

Contents

<i>Preface to the Second Edition</i>	vii
<i>Preface to the First Edition</i>	ix
1. Bioinformatics : An Overview	1–17
1.1. Introduction, 1	
1.2. Objectives of bioinformatics, 3	
1.3. What kind of data is used, 4	
1.4. Multiplicity of data and data mining, 5	
1.5. Major bioinformatics databases, 5	
1.6. Data integration, 11	
1.7. Data analysis, 12	
1.8. Careers in bioinformatics, 12	
1.9. Reference list of the major bioinformatics databases and tools, 14	
2. Molecular Biology and Bioinformatics	18–26
2.1. What is molecular biology, 18	
2.2. Systems approach in biology, 18	
2.3. Central dogma of molecular biology, 18	
2.4. Important definitions related to central dogma, 21	
2.5. Problems in molecular approach and the bioinformatics approach, 23	
2.6. Overview of bioinformatics applications, 23	
3. The Information Molecules and Information Flow	27–42
3.1. Introduction, 27	
3.2. Basic components, 27	
3.3. Basic chemistry of nucleic acids, 28	

3.4.	Structure of DNA, 29	
3.5.	Structure of RNA, 31	
3.6.	DNA replication is semi-conservative, 32	
3.7.	Denaturation and renaturation of DNA, 33	
3.8.	Genes - the functional elements in DNA, 34	
3.9.	Organisation of genes in eukaryotic chromosomes, 36	
3.10.	Structure of bacterial chromosome, 37	
3.11.	Analysing DNA, 37	
3.12.	Cloning methodology, 38	
3.13.	DNA sequencing and polymerase chain reaction (PCR), 39	
4.	Proteins—Their Structural Profiles and Properties	43–54
4.1.	Introduction, 43	
4.2.	Amino acids, 43	
4.3.	Protein structure, 46	
4.4.	Secondary structure elements, 48	
4.5.	Tertiary structure, 50	
4.6.	Quaternary structure, 51	
4.7.	Protein folding, 51	
4.8.	Protein function, 51	
4.9.	Proteins - purification and characterisation, 52	
5.	Using the Linux Operating System	55–73
5.1.	Introduction to linux, 55	
5.2.	The basics of linux system, 56	
5.3.	Using linux file system and directories, 58	
5.4.	Text processing, 63	
5.5.	Writing shell programs, 68	
6.	Programming with Perl	74–108
6.1.	Introduction to Perl, 74	
6.2.	Programming in Perl, 74	
6.3.	Illustrations of programming in Perl, 86	
6.4.	Operations with associative arrays (Hashes), 88	
6.5.	File input and output, 91	
6.6.	Perl applications for bioinformatics, 94	
6.7.	Perl application for bioinformatics - Bioperl, 106	
7.	Relational Databases for Biological Information	109–123
7.1.	Introduction, 109	
7.2.	Types of databases, 110	

8. Object-Oriented Databases	124–131
8.1. Introduction, <i>124</i>	
8.2. Object oriented databases, <i>125</i>	
8.3. Introduction to the Java clients, <i>128</i>	
8.4. Common object request broker architecture (CORBA), <i>130</i>	
8.5. Some projects, <i>131</i>	
9. Managing Biological Databanks	132–147
9.1. Introduction, <i>132</i>	
9.2. Submission of data, <i>132</i>	
9.3. Curation of databases, <i>134</i>	
9.4. Establishing databases on networks, <i>134</i>	
9.5. Integration of databases, <i>137</i>	
9.6. Mining of databases, <i>143</i>	
9.7. Management of workflow, <i>146</i>	
10. Alignment of Pairs of Sequences	148–172
10.1. Introduction to sequence analysis, <i>148</i>	
10.2. Sequence Analysis of Biological Data, <i>149</i>	
10.3. Models for sequence analysis and their biological motivation, <i>150</i>	
10.4. Methods of alignment, <i>152</i>	
10.5. Applications of dot matrices, <i>157</i>	
10.6. Methods for optimal alignments, <i>160</i>	
10.7. Using gap penalties and scoring matrices, <i>165</i>	
10.8. Sensitivity and specificity, <i>169</i>	
10.9. Illustrative examples, <i>170</i>	
11. Tools for Sequence Alignment	173–190
11.1. Introduction, <i>173</i>	
11.2. Fasta, <i>173</i>	
11.3. Blast, <i>179</i>	
11.4. Filtering and gapped blast, <i>183</i>	
11.5. PSI-blast, <i>187</i>	
11.6. Comparison of running time for various programs, <i>190</i>	
12. Alignment of Multiple Sequences	191–205
12.1. Introduction, <i>191</i>	
12.2. Tools for MSA, <i>193</i>	
12.3. Considerations in conducting MSA, <i>199</i>	
12.4. Applications of multiple alignment, <i>201</i>	
12.5. Viewing MSA, <i>204</i>	
12.6. Sequence detection efficiency measures : Sensitivity and specificity, <i>204</i>	

