

Chapter 13

Molecular Dynamic Simulation of Intrinsically Disordered Proteins and Relevant Forcefields



Prateek Kumar, Nitin Sharma, Amit Kumar, and Rajanish Giri

Abstract Intrinsically disordered proteins (IDPs) exist in every form of life, from bacteria to humans and viruses. They do not hold any well-defined or properly folded structure in the physiological state but have the ability to gain different structural conformations upon interactions with their physiological partners. The limitations of experimental techniques to study high structural dynamics of IDPs have led us to depend on computational simulations. The current scenario of interdisciplinary studies to understand biology with physics has been advantageous for exploring atomic-level dynamics of IDPs. To date, several physics-based forcefields have been developed that calculates the microscopic parameters of a biomacromolecule in an aqueous environment. In this chapter, we have discussed the conformational behavior of IDPs and induced structural properties through understanding the relevant forcefields for molecular dynamics simulations.

13.1 Introduction

By the end of twentieth-century and start of twenty-first century, the understanding of disordered or unstructured proteins started developing. At present, a large number of researchers from every corner of world have devoted their research to describe the proper structure and functioning of disordered regions. A large proportion of gene sequences appear to code not only for folded, globular proteins but also for long stretches of amino acids that are likely to be either unfolded in solution or adopt non-globular structures of unknown conformation (Wright and Dyson 1999). Approximately 44% of genes in humans that code for proteins contain disordered regions (Van Der Lee et al. 2014; Oates et al. 2013). Generally, these proteins or regions are termed as intrinsically disordered proteins or regions (IDPs/IDPRs).

P. Kumar · N. Sharma · A. Kumar · R. Giri (✉)
School of Basic Sciences, Indian Institute of Technology Mandi, Kamand, Himachal Pradesh,
India
e-mail: rajanishgiri@iitmandi.ac.in

These intrinsically disordered protein regions (IDPRs) can be highly conserved within various closely related families or domains of proteins in both composition and sequence (Van Der Lee et al. 2014; Chen et al. 2006). The disordered regions are partially or fully unstructured and are characterized based on various parameters. In other words, they do not possess a proper three-dimensional structure as they fail to acquire structural propensity measured through spectroscopy techniques such as X-ray, NMR, etc. (Dunker et al. 2001). The intrinsic lack of structure can confer functional advantages to a protein like IDPs provide larger interaction surface area, more conformational flexibility, and exposure to interaction prone structural motifs allows IDPs to interact with several other proteins (Babu et al. 2011).

Furthermore, distinct post-translational modifications alleviate regulation of their function and stability in a cell. Some IDPs can attain a fixed tertiary structure on interaction with other molecules known as folders. In contrast, other are called non-folders which do not possess any defined tertiary structure under any physiological conditions. They have ability to undergo partial folding on interaction with specific binding partner proteins (coupled folding and binding), whereas many others constitute flexible linkers that have a role in the assembly of macromolecular arrays (Nishimura et al. 2005). Their conformations may vary from random coils, partially extended globules to collapsed globules with different contents of secondary structure. These distinctly variable structural behaviors of IDPs led to propose multi-state protein structure theories such as trinity (collapsed, ordered, and extended disorder) and quartet (coil, pre-molten globule, molten globule, and folded structure) (Zhang et al. 2013; Dunker and Obradovic 2001).

To elucidate the structure of IDPs and detailed mechanistic insight into their function, firstly, IDPs differential conformations need to be determined. The molecular dynamics (MD) simulation is an excellent computational route for determination of proteins disordered states at atomic level. However, the peculiarities of MD simulation results depend on the accuracy of the physical model (i.e. forcefield) used (Robustelli et al. 2018). There are a number of force fields have been used for the description of folded proteins, but limited for disordered structure prediction (Nerenberg et al. 2012; Piana et al. 2015; Best et al. 2014; Mittal and Best 2010; Lindorff-Larsen et al. 2012, 2013; Beauchamp et al. 2012; Lange et al. 2010). Therefore, in this chapter, we are focused on the IDPs and how the computational method MD simulation exploring the structure disorder via different force fields.

13.2 IDPs and IDPRs: Structure-Function Relationship

The universal lock and key hypothesis for structure function paradigm changed the protein science for a longer time. The proteins 3-D structures were mapped mostly with X-ray crystallography. Despite that, most of the proteins lack complete structures and so-called missing electron density regions (Le Gall et al. 2007). These proteins and regions are unfolded, unstructured and inherent properties of proteins, hence named “intrinsically disordered proteins or intrinsically disordered protein

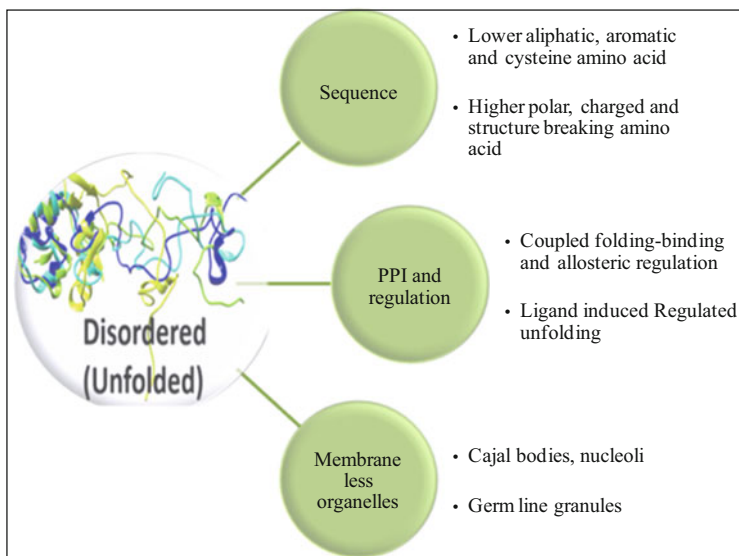


Fig. 13.1 The versatility of intrinsically disordered proteins

regions (IDPs/IDPRs)” (Dunker et al. 2013). The comparison studies of ordered and disordered proteins backbone revealed that the disorder proteins are rich in amino acid Ala, Arg, Gly, Gln, Ser, Glu, Lys, and Pro (Williams et al. 2000; Romero et al. 2001).

