

Contents

Part I Generating and Inferring Structures

1	Ab Initio Protein Structure Prediction	3
	Jooyoung Lee, Peter L. Freddolino and Yang Zhang	
1.1	Introduction	4
1.2	Energy Functions	5
1.2.1	Physics-Based Energy Functions	7
1.2.2	Knowledge-Based Energy Function Combined with Fragments	11
1.3	Conformational Search Methods	18
1.3.1	Monte Carlo Simulations	18
1.3.2	Molecular Dynamics	19
1.3.3	Genetic Algorithm	20
1.3.4	Mathematical Optimization	21
1.4	Model Selection	21
1.4.1	Physics-Based Energy Function	22
1.4.2	Knowledge-Based Energy Function	23
1.4.3	Sequence-Structure Compatibility Function	24
1.4.4	Clustering of Decoy Structures	25
1.5	Remarks and Discussions	25
	References	27
2	Protein Structures, Interactions and Function from Evolutionary Couplings	37
	Thomas A. Hopf and Debora S. Marks	
2.1	Introduction	38
2.2	Evolutionary Couplings from Sequence Alignments	42
2.2.1	The Global Model	42
2.3	Three-Dimensional Protein Structures from Evolutionary Couplings	46

2.3.1	Transmembrane Proteins	48
2.3.2	Protein Interactions and Complexes.	49
2.3.3	Conformational Plasticity and Disordered Proteins	51
2.4	Predicting the Effect of Mutations	52
2.5	Summary and Future Challenges	54
	References.	55
3	Fold Recognition	59
	Lawrence A. Kelley	
3.1	Introduction	59
3.1.1	The Importance of Blind Trials: The CASP Competition.	60
3.1.2	Ab Initio Structure Prediction Versus Homology Modelling	60
3.1.3	The Limits of Fold Space	62
3.2	Pushing Sequence Similarity to the Limits: The Power of Evolutionary Information	64
3.2.1	The Rise of Hidden Markov Models.	67
3.2.2	Using Predicted Structural Features.	68
3.2.3	Harnessing 3D Structure to Enhance Recognition	70
3.2.4	Knowledge-Based Potentials	70
3.2.5	Summary.	72
3.3	CASP: The Great Filter	72
3.3.1	The Leaders.	73
3.3.2	Individual Algorithms	73
3.3.3	Consensus Methods.	75
3.4	Post-processing	76
3.4.1	Choosing and Combining Candidate Models.	76
3.4.2	Post-processing in Practice	79
3.4.3	Use of Contacts.	82
3.5	Tools for Fold Recognition on the Web.	85
3.6	The Future	86
	References.	88
4	Comparative Protein Structure Modelling	91
	András Fiser	
4.1	Introduction	91
4.1.1	Structure Determines Function.	91
4.1.2	Sequences, Structures, Structural Genomics.	92
4.1.3	Approaches to Protein Structure Prediction	94
4.2	Steps in Comparative Protein Structure Modelling	96
4.2.1	Searching for Structures Related to the Target Sequence	98
4.2.2	Selecting Templates.	100

4.2.3	Sequence to Structure Alignment	102
4.2.4	Model Building	103
4.2.5	Model Evaluation	114
4.3	Performance of Comparative Modelling	116
4.3.1	Accuracy of Methods	116
4.3.2	Errors in Comparative Models	117
4.4	Applications of Comparative Modelling	119
4.4.1	Modelling of Individual Proteins	119
4.4.2	Comparative Modelling and the Protein Structure Initiative	119
4.5	Summary	120
	References	121
5	Advances in Computational Methods for Transmembrane Protein Structure Prediction	135
	Tim Nugent, David Jones and Sikander Hayat	
5.1	Introduction	136
5.2	Membrane Protein Structural Classes	136
5.2.1	α -Helical Bundles	137
5.2.2	Transmembrane β -Barrels	137
5.3	Databases	139
5.4	Multiple Sequence Alignments	140
5.5	Transmembrane Protein Topology Prediction	141
5.5.1	Early α -Helical Topology Prediction Approaches	142
5.5.2	Machine Learning Approaches for α -Helical Topology Prediction	142
5.5.3	Signal Peptides and Re-entrant Helices	144
5.5.4	Consensus Approaches for α -Helical Topology Prediction	145
5.5.5	Transmembrane β -Barrel Topology Prediction	146
5.5.6	Empirical Approaches for β -Barrel Topology Prediction	147
5.5.7	Machine Learning Approaches for β -Barrel Topology Prediction	148
5.5.8	Consensus Approaches for β -Barrel Topology Prediction	149
5.6	3D Structure Prediction	150
5.6.1	Homology Modelling of α -Helical Transmembrane Proteins	150
5.6.2	Homology Modelling of Transmembrane β -Barrel Proteins	151
5.6.3	De Novo Modelling of α -Helical Transmembrane Proteins	152
5.6.4	De Novo Modelling of Transmembrane β -Barrels	154

