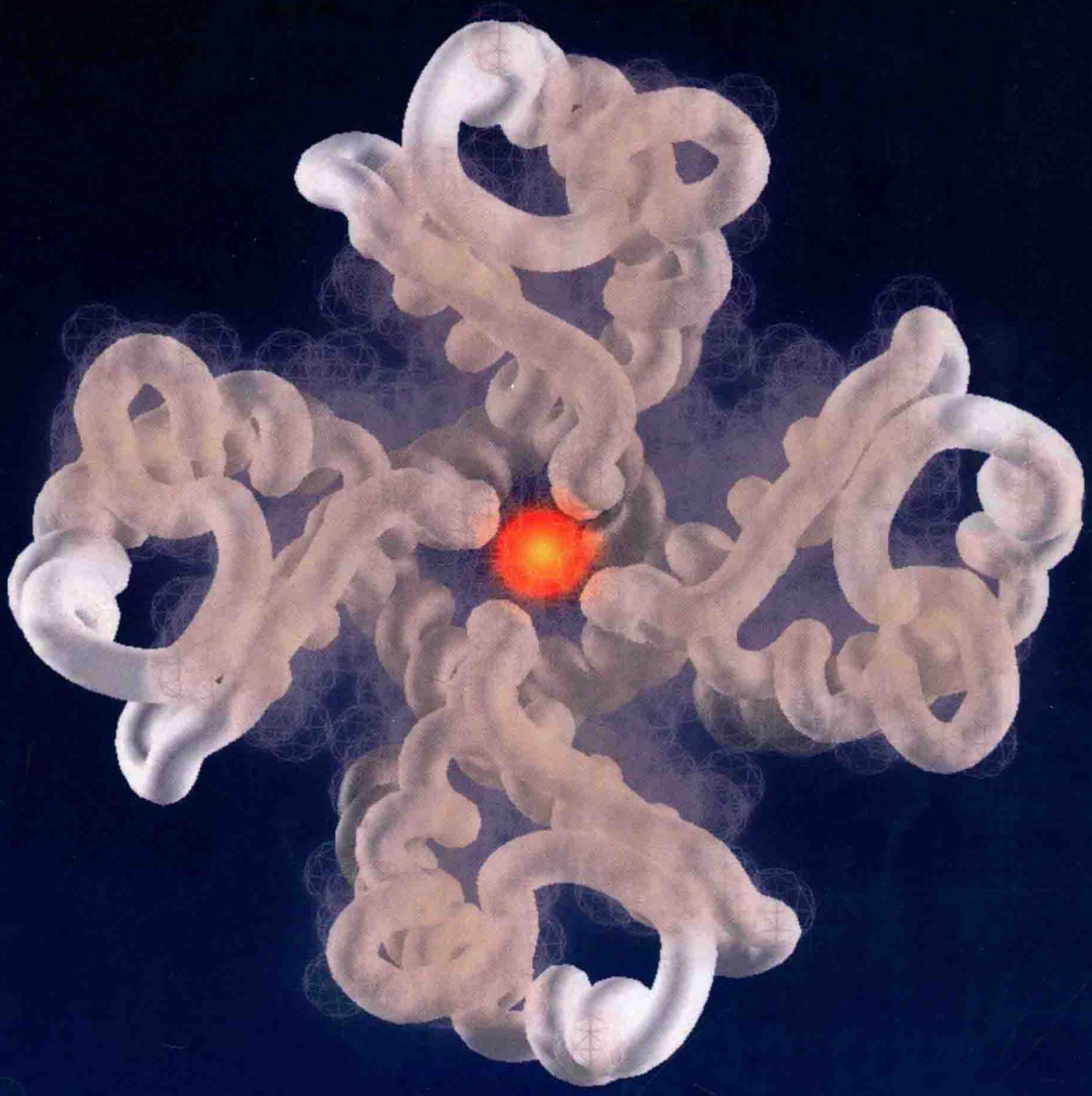
Introduction to Protein Structure

Second Edition



Carl Branden & John Tooze

Introduction to Protein Structure

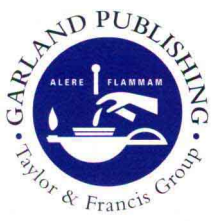
Second Edition

Carl Branden

Microbiology and Tumor Biology Center
Karolinska Institute
Stockholm
Sweden

John Tooze

Imperial Cancer Research Fund Laboratories
Lincolns Inn Fields
London
UK



THE COVER

Front: The structure of the potassium channel from *Streptomyces lividans*, determined by Rodney MacKinnon at the Rockefeller University, New York. As discussed in Chapter 12, this structure—the first of such an ion channel—shows how the channel allows the passage of potassium ions through cell membranes with high efficiency and selectivity. The view is looking down the protein as it sits in the cell membrane, as seen from outside the cell, with a potassium ion shown in gold. This image was produced using the GRASP program (A. Nicholls and B. Honig, Columbia University) from atomic coordinates kindly provided by Rodney MacKinnon.
Back: A hand-drawn image of the potassium channel, in the same view as on the front cover, with each subunit of the tetrameric protein shown in a different color.
Cover design by Christopher Thorpe and Nigel Orme.

Visit the *Introduction to Protein Structure* Web site:
http://www.proteinstructure.com/

For information on other textbooks available from Garland Publishing, visit:
http://www.garlandpub.com/

Library of Congress Cataloging-in-Publication Data
Branden, Carl–Ivar
Introduction to protein structure / Carl–Ivar Branden, John Tooze.
--2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-8153-2304-2 (hardcover). --ISBN 0-8153-2305-0 (pbk.)
1. Proteins--Structure. I. Tooze, John. II. Title.
QP551.B7635 1998
572'.633--dc21 98-34487
CIP

Published by Garland Publishing, Inc.
19 Union Square West, New York, NY 10003-3382

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

Introduction to Protein Structure

Second Edition

An aerial view of the European Synchrotron Radiation Facility at Grenoble, France, an advanced source of synchrotron x-ray radiation for use in the study of protein structure, as well as for use in the physical and material sciences. The synchrotron radiation is produced in the circular building in the lower left of the photograph. (Courtesy of ESRF.)

Preface

The determination of the atomic structures of proteins has seen an enormous increase in impetus since the first edition of this book was published in 1991. The number of new structures reported is close to doubling each year. Technical advances—for example the increased availability of synchrotron x-ray beams and methods for freezing crystals so as to reduce radiation damage to them, the development of multidimensional NMR and NMR machines with ever more powerful magnetic fields, and the exploitation of gene cloning, sequencing and expression systems have all contributed to the growth of protein structure determination. On the one hand, it is becoming increasingly easy to obtain relatively large amounts of naturally rare proteins, on the other hand the crystallographers can work with ever smaller crystals.

