Structural Predictions of Intrinsically Disordered Proteins with Computational Methods

E. Abad, G. Rossetti

published in

NIC Symposium 2016
K. Binder, M. Müller, M. Kremer, A. Schnurpfeil (Editors)

Forschungszentrum Jülich GmbH,
John von Neumann Institute for Computing (NIC),
Schriften des Forschungszentrums Jülich, NIC Series, Vol. 48,
ISBN 978-3-95806-109-5, pp. 89.
http://hdl.handle.net/2128/9842

Structural Predictions of Intrinsically Disordered Proteins with Computational Methods

Enrique Abad[1] **and Giulia Rossetti**[1,2]

[1] Computational Biomedicine, Institute for Advanced Simulation IAS-5 and Institute of Neuroscience and Medicine INM-9, Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: e.abad@grs-sim.de

[2] Jülich Supercomputing Centre, Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: g.rossetti@fz-juelich.de

Structural characterisation of intrinsically disordered proteins is highly non-trivial. They exist as dynamic, highly flexible structural ensembles that undergo conformational conversions on a wide range of timescales, spanning from picoseconds to milliseconds. Computational methods may be of great help to characterise these proteins. Here we review recent progress from our lab and other groups to develop and apply *in silico* methods for structural predictions of these highly relevant, challenging systems.

1 Introduction

In the following, we closely follow the more detailed original report in Rossetti, *et al.*[1] (Copyright Elsevier, 2015). Intrinsically Disordered proteins (IDPs) are an important class of functional proteins with high abundance in nature[2–4], specifically in humans, where they represent almost one third of the genome[5,2–4]. Notably IDPs are extensively associated with human diseases and amyloidosis[6,7]. Specifically, 79 % of cancer associated proteins, 57 % of the cardiovascular disease associated proteins and 55 % of neurodegenerative disease associated proteins are predicted to contain 30 or more consecutive disordered residues[8].

Studying the structural determinants of this class of proteins is the key to understand their role for cellular function and dysfunction in both healthy and altered-disease-associated pathways. Unfortunately, traditional computational and experimental approaches have been hampered so far by a variety of challenges. IDPs do not adopt a well-defined native three-dimensional structure[9] and they lack stable tertiary and/or secondary structures when isolated in solution under near-physiological conditions[10] and exist in an ensemble of states both in solution and when unbound to a ligand *in vivo*[10]. This means that an ensemble of inter-converting conformers is required to describe the conformational behaviour of IDPs[11].

Apparently IDPs do not simply occur as filler material amongst functional well-structured proteins, instead they are associated with a variety of biological functions[2]. IDPs are indeed enriched in signalling and regulatory functions because disordered segments permit interaction with several proteins and hence the re-use of the same protein in multiple pathways[12,2,13,14]. Moreover, IDPs can apply different molecular recognition mechanisms and functional modes[15,14] compared to the globular proteins. The majority of IDPs undergo a disorder-to-order transition upon functioning[16,17], a structural transition from a partially disordered state into a more highly ordered conformation in the complex[17,18], also called folding-upon-binding mechanism (Fig. 1A).

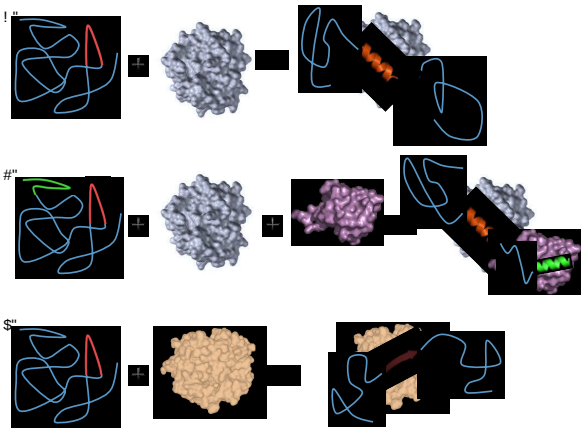

Figure 1. Schematic of IDPs molecular recognition mechanisms. A) disorder-to-order transition[14]. B) Binding commonality[19]. C) Binding diversity[14]. Figure from our original report in Rossetti, *et al.*[1] (Copyright Elsevier, 2015) .

