Disorder in Proteins

Airy Sanjeev, Tezpur University, Tezpur, Assam, India
Sandeep Kaushik, Headquarters of the European Institute of Excellence on Tissue Engineering and Regenerative Medicine, Guimaraes, Portugal
Biswa R Meher, Berhampur University, Berhampur, Odisha, India
Venkata SK Mattaparthi, Tezpur University, Tezpur, Assam, India

1 Background

1.1 Intrinsically Disordered Proteins

Intrinsically disordered proteins (IDPs) are important cellular components that play significant roles in various processes such as cell signaling, translation, transcription, and cell cycle regulation. Disorder in a polypeptide structure allows the same polypeptide to assume diverse interactions with different consequences (Dunker *et al.*, 2008; Dyson and Wright, 2005). IDPs are distinguished based on their low sequence complexity and biased amino acid composition: they typically contain a high proportion of charged and hydrophilic amino acids and a low proportion of bulky hydrophobic amino acids. These IDPs undertake transitions of *disorder-to-order* state on binding to the partners though they do not contain a stable tertiary structure (Uversky, 2013; Wright and Dyson, 2009). These proteins are not able to fold into well-defined, globular three-dimensional structures spontaneously even though they are perfectly functional. But they have the ability to rapidly fluctuate through a range of conformations that can cover a continuous conformational space that range from the extended statistical coils to collapsed globules and are dynamically disordered (Dyson and Wright, 2005). IDPs also play a significant role in the ordered assembly of macromolecular machines such as ribosomes, chromatin organization, assembly and disassembly of microfilaments and microtubules, nuclear pore transportation, binding and small molecule transportation, and functional protein domain separation (Frey *et al.*, 2006; Guharoy *et al.*, 2013; Tompa, 2005, 2010). In addition to IDPs, some proteins contain some disordered sequences known as Intrinsically Disordered Regions (IDRs). IDPs are known to interact with and function as *hubs* in the various protein-interaction networks (Dunker *et al.*, 2005; Kim *et al.*, 2008). IDPs get subjected to various processes such as alternative splicing of their mRNA and combinatorial Post-Translational Modifications (PTM) that can add complexity to the regulatory networks and provide a tissue-specific signaling mechanism. Abundance of these IDRs are tightly being regulated that are related to the disease for ascertaining the signaling both in time and space and also for mutation in the IDPs (Babu *et al.*, 2011; Gsponer *et al.*, 2008; Vavouri *et al.*, 2009). In the same manner, various folded proteins can control the transitions of order-to-disorder state that can mediate the various biological functions (Mitrea and Kriwacki, 2013; Schultz and Natarajan, 2013). IDPs are also known to take part in the signaling complexes assembly and in the self-assembly of dynamic cytoplasmic and membrane-less nuclear organelles. However the PTMs can control the IDPs function leading to structural changes (*folding*). The role of PTM is to stabilize or destabilize the individual secondary structural elements in the various IDPs (Espinoza-Fonseca *et al.*, 2008; Pufall *et al.*, 2005; Theillet *et al.*, 2012). IDPs also have the ability to bind to multiple interaction partners transiently in the dynamic regulatory networks that can accurately respond to cellular signals and possess potential for complex information processing when performing the signaling functions (Stein *et al.*, 2009). The interactions of IDPs with their binding partners are controlled by the PTMs allowing them to work as switches and rheostats (Borg *et al.*, 2007; Dyson and Wright, 2005; Gsponer and Babu, 2009; Lee *et al.*, 2010; Van Roey *et al.*, 2012, 2013). IDPs are known to exchange their binding partners for which they compete for the central hub of the protein that is present in very limited amount. The disordered regions in signaling and regulatory proteins however consist of many conserved sequence motifs which can interact with nucleic acids or other proteins (Dyson and Wright, 2002, 2005).

1.2 Intrinsically Disordered Protein Regions

Various neurodegenerative diseases such as Parkinson's disease and prion diseases have been linked with intrinsically disordered proteins. Also many glutamine-rich motifs transcription factors are involved in polyglutamine diseases (McCampbell and Fischbeck, 2001).

Even after binding to the targets (Baker *et al.*, 2007; Mittag *et al.*, 2010), some IDPs remain disordered and form fuzzy complexes (Tompa and Fuxreiter, 2008). In addition disordered dynamic sites on distinct, non-overlapping surfaces of the target protein have been observed (Ferreon *et al.*, 2009; Ishiyama *et al.*, 2010). These binding interactions mediate pathway crosstalk via the formation of ternary complexes and also enhance the target-binding affinity with various binding partners that can modulate allosteric interactions (Ferreon *et al.*, 2013). These proteins consist of multiple interaction motifs that can mediate binding to a diversity of targets which may act as *'central hubs'* in signaling networks (Dunker *et al.*, 2005; Kim *et al.*, 2008). They can facilitate the assembly of both ternary and higher-order complexes through multiple sites and also integrate diverse signaling pathways. These sites allow allosteric responses in biological signaling, and the large entropy variations between the free and bound states can act to tune the energetics of the binding process (Flock *et al.*, 2014; Hilser and Thompson, 2007; Motlagh *et al.*, 2014; Nussinov and Tsai, 2013). in addition to fully disordered proteins, a large number of proteins present in various organisms contain polypeptide segments that do not form a stable 3D structure but are however functional (Daughdrill *et al.*, 1997; Dunker *et al.*, 2008; Dyson and Wright, 2005; Fink, 2005; Gsponer and Babu, 2009; Kriwacki *et al.*, 1996; Tompa, 2002; Uversky *et al.*, 2000; Uversky and Dunker, 2010; Ward *et al.*, 2004; Wright and Dyson, 1999). In comparison to the original view of the globular and classical structure-function method, the functionality of the IDRs results in a very unique approach (Li *et al.*, 2001, 2012a,b; Romero *et al.*, 2001; Weatheritt *et al.*, 2014). For example, these regions are required for the initiation of programmed necrosis through the formation of heterodimeric amyloid fibril (Li *et al.*, 2012a,b). Recent reports suggest that disordered sequences of the RHIM (**R**IP **h**omotypic **i**nteraction **m**otif) are buried in the inactive state but on kinase activation they become exposed to amyloid, leading to necrosome formation (Li *et al.*, 2012a,b).

