stock of the experimental basis for our models. For this chapter, the "Experiments Behind the Facts" focuses on how biological structures are explored and measured.



EXPERIMENTS

Experiments Behind the Facts: Probing Biological Structure To size up cells and their organelles, we need to extract "typical" structural parameters from a variety of experimental studies. Though we leave a description of the design and setup of such experiments to more specialized texts, the goal is to provide at least enough detail so that the reader can see where some of the key structural facts that we will use throughout the book originated. We emphasize two broad categories of experiments: (i) those in which some form of radiation interacts with the structure of interest and (ii) those in which forces are applied to the structure of interest.

Figure 2.13 shows three distinct experimental strategies that feed into our estimates, all of which reveal different facets of biological structure. One of the mainstays of structural analysis is light microscopy. Figure 2.13(A) shows a schematic of the way in which light can excite fluorescence in samples that have some distribution of fluorescent molecules within them. In particular, this example shows a schematic of a microtubule that has some distribution of fluorophores along its length. Incident photons of one wavelength are absorbed by the fluorophores and this excitation leads them to emit light of a different wavelength, which is then detected. As a result of selective labeling of only the microtubules with fluorophores, it is only these structures that are observed when the cell is examined in the microscope. These experiments permit a determination of the size of various structures of interest, how many of them there are, and where they are localized. By calibrating the intensity from single fluorophores, it has become possible to take a single-molecule census for many of the important proteins in cells. For an example of this strategy, see Wu and Pollard (2005). A totally different window on the structure of the cell and its components is provided by tools such as the atomic-force microscope (AFM). As will be explained in Chapter 10, the AFM is a cantilever beam with a sharp tip on its end. The tip is brought very close to the surface where the structure of interest is present and is then scanned in the

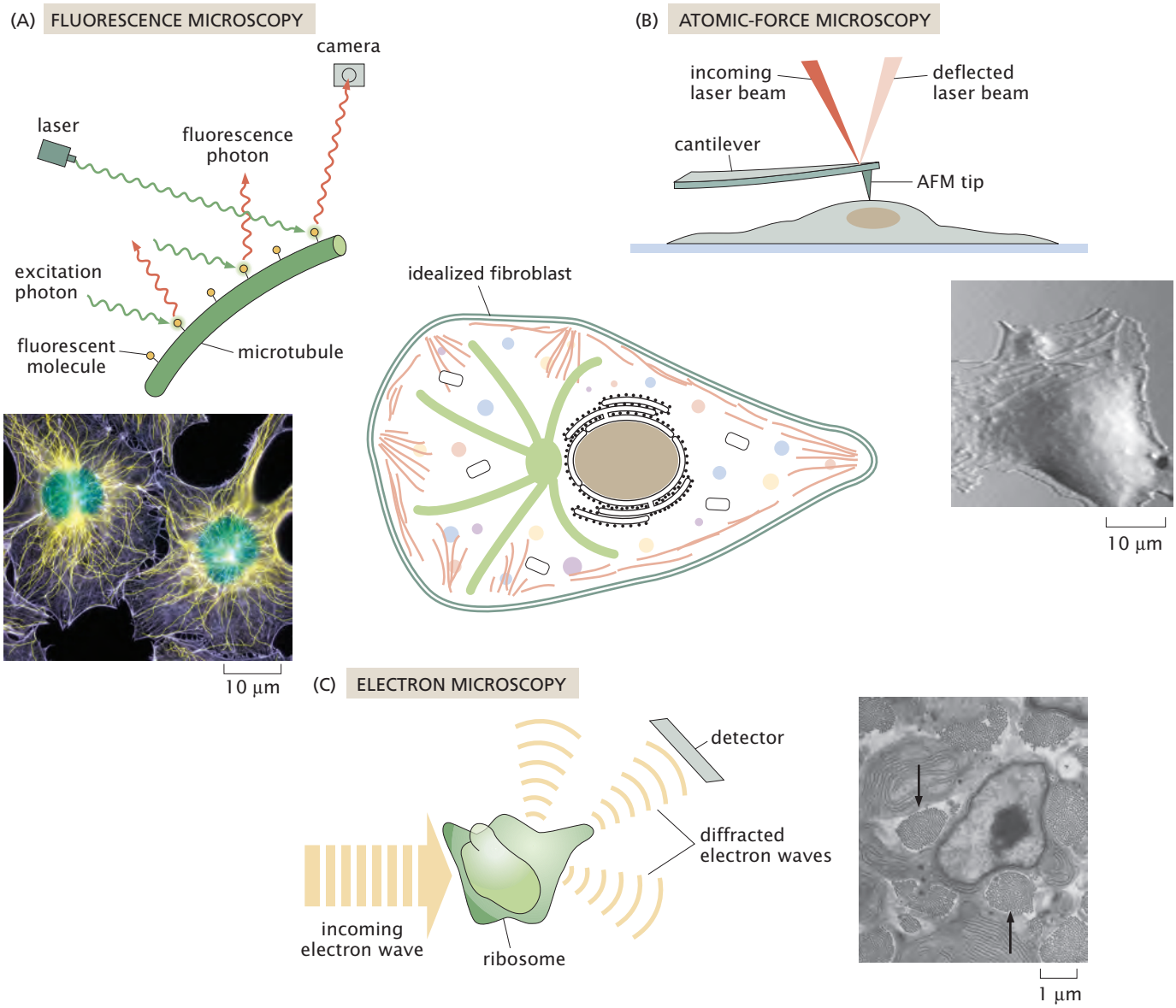


Figure 2.13: Experimental techniques that have revealed the structure of both cells and their organelles. (A) Fluorescence microscopy and the associated image of a fibroblast with labeled microtubules (yellow) and DNA (green). (B) Atomic-force microscopy schematic and the associated image of the surface topography of a fibroblast. (C) Electron microscopy schematic and image of cross-section through a fibroblast in an animal tissue. Arrows indicate bundles of collagen fibers. (A, courtesy of Torsten Wittman; B, adapted from M. Radmacher, *Meth. Cell Biol.* 83:347, 2007; C, adapted from D. E. Birk and R. L. Trelstad, *J. Cell Biol.* 103:231, 1986.)

plane of the sample. One way to operate the instrument is to move the cantilever up and down so that the force applied on the tip remains constant. Effectively, this demands a continual adjustment of the height as a function of the x–y position of the tip. The nonuniform pattern of cantilever displacements can be used to map out the topography of the structure of interest. Figure 2.13(B) shows a schematic of an AFM scanning a typical fibroblast cell as well as a corresponding image of the cell.

Figure 2.13(C) gives a schematic of the way in which X-rays or electrons are scattered off a biological sample. The schematic shows an incident plane wave of radiation that interacts with the biological specimen and results in the emergence

of radiation with the same wavelength but a new propagation direction. Each point within the sample can be thought of as a source of radiation, and the observed intensity at the detector reflects the interference from all of these different sources. By observing the pattern of intensity, it is possible to deduce something about the structure that did the scattering. This same basic idea is applicable to a wide variety of radiation sources, including X-rays, neutrons, and electrons.

An important variation on the theme of measuring the scattered intensity from irradiated samples is cryo-electron tomography. This technique is one of the centerpieces of structural biology and is built around uniting electron microscopy with sample preparation techniques that rapidly freeze the sample. The use of tomographic methods has made it possible to go beyond the planar sections seen in conventional electron microscopy images. The basis of the technique is indicated schematically in Figure 2.14, and relies on rotating the sample over a wide range of orientations and then assembling a corresponding three-dimensional reconstruction on the basis of the entirety of these images. This technique has already revolutionized our understanding of particular organelles and is now being used to image entire cells.

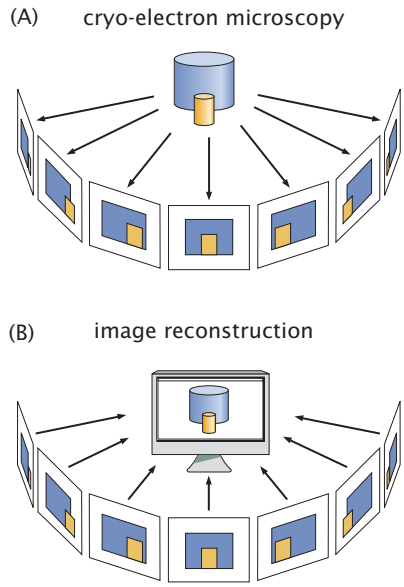


Figure 2.14: Schematic of tomographic reconstruction. (A) The three-dimensional sample is rotated in the electron microscope and imaged repeatedly to generate a series of two-dimensional images representing the pattern of radiation scattered at slightly different orientations. (B) Information from the image series is recombined computationally to generate a three-dimensional rendering of the original object.

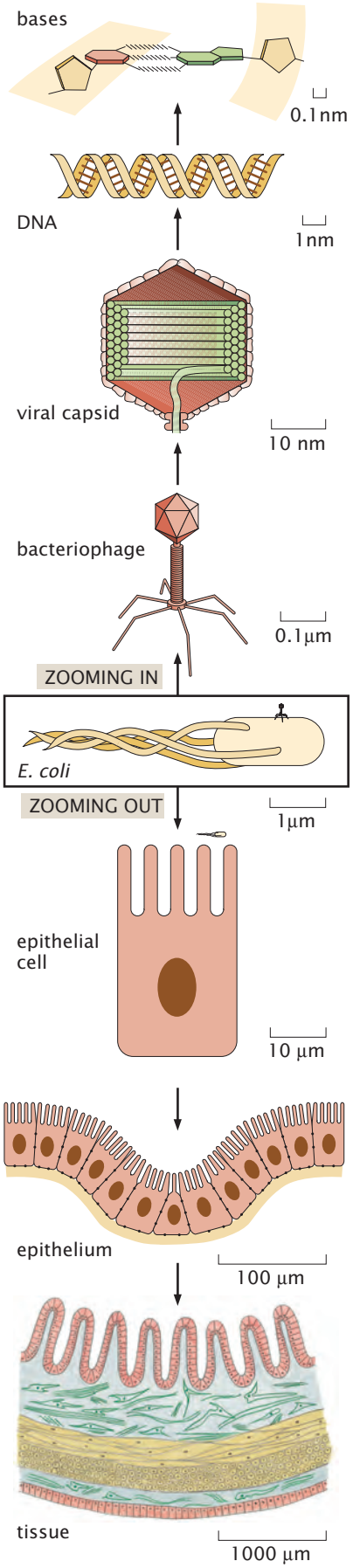
2.1.4 Where Does *E. coli* Fit?

Biological Structures Exist Over a Huge Range of Scales

The spatial scales associated with biological structures run from the nanometer scale of individual molecules all the way to the scale of the Earth itself. Where does *E. coli* fit into this hierachy of structures? Figure 2.15 shows the different structures that can be seen as we scale in and out from an *E. coli* cell. A roughly 10-fold increase in magnification relative to an individual bacterium reveals the viruses that attack bacteria. These viruses, known as bacteriophages, have a characteristic scale of roughly 100 nm. They are made up of a protein shell (the capsid) that is filled with the viral genome. Continuing our downward descent using yet higher magnification, we see the ordered packing of the viral genome within its capsid. These structures are intriguing because they involve the ordered arrangement of more than 10 μm of DNA in a capsid that is less than 100 nm across. Another rough factor-of-10 increase in resolution reveals the structure of the DNA molecule itself, with a characteristic cross-sectional radius of roughly 1 nm and a length of 3.4 nm per helical repeat.

A similar scaling out strategy reveals new classes of structures. As shown in Figure 2.15, a 10-fold increase in spatial scale brings us to the realm of eukaryotic cells in general, and specifically, to the scale of the epithelial cells that line the human intestine. We use this example because bacteria such as *E. coli* are central players as part of our intestinal ecosystem. Another 10-fold increase in spatial scale reveals one of the most important inventions of evolution, namely, multicellularity. In this case, the cartoon depicts the formation of planar sheets of epithelial cells. These planar sheets are themselves the building blocks of yet higher-order structures such as tissues and organs. Scaling out to larger scales would bring us to multicellular organisms and the structures they build.

The remainder of the chapter takes stock of the structures at each of these scales and provides a feeling for the molecular building blocks



that make up these different structures. Our strategy will be to build upon our cell-centered view and to first descend in length scale from that of cells to the molecules of which they are made. Once this structural descent is complete, we will embark on an analysis of biological structure in which we zoom out from the scale of individual cells to collections of cells.

2.2 Cells and Structures within Them

2.2.1 Cells: A Rogue's Gallery

All living organisms are based on cells as the indivisible unit of biological organization. However, within this general rule there is tremendous diversity among living cells. Several billion years ago, our last common ancestor gave rise to three different lineages of cells now commonly called Bacteria, Archaea, and Eukarya, a classification suggested by similarities and differences in ribosomal RNA sequences. Every living organism on Earth is a member of one of these groups. Most bacteria and archaea are small (3 μm or less) and extremely diverse in their preferred habitats and associated lifestyles ranging from geothermal vents at the bottom of the ocean to permafrost in Antarctica. Bacteria and Archaea look similar to one another and it has only been within the last few decades that molecular analysis has revealed that they are completely distinct lineages that are no more closely related to each other than the two are to Eukarya.

The organisms that we most often encounter in our everyday life and can see with the naked eye are members of Eukarya (individuals are called eukaryotes). These include all animals, all plants ranging from trees to moss, and also all fungi, such as mushrooms and mold. Thus far, we have focused on *E. coli* as a representative cell, although we must acknowledge that *E. coli*, as a member of the bacterial group, is in some ways very different from a eukaryotic or archaeal cell. The traditional definition of a eukaryotic cell is one that contains its DNA genome within a membrane-bound nucleus. Most bacteria and archaea lack this feature and also lack other elaborate intracellular membrane-bound structures such as the endoplasmic reticulum and the Golgi apparatus that are characteristic of the larger and more complex eukaryotic cells.

Cells Come in a Wide Variety of Shapes and Sizes and with a Huge Range of Functions

Cells come in such a wide variety of shapes, sizes, and lifestyles that choosing one representative cell type to tell their structural story is misleading. In Figure 2.16, we show a rogue's gallery illustrating a small segment of the variety of cell sizes and shapes found in the eukaryotic group, all referenced to the *E. coli* standard ruler. This gallery is by no means complete. There is much more variety than we can illustrate, but this covers a reasonable range of eukaryotic cell

Figure 2.15: Powers-of-10 representation of biological length scales. The hierarchy of scales is built around the *E. coli* standard ruler. Starting with *E. coli*, Section 2.2 considers a succession of 10-fold increases in resolution as are shown in the figure. Section 2.3 zooms out from the scale of an *E. coli* cell.

