Computational approaches for inferring the functions of intrinsically disordered proteins

Mihaly Varadi [1,2], Wim Vranken [1,2,3], Mainak Guharoy [1,2] and Peter Tompa [1,2]*

[1] Flemish Institute of Biotechnology, Brussels, Belgium, [2] Department of Structural Biology, VIB, Vrije Universiteit Brussels, Brussels, Belgium, [3] ULB-VUB - Interuniversity Institute of Bioinformatics in Brussels (IB)[2], Brussels, Belgium

Intrinsically disordered proteins (IDPs) are ubiquitously involved in cellular processes and often implicated in human pathological conditions. The critical biological roles of these proteins, despite not adopting a well-defined fold, encouraged structural biologists to revisit their views on the protein structure-function paradigm. Unfortunately, investigating the characteristics and describing the structural behavior of IDPs is far from trivial, and inferring the function(s) of a disordered protein region remains a major challenge. Computational methods have proven particularly relevant for studying IDPs: on the sequence level their dependence on distinct characteristics determined by the local amino acid context makes sequence-based prediction algorithms viable and reliable tools for large scale analyses, while on the structure level the *in silico* integration of fundamentally different experimental data types is essential to describe the behavior of a flexible protein chain. Here, we offer an overview of the latest developments and computational techniques that aim to uncover how protein function is connected to intrinsic disorder.

Keywords: intrinsically disordered proteins, IDP ensembles, IDP function, disorder prediction, protein ensemble database

Introduction

The traditional goal of protein structural biology is to relate the well-defined three-dimensional structure(s) of a protein to its biological function. This structure-function paradigm continues to facilitate many important discoveries, but has largely ignored the possible roles of conformational flexibility on function (Forman-Kay and Mittag, 2013). Yet, in the recent years it became apparent that structural disorder is ubiquitously present in diverse cellular processes, and has a particularly prominent role in regulation and signaling events occurring in the complex cellular environment (Tompa et al., 2006; Dunker et al., 2015). Proteins or protein regions that are enriched in conformational flexibility are referred to as intrinsically disordered proteins (IDPs) or protein regions (IDRs) (Dyson and Wright, 2005; Fink, 2005; Tompa, 2005). IDPs and IDRs lack a well-defined, stable three-dimensional fold, and therefore they populate ensembles of dynamically exchanging conformations, separated by low energy barriers. This dynamic behavior challenges the traditional structure-function paradigm (Wright and Dyson, 1999; Chouard, 2011), since it is far from trivial to describe the structural behavior of proteins that adopt such an extensive range of conformations, let alone infer their biological role (Tompa, 2011). Even though intrinsic disorder is occurring ubiquitously—more than 30% of the proteins in the known eukaryotic proteomes have disordered segments of 30 or more consecutive disordered residues

(Dunker et al., 2000)—we are only beginning to understand how protein function arises from the disordered state (Tompa, 2011). Here, we provide an overview of the recent developments in terms of computational methods and data resources that facilitate the understanding of intrinsic disorder and its connection to protein function.

Functional Consequences of Intrinsic Disorder

IDPs and IDRs can be viewed as having complementary functions to those of their folded counterparts. While the latter are often involved in enzymatic activities, molecular transportation or binding short peptides and small molecules, IDPs are mainly involved in signaling, regulation and enzymatic activity inhibition (Xie et al., 2007; Dunker et al., 2015), for example in cell cycle regulation (Yoon et al., 2012), cell division and differentiation (Ward et al., 2004; Xie et al., 2007).

There are a number of possible ways in which an IDP/IDR can realize its function. In perhaps the simplest scenarios, they serve as entropic chains, effectively influencing the orientation and distance between folded domains (Chong et al., 2010), and organizing the super-tertiary structure of the protein (Tompa, 2012). In some cases they are entropic springs or even timers, where the length and flexibility of the linker can determine stochastically how often two folded domains may encounter each other (Bentrop et al., 2001; Smagghe et al., 2010).

Another important role of conformational flexibility is in binding protein or nucleic acid partners. IDPs excel in establishing specific, but transient interactions (Dunker et al., 1998). From an energetic point of view, the reason behind their weaker binding affinities is that the entropic cost of stabilizing a single conformation from the dynamic ensemble that the IDP/IDR is sampling is relatively high (Dyson and Wright, 2005). However, in some cases the fine-tuning of favorable interactions is known to yield surprisingly strong affinities (Ferreon et al., 2013; Follis et al., 2013). An additional advantage of the high degree of conformational freedom is that an IDR can bind very diverse partners because it can easily adopt different conformations (Wang et al., 2011; Hsu et al., 2013) and is often enriched in short binding- and recognition motifs. It is therefore no surprise that IDPs are often hub- (Kim et al., 2008) or scaffold proteins (Dyson and Wright, 2005; Kim et al., 2008; Mittag et al., 2010a) that play essential roles in the cell by integrating signals (Lobley et al., 2007), so increasing the complexity of cellular networks (Dunker et al., 2005; Oldfield et al., 2008). Consequently, IDPs are often implicated in pathological conditions where loss of regulation is the major issue, such as different types of cancer (Andresen et al., 2012). Their involvement in diseases has recently turned IDPs into potential drug targets by either targeting the IDP, or its protein-protein interactions (Funk and Galloway, 1998; Metallo, 2010; Rezaei-Ghaleh et al., 2012).

Conformational flexibility implies high accessibility for potential binding partners and/or enzymes. Consequently, post-translational modification (PTM) sites are often found to be enriched in intrinsically disordered regions (Iakoucheva et al., 2004), with especially phosphorylation sites being prevalent (Gao et al., 2010). While IDPs often go through disorder-to-order transitions upon binding to their partners (Mohan et al., 2006; Wright and Dyson, 2009), in many cases they remain partially or fully flexible in their bound state, forming fuzzy complexes (Tompa and Fuxreiter, 2008; Fuxreiter and Tompa, 2012). One of the advantages of this fuzziness is that PTM sites within the chain can remain relatively accessible, allowing easier regulation of the IDP by modification enzymes (Mittag et al., 2010a). Such regulation by PTM sites is not limited to activation/deactivation of the protein; the modification of the surface of the IDR may also be the prerequisite of binding to a different partner (Oldfield et al., 2008), or even to the same partner, but with increased affinity (Mittag et al., 2010b).