The fundamental tenet of molecular biology, namely that one cannot really understand biological reactions without understanding the structure of the participating molecules, is at last being vindicated. As the database of known protein structures rapidly expands, so does the range of biological pathways about which we can ask meaningful questions at close to atomic levels of resolution. An understanding of the principles of protein structure is becoming of ever widening significance to molecular biology.

The pharmaceutical industry has over the past decade become a major user of the protein structure databases, and a major contributor of newly determined structures. Knowledge of an enzyme's or a receptor's atomic structure is invaluable in the search for specific and strongly binding inhibitors. For example the quest for effective inhibitors of HIV protease, to be used in combination therapy for AIDS, led many pharmaceutical companies to determine the structure of that protease with bound inhibitors. Over 120 of these structure determinations have been done so far and at least two inhibitors of HIV protease are now being regularly used to treat AIDS. It seems certain that the determination of the atomic structure of target molecules will play an increasingly important role in drug design.

The commercial exploitation of our increased understanding of protein structure will not, of course, be restricted to the pharmaceutical industry. The industrial use of enzymes in the chemical industry, the development of new and more specific pesticides and herbicides, the modification of enzymes in order to change the composition of plant oils and plant carbohydrates are all examples of other commercial developments that depend, in part, on understanding the structure of particular proteins at high resolution.

As the complete genomes of more and more species are sequenced, the determination of the function of previously unidentified open reading frames is becoming an increasing and challenging problem. The possibility of

setting up centers for automated high through-put structure determinations is being seriously discussed. In the absence of any recognizable sequence homology to proteins of known function, this approach, surprising though it may seem, could become an effective way of determining function via structural homology.

The growth in the interest in high-resolution protein structure over the past decade and the reception of the first edition have encouraged us to prepare a new edition of this book. Universities are devoting more time to courses specifically on protein structure, or increasing the amount of time given to protein structure in more generally based biology and biochemistry courses. We hope that this new edition of *Introduction to Protein Structure* will prove useful both to teachers and students.

