

Contents

| | | |
|----------|---|----------|
| 1 | Introduction | 1 |
| 1.1 | Bioinformatics | 1 |
| 1.2 | Bioinformatics-Related Terms | 2 |
| 1.3 | Journals Supporting Bioinformatics | 2 |
| 1.4 | Bioinformatics in Drug Discovery | 4 |
| 1.5 | Skills for Bioinformatics | 4 |
| 1.6 | Bioinformatics Warehousing in Drug Discovery | 5 |
| 1.7 | Bioinformatics Components | 5 |
| 1.8 | Bioinformation | 6 |
| 1.9 | Bioinformatics Variables | 7 |
| 1.10 | Cell Constituents | 7 |
| 1.10.1 | Nucleic Acids | 8 |
| 1.10.2 | Proteins | 9 |
| 1.10.3 | Classification of Amino Acids | 9 |
| 1.11 | Codon and Codon Usage Table | 9 |
| 1.12 | Bioinformation Discovery | 11 |
| 1.13 | Bioinformatics Principle | 11 |
| 1.14 | Bioinformatics Challenges | 12 |
| 1.15 | Biological Data | 12 |
| 1.16 | Data Explosion | 14 |
| 1.17 | Sequence Data | 14 |
| 1.18 | Structure Data | 14 |
| 1.19 | Small Molecules | 14 |
| 1.20 | Macromolecules | 16 |
| 1.21 | SCOP Dataset | 17 |
| 1.22 | CATH Dataset | 18 |
| 1.23 | Function Data | 20 |
| 1.24 | Pathways Data | 20 |
| 1.25 | Bioinformatics Developments | 21 |
| 1.26 | Discovery Environment | 21 |
| 1.27 | Sequence, Structure Alignment and Evolutionary Inferences | 22 |

| | | |
|----------|--|-----------|
| 1.28 | Molecular Modeling..... | 24 |
| 1.28.1 | Protein Modeling | 24 |
| 1.28.2 | Methods of Protein Modeling | 25 |
| 1.28.3 | Caveats on Homology Modeling | 26 |
| 1.29 | Exercises | 26 |
| 2 | Creating Datasets | 27 |
| 2.1 | Datasets | 27 |
| 2.2 | HLA-Binding Peptide Dataset | 29 |
| 2.3 | MHC–Peptide Structural Dataset..... | 29 |
| 2.4 | MHC–Peptide Structure Dataset Clustering | 32 |
| 2.5 | PDB Chain Identifier | 32 |
| 2.6 | Information Redundancy in Dataset | 36 |
| 2.7 | Information from MHC–Peptide Data..... | 36 |
| 2.8 | Structural Parameters for MHC–Peptide Dataset Analysis | 37 |
| 2.9 | Creation of Heterodimer and Homodimer Dataset | 37 |
| 2.10 | Homodimer Folding Dataset..... | 37 |
| 2.11 | Alanine-Mutated Interface Residues Dataset..... | 37 |
| 2.12 | Intronless Genes Dataset..... | 59 |
| 2.13 | Human Single Exon Gene (SEG) Dataset | 61 |
| 2.14 | Intron-Containing Genes Dataset..... | 61 |
| 2.15 | Fusion Protein Dataset | 62 |
| 2.16 | Exercises | 63 |
| | References | 64 |
| 3 | Tools and Techniques | 65 |
| 3.1 | Align | 65 |
| 3.2 | BIMAS..... | 66 |
| 3.3 | BLAST..... | 66 |
| 3.4 | ClustalW | 68 |
| 3.5 | DeCypher | 69 |
| 3.6 | DeepView | 69 |
| 3.7 | FASTA..... | 69 |
| 3.8 | Insight II..... | 69 |
| 3.9 | Genscan..... | 70 |
| 3.10 | GROMOS..... | 70 |
| 3.11 | HBPLUS | 70 |
| 3.12 | Lalign/Plalign..... | 71 |
| 3.13 | Ligplot..... | 71 |
| 3.14 | LOOK | 71 |
| 3.15 | Modeller..... | 72 |
| 3.16 | NACCESS..... | 73 |
| 3.17 | Phylip | 73 |
| 3.18 | ProtParam..... | 74 |
| 3.19 | Protorp..... | 74 |

