

CONTENTS

Preface to the first edition	xvii
Preface to the second edition	xx
1 Introduction	1
Proteins in their biological context	2
The amino acids	3
Dogmas—central and peripheral	5
The relationship between amino acid sequence and protein structure is robust	6
Regulation	7
The genetic code	7
With life so dependent on proteins, there is ample opportunity for things to go wrong	8
Genome sequences	11
Gene sequence determines amino acid sequence	12
Protein synthesis: the ribosome is the point of contact between genes and proteins—it is the fulcrum of genomics	14
Ribosomes were implicated in protein synthesis very early on	14
Structural studies of ribosomes by X-ray crystallography and electron microscopy	15
Protein stability, denaturation, aggregation, and turnover	15
Protein turnover	16
Description of protein structures	17
Primary structure	17
Secondary structure: helices and sheets are favourable conformations of the chain that recur in many proteins	17
Tertiary and quaternary structure	19
Folding patterns in native proteins—themes and variations	19
Modular proteins, and ‘mixing and matching’ as a mechanism of evolution	22
How do proteins develop new functions?	25
The study of proteins: in the laboratory, in the cell, in the computer	26
Spectroscopic methods of characterizing proteins in solution	26
Absorbance and fluorescence of proteins	29
Fluorescence is sensitive to the environment and dynamics of the chromophore	30
Fluorescence resonance energy transfer (FRET)	30
Circular dichroism	30
Protein expression patterns in space and time: proteomics	32
Subcellular localization	32
DNA microarrays	33
Mass spectrometry	34
Computing in protein science	34
Computer-instrumental partnerships in the laboratory	34

Simulations, including molecular dynamics	35
Bioinformatics	36
Databanks for protein science	36
Information-retrieval tools	37
Web access to the scientific literature	38
Useful websites	38
Recommended reading	39
Exercises, Problems, and Weblems	39
2 Protein structure	41
Introduction	42
Structures of the amino acids	42
Protein conformation	45
Conformational angles and the Sasisekharan–Ramakrishnan–Ramachandran plot	45
Sidechain conformation	47
Rotamer libraries	48
Stabilization of the native state	48
Protein folding patterns	52
Supersecondary structures	54
An album of small structures	55
Comparison of the folding patterns of acylphosphatase and a fungal toxin	60
Classification of protein structures	63
Databanks of protein structure classifications	63
SCOP	64
CATH	64
The DALI Database	68
A survey of protein structures and functions	68
Fibrous proteins	68
Enzymes—proteins that catalyse chemical reactions	72
Antibodies	74
Inhibitors	74
Carrier proteins	76
Membrane proteins	76
Receptors	78
Regulatory proteins	79
Motor proteins	79
Conformational change	80
Control of protein activity	81
Regulation of tyrosine hydroxylase illustrates several control mechanisms common to many proteins	82
Control cascades	82
Recommended reading	85
Exercises, Problems, and Weblems	85

3 Protein structure determination	95
Introduction	96
X-ray crystallography	97
X-ray structure determination	98
X-ray crystallography of proteins	99
Solving the phase problem in protein crystallography	104
Interpretation of the electron density: model building and improvement	107
The endgame—refinement	109
How accurate are the structures?	109
Nuclear magnetic resonance spectroscopy in structural biology	111
NMR spectra of proteins	113
Measurement of NMR spectra	114
Protein structure determination by NMR	114
Assignment of the spectrum	115
Transverse relaxation optimized spectroscopy	118
From the data to the structure	118
Low-temperature electron microscopy (cryo-EM)	119
Octameric pyruvate-ferredoxin oxidoreductase from <i>Desulfovibrio vulgaris</i> Hildenborough	119
Conformational change in activation of human integrin $\alpha V\beta 3$	120
The relationship between structure determinations of isolated proteins, and protein structure and function <i>in vivo</i>	121
Protein structure prediction and modelling	121
<i>A priori</i> methods of protein structure prediction	123
Empirical, or 'knowledge-based', methods of protein structure prediction	124
Critical assessment of structure prediction	126
Secondary structure prediction	129
Homology modelling	130
Fold recognition	132
Prediction of special categories of structures	134
Conformational energy calculations and molecular dynamics	137
ROSETTA	139
LINUS	140
Conformational change	141
The elastic network model accounts for conformational change in <i>Mycobacterium tuberculosis</i> thioredoxin reductase	143
Recommended reading	144
Exercises, Problems, and Weblems	145
4 Bioinformatics of protein sequence and structure	149
Introduction	149
Databases and information retrieval	150
Amino acid sequence databases	151
Protein databases at the U.S. National Center for Biotechnology Information	152

