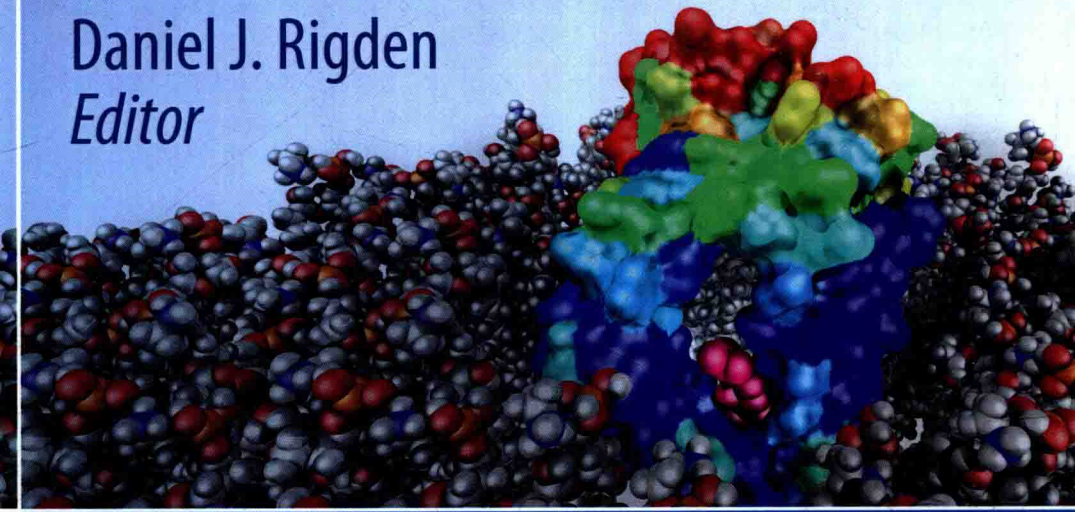


Daniel J. Rigden
Editor



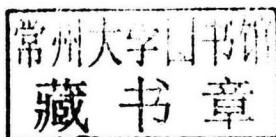
From Protein Structure to Function with Bioinformatics

Second Edition

Daniel J. Rigden
Editor

From Protein Structure to Function with Bioinformatics

Second Edition



 Springer

Editor

Daniel J. Rigden
Institute of Integrative Biology
University of Liverpool
Liverpool
UK

ISBN 978-94-024-1067-9

ISBN 978-94-024-1069-3 (eBook)

DOI 10.1007/978-94-024-1069-3

Library of Congress Control Number: 2017932416

© Springer Science+Business Media B.V. 2009, 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature

The registered company is Springer Science+Business Media B.V.

The registered company address is: Van Godewijckstraat 30, 3311 GX Dordrecht, The Netherlands

From Protein Structure to Function with Bioinformatics

Preface to the Second Edition

Welcome to the second edition! Since the publication of the first edition, the research area of protein structural informatics has continued to grow in volume and significance. A search of PubMed for 'protein structural bioinformatics' shows around 1000 papers in 2009 when the first edition was published, doubling to over 2000 in 2015. In the same period, the Protein Data Bank has similarly almost doubled, breaching 100,000 entries in 2014. Nevertheless, the gap between the protein sequences and structures continues to grow, as new technologies allow cheap and facile sequencing of previously intractable organisms and even of entire environments. Protein structural bioinformatics offers a computational route to bridge this gap by predicting structures for uncharacterised families. Those structures can then be analysed by a wide variety of further bioinformatics algorithms to shed light on their function. These two interlinking research areas are the topic of this book.

This second edition contains three chapters addressing areas not covered in the first edition. Each is contributed by world-leading experts in the field. The remaining chapters are all revised, many dramatically, to reflect seven years of fast-moving bioinformatics research with one chapter being entirely replaced. As previously, there are two sections covering first methods to generate or infer structure and secondly structure-based function annotation. Naturally, such a division is never clear-cut as prediction of a structure may simultaneously inform about its likely functions. For example, annotation of an intrinsically disordered region would immediately suggest, in eukaryotes at least, a role in protein-protein interaction since such stretches frequently harbour linear motifs bound by recognition modules on partner proteins.