5.6.5	Covariation-Based Approaches	154
5.6.6	Evolutionary Covariation-Based Methods for De Novo Modelling of α -Helical Membrane Proteins	155
5.6.7	Evolutionary Covariation-Based Methods for Transmembrane β -Barrel Structure Prediction	157
5.7	Future Directions	158
	References.	158
6	Bioinformatics Approaches to the Structure and Function of Intrinsically Disordered Proteins	167
	Zsuzsanna Dosztányi and Peter Tompa	
6.1	The Concept of Protein Disorder	168
6.2	Sequence Features of IDPs	169
6.2.1	The Unusual Amino Acid Composition of IDPs	169
6.2.2	Low Sequence Complexity and Disorder.	169
6.2.3	Flavours of Disorder	170
6.3	Prediction of Disorder.	171
6.3.1	Charge-Hydrophathy Plot	171
6.3.2	Propensity-Based Predictors.	171
6.3.3	Prediction Based on Simplified Biophysical Models	174
6.3.4	Machine Learning Algorithms	175
6.3.5	Related Approaches for the Prediction of Protein Disorder.	177
6.3.6	Comparison of Disorder Prediction Methods	178
6.4	Databases of IDPs	179
6.5	Structural Features of IDPs	180
6.6	Functional Classification of IDPs	181
6.6.1	Gene Ontology-Based Functional Classification of IDPs	182
6.6.2	Classification of IDPs Based on Their Mechanism of Action.	183
6.6.3	Functional Features of IDPs	185
6.7	Prediction of the Function of IDPs	188
6.7.1	Predicting Short Recognition Motifs in IDRs	190
6.7.2	Prediction of Disordered Binding Regions/MoRFs	191
6.7.3	Combination of Information on Sequence and Disorder: Phosphorylation Sites and CaM Binding Motifs	192
6.7.4	Correlation of Disorder Pattern and Function	193
6.8	Evolution of IDPs.	194
6.9	Conclusions	195
	References.	195

7	Prediction of Protein Aggregation and Amyloid Formation	205
	Ricardo Graña-Montes, Jordi Pujols-Pujol, Carlota Gómez-Picanyol and Salvador Ventura	
7.1	Introduction	206
7.2	The Physico-chemical and Structural Basis of Protein Aggregation	206
7.2.1	Intrinsic Determinants of Protein Aggregation	213
7.2.2	Extrinsic Determinants of Protein Aggregation	214
7.2.3	Specific Sequence Stretches Drive Aggregation	214
7.2.4	Structural Determinants of Amyloid-like Aggregation	215
7.3	Prediction of Protein Aggregation from the Primary Sequence	216
7.3.1	Phenomenological Approaches	221
7.3.2	Structure-Based Approaches	225
7.3.3	Consensus Methods	230
7.3.4	Applications of Sequence-Based Predictors	232
7.4	Prediction of Aggregation Propensity from the Tertiary Structure	242
7.5	Concluding Remarks	253
	References	254
8	Prediction of Biomolecular Complexes	265
	Anna Vangone, Romina Oliva, Luigi Cavallo and Alexandre M.J.J. Bonvin	
8.1	Introduction	266
8.2	Docking	268
8.2.1	Step 1: Searching	269
8.2.2	Step 2: Scoring	270
8.2.3	Data-Driven Docking	274
8.3	The Challenges of Docking: Flexibility and Binding Affinity	275
8.3.1	Changes upon Binding: The Flexible Docking Challenge	275
8.3.2	The ‘Perfect’ Scoring Function and the Binding Affinity Problem	276
8.4	Protein-Peptide Docking	278
8.5	Post-docking: Interface Prediction from Docking Results and Use of Docking-Derived Contacts for Clustering and Ranking	279
8.5.1	Web Tools for the Post-docking Processing	281
8.6	Concluding Remarks	283
	References	284

Part II From Structures to Functions

9	Function Diversity Within Folds and Superfamilies	295
	Benoit H. Dessailly, Natalie L. Dawson, Sayoni Das and Christine A. Orengo	
9.1	Defining Function	296
9.2	From Fold to Function	297
	9.2.1 Definition of a Fold	297
	9.2.2 Prediction of Function Using Fold Relationships	300
9.3	Function Diversity Between Homologous Proteins	303
	9.3.1 Definitions	303
	9.3.2 Evolution of Protein Superfamilies	307
	9.3.3 Function Divergence During Protein Evolution	308
9.4	Conclusion	320
	Bibliography	320
10	Function Prediction Using Patches, Pockets and Other Surface Properties	327
	Daniel J. Rigden	
10.1	Definitions of Protein Surfaces	328
10.2	Surface Patches	329
	10.2.1 Hydrophobic Patches	329
	10.2.2 Electrostatics	336
	10.2.3 Sequence Conservation	338
	10.2.4 Surface Atom Triplet Propensities	339
	10.2.5 Multiple Properties	340
10.3	Pockets	340
	10.3.1 Geometric Descriptions of Pockets	342
	10.3.2 Channels and Tunnels	343
	10.3.3 Distinguishing Functional Pockets	344
	10.3.4 Predicting Ligands for Pockets	345
10.4	Prediction of Catalytic Residues	347
10.5	Protein-Protein Interfaces	349
10.6	Other Specialised Binding Site Predictors	350
10.7	Medicinal Applications	352
10.8	Conclusions	353
	References	354
11	3D Motifs	361
	Jerome P. Nilmeier, Elaine C. Meng, Benjamin J. Polacco and Patricia C. Babbitt	
11.1	Background: Functional Annotation	362
	11.1.1 What Is Function?	363
	11.1.2 Genomics and Functional Annotation	363
	11.1.3 The Need for Structure-Based Methods	365