However, intrinsic disorder also has a dark side. In particular, the amino acid compositional bias of IDPs coupled with relatively high propensities to form β-sheets and turns leads to elevated aggregation potentials, and the formation of amyloid-type beta-structures (Levine et al., 2015). Indeed, IDPs have been implicated in aggregation-based diseases, such as Alzheimer's and Parkinson's (Huang and Stultz, 2009; Uversky, 2010).

Sequence-based Investigation of IDPs

There are a number of experimental techniques currently available for identifying and characterizing intrinsic disorder, such as circular dichroism (CD) (Weinreb et al., 1996), protease digestion (Johnson et al., 2012), Förster resonance energy transfer (FRET) (Haas, 2012), Electron Paramagnetic Resonance (EPR) spectroscopy (Drescher, 2012), small-angle X-ray and neutron scattering (SAXS and SANS) (Bernado and Svergun, 2012; Gabel, 2012) and nuclear magnetic resonance spectroscopy (NMR) (Kosol et al., 2013; Konrat, 2014). For initial and for high-throughput investigations, computational methods are however a very popular choice (Ward et al., 2004; Ishida and Kinoshita, 2007). Intrinsic disorder is associated with distinct sequence characteristics; IDPs/IDRs are enriched in "disorder promoting" amino acids, such as charged or polar residues, glycines and prolines, while hydrophobic residues are underrepresented (Uversky et al., 2000). Their conformational flexibility also implies that the local sequence context predominantly dictates the amino acid interactions that can take place, making IDPs more amendable to prediction of their characteristics from sequence. Throughout the last decade many disorder prediction algorithms were designed to exploit the information contained within the amino acid sequence of an IDP; there are more than 50 disorder predictors worldwide (He et al., 2009). The first disorder predictors, such as DisEMBL (Linding et al., 2003) were primarily based on the distinct compositional bias of IDPs. They were followed by faster and more reliable algorithms, such as IUPred (Dosztanyi et al., 2005), RONN (Yang et al., 2005), and Espritz (Walsh et al., 2012). Some of these more advanced methods rely on machine learning techniques (Bellay et al., 2012), or combine the results of several algorithms, such as the meta-predictor metaPrDOS (Ishida and Kinoshita, 2008). Overall, the accuracy of most predictors is consistently above 80%, with

the best methods currently peaking around 85% (Monastyrskyy et al., 2014). Alternatively, the novel Dynamine approach predicts backbone dynamics, which correlates (negatively) with intrinsic disorder (Cilia et al., 2014); interestingly, this approach is trained on estimations directly from NMR data and avoids structure-based information, complex machine-learning and evolutionary information (Cilia et al., 2013). The distribution of charged amino acids in the sequence of an IDP can also offer information on whether the protein chain is extended or collapsed (Das and Pappu, 2013).

Often it is unnecessary even to predict disorder, since there are a number of openly accessible online resources that store information of the disorder content of specific proteins. The Disordered Protein Database (DisProt) is the primary one of these sequence-based resources (Sickmeier et al., 2007). DisProt is manually curated, and stores information on proteins for which intrinsic disorder was experimentally determined. Where available, the proteins are also annotated with their known functions. However, DisProt houses data for 694 disordered proteins, which is only a minor fraction of the expected number of IDPs. MobiDB (Potenza et al., 2015) and D2P2 (Oates et al., 2013) on the other hand are online resources that store IDPs identified using prediction algorithms from the whole UniProt in addition to experimentally determined ones.

