Cellular and Molecular Life Sciences

REVIEW

CrossMark

A structural perspective of RNA recognition by intrinsically disordered proteins

Sushmita Basu[1] · **Ranjit Prasad Bahadur**[1]

Abstract Protein-RNA recognition is essential for gene expression and its regulation, which is indispensable for the survival of the living organism at one hand, on the other hand, misregulation of this recognition may lead to their extinction. Polymorphic conformation of both the inter-acting partners is a characteristic feature of such molecular recognition that promotes the assembly. Many RNA binding proteins (RBP) or regions in them are found to be intrinsically disordered, and this property helps them to play a central role in the regulatory processes. Sequence composition and the length of the flexible linkers between RNA binding domains in RBPs are crucial in making significant contacts with its partner RNA. Polymorphic conformations of RBPs can provide thermodynamic advantage to its binding partner while acting as a chaper-one. Prolonged extensions of the disordered regions in RBPs also contribute to the stability of the large cellular machines including ribosome and viral assemblies. The involvement of these disordered regions in most of the significant cellular processes makes RBPs highly associated with various human diseases that arise due to their misregulation.

Keywords RNA binding proteins · Protein-RNA recognition · Structural flexibility · Induced folding · Polymorphic conformation

✉ Ranjit Prasad Bahadur
r.bahadur@hijli.iitkgp.ernet.in;
ranjitp_bahadur@yahoo.com

1   Computational Structural Biology Lab, Department of Biotechnology, Indian Institute of Technology Kharagpur, Kharagpur 721302, India

Introduction

Protein-RNA interactions play a significant role in gene expression and its regulation. A recent study has reported a census of 1542 non-redundant human RNA binding proteins (RBP) which interact with all known RNA types [1]. The study also reports that RBPs form 7.5 % of all the coded proteins in human proteome. The RNA binding domains (RBD) in RBPs can recognise the RNA molecule in sequence dependent as well as independent manner [2]. There are many large protein-RNA complexes including ribosome, RNA polymerase and spliceosome, which bind to the target RNA as well as catalyses various reactions. Many RBPs have multiple RBDs that can independently recognise its partner RNA [3, 4]. Detailed analysis of RBDs reveals the inherent flexibility and plasticity of the binding surfaces [5]. These domains have higher specificity and affinity towards RNA, which are achieved via linkers that are disordered extensions of RBPs [6]. Moreover, it is likely that oligomerisation and the combinatorial modular architecture of RBPs enhances their specificity and affinity to its partner RNA [7]. In most of these protein-RNA interactions, conformational multiplicity is observed in either or both the partners [8]. In such cases, the RBDs and the partner RNA undergo binding-induced folding that finally gives rise to a biologically functional protein-RNA complex. In a recent computational study, Varadi et al. [9] has reported the abundance of disordered regions in RBPs and the significance of its conformational multiplicity in RNA binding. They not only found that most of the interface residues making direct contacts with RNA are disordered in nature, but also claim that these residues are highly conserved compared to the underlying sequence. A proteomic and bioinformatics study showed that in comparison to the whole cell extract, disordered proteins are

overrepresented in the cell nucleus [10]. Moreover, from the Gene Ontology analysis it is also revealed that most of the transcription factors are enriched in disordered regions.