In 1988 when we began writing the first edition, about 250 protein structures had been determined to medium to high resolution and in those days a professional protein crystallographer was familiar with most of them. We were not therefore faced with a severe problem of what to leave out as we wrote. Today, the coordinates of over 6500 proteins have been deposited in the Protein Data Bank at Brookhaven, New York. Both the number of structures and the variety of biological systems to which they relate are so high that the field of protein structure is becoming more fragmented and specialized. It is becoming increasingly difficult to keep sight of the wood amongst so many trees. The question of what to include and what to omit is, for today's authors, crucially important. We have tried to resist the temptation to describe more and more proteins, adding detail but not increasing understanding of the basic concepts. This edition is inescapably a little larger than its predecessor, but to contain the increase in size we have deleted two chapters while adding four. We run the risk of disappointing not a few structural biologists whose favourite proteins are not mentioned. To them we apologize and ask for their understanding.

Acknowledgements

In preparing the second edition of this book we have again relied heavily on and benefited greatly from the advice and constructive criticism of numerous colleagues. We are particularly grateful to Ken Holmes (Max-Planck Institute, Heidelberg), Lawrence Stern (MIT), Michelle Arkin (Sunesis Pharmaceuticals), and Watson Fuller (Keele University, UK) for their contributions to, respectively, Chapters 14 and 18, Chapter 15, Chapter 17 and Chapter 18. Stephen Harrison (Harvard University) and Paul Sigler (Yale University) provided extensive help and advice on Chapters 8–10, 13 and 16, and Chapters 8–10 and 13, respectively, for which we are especially grateful.

The following, in alphabetical order, have reviewed one or more chapters, correcting our errors of fact or interpretation and helping to ensure they have the appropriate balance and emphasis: Tom Alber (University of California, Berkeley), Tom Blundell (Cambridge University, UK), Stephen Burley (Rockefeller University), Charles Craik (University of California, San Francisco), Ken Dill (University of California, San Francisco), Chris Dobson (Oxford University, UK), Anthony Fink (Unversity of California, Santa Cruz), Robert Fletterick (University of California, San Francisco), Richard Henderson (LMB, Cambridge, UK), Werner Kühlbrandt (MPI, Frankfurt), David Parry (Massey University, New Zealand), Greg Petsko (Brandeis University), and David Trentham (NIMR, London, UK).

The book depends for its accessibility upon its illustrations and we are hugely indebted to Nigel Orme, who, as with the first edition, has converted sketches into lucid figures. Keith Roberts has again advised us on how best graphically to represent chemical and structural phenomena. Jane Richardson (Duke University) has generously produced the Kinemage supplement to this edition and the book relies upon Richardson-type diagrams throughout to render the structures discussed comprehensible. We thank our publishers Garland Publishing, now part of the Taylor and Francis Group, for their support, and in particular Matthew Day for his enthusiastic editing of the complete manuscript. Miranda Robertson, in her inimitable style, has again managed the entire project.

Contents

Part I Basic Structural Principles 1

1. The Building Blocks 3
- Proteins are polypeptide chains 4
- The genetic code specifies 20 different amino acid side chains 4
- Cysteines can form disulfide bridges 5
- Peptide units are building blocks of protein structures 8
- Glycine residues can adopt many different conformations 9
- Certain side-chain conformations are energetically favorable 10
- Many proteins contain intrinsic metal atoms 11
- Conclusion 12
- Selected readings 12

2. Motifs of Protein Structure 13
- The interior of proteins is hydrophobic 14
- The alpha (α) helix is an important element of secondary structure 14
- The α helix has a dipole moment 16
- Some amino acids are preferred in α helices 16
- Beta (β) sheets usually have their β strands either parallel or antiparallel 19
- Loop regions are at the surface of protein molecules 21
- Schematic pictures of proteins highlight secondary structure 22
- Topology diagrams are useful for classification of protein structures 23
- Secondary structure elements are connected to form simple motifs 24
- The hairpin β motif occurs frequently in protein structures 26
- The Greek key motif is found in antiparallel β sheets 27
- The β-α-β motif contains two parallel β strands 27
- Protein molecules are organized in a structural hierarchy 28
- Large polypeptide chains fold into several domains 29
- Domains are built from structural motifs 30
- Simple motifs combine to form complex motifs 30
- Protein structures can be divided into three main classes 31
- Conclusion 32
- Selected readings 33