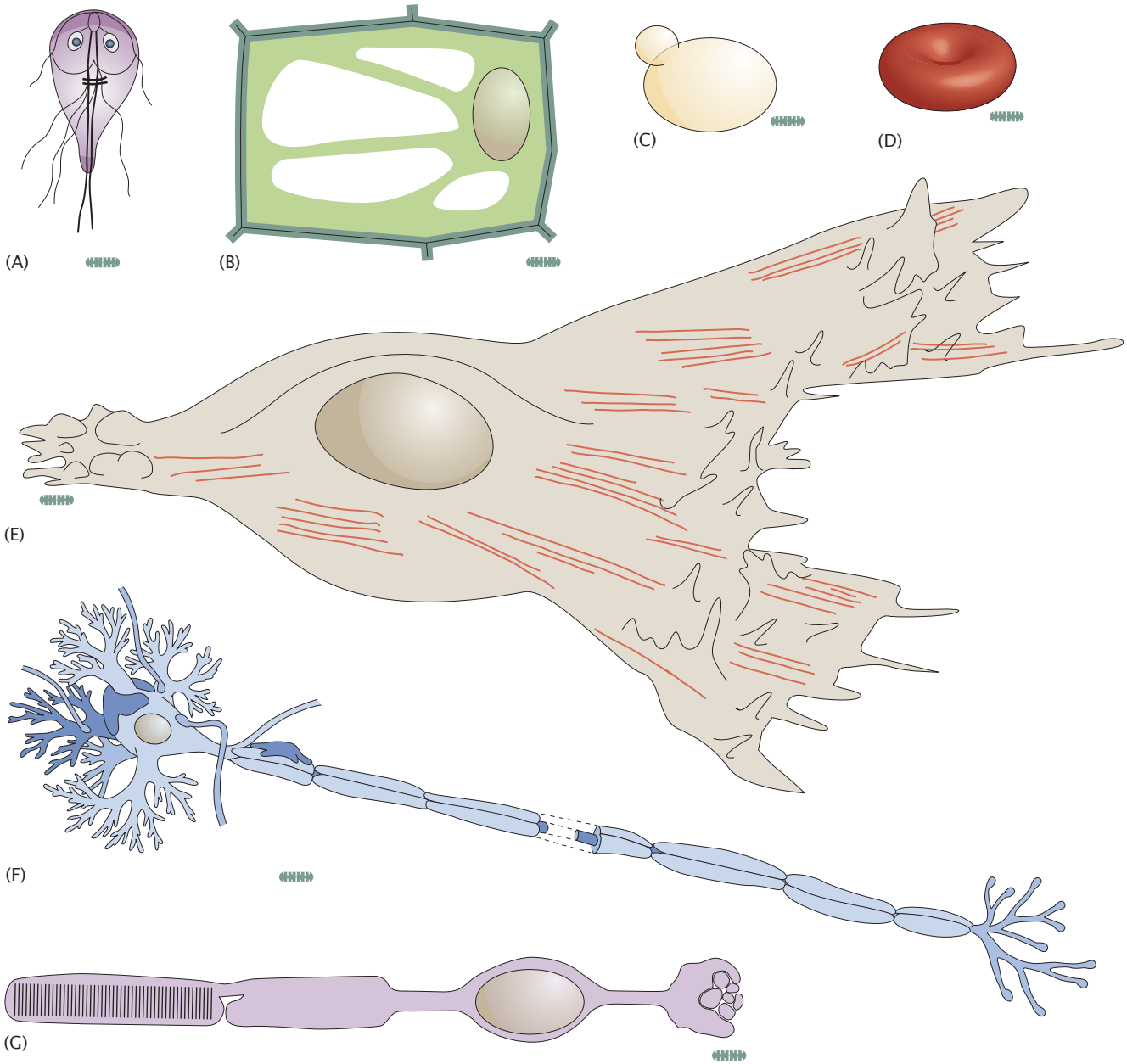


Figure 2.16: Cartoons of several different types of cells all referenced to the standard *E. coli* ruler. (A) The protist *Giardia lamblia*, (B) a plant cell, (C) a budding yeast cell, (D) a red blood cell, (E) a fibroblast cell, (F) a eukaryotic nerve cell, and (G) a retinal rod cell.

types that have been well studied by biologists. In this figure, we have chosen a variety of examples that represent experimental bias among biologists where more than half of the examples are human cells and the others represent the rest of the eukaryotic group.

The vast majority of eukaryotes are members of a group called protists, as shown in Figure 2.17. This poorly defined group encompasses all eukaryotes that are neither plants nor animals nor fungi. Protists are extremely diverse in their appearance and lifestyles, but they are all small (ranging from 0.002 mm to 2 mm). Some examples of protists are marine plankton such as *Emiliana huxleyi*, soil amoebae such as *Dictyostelium discoideum*, and the lovely creature *Paramecium* seen in any sample of pond water and familiar from many high-school biology



Figure 2.17: Protist diversity. This figure illustrates the morphological diversity of free-living protists. The various organisms are drawn to scale relative to the head of a pin. (Adapted from B. J. Finlay, *Science* 296:1061, 2002.)

classes. Another notable protist is the pathogen that causes malaria, called *Plasmodium falciparum*. Figure 2.16(A) shows the intriguing protist *Giardia lamblia*, a parasite known to hikers as a source of water contamination that causes diarrhea.

Although protists constitute the vast majority of eukaryotic cells on the planet, biologists are often inclined to study cells more closely related to or directly useful to us. This includes the plant kingdom, which is obviously important as a source of food and flowers. Plant cells like that shown in Figure 2.16(B) are characterized by a rigid cell wall, often giving them angular structures like that shown in the figure. The typical length scale associated with these cells is often tens of microns. One of the distinctive features of plant cells is their large vacuoles within the intracellular space that hold water and contribute to the mechanical properties of plant stems. These large vacuoles are very distinct from animal cells, where most of the intracellular space is filled with cytoplasm. Consequently, in comparing a plant and an animal cell of similar overall size, the plant cell will have typically 10-fold less cytoplasmic volume because most of its intracellular space is filled with vacuoles. Hydrostatic forces matter much more to plants than animals. For example, a wilting flower can be revived simply by application of water since this allows the vacuoles to fill and stiffen the plant stem.

Among the eukaryotes, the group most closely related to the animals (as proved by ribosomal RNA similarity and many other lines of evidence) is, surprisingly, the fungi. The representative fungus shown in Figure 2.16 is the budding yeast *Saccharomyces cerevisiae* (which we will refer to as *S. cerevisiae*). *S. cerevisiae* was domesticated by humans several thousand years ago and continues to serve as a treasured microbial friend that makes our bread rise and provides alcohol in our fermented beverages such as wine. Just as *E. coli* often serves as a key model prokaryotic system, the yeast cell often serves as the model single-celled eukaryotic organism. Besides the fact that humans are fond of *S. cerevisiae* for its own intrinsic properties, it is also useful to biologists as a representative fungus. Of all the other organisms on Earth, fungi are closest to animals in terms of evolutionary descent and similarity of protein functions. Although there are no single-celled animals, there are some single-celled fungi, including



S. cerevisiae. Therefore, many laboratory experiments relying on rapid replication of single cells are most easily performed on this organism and its relatives *Candida albicans* and *Schizosaccharomyces pombe* (a "fission" yeast that divides in the middle). Figure 2.18 shows a scanning electron microscope image of a yeast cell engaged in budding. As this image shows, the geometry of yeast is relatively simple compared with many other eukaryotic cells and it is also a fairly small member of this group, with a characteristic diameter of roughly 5 μm. Nonetheless, it possesses all the important structural hallmarks of the eukaryotes, including, in particular, a membrane-bound nucleus, separating the DNA genome from the cytoplasmic machinery that performs most metabolic functions.

Earlier, we estimated the molecular census of an *E. coli* cell. It will now be informative to compare those estimates with the corresponding model eukaryotic cell that will continue to serve as a comparative basis for all our eukaryotic estimates.

Figure 2.18: Microscopy images of yeast and their organelles. (A) Scanning electron micrograph of the yeast *Candida albicans* revealing the overall size scale of budding yeast. (B) Electron microscopy image of a cross-section through a budding *Candida albicans* yeast cell. (C) Confocal microscopy images of the mitochondria of *S. cerevisiae* under different growth conditions. (A, courtesy of Ira Herskowitz and Eric Schabtach; B, adapted from G. M. Walker, Yeast, Physiology and Biotechnology, John Wiley, 1998; C, adapted from A. Egner et al., *Proc. Natl Acad. Sci. USA* 99:3370, 2002.)

Estimate: Sizing Up Yeast The volume of a yeast cell can be computed in *E. coli* volume units, $V_{E.\ coli}$. In particular, if we recall that $V_{E.\ coli} \approx 1.0\,\mu m^3 = 1.0$ fL and think of yeast as a sphere of diameter 5 μm, then we have the relation $V_{yeast} \approx 60\,V_{E.\ coli}$; that is, roughly 60 *E. coli* cells would fit inside of a yeast cell. This result is consistent with experimental observations of the yeast cell size distribution such as shown in Figure 2.19. The surface area of a yeast cell can be estimated using a radius of $r_{yeast} \approx 2.5\,\mu m$, which yields $A_{yeast} \approx 80\,\mu m^2$. If we treat the yeast nucleus as a sphere with a diameter of roughly 2.0 μm, its volume is roughly $4\,\mu m^3$. Within this nucleus is housed the 1.2×10^7 bp of the yeast genome, which is divided among 16 chromosomes. The DNA in yeast is packed into higher-order structures mediated by protein assemblies known as histones. In particular, the DNA is wrapped around a series of cylindrical cores made up of eight such histone proteins each, with roughly 150 bp wrapped around each histone octamer, and approximately a 50 bp spacer between. As a result, we can estimate the number of nucleosomes (the histone–DNA complex) as

$$N_{nucleosome} \approx \frac{12 \times 10^6\,bp}{200\,bp/nucleosome} \approx 60{,}000. \qquad (2.13)$$

Experimentally, the measured number appears to be closer to 80,000, with a mean spacing between nucleosomes of the order

ESTIMATE

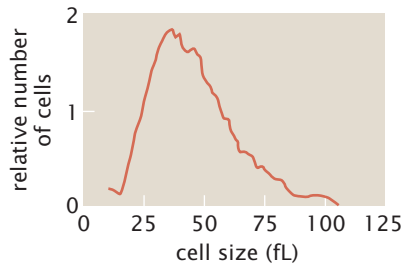


Figure 2.19: Yeast cell size distribution. Distribution of cell volumes measured for wild-type yeast cells. (Adapted from P. Jorgensen et al., *Science* 297:395, 2002.)

of 170 bp. The total volume taken up by the histones is roughly $230\,\text{nm}^3$ per histone (thinking of each histone octamer as a cylindrical disk of radius 3.5 nm and height 6 nm), resulting in a total volume of $14 \times 10^6\,\text{nm}^3$ taken up by the histones. The volume taken up by the genome itself is comparable at $1.2 \times 10^7\,\text{nm}^3$, where we have used the rule of thumb that the volume per base pair is $1\,\text{nm}^3$. The packing fraction (defined as the ratio of the volume taken up by the genome to the volume of the nucleus) associated with the yeast genomic DNA can be estimated by evaluating the ratio

$$\rho_{\text{pack}} \approx \frac{(1.2 \times 10^7\,\text{bp}) \times (1\,\text{nm}^3/\text{bp})}{4 \times 10^9\,\text{nm}^3} \approx 3 \times 10^{-3}. \tag{2.14}$$

Note that we have used the fact that the yeast genome is 1.2×10^7 bp in length and is packed in the nucleus, which has a volume $\approx 4\,\mu\text{m}^3$.

These geometric estimates may be used to make corresponding molecular estimates, such as the number of lipids and proteins in a typical yeast cell. The number of proteins can be estimated in several ways—perhaps the simplest is just to assume that the fractional occupancy of yeast cytoplasm is identical to that of *E. coli*, with the result that there will be 60 times as many proteins in yeast as in *E. coli* based strictly on scaling up the cytoplasmic volume. This simple estimate is obtained by *assuming* that the composition of the yeast interior is more or less the same as that of an *E. coli* cell. This strategy results in

$$N_{\text{protein}}^{\text{yeast}} \approx 60 \times N_{\text{protein}}^{\textit{E. coli}} \approx 2 \times 10^8. \tag{2.15}$$

The number of lipids associated with the plasma membrane of the yeast cell can be obtained as

$$N_{\text{lipid}} \approx \frac{2 \times 0.5 \times A_{\text{yeast}}}{A_{\text{lipid}}} \approx \frac{2 \times 0.5 \times (80 \times 10^6\,\text{nm}^2)}{0.5\,\text{nm}^2} \approx 2 \times 10^8, \tag{2.16}$$

where the factor of 0.5 is based on the idea that roughly half of the surface area is covered by membrane proteins rather than lipids themselves and the factor of 2 accounts for the fact that the membrane is a bilayer. This should be contrasted with the situation in *E. coli*, which has a double membrane surrounding the cytoplasm.

Another interesting estimate suggested by Figure 2.18(C) is associated with the organellar content of these cells. In particular, this figure shows the mitochondria of yeast that are being grown in two different media. These pictures suggest several interesting questions such as what fraction of the cellular volume is occupied by mitochondria and what is the surface area tied up with the mitochondrial outer membranes? Using the image of the cells grown on glucose in the left of Figure 2.18(C), we see that the mitochondria have a reticular form made up of several long tubes with a radius of roughly $r = 100$ nm. These tubes are nearly circular like the inner tubes of a bicycle with

a radius of curvature of approximately $R = 1\,\mu m$, meaning that each such tube has a length of roughly $L = 2\pi R \approx 6\,\mu m$, for a total tube length of 12 μm given that there appear to be two such tubes. Hence, the volume taken up by the mitochondria is $V = \pi r^2(2L) \approx 0.5\ \mu m^3$, with an area of $A = 2\pi r(2L) \approx 6\ \mu m^2$, meaning that about 1% of the volume of the cells grown on glucose is taken up by their mitochondria and an area is tied up in the mitochondrial outer membrane that is roughly 10% of the total plasma membrane area. In other cell types, the fraction of the area taken up by the mitochondria can be much larger, as will be seen in Chapter 11 (p. 427).

Our estimates are brought into sharpest focus when they are juxtaposed with actual measurements. The census of yeast cells has been performed in several distinct and fascinating ways. For a series of recent studies, the key idea is to generate thousands of different yeast strains, a strain library, each of which has a tag on a different one of the yeast gene products. Each yeast protein can then be specifically recognized by antibodies in the specific strain where it has been tagged. A second scheme is to construct protein fusions in which the protein of interest is attached to a fluorescent protein such as the green fluorescent protein (GFP). Then, by querying each and every yeast strain in the library either by examining the extent of antibody binding or fluorescence, it is possible to count up the numbers of each type of protein. Figure 2.20(B) shows a histogram of the number of proteins that occur with a given protein copy number in yeast. Although the average protein appears to be present in the cell with a few thousand copies, this histogram shows that some proteins are present with fewer than 50 copies and others with more than a million copies. Further, similarly quantifying mRNA as shown in Figure 2.20(C) reveals that many genes, including essential genes, are expressed with an average of less than one molecule of RNA per cell. By adding up the total number of proteins on the basis of this census, we estimate there are 50×10^6 proteins in a yeast cell, somewhat less than suggested by our crude estimate given above. Now that we have completed our first introduction to the important yeast as a representative eukaryote, we return to our tour of cell types in Figure 2.16.