The persistence of natively unfolded proteins throughout evolution may reside in advantages of flexible structure during disorder order transitions in comparison with rigid proteins[10,20,21]. The potential advantages of intrinsic lack of structure and function-related disorder-to-order transitions are:

i) Decoupling of specificity from binding strength. IDPs are capable to combine high specificity with low affinity[10]. This is due to the fact that folding and binding are coupled for IDPs. Therefore the change in enthalpy is compensated by a much larger loss of conformational entropy as compared to globular proteins. This results in a lower absolute value of free-energy, decreasing the stability of the resulting complex[22,3].

ii) Binding commonality in which multiple, distinct sequences fold differently yet each recognises a common binding surface[19] (Fig. 1B). These localised interacting regions allow IDPs to have an increased modularity as different binding regions can be incorporated into the same protein without excessively increasing protein length[23].

iii) Binding diversity. IDPs may folds differently to recognise differently shaped partners by several structural accommodations at the various binding interfaces[10,17,2,21] (Fig. 1C). This phenomenon known as one-to-many signalling, illustrate the complexity of the different binding modes of IDPs and enables an exceptional plasticity in cellular responses[14].

iv) The creation of very large interaction surfaces as the disordered protein wraps-up or surrounds its partner[24] making it possible to overcome steric restrictions[10], meaning that these proteins utilise a much larger fraction of their accessible surfaces compared to globular proteins[25].

v) A increased speed of interaction. IDPs display both faster rates of association by reducing dependence on orientation factors and by enlarging target sizes and faster

rates of dissociation[10]. The great conformational freedom of IDPs in multidirectional search permits the recognition of distant and/or discontinuous determinants on the target[6,15]. Moreover, their extended structure enables them to contact their partner(s) over a large binding surface area, which allows the same interaction potential to be realised by shorter proteins overall[15].

All these features collocate IDPs among the major cellular regulators, recognisers, and signal transducers[11]. Also IDPs reduced life-time in the cell, possibly represents a mechanism of rapid turnover of important regulatory molecules[21].

The inherent flexibility of IDPs calls upon new experimental and computational strategies for studying these proteins, since describing the ensemble of conformations of IDP at atomistic level remains a considerable challenge.

In this review we will first discuss the computational strategies that have been devised to tackle the conformational plasticity of IDPs, complemented by an application from our lab.

2 Computational Methods for IDPs

Computational methods using physics-based empirical molecular mechanics force fields increasingly release critical contributions in providing general insights into the behaviour of IDPs[26,11,22]. However, the dynamic and heterogeneous nature of IDPs presents substantial challenges, in terms of force field accuracy and of conformational sampling capability. MD simulations are indeed sensitive to the choice of the protein force field, which are typically parameterised to reproduce the behaviours of folded proteins rather than IDPs, and thus they may fail to capture important aspects of IDP conformational ensembles[27]. We will therefore describe all the different techniques so far used for IDPs.

2.1 Molecular Dynamics and Monte Carlo Simulations

Molecular dynamics (MD) and Monte Carlo (MC) simulations complement experiments by elucidating chemical details underlying the conformational dynamics of biological macromolecules[28]. Unfortunately, it is extremely difficult to adequately sample the conformational space accessible to IDPs. In details:

- MD in explicit solvent at room temperature is generally insufficient for achieving convergence in simulated structural ensembles of IDPs, due to their large conformational space and the so-called kinetic trapping, i.e., the system tends to be confined to local energy minima[29]. Such minima are separated by free-energy barriers, whose heights are often much larger than the thermal energy available to the system[29]. Therefore MD is not always suitable to sample the dynamical behaviour of IDPs.