IDPs have a greater impact on the constitutive pre-mRNA splicing and alternative splicing catalyzed by a dynamic ribonucleoprotein machine called the *spliceosome*. The components of this body are highly enriched in intrinsically disordered regions (Korneta and Bujnicki, 2012) that are considered as the spliced segments of the protein substrates alternatively (Romero *et al.*, 2006) (*'alternatively spliced'*).

1.3 Disordered Proteins and Disease

IDPs and IDRPs were considered as the two main variants of protein consisting of varying physico-chemical and functional properties. This arises mainly due to the varying percentage of disordered regions they contain. The IDRPs are the protein sequence that consists of regions which are locally disordered and they are featured by the existence of disordered domains in structure that are folded globally. These IDPs also contain some disordered regions in the relevant percentage of their residues.

The representation of the differential distribution of IDRPs in the *Homo sapiens* and in proteins annotated as disease-related, as a function of the minimal length of the disordered regions is depicted in **Fig. 1**. From the figure, the overall percentages of IDRPs present in the various diseases were depicted as a function of their disordered protein length in the case of *Homo sapiens*. We noticed among all the various diseases, the percentage of proteins in the IDRPs were higher in case of calcium homeostasis and lowest in cancer. In case of calcium homeostasis the range of IDRPs exceed by 46% while in the whole human proteome it is about 26%. Although further investigation is required for the remarkable case of IDRPs enrichment that is penetrated by the calcium homeostasis.

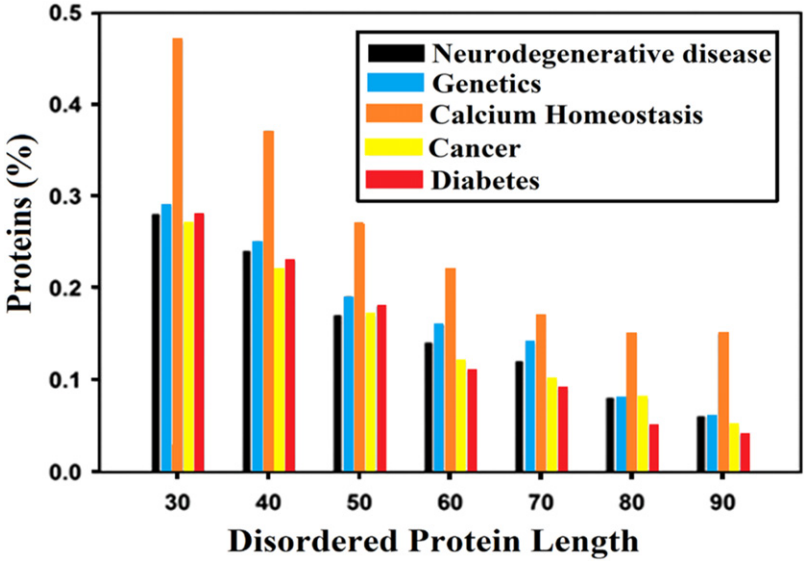

Fig. 1 The varying distribution of the overall IDRPs percentage in the Homo sapiens represented in diseases as a function of disordered protein length.

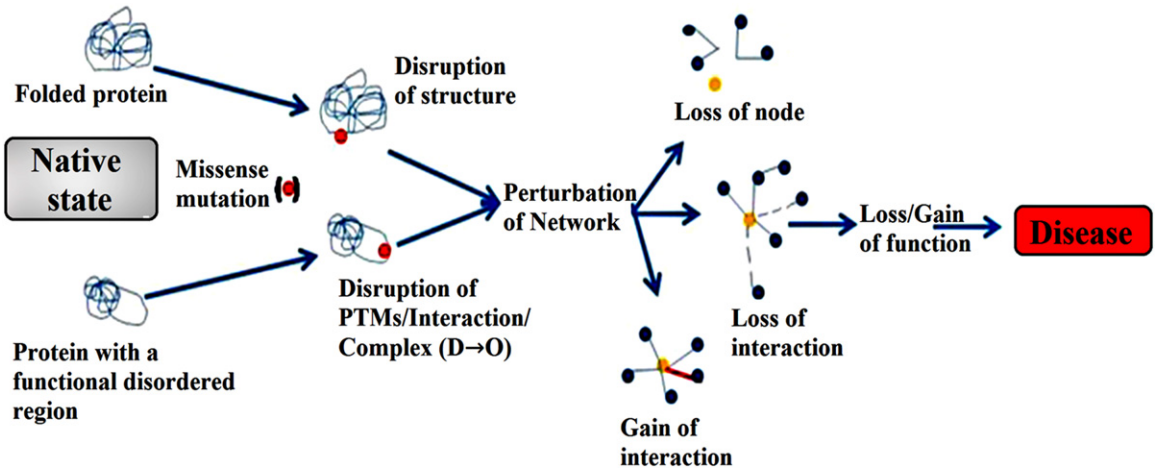

Fig. 2 Schematic diagram of the impact of diseased mutation on the ordered and disordered protein.

The occurrence of IDPs in the human proteome and genetic diseases is almost same. As the IDPs prevail in various diseases, it can be suggested that there is a lack of three-dimensional structure due to which the diseases occur reported earlier (Uversky *et al.*, 2008, 2009). But the development of diabetes is not influenced by the prevalence of IDPs as studied earlier (Uversky *et al.*, 2008, 2009).

The occurrence of IDPs in the various diseases is more when compared to the human proteome. But this fact cannot illustrate that IDPs are necessary for the occurrence of diseases as the maximum proteins are not IDPs connected to the various diseases. However IDPs are more common than the other proteomes. Hence we can infer that in the development of cancer and neurodegenerative diseases, the absence of a three dimensional structure play a significant role while it have no such effective role on the other diseases. In case of calcium homeostasis, though the protein is well structured with flexible loops but this flexibility is necessary for the development of the diseases (Lu *et al.*, 2015; Vacic *et al.*, 2012; Vacic and Iakoucheva, 2012).

