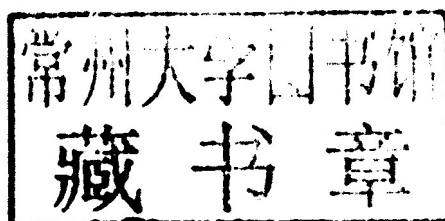


# HOW PROTEINS WORK

**Mike Williamson**

# HOW PROTEINS WORK

**Mike Williamson**



*Garland Science*

Vice President: Denise Schanck

Editor: Summers Scholl

Editorial Assistant: Kelly O'Connor

Production Editor and Layout: EJ Publishing Services

Illustrator and Cover Design: Matthew McClements, Blink Studio, Ltd.

Copyeditor: Bruce Goatly

Proofreader: Sally Livitt

Indexer: Medical Indexing Limited

© 2012 by Garland Science, Taylor & Francis Group, LLC

This book contains information obtained from authentic and highly regarded sources. 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 978-0-8153-4446-9

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

Williamson, Mike (Michael Paul)

How proteins work / Mike Williamson.

p. cm.

Includes bibliographical references.

ISBN 978-0-8153-4446-9

1. Proteins. 2. Proteins--Analysis. 3. Biochemistry. 4. Biophysics. I.

Title.

QD431.W536 2012

572'.636--dc23

201101758

Published by Garland Science, Taylor & Francis Group, LLC,  
an informa business, 711 Third Avenue, New York, NY 10017,  
USA, and 2 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

 **Garland Science**  
Taylor & Francis Group

Visit our website at <http://www.garlandscience.com>

# **HOW PROTEINS WORK**

# Preface

Proteins are endlessly fascinating. They carry out almost all the catalytic functions in the cell, as well as directing and forming most of the structural framework. They catalyze reactions many orders of magnitude faster than any system that humans can devise under comparable conditions. Proteins are also much larger than most human-designed catalysts. They make many interactions of widely varying strength and duration. Crucial to any understanding of protein function is their structure: although many of the principles governing how proteins work were understood many years ago, it is not until we have the structural details that we can really appreciate exactly what the proteins are doing. This is a major reason why we don't understand membrane proteins as well as we do globular proteins. However, the structure is merely the detail that enables us to reach toward the concepts that really explain how proteins work. I have therefore endeavored to look beyond the structural detail to understand the underlying principles.

This textbook grew out of my courses for intermediate and advanced undergraduates, and is inspired by the idea that proteins are a functional part of living and evolving systems. They have a certain form and function because it works, not necessarily as the most perfect solution to the biological problem, but certainly as a viable and successful solution. Advanced undergraduate and graduate students, as well as practitioners, interested in proteins should find the book useful. A basic foundation in chemistry and biology, as supplied by introductory undergraduate courses in biochemistry, should suffice. Students may have a background in chemistry, biology or physics, but I have tried to write the text so that it is accessible to all.

The book is written in a style that I like to read. This means the text is discursive; occasionally it goes off at a tangent; it has analogies and examples liberally scattered around; it simplifies systems as far as possible in an attempt to see the forest for the trees; and it places more emphasis on principles than on the experiments that were used to derive the principles.

There is considerable discussion of the role of evolution in tinkering with proteins to create something with a desired function. I am particularly interested in how proteins solve 'difficult' biological problems, such as catalysis, movement, and signaling. In the same way that you cannot really understand a foreign country without having some idea of its history, I believe that you cannot understand proteins unless you have some idea of how they got to their present form. The use of everyday analogies and emphasis of the physical environment around the protein enable the reader to understand proteins as well as merely know the facts about them.

Quantitative calculations are used to understand how proteins work. I strongly believe that the field of biochemistry in general, and protein science in particular, will need to place more emphasis on quantitative measurements as they mature. A holistic view—integrating structural, chemical, and biological data to try to understand how proteins help the cell to function properly—is key to this text. We are moving into a new era of biological science, where we have a good idea of many of the pieces, and we are starting to see how the pieces work together to achieve a functional whole (the idea behind Systems Biology). This book is an attempt to do exactly that.

*I have been occupied for some time with the study of the most essential substances of the animal kingdom: fibrin, albumin and gelatin. I conclude that the organic substance which is present in all constituents of the animal body, also as we shall soon see in the plant kingdom, could be named protein from πρωτεος [proteios], primarius, which has the composition  $C_{400}H_{620}N_{100}O_{120}\dots$*

**Gerhardus Johannes Mulder  
(1802–1880)**

I do not attempt to be comprehensive in the coverage of proteins. There is little coverage of medical aspects of proteins, though they are certainly described where relevant, as in signaling. I have often skipped over the experimental evidence for many of the facts presented, because I do not want to obscure the principles of how proteins work by inclusion of too much experimental detail.

Chapters 1-4 and 6 present the physical constraints that have resulted in proteins looking and working the way they do. These limitations include the structures and properties of amino acids and the forces that hold proteins together, which are discussed in Chapter 1, along with a detailed discussion of the way evolution shapes proteins. Chapter 2 discusses the domain, the fundamental structural and evolutionary building block of proteins, while Chapter 3 considers how domains associate together into oligomeric proteins; it also discusses consequences of oligomerization such as allostery and cooperativity. Chapter 4 covers an important topic that is not often discussed in textbooks, namely the cellular environment and how this influences proteins. It describes the crowded environment of the cell, how proteins bind rapidly and yet specifically to their targets, and natively unstructured proteins, as well as post-translational modifications and protein folding. Finally, Chapter 6 discusses the developing area of internal mobility within proteins.

The second half of the book, Chapters 5 and 7-10, covers various biological functions of proteins, and considers how they carry out these functions, and how their structure enables them to do so. These are enzyme catalysis in Chapter 5, movement and translocation in Chapter 7, signaling in Chapter 8, regulation (by the formation of complexes) in Chapter 9, and coordination of sequential reactions by multi-enzyme complexes in Chapter 10. Additionally, Chapter 9 looks at the results emerging from high-throughput technology. Finally, Chapter 11 discusses the techniques used in studying proteins, both experimental and theoretical.