Sequence information on intrinsic disorder can be exploited for more than merely the prediction of disordered residues. In fact, there are many recent algorithms that aim at predicting functional sites and/or the functional role of IDRs. For example, larger hydrophobic residues such as tryptophan and leucine are often found within peptide motifs that act as recognition units located within IDR segments, called molecular recognition features (MoRFs) (Mohan et al., 2006; Fuxreiter et al., 2007; Brown et al., 2010). Disordered motifs are generally short, 3-15 residue long segments; therefore identifying them poses a computational challenge (Gould et al., 2010). Consequently, predicting functional sites in IDRs is not straightforward, and prone to high false positive rates (Tompa, 2011). Additional layers of information can enhance the performance, for example MoRFpred, which uses order/disorder patterns (Cheng et al., 2007), ANCHOR, which estimates the interaction of a segment with a general partner (Meszaros et al., 2009) or DisCons, which takes into consideration the evolutionary conservation of both the amino acid sequence and of the disorder as a feature (Varadi et al., 2015).
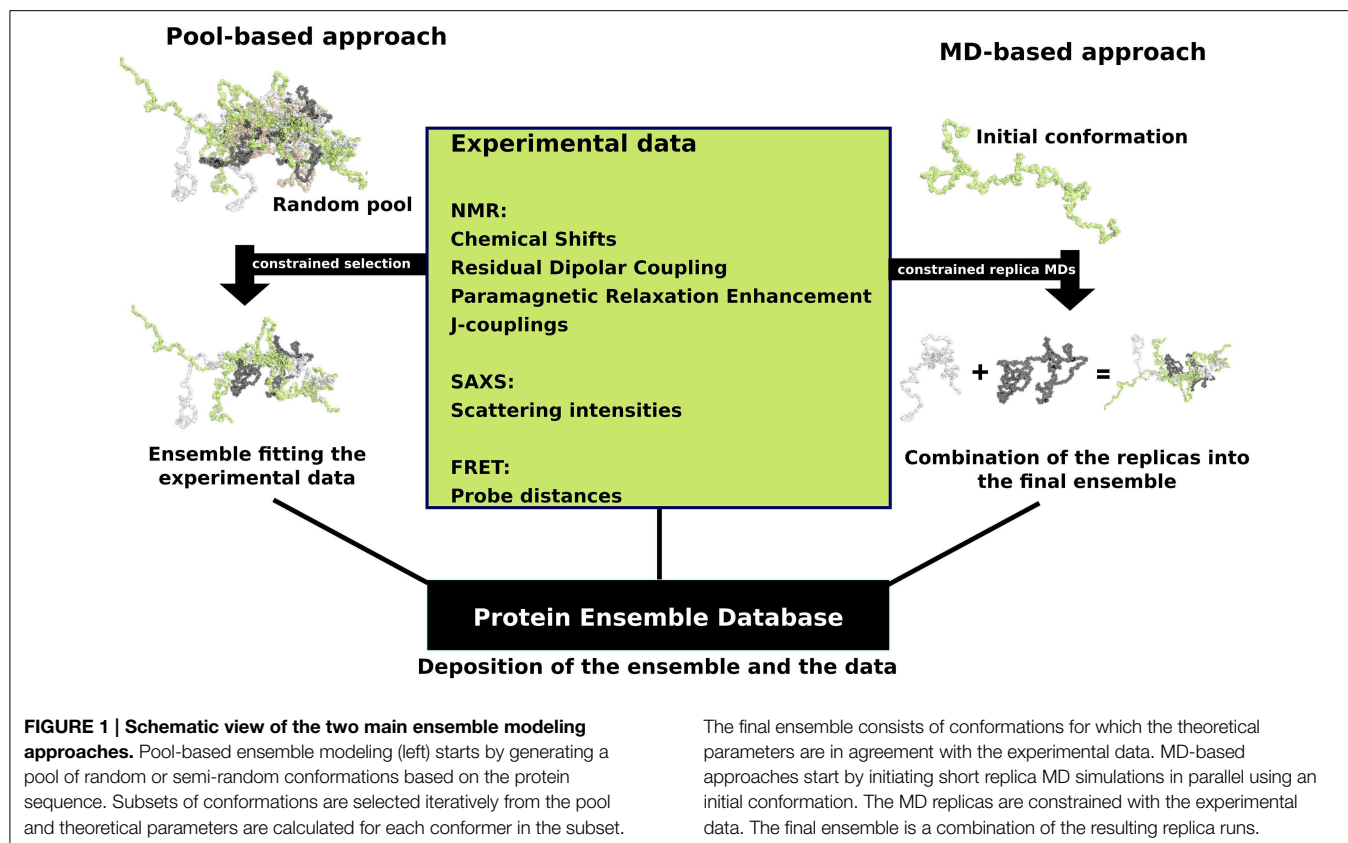
Structural Representation of IDPs

Ideally it would be possible to describe the structure of an IDP/IDR in full atomic detail. Indeed, the set of conformations IDPs sample is often not completely random, for example both p21 and p27 are known to sample distinct secondary structural elements that are biologically relevant and involved in conformation selection (Kriwacki et al., 1996; Sivakolundu et al., 2005). Due to their inherent conformational flexibility, however, the structure of an IDP cannot be described with a single, static conformation (Tompa and Varadi, 2014). This conformational diversity of IDPs precludes crystallization and examination by

X-ray crystallography is therefore not a viable option. NMR spectroscopy, while more attuned to conformational diversity, remains hindered by distinct difficulties such as peak overlap (Bellay et al., 2012), while traditional structure calculation protocols do not properly account for multiple conformations (Vranken, 2014).

In response to this challenge, a number of approaches were developed that combine experimental data with computational methodology with the aim to accurately describe the full conformational ensemble adopted by IDPs. Experimental data from techniques that rely on measurements performed in solution are particularly well suited for studying the dynamic structure of an IDP, even though they often represent an average over the different conformations that are adopted by the IDP. These experimental measurements predominantly include NMR-derived parameters, such as chemical shifts (CSs) (Jensen et al., 2011), residual dipolar couplings (RDCs) (Mittag et al., 2010b), paramagnetic relaxation enhancements (PREs) (Mittag et al., 2010b), and J-couplings (Mittag et al., 2010b), as well as scattering intensities from small-angle X-ray scattering (SAXS) (Allison et al., 2009) and probe distances from Forster resonance energy transfer (FRET) (Haas, 2012). These experimental data are then combined with computational methods to determine an ensemble of conformations for an IDP, with two main approaches being used; the first approach is referred to as pool-based modeling, while the second one is based on molecular dynamics (MD) simulations (Tompa and Varadi, 2014) (**Figure 1**).

When using a pool-based approach, such as the ensemble optimization method (EOM) (Bernado et al., 2007) which was designed to model ensembles based on SAXS data, the initial step is to generate a very large random or semi-random pool of conformations based on the amino acid sequence of the IDP/IDR with algorithms such as Flexible-Meccano (Ozenne et al., 2012). The sampling of conformations can be biased using experimental data, such as secondary structure propensities derived from chemical shifts. Next, theoretical parameters are estimated for each conformer in the pool, for example the theoretical scattering intensities calculated by software such as CRYSOL (Bernado and Svergun, 2012). Selection algorithms are then deployed in order to select subsets of the pool in a way that when the theoretical parameters are averaged over a subset, they are in excellent agreement with the experimental data (Sibille and Bernado, 2012). Another example is the ENSEMBLE methodology, which uses as input a large set of conformations together with relevant experimental data, and then prunes the ensemble of conformations to a smaller subset. During the filtering step, conformations are assigned weights so that the resulting ensemble average values fit the experimental input values. Structures that do not contribute to this fitting are discarded (Krzeminski et al., 2013).