As it is stated that the ordered protein follows the structure-function paradigm, i.e., sequence-structure-function, whereas disordered protein follows the disorder-function paradigm (Uversky 2013). The disordered proteins are abundant in all three kingdom of life and viruses, which speculate their importance (Gadhve et al. 2020; Garg et al. 2019; Giri et al. 2016, 2020; Singh et al. 2018a; Kumar et al. 2017, 2019, 2020a; Schad et al. 2011). The very interesting properties of these IDPs are their versatility of performing functions, which can be explained by the “fly casting mechanism” (Huang and Liu 2009; Shoemaker et al. 2000) (Fig. 13.1). Moreover, IDPs can perform function either in native disorder state or can bind to a partner to acquire folding state (Tompa 2005; Uversky and Dunker 2010, 2013; Tompa and Fuxreiter 2008; Dunker et al. 2002). This functional diversity of IDPs lies in its sequence heterogeneity, which allows it to bind with different partners and thus, different conformation and functions (Oldfield et al. 2008). Further, the IDPs/IDPRs possess larger surface area and structural flexibility. Due to the structural flexibility of IDPs they tend to expose peptide regions having molecular recognition features (MoRFs), which may fold while interacting with binding partners (Kumar et al. 2017, 2020a; Mohan et al. 2006; Oldfield et al. 2005; Uversky et al. 2005; Singh et al. 2018b; Mishra et al. 2018). A classic example of this IDPs-binding partner gaining multiple conformations can be illustrated with p53 C-terminal domain (p53-CTD). The p53 CTD (residue 374–388) bound to different partners and acquire

different conformations viz., cyclin A (coil), sirtuin (sheet), CBP (coil), and S100bb (helix) (Uversky 2009; Fadda and Nixon 2017; Kannan et al. 2016).

Besides the potential of IDPs/IDPRs of their multiple functions depending upon the binding partners, surrounding environment need to be considered. There are ample of IDPs which showed a change in conformation in the presence of varying pH, temperature, ions, detergent, organic solvent, crowding agents, and lipids (Kumar et al. 2020a; Uversky 2009; Lopes et al. 2013; Kjaergaard et al. 2010). The surrounding environment imparts electrostatic interaction, hydrophobic interaction, and osmophobic effect, which help IDPs to gain structural conformation (Uversky 2009).

13.3 IDPs in the Human Genome: Organizing Functions or Problems?

Despite being physiologically disordered, IDPs play crucial roles in biological activities. The abundance of IDPs in complex cellular organization displays its importance in regulatory processes (Uversky et al. 2008). These processes include molecular recognition, molecular assembly, entropic activities, and post-translational modifications. Various studies have reported the presence of IDPs in human regulatory proteins such as transcription factors and co-regulators. Eukaryotic proteins seem to use disorder for transient binding purposes (signaling and regulation), while prokaryotic proteins seem to use disorder for longer-lasting interactions, such as complex formation. A recent report suggested that functional misfolding can be induced by fugacious changes in protein environment, and structure can be reversed by restoring the environment or modifications. These induced nature and fugacious characters are important features of these IDPRs or conditionally disordered protein regions (Uversky 2015). Some interesting studies about the occurrence of IDPs in viruses have shown that it plays crucial roles in hijack of host cellular functional machinery (Gadhave et al. 2020; Garg et al. 2019; Giri et al. 2016, 2020; Xue et al. 2010; Kumar et al. 2020b).

The versatile nature of IDPs associated with folding, signaling, and many more, however, they are also implicated in many diseases. It is seen that selective mutations in IDPRs (i.e., amyloid β -peptide, α -synuclein, and huntingtin) may lead to structural complexity and enhanced aggregation propensity of these systems, which are associated with numerous neurodegenerative diseases (Babu et al. 2011; Uversky et al. 2008; Wu and Fuxreiter 2016). The IDPRs contain certain motif which are important for interaction and a slight change in these motif lead to altered cell signaling and thus cancer like diseases (Babu 2016; Hegyi et al. 2009; Colak et al. 2013).

According to the studies by disorder predictors, eukaryotic mammals are shown to contain nearly 75% of signaling proteins that contain long disordered regions with more than 30 residues, and about 25% of the predicted proteins are fully disordered

in nature (Dunker et al. 2008a). Furthermore, eukaryotic proteins utilize the disorder for transient binding purposes (signaling and regulation), while prokaryotic proteins seem to use disorder for complex formation (Dunker et al. 2008b). Another example of varied length intrinsically disordered proteins are Transmembrane proteins that contain extracellular or cytosolic disorder regions (Uversky 2013; Xue et al. 2009). A total of 40% of human integral plasma proteins is predicted to contain long stretched disordered regions (Minezaki et al. 2007; Yang et al. 2008; De Biasio et al. 2008). Disordered regions usually bind to multiple targets with low affinity, which is an ideal condition for signal transduction (Dunker et al. 2002). Some recent findings mention the functioning of ordered proteins on a decrease in the percentage of their ordered structure and need partial or complete functional misfolding (Uversky 2015).

13.4 Characterization of IDP and IDPRs

In the past two decades, rapid progress in the exploration of IDPs have radically changed the understanding and importance of the field. The high occurrence of IDPs in a cellular organization has increased the demand for new perspectives in structural and functional studies. The conformational flexibility of IDPs did not allow it to accurately study with old structural techniques. Therefore, it appeals to introduce new methods to study functional aspects in IDPs (Habchi et al. 2014). The structural studies of IDPs have shed light on the critical aspect that disorder lies in the amino-acid sequence of a protein. The thorough studies of IDP sequences and structural information suggest that disordered regions show low hydrophobicity and higher net charge, and characterized by low hydrodynamic radius, high structural heterogeneity, and poor secondary structure organization (Uversky 2019). But it may have a tendency to gain structural regions in presence of natural ligands. On the basis of these structural and sequence-based data, various algorithms have been designed to predict the structural disorder propensity of protein regions using disorder predictors. These bioinformatics disorder prediction tools are commonly used to characterize the protein disorderedness. The higher proportions of hydrophilic stretch of sequences are analyzed by a web server such as DISPROT (Megan Sickmeier et al. 2007), IUPred (Zsuzsanna Dosztányi et al. 2005), PONDR (Obradovic et al. 2003), PrDOS (Ishida and Kinoshita 2007), D²P² (Oates et al. 2013) and ESpritz (Walsh et al. 2012), which indicates the higher probability of disorder of those regions. Some of the test sets for structural predictions have been further confirmed by various experimental tools such as NMR, X-ray studies (Konrat 2014; Brutscher et al. 2015). This represents the high reliability of disorder predictors and improves the knowledge of the functional relevance of IDPs and IDPRs in various organisms.

