

Contents

Preface

xv

I INTRODUCTION

Overview	1
Computationally Intense Problems: A Central Theme in Modern Biology	2
Building the Public Infrastructure	6
The Human Genome's Several Layers of Complexity	7
Toward Personalized Medicine	10
Illnesses are Polygenic	11
New Science, New Infrastructure	16
The Proactive Future of Information-Based Medicine	18

2 INTRODUCTION TO BIOINFORMATICS

Introduction	23
The Emergence of Bioinformatics	24
The Public Database Infrastructure	27
Building Database Infrastructure for Bioinformatics	30
Traditional Bioinformatic Tools and Algorithms	79
Summary	100

3 GENE STRUCTURE

Introduction	105
The Central Dogma of Molecular Biology	107
The Genetic Code	108
Structure and Content of the Genome	118
Computational Techniques for the Identification of Genomic Features	132
High-Throughput Gene Sequencing	141
Summary	145

4 COMPUTATIONAL TECHNIQUES FOR SEQUENCE ANALYSIS

Introduction	149
Hidden Markov Models	150
Perceptrons and Neural Networks	156
Pattern Discovery, Single Nucleotide Polymorphisms, and Haplotype Identification	170
Summary	180

5 TRANSCRIPTION

Introduction	183
The Transcriptome	186
Technologies for Transcriptional Profiling	192
Hierarchical Clustering	200
Summary	234

6 OVERVIEW OF THE PROTEOME AND THE PROTEIN TRANSLATION PROCESS

Introduction	237
Ribosomal Structure and the Protein Translation Process	238
Special Features of the Eukaryotic-Translation Process	246
Summary	255

7 PROTEIN STRUCTURE PREDICTION

Introduction	259
Overview of <i>Ab Initio</i> and Database-Driven Approaches	261
Overview of Protein Structure	263
Protein Structure Databases	276
<i>Ab Initio</i> Structure Prediction	288
Predicting Lead—Target Interactions	296
Summary	299

8 MEDICAL INFORMATICS AND INFORMATION-BASED MEDICINE

Introduction	307
The Continuous Evolution in Understanding that Leads to Genomic Medicine	308
Electronic Medical Records	314
Grid Computing and Medical Informatics	323
Modeling and Predicting Disease	333
Summary	341

9 NEW THEMES IN BIOINFORMATICS

Introduction	347
Overview of Parallel Computing and Workflow Distribution in Bioinformatics	348
Workflow Computing	355
High-Performance Computing and Systems Biology	359
The Delineation of Metabolic Pathways	361
Systems Biology	367
Summary	374

FURTHER READING	377
-----------------	-----

INDEX	379
-------	-----