| | | |
|----------|---|-----------|
| 3.20 | Psap | 74 |
| 3.21 | Ppsearch | 74 |
| 3.22 | Pymol | 74 |
| 3.23 | Rasmol | 74 |
| 3.24 | Rosetta Design | 75 |
| 3.25 | Surfnet | 75 |
| 3.26 | Sybyl | 75 |
| 3.27 | T-Epitope Designer | 75 |
| 3.28 | Exercises | 75 |
| | References | 76 |
| 4 | Protein Subunits Interaction | 79 |
| 4.1 | Protein Subunit Interaction | 79 |
| 4.2 | Dimer Datasets in Literature | 80 |
| 4.3 | Parameters in Subunit Interaction | 81 |
| 4.3.1 | Hydrophobic Effect | 81 |
| 4.3.2 | Interface Size | 82 |
| 4.3.3 | Interface H-Bonds | 82 |
| 4.3.4 | Interface Residues | 84 |
| 4.3.5 | Interface Residue Type | 84 |
| 4.3.6 | Inferences on Protein Subunit Interactions | 85 |
| 4.4 | Exercises | 85 |
| | References | 86 |
| 5 | Homodimer Folding and Binding | 87 |
| 5.1 | Importance of Homodimers | 87 |
| 5.2 | Homodimer Folding | 88 |
| 5.3 | Homodimer Structures in Folding | 89 |
| 5.4 | Size, Interface Area and Structure | 89 |
| 5.5 | Interface to Total Residues | 92 |
| 5.6 | Exterior, Interior and Interface Hydrophobicity in 2S and 3S | 93 |
| 5.7 | Folding and Binding Mechanism | 94 |
| 5.8 | Concluding Remarks | 96 |
| 5.9 | Exercises | 96 |
| | References | 96 |
| 6 | Fusion Proteins | 97 |
| 6.1 | Gene Fusion | 97 |
| 6.2 | Operons in Prokaryotes as Human Fusion Proteins | 98 |
| 6.3 | Multiple Functions in Fusion Proteins | 100 |
| 6.4 | Alternative Splicing in Fusion Genes | 101 |
| 6.5 | Protein Subunit Interaction and Fusion Proteins | 101 |
| 6.6 | Mechanism of Gene Fusion | 102 |
| 6.7 | Hypothesis of Gene Fusion | 102 |

| | | |
|----------|--|------------|
| 6.8 | Structural Importance of Fusion Proteins | 102 |
| 6.8.1 | Fusion Protein IGPS Function | 103 |
| 6.8.2 | Fusion Protein IGPS Structure..... | 104 |
| 6.8.3 | IGPS Sequence, Structure and Properties | 105 |
| 6.8.4 | Interface Area in IGPS..... | 106 |
| 6.8.5 | Gap Volume in IGPS..... | 106 |
| 6.8.6 | Radius of Gyration in IGPS | 108 |
| 6.8.7 | Structural Features of Fusion Protein IGPS..... | 108 |
| 6.9 | Exercises | 109 |
| | References..... | 110 |
| 7 | Major Histocompatibility Complex (MHC) | |
| | and Peptide Binding..... | 111 |
| 7.1 | MHC Biology and Diversity | 111 |
| 7.2 | Promise of MHC in Medicine..... | 114 |
| 7.3 | MHC Structure and Function..... | 114 |
| 7.3.1 | Class I MHC Structure and Function..... | 115 |
| 7.3.2 | Class II MHC Structure and Function | 118 |
| 7.4 | MHC-Peptide Motifs | 120 |
| 7.4.1 | Class I MHC-Peptide Motifs | 120 |
| 7.4.2 | Class II MHC-Peptide Motifs | 121 |
| 7.5 | MHC-Peptide Binding | 122 |
| 7.6 | MHC Polymorphism and Specificity | 123 |
| 7.7 | MHC-Peptide Complex in T-Cell-Mediated Immune Response..... | 123 |
| 7.8 | MHC-Peptide Binding Predictions | 124 |
| 7.8.1 | Data-Driven Methods..... | 124 |
| 7.8.2 | Molecular Modeling Methods..... | 126 |
| 7.8.3 | Limitations in Molecular Modeling Methods | 127 |
| 7.9 | Application..... | 127 |
| 7.10 | Exercises | 128 |
| | References..... | 128 |
| 8 | HLA Supertypes..... | 131 |
| 8.1 | HLA Supertypes: Definition | 131 |
| 8.2 | Grouping of HLA Alleles by Several Research Groups | 133 |
| 8.3 | Perplexing Issues with HLA Supertypes | 134 |
| 8.4 | Structural Basis for HLA Supertypes | 134 |
| 8.5 | Predictive Grouping of HLA Supertypes..... | 136 |
| 8.6 | Grouping Using Electrostatic Distribution Maps | 137 |
| 8.7 | Concluding Remarks..... | 137 |
| 8.8 | Exercises | 138 |
| | References..... | 138 |

| | | |
|-----------|--|------------|
| 9 | T-Epitope Designer | 141 |
| 9.1 | HLA-Peptide Binding and Its Prediction | 141 |
| 9.2 | Available Prediction Servers | 142 |
| 9.3 | T-Epitope Designer | 142 |
| 9.4 | Model | 142 |
| 9.5 | User Interface..... | 144 |
| 9.6 | Input Data..... | 144 |
| 9.7 | Output Result | 145 |
| 9.8 | Conclusion | 145 |
| 9.9 | Cautionary Note | 145 |
| 9.10 | Exercises | 146 |
| | References..... | 146 |
| 10 | Eukaryotic Genes, Functions, Genomes, Design, and Evolution | 149 |
| 10.1 | Eukaryotic Genes and Genomes | 149 |
| 10.2 | SEGE..... | 151 |
| 10.3 | Genome SEGE..... | 151 |
| 10.4 | Human Single Exon Genes | 152 |
| 10.5 | U-Genome..... | 154 |
| 10.6 | ExInt..... | 154 |
| 10.7 | Alternative Splicing | 155 |
| 10.8 | Intron and Exon Content in Genomes..... | 156 |
| 10.9 | Exon–Intron Length Patterns | 157 |
| 10.10 | Intron Organization and Evolution | 158 |
| 10.11 | Exercises | 159 |
| | References..... | 160 |
| | Index..... | 163 |



<http://www.springer.com/978-1-4419-0519-2>

Bioinformation Discovery

Data to Knowledge in Biology

Kangueane, P.

2009, XXV, 166 p.,

ISBN: 978-1-4419-0519-2