11.2	3D Motif Matching Techniques	366
11.2.1	What Is a 3D Motif?	366
11.2.2	Historical Development of Motif Matching Methods	369
11.3	Algorithmic Approaches to Motif Matching	373
11.3.1	Methods Using 3D Motifs	374
11.3.2	Efficiency Considerations for 3D Motifs	375
11.3.3	Methods with Nonstandard Motif Information	376
11.3.4	Interpretation of Results	377
11.4	Methods for Deriving Motifs	378
11.4.1	Literature Search and Manual Curation	379
11.4.2	Annotated Sites in PDB Structures	379
11.4.3	Mining for Emergent Properties	380
11.5	Molecular Docking for Functional Annotation	383
11.6	Discussion and Conclusions	385
	References	386
12	Protein Dynamics: From Structure to Function	393
	Marcus B. Kubitzki, Bert L. de Groot and Daniel Seeliger	
12.1	Molecular Dynamics Simulations	393
12.1.1	Principles and Approximations	394
12.1.2	Applications	396
12.1.3	Limitations—Enhanced Sampling Algorithms	402
12.2	Principal Component Analysis	406
12.3	Collective Coordinate Sampling Algorithms	409
12.3.1	Essential Dynamics	409
12.3.2	TEE-REX	410
12.4	Methods for Functional Mode Prediction	413
12.4.1	Normal Mode Analysis	413
12.4.2	Elastic Network Models	414
12.4.3	CONCOORD	415
12.5	Summary and Outlook	419
	References	420
13	Integrated Servers for Structure-Informed Function Prediction	427
	Roman A. Laskowski	
13.1	Introduction	427
13.1.1	The Problem of Predicting Function from Structure	428
13.1.2	Structure-Function Prediction Methods	430
13.2	ProKnow	431
13.2.1	Fold Matching	432
13.2.2	3D Motifs	434
13.2.3	Sequence Homology	434

13.2.4	Sequence Motifs	434
13.2.5	Protein Interactions	434
13.2.6	Combining the Predictions.	435
13.2.7	Prediction Success.	435
13.3	ProFunc	436
13.3.1	ProFunc's Structure-Based Methods	437
13.3.2	Assessment of the Structural Methods.	442
13.4	Conclusion	444
	References.	445
14	Case Studies: Function Predictions of Structural Genomics Results.	449
	James D. Watson, Roman A. Laskowski and Janet M. Thornton	
14.1	Introduction	449
14.2	Function Prediction Case Studies	451
14.2.1	Teichman et al. (2001)	451
14.2.2	Kim et al. (2003).	451
14.2.3	Watson et al. (2007)	453
14.2.4	Lee et al. (2011)	456
14.3	Some Specific Examples.	456
14.3.1	Adams et al. (2007).	456
14.3.2	AF0491 Protein.	457
14.3.3	The GxGYxYP Family	459
14.4	Community Annotation.	460
14.5	Conclusions	461
	References.	462
15	Prediction of Protein Function from Theoretical Models	467
	Daniel J. Rigden, Iwona A. Cymerman and Janusz M. Bujnicki	
15.1	Background	467
15.2	Suitability of Protein 3D Models for Structure-Based Predictions	469
15.2.1	Surface Properties	470
15.2.2	Functional Sites.	472
15.2.3	Specific Binding Predictions	473
15.2.4	Small Molecule Binding	474
15.2.5	Protein-Protein Interactions	476
15.2.6	Protein Model Databases.	477
15.3	Function Prediction Examples.	478
15.3.1	Fold Prediction with Fragment-Based Ab Initio Models	478
15.3.2	Fold Prediction with Contact-Based Models	481
15.3.3	Plasticity of Catalytic Site Residues	483
15.3.4	Prediction of Ligand Specificity	484

15.3.5	Prediction of Cofactor Specificity Using an Entry from a Database of Models	485
15.3.6	Mutation Mapping	488
15.3.7	Protein Complexes.	489
15.3.8	Structure Modelling of Alternatively Spliced Isoforms	490
15.3.9	From Broad Function to Molecular Details	491
15.4	Conclusions	493
	References.	493
Index	499

From Protein Structure to Function with Bioinformatics

J. Rigden, D. (Ed.)

2017, XV, 503 p. 86 illus., 75 illus. in color., Hardcover

ISBN: 978-94-024-1067-9