

SECOND EDITION

# Bioinformatics Basics

Applications in Biological Science  
and Medicine

Edited by

Lukas K. Buehler  
Hooman H. Rashidi



Taylor & Francis  
Taylor & Francis Group

---

# Contents

---

<b>1</b>	<b>Biology and Information</b> .....	<b>1</b>
1.1	Bioinformatics—A Rapidly Maturing Science .....	1
1.1.1	From Genes to Proteins.....	2
1.1.2	Bioinformatics in the Public Domain.....	7
1.2	Computers in Biology and Medicine .....	12
1.2.1	Computational Tools.....	12
1.2.2	Limitations of Computational Tools .....	18
	References.....	23
1.3	The Virtual Doctor.....	23
1.3.1	Mapping the Human Brain .....	25
	References.....	30
1.4	Biological Macromolecules as Information Carriers.....	30
	References.....	34
1.5	Proteins: From Sequence to Structure to Function .....	34
1.5.1	Molecular Interaction in Protein Structures .....	35
1.5.1.1	The Peptide Bond .....	35
1.5.1.2	Characteristics of the Peptide Bond .....	37
1.5.1.3	The Hydrophobic Effect and How It Contributes to Protein Folding.....	39
1.5.1.4	Electrostatic Interactions.....	41
1.5.1.5	Hydrogen Bonding.....	41
1.5.1.6	Conformational Entropy.....	42
1.5.1.7	Van der Waals Interactions (Packing).....	42
1.5.1.8	Covalent Bonds (e.g., Disulfide Bridge).....	42
1.5.2	Protein Functions .....	42
1.5.2.1	Enzymes .....	42
1.5.2.2	Regulatory Proteins .....	43
1.5.2.3	Storage .....	43
1.5.2.4	Transportation .....	43
1.5.2.5	Signaling.....	43
1.5.2.6	Immunity.....	44
1.5.2.7	Structural.....	44
	References.....	44
1.6	DNA and RNA Structure .....	45
1.6.1	The DNA Double Helix.....	46
1.6.2	Genomic Size of DNA .....	48
	References.....	49
1.7	DNA Cloning and Sequencing.....	49
1.7.1	DNA Cloning .....	50

1.7.2	Transcriptional Profiling .....	51
1.7.3	Positional Cloning and Chromosome Mapping .....	52
1.7.4	Polymerase Chain Reaction (PCR) .....	54
1.7.5	Sequencing Technologies .....	55
	References.....	56
1.8	Genes, Taxonomy, and Evolution .....	56
	References.....	61
<b>2</b>	<b>Biological Databases</b> .....	<b>63</b>
2.1	Biological Database Organization.....	63
2.1.1	Database Content and Management.....	63
2.1.2	Data Submissions .....	64
2.1.3	The Growth of Public Databases .....	68
2.1.4	Data Retrieval .....	70
	References.....	74
2.1.5	Data Annotation and Database Connectivity.....	75
2.1.5.1	Annotation .....	76
2.1.5.2	Redundancy .....	78
2.2	Public Databases .....	79
2.2.1	National Center for Biotechnology Information (NCBI) .....	79
2.2.1.1	Who is Employed by NCBI?.....	81
2.2.1.2	What Kind of Research is Conducted at NCBI? .....	81
2.2.1.3	What Types of Databases are Supported by NCBI? .....	81
2.2.1.4	What Do We Mean by Redundancy? .....	82
2.2.1.5	What are Some of the Services Offered by NCBI? .....	84
	References.....	89
2.2.2	European Bioinformatics Institute (EBI) .....	90
2.2.2.1	Who is Employed by EBI? .....	90
2.2.2.2	What Kind of Research is Conducted at EBI? .....	90
2.2.2.3	What Are Some Of The Services Offered By EBI? .....	92
	References.....	101
2.2.3	Kyoto Encyclopedia of Genes and Genomes (KEGG).....	102
2.2.3.1	Classification of Biological Molecules .....	112
2.2.3.2	Cellular Processes at KEGG .....	114
	References.....	117
2.3	Database Mining Tools .....	118
2.3.1	Sequence Similarity Search Tools: BLAST and FASTA.....	118
2.3.1.1	Shared Characteristics in Both Sequence Alignment Tools.....	120
2.3.1.2	How are Sequence Alignments Useful?.....	121
2.3.1.3	Basic Local Alignment Search Tool (BLAST) .....	122
2.3.2	An Overview of Database Sequence Searching.....	129