In the last two decades, many experimental data showed that several proteins or regions of proteins are intrinsically disordered (IDP: Intrinsically Disordered Proteins; IDR: Intrinsically Disordered Regions) in nature under native, functional condition [11, 12]. A correlation analysis between the 'functional keywords' and 'IDPs function' in Swiss-Prot [13] revealed that IDPs are mainly involved in cell cycle, gene expression, signalling events and regulations; whereas, structured proteins are found majorly correlated with keywords related to enzymatic catalysis [14]. It is the structural pliability of IDPs that makes them an important player in various interactions within the living system. An extensive study on the functional roles of IDPs have come up with five different classes: effectors, assemblers, entropic, scavengers and display sites [15]. The increased interaction potential of the IDPs is achieved by various functional features. The IDRs are involved in many low-affinity and high-specificity interactions for example, cell-signalling and regulation, which are achieved via functional elements like small linear motifs (SLiMs), molecular recognitions features (MoRFs) and low complexity regions (LCR) [16]. SLiMs are short stretches of amino acids (3–10 residues long) involved in complex formation mostly by regulating the low-affinity interactions [17]. The compact binding of them to the surface of globular proteins promotes their multiple-occurrence in IDP [17–19]. On the other hand, MoRFs (10–70 residues long) are generally longer than SLiMs and are involved in specific protein–protein interactions [20, 21]. There are regions in IDPs called LCRs with repetitive amino acid residues along the sequence [22]. These motifs undergo disorder to order transition after binding to its partner, and in the unbound form, they are mostly biased to the conformation they adopt upon binding [23]. Considering the multi-functionality, the mode of interactions involving IDPs is always an interesting topic of research in protein science. A comparative study between globular proteins and IDPs revealed that the latter have unique molecular principles of interactions [24]. It was found that compared to the globular proteins, IDPs have larger surface area as well as interface area per residue. The exposed segments of IDPs forming the interface with its partner molecules are enriched in hydrophobic residues. Moreover, it was also observed that hydrophobic–hydrophobic interactions are dominant over polar–polar interactions in molecular recognition involving IDPs. Hydrophobic residues of IDPs have a typical tendency to be present at the exposed interacting segments and avoid getting collapsed in forming a regular folded structure. Based on these observations, it was suggested that IDPs follow 'inside-out' folding

mechanism where the partner binds to the interacting segments of the IDP mainly involving hydrophobic contacts, which further promotes folding [24].

Structural disorder can be studied by many experimental techniques such as X-ray crystallography, NMR spectroscopy, circular dichroism (CD) or small-angle X-ray scattering. While NMR can give reliable consensus, X-ray fails to give information about disordered regions, raising confusion whether the failure to resolve is because of the technical issues (missing electron density can arise from failure to solve the phase problem, crystal defects or even from unintentional proteolytic removal during protein purification) or because of the inherent property of IDPs. On the other hand, CD with a combination of far-UV and near-UV gives an idea whether a protein has disordered regions [12, 25, 26]. Intriguingly, thousands of the structures in PDB are now known to contain disordered regions [27].

Flexibility of binding partners is an indispensable feature in protein-RNA recognition. The involvement of IDRs in such recognition process facilitates RBPs to attain alternate conformations upon binding with the RNA. In this review, we are focusing on the structural basis of the polymorphic conformations involved in RBP-RNA interactions. This polymorphic conformations lead to the structural plasticity of the interface, thereby influencing the recognition, which eventually regulates different cellular processes.