The first new chapter, Chap. 2, covers arguably the most exciting development in protein bioinformatics of recent years, namely the new-found ability to accurately predict contacting residue pairs through covariance analysis of large multiple sequence alignments. These contact predictions have a wide and still expanding range of applications. Most obviously, the data allow for protein structure prediction in conjunction either with protein distance geometry methods or, more effectively, by synergistic incorporation into fragment assembly *ab initio* modelling

methods. The contact predictions also inform on the likely harmfulness of single amino acid polymorphisms (SAPs) and allow for better prediction of protein-protein interactions. Prediction of protein-protein complex structures, both between globular domains and between a domain and a short linear motif, is the subject of the new Chap. 8. A full accounting of protein-protein interactions in cells is crucial for the future prospects of integrative systems-level methods, while structural knowledge of interfaces again contributes to prediction of the consequences of SAPs. The third new arrival, Chap. 7, covers predictions of amyloid structure in proteins. Such structure is of huge biomedical interest, underlying diseases such as Parkinson's and Alzheimer's, but is equally intriguing for the normal physiological roles of 'functional amyloids'. Finally, the new Chap. 10 text covers the fascinating variety of means by which structural bioinformatics can mark up a structure, experimental or modelled, for likely functional pockets and patches on the protein surface.

The methods covered in this book comprise a comprehensive toolkit to address future challenges in protein structure, function and evolution. Recent papers open up new viewpoints on protein evolution (Alva et al. 2015; Edwards and Deane 2015) and on the amenability of different folds to functional innovation (Toth-Petroczy and Tawfik 2014), treat the biophysical consequences of protein ageing (de Graff et al. 2016) and even reveal oversights in our accounting of molecular interactions (Newberry and Raines 2016). Clearly, exciting times lie ahead for protein bioinformaticians!

Liverpool, UK

Daniel J. Rigden

References

- Alva V, Soding J, Lupas AN (2015) A vocabulary of ancient peptides at the origin of folded proteins. *Elife* 4:e09410
- de Graff AM, Hazoglou MJ, Dill KA (2016) Highly charged proteins: the achilles' heel of aging proteomes. *Structure* 24(2):329–336
- Edwards H, Deane CM (2015) Structural bridges through fold space. *PLoS Comput Biol* 11(9): e1004466
- Newberry RW, Raines RT (2016) A prevalent intraresidue hydrogen bond stabilizes proteins. *Nat Chem Biol* 12(12):1084–1088
- Toth-Petroczy A, Tawfik DS (2014) The robustness and innovability of protein folds. *Curr Opin Struct Biol* 26:131–138

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-Hydropathy 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

- Tompa P (2002) Intrinsically unstructured proteins. *Trends Biochem Sci* 27(10):527–533
- Tompa P (2005) The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett* 579(15):3346–3354. doi:10.1016/j.febslet.2005.03.072
- Tompa P (2012) Intrinsically disordered proteins: a 10-year recap. *Trends Biochem Sci* 37(12):509–516. doi:10.1016/j.tibs.2012.08.004
- Tompa P, Csermely P (2004) The role of structural disorder in the function of RNA and protein chaperones. *Faseb J* 18(11):1169–1175. doi:10.1096/fj.04-1584rev
- Tompa P, Fuxreiter M (2008) Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci* 33(1):2–8. doi:10.1016/j.tibs.2007.10.003
- Tompa P, Szasz C, Buday L (2005) Structural disorder throws new light on moonlighting. *Trends Biochem Sci* 30(9):484–489. doi:10.1016/j.tibs.2005.07.008
- Tompa P, Dosztányi Z, Simon I (2006) Prevalent structural disorder in *E. coli* and *S. cerevisiae* proteomes. *J Proteome Res* 5(8):1996–2000. doi:10.1021/pr0600881
- Tompa P, Fuxreiter M, Oldfield CJ, Simon I, Dunker AK, Uversky VN (2009) Close encounters of the third kind: disordered domains and the interactions of proteins. *BioEssays* 31(3):328–335. doi:10.1002/bies.200800151
- Tompa P, Davey NE, Gibson TJ, Babu MM (2014) A million peptide motifs for the molecular biologist. *Mol Cell* 55(2):161–169. doi:10.1016/j.molcel.2014.05.032
- Trizenberg SJ (1995) Structure and function of transcriptional activation domains. *Curr Opin Genet Dev* 5(2):190–196
- Trombitas K, Greaser M, Labeit S, Jin JP, Kellermayer M, Helmes M, Granzier H (1998) Titin extensibility in situ: entropic elasticity of permanently folded and permanently unfolded molecular segments. *J Cell Biol* 140(4):853–859
- Tucker MM, Robinson JB Jr, Stellwagen E (1981) The effect of proteolysis on the calmodulin activation of cyclic nucleotide phosphodiesterase. *J Biol Chem* 256(17):9051–9058
- Tuite MF, Koloteva-Levin N (2004) Propagating prions in fungi and mammals. *Mol Cell* 14(5):541–552. doi:10.1016/j.molcel.2004.05.012
- Uversky VN (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* 11(4):739–756. doi:10.1110/ps.4210102
- Uversky VN (2013) A decade and a half of protein intrinsic disorder: biology still waits for physics. *Protein Sci* 22(6):693–724. doi:10.1002/pro.2261
- Uversky VN, Gillespie JR, Fink AL (2000) Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins* 41(3):415–427
- Uversky VN, Oldfield CJ, Dunker AK (2005) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recogn* 18(5):343–384. doi:10.1002/jmr.747
- Vacic V, Oldfield CJ, Mohan A, Radivojac P, Cortese MS, Uversky VN, Dunker AK (2007) Characterization of molecular recognition features, MoRFs, and their binding partners. *J Proteome Res* 6(6):2351–2366. doi:10.1021/pr0701411
- van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, Fuxreiter M, Gough J, Gsponer J, Jones DT, Kim PM, Kriwacki RW, Oldfield CJ, Pappu RV, Tompa P, Uversky VN, Wright PE, Babu MM (2014) Classification of intrinsically disordered regions and proteins. *Chem Rev* 114(13):6589–6631. doi:10.1021/cr400525m
- Van Roey K, Gibson TJ, Davey NE (2012) Motif switches: decision-making in cell regulation. *Curr Opin Struct Biol* 22(3):378–385. doi:10.1016/j.sbi.2012.03.004
- Van Roey K, Dinkel H, Weatheritt RJ, Gibson TJ, Davey NE (2013) The switches. ELM resource: a compendium of conditional regulatory interaction interfaces. *Sci Signal* 6(269):rs7. doi:10.1126/scisignal.2003345
- Van Roey K, Uyar B, Weatheritt RJ, Dinkel H, Seiler M, Budd A, Gibson TJ, Davey NE (2014) Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem Rev* 114(13):6733–6778. doi:10.1021/cr400585q
- Varadi M, Kosol S, Lebrun P, Valentini E, Blackledge M, Dunker AK, Felli IC, Forman-Kay JD, Kriwacki RW, Pierattelli R, Sussman J, Svergun DI, Uversky VN, Vendruscolo M, Wishart D, Wright PE, Tompa P (2014) pE-DB: a database of structural ensembles of intrinsically

