INTERNATIONAL RESEARCH JOURNAL OF PHARMACY

www.irjponline.com

ISSN 2230 – 8407

Review Article

INTRINSICALLY DISORDERED PROTEINS (IDPs) IN HUMAN DISEASES: A REVIEW

Divya Shaji *
Independent Researcher, Kerala, India
*Corresponding Author Email: divyas.bioinfo@gmail.com

Article Received on: 06/09/18 Approved for publication: 12/10/18

DOI: 10.7897/2230-8407.0911247

ABSTRACT

Biologically active proteins without stable tertiary structure are called Intrinsically Disordered Proteins (IDPs) or Intrinsically Unstructured Proteins (IUPs). These IDPs are highly abundant in nature and are involved in regulation, signaling, and control. Structures and functions of IDPs are controlled and modulated by alternative splicing and posttranslational modifications. Many proteins associated with human diseases are IDPs. Each of these diseases originates from the misfolding or dysfunction of a specific protein. This review presents some of the IDPs involved in human diseases.

Keywords: Intrinsically disordered proteins; Neurodegenerative diseases; Diabetes; Cancer; Drug discovery; IDPs; IUPs; natively unfolded proteins; human diseases.

INTRODUCTION

Intrinsically Disordered Proteins (IDPs) are proteins that have no well-defined structures under physiological conditions but still have biological functions[1,2]. IDPs are also called natively unfolded, natively denatured, and intrinsically unstructured proteins (IUPs)[1, 3, 4]. Although they do not have ordered structures in the unbound (free) state under physiological conditions, they undergo a disorder-to-order transition upon binding to their biological partners via coupled folding and binding[5, 6]. Coupled folding and binding might involve just a few residues or an entire protein domain[2, 7]. These IDPs are highly abundant in nature and have numerous biological activities[8]. IDPs are more abundant in eukaryotic proteomes than in archaea and prokaryotes[9].

Some proteins are predicted to be entirely disordered, while others contain disordered sequences, named to as intrinsically disordered regions (IDRs), in combination with structured globular domains. The majority of proteins in eukaryotic proteomes contain both intrinsically disordered and ordered regions. There are two major classes of protein disorder; short regions and long regions. The short regions are typically <15-20 residues which serve as flexible linkers between or within domains, and long regions are >30–50 residues. These two classes have different amino acid propensities[10]. Intrinsically disordered proteins have low sequence complexity and amino acid compositional bias, with a low content of hydrophobic amino acids (Val, Leu, Ile, Met, Phe, Trp and Tyr), and a high content of particular polar and charged amino acids (Gln, Ser, Pro, Glu, Lys, Gly and Ala). The hydrophobic amino acids normally form the core of a folded globular protein[2, 11, 12].

INTERACTIONS OF IDPs

Disordered segments adopt a largely extended and open conformation in the complex. Generally shorter regions undergo disorder-to-order transition than long regions. These short regions are usually below 100 residues; in many cases the length of the disordered binding regions are less than 30 residues. Compared to the globular proteins, the interface of disordered proteins is more hydrophobic, and the interaction contacts are also significantly different. IDPs tend to favor hydrophobic-hydrophobic contacts with the partner proteins at the interface[13]. Polar and charged residues also play a larger role in the interfaces of intrinsically disordered proteins compared to the interfaces of globular proteins. This suggests that polar interactions are key contributors to the specificity of interactions that involve intrinsically disordered proteins[14]. Intrinsically disordered (ID) regions provide large surface area. By providing larger interaction area, ID regions can support interactions with several molecules to form large multimeric complexes[15].

IDPs IN HUMAN DISEASES

IDPs are abundantly involved in the development of various diseases, including cancer, cardiovascular diseases, amyloidoses, neurodegenerative diseases, and diabetes. Disease related IDPs are attractive targets for drugs modulating protein-protein interactions[16]. Each of these diseases originates from the misfolding or dysfunction of a particular protein. Some disease-related proteins have an intrinsic propensity to form pathologic conformation(s). For other proteins, interactions or impaired interactions with other proteins, small molecules, and other endogenous factors can induce conformational changes and increase the propensity to misfold. Misfolding and dysfunction can be caused by point mutation(s), impaired Post Translational Modifications (PTMs), an increased probability of degradation, impaired trafficking, loss of binding partners, or oxidative damage[17, 18].

Some examples of IDPs which are involved in Neurodegenerative diseases, Cancer, and Type-2 Diabetes are discussed below.

