

Contents

Preface to the French edition	vii
Preface to the English edition	ix
1 Genome sequencing	1
1.1 Automatic sequencing	1
1.2 Sequencing strategies	4
1.3 Fragmentation strategies	8
1.4 Sequence assembly	12
1.5 Filling gaps	14
1.6 Obstacles to reconstruction	16
1.7 Utilizing a complementary 'large' clone library	18
1.8 The first large-scale sequencing project: The <i>Haemophilus influenzae</i> genome	19
1.9 cDNA and EST	20
2 Sequence comparisons	25
2.1 Introduction: Comparison as a sequence prediction method	25
2.2 A sample molecule: the human androsterone receptor	26
2.3 Sequence homologies – functional homologies	27
2.4 Comparison matrices	28
2.5 The problem of insertions and deletions	33
2.6 Optimal alignment: the dynamic programming method	34
2.7 Fast heuristic methods	38
2.8 Sensitivity, specificity, and confidence level	46
2.9 Multiple alignments	50
3 Comparative genomics	61
3.1 General properties of genomes	61
3.2 Genome comparisons	67
3.3 Gene evolution and phylogeny: applications to annotation	75
4 Genetic information and biological sequences	85
4.1 Introduction: Coding levels	85
4.2 Genes and the genetic code	85
4.3 Expression signals	87

4.4	Specific sites	91
4.5	Sites located on DNA	91
4.6	Sites present on RNA	96
4.7	Pattern detection methods	96
5	Statistics and sequences	107
5.1	Introduction	107
5.2	Nucleotide base and amino acid distribution	107
5.3	The biological basis of codon bias	112
5.4	Using statistical bias for prediction	113
5.5	Modeling DNA sequences	116
5.6	Complex models	120
5.7	Sequencing errors and hidden Markov models	123
5.8	Hidden Markov processes: a general sequence analysis tool	127
5.9	The search for genes – a difficult art	127
6	Structure prediction	131
6.1	The structure of RNA	131
6.2	Properties of the RNA molecule	132
6.3	Secondary RNA structures	134
6.4	Thermodynamic stability of RNA structures	138
6.5	Finding the most stable structure	144
6.6	Validation of predicted secondary structures	149
6.7	Using chemical and enzymatic probing to analyze folding	150
6.8	Long-distance interactions and three-dimensional structure prediction	152
6.9	Protein structure	155
6.10	Secondary structure prediction	158
6.11	Three-dimensional modeling based on homologous protein structure	161
6.12	Predicting folding	166
7	Transcriptome and proteome: macromolecular networks	169
7.1	Introduction	169
7.2	Post-genomic methods	170
7.3	Macromolecular networks	182
7.4	Topology of macromolecular networks	193
7.5	Modularity and dynamics of macromolecular networks	199
7.6	Inference of regulatory networks	206
8	Simulation of biological processes in the genome context	211
8.1	Types of simulations	213
8.2	Prediction and explanation	213
8.3	Simulation of molecular networks	215
8.4	Generic post-genomic simulators	226
	Index	233