3. Alpha-Domain Structures 35
- Coiled-coil α helices contain a repetitive heptad amino acid sequence pattern 35
- The four-helix bundle is a common domain structure in α proteins 37
- Alpha-helical domains are sometimes large and complex 39
- The globin fold is present in myoglobin and hemoglobin 40
- Geometric considerations determine α-helix packing 40
- Ridges of one α helix fit into grooves of an adjacent helix 40
- The globin fold has been preserved during evolution 41
- The hydrophobic interior is preserved 42
- Helix movements accommodate interior side-chain mutations 43
- Sickle-cell hemoglobin confers resistance to malaria 43
- Conclusion 45
- Selected readings 45

4. Alpha/Beta Structures 47
- Parallel β strands are arranged in barrels or sheets 47
- Alpha/beta barrels occur in many different enzymes 48
- Branched hydrophobic side chains dominate the core of α/β barrels 49
- Pyruvate kinase contains several domains, one of which is an α/β barrel 51
- Double barrels have occurred by gene fusion 52
- The active site is formed by loops at one end of the α/β barrel 53

Alpha/beta barrels provide examples of evolution of new enzyme activities 54
Leucine-rich motifs form an α/β-horseshoe fold 55
Alpha/beta twisted open-sheet structures contain α helices on both sides of the β sheet 56
Open β-sheet structures have a variety of topologies 57
The positions of active sites can be predicted in α/β structures 57
Tyrosyl-tRNA synthetase has two different domains (α/β + α) 59
Carboxypeptidase is an α/β protein with a mixed β sheet 60
Arabinose-binding protein has two similar α/β domains 62
Conclusion 63
Selected readings 64

5. Beta Structures 67
Up-and-down barrels have a simple topology 68
The retinol-binding protein binds retinol inside an up-and-down β barrel 68
Amino acid sequence reflects β structure 69
The retinol-binding protein belongs to a superfamily of protein structures 70
Neuraminidase folds into up-and-down β sheets 70
Folding motifs form a propeller-like structure in neuraminidase 71
The active site is in the middle of one side of the propeller 72
Greek key motifs occur frequently in antiparallel β structures 72
The γ-crystallin molecule has two domains 74
The domain structure has a simple topology 74
Two Greek key motifs form the domain 74
The two domains have identical topology 75
The two domains have similar structures 76
The Greek key motifs in γ crystallin are evolutionarily related 76
The Greek key motifs can form jelly roll barrels 77
The jelly roll motif is wrapped around a barrel 77
The jelly roll barrel is usually divided into two sheets 78
The functional hemagglutinin subunit has two polypeptide chains 79
The subunit structure is divided into a stem and a tip 79
The receptor binding site is formed by the jelly roll domain 80
Hemagglutinin acts as a membrane fusogen 80
The structure of hemagglutinin is affected by pH changes 81
Parallel β-helix domains have a novel fold 84
Conclusion 85
Selected readings 87

6. Folding and Flexibility 89
Globular proteins are only marginally stable 90
Kinetic factors are important for folding 91
Molten globules are intermediates in folding 92
Burying hydrophobic side chains is a key event 93
Both single and multiple folding pathways have been observed 93
Enzymes assist formation of proper disulfide bonds during folding 96
Isomerization of proline residues can be a rate-limiting step in protein folding 98
Proteins can fold or unfold inside chaperonins 99
GroEL is a cylindrical structure with a central channel in which newly synthesized polypeptides bind 100
GroES closes off one end of the GroEL cylinder 102
The GroEL–GroES complex binds and releases newly synthesized polypeptides in an ATP-dependent cycle 102
The folded state has a flexible structure 104
Conformational changes in a protein kinase are important for cell cycle regulation 105
Peptide binding to calmodulin induces a large interdomain movement 109
Serpins inhibit serine proteinases with a spring-loaded safety catch mechanism 110
Effector molecules switch allosteric proteins between R and T states 113
X-ray structures explain the allosteric properties of phosphofructokinase 114
Conclusion 117
Selected readings 119

7. DNA Structures 121
The DNA double helix is different in A- and B-DNA 121
The DNA helix has major and minor grooves 122
Z-DNA forms a zigzag pattern 123
B-DNA is the preferred conformation *in vivo* 124
Specific base sequences can be recognized in B-DNA 124
Conclusion 125
Selected readings 126