- In MC approaches, stochastic conformational searches are used to efficiently sample conformations of the protein chain[30]. MC surmounts energy barriers by moving through successive discrete local minima in the energy landscape. In this way, MC sample all the minima of conformational without seeing the energy barriers[31].

2.2 Enhanced Sampling Techniques

These methods achieve a random walk in the potential-energy space, allowing the system to easily overcome the energy barriers that separate local minima. Three well-known approaches for carrying out generalised ensemble MD or MC simulations are the multi-canonical algorithm[32], the simulated tempering[33] and the replica exchange method[34]. They are very briefly summarised here.

- The multi-canonical algorithm (MUCA) method[32] assigns to each state with energy E a non-Boltzmann weight that is independent of temperature so that a uniform potential energy distribution is obtained ensuring that all the energy states are sampled with the same likelihood. This approach was applied to the coupled folding and binding of an IDP in order to generate the corresponding free-energy landscape[35].

- The simulated tempering (ST) performs a free random walk in temperature space. This random walk, in turn, induces a random walk in the potential energy space and allows the simulation to escape from states of local energy minima. In ST[33], the temperature of the system is randomly switched between several predefined values. ST was applied to study the binding mechanism of two IDPs in combination with classical MD[36].

- The replica-exchange method (REM) uses standard Boltzmann weight factors that are known *a priori*[34]. In this method, a number of non-interacting copies (or replicas) of the original system at different temperatures are simulated in parallel under different conditions[34]; at given time intervals, the simulation conditions are exchanged with a specific transition probability between replica pairs[34]. A variation is the replica exchange solute tempering method, REST2[37], in which only the protein and the ions (i.e. the solute) are simulated at different effective temperatures by applying an appropriate potential energy function to each replica. REM could also be coupled with Monte-Carlo simulations, (REMC)[38,39] to explore the conformation space of IDPs (see a recent application from our lab below).

New generalised-ensemble algorithms could be obtained by combining the merits of the above three methods (reviewed in Ref. 40). Another particularly attractive approach to overcome the sampling bottleneck is to combine large numbers of equilibrium and/or generalised ensemble simulations using network methods based on MC algorithms like Markov State Models recently applied also to IDPs[41,42].

2.3 Solvent Representation

In all the methods discussed in the previous sections, the solvent can be represented either as a continuum model, or as explicit molecules. Traditional explicit solvent protein force fields arguably provide the most realistic description of solvent, but also significantly increase the system size leading to prohibitive computational cost to sufficiently sample the immense conformational space of IDPs[43]. Moreover, explicit solvent force fields are know to have a tendency to over-stabilise helices[44] and overestimate the strength of protein-protein interactions. A substantial reduction in the computational cost could be obtained using implicit solvent models[45]. Recently important advances have been made to greatly

improve the efficiency and achievable accuracy of implicit solvent models based on generalised Born (GB) approximation, however they do not properly describe short-range effects where the detailed interplay of a few non-bulk-like water molecules is important and might be further limited by the specific methodology for calculating the solvation free energy as well as the physical parameters of the solvation model[46]. Despite these caveats, implicit solvent force field has been successfully applied to simulation of regulatory IDPs[47,48].