In contrast, ensemble modeling procedures are based on molecular dynamics (MD) simulations and begin with random conformations in parallel. Multiple "replica" simulations are initiated using these initial conformations, and constraints are applied over multiple models based on the experimental data, i.e., sets of conformations are required to satisfy experimentally determined constraints, such as pair-wise

FIGURE 1 | Schematic view of the two main ensemble modeling approaches. Pool-based ensemble modeling (left) starts by generating a pool of random or semi-random conformations based on the protein sequence. Subsets of conformations are selected iteratively from the pool and theoretical parameters are calculated for each conformer in the subset. The final ensemble consists of conformations for which the theoretical parameters are in agreement with the experimental data. MD-based approaches start by initiating short replica MD simulations in parallel using an initial conformation. The MD replicas are constrained with the experimental data. The final ensemble is a combination of the resulting replica runs.

distances or secondary structure propensities (Cavalli et al., 2013). Given the conformational heterogeneity of IDPs, which sample an extensive range of conformations during their biological lifetime, extensive simulations are required to ensure adequate sampling of relevant regions of the conformational space. This significantly increases the computational costs associated with IDP simulations and makes it difficult to achieve even for systems of modest size. Recent developments to address these issues include techniques such as multi-scale enhanced sampling (MSES) (Lee and Chen, 2015) and replica exchange with guided annealing (RE-GA) (Zhang and Chen, 2014). The MSES protocol combines coarse-grained, topology-based models with atomistic force fields to enhance sampling and was recently optimized for simulating IDP conformational ensembles, where it could capture reversible helix-coil transitions (Lee and Chen, 2015). RE-GA has been suggested to be suitable for systems with small conformational transition barriers (as is the case for IDPs), and helped the disordered kinase inducible domain (KID) protein to efficiently escape non-specific compact states while requiring less computation (Zhang and Chen, 2014).

Therefore, the main differences between these two classes of modeling techniques are that the pool-based approach is much faster; conformations can be easily generated, but the final results strongly depend on the quality and diversity of this initial pool of conformations. The ensemble modeling MD approaches are in contrast very slow; conformational sampling depends on the MD simulation, but they have the advantage that the experimental data are continuously applied, a timeline of conformational changes is available, and due to their more rigorous simulation of physical reality, they should give a better representation of the thermally accessible structural ensemble.

These two modeling approaches (i.e., pool and MD-based) are currently the state-of-the-art and have been applied to generate the structural ensemble of many IDPs (**Table 1**). These ensemble models are, however, not straightforward to interpret (Tompa and Varadi, 2014). The key issue is that the experimental information that is available to either filter or constrain during the calculation is very sparse compared to the immense degree of conformational freedom the IDP experiences in solution, resulting in a hugely underdetermined problem. As a direct consequence, many ensembles of models can describe the experimental data equally well, allowing multiple, ambiguous solutions that strongly depend on the calculation approach and the amount and type of experimental data available. In fact, one can model the ensemble of an IDP with an excellent fit to the data, then discard the ensemble and remodel another, unique and different ensemble with an equally good fit. In this sense, ensembles should be considered as a whole and their structural characteristics analyzed as the average over the ensemble, and over-interpretation of single conformations should be avoided. However, if certain characteristics or pre-formed secondary structural elements are consistently modeled in multiple ensembles, then such structural features might be functionally relevant.

Protein	Data type	Protocol	PED ID	References
Sic1/Cdc4	NMR and SAXS	Pool-based	PED9AAA	Mittag et al., 2010b
p15 PAF	NMR and SAXS	Pool-based	PED6AAA	De Biasio et al., 2014
MKK7	NMR	Pool-based	PED5AAB	Kragelj et al., 2015
Beta-synuclein	NMR	MD-based	PED1AAD	Allison et al., 2014
P27 KID	NMR	MD-based	PED2AAA	Sivakolundu et al., 2005

The field of ensemble modeling therefore still presents exciting opportunities for further development, and several important issues will have to be addressed before the techniques become more reliable. In our view, the first is increasing the number of experimentally derived constraints, which will lead to higher quality models, or cross-validation with new types of experimental data, which will also increase the power of ensembles. The second is the further incorporation of knowledge based information into the calculations, such as the results from reliable predictions or improved force fields. The third is that specific validation and evaluation approaches are required for these ensembles, likely starting from the currently well-developed NMR validation field (Rosato et al., 2013; Vuister et al., 2014) but with better accounting for multiple conformations (Vranken, 2014). Especially NMR CS values, whenever available, are very useful for the estimation of residue-level backbone and side-chain dynamics (Berjanskii and Wishart, 2005, 2013) as well as secondary structure populations (Shen and Sali, 2006; Camilloni et al., 2012), with new methods providing reference chemical shift values for IDPs (Tamiola et al., 2010). They are already effectively used to generate pools with predetermined conformations and as restraints in molecular dynamics simulations (Krieger et al., 2014), but have immense potential for the validation and evaluation of ensembles. Finally, the overwhelming majority of the already generated ensemble models were previously unavailable to the scientific community, impeding the establishment of standardized validation and evaluation protocol. The Protein Ensemble Database (PED) is an international initiative launched to address this issue, effectively making the experimental data and the ensemble models available to the scientific community (Varadi et al., 2014). This is expected to facilitate the development of the next generation of ensemble modeling techniques, and should provide a basis for defining standards of validation and evaluation.

Toward the Functional Interpretation of IDP Ensembles

While ensemble models do not yet possess a predictive power comparable to that of the structure of folded proteins/domains, these models can already offer insights regarding the function of an IDP. Through the integration of experimental data into an ensemble model, functionally important segments might be inferred. For example, transient secondary structural elements in the ensemble of an IDP are often important in terms of function. Such pre-formed elements are often molecular recognition units, playing major roles in binding to various partners. For example, thep27 protein samples transient helices are consistent with the secondary structure of the bound state p27-Cdk2-cyclin (Sivakolundu et al., 2005). Therefore, in accord with the notion of conformational selection, if a certain secondary structural element is sampled in the ensemble, it might be functionally relevant in the bound form as well (Yoon et al., 2012). Again, such interpretations have to be treated with caution given that the ensembles are based on lower resolution, averaged experimental observations, and the ensemble models should therefore be accurate on average, not by single conformations.