Part 2 Structure, Function, and Engineering 127

8. DNA Recognition in Procaryotes by Helix-Turn-Helix Motifs 129
A molecular mechanism for gene control 129
Repressor and Cro proteins operate a procaryotic genetic switch region 130
The x-ray structure of the complete lambda Cro protein is known 131
The x-ray structure of the DNA-binding domain of the lambda repressor is known 132
Both lambda Cro and repressor proteins have a specific DNA-binding motif 133
Model building predicts Cro–DNA interactions 134
Genetic studies agree with the structural model 135
The x-ray structure of DNA complexes with 434 Cro and repressor revealed novel features of protein–DNA interactions 136
The structures of 434 Cro and the 434 repressor DNA-binding domain are very similar 137
The proteins impose precise distortions on the B-DNA in the complexes 138
Sequence-specific protein–DNA interactions recognize operator regions 138

Protein–DNA backbone interactions determine DNA conformation 139
Conformational changes of DNA are important for differential binding of repressor and Cro to different operator sites 140
The essence of phage repressor and Cro 141
DNA binding is regulated by allosteric control 142
The *trp* repressor forms a helix-turn-helix motif 142
A conformational change operates a functional switch 142
Lac repressor binds to both the major and minor grooves inducing a sharp bend in the DNA 143
CAP-induced DNA bending could activate transcription 146
Conclusion 147
Selected readings 148

9. DNA Recognition by Eucaryotic Transcription Factors 151
Transcription is activated by protein–protein interactions 152
The TATA box-binding protein is ubiquitous 153
The three-dimensional structures of TBP–TATA box complexes are known 154
A β sheet in TBP forms the DNA-binding site 154
TBP binds in the minor groove and induces large structural changes in DNA 155
The interaction area between TBP and the TATA box is mainly hydrophobic 157
Functional implications of the distortion of DNA by TBP 158
TFIIA and TFIIB bind to both TBP and DNA 159
Homeodomain proteins are involved in the development of many eucaryotic organisms 159
Monomers of homeodomain proteins bind to DNA through a helix-turn-helix motif 160
In vivo specificity of homeodomain transcription factors depends on interactions with other proteins 162
POU regions bind to DNA by two tandemly oriented helix-turn-helix motifs 164
Much remains to be learnt about the function of homeodomains *in vivo* 166
Understanding tumorigenic mutations 166
The monomeric p53 polypeptide chain is divided in three domains 167
The oligomerization domain forms tetramers 167
The DNA-binding domain of p53 is an antiparallel β barrel 168
Two loop regions and one α helix of p53 bind to DNA 169
Tumorigenic mutations occur mainly in three regions involved in DNA binding 170
Conclusions 172
Selected readings 172

10. Specific Transcription Factors Belong to a Few Families 175
Several different groups of zinc-containing motifs have been observed 176
The classic zinc fingers bind to DNA in tandem along the major groove 177
The finger region of the classic zinc finger motif interacts with DNA 178
Two zinc-containing motifs in the glucocorticoid receptor form one DNA-binding domain 181
A dimer of the glucocorticoid receptor binds to DNA 183
An α helix in the first zinc motif provides the specific protein–DNA interactions 184
Three residues in the recognition helix provide the sequence-specific interactions with DNA 184
The retinoid X receptor forms heterodimers that recognize tandem repeats with variable spacings 185
Yeast transcription factor GAL4 contains a binuclear zinc cluster in its DNA-binding domain 187
The zinc cluster regions of GAL4 bind at the two ends of the enhancer element 188
The linker region also contributes to DNA binding 189
DNA-binding site specificity among the C_6-zinc cluster family of transcription factors is achieved by the linker regions 190
Families of zinc-containing transcription factors bind to DNA in several different ways 191
Leucine zippers provide dimerization interactions for some eucaryotic transcription factors 191
The GCN4 basic region leucine zipper binds DNA as a dimer of two uninterrupted α helices 193
GCN4 binds to DNA with both specific and nonspecific contacts 194
The HLH motif is involved in homodimer and heterodimer associations 196
The α-helical basic region of the b/HLH motif binds in the major groove of DNA 198
The b/HLH/zip family of transcription factors have both HLH and leucine zipper dimerization motifs 199
Max and MyoD recognize the DNA HLH consensus sequence by different specific protein–DNA interactions 201
Conclusion 201
Selected readings 203

11. An Example of Enzyme Catalysis: Serine Proteinases 205
Proteinases form four functional families 205
The catalytic properties of enzymes are reflected in K_m and k_{cat} values 206
Enzymes decrease the activation energy of chemical reactions 206
Serine proteinases cleave peptide bonds by forming tetrahedral transition states 208
Four important structural features are required for the catalytic action of serine proteinases 209
Convergent evolution has produced two different serine proteinases with similar catalytic mechanisms 210
The chymotrypsin structure has two antiparallel β-barrel domains 210