3 Applications from our Lab

A REMC-derived approach based on an implicit-solvent all-atom potential[49] was recently applied on the disordered N-terminal domain of Prion Protein (PrP)[38]. We were able to predict the conformational ensemble of the wild type (WT) and mutated mouse PrPC N-terminal domain. Importantly, the work shows how pathogenic mutations (PMs) affect the PrPC binding to functional interactors and/or the translocation[38]. In Dibenedetto *et al.*[50], we proposed a computational protocol based on classical MD simulations for investigating how the conformational space of the IDP alpha-synuclein (AS) is affected by the binding of an anti-aggregation drug, dopamine (DOP). Specifically, we analysed the conformational ensemble of AS, alone and in the presence of the drug, with a newly developed tool based on the dihedral angle distributions visited during MD[50]. The latter allows interpreting 2D 1H-15N Heteronuclear multiple-quantum correlation (HMQC) spectra of AS in the presence of the anti-aggregation drug by distinguishing variation of chemical shifts due to direct contacts with the drug from the ones due to conformational changes of the AS induced by long-range effects of the binding. Very recently [under review], we have extended this protocol for the study of the physiological, N-terminally acetylated form of alpha-synculein (AcAS). We have employed the REST2 method exposed in Sec. 2.2, with the same force field as in the work of Dibenedetto *et al.*[50]. Realistic starting geometries from this previous work underwent 15 ns of REST2 simulations with 32 replicas after manually adding the acetyl group to the N-terminus. The obtained ensemble compares well with experimental measurements of local properties (NMR-chemical shifts) as well as global properties (hydrodynamic radius, circular dichroism spectra).

4 Conclusions

We have presented recent computational investigations regarding proteins of biological relevance, whose structure determination poses challenges to experimental techniques. Our studies of IDPs suggest, as already pointed out, that methods at different resolution might give important insights in their biological function as well as ligand binding, a process which is so far not well understood.

Acknowledgements

The authors gratefully acknowledge the computing time granted by the John von Neumann Institute for Computing (NIC) and the JARA-HPC Vergabegremium (VGG) and provided on the supercomputers JUROPA and JURECA at Jülich Supercomputing Centre (JSC).

References

1. G. Rossetti, D. Dibenedetto, V. Calandrini, A. Giorgetti, and P. Carloni, *Structural predictions of neurobiologically relevant G-protein coupled receptors and intrinsically disordered proteins*, Archives of Biochemistry and Biophysics, **582**, 91–100, 2015.
2. H. J. Dyson and P. E. Wright, *Intrinsically unstructured proteins and their functions*, Nat Rev Mol Cell Biol, **6**, no. 3, 197–208, 2005.
3. P. Tompa, *Intrinsically unstructured proteins*, Trends Biochem Sci, **27**, no. 10, 527–33, 2002.
4. V. N. Uversky, *Natively unfolded proteins: a point where biology waits for physics*, Protein Sci, **11**, no. 4, 739–56, 2002.
5. A. K. Dunker, Z. Obradovic, P. Romero, E. C. Garner, and C. J. Brown, *Intrinsic protein disorder in complete genomes*, Genome informatics International Conference on Genome Informatics, **11**, 161–71, 2000.
6. L. M. Iakoucheva, C. J. Brown, J. D. Lawson, Z. Obradović, and A. K. Dunker, *Intrinsic disorder in cell-signaling and cancer-associated proteins*, Journal of Molecular Biology, **323**, no. 3, 573–84, 2002.
7. V. N. Uversky, C. J. Oldfield, and A. K. Dunker, *Intrinsically Disordered Proteins in Human Diseases: Introducing the D 2Concept*, Annual Review of Biophysics, **37**, no. 1, 215–246, 2008.
8. V. N. Uversky, C. J. Oldfield, and A. K. Dunker, *Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling*, Journal of Molecular Recognition, **18**, no. 5, 343–384, 2005.
9. V. N. Uversky and D. Eliezer, *Biophysics of Parkinsons Disease: Structure and Aggregation of-Synuclein*, Current Protein and Peptide Science, **10**, no. 5, 483–499, 2009.
10. A. K. Dunker, J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner, and Z. Obradovic, *Intrinsically disordered protein*, Journal of Molecular Graphics and Modelling, **19**, no. 1, 26–59, 2001.
11. A. K. Dunker, I. Silman, V. N. Uversky, and J. L. Sussman, *Function and structure of inherently disordered proteins*, Current opinion in structural biology, **18**, no. 6, 756–764, 2008.
12. A. K. Dunker, E. Garner, S. Guilliot, P. Romero, K. Albrecht, J. Hart, Z. Obradovic, C. Kissinger, and J. E. Villafranca, *Protein disorder and the evolution of molecular recognition: theory, predictions and observations*, Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing, pp. 473–484, 1998.
13. R. G. Smock and L. M. Gierasch, *Sending Signals Dynamically*, Science, **324**, no. 5924, 198–203, 2009.
14. V. N. Uversky, *Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes*, Chemical Society Reviews, **40**, no. 3, 1623–1634, 2011.
15. P. Tompa, *The interplay between structure and function in intrinsically unstructured proteins*, FEBS Letters, **579**, no. 15, 3346–3354, 2005.