IDPs IN NEURODEGENERATIVE DISEASES

α-synuclein

α-synuclein is one of the most studied disease-related IDPs[19], that is related to Parkinson's disease (PD). Recent studies suggested

that α-synuclein plays a central role in the pathogenesis of PD. α-synuclein contains 140 amino acids with three domains: an N-terminal domain (aa 1–65), a non-amyloid-β component of plaques (NAC) domain (aa 66–95), and a C-terminal domain (aa 96–140) [20]. α-Synuclein protein (PDB ID: 1xq8) is highly disordered when isolated in solution[21]. The three-dimensional structure of human α-synuclein is shown in Fig 1.

Tau protein

Tau protein is involved in the pathology of various neurodegenerative diseases, including Alzheimer's disease[21]. Human tau (PDB ID: 2N4R) is encoded by the *MAPT* gene, located on chromosome 17 which contains 441 amino acids with four major domains. Biophysical studies have revealed that Tau is an intrinsically disordered protein (contains a low proportion of hydrophobic amino acids) which maintains a highly flexible confirmation with a low content of secondary structure[22, 23]. The three-dimensional structure of human Tau protein is shown in Fig 2.

Amyloid β-protein

Amyloid deposits in Alzheimer's disease contain the amyloid β-protein (Aβ), which is a 40–42 residue peptide produced by endoproteolytic cleavage of the amyloid precursor protein (APP). Amyloid β-protein is an intrinsically disordered protein which contains a very low content of hydrophobic amino acids and a high content of charged residues[21]. NMR studies have shown that monomers of Aβ1–40, or Aβ1–42 possess no α-helical or β-sheet structure[21, 24].

Polyglutamine (PolyQ) repeat diseases

PolyQ diseases are a group of neurodegenerative disorders caused by the expansion of GAC trinucleotide repeats that code for polyQ in the gene products[25, 26]. These polyglutamine repeat diseases include Huntington's disease (HD), Kennedy disease (also known as spinal and bulbar muscular atrophy), spinocerebellar ataxia type 1 (SCA1), dentatorubral–pallidoluysian atrophy (DRPLA), spinocerebellar ataxia type 2 (SCA2), Machado–Joseph disease (MJD/SCA3), SCA6, SCA7 and SCA17. The polyQ repeat varies between 16 and 37 residues in healthy individuals, and individuals who are afflicted by disease have repeats of more than 38 residues. Several proteins responsible for the pathogenesis of polyQ repeat diseases were predicted to be either completely disordered or to contain long disordered regions. The examples include androgen receptor in SBMA, atrophin-1 in DRPLA, ataxin-2 in SCA2 , ataxin-3 in SCA3/MJD , P/Q-type calcium channel α1A subunit in SCA6 , ataxin-7 in SCA7 and TATA-box-binding protein in SCA17[19].

SOD1

SOD1 is an abundant protein found in the cytosol, the nucleus, peroxisomes and the mitochondrial intermembrane space of human cells. Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disorder that caused by mutations in the gene that encodes the antioxidant enzyme copper-zinc superoxide dismutase (SOD1). Structural analysis has revealed that the regions 22–30, 55–95 and 123–142 of human SOD1 are likely to be intrinsically disordered[19]. The three-dimensional structure of human SOD1 is shown in Fig. 3.

IDPs IN CANCER

p53

p53 is one of the most studied intrinsically disordered protein, consisting of ordered and disordered regulatory regions engaged in multiple interactions[27]. For each interaction, only a short region of p53 typically becomes structured upon binding[19]. p53 also known as tumor protein 53 or tumor antigen p53 is a tumor-suppressor protein encoded by the TP53 gene in humans[19, 28]. The p53 protein consists of 393 residues and three functional regions: (i) a N-terminal domain (residues 1–93) containing a transcriptional activation domain and a proline-rich domain; (ii) a core DNA-binding globular domain (residues 102–292) and (iii) a C-terminal domain (CTD) consisting of a tetramerization domain (residues 320–356) and a regulatory domain (residues 363–393)[29]. 50–55% of human cancers are associated with mutations in the p53 gene[19,28] . These mutations frequently occur in the DNA-binding domain of p53[19,27, 28]. The three-dimensional structure of human P53 protein is shown in Fig 4.

c-Myc

Myc proteins are transcription factors (TFs) that are involved in several physiological processes, such as cell proliferation, apoptosis, metabolism, and biosynthesis of proteins. c-Myc is the hot spot for understanding and developing therapeutics against cancers and cancer stem cells. Disorder prediction studies show that the C-terminal region (410-490) of c-Myc is the most conserved region in protein disorder pattern and in the absence of binding partner the N-terminal region is also unstructured. Computational and experimental researches have shown that c-Myc extensively uses its disorder regions to perform diverse interactions[30]. Deregulation of the c-Myc transcription factor is involved in almost 50% of human cancers[19,30]. A high percentage of cancer-associated proteins have long disordered regions[29]. Computational studies have revealed that a majority of the cancer/testis antigens are intrinsically disordered proteins[31].