The active site is formed by two loop regions from each domain 211
Did the chymotrypsin molecule evolve by gene duplication? 212
Different side chains in the substrate specificity pocket confer preferential cleavage 212
Engineered mutations in the substrate specificity pocket change the rate of catalysis 213
The Asp 189-Lys mutation in trypsin causes unexpected changes in substrate specificity 215
The structure of the serine proteinase subtilisin is of the α/β type 215
The active sites of subtilisin and chymotrypsin are similar 216
A structural anomaly in subtilisin has functional consequences 217
Transition-state stabilization in subtilisin is dissected by protein engineering 217
Catalysis occurs without a catalytic triad 217
Substrate molecules provide catalytic groups in substrate-assisted catalysis 218
Conclusion 219
Selected readings 220

12. Membrane Proteins 223
Membrane proteins are difficult to crystallize 224
Novel crystallization methods are being developed 224
Two-dimensional crystals of membrane proteins can be studied by electron microscopy 225
Bacteriorhodopsin contains seven transmembrane α helices 226
Bacteriorhodopsin is a light-driven proton pump 227
Porins form transmembrane channels by β strands 228
Porin channels are made by up and down β barrels 229
Each porin molecule has three channels 230
Ion channels combine ion selectivity with high levels of ion conductance 232
The K^+ channel is a tetrameric molecule with one ion pore in the interface between the four subunits 232
The ion pore has a narrow ion selectivity filter 233
The bacterial photosynthetic reaction center is built up from four different polypeptide chains and many pigments 234
The L, M, and H subunits have transmembrane α helices 236
The photosynthetic pigments are bound to the L and M subunits 237
Reaction centers convert light energy into electrical energy by electron flow through the membrane 239
Antenna pigment proteins assemble into multimeric light-harvesting particles 240
Chlorophyll molecules form circular rings in the light-harvesting complex LH2 241
The reaction center is surrounded by a ring of 16 antenna proteins of the light-harvesting complex LH1 242
Transmembrane α helices can be predicted from amino acid sequences 244
Hydrophobicity scales measure the degree of hydrophobicity of different amino acid side chains 245
Hydropathy plots identify transmembrane helices 245
Reaction center hydropathy plots agree with crystal structural data 246
Membrane lipids have no specific interaction with protein transmembrane α helices 246
Conclusion 247
Selected readings 248

13. Signal Transduction 251
G proteins are molecular amplifiers 252
Ras proteins and the catalytic domain of G_α have similar three-dimensional structures 254
G_α is activated by conformational changes of three switch regions 257
GTPases hydrolyze GTP through nucleophilic attack by a water molecule 259
The G_β subunit has a seven-blade propeller fold, built up from seven WD repeat units 261
The GTPase domain of G_α binds to G_β in the heterotrimeric $G_{\alpha\beta\gamma}$ complex 263
Phosducin regulates light adaptation in retinal rods 265
Phosducin binding to $G_{\beta\gamma}$ blocks binding of G_α 265
The human growth hormone induces dimerization of its cognate receptor 267
Dimerization of the growth hormone receptor is a sequential process 268
The growth hormone also binds to the prolactin receptor 269
Tyrosine kinase receptors are important enzyme-linked receptors 270
Small protein modules form adaptors for a signaling network 272
SH2 domains bind to phosphotyrosine-containing regions of target molecules 273
SH3 domains bind to proline-rich regions of target molecules 274
Src tyrosine kinases comprise SH2 and SH3 domains in addition to a tyrosine kinase 275
The two domains of the kinase in the inactive state are held in a closed conformation by assembly of the regulatory domains 277
Conclusion 278
Selected readings 280

14. Fibrous Proteins 283
Collagen is a superhelix formed by three parallel, very extended left-handed helices 284
Coiled coils are frequently used to form oligomers of fibrous and globular proteins 286
Amyloid fibrils are suggested to be built up from continuous β sheet helices 288
Spider silk is nature's high-performance fiber 289
Muscle fibers contain myosin and actin which slide against each other during muscle contraction 290

Myosin heads form cross-bridges between the actin and myosin filaments 291
Time-resolved x-ray diffraction of frog muscle confirmed movement of the cross-bridges 292
Structures of actin and myosin have been determined 293
The structure of myosin supports the swinging cross-bridge hypothesis 295
The role of ATP in muscular contraction has parallels to the role of GTP in G-protein activation 296
Conclusion 297
Selected readings 298

