Contents

Section I Generating and Inferring Structures

1 *Ab Initio* Protein Structure Prediction 3
Jooyoung Lee, Sitao Wu, and Yang Zhang

1.1 Introduction 3
1.2 Energy Functions 5
1.2.1 Physics-Based Energy Functions 5
1.2.2 Knowledge-Based Energy Function Combined with Fragments 9
1.3 Conformational Search Methods 13
1.3.1 Monte Carlo Simulations 14
1.3.2 Molecular Dynamics 15
1.3.3 Genetic Algorithm 15
1.3.4 Mathematical Optimization 16
1.4 Model Selection 16
1.4.1 Physics-Based Energy Function 17
1.4.2 Knowledge-Based Energy Function 17
1.4.3 Sequence-Structure Compatibility Function 18
1.4.4 Clustering of Decoy Structures 19
1.5 Remarks and Discussions 19

2 Fold Recognition 27
Lawrence A. Kelley

2.1 Introduction 27
2.1.1 The Importance of Blind Trials: The CASP Competition 28
2.1.2 *Ab Initio* Structure Prediction Versus Homology Modelling 28
2.1.3 The Limits of Fold Space 30
2.1.4 A Note on Terminology: 'Threading' and 'Fold Recognition' 31
2.2 Threading 31

2.2.1 Knowledge-Based Potentials 32
2.2.2 Finding an Alignment 34
2.2.3 Heuristics for Alignment 35
2.3 Remote Homology Detection Without Threading 38
2.3.1 Using Predicted Structural Features 39
2.3.2 Sequence Profiles and Hidden Markov Models 41
2.3.3 Fold Classification and Support Vector Machines 43
2.3.4 Consensus Approaches 45
2.3.5 Traversing the Homology Network 45
2.4 Alignment Accuracy, Model Quality and Statistical Significance 47
2.4.1 Algorithms for Alignment Generation and Assessment 47
2.4.2 Estimation of Statistical Significance 48
2.5 Tools for Fold Recognition on the Web 49
2.6 The Future 50

3 Comparative Protein Structure Modelling 57
András Fiser

3.1 Introduction 57
3.1.1 Structure Determines Function 57
3.1.2 Sequences, Structures, Structural Genomics 58
3.1.3 Approaches to Protein Structure Prediction 58
3.2 Steps in Comparative Protein Structure Modelling 60
3.2.1 Searching for Structures Related to the Target Sequence 62
3.2.2 Selecting Templates 64
3.2.3 Sequence to Structure Alignment 65
3.2.4 Model Building 67
3.2.5 Model Evaluation 76
3.3 Performance of Comparative Modelling 77
3.3.1 Accuracy of Methods 77
3.3.2 Errors in Comparative Models 78
3.4 Applications of Comparative Modelling 80
3.4.1 Modelling of Individual Proteins 80
3.4.2 Comparative Modelling and the Protein Structure Initiative 80
3.5 Summary 81

4 Membrane Protein Structure Prediction 91
Timothy Nugent and David T. Jones

4.1 Introduction 91
4.2 Structural Classes 92

4.2.1 Alpha-Helical Bundles 92
4.2.2 Beta-Barrels 92
4.3 Membrane Proteins Are Difficult to Crystallise 94
4.4 Databases 94
4.5 Multiple Sequence Alignments 96
4.6 Transmembrane Protein Topology Prediction 98
4.6.1 Alpha-Helical Proteins 98
4.6.2 Beta-Barrel Proteins 102
4.6.3 Whole Genome Analysis 102
4.6.4 Data Sets, Homology, Accuracy and Cross-Validation 103
4.7 3D Structure Prediction 105
4.8 Future Developments 107

5 Bioinformatics Approaches to the Structure and Function of Intrinsically Disordered Proteins 113
Peter Tompa

5.1 The Concept of Protein Disorder 113
5.2 Sequence Features of IDPs 115
5.2.1 The Unusual Amino Acid Composition of IDPs 115
5.2.2 Sequence Patterns of IDPs 115
5.2.3 Low Sequence Complexity and Disorder 116
5.3 Prediction of Disorder 116
5.3.1 Prediction of Low-Complexity Regions 116
5.3.2 Charge-Hydropathy Plot 117
5.3.3 Propensity-Based Predictors 117
5.3.4 Predictors Based on the Lack of Secondary Structure 118
5.3.5 Machine Learning Algorithms 119
5.3.6 Prediction Based on Contact Potentials 120
5.3.7 A Reduced Alphabet Suffices to Predict Disorder 121
5.3.8 Comparison of Disorder Prediction Methods 122
5.4 Functional Classification of IDPs 122
5.4.1 Gene Ontology-Based Functional Classification of IDPs 122
5.4.2 Classification of IDPs Based on Their Mechanism of Action 123
5.4.3 Function-Related Structural Elements in IDPs 126
5.5 Prediction of the Function of IDPs 128
5.5.1 Correlation of Disorder Pattern and Function 128
5.5.2 Predicting Short Recognition Motifs in IDRs 128
5.5.3 Prediction of MoRFs 129

5.5.4 Combination of Information on Sequence and Disorder: Phosphorylation Sites and CaM Binding Motifs 131
5.5.5 Flavours of Disorder 131
5.6 Limitations of IDP Function Prediction 132
5.6.1 Rapid Evolution of IDPs 132
5.6.2 Sequence Independence of Function and Fuzziness 133
5.6.3 Good News: Conservation and Disorder 134
5.7 Conclusions 135

