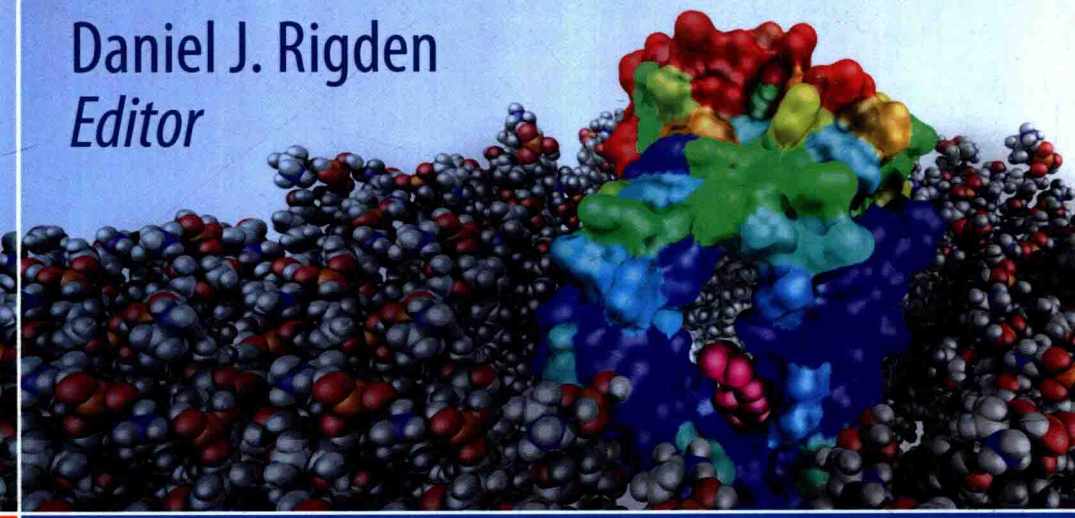
Daniel J. Rigden
*Editor*



# From Protein Structure to Function with Bioinformatics

*Second Edition*

Springer

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



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

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, Dosztanyi 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
Triezenberg 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