15. Recognition of Foreign Molecules by the Immune System 299
The polypeptide chains of antibodies are divided into domains 300
Antibody diversity is generated by several different mechanisms 302
All immunoglobulin domains have similar three-dimensional structures 303
The immunoglobulin fold is best described as two antiparallel β sheets packed tightly against each other 304
The hypervariable regions are clustered in loop regions at one end of the variable domain 305
The antigen-binding site is formed by close association of the hypervariable regions from both heavy and light chains 306
The antigen-binding site binds haptens in crevices and protein antigens on large flat surfaces 308
The CDR loops assume only a limited range of conformations, except for the heavy chain CDR3 311
An IgG molecule has several degrees of conformational flexibility 312
Structures of MHC molecules have provided insights into the molecular mechanisms of T-cell activation 312
MHC molecules are composed of antigen-binding and immunoglobulin-like domains 313
Recognition of antigen is different in MHC molecules compared with immunoglobulins 314
Peptides are bound differently by class I and class II MHC molecules 315
T-cell receptors have variable and constant immunoglobulin domains and hypervariable regions 316
MHC–peptide complexes are the ligands for T-cell receptors 318
Many cell-surface receptors contain immunoglobulin-like domains. 318
Conclusion 320
Selected readings 321

16. The Structure of Spherical Viruses 325
The protein shells of spherical viruses have icosahedral symmetry 327
The icosahedron has high symmetry 327
The simplest virus has a shell of 60 protein subunits 328
Complex spherical viruses have more than one polypeptide chain in the asymmetric unit 329
Structural versatility gives quasi-equivalent packing in T = 3 plant viruses 331
The protein subunits recognize specific parts of the RNA inside the shell 332
The protein capsid of picornaviruses contains four polypeptide chains 333
There are four different structural proteins in picornaviruses 334
The arrangement of subunits in the shell of picornaviruses is similar to that of T = 3 plant viruses 334
The coat proteins of many different spherical plant and animal viruses have similar jelly roll barrel structures, indicating an evolutionary relationship 335
Drugs against the common cold may be designed from the structure of rhinovirus 337
Bacteriophage MS2 has a different subunit structure 339
A dimer of MS2 subunits recognizes an RNA packaging signal 339
The core protein of alphavirus has a chymotrypsin-like fold 340
SV40 and polyomavirus shells are constructed from pentamers of the major coat protein with nonequivalent packing but largely equivalent interactions 341
Conclusion 343
Selected readings 344

17. Prediction, Engineering, and Design of Protein Structures 347
Homologous proteins have similar structure and function 348
Homologous proteins have conserved structural cores and variable loop regions 349
Knowledge of secondary structure is necessary for prediction of tertiary structure 350
Prediction methods for secondary structure benefit from multiple alignment of homologous proteins 351
Many different amino acid sequences give similar three-dimensional structures 352
Prediction of protein structure from sequence is an unsolved problem 352
Threading methods can assign amino acid sequences to known three-dimensional folds 353
Proteins can be made more stable by engineering 354
Disulfide bridges increase protein stability 355
Glycine and proline have opposite effects on stability 356
Stabilizing the dipoles of α helices increases stability 357
Mutants that fill cavities in hydrophobic cores do not stabilize T4 lysozyme 358
Proteins can be engineered by combinatorial methods 358
Phage display links the protein library to DNA 359
Affinity and specificity of proteinase inhibitors can be optimized by phage display 361

Structural scaffolds can be reduced in size while function is retained 363
Phage display of random peptide libraries identified agonists of erythropoetin receptor 364
DNA shuffling allows accelerated evolution of genes 365
Protein structures can be designed from first principles 367
A β structure has been converted to an α structure by changing only half of the sequence 368
Conclusion 370
Selected readings 371

18. Determination of Protein Structures 373
Several different techniques are used to study the structure of protein molecules 373
Protein crystals are difficult to grow 374
X-ray sources are either monochromatic or polychromatic 376
X-ray data are recorded either on image plates or by electronic detectors 377
The rules for diffraction are given by Bragg's law 378
Phase determination is the major crystallographic problem 379
Phase information can also be obtained by Multiwavelength Anomalous Diffraction experiments 381
Building a model involves subjective interpretation of the data 381
Errors in the initial model are removed by refinement 383
Recent technological advances have greatly influenced protein crystallography 383
X-ray diffraction can be used to study the structure of fibers as well as crystals 384
The structure of biopolymers can be studied using fiber diffraction 386
NMR methods use the magnetic properties of atomic nuclei 387
Two-dimensional NMR spectra of proteins are interpreted by the method of sequential assignment 389
Distance constraints are used to derive possible structures of a protein molecule 390
Biochemical studies and molecular structure give complementary functional information 391
Conclusion 391
Selected readings 392

Protein Structure on the World Wide Web 393

Part 1

Basic Structural Principles

The Building Blocks

1

Recombinant DNA techniques have provided tools for the rapid determination of DNA sequences and, by inference, the amino acid sequences of proteins from structural genes. The number of such sequences is now increasing almost exponentially, but by themselves these sequences tell little more about the biology of the system than a New York City telephone directory tells about the function and marvels of that city.