In regard to the presence of mutations, there exists a vast difference between the ordered and disordered regions (Lu *et al.*, 2015; Vacic *et al.*, 2012; Vacic and Iakoucheva, 2012). The illustration of the effect of diseased mutations on the ordered and disordered regions referred from *Vacic and Iakoucheva* is depicted in **Fig. 2**.

The IDPs can contribute to the signaling complexes assembly and also membrane-less nuclear and cytoplasmic organelles self-assembly. To identify and characterize the various disordered protein regions several experimental, bioinformatics and computational analysis are being performed that result in higher evolvement of the biological processes (Wright, 2015).

A typical example of IDP is the *α-synuclein* protein which is the principal component for the occurrence of debilitating disease, Parkinson's disease (PD). A schematic representation of the 3D structure of α-synuclein protein is depicted in **Fig. 3**.

2 Short Motifs Found in Disordered Regions and the Sequence-Based Bioinformatics

Currently the **P**rotein **D**ata **B**ank (PDB) consists of about 100,000 deposited structures that are noticed to be accumulated for the past decades (Berman *et al.*, 2013). But almost all the structures have been discovered till now. About 33% proteome present in the eukaryotes are noticed to possess putative long disordered segments (Ward *et al.*, 2004). Numerous peptide motifs known as **S**hort **L**inear **M**otifs (SLiMs) exist in the proteome of the disordered regions that are already discovered and also this SLiMs have been demonstrated experimentally (Tompa *et al.*, 2014). However significant importance has been given to sequence motifs that impact protein function. Sequence motifs are often observed at protein functional sites such as cleavage sites, binding sites, post-translational modification sites and sub-cellular targeting sites, and are involved in specific protein-protein interactions, regulatory functions and signal transduction (Van Roey *et al.*, 2014).

The SLiMs are recognized as the well-defined sequence patterns that represent graphically via the machine-readable **r**egular **e**xpression (REs) or via the sequence logos (Schneider and Stephens, 1990).This however consist the definition for the position-specific of the allowed residue types and also for the ambiguous positions. The ***Mo***lecular ***R***ecognition ***F***eatures (MoRFs) segments are the specific class of **i**ntrinsically **d**isordered **r**egions (IDRs) consisting of specific molecular recognition and binding functions. They are usually short segments which are interaction-prone and are found within the long IDRs (Vacic *et al.*, 2007). They exhibit disorder-to-order transitions when they bind to their partners (Mohan *et al.*, 2006). They are classified on the basis of their bound state structures into α-MoRFs, *β*-MoRFs and *ι*-MoRFs. They are not characterized based on the sequence pattern (RE) like SLiMs. Rather they are defined as an interaction-prone disordered segments which form ordered secondary structures when bind to the protein partner. As IDRs significantly helped in the protein interactions, the understanding of protein-protein interactions has effectively changed (Dinkel *et al.*, 2014; Edwards *et al.*, 2012; Petsalaki and Russell, 2008). The interactions via the SLiMs represent various functions in the diverse process which includes control of cell cycle progression, substrate selection for proteasomal degradation, targeting proteins to specific sub-cellular locations and for stabilizing scaffolding complexes.

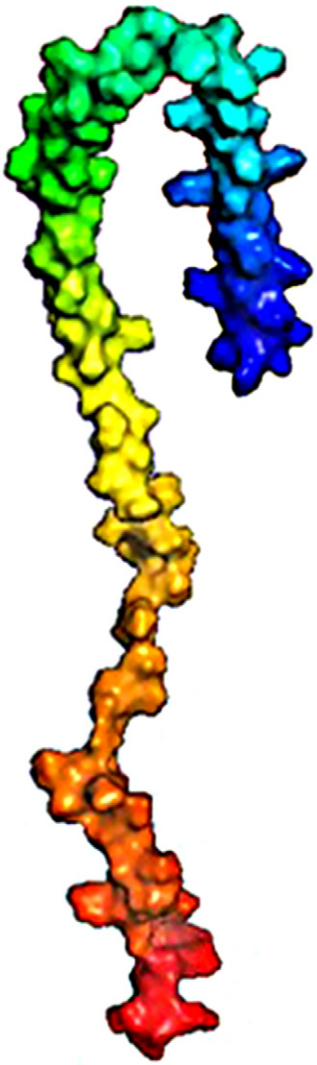

Fig. 3 A three-dimensional schematic view of α-synuclein protein visualized in PyMol software.

ScanSite is another database that is known to stores data in 12 different groups for 65 motifs (Obenauer *et al.*, 2003). Apart from ScanSite, Prosite is another database consisting of various data for about 1308 patterns or regular expression even though it comprises of domain signatures (Sigrist *et al.*, 2013). There are very less information related to the SLiMs although much information is there about the eukaryotic cell regulation (Tompa *et al.*, 2014). Hence from these, the need for developing computational methods for predicting novel SLiMs in the protein sequences have been urged effectively.

3 Structure-Based Computational Approaches for Modeling and Simulation of Disordered Regions

For describing the IDP structure and dynamics accurately, the best method is the all-atom molecular simulations consisting of explicit protein and solvent (Ithuralde *et al.*, 2016; Knott and Best, 2012; Stanley *et al.*, 2015). Molecular dynamics (MD) is a method that computationally simulates the time evolution of set of interacting atoms using Newton's equations of motion. These techniques are dependent on how the molecules will interact- *a force field* and are popular in materials chemistry, biochemistry and biophysics. It is a technique for producing dynamical trajectory for systems that are composed of N particles through integrating the Newton's equations of motion. For this, we need a set of initial conditions (positions and velocities of each particle), a good model to represent the forces acting between the particles and to define the boundary conditions to be employed (Hospital *et al.*, 2015). A force field is defined as a mathematical expression that describes the dependence of the energy of a system on the coordinates of its particles. Many force fields are available in the literature that have different degrees of complexity, and oriented to treat different kinds of systems. But there also exists two main drawbacks for the recent force fields on investigating the IDP properties. The first drawback is over-populating a β-sheet or α-helical structure from the older force fields that are eventually

noticed in the disordered chains not forming a long-range tertiary structure (Best *et al.*, 2008). Through the explicit optimization of the backbone torsion parameters opposite to the experimental NMR data, the problems related to the recent force field generation have been clarified. The updated versions of the force-fields are CHARMM 22* (Piana *et al.*, 2011), CHARMM 36 (Best *et al.*, 2012; Huang *et al.*, 2016), Amber ff03* (Best and Hummer, 2009), Amber ff99SB*-ILDN (Best and Hummer, 2009; Lindorff-Larsen *et al.*, 2010) and Amber ff99SBnmr (Li and Bruschweiler, 2011). These marked updated force fields have been reported in the literature by Piana *et al.* (2014).