16. A. P. Demchenko, *Recognition between flexible protein molecules: induced and assisted folding*, Journal of molecular recognition, **14**, no. 1, 42–61, 2001.
17. H. J. Dyson and P. E. Wright, *Coupling of folding and binding for unstructured proteins*, Current opinion in structural biology, **12**, no. 1, 54–60, 2002.
18. P. E. Wright and H. J. Dyson, *Linking folding and binding*, Current opinion in structural biology, **19**, no. 1, 31–38, 2009.
19. P. Radivojac, L. M. Iakoucheva, C. J. Oldfield, Z. Obradovic, V. N. Uversky, and A. K. Dunker, *Intrinsic disorder and functional proteomics*, Biophysical journal, **92**, no. 5, 1439–1456, 2007.
20. B. W. Pontius, *Close encounters: why unstructured, polymeric domains can increase rates of specific macromolecular association*, Trends Biochem Sci, **18**, no. 5, 181–6, 1993.
21. P. E. Wright and H. J. Dyson, *Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm*, Journal of Molecular Biology, **293**, no. 2, 321–31, 1999.
22. D. Eliezer, *Biophysical characterization of intrinsically disordered proteins*, Current opinion in structural biology, **19**, no. 1, 23–30, 2009.
23. K. Gunasekaran, C.-J. Tsai, S. Kumar, D. Zanuy, and R. Nussinov, *Extended disordered proteins: targeting function with less scaffold*, Trends in biochemical sciences, **28**, no. 2, 81–85, 2003.
24. Y. Choo and J. W. Schwabe, "All wrapped up", 00/01 1998, 4.
25. B. Mszros, P. Tompa, I. Simon, and Z. Dosztnyi, *Molecular principles of the interactions of disordered proteins*, Journal of Molecular Biology, **372**, no. 2, 549–561, 2007.
26. J. M. Bourhis, B. Canard, and S. Longhi, *Predicting protein disorder and induced folding: from theoretical principles to practical applications*, Current Protein and Peptide Science, **8**, no. 2, 135–149, 2007.
27. S. Piana, K. Lindorff-Larsen, and D. E. Shaw, *How robust are protein folding simulations with respect to force field parameterization?*, Biophysical journal, **100**, no. 9, L47–L49, 2011.
28. D. Frenkel, B. Smit, and M. A. Ratner, *Understanding molecular simulation: from algorithms to applications*, Physics Today, **50**, 66, 1997.
29. H. A. Scheraga, M. Khalili, and A. Liwo, *Protein-folding dynamics: overview of molecular simulation techniques*, Annu. Rev. Phys. Chem., **58**, 57–83, 2007.
30. A. Vitalis and R. V. Pappu, *Methods for Monte Carlo simulations of biomacromolecules*, Annual reports in computational chemistry, **5**, 49–76, 2009.
31. Z. Li and H. A. Scheraga, *Monte Carlo-minimization approach to the multiple-minima problem in protein folding*, Proceedings of the National Academy of Sciences, **84**, no. 19, 6611–6615, 1987.
32. N. Nakajima, H. Nakamura, and A. Kidera, *Multicanonical ensemble generated by molecular dynamics simulation for enhanced conformational sampling of peptides*, The Journal of Physical Chemistry B, **101**, no. 5, 817–824, 1997.
33. E. Marinari and G. Parisi, *Simulated tempering: a new Monte Carlo scheme*, EPL (Europhysics Letters), **19**, no. 6, 451, 1992.
34. Y. Sugita and Y. Okamoto, *Replica-exchange molecular dynamics method for protein folding*, Chemical Physics Letters, **314**, no. 1, 141–151, 1999.