13. Phylogenetic Analysis	206–222
13.1. Introduction, 206	
13.2. Concept of trees, 206	
13.3. Phylogenetic trees and multiple alignments, 207	
13.4. Distance matrix methods (MD), 209	
13.5. Character based methods, 212	
13.6. Methods of evaluating phylogenies, 215	
13.7. Summary of the phylogenetic methods, 216	
13.8. Steps in constructing alignments and phylogenies, 217	
13.9. Considerations in choice of the method, 217	
13.10. Working with phylogenetic trees - an illustration, 219	
14. Gene Prediction Methods	223–234
14.1. Introduction, 223	
14.2. Using patterns to predict genes, 224	
14.3. Methods of gene prediction, 225	
14.4. Gene prediction tools, 232	
14.5. Summary of tools for DNA/RNA structure and function analysis,	
15. Understanding and Using Biological Databases	235–247
15.1. Introduction, 235	
15.2. Overview of RNA secondary structure, 236	
15.3. Overview of RNA tertiary structure, 237	
15.4. Assumptions in RNA structure prediction, 237	
15.5. Methods of RNA structure prediction, 238	
16. Visualization and Prediction of Protein Structure	248–285
16.1. Protein structure overview, 248	
16.2. Different structural proteins, 250	
16.3. Protein structure databases and visualization tools, 252	
16.4. Protein classification, 256	
16.5. Protein structure prediction, 257	
16.6. Methods of structure prediction for known folds, 261	
16.7. Methods of structure prediction for unknown folds, 266	
16.8. Protein function prediction, 284	
16.9. Accuracy of prediction, 284	
17. Gene Mapping, Sequence Assembly and Gene Expression	286–308
17.1. Introduction, 286	
17.2. Gene mapping, 287	
17.3. Application of mapping, 290	

17.4. DNA sequencing, 293	
17.5. Algorithm for alignment of sequencing fragments, 298	
17.6. DNA microarrays, 301	
17.7. Microarray experiment design and data analysis, 304	
18. Proteomics.....	309–331
18.1. Introduction to proteomics, 309	
18.2. Proteome analysis, 311	
18.3. Tools for proteome analysis, 312	
18.4. Metabolic pathways, 321	
18.5. Genetic networks, 323	
19. Analyzing Metabolic Pathways	332–343
19.1. Introduction, 332	
19.2. Gene transcription networks, 332	
19.3. Metabolic pathways, 333	
19.4. Signalling network, 338	
19.5. Simulation of cellular metabolism, 340	
20. Methods of Statistical Analysis	344–364
20.1. Introduction, 344	
20.2. Fundamentals of probability and statistics, 345	
20.3. Applications of Statistical Tools, 355	
21. Problem Solving in Bioinformatics	365–393
21.1. Introduction, 365	
21.2. Genomic analysis for DNA sequences, 365	
21.3. Genomic analysis for protein sequences, 366	
21.4. Strategies and options for similarity search, 366	
21.5. Practical considerations in sequence analysis, 368	
21.6. Flowchart for protein structure prediction, 372	
21.7. Illustrations : Some problems and solutions, 373	
22. Problems for Self Assessment	394–404
Selected References	405–406
Glossary	407–418
Index	419–426