The second drawback for the all-atom force fields was evident due to the direct comparison of all-atom simulations with single-molecule FRET and SAXS experiments which showed that the simulated unfolded states were much too collapsed (Piana *et al.*, 2014; Nettels *et al.*, 2009) (being close to maximally compact globules). As such the secondary structures were also over-populated (Skinner *et al.*, 2014). Till the recent two years, the various issues related to the force fields were considered as same. However various modification related to the force field have already addressed this issue (Best *et al.*, 2014; Huang *et al.*, 2016; Nerenberg *et al.*, 2012; Piana *et al.*, 2015; Ploetz *et al.*, 2010; Yoo and Aksimentiev, 2016). From all these force fields, the first approach that was selected was the direct strengthening of protein-water interactions by upgrading the Lennard-Jones mixing rule (Best *et al.*, 2014; Huang *et al.*, 2016; Nerenberg *et al.*, 2012) for investigating the protein-water interactions, as stated by the Amber ff03 ws force field (Best *et al.*, 2014). The second approach evaluated the increasing the well-depth of Lennard-Jones from the novel water model of TIP4P-D (Piana *et al.*, 2015). But the overall impact for the disordered proteins dimension via using Amber ff03ws and TIP4P-D was noticed to be same. This is due to the fact that for both the models, the rise in the interaction for protein-water dispersion is same. Whereas for both models, the average bulk water-water interactions remain same as both the models match the experimental heat of water vaporization. These two models have been shown to play a significant role with the data from SAXS for the other related IDPs(Henriques *et al.*, 2015; Henriques and Skepo, 2016; Zheng *et al.*, 2016). The peptide reconfiguration dynamics is also precisely captured due to the less level of collapsed chain (Zerze *et al.*, 2016).

Apart from these two, the third promising is defined as the Kirkwood-Buff force field (KBFF) designed for proteins for emphasizing on the properties of solution (Ploetz *et al.*, 2010). The KBFF gives rise to results for FG-nucleoporins that are considered consistent with SAXS and FRET experiments (Mercadante *et al.*, 2015). Even after designing new force fields for IDPs, there exists some drawbacks such as folded peptides and proteins are destabilized (Mercadante *et al.*, 2015; Nerenberg *et al.*, 2012; Piana *et al.*, 2015) due to the favourable interactions with water in the denatured state (Best *et al.*, 2014). This effect is compensated presumably by specific stabilizing interactions that are absent in the current force fields (Nerenberg *et al.*, 2012). But however the improvements in the force field may not be considered universal as the presence of some molecules such as disordered arginine/serine (RS) peptide are comparatively more expanded than the experiment (Rauscher *et al.*, 2015). However the atomistic simulations are almost unaffordable and possibly unnecessary due to the larger length spanned by many IDPs in their functional states.

4 Emerging Computational Tools and Techniques

At present, there exist numerous resources which can help us to gain an insight for the function of IDR. The illustrations of the various resources available are summarized in **Table 1**. There are several areas where efforts are being upgraded and made to help us understand the protein disorder function. We however notice the combination of multiple existing classification schemes that can help us to retain the function for predicting the IDRs. This perhaps can help us to have an extensive knowledge of the function of protein and also to obtain enhanced coverage of function.

On the basis of sequence analysis, various web servers have been designed to predict the disordered regions in protein structures. Recently a database has been launched that is coded by the human genome which consists of consensus disorder predictions for all the proteins (Oates *et al.*, 2013). However, vast developments have been done in the field of experimental and computational tools that can generate the structural ensembles of disordered proteins. Based on the NMR, SAXS data and pE-DB, the modeling of ensembles have been made (Marsh and Forman-Kay, 2012; Salmon *et al.*, 2010; Sibille and Bernado, 2012).

Various approaches have been initiated for the design of disordered proteins and to measure the degree of disordered regions. **PrDOS** is one such software that was created by Ishida and Kinoshita at the University of Tokyo (Ishida and Kinoshita, 2007). In this software, a hybrid method is used along with a machine learning approach that depends on both the SVM model with a template-based model. The input is used as an evolutionary profile of the input sequence by the machine learning model. Whereas the template-based model explore the homologues in the PDB. Both the predictions made by the machine learning and template-based models are based on a weighted average of the results.

Another approach is the **MFDp** software developed by Kurgan's group at the University of Alberta (currently at the Virginia Commonwealth University) (Mizianty *et al.*, 2010).In this meta-predictor, three SVM models are combined in order to calculate long, short, and all-size IDRs. A variety of different inputs are being utilized by SVM that include information taken from the amino acid sequence. The putative disorders are predicted from three predictors that include: *evolutionary profile, putative B-factors, putative secondary structure and backbone dihedral torsion angles, putative solvent accessibility* and *putative annotation of globular domains*. **MFDp** was recently upgraded to **MFDp2**, in 2013 (Mizianty *et al.*, 2013, 2014) that combines predictions generated by **MFDp** with predictions that are based on computed alignment against disordered proteins database that are being extracted from DisProt. The predictions generated are corrected in a way that predicted disordered residues number matches putative disordered residues output number by DisCon method (Mizianty *et al.*, 2011).