Conclusions

IDPs are involved ubiquitously in biological processes, and play essential roles in the regulation of complex cellular systems (Tompa et al., 2006; Dunker et al., 2015). These multi-purpose proteins combine conformational flexibility with an enrichment of binding motifs and post-translational modification sites (Iakoucheva et al., 2004; Wang et al., 2011; Hsu et al., 2013). Due to their biological importance, it is imperative to characterize them and attempt to relate their sequence and structure with their physiological roles (Tompa, 2011). Such an endeavor has the potential to offer valuable insights that can be translated into new drugs and therapies (Funk and Galloway, 1998; Metallo, 2010; Rezaei-Ghaleh et al., 2012). Sequence-based *in silico* techniques such as disorder prediction algorithms are already comparable in terms of accuracy to that of the secondary structure prediction algorithms of folded proteins (Monastyrskyy et al., 2014), and functional prediction algorithms are also widely available (Cheng et al., 2007; Meszaros et al., 2009; Varadi et al., 2015). Yet, a major breakthrough is expected when ensemble models based on diverse experimental data prove to be biologically relevant, so that we can confidently infer specific protein function from the structural representation of an IDP (Tompa and Varadi, 2014). However, in order to realize this goal a number of challenges need to be tackled first (Tompa, 2011). Chief among these issues are improving the amount and available types of experimental data and establishing standardized protocols for the validation and evaluation of the ensemble modeling procedures. Only then can the field advance in terms of increasing the predictive power of ensemble models (Tompa and Varadi, 2014).

Acknowledgments

References

Allison, J. R., Rivers, R. C., Christodoulou, J. C., Vendruscolo, M., and Dobson, C. M. (2014). A relationship between the transient structure in the monomeric state and the aggregation propensities of alpha-synuclein and beta-synuclein. *Biochemistry* 53, 7170–7183. doi: 10.1021/bi5009326

Allison, J. R., Varnai, P., Dobson, C. M., and Vendruscolo, M. (2009). Determination of the free energy landscape of alpha-synuclein using spin label nuclear magnetic resonance measurements. *J. Am. Chem. Soc.* 131, 18314–18326. doi: 10.1021/ja904716h

Andresen, C., Helander, S., Lemak, A., Farès, C., Csizmok, V., Carlsson, J., et al. (2012). Transient structure and dynamics in the disordered c-Myc transactivation domain affect Bin1 binding. *Nucleic Acids Res.* 40, 6353–6366. doi: 10.1093/nar/gks263

Bellay, J., Michaut, M., Kim, T., Han, S., Colak, R., Myers, C. L., et al. (2012). An omics perspective of protein disorder. *Mol. Biosyst.* 8, 185–193. doi: 10.1039/C1MB05235G

Bentrop, D., Beyermann, M., Wissmann, R., and Fakler, B. (2001). NMR structure of the "ball-and-chain" domain of KCNMB2, the beta 2-subunit of large conductance Ca2+- and voltage-activated potassium channels. *J. Biol. Chem.* 276, 42116–42121. doi: 10.1074/jbc.M107118200

Berjanskii, M. V., and Wishart, D. S. (2005). A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc.* 127, 14970–14971. doi: 10.1021/ja054842f

Berjanskii, M. V., and Wishart, D. S. (2013). A simple method to measure protein side-chain mobility using NMR chemical shifts. *J. Am. Chem. Soc.* 135, 14536–14539. doi: 10.1021/ja407509z

Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and Svergun, D. I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* 129, 5656–5664. doi: 10.1021/ja069124n

Bernadó, P., and Svergun, D. I. (2012). Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol. Biosyst.* 8, 151–167. doi: 10.1039/C1MB05275F

Brown, C. J., Johnson, A. K., and Daughdrill, G. W. (2010). Comparing models of evolution for ordered and disordered proteins. *Mol. Biol. Evol.* 27, 609–621. doi: 10.1093/molbev/msp277

Camilloni, C., De Simone, A., Vranken, W. F., and Vendruscolo, M. (2012). Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry* 51, 2224–2231. doi: 10.1021/bi3001825

Cavalli, A., Camilloni, C., and Vendruscolo, M. (2013). Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *J. Chem. Phys.* 138, 094112. doi: 10.1063/1.4793625

Cheng, Y., Oldfield, C. J., Meng, J., Romero, P., Uversky, V. N., and Dunker, A. K. (2007). Mining alpha-helix-forming molecular recognition features with cross species sequence alignments. *Biochemistry* 46, 13468–13477. doi: 10.1021/bi7012273

Chong, P. A., Lin, H., Wrana, J. L., and Forman-Kay, J. D. (2010). Coupling of tandem Smad ubiquitination regulatory factor (Smurf) WW domains modulates target specificity. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18404–18409. doi: 10.1073/pnas.1003023107

Chouard, T. (2011). Structural biology: breaking the protein rules. *Nature* 471, 151–153. doi: 10.1038/471151a

Cilia, E., Pancsa, R., Tompa, P., Lenaerts, T., and Vranken, W. F. (2013). From protein sequence to dynamics and disorder with DynaMine. *Nat. Commun.* 4, 2741. doi: 10.1038/ncomms3741

Cilia, E., Pancsa, R., Tompa, P., Lenaerts, T., and Vranken, W. F. (2014). The DynaMine webserver: predicting protein dynamics from sequence. *Nucleic Acids Res.* 42, W264–W270. doi: 10.1093/nar/gku270

