

CONTENTS

Preface to the first edition	xvii
Preface to the second edition	xx
Preface to the third edition	xxi
1 Introduction	1
Proteins in their biological context	2
The amino acids	4
Dogmas—central and peripheral	5
The relationship between amino acid sequence and protein structure is robust	6
Disorder in proteins	7
Regulation	10
The genetic code	11
With life so dependent on proteins, there is ample opportunity for things to go wrong	12
Genome sequences	15
Gene sequence determines amino acid sequence	16
Protein synthesis: the ribosome is the point of contact between genes and proteins—it is the fulcrum of genomics	17
Ribosomes were implicated in protein synthesis very early on	18
Structural studies of ribosomes by X-ray crystallography and electron microscopy	18
Protein stability, denaturation, aggregation, and turnover	19
Protein turnover	19
Description of protein structures	20
Primary structure	21
Secondary structure: helices and sheets are favourable conformations of the chain that recur in many proteins	21
Tertiary and quaternary structure	23
Folding patterns in native proteins—themes and variations	23
Modular proteins, and 'mixing and matching' as a mechanism of evolution	25
How do proteins develop new functions?	27
The study of proteins: in the laboratory, in the cell, in the computer	29
Spectroscopic methods of characterizing proteins in solution	30
Absorbance and fluorescence of proteins	32
Fluorescence is sensitive to the environment and dynamics of the chromophore	34
Fluorescence resonance energy transfer (FRET)	34
Circular dichroism	34
Protein expression patterns in space and time: proteomics	35
Subcellular localization	35
The transcriptome	36

DNA microarrays	37
Mass spectrometry	38
Computing in protein science	38
Computer-instrument partnerships in the laboratory	38
Simulations, including molecular dynamics	39
Bioinformatics	40
Introduction to databanks for protein science	40
Information-retrieval tools	41
Web access to the scientific literature	41
Useful websites	42
Recommended reading	42
Exercises and Problems	43
 2 Protein structure	 46
Introduction	47
Structures of the amino acids	47
Protein conformation	50
Conformational angles and the Sasisekharan–Ramakrishnan–Ramachandran plot	51
Sidechain conformation	53
Rotamer libraries	53
Stabilization of the native state	54
Conformational change	59
Protein folding patterns	60
Supersecondary structures	60
An album of small structures	62
Comparison of the folding patterns of acylphosphatase and a fungal toxin	64
Classification of protein structures	67
Databanks of protein structure classifications	68
SCOP	68
SCOP2	70
CATH	71
The DALI Database	72
A survey of protein structures and functions	73
Fibrous proteins	73
Enzymes—proteins that catalyse chemical reactions	76
Antibodies	77
Inhibitors	77
Carrier proteins	77
Membrane proteins	78
Receptors	80
Regulatory proteins	81
Motor proteins	81

Control of protein activity	81
Regulation of tyrosine hydroxylase illustrates several control mechanisms common to many proteins	84
Control cascades	84
Recommended reading	85
Exercises and Problems	85
3 Protein structure determination	92
Introduction	93
X-ray crystallography	94
X-ray structure determination	95
X-ray crystallography of proteins	96
Interpretation of the electron density: model building and improvement	100
The endgame—refinement	102
How accurate are the structures?	102
X-ray crystallography—the theoretical background	104
Nuclear magnetic resonance spectroscopy in structural biology	110
NMR spectra of proteins	111
Measurement of NMR spectra	113
Protein structure determination by NMR	113
Assignment of the spectrum	114
Transverse relaxation optimized spectroscopy	117
From the data to the structure	117
Solid-state NMR: magic angle spinning	118
Near atomic-resolution low-temperature electron microscopy (cryo-EM)	119
Octameric pyruvate-ferredoxin oxidoreductase from <i>Desulfovibrio vulgaris</i> Hildenborough	120
Conformational change in activation of human integrin $\alpha V\beta 3$	120
Trajectories of conformational change	124
The elastic network model accounts for conformational change in <i>Mycobacterium tuberculosis</i> thioredoxin reductase	125
The relationship between structure determinations of isolated proteins, and protein structure and function <i>in vivo</i>	127
Protein structure prediction and modelling	127
<i>A priori</i> methods of protein structure prediction	128
Empirical, or 'knowledge-based', methods of protein structure prediction	129
Secondary structure prediction	131
Homology modelling	133
Fold recognition	135
Antibody modeling	137
Prediction of special categories of structures	139
Conformational energy calculations and molecular dynamics	140
ROSETTA	142

