
Table of Contents

Chapter 1	The Data: Storage and Retrieval.....	1
Glossary		1
1.1	Basic Principles.....	2
1.2	The Data.....	3
1.2.1	The Postgenomic Era.....	3
1.2.2	Nucleic Acid Data Banks.....	6
1.2.3	Protein Sequence Data Banks.....	7
1.2.4	Protein Structure Databases.....	14
1.2.5	Protein Interaction Databases	16
1.2.6	Derived Data Banks	17
1.2.7	Integration of Databases	18
1.3	Data Quality	18
1.4	Data Representation	18
1.4.1	Protein Architecture	20
References.....		24
Problems		26
Chapter 2	Genome Sequence Analysis.....	29
Glossary		29
2.1	Basic Concepts.....	30
2.2	Genome Sequencing	31
2.3	Finding the Genes	32
2.4	Statistical Methods to Search for Genes	34
2.4.1	Site-Specific Scoring Matrices.....	34
2.4.2	Artificial Neural Networks.....	37
2.4.3	Markov Models and Hidden Markov Models.....	40
2.4.4	Levels of Reliability.....	45
2.5	Comparative Genomics.....	47
2.6	A Virtual Window on Genomes: The World Wide Web	48
References.....		49
Problems		50
Chapter 3	Protein Evolution	53
Glossary		53
3.1	Basic Concepts.....	53
3.2	Molecular Evolution.....	54

3.3	How to Align Two Similar Sequences	55
3.4	Similarity Matrices.....	58
3.4.1	PAM Matrices	59
3.4.2	BLOSUM Matrices	60
3.5	Penalties for Insertions and Deletions	61
3.6	The Alignment Algorithm.....	61
3.7	Multiple Alignments	66
3.7.1	How to Add Sequences to a Pre-Existing Alignment	70
3.8	Phylogenetic Trees	71
	References.....	73
	Problems	74
Chapter 4	Similarity Searches in Databases.....	77
	Glossary	77
4.1	Basic Principles.....	77
4.2	The Methods	78
4.2.1	FASTA.....	79
4.2.2	BLAST	83
4.2.3	Profile Searches.....	86
4.2.4	PSI-BLAST	87
	References.....	89
	Problems	90
Chapter 5	Amino Acid Sequence Analysis	93
	Glossary	93
5.1	Basic Principles.....	93
5.1.1	Is It Really an Orphan Sequence?.....	94
5.2	Search for Sequence Patterns	95
5.3	Feature Extraction.....	96
5.4	Secondary Structure: Part One	96
	References.....	98
	Problems	99
Chapter 6	Prediction of the Three-Dimensional Structure of a Protein.....	103
	Glossary	103
6.1	Basic Principles.....	103
6.2	The CASP Experiment.....	105
6.3	Secondary Structure Prediction: Part Two.....	107
6.4	Long-Range Contact Prediction.....	109
6.5	Predicting Molecular Complexes: Docking Methods	110
	References.....	111
	Problems	113

Chapter 7	Homology Modeling	115
Glossary		115
7.1	Basic Principles	115
7.2	The Steps of Comparative Modeling	117
7.2.1	Template Selection	118
7.2.2	Sequence Alignment	119
7.2.3	Loops	121
7.2.4	Side Chain Modeling	124
7.2.5	Model Optimization	125
7.3	Accuracy of Homology Models	126
7.4	Manual versus Automatic Models	126
7.5	Practical Notes	126
7.6	Summing Up...	129
References		131
Problems		134
Chapter 8	Fold Recognition Methods	137
Glossary		137
8.1	Basic Principles	137
8.2	Profile-Based Methods	138
8.3	Threading Methods	140
8.4	The Fold Library	142
8.5	How Well Do These Methods Work?	142
References		143
Problems		143
Chapter 9	New Fold Modeling	147
Glossary		147
9.1	Basic Principles	147
9.2	Estimating the Energy of a Protein Conformation	148
9.3	Energy Minimization	150
9.4	Molecular Dynamics	151
9.4.1	The Monte Carlo Method	153
9.4.2	Genetic Algorithms	154
9.4.3	Combined Methodologies	156
References		156
Problems		157
Chapter 10	The “Omics” Universe	159
10.1	Basic Principles	159
10.2	Transcriptomics	159

10.3 Proteomics.....	162
10.4 Interactomics	165
10.5 Structural Genomics.....	166
10.6 Pharmacogenomics.....	166
10.7 But This Is Not All... ..	168
Useful Web Sites	169
Index	171