The main text is augmented with boxes referred to by numbered asterisks (\*) that provide more details on select topics, brief biographies of prominent scientists, and pedagogical analogies for further elucidation of concepts. There is also a glossary containing definitions to words that appear in bold throughout the main text.

## Online Resources

Accessible from [www.garlandscience.com](http://www.garlandscience.com), the Student and Instructor Resources websites provide learning and teaching tools created for *How Proteins Work*. The Student Resources Site is open to everyone, and users have the option to register in order to use book-marking and note-taking tools. The Instructor's Resources Site requires registration and access is available to instructors who have assigned the book to their course. To access the Instructor Resource Site please contact your local sales representative or email [science@garland.com](mailto:science@garland.com). Below is an overview of the resources available for this book. On the website, the resources may be browsed by individual chapters and there is a search engine. You can also access the resources available for other Garland Science titles.

## For Students

### **Animations and Videos**

The animations and videos dynamically illustrate important concepts from the book, and make many of the more difficult topics accessible.

### **Flashcards**

Each chapter contains a set of flashcards, built into the website, that allow students to review key terms from the text.

## **Glossary**

The complete glossary from the book is available on the website and can be searched and browsed as a whole or sorted by chapter.

## **Hints**

The hints provide strategies and clues for solving some of the more difficult end-of-chapter problems.

## **Solutions to Problems**

Solutions to the odd-numbered problems are provided for self-testing.

## **For Instructors**

### **Figures**

The images from the book are available in two convenient formats: PowerPoint® and JPEG. They have been optimized for display on a computer. Figures are searchable by figure number, figure name, or by keywords used in the figure legend from the book.

### **Animations and Videos**

The animations and videos that are available to students are also available on the Instructor's website in two formats. The WMV formatted movies are created for instructors who wish to use the movies in PowerPoint presentations on Windows® computers; the QuickTime formatted movies are for use in PowerPoint for Apple computers or Keynote® presentations. The movies can easily be downloaded to your computer using the "download" button on the movie preview page.

### **Power Point Presentations**

The PowerPoint presentations contain the figures and micrographs from the book. There is one presentation for each chapter.

### **Solutions Manual**

A complete solutions manual is provided for all problems in the text.

## **Acknowledgments**

I need to thank the many people who have helped in one way or another. These include my supervisors and mentors Dudley Williams and Kurt Wüthrich, as well as colleagues who have provided much needed insight. For advice and corrections: Pete Artymiuk and Per Bullough. For suggestions particularly on the problems: Abaigael Keegan, Hugh Dannatt, Rebecca Hill, Vicki Kent, Tacita Nye and Muhammed Qureshi. And of course the production team at Garland Science, particularly Summers Scholl who has nurtured the book through its gestation; Emma Jeffcock, who has ably managed the production process; Matt McClements, who has turned my sketches into awesome images; and Bruce Goatly, whose spot-on comments and corrections were appreciated.

Mike Williamson

# Contents

Chapter 1	Protein Structure and Evolution	1
Chapter 2	Protein Domains	61
Chapter 3	Oligomers	99
Chapter 4	Protein Interactions <i>in vivo</i>	129
Chapter 5	How Enzymes Work	179
Chapter 6	Protein Flexibility and Dynamics	215
Chapter 7	How Proteins Make Things Move	243
Chapter 8	How Proteins Transmit Signals	285
Chapter 9	Protein Complexes: Molecular Machines	325
Chapter 10	Multienzyme Complexes	349
Chapter 11	Techniques for Studying Proteins	379
Glossary		435
Index		445

# Detailed Contents

## Chapter 1 Protein Structure and Evolution

<b>1.1</b>	<b>STRUCTURES OF AMINO ACIDS AND PEPTIDES</b>	<b>1</b>
1.1.1	Proteins are composed of amino acids	1
1.1.2	Amino acids have only a few allowed conformations	4
1.1.3	The most populated conformation is in the $\beta$ -sheet region	9
1.1.4	The other main conformations are the $\alpha$ helix and the "random coil"	10
1.1.5	The $pK_a$ value describes the protonation behavior of side chains	12
<b>1.2</b>	<b>THE FORCES HOLDING PROTEINS TOGETHER</b>	<b>13</b>
1.2.1	Electrostatic forces can be strong	13
1.2.2	Hydrogen bonds are formed by electrostatic dipoles	14
1.2.3	Van der Waals forces are individually weak but collectively strong	14
1.2.4	The hydrophobic interaction is entropic in origin	15
1.2.5	Hydrogen bonds are uniquely directional	17
1.2.6	Cooperativity is a feature of large systems	17
1.2.7	The formation of a $\beta$ hairpin is cooperative	18
1.2.8	Hydrogen bond networks are cooperative	19
1.2.9	Proteins require a layer of water for their function	20
1.2.10	Entropy and enthalpy tend to change in compensatory directions	21
<b>1.3</b>	<b>THE STRUCTURE OF PROTEINS</b>	<b>23</b>
1.3.1	Proteins are composed of primary, secondary, tertiary, and quaternary structure	23
1.3.2	Secondary structures pack together in structure motifs	24
1.3.3	Membrane proteins are different from globular proteins	29
1.3.4	The structure of a protein is (more or less) determined by its sequence	31
1.3.5	Some proteins form metastable structures	32
1.3.6	Structure is conserved more than sequence	33
1.3.7	Structural homology can be used to identify function	33

<b>1.4</b>	<b>THE EVOLUTION OF PROTEINS</b>	<b>36</b>
1.4.1	What are the purposes of proteins?	36
1.4.2	Evolution is a tinker	37
1.4.3	Many proteins arose by gene duplication	38
1.4.4	Most new proteins arise by modification of duplicated genes	40
1.4.5	Evolutionary tinkering leaves its fingerprints behind	42
1.4.6	New proteins can arise by gene sharing	42
1.4.7	Evolution usually retains chemistry and alters binding	44
1.4.8	Convergent and divergent evolution are difficult to distinguish	44
1.4.9	New functions may often develop from promiscuous or moonlighting precursors	45
1.4.10	Retrograde evolution is not common	48
1.4.11	Proteins began in an RNA world	48
1.4.12	Most evolutionary innovation happened very early	50
<b>1.5</b>	<b>SUMMARY</b>	<b>52</b>
<b>1.6</b>	<b>FURTHER READING</b>	<b>53</b>
<b>1.7</b>	<b>WEBSITES</b>	<b>54</b>
<b>1.8</b>	<b>PROBLEMS</b>	<b>55</b>
<b>1.9</b>	<b>NUMERICAL PROBLEMS</b>	<b>56</b>
<b>1.10</b>	<b>REFERENCES</b>	<b>56</b>