Cells from Humans Have a Huge Diversity of Structure and Function

The remainder of the cells in Figure 2.16 are all human cells and show another interesting aspect of cellular diversity. To a first approximation, every cell in the human body contains the same DNA genome. And yet, individual human cells differ significantly with respect to their sizes (with sizes varying from roughly 5 μm for red blood cells to 1 m for the largest neurons), shapes, and functions. For example, rod cells in the retina are specialized to detect incoming light and transmit that information to the neural system so that we can see. Red blood cells are primarily specialized as carriers of oxygen and, in fact, are dramatically different from almost all other cells in having dispensed with their nucleus as part of their developmental process. As we will discuss extensively throughout the book, other cells have other unique classes of specialization.

Figure 2.16(D) is a schematic of the structure of a red blood cell. Note that the shapes of these cells are decidedly not spherical, raising interesting questions about the mechanisms of cell-shape maintenance. Despite their characteristic size of order 5 μm, these

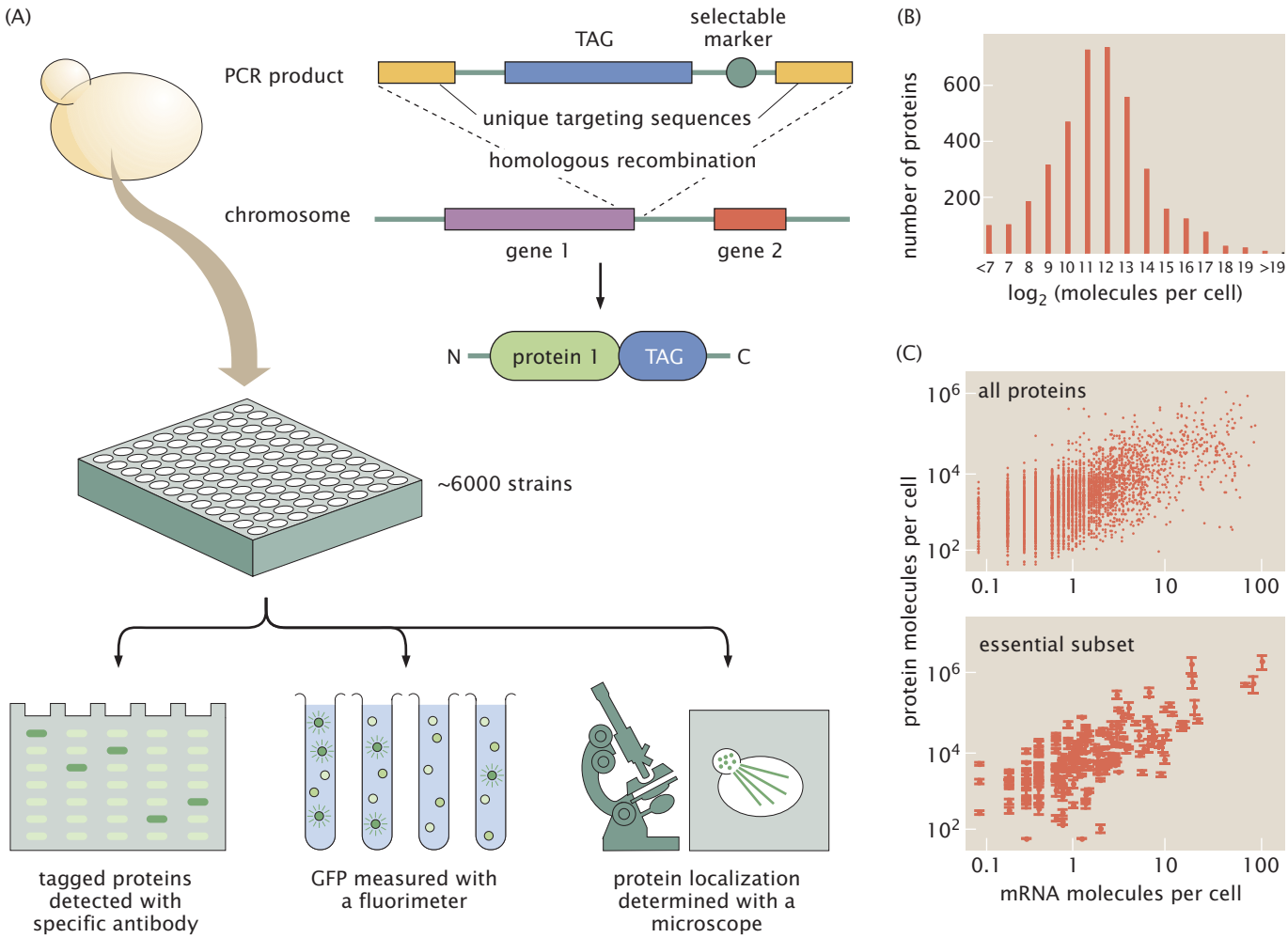


Figure 2.20: Protein copy numbers in yeast. (A) Schematic of constructs used to measure the protein census and several different methods for quantifying protein in labeled cells. The "tag" attached to each gene may encode a fluorescent protein or a site for antibody recognition, depending on the experiment. (B) Result of antibody detection of various proteins in yeast showing the number of proteins that have a given copy number. The number of copies of the protein per cell is expressed in powers of 2 as 2^N. (C) Abundances of various proteins as a function of their associated mRNA copy numbers. The bottom plot shows this result for essential soluble proteins. (A–C, adapted from S. Ghaemmaghami et al. *Nature* 425:737, 2003.)

cells easily pass through capillaries with less than half their diameter as shown in Figure 2.21, implying that their shape is altered significantly as part of their normal life cycle. While in capillaries (either artificial or *in vivo*), the red blood cell is severely deformed to pass through the narrow passage. In their role as the transport vessels for oxygen-rich hemoglobin, these cells will serve as an inspiration for our discussion of the statistical mechanics of cooperative binding. Red blood cells are a target of one of the most common infectious diseases suffered by humans, malaria, caused by the invasion of a protozoan. Malaria-infected red blood cells are much stiffer than normal cells and cannot deform to enter small capillaries. Consequently, people suffering from malaria experience severe pain and damage to tissues because of the inability of their red blood cells to enter those tissues and deliver oxygen.

One of the most commonly studied eukaryotic cells from multicellular organisms is the fibroblast, which is shown schematically in Figure 2.16(E) and in an AFM image in Figure 2.22. These cells will serve as a centerpiece for much of what we will have to say about

"typical" eukaryotic cells in the remainder of the book. Fibroblasts are associated with animal connective tissue and are notable for secreting the macromolecules of the extracellular matrix.

Cells in multicellular organisms can be even more exotic. For example, nerve cells (Figure 2.16F) and rod cells (Figure 2.16G) reveal a great deal more complexity than the examples highlighted above. In these cases, the cell shape is intimately related to its function. In the case of nerve cells, their sinewy appearance is tied to the fact that the various branches (also called "processes") known as dendrites and axons convey electrical signals that permit communication between distant parts of an animal's nervous system. Despite having nuclei with typical eukaryotic dimensions, the cells themselves can extend processes with characteristic lengths of up to tens of centimeters. The structural complexity of rod cells is tied to their primary function of light detection in the retina of the eye. These cells are highly specialized to perform transduction of light energy into chemical energy that can be used to communicate with other cells in the body and, in particular, with brain cells that permit us to be conscious of perceiving images. Rod cells accomplish this task using large stacks of membranes which are the antennas participating in light detection. Figure 2.16 only scratches the surface of the range of cellular size and shape, but at least conveys an impression of cell sizes relative to our standard ruler.

2.2.2 The Cellular Interior: Organelles

As we descend from the scale of the cell itself, a host of new structures known as organelles come into view. The presence of these membrane-bound organelles is one of the defining characteristics that distinguishes eukaryotes from bacteria and archaea. Figure 2.23 shows a schematic of a eukaryotic cell and an associated electron microscopy image revealing some of the key organelles. These organelles serve as the specialized apparatus of cell function in capacities ranging from genome management (the nucleus) to energy generation (mitochondria and chloroplasts) to protein synthesis and modification (endoplasmic reticulum and Golgi apparatus) and beyond. The compartments that are bounded by organellar membranes can have completely different protein and ion compositions. In addition, the membranes of each of these different membrane systems are characterized by distinct lipid and protein compositions.

A characteristic feature of many organelles is that they are compartmentalized structures that are separated from the rest of the cell by membranes. The nucleus is one of the most familiar examples since it is often easily visible using standard light microscopy. If we use

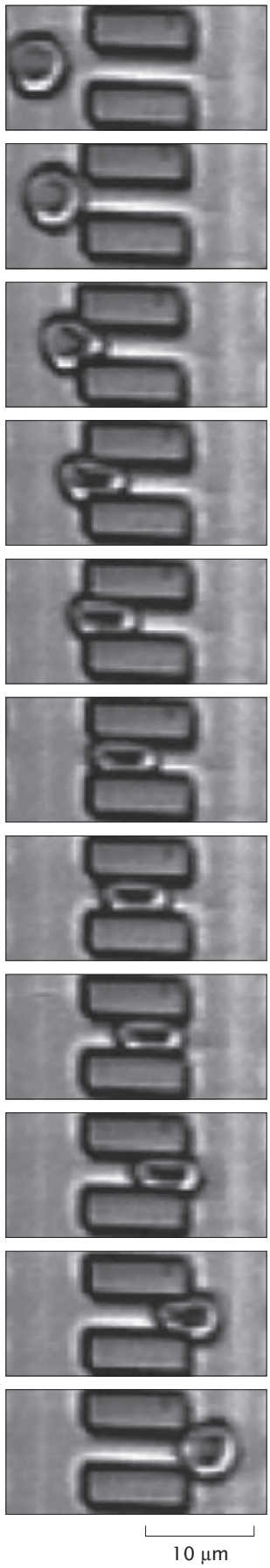


Figure 2.21: Deformability of red blood cells. To measure the deformability of human red blood cells, an array of blocks was fabricated in silicon, each block was $4\,\mu m \times 4\,\mu m \times 12\,\mu m$. The blocks were spaced by $4\,\mu m$ in one direction and $13\,\mu m$ in the other. A glass coverslip covered the top of this array of blocks. A dilute suspension of red blood cells in a saline buffer was introduced to the system. A slight pressure applied at one end of the array of blocks provided bulk liquid flow, from left to right in the figure. This liquid flow carried the red blood cells through the narrow passages. Video microscopy captured the results. The figure shows consecutive video fields with the total elapsed time just over one-third of a second. (Courtesy of J. Brody.)

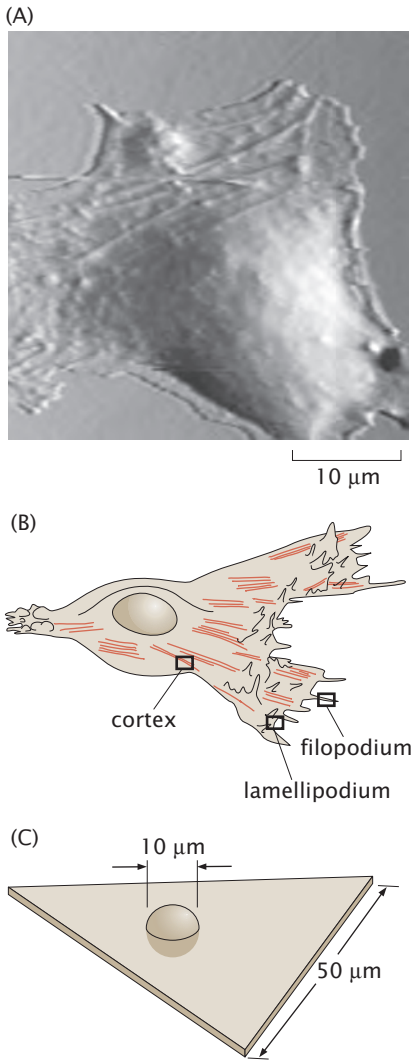


Figure 2.22: Structure of a fibroblast. (A) Atomic-force microscopy image of a fibroblast, (B) cartoon of the external morphology of a fibroblast, and (C) characteristic dimensions of a "typical" fibroblast. (A, adapted from M. Radmacher, *Meth. Cell Biol.* 83:347, 2007.)

the fibroblast as an example, then the cell itself has dimensions of roughly 50 μm, while the nucleus has a characteristic linear dimension of roughly 10 μm, as shown schematically in Figure 2.22. From a functional perspective, the nucleus is much more complex than simply serving as a storehouse for the genetic material. Chromosomes are organized within the nucleus, forming specific domains as will be discussed in more detail in Chapter 8. Transcription as well as several kinds of RNA processing occur in the nucleus. There is a flux of molecules such as transcription factors moving in and completed RNA molecules moving out through elaborate gateways in the nuclear membrane known as nuclear pores. Portions of the genome involved in synthesis of ribosomal RNA are clustered together, forming striking spots that can be seen in the light microscope and are called nucleoli.

Moving outward from the nucleus, the next membranous organelle we encounter is often the endoplasmic reticulum. Indeed, the membrane of the endoplasmic reticulum is contiguous with the membrane of the nuclear envelope. In some cells such as the pancreatic cell shown in Figure 2.24, the endoplasmic reticulum takes up the bulk of the cell interior. This elaborate organelle is the site of lipid synthesis and also the site of synthesis of proteins that are destined to be secreted or incorporated into membranes. It is clear from images such as those in Figures 2.24 and 2.25 that the endoplasmic reticulum can assume different geometries in different cell types and under different conditions. How much total membrane area is taken up by the endoplasmic reticulum? How strongly does the specific membrane morphology affect the total size of the organelle?

ESTIMATE

Estimate: Membrane Area of the Endoplasmic Reticulum

One of the most compelling structural features of the endoplasmic reticulum is its enormous surface area. To estimate the area associated with the endoplasmic reticulum, we take our cue from Figure 2.24, which suggests that we think of the endoplasmic reticulum as a series of concentric spheres centered about the nucleus. We follow Fawcett (1966), who characterizes the endoplasmic reticulum as forming "lamellar systems of flat cavities, rather uniformly spaced and parallel to one another" as shown in Figure 2.24.

