
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



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	Both single and multiple folding pathways have been observed	93
Leucine-rich motifs form an α/β -horseshoe fold	55	Enzymes assist formation of proper disulfide bonds during folding	96
Alpha/beta twisted open-sheet structures contain α helices on both sides of the β sheet	56	Isomerization of proline residues can be a rate-limiting step in protein folding	98
Open β -sheet structures have a variety of topologies	57	Proteins can fold or unfold inside chaperonins	99
The positions of active sites can be predicted in α/β structures	57	GroEL is a cylindrical structure with a central channel in which newly synthesized polypeptides bind	100
Tyrosyl-tRNA synthetase has two different domains ($\alpha/\beta + \alpha$)	59	GroES closes off one end of the GroEL cylinder	102
Carboxypeptidase is an α/β protein with a mixed β sheet	60	The GroEL–GroES complex binds and releases newly synthesized polypeptides in an ATP-dependent cycle	102
Arabinose-binding protein has two similar α/β domains	62	The folded state has a flexible structure	104
Conclusion	63	Conformational changes in a protein kinase are important for cell cycle regulation	105
Selected readings	64	Peptide binding to calmodulin induces a large interdomain movement	109
5. Beta Structures	67	Serpins inhibit serine proteinases with a spring-loaded safety catch mechanism	110
Up-and-down barrels have a simple topology	68	Effector molecules switch allosteric proteins between R and T states	113
The retinol-binding protein binds retinol inside an up-and-down β barrel	68	X-ray structures explain the allosteric properties of phosphofructokinase	114
Amino acid sequence reflects β structure	69	Conclusion	117
The retinol-binding protein belongs to a superfamily of protein structures	70	Selected readings	119
Neuraminidase folds into up-and-down β sheets	70	7. DNA Structures	121
Folding motifs form a propeller-like structure in neuraminidase	71	The DNA double helix is different in A- and B-DNA	121
The active site is in the middle of one side of the propeller	72	The DNA helix has major and minor grooves	122
Greek key motifs occur frequently in antiparallel β structures	72	Z-DNA forms a zigzag pattern	123
The γ -crystallin molecule has two domains	74	B-DNA is the preferred conformation <i>in vivo</i>	124
The domain structure has a simple topology	74	Specific base sequences can be recognized in B-DNA	124
Two Greek key motifs form the domain	74	Conclusion	125
The two domains have identical topology	75	Selected readings	126
The two domains have similar structures	76	Part 2 Structure, Function, and Engineering	127
The Greek key motifs in γ crystallin are evolutionarily related	76	8. DNA Recognition in Prokaryotes by Helix-Turn-Helix Motifs	129
The Greek key motifs can form jelly roll barrels	77	A molecular mechanism for gene control	129
The jelly roll motif is wrapped around a barrel	77	Repressor and Cro proteins operate a prokaryotic genetic switch region	130
The jelly roll barrel is usually divided into two sheets	78	The x-ray structure of the complete lambda Cro protein is known	131
The functional hemagglutinin subunit has two polypeptide chains	79	The x-ray structure of the DNA-binding domain of the lambda repressor is known	132
The subunit structure is divided into a stem and a tip	79	Both lambda Cro and repressor proteins have a specific DNA-binding motif	133
The receptor binding site is formed by the jelly roll domain	80	Model building predicts Cro–DNA interactions	134
Hemagglutinin acts as a membrane fusogen	80	Genetic studies agree with the structural model	135
The structure of hemagglutinin is affected by pH changes	81	The x-ray structure of DNA complexes with 434 Cro and repressor revealed novel features of protein–DNA interactions	136
Parallel β -helix domains have a novel fold	84	The structures of 434 Cro and the 434 repressor DNA-binding domain are very similar	137
Conclusion	85	The proteins impose precise distortions on the B-DNA in the complexes	138
Selected readings	87	Sequence-specific protein–DNA interactions recognize operator regions	138
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		

Protein–DNA backbone interactions determine DNA conformation	139	The finger region of the classic zinc finger motif interacts with DNA	178
Conformational changes of DNA are important for differential binding of repressor and Cro to different operator sites	140	Two zinc-containing motifs in the glucocorticoid receptor form one DNA-binding domain	181