13.5 Molecular Dynamics Simulations: Relevance with Structure Biology

The three-dimensional (3D) structure of biological macromolecules (e.g., proteins) or chemically synthesized polymers are essential for structural biology and applications in drug discovery. These days, the structure elucidation through X-ray, NMR, Cryo-EM techniques have been an advantage to understand their Structure-Function-Paradigm. However, there are many proteins which can not form rigid three-dimensional structures. So their thermodynamic properties, microscopic energies, and specific interaction with other molecules at the atomic-level cannot be understood well through experimental methods (Chong et al. 2017). Therefore, the characterization of proteins at the atomic level is more feasible through atomistic computational simulations rather than experiments. Molecular dynamics (MD) simulations are capable of determining conformational dynamics, structure compositions, and organization of proteins in an aqueous environment (Hollingsworth and Dror 2018). Additionally, the interaction of proteins with lipid molecules, inhibitors, with partner proteins, etc. can also be determined through MD simulations. Due to advancements in computer hardware, it is now possible to explore such macromolecules in a deeper level for longer timescale up to seconds to meet the experimental observations (Perilla et al. 2015). Various simulation packages such as Desmond (Bowers et al. 2006), Gromacs (Berendsen et al. 1995), Amber (Ponder and Case 2003), NAMD (Phillips et al. 2005), etc. are available with different optimized forcefields. Generally, a Forcefield (FF) can be explained as the interatomic potential energy of a system, which is calculated along with several parameters such as bonds, angle, torsion, dihedral, etc., defined on the atomic coordinates (Jorgensen and Tirado-Rives 2005; González 2011). The atoms which are held together by simple harmonic or elastic forces represent a molecule within the specified region for simulation (González 2011). Also, van der waal interactions and electrostatic interactions are the integral constituents of a forcefield. The overall equation that defines a forcefield is,

$$\begin{aligned}
 U(\text{potential energy}) = & \sum_{\text{bonds}} \frac{1}{2} k_b (r - r_0)^2 + \sum_{\text{angles}} \frac{1}{2} k_a (\theta - \theta_0)^2 + \sum_{\text{torsions}} \frac{V_n}{2} \\
 & \times [1 + \cos(n\theta - \delta)] + \sum_{\text{improper}} V_{\text{imp}} + \sum_{LJ} 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^6}{r_{ij}^6} \right) \\
 & + \sum_{\text{electrostatic}} \frac{q_i q_j}{r_{ij}}
 \end{aligned}$$

Where, oscillations about equilibrium bond length, bond angle, torsional rotation of 4 atoms about a central bond, and nonbonded energy terms (electrostatics and Lenard-Jones (LJ)) are summed up for calculating potential energy.

13.6 MD Force Fields and Their Role in Conformation Dynamics

Several forcefields have been developed to date and are being used with different purposes for investigation in almost every field of science. For biological macromolecules, OPLS (AA) (Jorgensen and Tirado-Rives 1988), GROMOS (Berendsen et al. 1995), CHARMM (Vanommeslaeghe et al. 2009), AMBER (Ponder and Case 2003), Drude (Li et al. 2017) forcefields are optimized to deduce the conformational change, structural composition, protein aggregation, binding efficacy with respect to time in a given environment. All these forcefields have a different level of tendency to estimate structural composition. However, for disordered proteins, it is essential to be picky for selection of accurate forcefields. A wonderful comparison has been made by Ham and colleagues between GROMOS, CHARMM, AMBER, and OPLS forcefields for correct selection of forcefield to perform MD of an IDP. Among all of them, OPLS-AA has proper balanced tendency to evaluate the helical and beta property of protein (Chong et al. 2017). Also, for IDPs, OPLS, and a recently introduced CHARMM36 (Huang et al. 2016) are used to simulate disordered regions properly, which allow them to gain a proper helical or beta structure, if induces. Two IDP models amyloid-beta and p53, have been extensively used as model systems for testing of different forcefields and correlating them with experiments. Pacheco and Strodel have investigated the accuracy of five forcefields on amyloid-beta (A β ; one of the responsible protein for Alzheimer's Disease) where CHARMM22, OPLS, Amber's 99sb, 99sb-ildn, had high accuracy with the NMR results than 99sbildn-NMR forcefield (Carballo-Pacheco and Strodel 2017). Recently, D.E. Shaw and colleagues have modified Amber-ff99sb forcefield for disordered proteins which correlate well with experimental observations. The improved forcefield ff99sb-disp accurately calculates the transition states between ordered and disordered states (Robustelli et al. 2018). Additionally, a short disordered peptide of 24 amino acids, Histatin 5, was investigated through verities of Amber and Gromos forcefields and compared with experiments (Henriques et al. 2015).

Along with a forcefield, the selection of water model for simulation is essential for precise evaluation of interaction and behavior of a protein in an aqueous environment. Water models are defined based on the interaction sites, which is centered on the nuclei of water molecule. Most commonly, water models are SPC (simple point charge; with HOH angle 109.47°), TIP3P (104.5°), and TIP4P (104.5°) for protein simulations. TIP3P and TIP4P water models are basically based on transferrable interaction potentials (TIPS) with three- and four-point charge (Ouyang and Bettens 2015). Three-point charge has three interaction sites as water molecule has three atoms, while four-point charge has an additional dummy atom to improve the electrostatic distribution (Jorgensen and Tirado-Rives 2005). Moreover, five-point and six-point charge water models are also available, which has dummy atoms representing the lone pairs and one extra site for interaction (Fig. 13.2). These water models are placed around the protein structure, which is centered in a

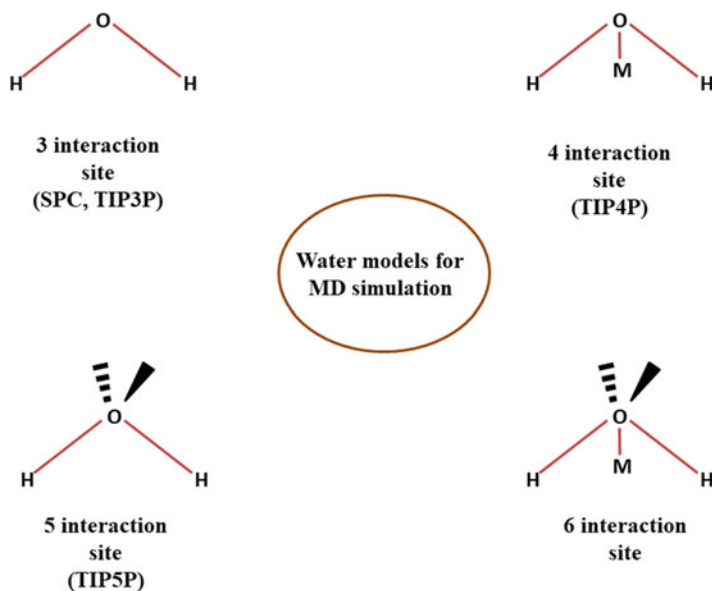


Fig. 13.2 Representation of water models based on different interaction sites. Here, O and H are oxygen and Hydrogen atoms while M represents the dummy atom in water models with 4, 5, and 6 interactions sites

simulation box (e.g. cubic, dodecahedron, etc.) with a defined size in the periodic cell.