An estimate can be made by adding up the areas from each of the concentric spheres making up our model endoplasmic reticulum. This can be done by simply noticing that the volume enclosed by the endoplasmic reticulum can be written as

$$V_{\mathrm{ER}} = \sum_i A_i d, \tag{2.17}$$

where A_i is the area of the ith concentric sphere and d is the distance between adjacent cisternae. Since two membranes bound each cisterna, the total area of the endoplasmic reticulum membrane is $A_{\mathrm{ER}} = 2 \times \sum_i A_i$. In our model, the total volume of the endoplasmic reticulum can be written as the difference between the volume taken up by the outermost sphere and the volume of the innermost concentric sphere (which is the same as the volume of the nucleus). This results in

$$V_{\mathrm{ER}} = \frac{4\pi}{3} R_{\mathrm{out}}^3 - \frac{4\pi}{3} R_{\mathrm{nucleus}}^3. \tag{2.18}$$

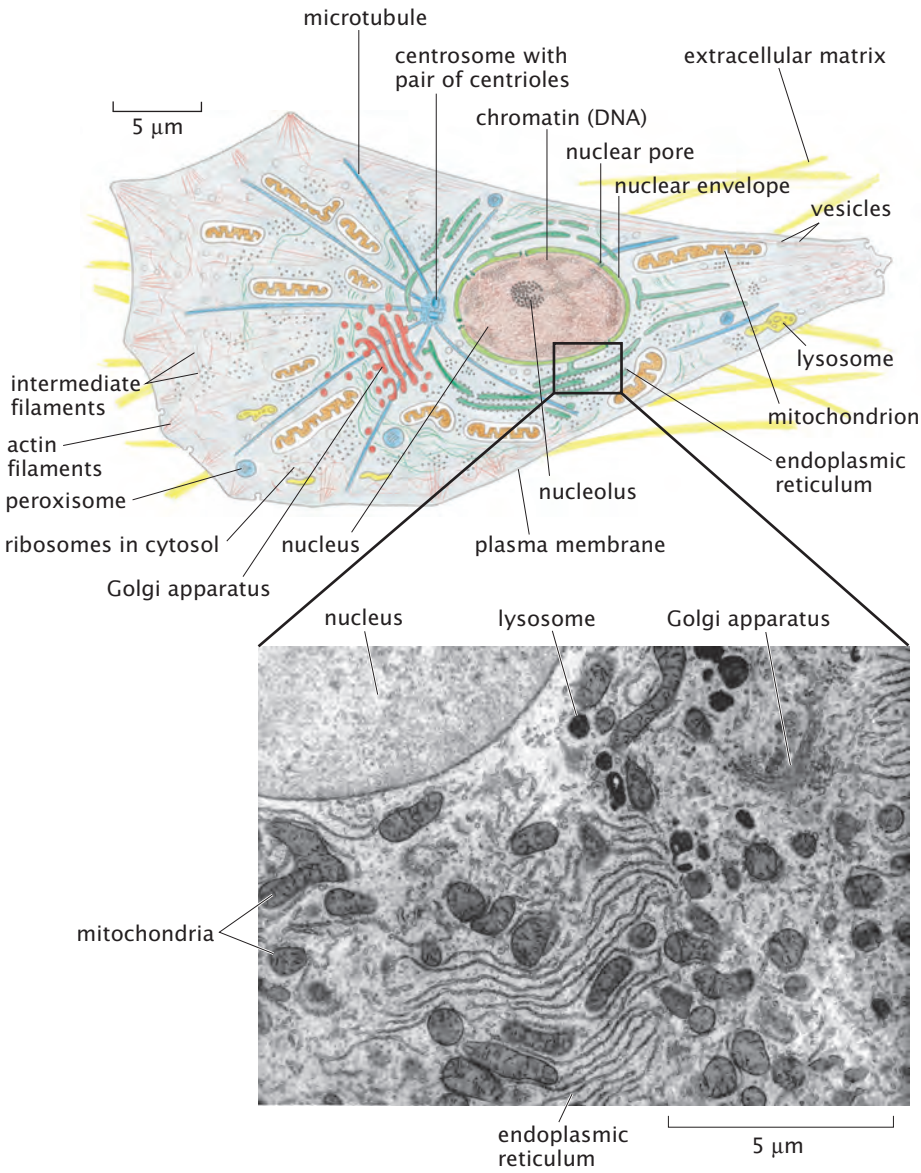


Figure 2.23: Eukaryotic cell and its organelles. The schematic shows a eukaryotic cell and a variety of membrane-bound organelles. A thin-section electron microscopy image shows a portion of a rat liver cell approximately equivalent to the boxed area on the schematic. A portion of the nucleus can be seen in the upper left-hand corner. The most prominent organelles visible in the image are mitochondria, lysosomes, the rough endoplasmic reticulum, and the Golgi apparatus. (Eukaryotic cell from Alberts et al., Molecular Biology of the Cell, 5th ed., New York, Garland Science, 2008; electron micrograph from D. W. Fawcett, The Cell, Its Organelles and Inclusions: An Atlas of Fine Structure. W. B. Saunders, 1966.)

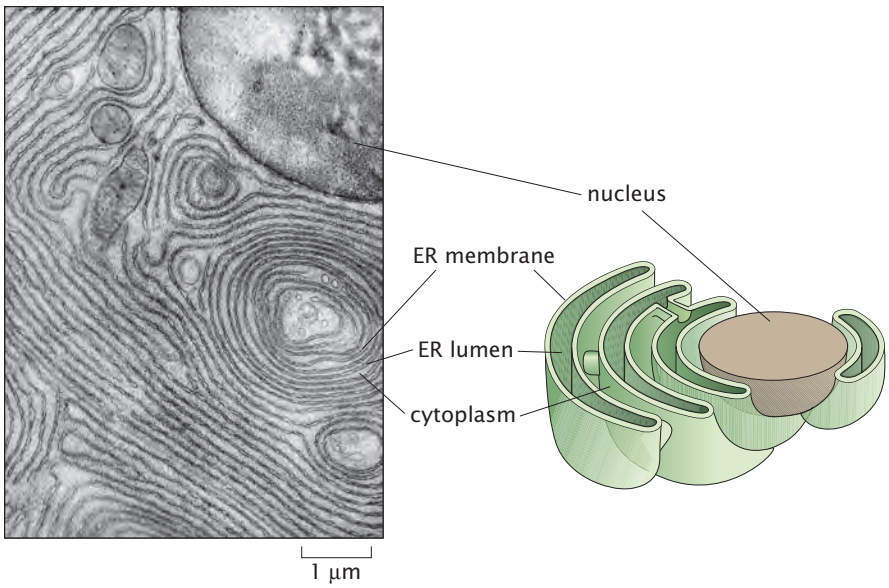


Figure 2.24: Electron micrograph and associated schematic of the endoplasmic reticulum (ER). The left-hand panel shows a thin-section electron micrograph of an acinar cell from the pancreas of a bat. The nucleus is visible at the upper right and the dense and elaborate endoplasmic reticulum structure is strikingly evident. The right-hand panel shows a schematic diagram of a model for the three-dimensional structure of the endoplasmic reticulum in this cell. Notice that the size of the lumen in the endoplasmic reticulum in the schematic is exaggerated for ease of interpretation. (Electron micrograph from D. W. Fawcett, The Cell, Its Organelles and Inclusions: An Atlas of Fine Structure. W. B. Saunders, 1966.)

Combining the two ways of computing the volume of the endoplasmic reticulum, Equations 2.17 and 2.18, we arrive at an expression for the endoplasmic reticulum area,

$$A_{ER} = \frac{8\pi}{3d}\left(R_{out}^3 - R_{nucleus}^3\right). \tag{2.19}$$

Using the values $R_{nucleus} = 5\,\mu m$, $R_{out} = 10\,\mu m$, and $d = 0.05\,\mu m$, we get, at an estimate, $A_{ER} = 15 \times 10^4\,\mu m^2$. This result should be contrasted with a crude estimate for the outer surface area of a fibroblast, which can be obtained by using the dimensions in Figure 2.22(C) and which yields an area of $10^4\,\mu m^2$ for the cell membrane itself. To estimate the area of the endoplasmic reticulum when it is in reticular form, we describe its structure as interpenetrating cylinders of diameter $d \approx 10\,nm$ separated by a distance $a \approx 60\,nm$, as shown in Figure 2.25. The completion of the estimate is left to the problems, but results in a comparable membrane area to the calculation here assuming concentric spheres.

The other major organelles found in most cells and visible in Figure 2.23 include the Golgi apparatus, mitochondria, and lysosomes. The Golgi apparatus, similar to the endoplasmic reticulum, is largely involved in processing and trafficking of membrane-bound and secreted proteins (for another view, see Figure 11.38 on p. 463). The Golgi apparatus is typically seen as a pancake-like stack of flattened compartments, each of which contains a distinct set of enzymes. As proteins are processed for secretion, for example by addition and remodeling of attached sugars, they appear to pass in an orderly fashion through each element in the Golgi stack. The mitochondria are particularly striking organelles with a smooth outer surface housing an elaborately folded system of internal membrane structures (for a more detailed view, see Figure 11.39 on p. 464). The mitochondria are the primary site of ATP synthesis for cells growing in the presence of oxygen, and their fascinating physiology and structure have been well studied. We will return to the topic of mitochondrial structure in Chapter 11 and discuss the workings of the tiny machine responsible for ATP synthesis in Chapter 16. Lysosomes serve a major role in the degradation of cellular components. In some specialized cells such as macrophages, lysosomes also serve as the compartment where bacterial invaders can be degraded. These membrane-bound organelles are filled with acids, proteases, and other degradative enzymes. Their shapes are polymorphous; resting lysosomes are simple and nearly spherical, whereas lysosomes actively involved in degradation of cellular components or of objects taken in from the outside may be much larger and complicated in shape.

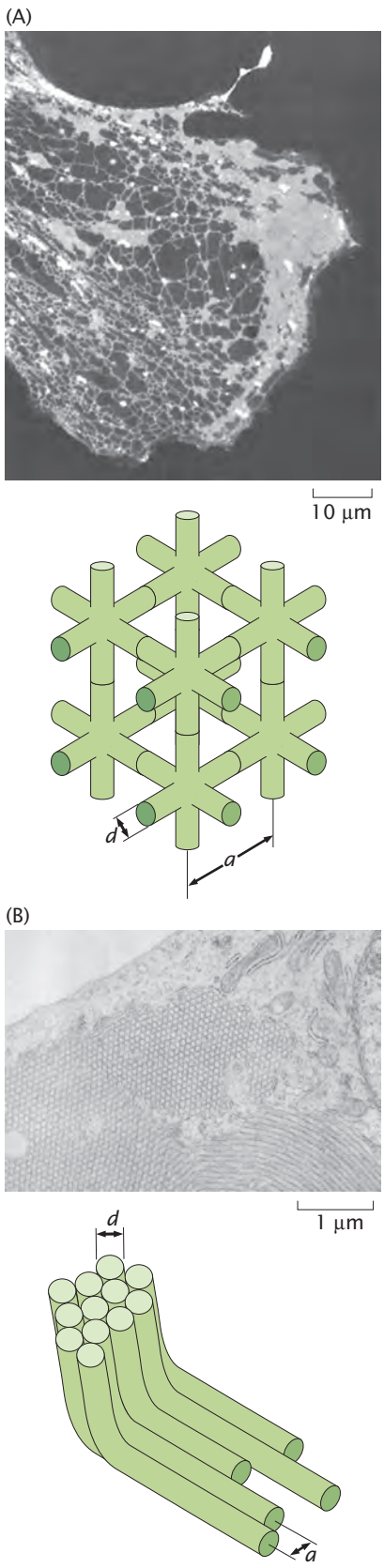


Figure 2.25: Variable morphology of the endoplasmic reticulum. (A) In most cultured cells, the endoplasmic reticulum is a combination of a web-like reticular network of tubules and larger flattened cisternae. In this image, a cultured fibroblast was stained with a fluorescent dye called DiOC6 that specifically labels endoplasmic reticulum membrane. On the bottom is a schematic of an idealized three-dimensional reticular network. (B) Some specialized cells and those treated with drugs that regulate the synthesis of lipids reorganize their endoplasmic reticulum to form tightly-packed, nearly crystalline arrays that resemble piles of pipes. (A, adapted from M. Terasaki et al., *J. Cell. Biol.* 103:1557, 1986; B, adapted from D. J. Chin et al., *Proc. Natl Acad. Sci. USA* 79:1185, 1982.)

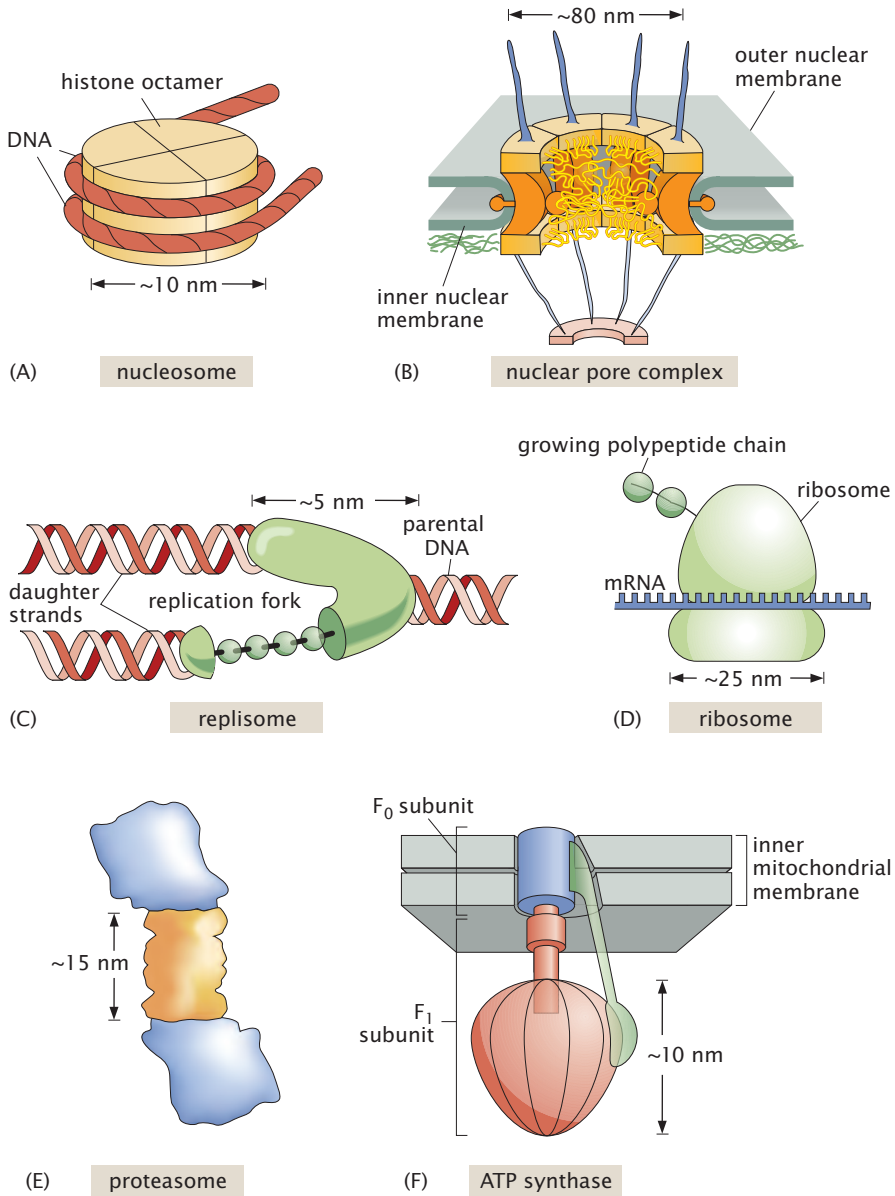


Figure 2.26: The macromolecular assemblies of the cell. (A) The nucleosome is a complex of eight protein units and DNA. (B) The nuclear pore complex mediates the transport of material in and out of the nucleus. (C) The replisome is a complex of proteins that mediate the copying of the genetic material. (D) The ribosome reads mRNA and synthesizes the corresponding polypeptide chain. (E) The proteasome degrades proteins. (F) ATP synthase is a large complex that synthesizes new ATP molecules.