Section II From Structures to Functions

6 Function Diversity Within Folds and Superfamilies 143
Benoit H. Dessailly and Christine A. Orengo

6.1 Defining Function 143
6.2 From Fold to Function 145
6.2.1 Definition of a Fold 145
6.2.2 Prediction of Function Using Fold Relationships 148
6.3 Function Diversity Between Homologous Proteins 151
6.3.1 Definitions 151
6.3.2 Evolution of Protein Superfamilies 152
6.3.3 Function Divergence During Protein Evolution 154
6.4 Conclusion 162

7 Predicting Protein Function from Surface Properties 167
Nicholas J. Burgoyne and Richard M. Jackson

7.1 Surface Descriptions 167
7.1.1 The van der Waals Surface 167
7.1.2 Molecular Surface (Solvent Excluded Surface) 168
7.1.3 The Solvent Accessible Surface 168
7.2 Surface Properties 169
7.2.1 Hydrophobicity 169
7.2.2 Electrostatics Properties 170
7.2.3 Surface Conservation 171
7.3 Function Predictions Using Surface Properties 171
7.3.1 Hydrophobic Surface 172
7.3.2 Electrostatic Surface 172
7.3.3 Surface Conservation 173
7.3.4 Combining Surface Properties for Function Prediction 174
7.4 Protein-Ligand Interactions 174
7.4.1 Properties of Protein-Ligand Interactions 174

7.4.2 Predicting Binding Site Locations 175
7.4.3 Predictions of Druggability 178
7.4.4 Annotation of Ligand Binding Sites 178
7.5 Protein-Protein Interfaces 180
7.5.1 Properties of Protein-Protein Interfaces 180
7.5.2 Hot-Spot Regions in Protein Interfaces 181
7.5.3 Predictions of Interface Location 182
7.6 Summary 184

8 3D Motifs 187
Elaine C. Meng, Benjamin J. Polacco, and Patricia C. Babbitt

8.1 Background and Significance 188
8.1.1 What Is Function? 189
8.1.2 Three-Dimensional Motifs: Definition and Scope 190
8.2 Overview of Methods 190
8.2.1 Motif Discovery 190
8.2.2 Motif Description and Matching 191
8.2.3 Interpretation of Results 193
8.3 Specific Methods 196
8.3.1 User-Defined Motifs 197
8.3.2 Motif Discovery 201
8.4 Related Methods 208
8.4.1 Hybrid (Point-Surface) Descriptions 208
8.4.2 Single-Point-Centred Descriptions 208
8.5 Docking for Functional Annotation 210
8.6 Discussion 212
8.7 Conclusions 212

9 Protein Dynamics: From Structure to Function 217
Marcus B. Kubitzki, Bert L. de Groot, and Daniel Seeliger

9.1 Molecular Dynamics Simulations 217
9.1.1 Principles and Approximations 218
9.1.2 Applications 220
9.1.3 Limitations – Enhanced Sampling Algorithms 226
9.2 Principal Component Analysis 230
9.3 Collective Coordinate Sampling Algorithms 233
9.3.1 Essential Dynamics 233
9.3.2 TEE-REX 234
9.4 Methods for Functional Mode Prediction 237
9.4.1 Normal Mode Analysis 237

9.4.2 Elastic Network Models 238
9.4.3 CONCOORD 239
9.5 Summary and Outlook 242

10 Integrated Servers for Structure-Informed Function Prediction 251
Roman A. Laskowski

10.1 Introduction 251
10.1.1 The Problem of Predicting Function from Structure 252
10.1.2 Structure-Function Prediction Methods 253
10.2 ProKnow 254
10.2.1 Fold Matching 254
10.2.2 3D Motifs 256
10.2.3 Sequence Homology 257
10.2.4 Sequence Motifs 257
10.2.5 Protein Interactions 258
10.2.6 Combining the Predictions 258
10.2.7 Prediction Success 258
10.3 ProFunc 259
10.3.1 ProFunc's Structure-Based Methods 259
10.3.2 Assessment of the Structural Methods 267
10.4 Conclusion 269

11 Case Studies: Function Predictions of Structural Genomics Results 273
James D. Watson and Janet M. Thornton

11.1 Introduction 273
11.2 Large Scale Function Prediction Case Studies 275
11.3 Some Specific Examples 281
11.4 Community Annotation 287
11.5 Conclusions 288

12 Prediction of Protein Function from Theoretical Models 293
Iwona A. Cymerman, Daniel J. Rigden, and Janusz M. Bujnicki

12.1 Background 293
12.2 Protein Models as a Community Resource 295
12.2.1 Model Quality 296
12.2.2 Databases of Models 297

12.3 Accuracy and Added Value of Model-Derived Properties 298
12.3.1 Implementation 300
12.4 Practical Application 302
12.4.1 Plasticity of Catalytic Site Residues 302
12.4.2 Mutation Mapping 304
12.4.3 Protein Complexes 305
12.4.4 Function Predictions from Template-Free Models 306
12.4.5 Prediction of Ligand Specificity 309
12.4.6 Structure Modelling of Alternatively Spliced Isoforms 310
12.4.7 From Broad Function to Molecular Details 312
12.5 What Next? 314

Index 319