Das, R. K., and Pappu, R. V. (2013). Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U.S.A.* 110, 13392–13397. doi: 10.1073/pnas.1304749110

De Biasio, A., Ibáñez de Opakua, A., Cordeiro, T. N., Villate, M., Merino, N., Sibille, N., et al. (2014). p15PAF is an intrinsically disordered protein with nonrandom structural preferences at sites of interaction with other proteins. *Biophys. J.* 106, 865–874. doi: 10.1016/j.bpj.2013.12.046

Dosztányi, Z., Csizmók, V., Tompa, P., and Simon, I. (2005). The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J. Mol. Biol.* 347, 827–839. doi: 10.1016/j.jmb.2005.01.071

Drescher, M. (2012). EPR in protein science: intrinsically disordered proteins. *Top. Curr. Chem.* 321, 91–119. doi: 10.1007/128_2011_235

Dunker, A. K., Bondos, S. E., Huang, F., and Oldfield, C. J. (2015). Intrinsically disordered proteins and multicellular organisms. *Semin. Cell Dev. Biol.* 37, 44–55. doi: 10.1016/j.semcdb.2014.09.025

Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M., and Uversky, V. N. (2005). Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* 272, 5129–5148. doi: 10.1111/j.1742-4658.2005.04948.x

Dunker, A. K., Garner, E., Guilliot, S., Romero, P., Albrecht, K., Hart, J., et al. (1998). Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pac. Symp. Biocomput.* 473–484.

Dunker, A. K., Obradovic, Z., Romero, P., Garner, E. C., and Brown, C. J. (2000). Intrinsic protein disorder in complete genomes. *Genome Inform. Ser. Workshop Genome Inform.* 11, 161–171.

Dyson, H. J., and Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. Nature reviews. *Mol. Cell Biol.* 6, 197–208. doi: 10.1038/nrm1589

Ferreon, A. C., Ferreon, J. C., Wright, P. E., and Deniz, A. A. (2013). Modulation of allostery by protein intrinsic disorder. *Nature* 498, 390–394. doi: 10.1038/nature12294

Fink, A. L. (2005). Natively unfolded proteins. *Curr. Opin. Struct. Biol.* 15, 35–41. doi: 10.1016/j.sbi.2005.01.002

Follis, A. V., Chipuk, J. E., Fisher, J. C., Yun, M. K., Grace, C. R., Nourse, A., et al. (2013). PUMA binding induces partial unfolding within BCL-xL to disrupt p53 binding and promote apoptosis. *Nat. Chem. Biol.* 9, 163–168. doi: 10.1038/nchembio.1166

Forman-Kay, J. D., and Mittag, T. (2013). From sequence and forces to structure, function, and evolution of intrinsically disordered proteins. *Structure* 21, 1492–1499. doi: 10.1016/j.str.2013.08.001

Funk, J. O., and Galloway, D. A. (1998). Inhibiting CDK inhibitors: new lessons from DNA tumor viruses. *Trends Biochem. Sci.* 23, 337–341. doi: 10.1016/S0968-0004(98)01242-0

Fuxreiter, M., and Tompa, P. (2012). Fuzzy complexes: a more stochastic view of protein function. *Adv. Exp. Med. Biol.* 725, 1–14. doi: 10.1007/978-1-4614-0659-4_1

Fuxreiter, M., Tompa, P., and Simon, I. (2007). Local structural disorder imparts plasticity on linear motifs. *Bioinformatics* 23, 950–956. doi: 10.1093/bioinformatics/btm035

Gabel, F. (2012). Small angle neutron scattering for the structural study of intrinsically disordered proteins in solution: a practical guide. *Methods Mol. Biol.* 896, 123–135. doi: 10.1007/978-1-4614-3704-8_8

Gao, J., Thelen, J. J., Dunker, A. K., and Xu, D. (2010). Musite, a tool for global prediction of general and kinase-specific phosphorylation sites. *Mol. Cell. Proteom.* 9, 2586–2600. doi: 10.1074/mcp.M110.001388

Gould, C. M., Diella, F., Via, A., Puntervoll, P., Gemund, C., Chabanis-Davidson, S., et al. (2010). ELM: the status of the 2010 eukaryotic linear motif resource. *Nucleic Acids Res.* 38, D167–D180. doi: 10.1093/nar/gkp1016

Haas, E. (2012). Ensemble FRET methods in studies of intrinsically disordered proteins. *Methods Mol. Biol.* 895, 467–498. doi: 10.1007/978-1-61779-927-3_28

He, B., Wang, K., Liu, Y., Xue, B., Uversky, V. N., and Dunker, A. K. (2009). Predicting intrinsic disorder in proteins: an overview. *Cell Res.* 19, 929–949. doi: 10.1038/cr.2009.87

Hsu, W. L., Oldfield, C. J., Xue, B., Meng, J., Huang, F., Romero, P., et al. (2013). Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding. *Prot. Sci.* 22, 258–273. doi: 10.1002/pro.2207

Huang, A., and Stultz, C. M. (2009). Finding order within disorder: elucidating the structure of proteins associated with neurodegenerative disease. *Future Med. Chem.* 1, 467–482. doi: 10.4155/fmc.09.40

Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z., et al. (2004). The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* 32, 1037–1049. doi: 10.1093/nar/gkh253

Ishida, T., and Kinoshita, K. (2007). PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* 35, W460–W464. doi: 10.1093/nar/gkm363

Ishida, T., and Kinoshita, K. (2008). Prediction of disordered regions in proteins based on the meta approach. *Bioinformatics* 24, 1344–1348. doi: 10.1093/bioinformatics/btn195

Jensen, M. R., Communie, G., Ribeiro, E. A. Jr., Martinez, N., Desfosses, A., Salmon, L., et al. (2011). Intrinsic disorder in measles virus nucleocapsids. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9839–9844. doi: 10.1073/pnas.1103270108