BRCA1

Mutations in BRCA1 (breast cancer type 1 susceptibility protein) gene are involved in the formation of breast and ovarian cancer. Moreover, this protein undergoes alternative splicing events associated with cancer[31,32]. BRCA1 contains 1863 amino acids with an amino terminal zinc ring finger motif, two nuclear localization signals, and two C-terminally located BRCT domains[31]. Structural and functional analysis have shown that the long central region of BRCA1 (residues 170–1645) is an intrinsically disordered scaffold which are involved in multiple protein-protein and protein-DNA interactions[31,33].

PTEN

PTEN (phosphatase and tensin homolog, deleted on chromosome TEN), is a tumor suppressor which is involved in various types of cancers[31, 34]. PTEN is the second most frequently mutated tumor suppressor gene after p53[17, 31]. Intron 3 (SV3) and intron 5 (SV5) regions of PTEN splice variants have been found in breast cancers[31, 32]. PTEN consists of the intrinsically disordered N-terminal phosphatidyl inositol-bisphosphate (PIP2) binding module (PBM), the dual-specificity lipid and protein phosphatase domain (PD), and the C-terminal intrinsically disordered region (C-tail). PTEN-long (a longer variant of PTEN) contains additional 173 amino acids at its N-terminus (the N-173 region)[31]. This N-173 region is highly disordered[31, 35].

IDPs IN TYPE 2 DIABETES MELLITUS (T2DM)

MAF

Musculoaponeurotic fibrosarcoma (MAF) proteins are transcription factors that are involved in cell differentiation. MAF family is classified into large MAF transcription factors and small MAF transcription factors. Large MAFs consists of four proteins: MAFA, MAFB, c-MAF and Nrl (neural retina-specific leucine). Previous computational research has revealed that human MAFA possess a high level of intrinsic disorder and contains multiple sites of various posttranslational modifications (PTMs)[36]. Earlier research has revealed that the C- terminal DNA binding domain of human MAFA (residues 227-301) is intrinsically disordered in solution and undergoes a disorder to order transition after binding with insulin MARE[36,37].

IRS

Insulin receptor substrates (IRSs) are a family of cytoplasmic adaptor proteins that mediate many of the key metabolic actions of insulin. IRS family consists of four proteins IRS1, IRS2, IRS3, and IRS4. Computational analysis has indicated that IRS1 (residues 160–264), IRS2 (residues 194–298), and IRS4 (residues 231–235) are highly disordered. These proteins are involved in various PTMs and these PTMS are located within the highly disordered regions of IRS proteins. These three proteins (IRS1, IRS2 and IRS4) contain multiple YXXM motifs and majority of these YXXM motifs in all three proteins are located within their short disordered regions[36]. The short disordered regions are called MoRFs[38] or ProSs[39, 40]. IRS1, IRS2, and IRS4 orthologies (orthology K16172) are engaged in various physiological and pathological pathways and therefore, these proteins are also involved in the pathogenesis of cancer and several endocrine and metabolic diseases[36].

SoCS

Suppressors of cytokine signaling (SoCS) have important roles in mediating inflammatory responses in both immune and non-immune systems. The SoCS family contains eight proteins: SoCS1 to SoCS7, and cytokine-inducible SH2 domain-containing protein (CIS)[36]. Because of their involvement in regulation of the insulin signaling and pancreatic beta-cell function, SoCS proteins are promising targets for the treatment of type 2 diabetes[36, 41]. Computational analysis have revealed that Kinase Inhibitory Regions (KIR) of SoCS1 (residues 55–66) and SoCS3 (residues 22–33) are highly disordered [36].

PDX1

PDX1 (pancreatic and duodenal homeobox1) is a transcription factor and the expression of PDX1 causes type2 diabetes. Computational analysis revealed that human PDX1 is highly disordered and contain several PTMs and disorder-based binding sites. Human PDX-1 represents an example of a T2DM-related intrinsically disordered hub because this protein can interact with 20 partners[36].

Fig. 1: Three-dimensional structure of human α-synuclein (PDB ID: 1XQ8).