Specialized, or 'boutique', databases	153
Nucleic acid sequence databases	153
Genome databases and genome browsers	154
Ensembl	154
Expression and proteomics databases	155
Databases of macromolecular structure	156
Organization of wwPDB entries	156
Retrieval of sequences and structures	158
Retrieval of amino acid sequences by keyword	158
The Sequence Retrieval System (SRS)	159
The Protein Information Resource (PIR) and associated databases	159
Retrieval of structures by keyword	160
Probing databanks with sequence information	161
Sequence alignment	162
The dotplot	163
Dotplots and alignments	164
BLAST and PSI-BLAST	165
Significance of alignments	168
Multiple sequence alignment	170
A multiple sequence alignment of thioredoxins shows the importance of conservation patterns	170
Analysis of structures	173
Superposition of structures	173
Structural alignment	174
Multiple structure alignment	175
Database searching for structures or fragments	175
Databases of protein families	177
Classifications of protein structures	177
Classification and assignment of protein function	178
The Enzyme Commission	178
The Gene Ontology™ Consortium protein function classification	178
The ENZYME database and PROSITE	181
Databases of metabolic networks	181
Recommended reading	183
Exercises, Problems and Weblems	183
5 Proteins as catalysts: enzyme structure, kinetics, and mechanism	187
Introduction	188
What are the crucial features of enzymes?	188
Reaction rates and transition states	191
The activated complex	193

Measurement of reaction rates	194
Slow the reaction down	194
Fast methods of data collection	195
Active sites	195
Cofactors	196
Protein–ligand binding equilibria	196
The Scatchard plot	197
Catalysis by enzymes	198
Enzyme kinetics	198
Derivation of K_M and V_{max} from rate data	200
Measures of effectiveness of enzymes	201
Inhibitors	201
Irreversible inhibitors	202
Multisubstrate reactions	203
Enzyme mechanisms	204
The mechanism of action of thymidylate synthase	205
Computational approaches to enzyme mechanisms	208
The mechanism of action of chymotrypsin	210
The evidence from kinetics	211
The evidence from crystallography	211
Blood coagulation	212
Thrombosis	212
Serpins: serine proteinase inhibitors—conformational disease	216
Several conformational states of serpins are known	217
Mechanism of proteinase inhibition by serpins	219
Evolutionary divergence of enzymes	219
The mechanism of action of malate and lactate dehydrogenases	219
Enolase, mandelate racemase, and muconate lactonizing enzyme catalyse different reactions but have related mechanisms	220
Enzymes that process DNA and RNA	221
The structure and mechanism of <i>E. coli</i> topoisomerase III	221
Motor proteins	224
The sliding filament mechanism of muscle contraction	224
ATPase	224
Membrane transport	227
Specificity of the potassium channel from <i>Streptomyces lividans</i> —room to swing a cation?	228
Allosteric regulation of protein activity	228
The allosteric change of haemoglobin	231
Recommended reading	236
Exercises, Problems, and Weblems	236

6	Proteins with partners	241
	Introduction	241
	General properties of protein–protein interfaces	243
	Burial of protein surface	243
	The composition of the interface	243
	Complementarity	243
	Specific interactions at protein–protein interfaces	244
	Phage M13 gene III protein and <i>E. coli</i> TolA	244
	Multisubunit proteins	245
	Diseases of protein aggregation	246
	Amyloidoses	246
	Alzheimer disease	247
	Prion diseases—spongiform encephalopathies	247
	The immune system	249
	Antibody structure	249
	Antibody maturation	256
	Catalytic antibodies—‘abzymes’	256
	Proteins of the major histocompatibility complex	258
	T-cell receptors	263
	Virus structures	264
	Tomato bushy stunt virus	269
	Bacteriophage HK97: protein chain-mail	269
	Photosynthetic reaction centres	270
	Protein–DNA interactions	271
	Structural themes in protein–DNA binding and sequence recognition	272
	Bacteriophage T7 DNA polymerase	274
	Some protein–DNA complexes that regulate gene transcription	274
	Recommended reading	280
	Exercises, Problems, and Weblems	281
7	Evolution of protein structure and function	285
	Introduction	285
	Protein structure classification	288
	Superpositions and alignments of pairs of proteins with increasingly more distant relationships	291
	Structural relationships among homologous domains	294
	Changes in proteins during evolution give clues to the roles of residues at different positions	298
	To what constraints are pathways of protein evolution subject?	298
	Closed β -barrel structures	299
	The TIM barrel	300
	Evolution of the globins	304
	Mammalian globins	305