The essence of phage repressor and Cro	141	A dimer of the glucocorticoid receptor binds to DNA	183
DNA binding is regulated by allosteric control	142	An α helix in the first zinc motif provides the specific protein–DNA interactions	184
The <i>trp</i> repressor forms a helix-turn-helix motif	142	Three residues in the recognition helix provide the sequence-specific interactions with DNA	184
A conformational change operates a functional switch	142	The retinoid X receptor forms heterodimers that recognize tandem repeats with variable spacings	185
Lac repressor binds to both the major and minor grooves inducing a sharp bend in the DNA	143	Yeast transcription factor GAL4 contains a binuclear zinc cluster in its DNA-binding domain	187
CAP-induced DNA bending could activate transcription	146	The zinc cluster regions of GAL4 bind at the two ends of the enhancer element	188
Conclusion	147	The linker region also contributes to DNA binding	189
Selected readings	148	DNA-binding site specificity among the C ₆ -zinc cluster family of transcription factors is achieved by the linker regions	190
9. DNA Recognition by Eucaryotic Transcription Factors	151	Families of zinc-containing transcription factors bind to DNA in several different ways	191
Transcription is activated by protein–protein interactions	152	Leucine zippers provide dimerization interactions for some eucaryotic transcription factors	191
The TATA box-binding protein is ubiquitous	153	The GCN4 basic region leucine zipper binds DNA as a dimer of two uninterrupted α helices	193
The three-dimensional structures of TBP–TATA box complexes are known	154	GCN4 binds to DNA with both specific and nonspecific contacts	194
A β sheet in TBP forms the DNA-binding site	154	The HLH motif is involved in homodimer and heterodimer associations	196
TBP binds in the minor groove and induces large structural changes in DNA	155	The α -helical basic region of the b/HLH motif binds in the major groove of DNA	198
The interaction area between TBP and the TATA box is mainly hydrophobic	157	The b/HLH/zip family of transcription factors have both HLH and leucine zipper dimerization motifs	199
Functional implications of the distortion of DNA by TBP	158	Max and MyoD recognize the DNA HLH consensus sequence by different specific protein–DNA interactions	201
TFIIA and TFIIB bind to both TBP and DNA	159	Conclusion	201
Homeodomain proteins are involved in the development of many eucaryotic organisms	159	Selected readings	203
Monomers of homeodomain proteins bind to DNA through a helix-turn-helix motif	160	11. An Example of Enzyme Catalysis: Serine Proteinases	205
<i>In vivo</i> specificity of homeodomain transcription factors depends on interactions with other proteins	162	Proteinases form four functional families	205
POU regions bind to DNA by two tandemly oriented helix-turn-helix motifs	164	The catalytic properties of enzymes are reflected in K_m and k_{cat} values	206
Much remains to be learnt about the function of homeodomains <i>in vivo</i>	166	Enzymes decrease the activation energy of chemical reactions	206
Understanding tumorigenic mutations	166	Serine proteinases cleave peptide bonds by forming tetrahedral transition states	208
The monomeric p53 polypeptide chain is divided in three domains	167	Four important structural features are required for the catalytic action of serine proteinases	209
The oligomerization domain forms tetramers	167	Convergent evolution has produced two different serine proteinases with similar catalytic mechanisms	210
The DNA-binding domain of p53 is an antiparallel β barrel	168	The chymotrypsin structure has two antiparallel β -barrel domains	210
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 active site is formed by two loop regions from each domain	211	Transmembrane α helices can be predicted from amino acid sequences	244
Did the chymotrypsin molecule evolve by gene duplication?	212	Hydrophobicity scales measure the degree of hydrophobicity of different amino acid side chains	245
Different side chains in the substrate specificity pocket confer preferential cleavage	212	Hydropathy plots identify transmembrane helices	245
Engineered mutations in the substrate specificity pocket change the rate of catalysis	213	Reaction center hydropathy plots agree with crystal structural data	246
The Asp 189-Lys mutation in trypsin causes unexpected changes in substrate specificity	215	Membrane lipids have no specific interaction with protein transmembrane α helices	246
The structure of the serine proteinase subtilisin is of the α/β type	215	Conclusion	247