35. J. Higo, Y. Nishimura, and H. Nakamura, *A free-energy landscape for coupled folding and binding of an intrinsically disordered protein in explicit solvent from detailed all-atom computations*, Journal of the American Chemical Society, **133**, no. 27, 10448–10458, 2011.
36. I. Staneva, Y. Huang, Z. Liu, and S. Wallin, *Binding of two intrinsically disordered peptides to a multi-specific protein: a combined Monte Carlo and molecular dynamics study*, PLoS computational biology, **8**, no. 9, e1002682, 2012.
37. L. Wang, R. A. Friesner, and B. J. Berne, *Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2)*, J. Phys. Chem. B, **115**, no. 30, 9431–9438, 2011.
38. X. Cong, N. Casiraghi, G. Rossetti, S. Mohanty, G. Giachin, G. Legname, and P. Carloni, *Role of Prion Disease-Linked Mutations in the Intrinsically Disordered N-Terminal Domain of the Prion Protein*, Journal of Chemical Theory and Computation, **9**, no. 11, 5158–5167, 2013.
39. R. K. Das and R. V. Pappu, *Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues*, Proceedings of the National Academy of Sciences, 2013.
40. Y. Okamoto, *Generalized-ensemble algorithms: enhanced sampling techniques for Monte Carlo and molecular dynamics simulations*, Journal of Molecular Graphics and Modelling, **22**, no. 5, 425–439, 2004.
41. C. M. Baker and R. B. Best, *Insights into the binding of intrinsically disordered proteins from molecular dynamics simulation*, Wiley Interdisciplinary Reviews: Computational Molecular Science, 2013.
42. Q. Qiao, G. R. Bowman, and X. Huang, *Dynamics of an intrinsically disordered protein reveal metastable conformations that potentially seed aggregation*, Journal of the American Chemical Society, **135**, no. 43, 16092–16101, 2013.
43. T. H. Click, D. Ganguly, and J. Chen, *Intrinsically Disordered Proteins in a Physics-Based World*, Int J Mol Sci, **11**, no. 12, 5292–5309, 2010.
44. R. B. Best, N.-V. Buchete, and G. Hummer, *Are current molecular dynamics force fields too helical?*, Biophysical journal, **95**, no. 1, L07–L09, 2008.
45. M. Feig and C. L. Brooks Iii, *Recent advances in the development and application of implicit solvent models in biomolecule simulations*, Current opinion in structural biology, **14**, no. 2, 217–224, 2004.
46. J. Chen and C. L. Brooks, *Critical importance of length-scale dependence in implicit modeling of hydrophobic interactions*, Journal of the American Chemical Society, **129**, no. 9, 2444–2445, 2007.
47. J. Chen, *Intrinsically disordered p53 extreme C-terminus binds to S100B ($\beta\beta$) through "fly-casting"*, Journal of the American Chemical Society, **131**, no. 6, 2088–2089, 2009.
48. D. Ganguly and J. Chen, *Atomistic details of the disordered states of KID and pKID. Implications in coupled binding and folding*, Journal of the American Chemical Society, **131**, no. 14, 5214–5223, 2009.
49. *Folding thermodynamics of peptides*, Biophys J, **88**, no. 3, 1560–9, 2005.
50. D. Dibenedetto, G. Rossetti, R. Caliandro, and P. Carloni, *A molecular dynamics simulations-based interpretation of NMR multidimensional heteronuclear spectra of alpha-synuclein/dopamine adducts*, Biochemistry, **52**, no. 38, 6672–83, 2013.