Protein structure prediction from contact maps derived from correlated mutations in multiple sequence alignments	144
Critical Assessment of Structure Prediction (CASP)	146
CAPRI	153
Recommended reading	153
Exercises and Problems	154
 4 Bioinformatics of protein sequence and structure	 160
Introduction	160
Databases and information retrieval	161
Amino acid sequence databases	162
Protein databases at the U.S. National Center for Biotechnology Information	163
Specialized, or 'boutique', databases	164
Nucleic acid sequence databases	165
Genome databases and genome browsers	166
Ensembl	166
Expression and proteomics databases	166
Databases of macromolecular structure	168
Organization of wwPDB entries	169
Retrieval of sequences and structures	170
Retrieval of amino acid sequences by keyword	171
The Protein Information Resource (PIR) and associated databases	171
Retrieval of structures by keyword	172
Probing databanks with sequence information	173
Sequence alignment	174
The dotplot	175
Dotplots and alignments	176
BLAST and PSI-BLAST	177
Significance of alignments	179
Multiple sequence alignment	181
A multiple sequence alignment of thioredoxins shows the importance of conservation patterns	182
Analysis of structures	184
Superposition of structures	184
Structural alignment	185
Multiple structure alignment	187
Database searching for structures or fragments	187
Databases of protein families	188
Classifications of protein structures	189
Classification and assignment of protein function	189

The Enzyme Commission	189
The Gene Ontology™ Consortium protein function classification	190
The ENZYME database and PROSITE	192
Databases of metabolic networks	193
Recommended reading	194
Exercises and Problems	195
5 Proteins as catalysts: enzyme structure, kinetics, and mechanism	197
Introduction	197
What are the crucial features of enzymes?	198
Reaction rates and transition states	201
The activated complex	203
Measurement of reaction rates	204
Slow the reaction down	205
Fast methods of data collection	205
Active sites	206
Cofactors	206
Protein–ligand binding equilibria	207
The Scatchard plot	207
Catalysis by enzymes	208
Enzyme kinetics	209
Derivation of K_M and V_{max} from rate data	210
Measures of effectiveness of enzymes	211
Inhibitors	212
Irreversible inhibitor binding	212
Multisubstrate reactions	213
Enzyme mechanisms	214
The mechanism of action of thymidylate synthase	216
Computational approaches to enzyme mechanisms	218
The mechanism of action of chymotrypsin	220
The evidence from kinetics	221
The evidence from crystallography	221
Blood coagulation	222
Thrombosis	222
Serpins: serine proteinase inhibitors–conformational disease	226
Several conformational states of serpins are known	227
Mechanism of proteinase inhibition by serpins	228
Evolutionary divergence of enzymes	229
The mechanism of action of malate and lactate dehydrogenases	229

Enolase, mandelate racemase, and muconate lactonizing enzyme catalyse different reactions but have related mechanisms	230
The structure and mechanism of <i>E. coli</i> topoisomerase III	231
Motor proteins	233
The sliding filament mechanism of muscle contraction	233
ATP synthase	234
Membrane transport	238
Specificity of the potassium channel from <i>Streptomyces lividans</i> —room to swing a cation?	238
Allosteric regulation of protein activity	239
The allosteric structural change of haemoglobin	242
Recommended reading	246
Exercises and Problems	246
6 Proteins with partners	252
Introduction	252
General properties of protein–protein interfaces	254
Burial of protein surface	254
The composition of the interface	254
Complementarity	254
Specific interactions at protein–protein interfaces	255
Phage M13 gene III protein and <i>E. coli</i> TolA	255
Multisubunit proteins	256
Diseases of protein aggregation	257
Amyloidoses	258
Alzheimer's disease	258
Prion diseases—spongiform encephalopathies	259
The immune system	260
Antibody structure	261
Antibody maturation	266
Catalytic antibodies—'abzymes'	267
Proteins of the major histocompatibility complex	268
T-cell receptors	273
Virus structures	274
Tomato bushy stunt virus	278
Bacteriophage HK97: protein chain-mail	278
Photosynthetic reaction centres	279
Protein–DNA interactions	280
Structural themes in protein–DNA binding and sequence recognition	281
Bacteriophage T7 DNA polymerase	282
Some protein–DNA complexes that regulate gene transcription	283
Recommended reading	287
Exercises and Problems	288