The active sites of subtilisin and chymotrypsin are similar	216	Selected readings	248
A structural anomaly in subtilisin has functional consequences	217	13. Signal Transduction	251
Transition-state stabilization in subtilisin is dissected by protein engineering	217	G proteins are molecular amplifiers	252
Catalysis occurs without a catalytic triad	217	Ras proteins and the catalytic domain of G_{α} have similar three-dimensional structures	254
Substrate molecules provide catalytic groups in substrate-assisted catalysis	218	G_{α} is activated by conformational changes of three switch regions	257
Conclusion	219	GTPases hydrolyze GTP through nucleophilic attack by a water molecule	259
Selected readings	220	The G_{β} subunit has a seven-blade propeller fold, built up from seven WD repeat units	261
12. Membrane Proteins	223	The GTPase domain of G_{α} binds to G_{β} in the heterotrimeric $G_{\alpha\beta\gamma}$ complex	263
Membrane proteins are difficult to crystallize	224	Phosducin regulates light adaptation in retinal rods	265
Novel crystallization methods are being developed	224	Phosducin binding to $G_{\beta\gamma}$ blocks binding of G_{α}	265
Two-dimensional crystals of membrane proteins can be studied by electron microscopy	225	The human growth hormone induces dimerization of its cognate receptor	267
Bacteriorhodopsin contains seven transmembrane α helices	226	Dimerization of the growth hormone receptor is a sequential process	268
Bacteriorhodopsin is a light-driven proton pump	227	The growth hormone also binds to the prolactin receptor	269
Porins form transmembrane channels by β strands	228	Tyrosine kinase receptors are important enzyme-linked receptors	270
Porin channels are made by up and down β barrels	229	Small protein modules form adaptors for a signaling network	272
Each porin molecule has three channels	230	SH2 domains bind to phosphotyrosine-containing regions of target molecules	273
Ion channels combine ion selectivity with high levels of ion conductance	232	SH3 domains bind to proline-rich regions of target molecules	274
The K^{+} channel is a tetrameric molecule with one ion pore in the interface between the four subunits	232	Src tyrosine kinases comprise SH2 and SH3 domains in addition to a tyrosine kinase	275
The ion pore has a narrow ion selectivity filter	233	The two domains of the kinase in the inactive state are held in a closed conformation by assembly of the regulatory domains	277
The bacterial photosynthetic reaction center is built up from four different polypeptide chains and many pigments	234	Conclusion	278
The L, M, and H subunits have transmembrane α helices	236	Selected readings	280
The photosynthetic pigments are bound to the L and M subunits	237	14. Fibrous Proteins	283
Reaction centers convert light energy into electrical energy by electron flow through the membrane	239	Collagen is a superhelix formed by three parallel, very extended left-handed helices	284
Antenna pigment proteins assemble into multimeric light-harvesting particles	240	Coiled coils are frequently used to form oligomers of fibrous and globular proteins	286
Chlorophyll molecules form circular rings in the light-harvesting complex LH2	241	Amyloid fibrils are suggested to be built up from continuous β sheet helices	288
The reaction center is surrounded by a ring of 16 antenna proteins of the light-harvesting complex LH1	242	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	Complex spherical viruses have more than one polypeptide chain in the asymmetric unit	329
Time-resolved x-ray diffraction of frog muscle confirmed movement of the cross-bridges	292	Structural versatility gives quasi-equivalent packing in T = 3 plant viruses	331
Structures of actin and myosin have been determined	293	The protein subunits recognize specific parts of the RNA inside the shell	332
The structure of myosin supports the swinging cross-bridge hypothesis	295	The protein capsid of picornaviruses contains four polypeptide chains	333
The role of ATP in muscular contraction has parallels to the role of GTP in G-protein activation	296	There are four different structural proteins in picornaviruses	334
Conclusion	297	The arrangement of subunits in the shell of picornaviruses is similar to that of T = 3 plant viruses	334
Selected readings	298	The coat proteins of many different spherical plant and animal viruses have similar jelly roll barrel structures, indicating an evolutionary relationship	335
15. Recognition of Foreign Molecules by the Immune System	299	Drugs against the common cold may be designed from the structure of rhinovirus	337