The proteins we observe in nature have evolved, through selective pressure, to perform specific functions. The functional properties of proteins depend upon their three-dimensional structures. The three-dimensional structure arises because particular sequences of amino acids in polypeptide chains fold to generate, from linear chains, compact domains with specific three-dimensional structures (Figure 1.1). The folded domains can serve as modules for building up large assemblies such as virus particles or muscle fibers, or they can provide specific catalytic or binding sites, as found in enzymes or proteins that carry oxygen or that regulate the function of DNA.

To understand the biological function of proteins we would therefore like to be able to deduce or predict the three-dimensional structure from the amino acid sequence. This we cannot do. In spite of considerable efforts over the past 25 years, this folding problem is still unsolved and remains one of the most basic intellectual challenges in molecular biology.

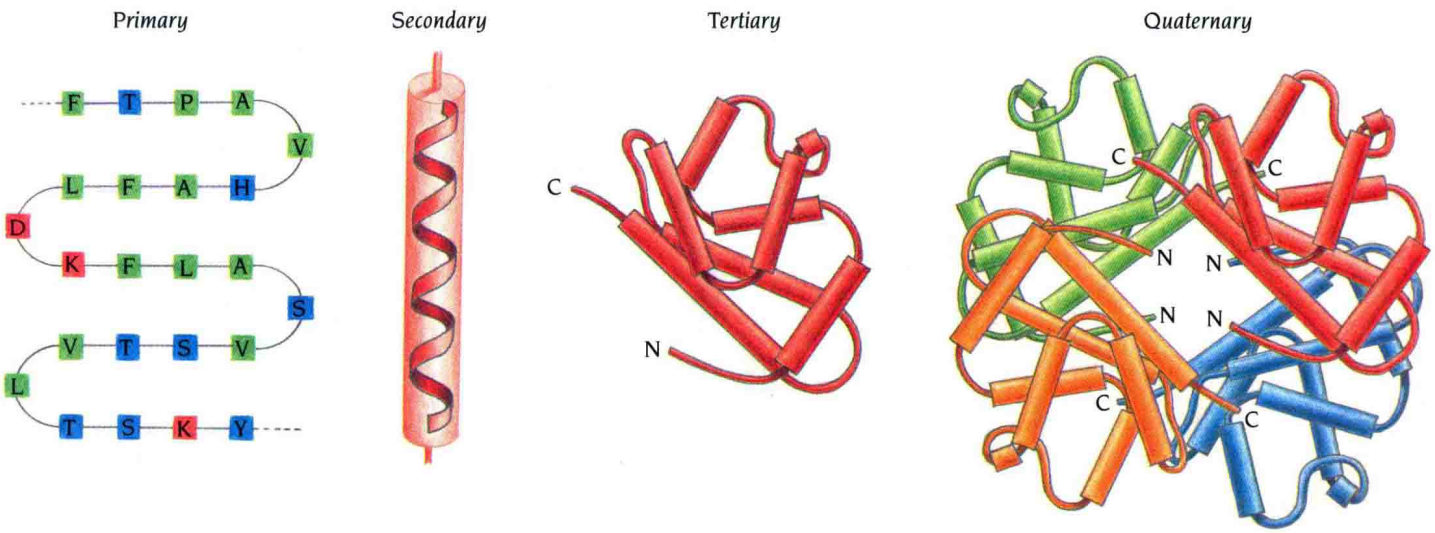


Figure 1.1 The amino acid sequence of a protein's polypeptide chain is called its **primary** structure. Different regions of the sequence form local regular **secondary** structures, such as alpha (α) helices or beta (β) strands. The **tertiary** structure is formed by packing such structural elements into one or several compact globular units called domains. The final protein may contain several polypeptide chains arranged in a **quaternary** structure. By formation of such tertiary and quaternary structure amino acids far apart in the sequence are brought close together in three dimensions to form a functional region, an **active site**.