13.7 MD Simulation Terminology: Structural Conformation Assessment

13.7.1 Energy Minimization, Equilibration, and Timescale

Building a simulation system to final production run, there are a number of steps for successfully trajectory generation. Before simulating a protein, the minimized structural conformation is vital as the system might produce erroneous results due to excess heat caused by unwanted and huge forces without minimization (Mackay et al. 1989). For simulating IDPs, a very popular method, steepest descent method is used for certain steps till it converged under required energy for the system.

Afterward, the simulation setup is subjected to equilibration as the minimized system has unoptimized solvents (e.g., water) around the proteins. Generally, the system is equilibrated under two conditions in constant temperature and pressure because for simulating a protein at a temperature, the system needs to be in proper positions without any unrestrained atoms. Two ensembles NPT and NVT are

Table 13.1 Tabulation of the timescale of different simulation techniques based on their applications

Sr. no.	Simulation methods	Applications	Timescale (seconds)
1.	Quantum mechanics (QM/MD)	Atomic motion (e.g., bond stretching, vibration)	10^{-15} to 10^{-12}
2.	All-atom MD	Structural transitions, side-chain rotation, loop movement, ligand binding	10^{-12} to 10^{-6}
3.	Coarse-grained MD	Biological assembly, protein folding, protein interaction	10^{-9} to 10^3

commonly used for equilibration in which the number of atoms (N), pressure (P), temperature (T), and volume (V) are kept constant and the system is processed for small simulation run upto few picoseconds or nanoseconds or till the system is equilibrated.

Finally, the production run for all-atom classical MD is performed at a constant temperature and pressure, which are maintained by specific thermostats and barostats, respectively. Thermostats like Nose Hoover, Berendsen, and Langevin while barostats such as Berendsen (Berendsen et al. 1984), Martyna-Tobias-Klein (MTK) (Lippert et al. 2013), Parrinello Rahman (Parrinello and Rahman 1980, 1981) are used as per the simulation setup requirements. In the case of IDPs, Nose Hoover (Posch et al. 1986) and Berendsen thermostats are preferred, and barostats MTK or Parrinello Rahman are considered for controlling temperature and pressure at an average value.

The simulation time is another important aspect to be looked carefully. As shown in above Table 13.1, the purpose of performing any simulation should be accomplished successfully within an adequate range of time. This timescale also depends on the number of atoms in a simulation setup and computer hardware. More number of atoms require more time and vice-versa, therefore, all atoms MD demands time from nanoseconds to few microseconds. However, apart from all-atoms MD simulation, Coarse-grain (CG) MD is one of the examples where a group of atoms are considered as a single bead, which reduces the degree of freedom among the atoms and allow the system for longer simulation in comparatively less time (Kmieciak et al. 2016).

13.7.2 RMSD, RMSF and Radius of Gyration

For the conformational analysis from the simulation trajectory, Root mean square deviation (RMSD) is the most critical parameter to be investigated. RMSD is measured between two sets of atoms (backbone, c-alpha, heavy, side-chain atoms) at a given time with reference to the initial one or the desired reference set of atoms (Kuzmanic and Zagrovic 2010). Similarly, the fluctuation in residues with respect to time can also be calculated as average of all simulated frames in the trajectory, it is

known as Root mean square fluctuation (Martínez 2015). Both these values deduce the conformational change of protein in the simulation environment over the time course. Another parameter, Radius of gyration (RoG) states the compactness of a protein structure. A well-folded or tightly packed structure will have lowest RoG value and vice-versa.

Many IDPs, which tend to get converted into an ordered form upon interaction with some physiological partners, are also vulnerable to different forcefields. In last two decades, the computer hardware power has been increased gradually and consequently, MD simulations have been very effective to explore the folding and unfolding of a protein in the presence of different conditions such as mixture of multiple solvents, membranes, ions, etc. at a constant or varying temperature and pH ranges. Also, various categories of simulation have made it easier to investigate the structural dynamics for longer timescale. A method where conformational swapping occurs between a number of MD replicas to obtain a conformation with minimum energy. These are known as Replica Exchange (RE) MD in which multiple replicas run simultaneously which are formed based on temperatures (lower to higher), required after selection through literature or experimental evidences (Sugita and Okamoto 1999). In next section, we have discussed the REMD with a well-suited example of p53, a tumor suppressor gene.

13.8 IDPs and Replica Exchange MD: In Perspective of p53-CTD

In our recent study, we performed a Replica exchange molecular dynamics simulation on p53-CTD using OPLS 2005 forcefield, embedded in Desmond simulation package (Bowers et al. 2006). As aforementioned, OPLS forcefield has a proper balance for alpha and beta propensity estimation in simulation. As illustrated in this study, the temperature induces changes in structural conformation of IDP (p53-CTD), which showed it's highly dynamics/flexible nature (Kumar et al. 2020a). The hydrophobic and electrostatic interactions play an important role in structural conformation of p53-CTD. The circular dichroism studies showed that the higher temperature leads to the compaction of a peptide, which is associated with the helical structural conformations (Kumar et al. 2020a; Kjaergaard et al. 2010). Previously, NMR studies showed that the temperature-induced structural conformation is associated with the random stretch of amino acid (non-helical) in a peptide (Kjaergaard et al. 2010). Our result also corroborated with these finding where MD simulation showed that the random stretch of amino acid or non-helical regions is responsible for change in structural transformations.

The p53-CTD adopts random coil conformations and have a tendency to gain structural conformation. According to REMD analysis, the highest structural compaction occurred at 80 °C where two major helical regions were formed (Fig. 13.3). The total potential energy of p53-CTD in aqueous system was calculated to be more

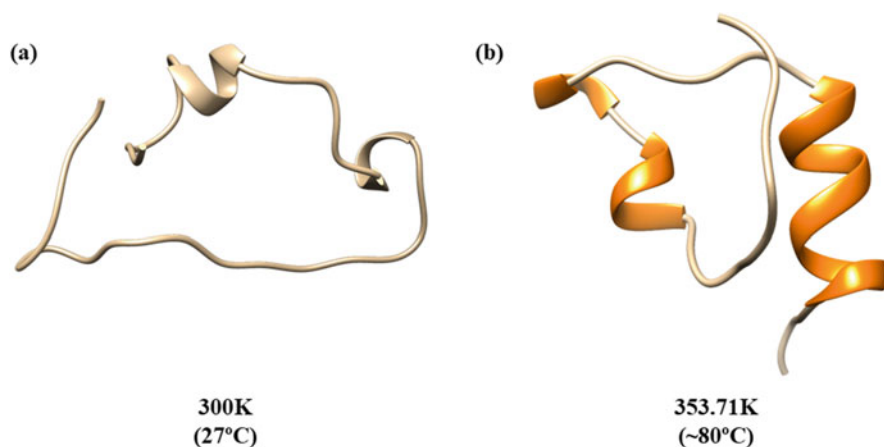


Fig. 13.3 Induction of helicity at two regions in p53-CTD at high temperature as obtained from Replica Exchange MD, using OPLS 2005 forcefield

negative (-33914.25 kcal/mol) at high temperature (353.71 K) while at 300 K, the potential energy of the system was -30382.4 kcal/mol. The p53 CTD could not fold spontaneously due to high intrinsic free energy. In the presence of binding partners, the p53-CTD possesses the minimum free energy and minimum binding energy from partner. Thus, the competitive effect of energy minimization results in different characteristic states, where each mechanism determines the sampling frequency of each characteristic state (Han et al. 2017).