References.....	131
2.3.3 Pattern Recognition Tools (Prosite).....	132
2.3.3.1 The Significance of Embedded Symbols within Each Signature and How to Read and Construct Signatures.....	133
References.....	134
2.3.4 Multiple Alignment and Phylogenetic Tree Analysis.....	134
References.....	137
<b>3 Genome Analysis.....</b>	<b>139</b>
3.1 The Genomic Organization of Genes.....	139
3.1.1 What are Genomes?.....	139
3.1.2 Mapping and Navigating Genomes.....	142
3.1.2.1 Genetic Linkage Maps.....	147
3.1.2.2 Physical Maps.....	149
3.1.2.3 From Sequence Maps to Gene Function Maps.....	154
References.....	154
3.1.2 The Genome Projects.....	155
3.1.2.1 How Many Genes are in a Genome?.....	158
References.....	164
3.1.3 The Human Genome.....	164
References.....	169
3.2 Comparative Genomics.....	170
3.2.1 Cluster of Orthologous Groups (COGs).....	170
3.2.2 Homologene at NCBI.....	176
References.....	182
3.2.2.1 Gene Order and Chromosome Rearrangements.....	182
3.2.2.2 MapViewer.....	183
References.....	185
3.3 Functional Genomics.....	185
3.3.1 The Transcriptome.....	185
3.4 Microarray and Bioarray Technology.....	190
3.4.1 Concept and Use.....	190
3.4.2 Summary of a Typical Experiment Using Microarray Technology.....	191
3.4.3 Microarray Bioinformatics.....	191
3.4.4 Image Processing.....	194
References.....	195
3.4.5 Data Annotation.....	195
References.....	196
3.4.6 Data Analysis.....	196
3.4.6.1 Experiment Design/Plan.....	197
3.4.6.2 Volume of Data.....	197

3.4.6.3	Dimensionality of Data.....	199
3.4.6.4	Quality of Data.....	199
3.4.7	Normalization.....	199
3.4.8	Statistical Analysis.....	199
3.4.9	Explorative Analysis.....	202
3.4.9.1	Aim of Clustering.....	202
3.4.9.2	Biological Interpretation of Clustering Results.....	202
3.4.9.3	Theory of Clustering.....	202
3.4.9.4	Clustering vs. Classification (Unsupervised vs. Supervised).....	203
3.4.10	Main Types of Clustering.....	203
3.4.10.1	Hierarchical Clustering.....	203
3.4.10.2	Nonhierarchical Clustering.....	204
3.4.10.3	Other Clustering or Classification Algorithms.....	206
3.4.10.4	Advice on Using Clustering.....	206
References	.....	207
3.4.11	Data Storage.....	208
References	.....	209
3.4.12	Data Mining.....	209
3.4.13	Protein Arrays.....	209
3.4.14	Concluding Remarks.....	211
3.5	Genomes as Gene Networks.....	211
References	.....	216
<b>4</b>	<b>Proteome Analysis.....</b>	<b>217</b>
4.1	Proteomics.....	217
4.1.1	What is a Proteome?.....	217
4.1.1.1	2-D Gels and Mass Spectrometry Tools.....	219
4.1.1.2	2-D PAGE at Expasy (Swiss Bioinformatics Institute).....	221
References	.....	225
4.2	Hydrodynamic Methods.....	226
4.2.1	Introduction.....	226
4.2.2	Analytical Ultracentrifugation.....	228
4.2.2.1	Experimental Setup and Instrumentation.....	228
4.2.2.2	Transport Processes in The AUC Cell.....	229
4.2.2.3	Analytical Ultracentrifuge (AUC) Experiments.....	234
4.2.3	Light Scattering.....	246
4.2.3.1	Experimental Setup and Instrumentation.....	246
4.2.3.2	Dynamic Light Scattering.....	247
4.2.3.3	Static Light Scattering.....	250
4.2.4	Global Analysis.....	252
4.2.5	Appendix.....	254
References	.....	255

4.3	Predictive Biology.....	256
4.3.1	Protein Structure Prediction.....	256
4.3.1.1	Structure Prediction Software.....	266
	References.....	269
4.3.2	Structural Genomics.....	269
	References.....	276
4.3.3	Rational Drug Design.....	276
	References.....	285
4.4	Systems Biology.....	285
4.4.1	Protein Interaction Networks.....	285
	References.....	290
4.4.2	Metabolic Reconstruction.....	290
	References.....	295
<b>5</b>	<b>The Bioinformatics Revolution in Medicine.....</b>	<b>297</b>
5.1	Genes and Diseases.....	297
5.1.1	From Molecules to Diseases.....	297
	References.....	303
5.1.2	Online Mendelian Inheritance in Man (OMIM).....	303
	References.....	308
5.1.3	Pharmacogenomics.....	308
5.2	Agricultural Genomics.....	309
5.2.1	Genetically Modified Organisms.....	312
	References.....	313
5.2.2	Biopharming.....	313
	References.....	316
	<b>Appendix A Glossary of Biological Terms.....</b>	<b>317</b>
	<b>Appendix B Bioinformatics Web Sites.....</b>	<b>323</b>
	<b>Index.....</b>	<b>325</b>