## Chapter 2 Protein Domains 61

<b>2.1</b>	<b>DOMAINS: THE FUNDAMENTAL UNIT OF PROTEIN STRUCTURE</b>	<b>61</b>
2.1.1	Domains can be defined in a variety of ways	61
2.1.2	Domains can usually be associated with specific functions	63
2.1.3	Domains are the basic building blocks of proteins	66
2.1.4	Modules are transposable domains	67
<b>2.2</b>	<b>THE KEY ROLE OF DOMAINS IN PROTEIN EVOLUTION</b>	<b>69</b>
2.2.1	Multidomain proteins are produced by exon shuffling	69
2.2.2	Multidomain proteins are also produced by other genetic mechanisms	71

2.2.3	Evolution can proceed in jumps by three-dimensional domain swapping	71	3.1.3	Oligomerization makes symmetric dimers	103
2.2.4	Three-dimensional domain swapping still occurs	72	3.1.4	Coding errors, coding efficiency and linkers are not convincing reasons	103
2.2.5	Increased binding specificity is conferred by additional domain interactions	73	<b>3.2</b>	<b>ALLOSTERY</b>	<b>105</b>
2.2.6	Intramolecular binding is strong because the effective concentration is high	76	3.2.1	Most enzymes are not allosteric	105
2.2.7	Intramolecular interactions lead to cooperative hydrogen bonding	77	3.2.2	Hemoglobin is the classic example of allostery	106
2.2.8	Intramolecular domain:peptide binding facilitates autoinhibition	77	3.2.3	Oxygen affinity in hemoglobin is fine-tuned by other effectors	107
2.2.9	Intramolecular domain:peptide binding facilitates evolutionary change	79	3.2.4	There are two main models for allostery	108
2.2.10	Binding specificity is increased by scaffold proteins	80	3.2.5	Glycogen phosphorylase is another good example of allostery	110
2.2.11	Intramolecular binding is strong because it has less unfavorable entropy	81	<b>3.3</b>	<b>COOPERATIVE BINDING OF DIMERS TO DNA</b>	<b>112</b>
<b>2.3</b>	<b>POTENTIAL ADVANTAGES OF MULTIDOMAIN CONSTRUCTION</b>	<b>83</b>	3.3.1	Cooperativity can be understood by using thermodynamics	112
2.3.1	Multidomain construction makes it simple to evolve a new function	83	3.3.2	Sequence-specific binding to DNA is a problem	113
2.3.2	Multidomain construction makes it simple to introduce control and regulation	84	3.3.3	The <i>trp</i> repressor recognizes DNA by hinge bending	115
2.3.3	Multidomain construction makes an effective enzyme	87	3.3.4	CAP recognizes DNA by rotation around the dimer interface	116
2.3.4	Multidomain construction simplifies folding and assembly and stabilizes the protein	88	3.3.5	DNA recognition by a symmetric leucine zipper	117
<b>2.4</b>	<b>PROTEINS AS TOOLS</b>	<b>89</b>	3.3.6	DNA recognition by a heterodimeric leucine zipper	119
2.4.1	Tools have independently moving parts	91	3.3.7	Max and Myc form a heterodimeric zipper with alternative partners	120
2.4.2	Tools have a common design but different sizes	91	3.3.8	DNA recognition by a tandem dimer	121
2.4.3	Tools have common parts with variable "ends"	92	<b>3.4</b>	<b>ISOZYMES</b>	<b>122</b>
2.4.4	Some tools are symmetric	92	<b>3.5</b>	<b>SUMMARY</b>	<b>124</b>
2.4.5	Some tools have a specialist use	92	<b>3.6</b>	<b>FURTHER READING</b>	<b>124</b>
<b>2.5</b>	<b>SUMMARY</b>	<b>93</b>	<b>3.7</b>	<b>WEBSITES</b>	<b>124</b>
<b>2.6</b>	<b>FURTHER READING</b>	<b>93</b>	<b>3.8</b>	<b>PROBLEMS</b>	<b>125</b>
<b>2.7</b>	<b>WEBSITES</b>	<b>94</b>	<b>3.9</b>	<b>NUMERICAL PROBLEMS</b>	<b>125</b>
<b>2.8</b>	<b>PROBLEMS</b>	<b>94</b>	<b>3.10</b>	<b>REFERENCES</b>	<b>126</b>
<b>2.9</b>	<b>NUMERICAL PROBLEMS</b>	<b>95</b>			
<b>2.10</b>	<b>REFERENCES</b>	<b>95</b>			
<b>Chapter 3</b>	<b>Oligomers</b>	<b>99</b>	<b>Chapter 4</b>	<b>Protein Interactions <i>in vivo</i></b>	<b>129</b>
<b>3.1</b>	<b>WHY DO PROTEINS OLIGOMERIZE?</b>	<b>99</b>	<b>4.1</b>	<b>FACTORS INFLUENCING COLLISION RATES</b>	<b>129</b>
3.1.1	Oligomerization shelters and regulates the active site	99	4.1.1	On a small scale, random processes have much more significant effects	129
3.1.2	Oligomerization provides improved enzyme functionality	102	4.1.2	Diffusion occurs by a random walk	130
			4.1.3	The collision rate is limited by geometrical factors	130
			4.1.4	Collision rates can be increased by electrostatic attraction	132
			4.1.5	Collision rates are also increased by electrostatic steering	133
			4.1.6	Protein binding takes place via an encounter complex	135