13.9 Future Prospects

In general, IDPs are very challenging to study at atomistic detail with current computational simulation forcefields to achieve its accuracy with experimental models. A great advancement has been made to counter these challenges, but still, a lot of improvement needs to be done. However, Bioinformatics has made it possible to a large extent, and specifically, molecular dynamics simulations are being used extensively to explore IDPs at atomic level. The exact evaluation of structural properties of IDPs/IDPRs have been very distinguished among different forcefield. In this context, it can be seen as a high opportunity for development of new FF or improvement of current FF, which unravel the atomistic details on IDPs' conformational dynamics with respect to experimental measures. The outcome will certainly lead to a better understanding of biophysics of IDPs and pave a potential role in drug discovery.

Acknowledgements All the authors would like to thank IIT Mandi for providing facilities. The work was partially supported from DBT, Government of India (BT/11/IYBA/2018/06) to

RG. Funding. PK and NS were supported by MHRD for funding. AK was supported by DBT, Government of India (BT/11/YBA/2018/06).

Conflict of Interest All authors declare that there is no financial competing interest.

References

- Babu, M. M. (2016). The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochemical Society Transactions*, *44*, 1185–1200. <https://doi.org/10.1042/BST20160172>
- Babu, M. M., van der Lee, R., de Groot, N. S., & Gsponer, J. (2011). Intrinsically disordered proteins: Regulation and disease. *Current Opinion in Structural Biology*, *21*, 432–440. <https://doi.org/10.1016/j.sbi.2011.03.011>
- Beauchamp, K. A., Lin, Y. S., Das, R., & Pande, V. S. (2012). Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements. *Journal of Chemical Theory and Computation*, *8*, 1409–1414. <https://doi.org/10.1021/ct2007814>
- Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., Dinola, A., & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, *81*, 3684–3690. <https://doi.org/10.1063/1.448118>
- Berendsen, H. J. C., van der Spoel, D., & van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, *91*, 43–56. [https://doi.org/10.1016/0010-4655\(95\)00042-E](https://doi.org/10.1016/0010-4655(95)00042-E)
- 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. <https://doi.org/10.1021/ct500569b>
- Bowers, K. J., Bowers, K. J., Chow, E., Xu, H., Dror, R. O., Eastwood, M. P., et al. (2006). Scalable algorithms for molecular dynamics simulations on commodity clusters. In *Proceedings of the 2006 ACM/IEEE conference on supercomputing* (pp. 43–43). <https://doi.org/10.1109/SC.2006.54>
- Brutscher, B., Felli, I. C., Gil-Caballero, S., Hošek, T., Kümmerle, R., Piai, A., et al. (2015). NMR methods for the study of intrinsically disordered proteins structure, dynamics, and interactions: General overview and practical guidelines. *Advances in Experimental Medicine and Biology*, *870*, 49–122. https://doi.org/10.1007/978-3-319-20164-1_3
- Carballo-Pacheco, M., & Strodel, B. (2017). Comparison of force fields for Alzheimer's A β 42: A case study for intrinsically disordered proteins. *Protein Science*, *26*, 174–185. <https://doi.org/10.1002/pro.3064>
- Chen, J. W., Romero, P., Uversky, V. N., & Dunker, A. K. (2006). Conservation of intrinsic disorder in protein domains and families: II. Functions of conserved disorder. *Journal of Proteome Research*, *5*, 888–898. <https://doi.org/10.1021/pr060049p>
- Chong, S.-H., Chatterjee, P., & Ham, S. (2017). Computer simulations of intrinsically disordered proteins. *Annual Review of Physical Chemistry*, *68*, 117–134. <https://doi.org/10.1146/annurev-physchem-052516-050843>
- Colak, R., Kim, T., Michaut, M., Sun, M., Irimia, M., Bellay, J., et al. (2013). Distinct types of disorder in the human proteome: Functional implications for alternative splicing. *PLoS Computational Biology*, *9*, e1003030. <https://doi.org/10.1371/journal.pcbi.1003030>
- De Biasio, A., Guarnaccia, C., Popovic, M., Uversky, V. N., Pintar, A., & Pongor, S. (2008). Prevalence of intrinsic disorder in the intracellular region of human single-pass type I proteins: The case of the notch ligand delta-4. *Journal of Proteome Research*, *7*, 2496–2506. <https://doi.org/10.1021/pr800063u>