Johnson, D. E., Xue, B., Sickmeier, M. D., Meng, J., Cortese, M. S., Oldfield, C. J., et al. (2012). High-throughput characterization of intrinsic disorder in proteins from the Protein Structure Initiative. *J. Struct. Biol.* 180, 201–215. doi: 10.1016/j.jsb.2012.05.013

Kim, P. M., Sboner, A., Xia, Y., and Gerstein, M. (2008). The role of disorder in interaction networks: a structural analysis. *Mol. Syst. Biol.* 4, 179. doi: 10.1038/msb.2008.16

Konrat, R. (2014). NMR contributions to structural dynamics studies of intrinsically disordered proteins. *J. Magn. Reson.* 241, 74–85. doi: 10.1016/j.jmr.2013.11.011

Kosol, S., Contreras-Martos, S., Cedeño, C., and Tompa, P. (2013). Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules* 18, 10802–10828. doi: 10.3390/molecules180910802

Kragelj, J., Palencia, A., Nanao, M. H., Maurin, D., Bouvignies, G., Blackledge, M., et al. (2015). Structure and dynamics of the MKK7-JNK signaling complex. *Proc. Natl. Acad. Sci. U.S.A.* 112, 3409–3414. doi: 10.1073/pnas.1419528112

Krieger, J. M., Fusco, G., Lewitzky, M., Simister, P. C., Marchant, J., Camilloni, C., et al. (2014). Conformational recognition of an intrinsically disordered protein. *Biophys. J.* 106, 1771–1779. doi: 10.1016/j.bpj.2014.03.004

Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I., and Wright, P. E. (1996). Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc. Natl. Acad. Sci. U.S.A.* 93, 11504–11509. doi: 10.1073/pnas.93.21.11504

Krzeminski, M., Marsh, J. A., Neale, C., Choy, W. Y., and Forman-Kay, J. D. (2013). Characterization of disordered proteins with ENSEMBLE. *Bioinformatics* 29, 398–399. doi: 10.1093/bioinformatics/bts701

Lee, K. H., and Chen, J. (2015). Multiscale enhanced sampling of intrinsically disordered protein conformations. *J. Comp. Chem.* doi: 10.1002/jcc.23957. [Epub ahead of print].

Levine, Z. A., Larini, L., LaPointe, N. E., Feinstein, S. C., and Shea, J. E. (2015). Regulation and aggregation of intrinsically disordered peptides. *Proc. Natl. Acad. Sci. U.S.A.* 112, 2758–2763. doi: 10.1073/pnas.1418155112

Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J., and Russell, R. B. (2003). Protein disorder prediction: implications for structural proteomics. *Structure* 11, 1453–1459. doi: 10.1016/j.str.2003.10.002

Lobley, A., Swindells, M. B., Orengo, C. A., and Jones, D. T. (2007). Inferring function using patterns of native disorder in proteins. *PLoS Comput. Biol.* 3:e162. doi: 10.1371/journal.pcbi.0030162

Mészáros, B., Simon, I., and Dosztányi, Z. (2009). Prediction of protein binding regions in disordered proteins. *PLoS Comput. Biol.* 5:e1000376. doi: 10.1371/journal.pcbi.1000376

Metallo, S. J. (2010). Intrinsically disordered proteins are potential drug targets. *Curr. Opin. Chem. Biol.* 14, 481–488. doi: 10.1016/j.cbpa.2010.06.169

Mittag, T., Kay, L. E., and Forman-Kay, J. D. (2010a). Protein dynamics and conformational disorder in molecular recognition. *J. Mol. Recogn.* 23, 105–116. doi: 10.1002/jmr.961

Mittag, T., Marsh, J., Grishaev, A., Orlicky, S., Lin, H., Sicheri, F., et al. (2010b). Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase. *Structure* 18, 494–506. doi: 10.1016/j.str.2010.01.020

Mohan, A., Oldfield, C. J., Radivojac, P., Vacic, V., Cortese, M. S., Dunker, A. K., et al. (2006). Analysis of molecular recognition features (MoRFs). *J. Mol. Biol.* 362, 1043–1059. doi: 10.1016/j.jmb.2006.07.087

Monastyrskyy, B., Kryshtafovych, A., Moult, J., Tramontano, A., and Fidelis, K. (2014). Assessment of protein disorder region predictions in CASP10. *Proteins* 82(Suppl. 2), 127–137. doi: 10.1002/prot.24391

Oates, M. E., Romero, P., Ishida, T., Ghalwash, M., Mizianty, M. J., Xue, B., et al. (2013). D(2)P(2): database of disordered protein predictions. *Nucleic Acids Res.* 41, D508–D516. doi: 10.1093/nar/gks1226

Oldfield, C. J., Meng, J., Yang, J. Y., Yang, M. Q., Uversky, V. N., and Dunker, A. K. (2008). Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics* 9(Suppl. 1):S1. doi: 10.1186/1471-2164-9-S1-S1

Ozenne, V., Bauer, F., Salmon, L., Huang, J. R., Jensen, M. R., Segard, S., et al. (2012). Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics* 28, 1463–1470. doi: 10.1093/bioinformatics/bts172

Potenza, E., Di Domenico, T., Walsh, I., and Tosatto, S. C. (2015). MobiDB 2.0: an improved database of intrinsically disordered and mobile proteins. *Nucleic Acids Res.* 43, D315–D320. doi: 10.1093/nar/gku982

Rezaei-Ghaleh, N., Blackledge, M., and Zweckstetter, M. (2012). Intrinsically disordered proteins: from sequence and conformational properties toward drug discovery. *Chembiochem* 13, 930–950. doi: 10.1002/cbic.201200093

Rosato, A., Tejero, R., and Montelione, G. T. (2013). Quality assessment of protein NMR structures. *Curr. Opin. Struct. Biol.* 23, 715–724. doi: 10.1016/j.sbi.2013.08.005

