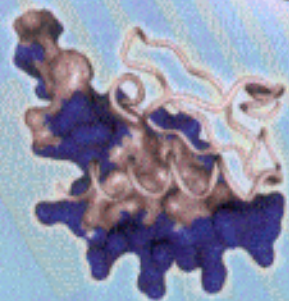


Instant Notes

Bioinformatics



D.R. Westhead, J.H. Parish and R.M. Twyman

CONTENTS

Abbreviations	vii
Preface	ix
Section A – Overview of bioinformatics	1
A1 The scope of bioinformatics	1
A2 Bioinformatics and the Internet	4
A3 Useful bioinformatics sites on the WWW	7
Section B – Data acquisition	9
B1 Sequencing DNA, RNA and proteins	9
B2 Determination of protein structure	15
B3 Gene and protein expression data	17
B4 Protein interaction data	22
Section C – Databases – contents, structure and annotation	29
C1 File formats	29
C2 Annotated sequence databases	35
C3 Genome and organism-specific databases	41
C4 Miscellaneous databases	45
Section D – Retrieval of biological data	51
D1 Data retrieval with Entrez and DBGET/LinkDB	51
D2 Data retrieval with SRS (sequence retrieval system)	55
Section E – Searching sequence databases by sequence similarity criteria	59
E1 Sequence similarity searches	59
E2 Amino acid substitution matrices	64
E3 Database searches: FASTA and BLAST	69
E4 Sequence filters	78
E5 Iterative database searches and PSI-BLAST	80
Section F – Multiple sequence alignment: gene and protein families	83
F1 Multiple sequence alignment and family relationships	83
F2 Protein families and pattern databases	86
F3 Protein domain families	91
Section G – Phylogenetics	97
G1 Phylogenetics, cladistics and ontology	97
G2 Building phylogenetic trees	103
G3 Evolution of macromolecular sequences	109
Section H – Sequence annotation	113
H1 Principles of genome annotation	113
H2 Annotation tools and resources	118

Section I – Structural bioinformatics	123
I1 Conceptual models of protein structure	123
I2 The relationship of protein three-dimensional structure to protein function	129
I3 The evolution of protein structure and function	131
I4 Obtaining, viewing and analyzing structural data	136
I5 Structural alignment	143
I6 Classification of proteins of known three-dimensional structure: CATH and SCOP	146
I7 Introduction to protein structure prediction	149
I8 Structure prediction by comparative modeling	151
I9 Secondary structure prediction	156
I10 Advanced protein structure prediction and prediction strategies	161
Section J – Microarray data analysis	165
J1 Microarray data: analysis methods	165
J2 Microarray data: tools and resources	170
J3 Sequence sampling and SAGE	175
Section K – Proteomic data analysis	179
K1 Analyzing data from 2D-PAGE gels	179
K2 Analyzing protein mass spectrometry data	185
Section L – Higher-order systems	189
L1 Modeling and reconstructing molecular pathways	189
L2 Protein interaction informatics	194
L3 Higher-order models	199
Section M – Cheminformatics in biology	203
M1 Conventions for representing molecules	203
M2 Cheminformatics resources	209
Section N – Bioinformatics in the pharmaceutical industry	215
N1 Bioinformatics and drug discovery	215
N2 Pharmainformatics resources	219
Section O – Basic principles of computing in bioinformatics	227
O1 Running computer software	227
O2 Computer operating systems	230
O3 Software downloading and installation	232
O4 Database management	236
Further reading	239
Glossary	243
Index	255