4.1.7	Electrostatic repulsion is also important for limiting interactions	137
4.1.8	Macromolecular crowding increases the amount of protein association but slows its rate	138
4.1.9	Larger proteins diffuse more slowly	142
<b>4.2</b>	<b>HOW PROTEINS CAN FIND THEIR PARTNERS FASTER</b>	<b>144</b>
4.2.1	Processivity decreases the off-rate from polymeric substrates	144
4.2.2	Searching is much faster in two dimensions	145
4.2.3	Searching is slightly faster again in one dimension	148
4.2.4	Searching is faster in smaller compartments	149
4.2.5	Sticky arms are useful for short-range searching	149
4.2.6	Proline-rich sequences make good sticky arms	150
4.2.7	Sticky-arm interactions have fast on- and off-rates	151
4.2.8	Sticky arms are fast because they zip up rather than lock on	154
<b>4.3</b>	<b>NATIVELY UNSTRUCTURED PROTEINS</b>	<b>155</b>
4.3.1	Natively unstructured proteins are common	155
4.3.2	Natively unstructured proteins permit specific binding with fast on-rate	156
4.3.3	Natively unstructured proteins provide specific binding without strong binding	157
4.3.4	Natively unstructured proteins may provide other benefits	157
<b>4.4</b>	<b>POST-TRANSLATIONAL MODIFICATION OF PROTEINS</b>	<b>158</b>
4.4.1	Covalent modifications modify protein function	158
4.4.2	Phosphorylation	158
4.4.3	Methylation and acetylation	164
4.4.4	Glycosylation	166
<b>4.5</b>	<b>PROTEIN FOLDING AND MISFOLDING</b>	<b>167</b>
4.5.1	Protein folding is often rapid and thermodynamically controlled	167
4.5.2	All proteins have a limited lifespan, especially unfolded ones	169
4.5.3	Amyloid is a consequence of protein misfolding	171
<b>4.6</b>	<b>SUMMARY</b>	<b>173</b>
<b>4.7</b>	<b>FURTHER READING</b>	<b>174</b>
<b>4.8</b>	<b>WEBSITES</b>	<b>174</b>
<b>4.9</b>	<b>PROBLEMS</b>	<b>174</b>
<b>4.10</b>	<b>NUMERICAL PROBLEMS</b>	<b>175</b>
<b>4.11</b>	<b>REFERENCES</b>	<b>176</b>

## Chapter 5 How Enzymes Work 179

<b>5.1</b>	<b>ENZYMES LOWER THE ENERGY OF THE TRANSITION STATE</b>	<b>179</b>
5.1.1	What is the transition state?	179
5.1.2	Enzymes lower both enthalpy and entropy barriers in the transition state	182
5.1.3	Catalytic antibodies demonstrate the strong entropic contribution	185
<b>5.2</b>	<b>CHEMICAL CATALYSIS</b>	<b>185</b>
5.2.1	Chemical reactions involve movement of electrons	185
5.2.2	A good leaving group is important	189
5.2.3	General acid and general base catalysis are ubiquitous	190
5.2.4	Electrophilic catalysis is also common	192
5.2.5	Thermolysin uses all these mechanisms	192
5.2.6	Nucleophilic catalysis implies a change in mechanism	194
5.2.7	Enzymes often use cofactors and coenzymes	195
5.2.8	Enzymes control water in the active site	197
<b>5.3</b>	<b>ENZYMES RECOGNIZE THE TRANSITION STATE, NOT THE SUBSTRATE</b>	<b>199</b>
5.3.1	The lock and key and induced-fit models	199
5.3.2	An enzyme should not bind strongly to its substrate	201
5.3.3	Binding and catalytic rate are closely interrelated	204
5.3.4	Transition-state analogs make good enzyme inhibitors	205
<b>5.4</b>	<b>TRIOSEPHOSPHATE ISOMERASE</b>	<b>206</b>
5.4.1	Triosephosphate isomerase uses many catalytic mechanisms	206
5.4.2	Triosephosphate isomerase is an evolutionarily perfect enzyme	209
<b>5.5</b>	<b>SUMMARY</b>	<b>211</b>
<b>5.6</b>	<b>FURTHER READING</b>	<b>211</b>
<b>5.7</b>	<b>PROBLEMS</b>	<b>212</b>
<b>5.8</b>	<b>NUMERICAL PROBLEMS</b>	<b>213</b>
<b>5.8</b>	<b>REFERENCES</b>	<b>213</b>

## Chapter 6 Protein Flexibility and Dynamics 215

<b>6.1</b>	<b>TIMESCALES AND DISTANCE SCALES OF MOTIONS</b>	<b>215</b>
6.1.1	Rapid motions are local and uncorrelated	215
6.1.2	Local motions produce global disorder	218