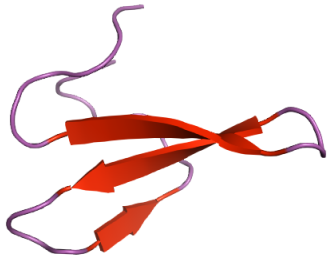

Fig. 2: Three-dimensional structure of human Tau protein (PDB ID: 2N4R).

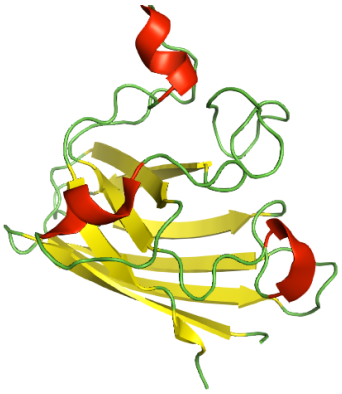

Fig. 3: Three-dimensional structure of human SOD1 protein (PDB ID : 2V0A).

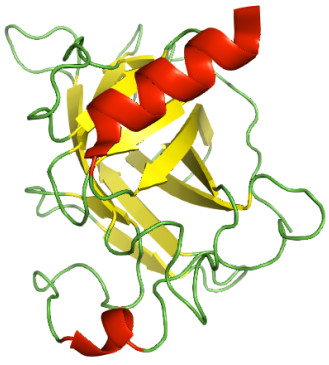

Fig. 4: Three-dimensional structure of human P53 protein (PDB ID: 1UOL).

CONCLUSION

Intrinsic disorder is common in proteins, and plays an important role in numerous protein functions. The involvement of IDPs in the pathogenesis of various diseases is determined by the unique structural and functional properties of these proteins. The abundance of intrinsic disorder in human diseases strongly suggests that disorder should be seriously evaluated and IDPs need to be considered as potential novel drug targets in the drug discovery process for the development of novel therapeutic compounds.

REFERENCES

1. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS et al. Intrinsically disordered protein. J Mol Graph Model 2001; 19:26–59.
2. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nature reviews Mol Cell Biol 2005; 6: 197–208.
3. Uversky VN. Natively unfolded proteins: A point where biology waits for physics. Protein Sci 2002; 11:739–756.
4. Wright PE, Dyson HJ. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. J Mol Biol 1999; 293:321–331.