- Dunker, A. K., & Obradovic, Z. (2001). The protein trinity – Linking function and disorder. *Nature Biotechnology*, *19*, 805–806. <https://doi.org/10.1038/nbt0901-805>
- Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., et al. (2001). Intrinsically disordered protein. *Journal of Molecular Graphics & Modelling*, *19*, 26–59. [https://doi.org/10.1016/S1093-3263\(00\)00138-8](https://doi.org/10.1016/S1093-3263(00)00138-8).
- Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., & Obradović, Z. (2002). Intrinsic disorder and protein fluctuations. *Biochemistry*, *41*, 6573–6582. <https://doi.org/10.1021/BI012159+>
- Dunker, A. K., Silman, I., Uversky, V. N., & Sussman, J. L. (2008a). Function and structure of inherently disordered proteins. *Current Opinion in Structural Biology*, *18*, 756–764. <https://doi.org/10.1016/j.sbi.2008.10.002>
- Dunker, A. K., Oldfield, C. J., Meng, J., Romero, P., Yang, J. Y., Chen, J. W., et al. (2008b). The unfoldomics decade: An update on intrinsically disordered proteins. *BMC Genomics*, *9*, 1–26. <https://doi.org/10.1186/1471-2164-9-S2-S1>
- Dunker, A. K., Babu, M. M., Barbar, E., Blackledge, M., Bondos, S. E., Dosztányi, Z., et al. (2013). What’s in a name? Why these proteins are intrinsically disordered. *Intrinsically Disordered Proteins*, *1*, e24157. <https://doi.org/10.4161/idp.24157>
- Fadda, E., & Nixon, M. G. (2017). The transient manifold structure of the p53 extreme C-terminal domain: Insight into disorder, recognition, and binding promiscuity by molecular dynamics simulations. *Physical Chemistry Chemical Physics*, *19*, 21287–21296. <https://doi.org/10.1039/c7cp02485a>
- Gadhave, K., Gehi, B. R., Kumar, P., Xue, B., Uversky, V. N., & Giri, R. (2020). The dark side of Alzheimer’s disease: Unstructured biology of proteins from the amyloid cascade signaling pathway. *Cellular and Molecular Life Sciences*, 1–46. <https://doi.org/10.1007/s00018-019-03414-9>
- Garg, N., Kumar, P., Gadhave, K., & Giri, R. (2019). The dark proteome of cancer: Intrinsic disorderedness and functionality of HIF-1 α along with its interacting proteins. *Progress in Molecular Biology and Translational Science*, *166*, 371–403. <https://doi.org/10.1016/BS.PMBTS.2019.05.006>
- Giri, R., Kumar, D., Sharma, N., & Uversky, V. N. (2016). Intrinsically disordered side of the zika virus proteome. *Frontiers in Cellular and Infection Microbiology*, *6*, 144. <https://doi.org/10.3389/fcimb.2016.00144>
- Giri, R., Bhardwaj, T., Shegane, M., Gehi, B. R., Kumar, P., & Gadhave, K. (2020). Dark proteome of newly emerged SARS-CoV-2 in comparison with human and bat coronaviruses. *BioRxiv*. <https://doi.org/10.1101/2020.03.13.990598>
- González, M. A. (2011). Force fields and molecular dynamics simulations. *Collection SFN*, *12*, 169–200. <https://doi.org/10.1051/sfn/201112009>
- Habchi, J., Tompa, P., Longhi, S., & Uversky, V. N. (2014). Introducing protein intrinsic disorder. *Chemical Reviews*, *114*(13), 6561–6588. <https://doi.org/10.1021/cr400514h>
- Han, M., Xu, J., & Ren, Y. (2017). Compromise in competition between free energy and binding effect of intrinsically disordered protein p53 C-terminal domain. *Molecular Simulation*, *43*, 110–120. <https://doi.org/10.1080/08927022.2016.1237023>
- Hegyí, H., Buday, L., & Tompa, P. (2009). Intrinsic structural disorder confers cellular viability on oncogenic fusion proteins. *PLoS Computational Biology*, *5*, e1000552. <https://doi.org/10.1371/journal.pcbi.1000552>
- Henriques, J., Cragnell, C., & Skepö, 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. <https://doi.org/10.1021/ct501178z>
- Hollingsworth, S. A., & Dror, R. O. (2018). Molecular dynamics simulation for all. *Neuron*, *99*, 1129–1143. <https://doi.org/10.1016/j.neuron.2018.08.011>
- Huang, Y., & Liu, Z. (2009). Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: A critical assessment of the “fly-casting” mechanism. *Journal of Molecular Biology*, *393*, 1143–1159. <https://doi.org/10.1016/j.jmb.2009.09.010>

- Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., De Groot, B. L., et al. (2016). CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nature Methods*, *14*, 71–73. <https://doi.org/10.1038/nmeth.4067>
- Ishida, T., & Kinoshita, K. (2007). PrDOS: Prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Research*, *35*, W460–W464. https://academic.oup.com/nar/article/35/suppl_2/W460/2923303. Accessed 30 May 2020
- Jorgensen, W. L., & Tirado-Rives, J. (1988). The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, *110*, 1657–1666. <https://doi.org/10.1021/ja00214a001>
- Jorgensen, W. L., & Tirado-Rives, J. (2005). Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 6665–6670. <https://doi.org/10.1073/pnas.0408037102>
- Kannan, S., Lane, D. P., & Verma, C. S. (2016). Long range recognition and selection in IDPs: The interactions of the C-terminus of p53. *Scientific Reports*, *6*, 1–13. <https://doi.org/10.1038/srep23750>
- Kjaergaard, M., Nørholm, A.-B. B., Hendus-Altenburger, R., Pedersen, S. F., Poulsen, F. M., & Kragelund, B. B. (2010). Temperature-dependent structural changes in intrinsically disordered proteins: Formation of α -helices or loss of polyproline II? *Protein Science*, *19*, 1555–1564. <https://doi.org/10.1002/pro.435>
- Kmieciak, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A. E., & Kolinski, A. (2016). Coarse-grained protein models and their applications. *Chemical Reviews*, *116*, 7898–7936. <https://doi.org/10.1021/acs.chemrev.6b00163>
- Konrat, R. (2014). NMR contributions to structural dynamics studies of intrinsically disordered proteins. *Journal of Magnetic Resonance*, *241*, 74–85. <https://doi.org/10.1016/j.jmr.2013.11.011>
- Kumar, D., Sharma, N., & Giri, R. (2017). Therapeutic interventions of cancers using intrinsically disordered proteins as drug targets: C-Myc as model system. *Cancer Informatics*, *16*, 117693511769940. <https://doi.org/10.1177/1176935117699408>
- Kumar, P., Saumya, K. U., & Giri, R. (2019). Identification of peptidomimetic compounds as potential inhibitors against MurA enzyme of *Mycobacterium tuberculosis*. *Journal of Biomolecular Structure & Dynamics*, 1–21. <https://doi.org/10.1080/07391102.2019.1696231>
- Kumar, A., Kumar, P., Kumari, S., Uversky, V. N., & Giri, R. (2020a). Folding and structural polymorphism of p53 C-terminal domain: One peptide with many conformations. *Archives of Biochemistry and Biophysics*, *684*, 108342. <https://doi.org/10.1016/j.abb.2020.108342>
- Kumar, D., Singh, A., Kumar, P., Uversky, V. N., Rao, C. D., & Giri, R. (2020b). Understanding the penetrance of intrinsic protein disorder in rotavirus proteome. *International Journal of Biological Macromolecules*, *144*, 892–908. <https://doi.org/10.1016/j.ijbiomac.2019.09.166>
- Kuzmanic, A., & Zagrovic, B. (2010). Determination of ensemble-average pairwise root mean-square deviation from experimental B-factors. *Biophysical Journal*, *98*, 861–871. <https://doi.org/10.1016/j.bpj.2009.11.011>
- Lange, O. F., Van Der Spoel, D., & De Groot, B. L. (2010). Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR data. *Biophysical Journal*, *99*, 647–655. <https://doi.org/10.1016/j.bpj.2010.04.062>
- Le Gall, T., Romero, P. R., Cortese, M. S., Uversky, V. N., & Dunker, A. K. (2007). Intrinsic disorder in the protein data bank. *Journal of Biomolecular Structure & Dynamics*, *24*, 325–341. <https://doi.org/10.1080/07391102.2007.10507123>
- Li, H., Chowdhary, J., Huang, L., He, X., MacKerell, A. D., & Roux, B. (2017). Drude polarizable force field for molecular dynamics simulations of saturated and unsaturated zwitterionic lipids. *Journal of Chemical Theory and Computation*, *13*, 4535–4552. <https://doi.org/10.1021/acs.jctc.7b00262>