6.1.3	Larger-scale motions are more correlated and are therefore slower	219	7.2.7	Bacterial flagella are related to ATP synthase	258
6.1.4	Slower motions are more protein-specific than fast motions	220	7.2.8	Many membrane pumps and transporters are based on a symmetric switch	259
6.1.5	Correlated motions can occur over several hydrogen bonds	224	7.2.9	The light-driven proton pump rhodopsin is a seven-transmembrane-helix G-protein-coupled receptor	261
<b>6.2</b>	<b>CONFORMATIONAL SELECTION</b>	<b>225</b>	<b>7.3</b>	<b>MOVEMENT ALONG ACTIN AND TUBULIN FIBERS</b>	<b>263</b>
6.2.1	Proteins populate a conformational landscape	225	7.3.1	Actin and tubulin fibers continually assemble and disassemble	263
6.2.2	Conformational selection is a better model than induced fit	227	7.3.2	Cells tightly regulate fiber growth	264
6.2.3	Conformational selection and induced fit are two ends of a continuum	229	7.3.3	How cells move	267
6.2.4	Enzymes have a small population in an "activated conformation"	231	7.3.4	Vesicles are transported along microtubules	268
<b>6.3</b>	<b>FUNCTIONAL MOTION</b>	<b>233</b>	7.3.5	Large cells require more directional intracellular transport	270
6.3.1	Enzymes do not catalyze mobility along the reaction coordinate	233	7.3.6	Mitosis requires major intracellular movement	270
6.3.2	Segmental motion is essential for binding and catalysis	235	<b>7.4</b>	<b>NUCLEAR TRANSPORT</b>	<b>272</b>
6.3.3	Buried waters are important for internal mobility	236	<b>7.5</b>	<b>TRANSPORT ACROSS AND INTO MEMBRANES</b>	<b>274</b>
6.3.4	Internal dynamics can produce allostery	238	7.5.1	Transport into membranes requires a signal sequence	274
<b>6.4</b>	<b>SUMMARY</b>	<b>238</b>	7.5.2	The channel in the ER membrane is Sec61	276
<b>6.5</b>	<b>FURTHER READING</b>	<b>239</b>	7.5.3	Transport into mitochondria and chloroplasts is similar	279
<b>6.6</b>	<b>WEBSITES</b>	<b>239</b>	7.5.4	Transport requires energy	280
<b>6.7</b>	<b>PROBLEMS</b>	<b>239</b>	<b>7.6</b>	<b>SUMMARY</b>	<b>281</b>
<b>6.8</b>	<b>NUMERICAL PROBLEMS</b>	<b>240</b>	<b>7.7</b>	<b>FURTHER READING</b>	<b>281</b>
<b>6.9</b>	<b>REFERENCES</b>	<b>240</b>	<b>7.8</b>	<b>WEBSITES</b>	<b>281</b>

## Chapter 7 How Proteins Make Things Move **243**

<b>7.1</b>	<b>HOW PROTEIN MOTORS WORK</b>	<b>243</b>
7.1.1	Most intracellular motion occurs by random diffusion	243
7.1.2	Unidirectional movement requires a ratchet	245
7.1.3	Ras GTPase is the archetypal switch	246
<b>7.2</b>	<b>MOTORS, PUMPS, AND TRANSPORTERS</b>	<b>248</b>
7.2.1	Myosin is the linear motor of muscle	248
7.2.2	Myosin works by linking actin binding to head rotation	250
7.2.3	Dynein moves toward the minus end of microtubules	251
7.2.4	Kinesin moves toward the plus end of microtubules	253
7.2.5	ATP synthase is a circular motor	254
7.2.6	ATP synthase links the circular motor to a proton pump	257

<b>7.9</b>	<b>PROBLEMS</b>	<b>282</b>
------------	-----------------	------------

<b>7.10</b>	<b>NUMERICAL PROBLEMS</b>	<b>282</b>
-------------	---------------------------	------------

<b>7.11</b>	<b>REFERENCES</b>	<b>283</b>
-------------	-------------------	------------

## Chapter 8 How Proteins Transmit Signals **285**

<b>8.1</b>	<b>AN OUTLINE OF THE PROBLEMS AND SOLUTIONS</b>	<b>285</b>
8.1.1	Signaling pathways have to overcome several problems	285
8.1.2	The membrane barrier can be crossed by lipophilic signals	286
8.1.3	The membrane barrier can be overcome by receptor dimerization	286
8.1.4	The membrane barrier can be overcome by helix rotation	287
8.1.5	The membrane barrier can be crossed by opening a channel	288

8.1.6	Signaling pathways make use of specialized protein modules	288
8.1.7	Signaling pathways use these modules to achieve specificity	291
8.1.8	Signaling pathways make use of colocation to achieve specificity	292
8.2	<b>DIMERIZING RECEPTOR KINASE SYSTEMS</b>	293
8.2.1	The Jak/Stat system is a simple pathway	293
8.2.2	Receptor dimerization takes a variety of forms	296
8.2.3	Ras is the immediate target of the receptor tyrosine kinase (RTK) system	299
8.2.4	Ras activates the kinase Raf	302
8.2.5	The downstream pathway from Raf is a kinase cascade	304
8.2.6	Colocation provides extra control	305
8.2.7	Autoinhibition provides extra control	307
8.2.8	The bacterial two-component signaling system has a histidine kinase	309
8.2.9	An evolutionary perspective provides a unifying explanation	311
8.2.10	Switching off the signal	313
8.3	<b>G-PROTEIN-COUPLED RECEPTOR SIGNALING</b>	313
8.4	<b>ION CHANNELS</b>	315
8.5	<b>SIGNALING VIA PROTEOLYSIS OF A LATENT GENE REGULATORY PROTEIN</b>	317
8.5.1	Notch receptor directly activates gene transcription	317
8.5.2	Hedgehog prevents proteolysis of an intracellular signal	318
8.6	<b>SUMMARY</b>	319
8.7	<b>FURTHER READING</b>	319
8.8	<b>WEBSITES</b>	320
8.9	<b>PROBLEMS</b>	320
8.10	<b>NUMERICAL PROBLEMS</b>	321
8.11	<b>REFERENCES</b>	321

## Chapter 9 Protein Complexes: Molecular Machines 325

9.1	<b>THE CELLULAR INTERACTOME</b>	326
9.1.1	Interactomes have similar structures	326
9.1.2	The picture is still far from clear	327
9.1.3	Interacting complexes have defined but transient structures	328
9.1.4	Interactomes constitute molecular machines	329
9.2	<b>THE EXOSOME</b>	330

9.3	<b>THE RNA POLYMERASE II COMPLEX</b>	334
9.3.1	Pol II assembles sequentially	334
9.3.2	There is an electron microscopy structure of the pre-initiation complex	336
9.3.3	The C-terminal domain is a key component during elongation	337
9.4	<b>THE METABOLON CONCEPT</b>	338
9.4.1	Metabolons are controversial	338
9.4.2	Colocalization provides evidence for metabolons	340
9.4.3	Channeling provides evidence for metabolons	340
9.4.4	High-throughput methods provide no evidence for metabolons	341
9.4.5	There is reasonably good evidence for a glycolytic metabolon	342
9.5	<b>SUMMARY</b>	344
9.6	<b>FURTHER READING</b>	345
9.7	<b>WEBSITES</b>	345
9.8	<b>PROBLEMS</b>	345
9.9	<b>NUMERICAL PROBLEMS</b>	346
9.10	<b>REFERENCES</b>	347

