

Contents

Abbreviations	ix
Contributors	xii
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>	18
Concepts	18
Finding genes in bacterial genomes	19
Finding genes in higher eukaryotes	20
Detecting non-coding RNA genes	26
3. Sequence comparison methods. <i>Christine Orengo</i>	28
Concepts	28
Data resources	29
Algorithms for pairwise sequence comparison	30
Fast database search methods	38
Assessing the statistical significance of sequence similarity	41
Intermediate sequence searching	43
Validation of sequence alignment methods by structural data	44
Multiple sequence alignment	44

4. Amino acid residue conservation.	48
<i>William S.J. Valdar and David T. Jones</i>	
Concepts	48
Models of molecular evolution	48
Substitution matrices	49
Scoring residue conservation	56
Methods for scoring conservation	58
Insights and conclusions	63
5. Function prediction from protein sequence.	64
<i>Sylvia B. Nagl</i>	
Overview	64
The similar sequence-similar structure-similar function paradigm	65
Functional annotation of biological sequences	65
Outlook: context-dependence of protein function	76
6. Protein structure comparison.	79
<i>Ian Sillitoe and Christine Orengo</i>	
Concepts	79
Data resources	83
Algorithms	83
Statistical methods for assessing structural similarity	98
Multiple structure comparison and 3-D templates for structural families	99
Conclusions	100
7. Protein structure classifications.	101
<i>Frances Pearl and Christine Orengo</i>	
Concepts	101
Data resources	102
Protocols used in classifying structures	103
Descriptions of the structural classification hierarchy	109
Overview of the populations in the different structural classifications and insights provided by the classifications	116
8. Comparative modeling.	119
<i>Andrew C.R. Martin</i>	

Concepts	119
Why do comparative modeling?	120
Experimental methods	121
Evaluation of model quality	129
Factors influencing model quality	131
Insights and conclusions	132
9. Protein structure prediction.	134
<i>David T. Jones</i>	
Concepts	134
Strategies for protein structure prediction	135
Secondary structure prediction	137
Fold recognition methods	145
<i>Ab initio</i> prediction methods	148
Critically assessing protein structure prediction	149
Conclusions	150
10. From protein structure to function.	151
<i>Annabel E. Todd</i>	
Introduction	151
What is function?	152
Challenges of inferring function from structure	152
Methods of functional evolution	153
Functional classifications	155
From structure to function	157
Evolution of protein function from a structural perspective	164
Structural genomics	172
Conclusions	174
11. From structure-based genome annotation to understanding genes and proteins.	176
<i>Sarah A. Teichmann</i>	
Concepts	176
Computational structural genomics: structural assignment of genome sequences	177
Methods and data resources for computational structural genomics	177

Proteome and protein evolution by computational structural genomics	182
Evolution of enzymes and metabolic pathways by structural annotation of genomes	186
Summary and outlook	192
12. Global approaches for studying protein-protein interactions.	194
<i>Sarah A. Teichmann</i>	
Concepts	194
Protein-protein interactions	195
Experimental approaches for large-scale determination of protein-protein interactions	196
Structural analyses of domain interactions	197
The use of gene order to predict protein-protein interactions	199
The use of phylogeny to predict protein-protein interactions	201
Summary and outlook	201
13. Predicting the structure of protein-biomolecular interactions.	203
<i>Richard M. Jackson</i>	
Concepts	203
Why predict molecular interactions?	203
Practical considerations	204
Molecular complementarity	205
The search problem	209
Conformational flexibility	212
Evaluation of models	215
Visualization methods	216
14. Experimental use of DNA arrays.	217
<i>Paul Kellam and Xiaohui Liu</i>	
Concepts	217
Methods for large-scale analysis of gene expression	219
Using microarrays	219
Properties and processing of array data	221
Data normalization	223
Microarray standards and databases	226
15. Mining gene expression data.	229
<i>Xiaohui Liu and Paul Kellam</i>	

Concepts	229
Data mining methods for gene expression analysis	231
Clustering	231
Classification	242
Conclusion and future research	244
16. Proteomics.	246
<i>Malcolm P. Weir, Walter P. Blackstock and Richard M. Twyman</i>	
The proteome	246
Proteomics	246
Technology platforms in proteomics	248
Case studies	256
Summary	258
17. Data management of biological information.	260
<i>Nigel J. Martin</i>	
Concepts	260
Data management concepts	261
Data management techniques	265
Challenges arising from biological data	272
Conclusions	273
18. Internet technologies for bioinformatics.	274
<i>Andrew C.R. Martin</i>	
Concepts	274
Methods and standards	275
Insights and conclusions	282
Glossary	284
Index	292
Colour plates can be found between pages 50 and 51, 210 and 211, 242 and 243	