What determines the globin folding pattern?	308
Truncated globins	310
Phycocyanins and the globins	311
Evolution of NAD-binding domains of dehydrogenases	313
Comparison of NAD-binding domains of dehydrogenases	315
The sequence motif G*G**G	320
Structure and evolution of serine proteinases of the chymotrypsin family	321
Structures of individual domains	322
The domain/domain interface	324
The specificity pocket	324
The β -barrels in serine proteinase domains and the packing of residues in their interiors	324
Evolution of visual pigments and related molecules	330
Selection has tuned vertebrate opsins so that the absorption maximum matches the light environment	333
How do proteins evolve new functions?	335
Protein evolution at the level of domain assembly	337
Domain swapping is a general mechanism for forming an oligomer from a multidomain protein	337
Directed evolution	338
Directed evolution of subtilisin E	339
Enhancement of thermal stability	339
Recommended reading	341
Exercises, Problems, and Weblems	341
8 Protein folding and design	347
Introduction	347
Why is protein folding so fast?	348
Thermodynamics—key concepts	349
Entropy	350
Spontaneity and equilibrium	350
Kinetics	351
Thermodynamics of protein folding	351
Thermodynamics of mutated proteins	352
Experimental characterization of events in protein folding	353
The molten globule	354
Folding funnels	355
The effect of denaturants on rates of folding and unfolding: chevron plots	356
The kinetics of folding of mutated proteins gives clues to the structure of the transition state for folding	356
Comparison of folding pathways of a natural protein and a circular permutant	357

Relationship between native structure and folding	360
The hierarchical model of protein folding	362
How fast could a protein fold?	363
Protein misfolding and the GroEL–GroES chaperone protein	364
The GroEL–GroES conformational change	365
Protein engineering	368
Protein design	368
<i>ab initio</i> design of a hyperstable variant of Streptococcal protein G, β 1 domain	368
Expanding and contracting the genetic code	370
Expansion of the genetic code	370
Contraction of the genetic code	371
Understanding the contents and layout of the common genetic code	375
Recommended reading	375
Exercises, Problems, and Weblems	375
9 Proteomics and systems biology	380
Introduction	381
Separation and analysis of proteins	382
Polyacrylamide gel electrophoresis	382
Two-dimensional polyacrylamide gel electrophoresis	383
Difference gel electrophoresis	383
Mass spectrometry	386
Identification of components of a complex mixture	386
Protein sequencing by mass spectrometry	388
Quantitative analysis of relative abundance	388
Measuring deuterium exchange in proteins	391
‘Ome, ‘ome, on the range—environmental genomics and proteomics	391
Metagenomics	391
Metaproteomics	391
Dynamic proteomics of the response to cadmium challenge	393
Microarrays	394
Microarray data are semiquantitative	396
Will high-throughput sequencing replace microarrays?	396
Applications of DNA microarrays	396
Analysis of microarray data	397
Expression patterns in different physiological states	400
Expression pattern changes in development: the life cycle of <i>Drosophila melanogaster</i>	400
Systems biology	402
Two parallel networks: physical and logical	403

Networks and graphs	404
Robustness and redundancy	405
Connectivity in networks	405
Dynamics, stability, and robustness	407
Protein complexes and aggregates	408
Protein interaction networks	409
Regulatory networks	413
Structures of regulatory networks	414
Structural biology of regulatory networks	414
Gene regulation	415
The transcriptional regulatory network of <i>E. coli</i>	415
Regulation of the lactose operon in <i>E. coli</i>	418
The genetic regulatory network of <i>Saccharomyces cerevisiae</i>	420
Adaptability of the yeast regulatory network	422
Recommended reading	425
Exercises, Problems, and Weblems	425
Epilogue	429
List of Abbreviations	430
Glossary	432
Index	443