## Chapter 10 Multienzyme Complexes 349

10.1	<b>SUBSTRATE CHANNELING</b>	350
10.1.1	Tryptophan synthase is the best example of substrate channeling	350
10.1.2	Most other examples of substrate channeling involve toxic intermediates	353
10.2	<b>CYCLIC REACTIONS</b>	354
10.2.1	Cyclic reactions require coordination	354
10.2.2	PDH has a large and complicated structure	355
10.2.3	PDH shows active-site coupling	357
10.2.4	Fatty acid synthase involves multiple rounds of a cyclic reaction	359
10.2.5	Structures of FAS contain a large cavity where the cyclic reaction occurs	360
10.2.6	$\beta$ -Oxidation is approximately the reverse of fatty acid synthesis	363
10.3	<b>ENZYME COMPLEXES THAT ARE ALMOST MECS</b>	364
10.3.1	Type I polyketide synthase is chemically similar to FAS but is not an MEC	364
10.3.2	Some polyketide synthases are proper MECs	366
10.3.3	Nonribosomal peptide biosynthesis is similar to polyketide synthase	367
10.3.4	Aromatic amino acid synthesis is a poor MEC	367

10.3.5	Integral membrane complexes are not MEC-like	369	11.4.4	Structure, electron density, and resolution	408
10.4	<b>POSSIBLE ADVANTAGES OF MULTIENZYME COMPLEXES</b>	374	11.4.5	Measures of quality: <i>R</i> factor and <i>B</i> factor	409
10.4.1	Cycling of substrates	374	11.4.6	Solvent and other molecules in protein crystals	410
10.4.2	Substrate channeling	374	11.4.7	The practicalities of protein X-ray diffraction	411
10.4.3	Speeding up reactions	374	11.4.8	Structures of membrane proteins	412
10.4.4	Faster response time	375	11.4.9	Fiber diffraction	413
10.4.5	Active-site coupling	375	11.4.10	Neutron diffraction	413
10.4.6	Increase in solvent capacity	375	11.4.11	Electron diffraction	413
10.4.7	Conclusion	375	11.5	<b>MICROSCOPY</b>	415
10.5	<b>SUMMARY</b>	376	11.5.1	Cryo-electron microscopy	415
10.6	<b>FURTHER READING</b>	376	11.5.2	Atomic force microscopy (AFM)	416
10.7	<b>WEBSITES</b>	376	11.6	<b>METHODS FOR STUDYING INTERACTIONS</b>	418
10.8	<b>PROBLEMS</b>	376	11.6.1	Surface plasmon resonance (SPR)	418
10.9	<b>NUMERICAL PROBLEMS</b>	377	11.6.2	Isothermal titration calorimetry (ITC)	419
10.10	<b>REFERENCES</b>	377	11.6.3	The Scatchard plot: an object lesson	419
			11.7	<b>MASS SPECTROMETRY</b>	421
<b>Chapter 11</b>	<b>Techniques for Studying Proteins</b>	<b>379</b>	11.8	<b>HIGH-THROUGHPUT METHODS</b>	423
11.1	<b>EXPRESSION AND PURIFICATION</b>	379	11.8.1	Proteomic analysis	423
11.2	<b>SPECTROSCOPIC METHODS</b>	383	11.8.2	Protein:protein interactions—yeast two-hybrid screens	424
11.2.1	An introduction to spectroscopic methods	383	11.8.3	Protein:protein interactions—TAP-tagging	425
11.2.2	UV/vis absorbance	384	11.9	<b>COMPUTATIONAL METHODS</b>	426
11.2.3	Circular dichroism (CD)	385	11.9.1	Bioinformatics	426
11.2.4	Fluorescence	386	11.9.2	Dynamics simulations	427
11.2.5	Single-molecule methods	389	11.9.3	Systems biology	429
11.2.6	Hydrodynamic measurements	390	11.10	<b>SUMMARY</b>	430
11.3	<b>NMR</b>	391	11.11	<b>FURTHER READING</b>	430
11.3.1	Nuclear spin and magnetization	391	11.12	<b>PROBLEMS</b>	431
11.3.2	Chemical shift	393	11.13	<b>NUMERICAL PROBLEMS</b>	432
11.3.3	Dipolar coupling	394	11.14	<b>REFERENCES</b>	433
11.3.4	<i>J</i> coupling	395			
11.3.5	Two-dimensional, three-dimensional, and four-dimensional spectra	396	<b>Glossary</b>		<b>435</b>
11.3.6	An example: the heteronuclear single-quantum coherence experiment	397	<b>Index</b>		<b>445</b>
11.3.7	Assignment of protein NMR spectra	399			
11.3.8	Chemical shift mapping	400			
11.3.9	Relaxation	401			
11.3.10	Protein structure calculation from NMR data	403			
11.4	<b>DIFFRACTION</b>	404			
11.4.1	Microscopy and the diffraction limit	404			
11.4.2	X-ray diffraction	405			
11.4.3	The phase problem in X-ray diffraction	407			

## CHAPTER 1

# Protein Structure and Evolution

Structural biology has had an enormous influence on biochemistry in general, and on the study of proteins in particular. It can almost be said that unless we know a protein's three-dimensional structure we cannot understand how it functions. However, when the crystallographer John Kendrew determined the structure of the first protein to be described in detail (myoglobin, in 1958), the most striking feature was its irregularity and complexity (or, as Max Perutz wrote, a "hideous and visceral-looking object"—**Figure 1.1** [2]). It soon became clear that proteins require this level of complexity to bind ligands and catalyze reactions specifically. But as soon as we start looking in detail at proteins, we see that there are regular patterns to the way in which proteins fold up, patterns that are determined by the underlying structures of amino acids and by the forces that dictate how they pack together. When we look at the human body, we can identify a hierarchy of structural and functional units, each dependent on the next: limbs, organs, cells, and cellular components. The same is true of proteins—each level of structure (quaternary, tertiary, secondary, and primary) depends on the one below.