Table 1 Function prediction methods and techniques for the intrinsically disordered regions and proteins

Method basis	*Detailed explanation*	*Technique*	*URL of the Web site*
Method for linear motifs	Description regarding well defined linear motifs that are mapped onto the other sequences of protein	ELM	http://elm.eu.org/
		MiniMotif	http://mnm.engr.uconn.edu/
	Recognition of putative uncharacterized motifs in the sequences of protein	SLiMPrints	http://bioware.ucd.ie/slimprints.html
		phylo-HMM	http://www.moseslab.csb.utoronto.ca/phylo_HMM/
		DiliMot	http://dilimot.russelllab.org/
		SLiMFinder	http://bioware.ucd.ie/slimfinder.html
Method for the sites of PTM	Description of various resources that are experimentally being verified by the PTM sites (phosphorylation)	Phospho.ELM	http://phospho.elm.eu.org/
		PhosphoSite	http://www.phosphosite.org/
		PHOSIDA	http://www.phosida.com/
	Detection and compilation of the peptide motifs that can forward the post-translational modification	ScanSite	http://scansite.mit.edu/
		NetPhorest	http://netphorest.info/
		NetworKIN	http://networkin.info/
		PhosphoNET	http://www.phosphonet.ca/
Features of molecular recognition	Assortment of the verified sequence elements that go through coupled folding and binding process	IDEAL	http://www.ideal.force.cs.is.nagoya-u.ac.jp/IDEAL/
	Determination of protein sequences that experience disorder-to-order transitions	MoRFpred	http://biomine.ece.ualberta.ca/MoRFpred/
		ANCHOR	http://anchor.enzim.hu/
Domain for intrinsically disordered proteins	Description regarding disordered protein domains that are detected by sequence profiles	Pfam	http://pfam.sanger.ac.uk/
Other	Calculation of gene ontology functions with features of protein sequence such as *intrinsic disorder*	FFPred	http://bioinf.cs.ucl.ac.uk/psipred/
	Description of experimentally confirmed disordered protein regions	DisProt	http://www.disprot.org/

A last approach was being developed named as **DISOPRED** that was released by Jones's group at the University College London (Jones and Cozzetto, 2015). The first version of this method, **DISOPRED** was published in 2003 (Jones and Ward, 2003), the second version, **DISOPRED2**, in 2004 (Ward *et al.*, 2004a) and the latest third version, **DISOPRED3**, in 2015 (Jones and Cozzetto, 2015). The latest version i.e. **DISOPRED3** is basically a machine learning method that is implemented as a two-stage neural network. It uses the predictions from three main predictors: **DISOPRED2** which is a specialized predictor for the long IDRs and a nearest neighbor-based model with similarities uses to a set of proteins by the IDRs. The design of this method is somewhat similar to **MFDp2**. It is from the fact this model includes a module that predicts long IDRs and an alignment-based module. However the main difference lies in the input method consisting of an evolutionary profile and the second stage using a neural network mixes these three predictions. The **DISOPRED3** software is also known to predict the protein binding site that is defined as the protein binding regions located mainly inside the IDRs.

5 Future Perspectives and Closing Remarks

Still many works are there to characterize the unfolded functional proteins. However various computational techniques for screening the protein sequences, entire genome sequences for the intrinsic disorder proteins will render more proteins belonging to this class. Our understanding will advance for the functions of these disordered regions of proteins in parallel progress of the functional genomics. Although researches in this field have reached the peak point of maturity but recently it has been switched to the prediction of the disordered regions functions. The low complexity sequences situated between the structured domains is however started to be addressed. But the mobility of polypeptide degree consistent with the function of protein is still not clear. Hence proteins and their functions will however progress to a continuum from a fully unstructured protein that folds on binding to the target to a relatively rigid protein that contain varied mobile functional regions. Hence the functional proteins should progress from a static state to a dynamic state where several conformations are consistent with their function.

Hence by understanding the functionality and unique features of IDPs, we can gain an insight onto the biologically active conformations that might play an important role for the diagnosis of several neurodegenerative diseases. Also from the newly developed databases for the intrinsically disordered protein, the determination of structure and their properties can be well understood.

Acknowledgement

The authors thank Central Libraries of Tezpur University and Central University of Jharkhand (CUJ) for providing us resources to complete the article.BRM is thankful to DBT-BUILDER programme (BT/PR-9028/INF/22/193/2013) for financial support.

