

Contents

Abbreviations	ix
Contributors	xi
Foreword. <i>Professor Janet Thornton</i>	xiii
1. Molecular evolution. <i>Sylvia Nagl</i>	1
Molecular evolution is a fundamental part of bioinformatics	1
Evolution of protein families	9
Outlook: Evolution takes place at all levels of biological organization	15
2. Gene finding. <i>John G. Sgouros and Richard M. Twyman</i>	19
Concepts	19
Finding genes in bacterial genomes	20
Finding genes in higher eukaryotes	21
Detecting non-coding RNA genes	27
3. Sequence comparison methods. <i>Christine Orengo</i>	29
Concepts	29
Data resources	30
Algorithms for pairwise sequence comparison	32
Fast database search methods	38
Assessing the statistical significance of sequence similarity	42
Intermediate sequence searching	44
Validation of sequence alignment methods by structural data	44
Multiple sequence alignment	45
4. Amino acid residue conservation. <i>William S.J. Valdar and David T. Jones</i>	49
Concepts	49
Models of molecular evolution	49
Substitution matrices	50
Scoring residue conservation	57
Methods for scoring conservation	59
Insights and conclusions	64
5. Function prediction from protein sequence. <i>Sylvia B. Nagl</i>	65
Overview	65
The similar sequence–similar structure–similar function paradigm	65
Functional annotation of biological sequences	66
Outlook: context-dependence of protein function	77
6. Protein structure comparison. <i>Ian Sillitoe and Christine Orengo</i>	81
Concepts	81
Data resources	84
Algorithms	84
Statistical methods for assessing structural similarity	99
Multiple structure comparison and 3-D templates for structural families	100
Conclusions	101

7. Protein structure classifications. <i>Frances Pearl and Christine Orengo</i>	103
Concepts	103
Data resources	104
Protocols used in classifying structures	104
Descriptions of the structural classification hierarchy	111
Overview of the populations in the different structural classifications and insights provided by the classifications	118
8. Comparative modeling. <i>Andrew C.R. Martin</i>	121
Concepts	121
Why do comparative modeling?	121
Experimental methods	123
Evaluation of model quality	131
Factors influencing model quality	132
Insights and conclusions	133
9. Protein structure prediction. <i>David T. Jones</i>	135
Concepts	135
Strategies for protein structure prediction	135
Secondary structure prediction	138
Fold recognition methods	145
<i>Ab initio</i> prediction methods	149
Critically assessing protein structure prediction	149
Conclusions	150
10. From protein structure to function. <i>Annabel E. Todd</i>	151
Introduction	151
What is function?	152
Challenges of inferring function from structure	152
Methods of functional evolution	152
Functional classifications	154
From structure to function	156
Evolution of protein function from a structural perspective	164
Structural genomics	171
Conclusions	174
11. From structure-based genome annotation to understanding genes and proteins. <i>Sarah A. Teichmann</i>	175
Concepts	175
Computational structural genomics: structural assignment of genome sequences	175
Methods and data resources for computational structural genomics	176
Proteome and protein evolution by computational structural genomics	181
Evolution of enzymes and metabolic pathways by structural annotation of genomes	186
Summary and outlook	191
12. Global approaches for studying protein–protein interactions. <i>Sarah A. Teichmann</i>	193
Concepts	193
Protein–protein interactions	193
Experimental approaches for large-scale determination of protein–protein interactions	195
Structural analyses of domain interactions	196
The use of gene order to predict protein–protein interactions	198

The use of phylogeny to predict protein–protein interactions	200
Summary and outlook	200
13. Predicting the structure of protein–biomolecular interactions. <i>Richard M. Jackson</i>	203
Concepts	203
Why predict molecular interactions?	203
Practical considerations	204
Molecular complementarity	204
The search problem	209
Conformational flexibility	211
Evaluation of models	214
Visualization methods	215
14. Experimental use of DNA arrays. <i>Paul Kellam and Xiaohui Liu</i>	217
Concepts	217
Methods for large-scale analysis of gene expression	218
Using microarrays	219
Properties and processing of array data	221
Data normalization	223
Microarray standards and databases	226
15. Mining gene expression data. <i>Xiaohui Liu and Paul Kellam</i>	229
Concepts	229
Data mining methods for gene expression analysis	230
Clustering	231
Classification	241
Conclusion and future research	244
16. Proteomics. <i>Malcolm P. Weir, Walter P. Blackstock and Richard M. Twyman</i>	245
The proteome	245
Proteomics	246
Technology platforms in proteomics	246
Case studies	254
Summary	257
17. Data management of biological information. <i>Nigel J. Martin</i>	259
Concepts	259
Data management concepts	260
Data management techniques	263
Challenges arising from biological data	270
Conclusions	271
18. Internet technologies for bioinformatics. <i>Andrew C.R. Martin</i>	273
Concepts	273
Methods and standards	274
Insights and conclusions	281
Glossary	283
Index	293