- Lindorff-Larsen, K., Maragakis, P., Piana, S., Eastwood, M. P., Dror, R. O., & Shaw, D. E. (2012). Systematic validation of protein force fields against experimental data. *PLoS One*, 7, e32131. <https://doi.org/10.1371/journal.pone.0032131>
- Lindorff-Larsen, K., Piana, S., Dror, R. O., & Shaw, D. E. (2013). How fast-folding proteins fold. *Science*, 334, 517–520. <https://doi.org/10.1126/science.1208351>
- Lippert, R. A., Predescu, C., Ierardi, D. J., Mackenzie, K. M., Eastwood, M. P., Dror, R. O., et al. (2013). Accurate and efficient integration for molecular dynamics simulations at constant temperature and pressure. *The Journal of Chemical Physics*, 139, 164106. <https://doi.org/10.1063/1.4825247>
- Lopes, J. L. S., Orcia, D., Araujo, A. P. U., Demarco, R., & Wallace, B. A. (2013). Folding factors and partners for the intrinsically disordered protein micro-exon gene 14 (MEG-14). *Biophysical Journal*, 104, 2512–2520. <https://doi.org/10.1016/j.bpj.2013.03.063>
- Mackay, D. H. J., Cross, A. J., & Hagler, A. T. (1989). The role of energy minimization in simulation strategies of biomolecular systems. In *Prediction of protein structure and the principles of protein conformation* (pp. 317–358). Boston: Springer. https://doi.org/10.1007/978-1-4613-1571-1_7
- Martínez, L. (2015). Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis. *PLoS One*, 10, e0119264. <https://doi.org/10.1371/journal.pone.0119264>
- Megan Sickmeier, A. K. D., Hamilton, J. A., LeGall, T., Vacic, V., Cortese, M. S., Tantos, A., et al. (2007). DisProt: The database of disordered proteins. *Nucleic Acids Research*, 35, D786–D793. https://pubmed.ncbi.nlm.nih.gov/17145717/?from_single_result=17145717&expanded_search_query=17145717. Accessed 30 May 2020
- Minezaki, Y., Homma, K., & Nishikawa, K. (2007). Intrinsically disordered regions of human plasma membrane proteins preferentially occur in the cytoplasmic segment. *Journal of Molecular Biology*, 368, 902–913. <https://doi.org/10.1016/j.jmb.2007.02.033>
- Mishra, P. M., Uversky, V. N., & Giri, R. (2018). Molecular recognition features in zika virus proteome. *Journal of Molecular Biology*, 430, 2372–2388. <https://doi.org/10.1016/j.jmb.2017.10.018>
- Mittal, J., & Best, R. B. (2010). Tackling force-field bias in protein folding simulations: Folding of Villin HP35 and Pin WW domains in explicit water. *Biophysical Journal*, 99, L26–L28. <https://doi.org/10.1016/j.bpj.2010.05.005>
- Mohan, A., Oldfield, C. J., Radivojac, P., Vacic, V., Cortese, M. S., Dunker, A. K., et al. (2006). Analysis of molecular recognition features (MoRFs). *Journal of Molecular Biology*, 362, 1043–1059. <https://doi.org/10.1016/j.jmb.2006.07.087>
- Nerenberg, P. S., Jo, B., So, C., Tripathy, A., & Head-Gordon, T. (2012). Optimizing solute-water van der Waals interactions to reproduce solvation free energies. *The Journal of Physical Chemistry B*, 116, 4524–4534. <https://doi.org/10.1021/jp2118373>
- Nishimura, C., Lietzow, M. A., Dyson, H. J., & Wright, P. E. (2005). Sequence determinants of a protein folding pathway. *Journal of Molecular Biology*, 351, 383–392. <https://doi.org/10.1016/j.jmb.2005.06.017>
- Oates, M. E., Romero, P., Ishida, T., Ghalwash, M., Mizianty, M. J., Xue, B., et al. (2013). D²P²: Database of disordered protein predictions. *Nucleic Acids Research*, 41, D508–D516. <https://pubmed.ncbi.nlm.nih.gov/23203878/>. Accessed 30 May 2020.
- Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C. J., & Dunker, A. K. (2003). Predicting intrinsic disorder from amino acid sequence. *Proteins: Structure, Function, and Bioinformatics*, 53, 566–572. <https://doi.org/10.1002/prot.10532>
- Oldfield, C. J., Cheng, Y., Cortese, M. S., Romero, P., Uversky, V. N., & Dunker, A. K. (2005). Coupled folding and binding with α -helix-forming molecular recognition elements. *Biochemistry*, 44, 12454–12470. <https://doi.org/10.1021/bi050736e>
- Oldfield, C. J., Meng, J., Yang, J. Y., Qu, M. Q., Uversky, V. N., & 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, S1. <https://doi.org/10.1186/1471-2164-9-S1-S1>