References

Babu, M.M., van der Lee, R., de Groot, N.S., Gsponer, Jr., 2011. Intrinsically disordered proteins: Regulation and disease. Current Opinion in Structural Biology 21, 432–440.
Baker, J.M.R., *et al.*, 2007. CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. Nature Structural and Molecular Biology 14, 738–745.
Berman, H.M., Kleywegt, G.J., Nakamura, H., *et al.*, 2013. The future of the protein data bank. Biopolymers 99 (3), 218–222. doi:10.1002/bip.22132.
Best, R.B., Buchete, N.V., Hummer, G., 2008. Are current molecular dynamics force fields too helical? Biophysical Journal 95, L07–L09.
Best, R.B., Hummer, G., 2009. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. Journal of Physical Chemistry B 113, 9004–9015.
Best, R.B., Zheng, W., Mittal, J., 2014. Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. Journal of Chemical Theory and Computation 10, 5113–5124.
Best, R.B., Zhu, X., Shim, J., *et al.*, 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone w, c and side-chain x_1 and x_2 dihedral angles. Journal of Chemical Theory and Computation 8, 3257–3273.
Borg, M., *et al.*, 2007. Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. Proceedings of the National Academy of Sciences of the United States of America 104, 9650–9655.
Daughdrill, G.W., Chadsey, M.S., Karlinsey, J.E., Hughes, K.T., Dahlquist, F.W., 1997. The C-terminal half of the anti-sigma factor, FlgM, becomes structured when bound to its target, sigma 28. Nature Structural & Molecular Biology 4, 285.
Dinkel, H., Van Roey, K., Michael, S., *et al.*, 2014. The eukaryotic linear motif resource ELM: 10 years and counting. Nucleic Acids Research 42, D259–D266. doi:10.1093/nar/gkt1047.
Dunker, A.K., Cortese, M.S., Romero, P., lakoucheva, L.M., Uversky, V.N., 2005. The roles of intrinsic disorder in protein interaction networks. FEBS Journal 272, 5129–5148.
Dunker, A.K., Oldfield, C.J., Meng, J., *et al.*, 2008. The unfoldomics decade: An update on intrinsically disordered proteins. BMC Genomics 9. doi:10.1186/1471-2164-9-S2-S1.
Dunker, A.K., Silman, I., Uversky, V.N., Sussman, J.L., 2008. Function and structure of inherently disordered proteins. Current Opinion in Structural Biology 18 (6), 756–764. doi:10.1016/j.sbi.2008.10.002.
Dyson, H.J., Wright, P.E., 2002. Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. Advances in Protein Chemistry 62, 311–340.
Dyson, H.J., Wright, P.E., 2005. Intrinsically unstructured proteins and their functions. Nature Reviews Molecular Cell Biology 6, 197–208.
Edwards, R.J., Davey, N.E., O'Brien, K., *et al.*, 2012. Interactome-wide prediction of short, disordered protein interaction motifs in humans. Molecular BioSystems 8 (1), 282–295. doi:10.1039/c1mb05212h.
Espinoza-Fonseca, L.M., Kast, D., Thomas, D.D., 2008. Thermodynamic and structural basis of phosphorylation-induced disorder-to-order transition in the regulatory light chain of smooth muscle myosin. Journal of the American Chemical Society 130 (37), 12208–12209. doi:10.1021/ja803143g.
Ferreon, A.C., Ferreon, J.C., Wright, P.E., Deniz, A.A., 2013. Modulation of allostery by protein intrinsic disorder. Nature 498, 390–394.
Ferreon, J.C., Martinez-Yamout, M.A., Dyson, H.J., Wright, P.E., 2009. Structural basis for subversion of cellular control mechanisms by the adenoviral E1A oncoprotein. Proceedings of the National Academy of Sciences of the United States of America 106, 13260–13265.
Fink, A.L., 2005. Natively unfolded proteins. Current Opinion in Structural Biology 15, 35–41.
Flock, T., Weatheritt, R.J., Latysheva, N.S., Babu, M.M., 2014. Controlling entropy to tune the functions of intrinsically disordered regions. Current Opinion in Structural Biology 26, 62–72.
Frey, S., Richter, R.P., Gorlich, D., 2006. FG-rich repeats of nuclear pore proteins form a three dimensional meshwork with hydrogel-like properties. Science 314, 815–817.
Gsponer, J., Babu, M.M., 2009. The rules of disorder or why disorder rules. Progress in Biophysics and Molecular Biology 99, 94–103.
Gsponer, J., Futschik, M.E., Teichmann, S.A., Babu, M.M., 2008. Tight regulation of unstructured proteins: From transcript synthesis to protein degradation. Science 322, 1365–1368.
Guharoy, M., Szabo, B., Martos, S.C., Kosol, S., Tompa, P., 2013. Intrinsic structural disorder in cytoskeletal proteins. Cytoskeleton 70, 550–571.
Henriques, J., Cragnell, C., Skepo, M., 2015. Molecular dynamics simulations of intrinsically disordered proteins: Force field evaluation and comparison with experiment. Journal of Chemical Theory and Computation 11, 3420–3431.
Henriques, J., Skepo, M., 2016. Molecular dynamics simulations of intrinsically disordered proteins: On the accuracy of the TIP4P-D water model and the representativeness of protein disorder models. Journal of Chemical Theory and Computation 12.
Hilser, V.J., Thompson, E.B., 2007. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. Proceedings of the National Academy of Sciences of the United States of America 104, 8311–8315.
Hospital, A., Goni, J.R., Orozco, M., Gelpi, J.L., 2015. Molecular dynamics simulations: Advances and applications. Advances and Applications in Bioinformatics and Chemistry 8, 37–47. Available at: https://doi.org/10.2147/AABC.S70333.
Huang, J., Rauscher, S., Nawrocki, G., *et al.*, 2016. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. Nature Methods 14, 71–73.
Ishida, T., Kinoshita, K., 2007. PrDOS: Prediction of disordered protein regions from amino acid sequence. Nucleic Acids Research 35, W460–W464.
Ishiyama, N., *et al.*, 2010. Dynamic and static interactions between p120 catenin and E-cadherin regulate the stability of cell-cell adhesion. Cell 141, 117–128.
Ithuralde, R.E., Roitberg, A.E., Turjanski, A.G., 2016. Structured and unstructured binding of an intrinsically disordered protein as revealed by atomistic simulations. Journal of the American Chemical Society 138, 8742–8751.
Jones, D.T., Cozzetto, D., 2015. DISOPRED3: Precise disordered region predictions with annotated protein-binding activity. Bioinformatics 31, 857–863.
Jones, D.T., Ward, J.J., 2003. Prediction of disordered regions in proteins from position specific score matrices. Proteins Structure Function and Bioinformatics 53, 573–578.
Kim, P.M., Sboner, A., Xia, Y., Gerstein, M., 2008. The role of disorder in interaction networks: A structural analysis. Molecular Systems Biology 4, 179.
Knott, M., Best, R.B., 2012. A preformed binding interface in the unbound ensemble of an intrinsically disordered protein: Evidence from molecular simulations. PLoS Computational Biology 8, e1002605.
Korneta, I., Bujnicki, J.M., 2012. Intrinsic disorder in the human spliceosomal proteome. PLoS Computational Biology 8, e1002641.
Kriwacki, R.W., Hengst, L., Tennant, L., Reed, S.I., Wright, P.E., 1996. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: Conformational disorder mediates binding diversity. Proceedings of the National Academy of Sciences of the United States of America 93, 11504.
Lee, C.W., Ferreon, J.C., Ferreon, A.C., Arai, M., Wright, P.E., 2010. Graded enhancement of p53 binding to CREB-binding protein (CBP) by multisite phosphorylation. Proceedings of the National Academy of Sciences of the United States of America 107, 19290–19295.
Lindorff-Larsen, K., Piana, S., Palmo, K., *et al.*, 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 78, 1950–1958.
Li, D.W., Bruschweiler, R., 2011. Iterative optimization of molecular mechanics force fields from NMR data of full-length proteins. Journal of Chemical Theory and Computation 7, 1773–1782.