These common organelles are only a few of those that can be found in eukaryotic cells. Some specialized cells have remarkable and highly specialized organelles that can be found nowhere else, such as the stacks of photoreceptive membranes found in the rod cells of the visual system as indicated schematically in Figure 2.16(G) (p. 53). The common theme is that all organelles are specialized subcompartments of the cell that perform a particular subset of cellular tasks and represent a smaller, discrete layer of organization one step down from the whole cell.

2.2.3 Macromolecular Assemblies: The Whole is Greater than the Sum of the Parts

Macromolecules Come Together to Form Assemblies

Proteins, nucleic acids, sugars, and lipids often work as a team. Indeed, as will become clear throughout the remainder of this book,

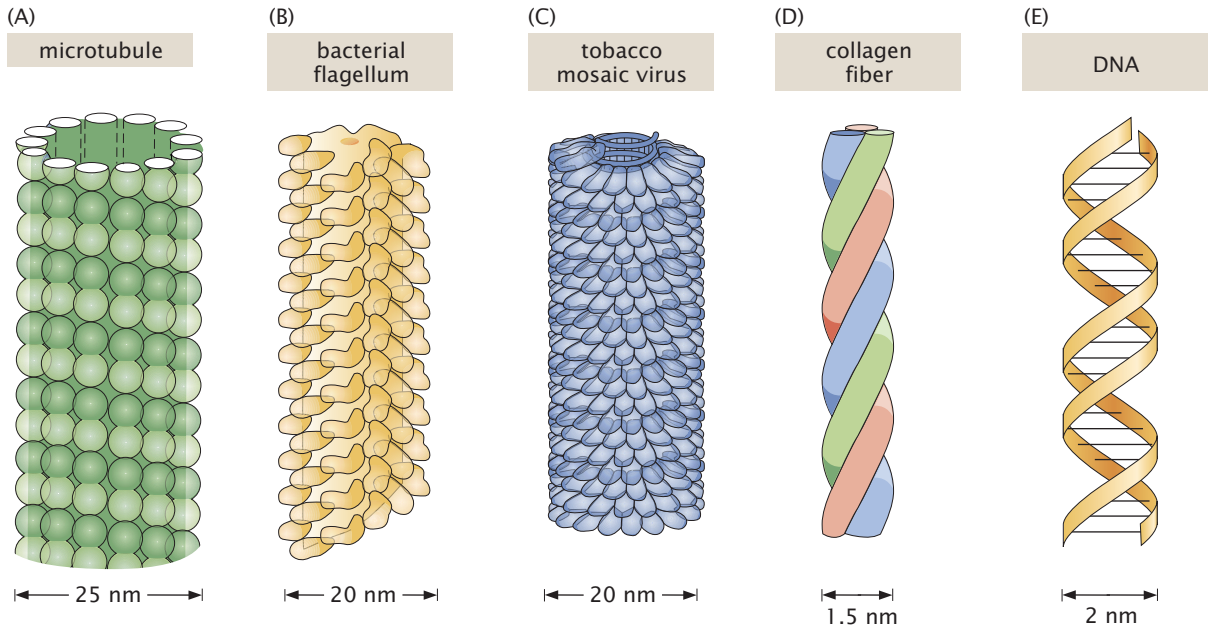


Figure 2.27: Helical motifs of molecular assemblies. Macromolecular assemblies have a variety of different helical structures, some formed from individual monomeric units (such as (A)–(C)), others resulting from coils of proteins (D), and yet others made up of paired nucleotides (E).

these macromolecules often come together to make assemblies. We think of yet another factor-of-10 magnification relative to the previous section, and with this increase of magnification we see assemblies such as those shown in cartoon form in Figure 2.26. The genetic material in the eukaryotic nucleus is organized into chromatin fibers, which themselves are built up of protein–DNA assemblies known as nucleosomes (Figure 2.26A). Traffic of molecules such as completed RNA molecules out of the nucleus is mediated by the elaborate nuclear pore complex (Figure 2.26B). The replication complex that copies DNA before cell division is similarly a collection of molecules, which has been dubbed the replisome (Figure 2.26C). When the genetic message is exported to the cytoplasm for translation into proteins, the ribosome (an assembly of proteins and nucleic acids) serves as the universal translating machine that converts the nucleic acid message from the RNA into the protein product written in the amino acid alphabet (Figure 2.26D). When proteins have been targeted for degradation, they are sent to another macromolecular assembly known as the proteasome (Figure 2.26E). The production of ATP in mitochondria is similarly mediated by a macromolecular complex known as ATP synthase (Figure 2.26F) . The key idea of this subsection is to show that there is a very important level of structure in cells that is built around complexes of individual macromolecules and with a characteristic length scale of 10 nm.

Helical Motifs Are Seen Repeatedly in Molecular Assemblies

A second class of macromolecular assemblies, characterized not by function but rather by structure, is the wide variety of helical macromolecular complexes. Several representative examples are depicted in Figure 2.27. In Figure 2.27(A), we show the geometric structure of microtubules. As will be described in more detail later, these structures are built up of individual protein units called tubulin. A second example shown in Figure 2.27(B) is the bacterial flagellum of *E. coli*. Here too, the same basic structural idea is repeated, with the helical geometry built up from individual protein units, in this case

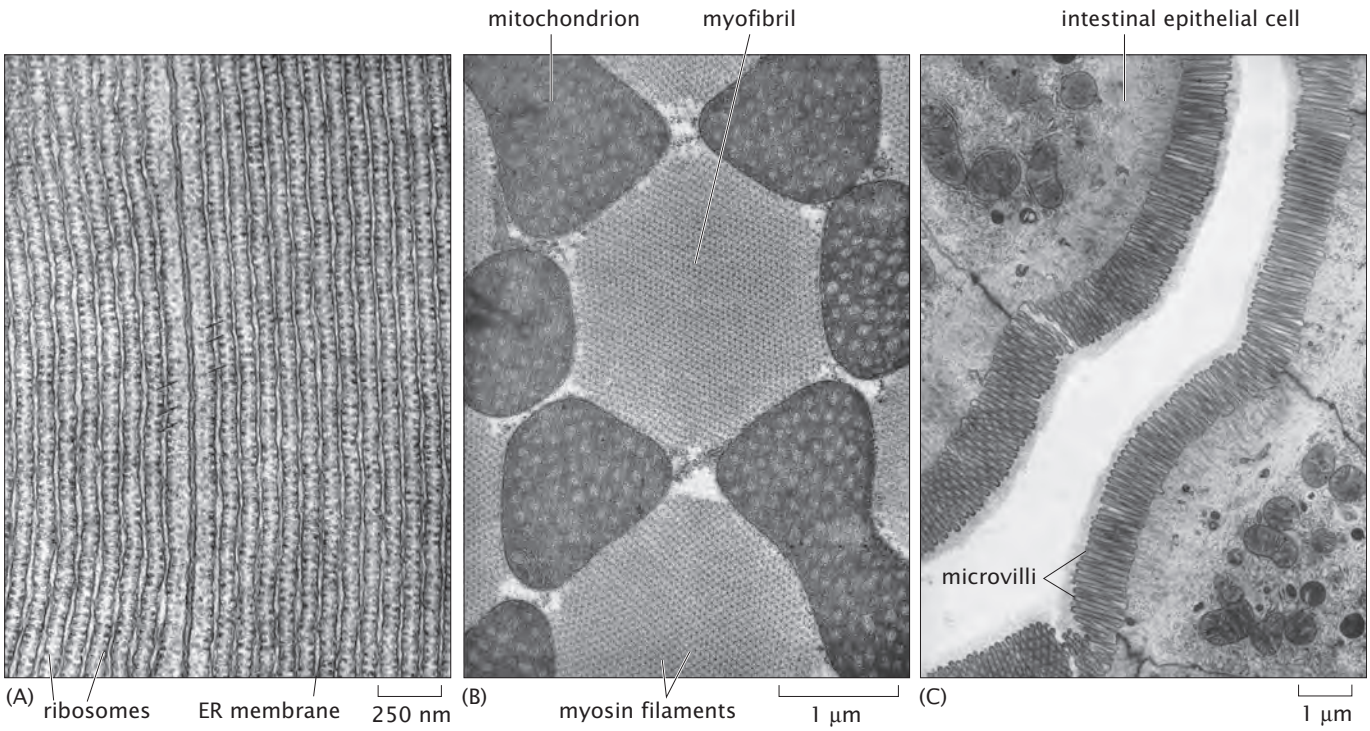


Figure 2.28: Ordered macromolecular assemblies. Collage of examples of macromolecules organized into superstructures. (A) Ribosomes on the endoplasmic reticulum ("rough ER"), (B) myofibrils in the flight muscle, and (C) microvilli at the epithelial surface. (A–C, adapted from D. W. Fawcett, The Cell, Its Organelles and inclusions: An Atlas of Fine Structure. W. B. Saunders, 1966.)

flagellin. The third example given in the figure is that of a filamentous virus, with tobacco mosaic virus (TMV) chosen as one of the most well-studied viruses.

The helical assemblies described above are characterized by individual protein units that come together to form helical filaments. An alternative and equally remarkable class of filaments are those in which α-helices (chains of amino acids forming protein subunits with a precise, helical geometry) wind around each other to form superhelices. The particular case study that will be of most interest in subsequent discussions is that of collagen, which serves as one of the key components in the extracellular matrix of connective tissues and is one of the major protein products of the fibroblast cells introduced earlier in the chapter (see Figure 2.16 on p. 53). The final example is the iconic figure of the DNA molecule itself.

Macromolecular Assemblies Are Arranged in Superstructures

Assemblies of macromolecules can interact with each other to create striking instances of cellular hardware with a size comparable to organelles themselves. Figure 2.28 depicts several examples. Figure 2.28(A) shows the way in which ribosomes are organized on the endoplasmic reticulum with a characteristic spacing that is comparable to the size of the ribosomes (≈20 nm). A second stunning example is the organization of myofibrils in muscles as shown in Figure 2.28(B). This figure shows the juxtaposition of the myofibrils and mitochondria. The myofibrils themselves are an ordered arrangement of actin filaments and myosin motors, as will be discussed in more detail in

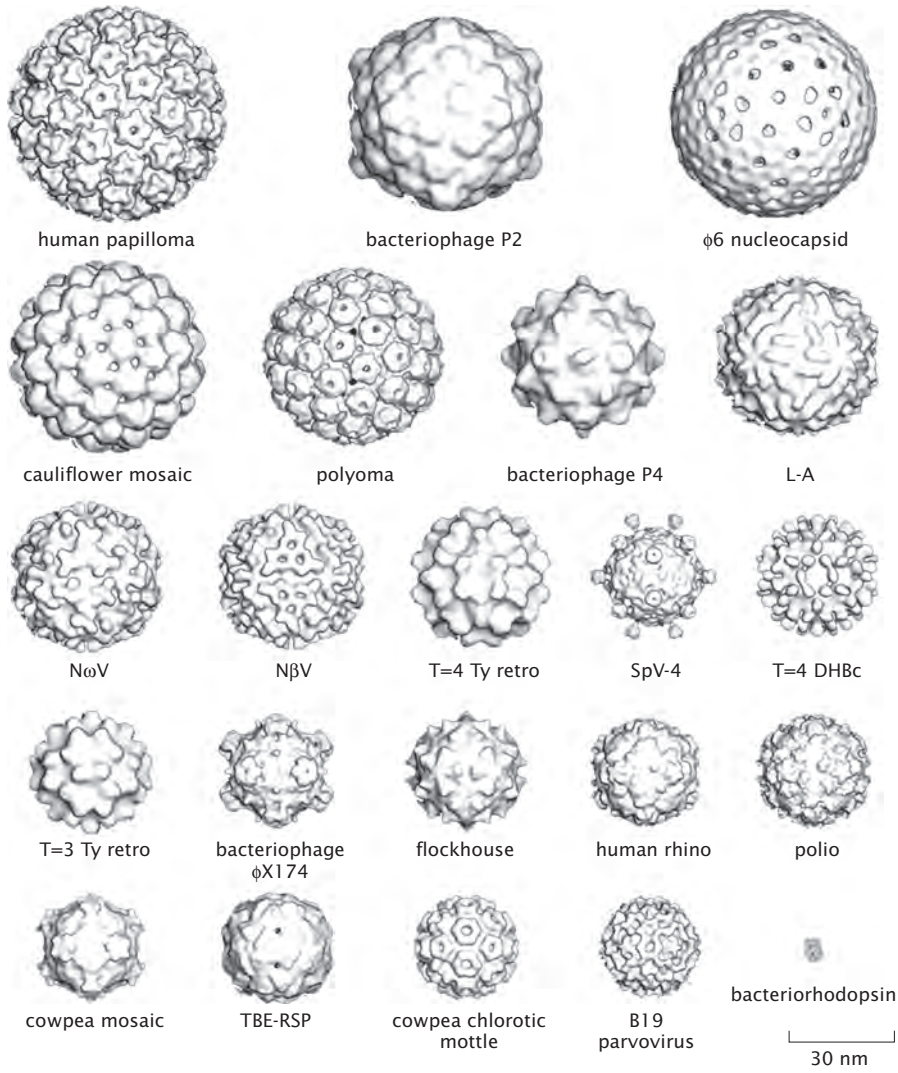


Figure 2.29: Structures of viral capsids. The regularity of the structure of viruses has enabled detailed, atomic-level analysis of their construction patterns. This gallery shows a variety of the different geometries explored by the class of nearly spherical viruses. For size comparison, a large protein, bacteriorhodopsin, is shown in the bottom right. (Adapted from T. S. Baker et al., *Microbiol. Mol. Biol. Rev.* 63:862, 1999.)

Chapter 16. The last example, shown in Figure 2.28(C), is of the protrusions of microvilli at the surface of an epithelial cell. These microvilli are the result of collections of parallel actin filaments. The list of examples of orchestration of collections of macromolecules could go on and on and should serve as a reminder of the many different levels of structural organization found in cells.

2.2.4 Viruses as Assemblies

Viruses are one of the most impressive and beautiful examples of macromolecular assembly. These assemblies are a collection of proteins and nucleic acids (though many viruses have lipid envelopes as well) that form highly ordered and symmetrical objects with characteristic sizes of tens to hundreds of nanometers. The architecture of viruses is usually a protein shell where the so-called capsid is made up of a repetitive packing of the same protein subunits over and over to form an icosahedron. Within the capsid, the virus packs its genetic material, which can be either single-stranded or double-stranded DNA or RNA depending upon the type of virus. Figure 2.29 is a gallery of the capsids of a number of different viruses. Different viruses have different elaborations on this basic structure and can include lipid coats,

surface receptors, and internal molecular machines such as polymerases and proteases. One of the most amazing features of viruses is that by hijacking the machinery of the host cell, the viral genome commands the construction of its own inventory of parts within the host cell and then, in the crowded environment of that host, assembles into infectious agents prepared to repeat the life cycle elsewhere.

Human immunodeficiency virus (HIV) is one of the viruses that has garnered the most attention in recent years. Figure 2.30 shows cryo-electron microscopy images of mature HIV virions and gives a sense of both their overall size (roughly 130 nm) and their internal structure. In particular, note the presence of an internal capsid shaped like an ice-cream cone. This internal structure houses the roughly 10 kilobase (kb) RNA viral genome. As with our analysis of the inventory of a cell considered earlier in the chapter, part of developing a "feeling for the organism" is to get a sense of the types and numbers of the different molecules that make up that organism. In the case of HIV, these numbers are interesting for many reasons, including that they say something about the "investment" that the infected cell has to make in order to construct new virions.