- Ouyang, J. F., & Bettens, R. P. A. (2015). Modelling water: A lifetime enigma. *CHIMIA International Journal for Chemistry*, 69, 104–111. <https://doi.org/10.2533/chimia.2015.104>
- Parrinello, M., & Rahman, A. (1980). Crystal structure and pair potentials: A molecular-dynamics study. *Physical Review Letters*, 45, 1196–1199. <https://doi.org/10.1103/PhysRevLett.45.1196>
- Parrinello, M., & Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52, 7182–7190. <https://doi.org/10.1063/1.328693>
- Perilla, J. R., Goh, B. C., Cassidy, C. K., Liu, B., Bernardi, R. C., Rudack, T., et al. (2015). Molecular dynamics simulations of large macromolecular complexes. *Current Opinion in Structural Biology*, 31, 64–74. <https://doi.org/10.1016/j.sbi.2015.03.007>
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26, 1781–1802. <https://doi.org/10.1002/jcc.20289>
- Piana, S., Donchev, A. G., Robustelli, P., & Shaw, D. E. (2015). Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *The Journal of Physical Chemistry. B*, 119, 5113–5123. <https://doi.org/10.1021/jp508971m>
- Ponder, J. W., & Case, D. A. (2003). *Force fields for protein simulations*. Hoboken: Elsevier.
- Posch, H. A., Hoover, W. G., & Vesely, F. J. (1986). Canonical dynamics of the Nosé oscillator: Stability, order, and chaos. *Physical Review A*, 33, 4253–4265. <https://doi.org/10.1103/PhysRevA.33.4253>
- Robustelli, P., Piana, S., & Shaw, D. E. (2018). Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences of the United States of America*, 115, E4758–E4766. <https://doi.org/10.1073/pnas.1800690115>
- Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., & Dunker, A. K. (2001). Sequence complexity of disordered protein. *Proteins*, 42, 38–48. [https://doi.org/10.1002/1097-0134\(20010101\)42:1<38::AID-PROT50>3.0.CO;2-3](https://doi.org/10.1002/1097-0134(20010101)42:1<38::AID-PROT50>3.0.CO;2-3)
- Schad, E., Tompa, P., & Hegyi, H. (2011). The relationship between proteome size, structural disorder and organism complexity. *Genome Biology*, 12, R120. <https://doi.org/10.1186/gb-2011-12-12-r120>
- Shoemaker, B. A., Portman, J. J., & Wolynes, P. G. (2000). Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 8868–8873. <https://doi.org/10.1073/pnas.160259697>
- Singh, A., Kumar, A., Yadav, R., Uversky, V. N., & Giri, R. (2018a). Deciphering the dark proteome of Chikungunya virus. *Scientific Reports*, 8, 5822. <https://doi.org/10.1038/s41598-018-23969-0>
- Singh, A., Kumar, A., Uversky, V. N., & Giri, R. (2018b). Understanding the interactability of chikungunya virus proteins via molecular recognition feature analysis. *RSC Advances*, 8, 27293–27303. <https://doi.org/10.1039/C8RA04760J>
- Sugita, Y., & Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314, 141–151. [https://doi.org/10.1016/S0009-2614\(99\)01123-9](https://doi.org/10.1016/S0009-2614(99)01123-9)
- Tompa, P. (2005). The interplay between structure and function in intrinsically unstructured proteins. *FEBS Letters*, 579, 3346–3354. <https://doi.org/10.1016/j.febslet.2005.03.072>
- Tompa, P., & Fuxreiter, M. (2008). Fuzzy complexes: Polymorphism and structural disorder in protein-protein interactions. *Trends in Biochemical Sciences*, 33, 2–8. <https://doi.org/10.1016/j.tibs.2007.10.003>
- Uversky, V. N. (2009). *Intrinsically disordered proteins and their environment: Effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding*. Berlin: Springer. <https://doi.org/10.1007/s10930-009-9201-4>
- Uversky, V. N. (2013). A decade and a half of protein intrinsic disorder: Biology still waits for physics. *Protein Science*, 22, 693–724. <https://doi.org/10.1002/pro.2261>

- Uversky, V. N. (2015). Intrinsically disordered proteins and their (disordered) proteomes in neurodegenerative disorders. *Frontiers in Aging Neuroscience*, 7, 18. <https://doi.org/10.3389/fnagi.2015.00018>
- Uversky, V. N. (2019). Intrinsically disordered proteins and their “mysterious” (meta)physics. *Frontiers of Physics*, 7, 10. <https://doi.org/10.3389/fphy.2019.00010>
- Uversky, V. N., & Dunker, A. K. (2010). Understanding protein non-folding. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1804, 1231–1264. <https://doi.org/10.1016/j.bbapap.2010.01.017>
- Uversky, V. N., & Dunker, A. K. (2013). The case for intrinsically disordered proteins playing contributory roles in molecular recognition without a stable 3D structure. *F1000 Biology Reports*, 5, 1. <https://doi.org/10.3410/B5-1>
- Uversky, V. N., Oldfield, C. J., & Dunker, A. K. (2005). Showing your ID: Intrinsic disorder as an ID for recognition, regulation and cell signaling. *Journal of Molecular Recognition*, 18, 343–384. <https://doi.org/10.1002/jmr.747>
- Uversky, V. N., Oldfield, C. J., & Dunker, A. K. (2008). Intrinsically disordered proteins in human diseases: Introducing the D 2 concept. *Annual Review of Biophysics*, 37, 215–246. <https://doi.org/10.1146/annurev.biophys.37.032807.125924>
- Van Der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., et al. (2014). Classification of intrinsically disordered regions and proteins. *Chemical Reviews*, 114, 6589–6631. <https://doi.org/10.1021/cr400525m>
- Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., et al. (2009). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry*, 31, NA–NA. <https://doi.org/10.1002/jcc.21367>
- Walsh, I., Martin, A. J., Di Domenico, T., & Tosatto, S. C. (2012). ESpritz: Accurate and fast prediction of protein disorder. *Bioinformatics*, 28, 503–509. <https://pubmed.ncbi.nlm.nih.gov/22190692/>. Accessed 30 May 2020
- Williams, R. M., Obradovic, Z., Mathura, V., Braun, W., Garner, E. C., Young, J., et al. (2000). The protein non-folding problem: Amino acid determinants of intrinsic order and disorder. In *Biocomputing 2001* (pp. 89–100). Singapore: World Scientific. https://doi.org/10.1142/9789814447362_0010.
- Wright, P. E., & Dyson, H. J. (1999). Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *Journal of Molecular Biology*, 293, 321–331. <https://doi.org/10.1006/jmbi.1999.3110>
- Wu, H., & Fuxreiter, M. (2016). The structure and dynamics of higher-order assemblies: Amyloids, signalosomes, and granules. *Cell*, 165, 1055–1066. <https://doi.org/10.1016/j.cell.2016.05.004>
- Xue, B., Li, L., Meroueh, S. O., Uversky, V. N., & Dunker, A. K. (2009). Analysis of structured and intrinsically disordered regions of transmembrane proteins. *Molecular BioSystems*, 5, 1688–1702. <https://doi.org/10.1039/b905913j>
- Xue, B., Williams, R., Oldfield, C., Kian-Meng Goh, G., Keith Dunker, A., & N. Uversky, V. (2010). Viral disorder or disordered viruses: Do viral proteins possess unique features? *Protein and Peptide Letters*, 17, 932–951. <https://doi.org/10.2174/092986610791498984>
- Yang, J. Y., Yang, M. Q., Dunker, A. K., Deng, Y., & Huang, X. (2008). Investigation of transmembrane proteins using a computational approach. *BMC Genomics*, 9, S7. <https://doi.org/10.1186/1471-2164-9-S1-S7>
- Zhang, T., Faraggi, E., Li, Z., & Zhou, Y. (2013). Intrinsically semi-disordered state and its role in induced folding and protein aggregation. *Cell Biochemistry and Biophysics*, 67, 1193–1205. <https://doi.org/10.1007/s12013-013-9638-0>
- Zsuzsanna Dosztányi, I. S., Csizmok, V., & Tompa, P. (2005). IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, 21, 3433–3434. https://pubmed.ncbi.nlm.nih.gov/15955779/?from_single_result=15955779&expanded_search_query=15955779. Accessed 30 May 2020