Li, Y., Hough, C.J., Suh, S.W., Sarvey, J.M., Frederickson, C.J., 2001. Rapid translocation of Zn(2+) from presynaptic terminals into postsynaptic hippocampal neurons after physiological stimulation. Journal of Neurophysiology 86, 2597–2604.

Li, J., *et al.*, 2012a. The RIP1/RIP3 necrosome forms a functional amyloid signaling complex required for programmed necrosis. Cell 150, 339–350.

Li, P., *et al.*, 2012b. Phase transitions in the assembly of multivalent signalling proteins. Nature 483, 336–340.

Lu, H.C., Chung, S.S., Fornili, A., Fraternali, F., 2015. Anatomy of protein disorder, flexibility and disease-related mutations. Frontiers in Molecular Biosciences 2, 47. https://doi.org/10.3389/fmolb.2015.00047.

Marsh, J.A., Forman-Kay, J.D., 2012. Ensemble modeling of protein disordered states: Experimental restraint contributions and validation. Proteins Structure Function and Bioinformatics 80, 556–572.

McCampbell, A., Fischbeck, K.H., 2001. Polyglutamine and CBP: Fatal attraction? Nature Medicine 7, 528–530.

Mercadante, D., Milles, S., Fuertes, G., *et al.*, 2015. Kirkwood-buff approach rescues overcollapse of a disordered protein in canonical protein force fields. Journal of Physical Chemistry B 119, 7975–7984.

Mitrea, D.M., Kriwacki, R.W., 2013. Regulated unfolding of proteins in signaling. FEBS Letters 17 (587(8)), 1081–1088. doi:10.1016/j.febslet.2013.02.024.

Mittag, T., *et al.*, 2010. 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.

Mizianty, M.J., Peng, Z.L., Kurgan, L., 2013. MFDp2: Accurate predictor of disorder in proteins by fusion of disorder probabilities, content and profiles. Intrinsically Disordered Proteins 1, e24428.

Mizianty, M.J., Stach, W., Chen, K., *et al.*, 2010. Improved sequence-based prediction of disordered regions with multilayer fusion of multiple information sources. Bioinformatics 26, i489–i496.

Mizianty, M.J., Uversky, V., Kurgan, L., 2014. Prediction of intrinsic disorder in proteins using MFDp2. Methods in Molecular Biology 1137, 147–162.

Mizianty, M.J., Zhang, T., Xue, B., *et al.*, 2011. In-silico prediction of disorder content using hybrid sequence representation. BMC Bioinformatics 12, 245.

Motlagh, H.N., Wrabl, J.O., Li, J., Hilser, V.J., 2014. The ensemble nature of allostery. Nature 508, 331–339.

Mohan, A., Oldfield, C.J., Radivojac, P., *et al.*, 2006. Analysis of molecular recognition features (MoRFs). Journal of Molecular Biology 362, 1043–1059.

Nerenberg, P.S., Jo, B., Tripathy, A., Head-Gordon, T., 2012. Optimizing solute-water van der Waals interactions to reproduce solvation free energies. Journal of Physical Chemistry B 116, 4524–4534.

Nettels, D., Muller-Spath, S., Kuster, F., *et al.*, 2009. Single molecule spectroscopy of the temperature-induced collapse of unfolded proteins. Proceedings of the National Academy of Sciences of the United States of America 106, 20740–20745.

Nussinov, R., Tsai, C.J., 2013. Allostery in disease and in drug discovery. Cell 153, 293–305.

Oates, M.E., Romero, P., Ishida, T., *et al.*, 2013. D (2)P(2): Database of disordered protein predictions. Nucleic Acids Research 41, D508–D516.

Obenauer, J.C., Cantley, L.C., Yaffe, M.B., 2003. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. Nucleic Acids Research 31 (13), 3635–3641.

Petsalaki, E., Russell, R.B., 2008. Peptide-mediated interactions in biological systems: New discoveries and applications. Current Opinion in Biotechnology 19 (4), 344–350. doi:10.1016/j.copbio.2008.06.004.

Piana, S., Donchev, A.G., Robustelli, P., Shaw, D.E., 2015. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. Journal of Physical Chemistry B 119, 5113–5123.

Piana, S., Klepeis, J.L., Shaw, D.E., 2014. Assessing the accuracy of physical models used in protein-folding simulations: Quantitative evidence from long molecular dynamics simulations. Current Opinion in Structural Biology 24, 98–105.

Piana, S., Lindorff-Larsen, K., Shaw, D.E., 2011. How robust are protein folding simulations with respect to force field parameterization. Biophysical Journal 100, L47–L49.

Ploetz, E.A., Bentenitis, N., Smith, P.E., 2010. Developing force fields from the microscopic structure of solutions. Fluid Phase Equilibria 290, 43–47.

Pufall, M.A., Lee, G.M., Nelson, M.L., *et al.*, 2005. Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. Science 309, 142–145.

Rauscher, S., Gapsys, V., Gajda, M., *et al.*, 2015. Structural ensembles of intrinsically disordered proteins depend strongly on force field: A comparison to experiment. Journal of Chemical Theory and Computation 11, 5513–5524.

Van Roey, K., Dinkel, H., Weatheritt, R.J., Gibson, T.J., Davey, N.E., 2013. The switches.ELM resource: A compendium of conditional regulatory interaction interfaces. Science Signaling 6, rs7.

Van Roey, K., Gibson, T.J., Davey, N.E., 2012. Motif switches: Decision-making in cell regulation. Current Opinion in Structural Biology 22, 378–385.

Van Roey, K., Uyar, B., Weatheritt, R.J., *et al.*, 2014. Short linear motifs: Ubiquitous and functionally diverse protein interaction modules directing cell regulation. Chemical Reviews 114, 6733–6778. doi:10.1021/cr400585q.

Romero, P., Obradovic, Z., Li, X., *et al.*, 2001. Sequence complexity of disordered protein. Proteins 42, 38–48.