- disordered and of unfolded proteins. *Nucleic Acids Res* 42(Database Issue):D326–335. doi:10.1093/nar/gkt960
- Varadi M, Guharoy M, Zsolyomi F, Tompa P (2015) DisCons: a novel tool to quantify and classify evolutionary conservation of intrinsic protein disorder. *BMC Bioinf* 16:153. doi:10.1186/s12859-015-0592-2
- Via A, Gould CM, Gemund C, Gibson TJ, Helmer-Citterich M (2009) A structure filter for the Eukaryotic Linear Motif Resource. *BMC Bioinf* 10:351. doi:10.1186/1471-2105-10-351
- Vucetic S, Brown CJ, Dunker AK, Obradovic Z (2003) Flavors of protein disorder. *Proteins* 52(4):573–584. doi:10.1002/prot.10437
- Waizenegger I, Gimenez-Abian JF, Wernic D, Peters JM (2002) Regulation of human separase by securin binding and autocleavage. *Curr Biol* 12(16):1368–1378
- Wang L, Sauer UH (2008) OnD-CRF: predicting order and disorder in proteins using conditional random fields. *Bioinformatics* 24(11):1401–1402. doi:10.1093/bioinformatics/btn132
- Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* 337(3):635–645. doi:10.1016/j.jmb.2004.02.002
- Weinreb PH, Zhen W, Poon AW, Conway KA, Lansbury PT Jr (1996) NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry* 35(43):13709–13715. doi:10.1021/bi961799n
- Wootton JC (1994) Non-globular domains in protein sequences: automated segmentation using complexity measures. *Comput Chem* 18(3):269–285
- Wootton JC, Federhen S (1996) Analysis of compositionally biased regions in sequence databases. *Methods Enzymol* 266:554–571
- Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* 293(2):321–331. doi:10.1006/jmbi.1999.3110
- Xie Q, Arnold GE, Romero P, Obradovic Z, Garner E, Dunker AK (1998) The sequence attribute method for determining relationships between sequence and protein disorder. *Genome Inf* 9:193–200 (Workshop on Genome Informatics)
- Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Uversky VN, Obradovic Z (2007) Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J Proteome Res* 6(5):1882–1898. doi:10.1021/pr060392u
- Xue B, Dunbrack RL, Williams RW, Dunker AK (1804) Uversky VN (2010a) PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta* 4:996–1010. doi:10.1016/j.bbapap.2010.01.011
- Xue B, Williams RW, Oldfield CJ, Dunker AK, Uversky VN (2010) Archaic chaos: intrinsically disordered proteins in Archaea. *BMC Syst Biol* 4(Suppl 1):S1. doi:10.1186/1752-0509-4-S1-S1
- Yang ZR, Thomson R, McNeil P, Esnouf RM (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* 21(16):3369–3376. doi:10.1093/bioinformatics/bti534