For our census of an HIV virion, we need to examine the assembly of the virus. In particular, one of the key products of its roughly 10 kb genome is a polyprotein, a single polypeptide chain containing what will ultimately be distinct proteins making up the capsid, matrix, and nucleocapsid, known as Gag and shown schematically in Figure 2.31. The formation of the *immature* virus occurs through the association of

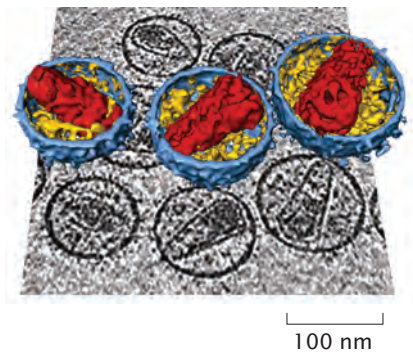


Figure 2.30: Structure of HIV viruses. The planar image shows a single frame from an electron microscopy tilt series. The three-dimensional images show reconstructions of the mature viruses featuring the ice-cream-cone-shaped capsid on the interior. (Adapted from J. A. G. Briggs et al., *Structure* 14:15, 2006.)

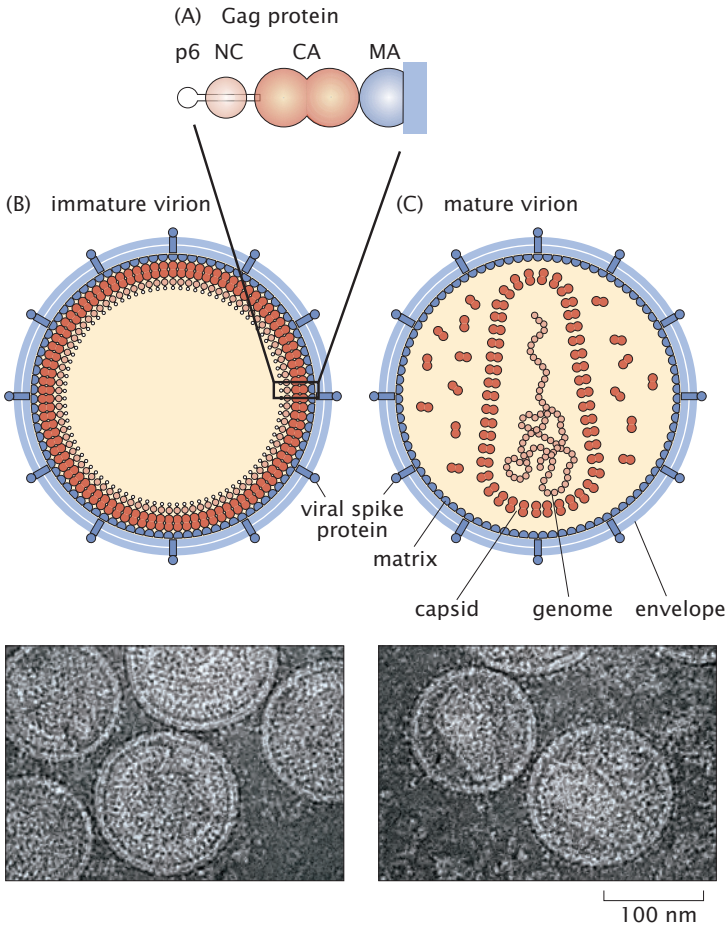


Figure 2.31: HIV architecture.
(A) Schematic of the Gag polyprotein, a 41,000 Da architectural building block.
(B) Immature virions showing the lipid bilayer coat and the uncut Gag shell on the interior.
(C) Mature virions in which the Gag protein has been cut by proteases and the separate components have assumed their architectural roles in the virus. The associated electron microscopy images show actual data for each of the cartoons. (Adapted from J. A. G. Briggs et al., *Nat. Struct. Mol. Biol.* 11:672, 2004.)

the N-terminal ends of these Gag proteins with the lipid bilayer of the host cell, with the C-termini pointing radially inward like the spokes of a three-dimensional wheel. (N-terminus and C-terminus refer to the two structurally distinct ends of the polypeptide chain. During protein synthesis, translation starts at the N-terminus and finishes at the C-terminus.) As more of these proteins associate on the cell surface, the nascent virus begins to form a bud on the cell surface, ultimately resulting in spherical structures like those shown in Figure 2.31. During the process of viral maturation, a viral protease (an enzyme that cuts proteins) clips the Gag protein into its component pieces known as matrix (MA), capsid (CA), nucleocapsid (NC), and p6. The matrix forms a shell of proteins just inside of the lipid bilayer coat. The capsid proteins form the ice-cream-cone-shaped object that houses the genetic material and the nucleocapsid protein is complexed with the viral RNA.



ESTIMATE

Estimate: Sizing Up HIV Unlike many of their more ordered viral counterparts, HIV virions have the intriguing feature that the structure from one to the next is not exactly the same. Indeed, they come in both different shapes and different sizes. As a result, our attempt to "size up" HIV will be built around some representative numbers for these viruses, but the reader is cautioned to think of a statistical distribution of sizes and shapes. As shown in the cryo-electron microscopy picture of Figure 2.30, the diameter of the virion is between 120 nm and 150 nm and we take a "canonical" diameter of 130 nm.

We begin with the immature virion. To find the number of Gag proteins within a given virion, we resort to simple geometrical reasoning. Since the radius of the overall virion is roughly 65 nm, and the outer 5 nm of that radius is associated with the lipid bilayer, we imagine a sphere of radius 60 nm that is decorated on the inside with the inward facing spokes of the Gag proteins. If we think of each such Gag protein as a cylinder of radius 2 nm, this means they take up an area $A_{\text{Gag}} \approx 4\pi\text{nm}^2$. Using this, we can find the number of such Gag proteins as

$$N_{\text{Gag}} = \frac{\text{surface area of virion}}{\text{area per Gag protein}} \approx \frac{4\pi(60\,\text{nm})^2}{4\pi\,\text{nm}^2} \approx 3500. \quad (2.20)$$

The total mass of these Gag proteins is roughly

$$M_{\text{Gag}} \approx 3500 \times 40{,}000\ \text{Da} \approx 150\ \text{MDa}, \quad (2.21)$$

where we have used the fact that the mass of each Gag polyprotein is roughly 40 kDa. This estimate for the number of Gag proteins is of precisely the same magnitude as those that have emerged from cryo-electron microscopy observations (see Briggs et al., 2004).

The number of lipids associated with the HIV envelope can be estimated similarly as

$$N_{\text{lipids}} \approx \frac{2 \times 4\pi(65\,\text{nm})^2}{1/2\,\text{nm}^2} \approx 200{,}000\ \text{lipids}, \quad (2.22)$$

where the factor of 2 accounts for the fact that the lipids form a bilayer, and we have used a typical area per lipid of $0.5\,\text{nm}^2$. The lipid census of HIV has been taken using mass spectrometry, which permits the measurement of each of the

different types of lipids forming the viral envelope (see Brügger et al., 2006). Interestingly, the diversity of lipids in the HIV envelope is enormous, with the lipid composition of the viral envelope distinct from that of the host cell membrane. The measured total number of different lipids is roughly 300,000. Further analysis of the parts list of HIV is left to the problems at the end of the chapter.

Ultimately, viruses are one of the most interesting cases of the power of macromolecular assembly. These intriguing machines occupy a fuzzy zone at the interface between the living and the nonliving.

2.2.5 The Molecular Architecture of Cells: From Protein Data Bank (PDB) Files to Ribbon Diagrams

If we continue with another factor of 10 in our powers-of-10 descent, we find the individual macromolecules of the cell. In particular, this increase in spatial resolution reveals four broad categories of macromolecules: lipids, carbohydrates, nucleic acids, and proteins. As was shown in Chapter 1 (p. 3), these four classes of molecules make up the stuff of life and have central status in building cells both architecturally and functionally. Though often these molecules are highly anisotropic (for example, a DNA molecule is usually many orders of magnitude longer than it is wide), their characteristic scale is between 1 and 10 nm. For example, as shown earlier in the chapter, a "typical" protein has a size of several nanometers. Lipids are more anisotropic, with lengths of 2–3 nm and cross-sectional areas of roughly 0.5 nm^2.

The goal of this section is to provide several different views of the molecules of life and how they fit into the structural hierarchy described throughout the chapter.

Macromolecular Structure Is Characterized Fundamentally by Atomic Coordinates

The conjunction of X-ray crystallography, nuclear magnetic resonance, and cryo-electron microscopy has revealed the atomic-level structures of a dazzling array of macromolecules of central importance to the function of cells. The list of such structures includes molecular motors, ion channels, DNA-binding proteins, viral capsid proteins, and various nucleic acid structures. The determination of new structures is literally a daily experience. Indeed, as will be asked of the reader in the problems at the end of the chapter, a visit to websites such as the Protein Data Bank or VIPER reveals just how many molecular and macromolecular structures are now known.

Though the word *structure* can mean different things to different people (indeed, that is one of the primary messages of this chapter and Chapter 8), most self-identified structural biologists would claim that the determination of structure ultimately refers to a list of atomic coordinates for the various atoms making up the structure of interest. As an example, Figure 1.3 (p. 6) introduced detailed atomic portraits of nucleic acids, proteins, lipids, and sugars. In such descriptions, the structural characterization of the system amounts to a set of coordinates

$$\mathbf{r}_i = x_i\mathbf{i} + y_i\mathbf{j} + z_i\mathbf{k}, \tag{2.23}$$

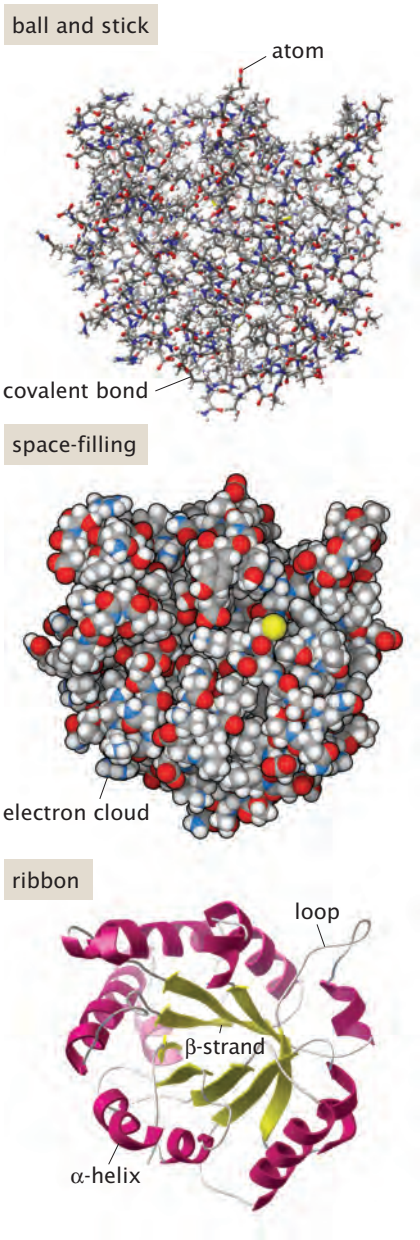


Figure 2.32: Three representations of triose phosphate isomerase. This enzyme is one of the enzymes in the glycolysis pathway (PDB 3tim). (Courtesy of D. Goodsell.)

where, having chosen some origin of coordinates, the coordinates of the ith atom in the structure are given by (x_i, y_i, z_i). That is, we have some origin of Cartesian coordinates and every atomic position is an address on this three-dimensional grid.

Because the macromolecules of the cell are subject to incessant jiggling due to collisions with each other and the surrounding water, a static picture of structure is incomplete. The structural snapshots embodied in atomic coordinates for a given structure miss the fact that each and every atom is engaged in a constant thermal dance. Hence, the coordinates of Equation 2.23 are really of the form

$$\mathbf{r}_i(t) = x_i(t)\mathbf{i} + y_i(t)\mathbf{j} + z_i(t)\mathbf{k}, \tag{2.24}$$

where the t reminds us that the coordinates depend upon time and what is measured in experiments might be best represented as $\langle \mathbf{r}_i(t) \rangle_{\text{time}}$, where the brackets $\langle\ \rangle_{\text{time}}$ signify an average over time.

An example of an atomic-level representation of one of the key proteins of the glycolysis pathway is shown in Figure 2.32. We choose this example because glycolysis will arise repeatedly throughout the book (see p. 191 for example) as a canonical metabolic pathway. The figure also shows several alternative schemes for capturing these structures such as using ribbon diagrams which highlight the ways in which the different amino acids come together to form elements of secondary structure such as α-helices and β-sheets.

Chemical Groups Allow Us to Classify Parts of the Structure of Macromolecules

When thinking about the structures of the macromolecules of the cell, one of the most important ways to give those structures functional meaning (as opposed to just a collection of coordinates) is through reference to the chemical groups that make them up. For example, the structure of the protein shown in Figures 1.3 and 2.32 is not just an arbitrary arrangement of carbons, nitrogens, oxygens, and hydrogens. Rather, this structure reflects the fact that the protein is made up of a linear sequence of amino acids that each have their own distinct identity as shown in Figure 2.33. The physical and chemical properties of these amino acids dictate the folded shape of the protein as well as how it functions.