Romero, P.R., Zaidi, S., Fang, Y.Y., *et al.*, 2006. Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. Proceedings of the National Academy of Sciences of the United States of America 103, 8390–8395. doi:10.1073/pnas.0507916103.

Salmon, L., Nodet, G., Ozenne, V., *et al.*, 2010. NMR characterization of long-range order in intrinsically disordered proteins. Journal of the American Chemical Society 132, 8407–8418.

Schneider, T.D., Stephens, R.M., 1990. Sequence logos: A new way to display consensus sequences. Nucleic Acids Research 18 (20), 6097–6100.

Schultz, J.E., Natarajan, J., 2013. Regulated unfolding: A basic principle of intraprotein signaling in modular proteins. Trends in Biochemical Sciences 38 (11), 538–545. doi:10.1016/j.tibs.2013.08.005.

Sibille, N., Bernado, P., 2012. Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS. Biochemical Society Transactions 40, 955–962.

Sigrist, C.J., de Castro, E., Cerutti, L., *et al.*, 2013. New and continuing developments at PROSITE. Nucleic Acids Research 41, D344–D347. doi:10.1093/nar/gks1067.

Skinner, J.J., Yu, W., Gichana, E.K., *et al.*, 2014. Benchmarking all-atom simulations using hydrogen exchange. Proceedings of the National Academy of Sciences of the United States of America 111, 15975–15980.

Stanley, N., Esteban-Martin, S., De Fabritiis, G., 2015. Progress in studying intrinsically disordered proteins with atomistic simulations. Progress in Biophysics & Molecular Biology 119, 47–52.

Stein, A., Pache, R.A., Bernado, P., Pons, M., Aloy, P., 2009. Dynamic interactions of proteins in complex networks: A more structured view. FEBS Journal 276, 5390–5405.

Theillet, F.X., Smet-Nocca, C., Liokatis, S., *et al.*, 2012. Cell signaling, post-translational protein modifications and NMR spectroscopy. Journal of Biomolecular NMR 54, 217–236.

Tompa, P., 2002. Intrinsically unstructured proteins. Trends in Biochemical Sciences 27, 527–533.

Tompa, P., 2005. The interplay between structure and function in intrinsically unstructured proteins. FEBS Letters 579, 3346–3354.

Tompa, P., Davey, N.E., Gibson, T.J., *et al.*, 2014. A million peptide motifs for the molecular biologist. Molecular Cell 55 (2), 161–169. doi:10.1016/j.molcel.2014.05.032.

Tompa, P., Fuxreiter, M., 2008. Fuzzy complexes: Polymorphism and structural disorder in protein-protein interactions. Trends in Biochemical Sciences 33, 2–8.

Tompa, P., 2010. Structure and Function of Intrinsically Disordered Proteins. Boca Raton: Chapman & Hall.

Uversky, V.N., 2013. A decade and a half of protein intrinsic disorder: Biology still waits for physics. Protein Science 22 (6), 693–724.

Uversky, V.N., Dunker, A.K., 2010. Understanding protein non-folding. Biochimica et Biophysica Acta 1804, 1231–1264.

Uversky, V.N., Gillespie, J.R., Fink, A.L., 2000. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 41, 415–427.

Uversky, V.N., Oldfield, C.J., Dunker, A.K., 2008. Intrinsically disordered proteins in human diseases: Introducing the D2 concepts. Annual Review of Biophysics 37, 215–246.

Uversky, V.N., Oldfield, C.J., Midic, U., *et al*., 2009. Unfoldomics of human diseases: Linking protein intrinsic disorder with diseases. BMC Genomics 10, S7–S24.
Vacic, V., Iakoucheva, L.M., 2012. Disease mutations in disordered regions – Exception to the rule? Molecular Biosystems 8 (1), 27–32. https://doi.org/10.1039/c1mb05251a.
Vacic, V., Markwick, P.R.L., Oldfield, C.J., *et al*., 2012. Disease-Associated Mutations Disrupt Functionally Important Regions of Intrinsic Protein Disorder. PLoS Computational Biology 8 (10), e1002709. doi:10.1371/journal.pcbi.1002709.
Vacic, V., Oldfield, C.J., Mohan, A., *et al*., 2007. Characterization of molecular recognition features, MoRFs, and their binding partners. Journal of Proteome Research 6 (6), 2351–2366. doi:10.1021/pr0701411.
Vavouri, T., Semple, J.I., Garcia-Verdugo, R., Lehner, B., 2009. Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. Cell 138, 198–208.
Ward, J.J., McGuffin, L.J., Bryson, K., Buxton, B.F., Jones, D.T., 2004a. The DISOPRED server for the prediction of protein disorder. Bioinformatics 20, 2138–2139.
Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., Jones, D.T., 2004. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. Journal of Molecular Biology 337, 635–645.
Weatheritt, R.J., Gibson, T.J., Babu, M.M., 2014. Asymmetric mRNA localization contributes to fidelity and sensitivity of spatially localized systems. Nature Structural and Molecular Biology 21, 833–839.
Wright, P.E., Dyson, H.J., 2009. Linking folding and binding. Current Opinion in Structural Biology 19 (1), 31–38. doi:10.1016/j.sbi.2008.12.003.
Wright, P.E., Dyson, H.J., 1999. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. Journal of Molecular Biology 293, 321–331.
Wright, P.E., Dyson, H.J., 2015. Intrinsically Disordered Proteins in Cellular Signaling and Regulation. Nature reviews. Molecular cell biology 16, 18–29. doi:10.1038/nrm3920.
Yoo, J., Aksimentiev, A., 2016. Refined parameterisation of nonbonded interactions improves conformational sampling and kinetics of protein folding simulations. Journal of Physical Chemistry Letters 7, 3812–3818.
Zerze, G.H., Mittal, J., Best, R.B., 2016. Diffusive dynamics of contact formation in disordered polypeptides. Physical Review Letters 2016.116:068102.
Zheng, W., Borgia, A., Buholzer, K., *et al*., 2016. Probing the action of chemical denaturant on an intrinsically disordered protein by simulation and experiment. Journal of the American Chemical Society 138, 11702–11713.