7	Evolution of protein structure and function	291
	Introduction	291
	Protein structure classification	294
	A case study: superpositions and alignments of pairs of proteins with increasingly more-distant relationships	296
	Structural relationships among homologous domains	298
	Changes in proteins during evolution give clues to the roles of residues at different positions	302
	To what constraints are pathways of protein evolution subject?	302
	Closed β -barrel structures	303
	The TIM barrel	303
	Evolution of the globins	307
	Mammalian globins	308
	What determines the globin folding pattern?	310
	Truncated globins	312
	Expansion of the globin family	313
	Classification of the globins	313
	Globin functions	316
	Phycocyanins and the globins	316
	Evolution of NAD-binding domains of dehydrogenases	318
	Comparison of NAD-binding domains of dehydrogenases	320
	The sequence motif G*G**G	323
	Structure and evolution of serine proteinases of the chymotrypsin family	324
	Structures of individual domains	324
	The domain/domain interface	326
	The specificity pocket	327
	The β -barrels in serine proteinase domains and the packing of residues in their interiors	328
	Evolution of visual pigments and related molecules	331
	Selection has tuned vertebrate opsins so that the absorption maximum varies with the light environment	335
	How do proteins evolve new functions?	338
	Pathways and limits in the divergence of sequence, structure and function	339
	Evolution of functional change in the enolase superfamily	342
	Protein evolution at the level of domain assembly	345
	Domain swapping is a general mechanism for forming an oligomer from a multidomain protein	345
	Directed evolution	346
	Directed evolution of subtilisin E	347
	Enhancement of thermal stability	347
	Activity in organic solvents	348
	Affinity selectivity by phage display	349
	Recommended reading	350
	Exercises and Problems	350

8	Protein folding and design	356
	Introduction	356
	Why is protein folding so fast?	357
	Thermodynamics—key concepts	358
	Entropy	359
	Spontaneity and equilibrium	359
	Kinetics	360
	Thermodynamics of protein folding	360
	Thermodynamics of mutated proteins	361
	Experimental characterization of events in protein folding	362
	The molten globule	363
	Folding funnels	364
	The effect of denaturants on rates of folding and unfolding: chevron plots	365
	The kinetics of folding of mutated proteins gives clues to the structure of the transition state for folding	365
	Comparison of folding pathways of a natural protein and a circular permutant	366
	Relationship between native structure and folding	369
	The hierarchical model of protein folding	371
	How fast could a protein fold?	372
	Protein misfolding and the GroEL–GroES chaperone protein	373
	The GroEL–GroES conformational change	375
	Protein engineering	376
	Protein design	376
	<i>ab initio</i> design of a hyperstable variant of Streptococcal protein G, β 1 domain	376
	Expanding and contracting the genetic code	378
	Expansion of the genetic code	378
	Contraction of the genetic code	381
	Understanding the contents and layout of the common genetic code	382
	Recommended reading	382
	Exercises and Problems	383
9	Proteomics and systems biology	387
	Introduction	388
	Separation and analysis of proteins	389
	Polyacrylamide gel electrophoresis	389
	Two-dimensional polyacrylamide gel electrophoresis	390
	Difference gel electrophoresis	390

Mass spectrometry	393
Identification of components of a complex mixture	393
Protein sequencing by mass spectrometry	395
Quantitative analysis of relative abundance	395
Measuring deuterium exchange in proteins	398
'Ome, 'ome, on the range—environmental genomics and proteomics	398
Metagenomics	398
Metaproteomics	398
Dynamic proteomics of the response to cadmium challenge	399
Microarrays	401
Microarray data are semiquantitative	401
Applications of DNA microarrays	403
Analysis of microarray data	404
Expression patterns in different physiological states	406
Expression pattern changes in development: the life cycle of <i>Drosophila melanogaster</i>	406
RNAseq	408
RNAseq v. microarrays	409
Systems biology	411
Two parallel networks: physical and logical	411
Networks and graphs	412
Robustness and redundancy	413
Connectivity in networks	414
Dynamics, stability, and robustness	416
Protein complexes and aggregates	417
Protein interaction networks	417
Regulatory networks	421
Structures of regulatory networks	422
Structural biology of regulatory networks	423
Gene regulation	424
The transcriptional regulatory network of <i>E. coli</i>	424
Regulation of the lactose operon in <i>E. coli</i>	427
The genetic regulatory network of <i>Saccharomyces cerevisiae</i>	429
Adaptability of the yeast regulatory network	430
Recommended reading	433
Exercises and Problems	434
Epilogue	437
List of Abbreviations	438
Glossary	440
Index	453