Amino acids are but one example of a broader class of sub-nanometer-scale structural building blocks known as "chemical groups." These chemical groups occur with great frequency in different macromolecules and, like the amino acids, each has its own unique chemical identity. Figure 2.34 shows a variety of chemical groups that are of interest in biochemistry and molecular biology. These are all biologically important chemical functional groups that can be attached to a carbon atom as shown in Figure 2.34 and are all found in protein structures. The methyl and phenyl groups contain only carbon and hydrogen and are hence hydrophobic (unable to form hydrogen bonds with water). To the right of these are shown two chemically similar groups, alcohol and thiol, consisting of oxygen or sulfur plus a single hydrogen. The key feature of these two groups is that they are highly reactive and can participate in chemical reactions forming new covalent bonds. Amino acids containing these functional groups (serine, threonine, tyrosine, and cysteine) are frequently important active site residues in enzyme-catalyzed reactions. The next row starts with a nitrogen-containing amino group, which is

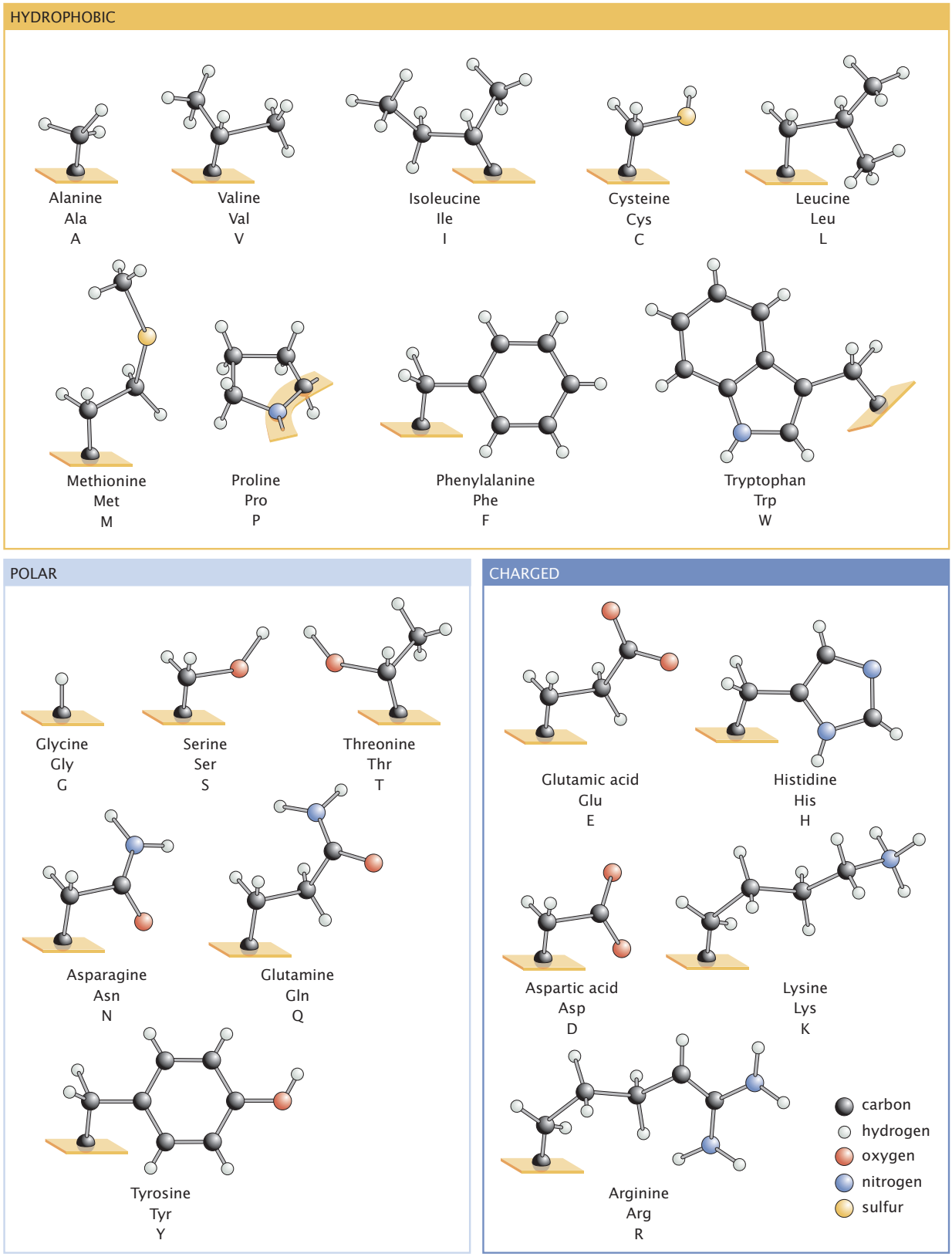


Figure 2.33: Amino acid side chains. The amino acids are represented here in ball-and-stick form, where a black ball indicates a carbon atom, a small white ball a hydrogen atom, a red ball an oxygen atom, a blue ball a nitrogen atom, and a yellow ball a sulfur atom. Only the side chains are shown. The peptide backbone of the protein to which these side chains are attached is indicated by an orange tile. The amino acids are subdivided based upon their physical properties. The group shown at the top are hydrophobic and tend to be found on the interior of proteins. Those at the bottom are able to form hydrogen bonds with water and are often found on protein surfaces.

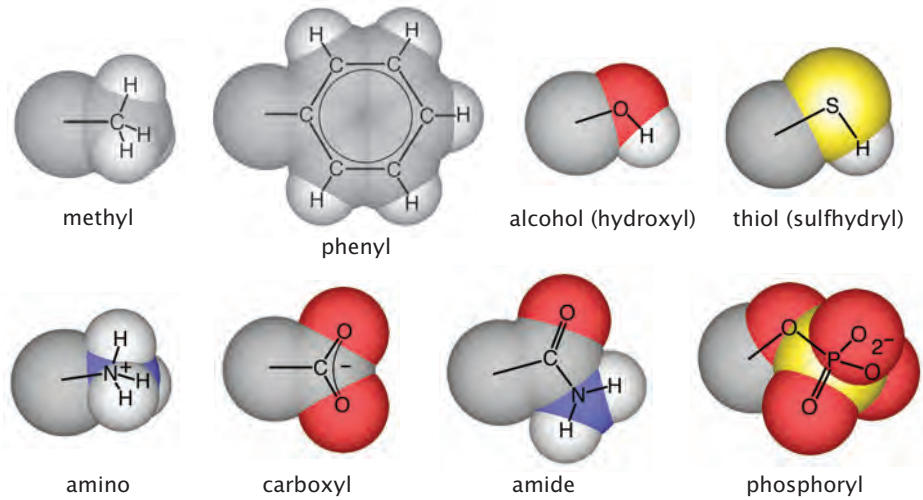


Figure 2.34: Chemical groups. These are some of the most common groups found in organic molecules such as proteins. (Courtesy of D. Goodsell.)

usually positively charged at neutral pH, and a carboxylic acid, which is usually negatively charged. All amino acids in monomeric form have both of these groups. In a protein polymer, there is a free amino group at the N-terminus of the protein and a free carboxylic acid group at the C-terminus of the protein. Several amino acids also contain these groups as part of their side chains and the charge-based interactions are frequently responsible for chemical specificity in molecular recognition as well as some kinds of catalysis. An amide group is shown next. This group is not generally charged but is able to participate in a variety of hydrogen bonds. The last group shown is a phosphoryl group, which is not part of any amino acid that is incorporated by the ribosome in a polypeptide chain during translation. However, phosphoryl groups are frequently added to proteins as a post-translational modification and perform extremely important regulatory functions.

Nucleic acids can similarly be thought of from the point of view of chemical groups. In Figure 1.3 (p. 9), we showed the way in which individual groups can be seen as the building blocks of DNA structures such as that shown in Figure 1.3(A). In particular, we note that the backbone of the double helix is built up of sugars (represented as pentagons) and phosphates. Similarly, the nitrogenous bases that mediate the pairing between the complementary strands of the backbone are represented diagrammatically via hexagons and pentagons, with hydrogen bonds depicted as shown in the figure.

This brief description of the individual molecular units of the machines of the cell brings us to the end of our powers-of-10 descent that examined the structures of the cell. Our plan now is to zoom out from the scale of individual cells to examine the structures they form together.

2.3 Telescoping Up in Scale: Cells Don't Go It Alone

Our powers-of-10 journey has thus far shown us the way in which cells are built from structural units going down from organelles to macromolecular assemblies to individual macromolecules to chemical groups, atoms, and ions. Equally interesting hierarchies of structures are revealed as we reduce the resolution of our imaginary camera and zoom out from the scale of individual cells. What we see once we

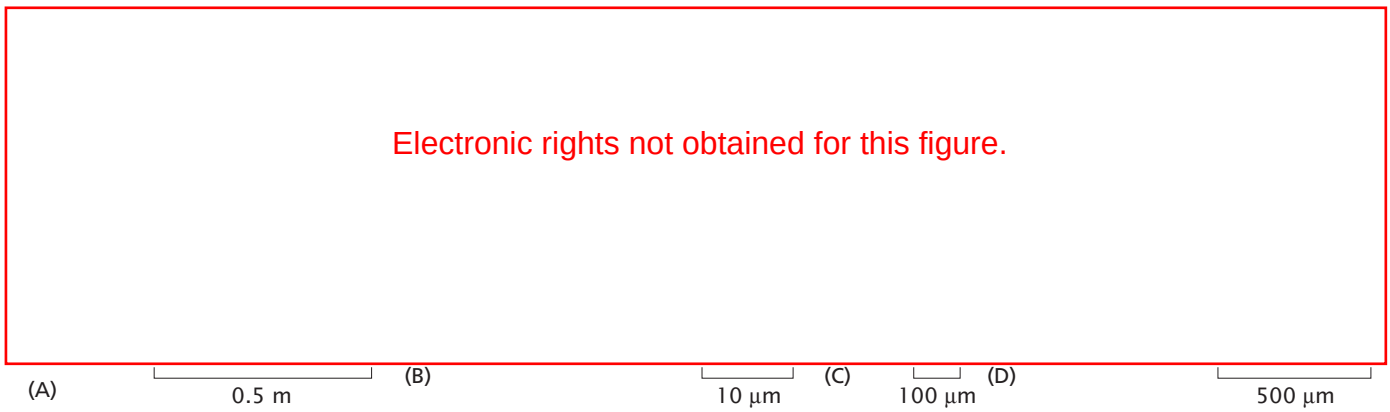


Figure 2.35: Representative examples of different communities of cells. (A) These stromatolites living in Shark Bay (Western Australia) are layered structures built by cyanobacteria. Stromatolites dominate the fossil record between 1 and 2 billion years ago, but are rare today. (B) This scanning electron micrograph showing a bacterial biofilm growing on an implanted medical device illustrates the morphological variety of bacterial species that can cooperate to form an interdependent community. (C) The social amoeba *Dictyostelium discoideum* forms fruiting bodies. The picture shows a fruiting body with spores—the tall stalk with a bulb at the top is a collection of amoebae. (D) A close-up photograph of the head of the robber fly, *Holocephala fusca*, shows the ordered structure of the compound eyes found in most insects. (A, Courtesy of Bellarmine University Department of Biology; B, Reproduced with permission of the ASM Microbe Library; C, Courtesy of Pauline Schaap; D, Courtesy of Thomas Shahan.)

begin to zoom out from the scale of single cells is the emergence of communities in which cells do not act independently.

2.3.1 Multicellularity as One of Evolution's Great Inventions

Life has been marked by several different evolutionary events that each wrought a wholesale change in the way that cells operate. One important category of such events is the acquisition of the ability of cells to communicate and cooperate with one another to form multicellular communities with common goals. This has happened many times throughout all branches of life and has culminated in extremely large organisms such as redwood trees and giraffes. In this section, we explore the ways in which cell–cell communication and cooperation have given rise to new classes of biological structures. Figure 2.35 shows a variety of different examples of cellular communities, some of which form the substance of the remainder of the chapter.

Bacteria Interact to Form Colonies such as Biofilms

The oldest known cellular communities recorded in the fossil record are essentially gigantic bacterial colonies called stromatolites. Although most stromatolites were outcompeted in their ecological niches by subsequent fancier forms of multicellular life, a few can still be found today taking essentially the same form as their 2-billion-year-old fossil brethren. Bacterial colonies of this type can still be found today in Australia, such as those shown in Figure 2.35(A). These communities have a characteristic size of about a meter and reflect collections of bacterial cells held together by an extracellular matrix secreted by these cells.

Many interesting kinds of bacterial communities consist of more than one species. Indeed, through a sophisticated system of signaling,

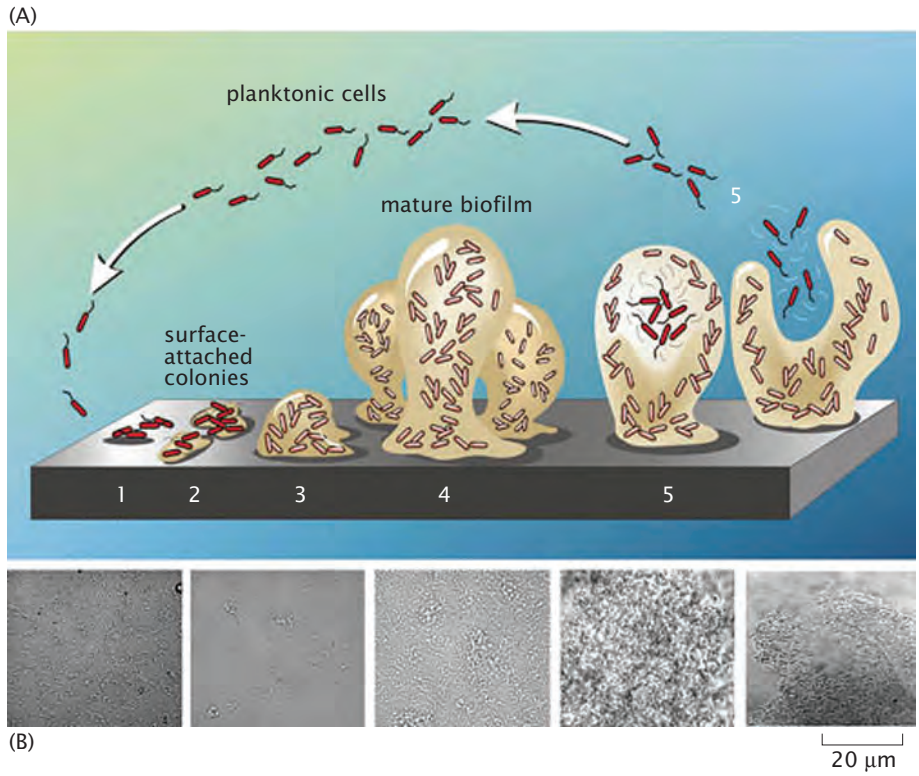


Figure 2.36: Schematic of the formation of a biofilm by bacteria. (A) The various stages in the formation of the biofilm are (1) attachment to surface, (2) secretion of extracellular polymeric substance (EPS), (3) early development, (4) maturation, and (5) shedding of cells from the biofilm. (B) The microscopy images show biofilms in various stages of the film formation process. (Adapted from P. Stoodley et al., *Annu. Rev. Microbiol.* 56:187, 2002. Original figure created by Peg Dirckx, MSU-CBE.)

detection, and organization, bacteria form colonies of all kinds ranging from biofilms to the ecosystems within animal guts. Bacterial biofilms are familiar to us all as the basis of the dentist's warning to floss our teeth every night. These communities are functionally as well as structurally interdependent. Other biofilms are noted for their destructive force when they attach to the surfaces of materials. Figure 2.35(B) shows a biofilm that grew on a silicon rubber voice prosthesis that had been implanted in a patient for about three months.

Structurally, a biofilm is formed as shown schematically in Figure 2.36. The key building blocks of such structures are a population of bacteria, a surface onto which these cells may adhere, and an aqueous environment. The formation of a biofilm results in a population of bacterial cells that are attached to a surface and enclosed in a polymeric matrix built up of molecules produced by these very same bacteria. The early stages of biofilm formation involve the adhesion of the bacteria to a surface, followed by changes in the characteristics of these bacteria such as the loss of flagella and micro-colony formation mediated by pili-twitching motility. At a larger scale, these changes at the cellular level are attended by the formation of colonies of cells and differentiation of the colonies into structures that are embedded in extracellular polysaccharides. Though there are a variety of different morphologies that are adopted by such films, roughly speaking, these biofilms are relatively porous structures (presumably to provide a conduit for import and export of nutrients and waste, respectively) that typically take on mushroom-like structures such as indicated schematically in Figure 2.36. These films have a relative proportion of something like 85% of the mass taken up by extracellular matrix, while the remaining 15% is taken up by cells themselves. A typical thickness for such films ranges from 10 to 50 μm.