The polypeptide chains of antibodies are divided into domains	300	Bacteriophage MS2 has a different subunit structure	339
Antibody diversity is generated by several different mechanisms	302	A dimer of MS2 subunits recognizes an RNA packaging signal	339
All immunoglobulin domains have similar three-dimensional structures	303	The core protein of alphavirus has a chymotrypsin-like fold	340
The immunoglobulin fold is best described as two antiparallel β sheets packed tightly against each other	304	SV40 and polyomavirus shells are constructed from pentamers of the major coat protein with nonequivalent packing but largely equivalent interactions	341
The hypervariable regions are clustered in loop regions at one end of the variable domain	305	Conclusion	343
The antigen-binding site is formed by close association of the hypervariable regions from both heavy and light chains	306	Selected readings	344
The antigen-binding site binds haptens in crevices and protein antigens on large flat surfaces	308	17. Prediction, Engineering, and Design of Protein Structures	347
The CDR loops assume only a limited range of conformations, except for the heavy chain CDR3	311	Homologous proteins have similar structure and function	348
An IgG molecule has several degrees of conformational flexibility	312	Homologous proteins have conserved structural cores and variable loop regions	349
Structures of MHC molecules have provided insights into the molecular mechanisms of T-cell activation	312	Knowledge of secondary structure is necessary for prediction of tertiary structure	350
MHC molecules are composed of antigen-binding and immunoglobulin-like domains	313	Prediction methods for secondary structure benefit from multiple alignment of homologous proteins	351
Recognition of antigen is different in MHC molecules compared with immunoglobulins	314	Many different amino acid sequences give similar three-dimensional structures	352
Peptides are bound differently by class I and class II MHC molecules	315	Prediction of protein structure from sequence is an unsolved problem	352
T-cell receptors have variable and constant immunoglobulin domains and hypervariable regions	316	Threading methods can assign amino acid sequences to known three-dimensional folds	353
MHC-peptide complexes are the ligands for T-cell receptors	318	Proteins can be made more stable by engineering	354
Many cell-surface receptors contain immunoglobulin-like domains.	318	Disulfide bridges increase protein stability	355
Conclusion	320	Glycine and proline have opposite effects on stability	356
Selected readings	321	Stabilizing the dipoles of α helices increases stability	357
16. The Structure of Spherical Viruses	325	Mutants that fill cavities in hydrophobic cores do not stabilize T4 lysozyme	358
The protein shells of spherical viruses have icosahedral symmetry	327	Proteins can be engineered by combinatorial methods	358
The icosahedron has high symmetry	327	Phage display links the protein library to DNA	359
The simplest virus has a shell of 60 protein subunits	328	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	Building a model involves subjective interpretation of the data	381
Phage display of random peptide libraries identified agonists of erythropoietin receptor	364	Errors in the initial model are removed by refinement	383
DNA shuffling allows accelerated evolution of genes	365	Recent technological advances have greatly influenced protein crystallography	383
Protein structures can be designed from first principles	367	X-ray diffraction can be used to study the structure of fibers as well as crystals	384
A β structure has been converted to an α structure by changing only half of the sequence	368	The structure of biopolymers can be studied using fiber diffraction	386
Conclusion	370	NMR methods use the magnetic properties of atomic nuclei	387
Selected readings	371	Two-dimensional NMR spectra of proteins are interpreted by the method of sequential assignment	389
18. Determination of Protein Structures	373	Distance constraints are used to derive possible structures of a protein molecule	390
Several different techniques are used to study the structure of protein molecules	373	Biochemical studies and molecular structure give complementary functional information	391
Protein crystals are difficult to grow	374	Conclusion	391
X-ray sources are either monochromatic or polychromatic	376	Selected readings	392
X-ray data are recorded either on image plates or by electronic detectors	377	Protein Structure on the World Wide Web	393
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		