5. Huang YQ, Liu ZR. Kinetic advantage of intrinsically disordered proteins in coupled folding binding process: a critical assessment of the "fly-casting" mechanism. J Mol Biol 2009; 393:1143–1159.
6. Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. Curr Opin Struct Biol 2002; 12:54–60.
7. Zhou P, Lugovskoy aa, McCarty JS, Li P, Wagner G. Solution structure of DFF40 and DFF45 N-terminal domain complex and mutual chaperone activity of DFF40 and DFF45. In Proc Natl Acad Sci 2001; 98:6051–6055.
8. Uversky VN. Introduction to intrinsically disordered proteins (IDPs). Chem Rev 2014; 14: 6557–6560.
9. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol 2004; 337:635–645.
10. Van Der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, et al. Classification of intrinsically disordered regions and proteins. Chem Rev 2014; 114: 6589–6631.
11. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins Struct Funct Genet 2001; 42: 38–48.
12. Vucetic S, Brown CJ, Dunker AK, Obradovic Z. Flavors of protein disorder. Proteins Struct Funct Genet 2003; 52:573–584.
13. Mészáros B, Tompa P, Simon I, Dosztányi Z. Molecular principles of the interactions of disordered proteins. J Mol Biol 2007; 372:549–561.
14. Wong E.T.C, Na D, Gsponer J. On the importance of polar interactions for complexes containing intrinsically disordered proteins. PLoS Comput Biol 2013; 9.8:e1003192.
15. Gsponer J, Madan Babu M. The rules of disorder or why disorder rules. Prog Biophys Mol Biol 2009; 99:94-103.
16. Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Biophys 2008; 37:215–246.
17. Uversky VN, Davé V, Iakoucheva LM, Malaney P, Metallo SJ, Pathak RR, et al. Pathological unfoldomics of uncontrolled chaos: Intrinsically disordered proteins and human diseases. Chem Rev 2014; 114:6844-6879.
18. Uversky VN. Intrinsically disordered proteins and novel strategies for drug discovery. Expert Opin Drug Discov 2012; 7:475–488.
19. Uversky VN. Targeting intrinsically disordered proteins in neurodegenerative and protein dysfunction diseases: another illustration of the D2 concept. Expert Rev Proteomics 2010; 7:543–64.
20. Xu L, Pu J. Alpha-Synuclein in Parkinson's Disease: From Pathogenetic Dysfunction to Potential Clinical Application. Parkinsons Dis 2016; 1:10.
21. Wang J, Cao Z, Zhao L, Li S. Novel strategies for drug discovery based on intrinsically disordered proteins (IDPs). Int J Mol Sci 2011; 12:3205-3219.
22. Guo T, Noble W, Hanger DP. Roles of tau protein in health and disease. Acta Neuropathologica 2017; 6:1-40.
23. Von Bergen M, Barghorn S, Jeganathan S, Mandelkow EM, Mandelkow E. Spectroscopic approaches to the conformation of tau protein in solution and in paired helical filaments. Neurodegener Dis 2006; 3:197–206.
24. Kirkitadze MD, Condron MM, Teplow DB. Identification and characterization of key kinetic intermediates in amyloid β-protein fibrillogenesis. J Mol Biol 2001; 312:1103–1119.
25. Zoghbi HY, Orr HT. Polyglutamine diseases: protein cleavage and aggregation. Curr Opin Neurobiol 1999; 9:566–570.
26. Uversky VN. Intrinsically disordered proteins and their (disordered) proteomes in neurodegenerative disorders. Front Aging Neurosci 2015; 7:18.
27. Kannan S, Lane DP, Verma CS. Long range recognition and selection in IDPs: the interactions of the C-terminus of p53. Scientific reports 2016; 6.
28. Joerger AC, Fersht AR. Structural biology of the tumor suppressor p53. Annu Rev Biochem 2008; 77:557–582.
29. Iakoucheva LM, Brown CJ, Lawson JD, Obradović Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 2002; 323:573-84.
30. Kumar D, Sharma N, Giri R. Therapeutic Interventions of Cancers Using Intrinsically Disordered Proteins as Drug Targets: c-Myc as Model System. Cancer Inform 2017; 16
31. Uversky VN. Wrecked regulation of intrinsically disordered proteins in diseases: pathogenicity of deregulated regulators. Front Mol Biosci 2014; 1: 6.
32. Okumura N, Yoshida H, Kitagishi Y, Nishimura Y, Matsuda S. Alternative splicings on p53,BRCA1 and PTEN genes involved in breast cancer. Biochem Biophys Res Commun 2011; 413:395–399.
33. Mark WY, Liao JC, Lu Y, Ayed A, Laister R, Szymczyna B, et al. Characterization of segments from the central region of BRCA1: an intrinsically disordered scaffold for multiple protein–protein and protein–DNA interactions? J Mol Biol 2005; 345:275-87.
34. Hollander MC, Blumenthal GM, Dennis PA, PTEN loss in the continuum of common cancers, rare syndromes and mouse models. Nat Rev Cancer 2011; 11:289-301.
35. Malaney P, Uversky VN. The PTEN Long N-tail is intrinsically disordered: increased viability for PTEN therapy. Mol Biosyst 2013; 9:2877-88.
36. Du Z, Uversky VN. A Comprehensive Survey of the Roles of Highly Disordered Proteins in Type 2 Diabetes. Int J Mol Sci 2017; 18
37. Pellegrino S, Annoni C, Contini A, Clerici F, Gelmi M.L. Expedient chemical synthesis of 75mer DNA binding domain of MafA: An insight on its binding to insulin enhancer. Amino Acids 2012; 43: 1995–2003.
38. Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK et al. Analysis of molecular recognition features (MoRFs). J Mol Biol 2006; 362:1043-59.
39. Shaji D**,** Amemiya T, Koike R, Ota M. Interface property responsible for effective interactions of protean segments: Intrinsically disordered regions that undergo disorder-to-order transitions upon binding, Biochem Biophys Res Commun 2016; 478:123-127.
40. Shaji D**.** The relationship between relative solvent accessible surface area (rASA) and irregular structures in Protean Segments (ProSs). Bioinformation 2016; 12:381-387.
41. Feng X, Tang H, Leng J, Jiang Q. Suppressors of cytokine signaling (SOCS) and type 2 diabetes. Mol Biol Rep 2014; 41:2265–2274.

Cite this article as:

Divya Shaji. Intrinsically disordered proteins (IDPs) in human diseases: A review. Int. Res. J. Pharm. 2018;9(11):1-4 http://dx.doi.org/10.7897/2230-8407.0911247

Source of support: Nil, Conflict of interest: None Declared