Shen, M. Y., and Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Prot. Sci.* 15, 2507–2524. doi: 10.1110/ps.062416606

Sibille, N., and Bernadó, P. (2012). Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS. *Biochem. Soc. Trans.* 40, 955–962. doi: 10.1042/BST20120149

Sickmeier, M., Hamilton, J. A., LeGall, T., Vacic, V., Cortese, M. S., Tantos, A., et al. (2007). DisProt: the database of disordered proteins. *Nucleic Acids Res.* 35, D786–D793. doi: 10.1093/nar/gkl893

Sivakolundu, S. G., Bashford, D., and Kriwacki, R. W. (2005). Disordered p27Kip1 exhibits intrinsic structure resembling the Cdk2/cyclin A-bound conformation. *J. Mol. Biol.* 353, 1118–1128. doi: 10.1016/j.jmb.2005.08.074

Smagghe, B. J., Huang, P. S., Ban, Y. E., Baker, D., and Springer, T. A. (2010). Modulation of integrin activation by an entropic spring in the {beta}-knee. *J. Biol. Chem.* 285, 32954–32966. doi: 10.1074/jbc.M110.145177

Tamiola, K., Acar, B., and Mulder, F. A. (2010). Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J. Am. Chem. Soc.* 132, 18000–18003. doi: 10.1021/ja105656t

Tompa, P. (2005). The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.* 579, 3346–3354. doi: 10.1016/j.febslet.2005.03.072

Tompa, P. (2011). Unstructural biology coming of age. *Curr. Opin. Struct. Biol.* 21, 419–425. doi: 10.1016/j.sbi.2011.03.012

Tompa, P. (2012). On the supertertiary structure of proteins. *Nat. Chem. Biol.* 8, 597–600. doi: 10.1038/nchembio.1009

Tompa, P., Dosztanyi, Z., and Simon, I. (2006). Prevalent structural disorder in *E. coli* and *S. cerevisiae* proteomes. *J. Proteom. Res.* 5, 1996–2000. doi: 10.1021/pr0600881

Tompa, P., and Fuxreiter, M. (2008). Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci.* 33, 2–8. doi: 10.1016/j.tibs.2007.10.003

Tompa, P., and Varadi, M. (2014). Predicting the predictive power of IDP ensembles. *Structure* 22, 177–178. doi: 10.1016/j.str.2014.01.003

Uversky, V. N. (2010). Targeting intrinsically disordered proteins in neurodegenerative and protein dysfunction diseases: another illustration of the D(2) concept. *Exp. Rev. Proteom.* 7, 543–564. doi: 10.1586/epr.10.36

Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* 41, 415–427. doi: 10.1002/1097-0134(20001115)41:3<415::AID-PROT130>3.0.CO;2-7

Varadi, M., Guharoy, M., Zsolyomi, F., and Tompa, P. (2015). DisCons: a novel tool to quantify and classify evolutionary conservation of intrinsic protein disorder. *BMC Bioinform.* 16:153. doi: 10.1186/s12859-015-0592-2

Varadi, M., Kosol, S., Lebrun, P., Valentini, E., Blackledge, M., Dunker, A. K., et al. (2014). pE-DB: a database of structural ensembles of intrinsically disordered and of unfolded proteins. *Nucleic Acids Res.* 42, D326–D335. doi: 10.1093/nar/gkt960

Vranken, W. F. (2014). NMR structure validation in relation to dynamics and structure determination. *Prog. Nuclear Magn. Reson. Spectr.* 82, 27–38. doi: 10.1016/j.pnmrs.2014.08.001

Vuister, G. W., Fogh, R. H., Hendrickx, P. M., Doreleijers, J. F., and Gutmanas, A. (2014). An overview of tools for the validation of protein NMR structures. *J. Biomol. NMR* 58, 259–285. doi: 10.1007/s10858-013-9750-x

Walsh, I., Martin, A. J., Di Domenico, T., and Tosatto, S. C. (2012). ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics* 28, 503–509. doi: 10.1093/bioinformatics/btr682

Wang, Y., Fisher, J. C., Mathew, R., Ou, L., Otieno, S., Sublet, J., et al. (2011). Intrinsic disorder mediates the diverse regulatory functions of the Cdk inhibitor p21. *Nat. Chem. Biol.* 7, 214–221. doi: 10.1038/nchembio.536

Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F., and Jones, D. T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* 337, 635–645. doi: 10.1016/j.jmb.2004.02.002

Weinreb, P. H., Zhen, W., Poon, A. W., Conway, K. A., and Lansbury, P. T. Jr. (1996). NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry* 35, 13709–13715. doi: 10.1021/bi961799n

Wright, P. E., and Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* 293, 321–331. doi: 10.1006/jmbi.1999.3110

Wright, P. E., and Dyson, H. J. (2009). Linking folding and binding. *Curr. Opin. Struct. Biol.* 19, 31–38. doi: 10.1016/j.sbi.2008.12.003

Xie, H., Vucetic, S., Iakoucheva, L. M., Oldfield, C. J., Dunker, A. K., Uversky, V. N., et al. (2007). Functional anthology of intrinsic disorder. 1.Biological processes and functions of proteins with long disordered regions. *J. Proteom. Res.* 6, 1882–1898. doi: 10.1021/pr060392u

Yang, Z. R., Thomson, R., McNeil, P., and Esnouf, R. M. (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* 21, 3369–3376. doi: 10.1093/bioinformatics/bti534

Yoon, M. K., Mitrea, D. M., Ou, L., and Kriwacki, R. W. (2012). Cell cycle regulation by the intrinsically disordered proteins p21 and p27. *Biochem. Soc. Trans.* 40, 981–988. doi: 10.1042/BST20120092

Zhang, W., and Chen, J. (2014). Replica exchange with guided annealing for accelerated sampling of disordered protein conformations. *J. comput. Chem.* 35, 1682–1689. doi: 10.1002/jcc.23675