Even more importantly, the structure and function of proteins are a product of evolution. This is again true of the human body: we cannot hope to understand its functions, malfunctions, and development without understanding something about the evolutionary processes that shaped it. This is why an evolutionary viewpoint pervades this book, and why a considerable part of the first chapter has been set aside to consider the implications of evolution.

Chapter 1 lays down a framework and sets the scene for the rest of the book. It is, however, far from being just an introduction, and contains some advanced material.

## 1.1 STRUCTURES OF AMINO ACIDS AND PEPTIDES

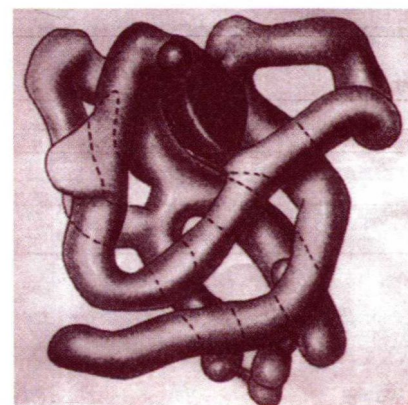
### 1.1.1 Proteins are composed of amino acids

There are 20 common amino acids coded for by DNA and translated into proteins from mRNA on ribosomes, as listed in **Table 1.1**. These are all **L-amino acids (\*1.1)**. In addition, selenocysteine is coded for by UGA, the umber codon, which is normally a termination codon; an extra nucleotide sequence slightly downstream in the mRNA directs the cell to insert selenocysteine here. Bacteria can also produce D-amino acids and unusual amino acids by using nonribosomal synthesis, which does not concern us here but is discussed further in Chapter 10. The amino acids are known both by their three-letter abbreviations and also by one-letter codes, which match the three-letter name where possible (see Table 1.1).

An amino acid consists of a carboxylic acid, which is attached to a carbon atom called the  $\alpha$ -carbon because it is adjacent to the carboxylate. In turn, the  $\alpha$ -carbon is attached to an amine (hence the name *amino acid*). In the smallest amino acid, glycine, this is all there is. In all the others, the  $\alpha$ -carbon is attached to a  $\beta$ -carbon, which in turn is often attached to further atoms. These are given succeeding letters from the Greek alphabet:  $\gamma$ ,  $\delta$ , etc. The carbonyl, C $\alpha$  and amine are called the **backbone (\*1.3)**, the other atoms being the **side chain (\*1.4)**.

*The basic laws of physics can usually be expressed in exact mathematical form, and they are probably the same throughout the universe. The "laws" of biology, by contrast, are often only broad generalizations, since they describe rather elaborate chemical mechanisms that natural selection has evolved over billions of years.*

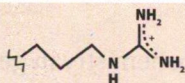
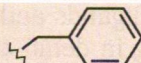
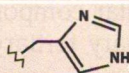


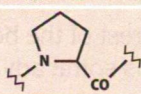
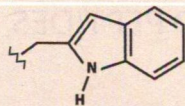
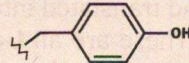

**Francis Crick (1988), [1]**



**FIGURE 1.1**

The first view of a protein structure was Kendrew's "hideous and visceral-looking object": the low-resolution crystal structure of myoglobin, obtained in 1958. At this resolution it is only possible to see the course of the peptide chain, much of which is in the form of  $\alpha$  helices. Although the internal structure of an  $\alpha$  helix is regular, the rest of the protein (the tertiary structure) is strikingly irregular. The darker region near the top is the heme, which should of course be almost completely flat; in higher-resolution structures it is indeed flat. (From J.C. Kendrew et al., *Nature* 181:662–666, 1958. With permission from Macmillan Publishers Ltd.)

TABLE 1.1 The 20 common amino acids plus selenocysteine

Name	Three-letter code	One-letter code <sup>b</sup>	Side-chain structure <sup>c</sup>	pK <sub>a</sub> of side chain	Range of pK <sub>a</sub> in proteins	Comments
Alanine	Ala	A	CH <sub>3</sub>			Hydrophobic, small
Arginine	Arg	R		12.5		Hydrophobic in middle, basic at end
Asparagine <sup>a</sup>	Asn	N	CH <sub>2</sub> -CONH <sub>2</sub>			Polar
Aspartic acid <sup>a</sup>	Asp	D	CH <sub>2</sub> -CO <sub>2</sub> <sup>-</sup>	3.9	2.0–6.7	Acidic
Cystine/cysteine	Cys	C	CH <sub>2</sub> -S-; CH <sub>2</sub> -SH	8.3	2.9–10.5	Hydrophobic Reduced (SH) is called cysteine; oxidized (S-S) is called cystine.
Phenylalanine	Phe	F				Hydrophobic, aromatic
Glutamine <sup>a</sup>	Gln	Q	CH <sub>2</sub> -CH <sub>2</sub> -CONH <sub>2</sub>			Polar
Glutamic acid <sup>a</sup>	Glu	E	CH <sub>2</sub> -CH <sub>2</sub> -CO <sub>2</sub> <sup>-</sup>	3.2	2.0–6.7	Acidic
Glycine	Gly	G	H			Hydrophobic
Histidine	His	H		6.0	2.3–9.2	Basic, aromatic
Isoleucine	Ile	I				Hydrophobic
Leucine	Leu	L				Hydrophobic
Lysine	Lys	K	CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub> -NH <sub>3</sub> <sup>+</sup>	10.5	6.0	Basic
Methionine	Met	M	CH <sub>2</sub> -CH <sub>2</sub> -S-CH <sub>3</sub>			Hydrophobic
Proline	Pro	P				Hydrophobic and hydrophilic <sup>d</sup>
Serine	Ser	S	CH <sub>2</sub> -OH	14.0		Polar
Threonine	Thr	T	CH(OH)-CH <sub>3</sub>	15.0		Polar
Tryptophan	Trp	W				Hydrophobic, aromatic
Tyrosine	Tyr	Y		9.7	6.1	Aromatic
Valine	Val	V				Hydrophobic
Selenocysteine	-	-	CH <sub>2</sub> -SeH			Hydrophobic <sup>e</sup>

Amino acids have the common structure  $^+H_3N-CH(R)-CO_2^-$ , where R is the **side chain** and the rest is the **backbone**. The table gives the structure of R.