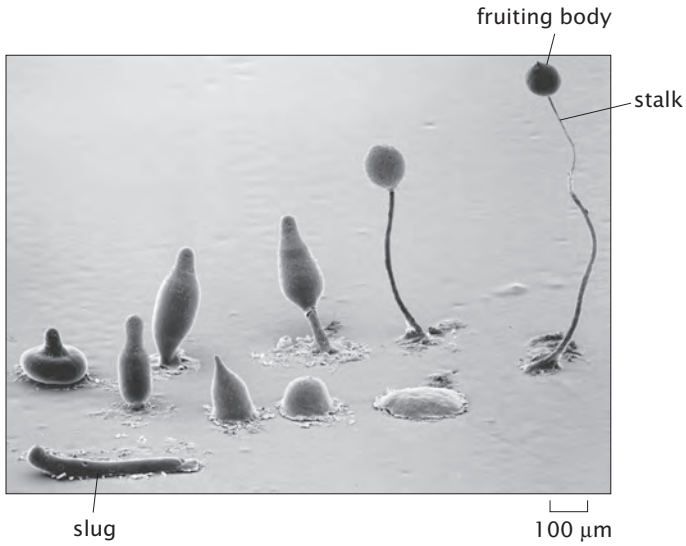


Figure 2.37: Formation of a multicellular structure during starvation. The social amoeba *Dictyostelium discoideum* responds to starvation by forming a structure made up of tens of thousands of cells in which individual cells suffer different fates. Cells near the top of the structure form spores, which are resurrected once conditions are favorable. The figure shows the developmental stages that take place on the way to making fruiting bodies, starting in the bottom right and proceeding clockwise. (Copyright, M. J. Grimson and R. L. Blanton, Biological Sciences Electron Microscopy Laboratory, Texas Tech University.)

Teaming Up in a Crisis: Lifestyle of *Dictyostelium discoideum*

Although bacteria can form communities, eukaryotes have clearly raised this to a high art. One particularly fascinating example that may give clues as to the origin of eukaryotic multicellularity is the cellular slime mold *Dictyostelium discoideum*, as shown in Figure 2.35(C). This small, soil-dwelling amoeba pursues a solitary life when times are good but seeks the comfort of its fellows during times of starvation. *Dictyostelium* is usually content to wander around as an individual with a characteristic size between 5 and 10 μm. However, when deprived of their bacterial diet, these cells undertake a radical change in lifestyle, which involves both interaction and differentiation through a series of fascinating intermediate steps as shown in Figure 2.37. A group of *Dictyostelium* amoebae in a soil sample that find themselves faced with starvation send chemical signals to one another resulting in the coalescence of thousands of separate amoebae to form a slug that looks like a small nematode worm. These cells appear to be poised on the brink between unicellular and multicellular lifestyles and can readily convert between them. Ultimately, as shown in the figure, the slug stops moving and begins to form a stalk. At the tip of the stalk is a nearly spherical bulb that contains many thousands of spores, essentially cells in a state of suspended animation. When environmental conditions are appropriate for individual amoeba to thrive, the spores undergo the process of germination, with each spore becoming a functional, individual amoeba.

Estimate: Sizing Up the Slug and the Fruiting Body The relation between the number of cells in a slug and its size is shown in Figure 2.38, where we see that these slugs can range in length between several hundred and several thousand microns with the number of cells making up the slug being between tens of thousands and several million. The next visible stage in the development is the sprouting of a stalk with a bulb at the top known as a fruiting body. The stalk is of the order of a millimeter in length, while the fruiting body itself is several hundred microns across. This fruiting body is composed of thousands of spores, that are effectively

ESTIMATE

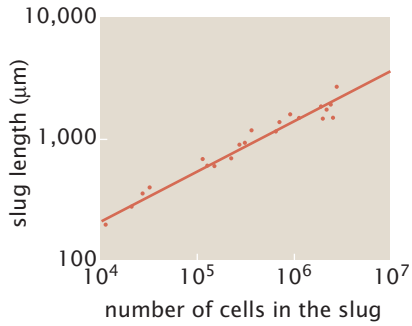


Figure 2.38: Slug size in *Dictyostelium discoideum*. The plot shows a relation between the size of the *Dictyostelium* slug and the corresponding number of cells. Points represent different individuals and the line shows the overall trend. (Adapted from Bonner, www.dictybase.org.)

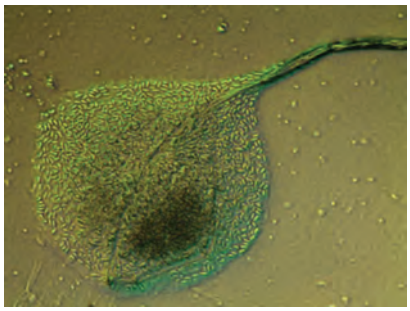


Figure 2.39: Microscopy image of a fruiting body. The fruiting body has been squished on a microscope slide, revealing both the size of the spores and their numbers.

in a state of suspended animation. An example of such a fruiting body that has been squished on a microscope slide is shown in Figure 2.39. This structure has functional consequences. In particular, those cells that become spores remain poised to respond to a better day, while the cells that formed the stalk have effectively ended their lives for the good of those that survive.

An immediate question of interest concerning the multicellular fruiting bodies shown in Figure 2.37 is how many cells conspire to make up such structures. An estimate of the number of cells in a fruiting body can be constructed by examining the nearly hemispherical collection of cells shown in Figure 2.39. Note that we treat the fruiting body as a hemisphere since, once it is squashed on the microscope slide, its shape is flattened and resembles a hemisphere more closely than a sphere. The diameter of this hemisphere is roughly $200\,\mu\text{m}$. Our rough estimate for the number of spores in the fruiting body is obtained by evaluating the ratio

$$\text{number of cells} = \frac{V_{\text{body}}}{V_{\text{cell}}}, \tag{2.25}$$

where we assume that the entirety of the fruiting body volume is made up of cells. If we assume that the cell is $4\,\mu\text{m}$ in diameter, this yields

$$\text{number of cells} = \frac{\frac{2}{3}\pi(100\,\mu\text{m})^3}{\frac{4}{3}\pi(2\,\mu\text{m})^3} \approx 6 \times 10^4 \text{ cells}. \tag{2.26}$$

Note that the size of the ball of cells in a fruiting body can vary dramatically from the $200\,\mu\text{m}$ scale shown here to several times larger, and, as a result, the estimate for the number of cells in a fruiting body can vary. Also, note that a factor-of-2 error in our estimate of the size of an individual cell will translate into a factor-of-8 error in our count of the number of cells in the slug or fully formed fruiting body.

Multicellular Organisms Have Many Distinct Communities of Cells

The three branches of life that have most notably exploited the potential of the multicellular lifestyle are animals, plants, and fungi. While bacterial and protist colonial organisms rarely form communities with characteristic dimensions of more than a few millimeters (with the exception of stromatolites), individuals in these three groups routinely grow to more than a meter in size. Their enormous size and corresponding complexity can be attributed to at least three factors: (i) production of extracellular matrix material that can provide structural support for large communities of cells, (ii) a predilection towards cellular specialization such that many copies of cells with the same genomic content can develop to perform distinct functions, and (iii) highly sophisticated mechanisms for the cells to communicate with one another within the organism. We emphasize that these traits are not unique to animals, plants, and fungi, but they are used more extensively there than elsewhere. A beautiful illustration of these principles is seen in the eye of a fly as shown in Figure 2.35(D). The eye is made up of hundreds of small units called ommatidia, each of which contains a group of eight photoreceptor cells, support cells,

and a cornea. During development of the eye, these cells signal to one another to establish their identities and relative positions to create a stereotyped structure that is repeated many times. The overarching theme of the remainder of this chapter is the exploration of how cells come together to form higher-order structures and how these structures fit into the overall hierarchy of structures formed by living organisms.

2.3.2 Cellular Structures from Tissues to Nerve Networks

Multicellular structures are as diverse as cells themselves. Often, the nature of these structures is a reflection of their underlying function. For example, the role of epithelia as barriers dictates their tightly packed, planar geometries. By way of contrast, the informational role of the network of neurons dictates an entirely different type of multicellular structure.

One Class of Multicellular Structures is the Epithelial Sheets

Epithelial sheets form part of the structural backdrop in organs ranging from the skin to the bladder. Functionally, the cells in these structures have roles such as serving as a barrier to transport of molecules, providing an interface at which molecules can be absorbed into cells, and serving as the seat of certain secretions. Several different views of these structures are shown in Figure 2.40.

The morphology of epithelial sheets is diverse in several ways. First, the morphology of the individual cells can be different (isotropic versus anisotropic). In addition, the assemblies of cells themselves have different shapes. For example, the different structures can be broadly classified into those that are a monolayer sheet (simple epithelium) and those that are a multilayer (stratified epithelium). Within these two broad classes of structures, the cells themselves have different morphologies. The cells making up a given epithelial sheet can be flat, pancake-like cells, denoted as *squamous* epithelium. If the cells making up the epithelial sheet have no preferred orientation, they are referred to as *cuboidal*, while those that are elongated perpendicular to the extracellular support matrix are known as *columnar* epithelia. Epithelial sheets have as one of their functions (as do lipid bilayers) the segregation of different media, which can have highly different ionic concentrations, pH, macromolecular concentrations, and so on.

Tissues Are Collections of Cells and Extracellular Matrix

We have seen that cells can interact to form complexes. An even more intriguing example of a multicellular structure is provided by tissues such as that shown in Figure 2.41. These connective tissues are built up from a diverse array of cells and materials they secrete. Beneath the epithelial surface, fibroblasts construct extracellular matrix. This connective tissue is built up of three main components: cells such as fibroblasts and macrophages, connective fibers, and a supporting substance made up of glycosaminoglycans (GAGs). What is especially appealing and intriguing about these tissues is the orchestration, in both space and time, leading to the positioning of cells and fibers. The fibroblasts, which were featured earlier in this chapter, serve as factories for the proteins that make up the extracellular matrix. In

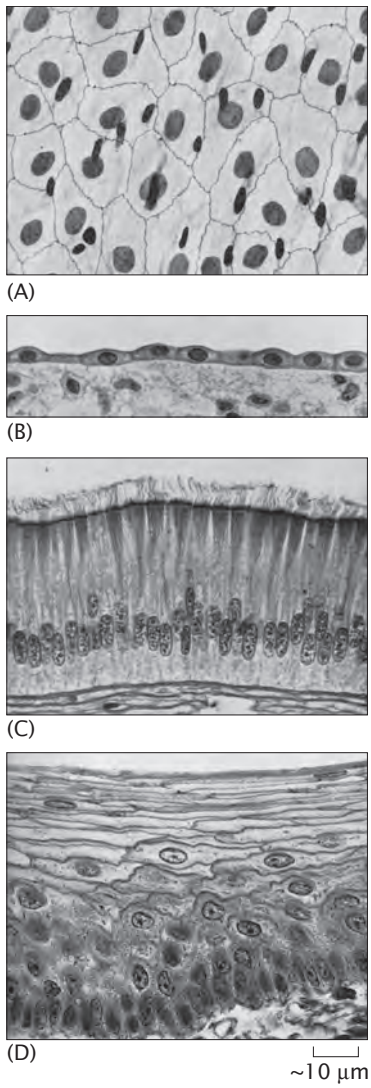


Figure 2.40: Shapes and architecture of epithelial cells. Epithelia are tissues formed by continuous sheets of cells that make tight contacts with one another. (A) Viewed from above, a simple epithelial sheet resembles a tiled mosaic. The dark ovals are the cell nuclei stained with silver. (B) Viewed from the side, simple epithelia such as this from the dog kidney may form a single layer of flattened cells above a loose, fibrous connective tissue. (C) In some specialized epithelia, the cells may extend upwards forming elongated columns and develop functional specializations at the top surface such as the beating cilia shown here. This particular ciliated columnar epithelium is from the alimentary tract of a freshwater mussel. (D) In other epithelia such as skin or the kitten's gum shown here, epithelial cells may form multiple layers. (A–D, adapted from D. W. Fawcett, The Cell, Its Organelles and Inclusions: An Atlas of Fine Structure. W. B. Saunders, 1966.)



Figure 2.41: Tissue organization. This thin section of a villus in the human small intestine was stained with hematoxylin and eosin, a mixture of dyes that stains DNA and other negatively charged macromolecules blue or dark purple, and stains most proteins pink. In this image, a finger-shaped intestinal villus sticks up into the lumen of the intestine, covered with a single continuous monolayer of epithelial cells, whose function is to absorb nutrients. The supporting structure underneath the epithelial layer is composed of extracellular matrix secreted by fibroblasts. Numerous immune system cells also patrol this territory. (Adapted from L. G. van der Flier and H. Clevers, *Annu. Rev. Physiol.* 71:241, 2009.)

particular, they synthesize proteins such as collagen and elastin that, when secreted, assemble to form fibrous structures that can support mechanical loads. The medium within which these fibers (and the cells) are embedded is made up of the third key component of the connective tissue, namely, the hydrated gel of GAGs.

Nerve Cells Form Complex, Multicellular Complexes

A totally different example of structural organization involving multiple cells is illustrated by collections of neurons. Neurons are the specialized cells in animals that we associate with thinking and feeling. These cells allow for the transmission of information over long distances in the form of electrical signals. Neurons are constructed with many input terminals known as dendrites and a single output path known as the axon. The fascinating structural feature of these cells is that they assemble into complex networks that are densely connected in patterns where dendrites from a given cell reach out to many others, from which they take various inputs. An example of a collection of fluorescently labeled neurons is shown in Figure 2.42. Note that the branches (dendrites and axons) that reach out from the various cells have lengths far in excess of the 10 μm scale characteristic of typical eukaryotes. Indeed, axons of some neurons can have lengths of centimeters and more.

One fascinating example of neuronal contact is offered by the so-called neuromuscular junction as shown in Figure 2.43. These junctions are the point of contact between motor neurons, which convey the marching orders for a given muscle, and the muscle fiber itself. As is seen in the figure, the axon from a given motor neuron makes contacts with various muscle fibers. As will be described in more detail in Chapter 17, when an electrical signal (action potential) arrives at the contact point known as a synapse, chemical neurotransmitters are released into the space between the nerve and muscle. These neurotransmitters result in the opening of ligand-gated ion channels in the muscle, which results in a change in the electrical state of the muscle and leads to its contraction. These contrasting examples provide a small window into the diversity of cell–cell contacts.

2.3.3 Multicellular Organisms

The highest level in the structural hierarchy to be entertained here is that of individual multicellular organisms. The diversity of multicellular organisms is legendary, ranging from roses to hummingbirds, from Venus flytraps to the giant squid. What is especially remarkable about this diversity is contained in the simple statement of the cell theory: each and every one of these organisms is a collection of cells and their products. And, each and every one of these organisms is the result of a long history of evolution resulting in specialization and diversification of the various cells that make up that organism.

Cells Differentiate During Development Leading to Entire Organisms

The fruit fly *Drosophila melanogaster* has had a long and rich history as one of the key "model" organisms of biology and is a useful starting point for thinking about the size of organisms. As shown in