<sup>a</sup>In addition, Asp and Asn are collectively called Asx with one-letter code B, and Glu and Gln are called Glx with code Z.

<sup>b</sup>The one-letter code for any of the 20 amino acids is usually X. The one-letter code matches the first letter of the amino acid where this is unique (C, H, I, M, S, V). Where more than one amino acid starts with the same letter, the code is assigned to the more common amino acid (A, G, L, P, T). The rest are phonetic where possible (Fenylalanine, asparagiNe, aRginine, Qtamine, tYrosine). Tryptophan has a double ring (double-u or W), and the others have a letter somewhere near the letter that the amino acid starts with (Asp D, Glu E, Lys K).

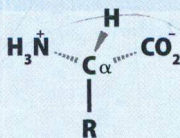
<sup>c</sup>The backbone CH carbon is the alpha carbon C $\alpha$ , and its attached proton is H $\alpha$ . The side-chain atoms are given succeeding letters from the Greek alphabet:  $\beta$  (beta),  $\gamma$  (gamma),  $\delta$  (delta),  $\epsilon$  (epsilon),  $\zeta$  (zeta),  $\eta$  (eta). In computer files such as coordinate files, these labels are given in capital letters: A, B, G, D, E, Z, H. Where there is more than one heavy atom the same distance out from C $\alpha$ , they are numbered 1 and 2; so for example the two methyl groups of a leucine are called C $\delta$ 1 and C $\delta$ 2. The dihedral angles along the side chain are called  $\chi_1$  (chi-1, pronounced kai, the angle formed by the four atoms N, C $\alpha$ , C $\beta$ , and C $\gamma$ ),  $\chi_2$ , and so on.

<sup>d</sup>The entire amino acid is drawn here. Strictly, proline is not an amino acid but an imino acid because it has an NH group, not an NH<sub>2</sub> group. As discussed in Chapter 4, the ring is hydrophobic but the main chain is unusually hydrophilic, making polyproline, for example, soluble in water.

<sup>e</sup>Selenocysteine is not normally counted as one of the standard amino acids (see the text).

### \*1.1 L-Amino acid

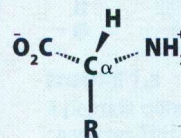
An L-amino acid is an  $\alpha$ -amino acid with L chirality at the  $\alpha$  carbon (**Figure 1.1.1**). The prefix L stands for levo and means that the related compound L-glyceraldehyde rotates polarized light to the left.



**FIGURE 1.1.1**

The  $C_\alpha$  carbon of amino acids is chiral. This figure shows an L-amino acid.

A D-amino acid (**Figure 1.1.2**) has the opposite **chirality** (\*1.2): D-glyceraldehyde rotates polarized light to the right (dextro).



**FIGURE 1.1.2**

A D-amino acid.

### \*1.2 Chirality

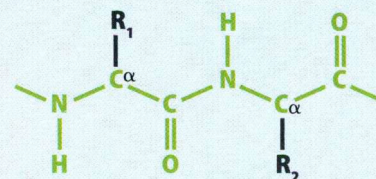
Any molecule whose reflection in a mirror cannot be superimposed is asymmetric or *chiral*. The two mirror images are called *enantiomers*, or more generally but less specifically *isomers*. Their physical and chemical properties are identical, except that one rotates plane-polarized light to the left and the other rotates it to the right. The most common origin of chirality is carbon atoms that have four nonidentical groups attached to them: for example  $C_\alpha$  carbons in

amino acids (see Figure 1.1.1). [The exception is glycine, which is not chiral because the  $C_\alpha$  has two hydrogens attached and is therefore symmetrical.] The two enantiomers are called L and D. The formal definition of L is as follows: view the  $C_\alpha$  with the  $H_\alpha$  toward you. If  $C=O$ , side chain, N go in a clockwise direction, the amino acid is L, whereas if they are anticlockwise it is D. This nomenclature is related to the organic chemistry (Cahn–Ingold–Prelog) definitions of S and R: all L-amino acids except cystine are also S.

The 20 amino acids are conveniently divided into groups. Four (Asp, Glu, Arg, and Lys) carry a charge at neutral pH: two are positive (basic: Arg and Lys) and two negative (acidic: Asp and Glu). Seven are **hydrophobic** (eight if we include glycine), and the remaining eight have polar groups. Of these, histidine is noteworthy because its  $pK_a$  is close to 7. Therefore in a protein at neutral pH it can be either protonated or not, depending on its local environment. Cysteine is also “special” because the side chain is easily oxidized to form the S–S disulfide form, where it is known as cystine. In an extracellular environment, including in the blood, cysteine is usually oxidized to cystine. However, the intracellular environment is normally sufficiently reducing that the dominant form is cysteine. Therefore one commonly finds extracellular proteins that are stabilized by disulfide bridges, whereas disulfides are not usually found in intracellular proteins. (In intracellular proteins, a similar stabilizing role is played by zinc, which binds to a combination of four cysteine or histidine side chains, forming a variety of “zinc finger” structures.) Cysteine also has a fairly low  $pK_a$ , making it a good nucleophile (\*5.7). It is therefore often found in enzyme active sites.

### \*1.3 Backbone

This is generally taken to mean the N,  $C_\alpha$  and carbonyl CO groups in a protein (**Figure 1.3.1**).

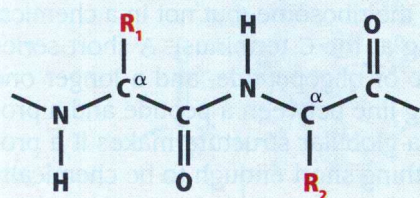


**FIGURE 1.3.1**

A protein backbone (green).

### \*1.4 Side chain

The side chain is those parts of a protein that are not the backbone (**Figure 1.4.1**). Each amino acid except glycine has a side chain (Table 1.1).



**FIGURE 1.4.1**

Protein side chains.