RNA recognition involving IDPs

Protein-RNA recognition is essential for post-transcriptional gene regulation [28]. The recognition can be both sequence and structure dependent [29]. Besides, recognition may involve induced-folding of protein, RNA or both [8, 30]. The diversified mode of RNA binding is a major characteristic of the recognition, and the low energy requirement of the RNA to deform and unfold facilitates the process [31]. The major benefit of a protein domain having structural flexibility is its potential to mould into a binding surface according to its partner molecule, and likewise can facilitate to bind multiple partners [30]. This has been observed in a complex between TruB and its partner RNA. Here, the thumb loop in TruB is disordered, which undergoes disorder to order transition and gains helical conformation upon binding with the RNA (Fig. 1) [32]. RBPs have characteristic sequence features, which particularly bind to single stranded or double stranded RNA. Some of the very common RNA binding domains are zinc finger [33], RNA recognition motif (RRM) [34] and K homology domain [35]. RGG/RG motif is a SLiM, which is also found abundantly in RBPs [36]. This RGG motif is found to have high binding affinity towards G-rich RNA sequences [37]. The RG-rich regions take part in RNA metabolic processes via both selective and non-selective
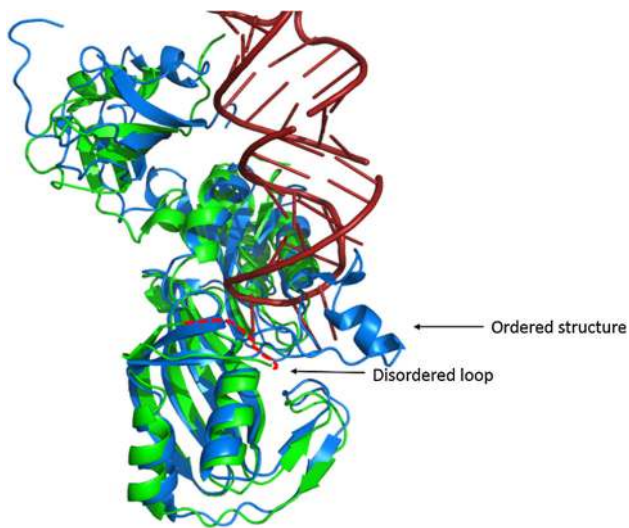
Fig. 1 Superposed structure of TruB with its partner RNA in bound (*coloured in blue cartoon*, PDB ID: 1R3E) and unbound (*coloured in green cartoon*, PDB ID: 1R3F) conformations [32]. The disordered thumb loop (*red dashed lines*) of TruB undergoes conformational transitions and become ordered upon binding with its partner RNA

binding [7]. A very recent computational analysis showed that the nucleiome of archae, bacteria and eukaryota have abundance of disordered regions [38]. Disordered regions promote the formation of ribo-nucleoproteins, which can further act as assembly domains [39]. The flexibility owing to the polymorphic conformations in RBD is a highly facilitating factor for RNA recognition promoting IDP-RNA interactions in various cellular processes [40].

RNA chaperones

One of the major concerns in RNA folding is the kinetically trapping of alternate RNA conformers. This can be resolved by the RNA chaperones, which make non-specific interactions with the RNA thereby assisting the RNA folding [8]. In many cases, it has been found that the protein domains crucial for the function of RNA chaperone are disordered in nature [15, 31, 41]. These disordered regions render interaction multiplicity along with higher momentum of interaction with the partner molecule. A comparative study between protein and RNA chaperones revealed that the occurrence of disordered regions in RNA chaperones are much higher than that in protein chaperones [41]. This study also suggested a probable mechanism of entropy transfer, which assists the misfolded RNA to come out of the local energy minima trough and rearrange to attain the favourable conformation. The flexible domains of the RNA chaperones interact with the RNA molecule and undergo disorder to order transition, which in turn provides a thermodynamic advantage to the trapped RNA molecule to get a minimum

energy conformation. The folding of RNA is enthalpy driven and followed by the thermodynamic equation:

$$\Delta G = \Delta H - T\Delta S$$

The chaperone molecule binds to the RNA and transfers its entropy to the RNA molecule making the overall process entropy driven. This brings an overall change in ΔG, which is favourable for spontaneous folding of RNA. A prominent example of such mutual induced-fit phenomenon is observed in the binding of ribosomal protein L5 of *Xenopus* oocyte with 5S rRNA [42]. Here, the flexible N and C terminal regions of the protein, enriched with non-polar amino acids, increase the interacting surface area by making multiple contacts. Ribosomal protein S12 is also found to assist the folding of phage T4 group I intron, and is assumed to stabilise the correct conformation of the RNA in due of its flexible binding region [43]. The disordered N terminal of prion protein, huPrP, also shows chaperoning function in nucleic acid annealing, viral RNA dimerisation and binding to complementary tRNA [44].

Ribosomal assembly

The interactive features of IDPs like MoRF, SLiM and LCR provide them multiple binding competence, which promotes them to involve in macromolecular assemblies [45]. Polymorphic conformation translated into structural flexibility provides an advantage in complex formation by reducing the steric hindrances [16]. One such higher-order complex is the ribosome assembly, which involves many RBPs having IDRs. Genome wide study has revealed the widespread prevalence of unstructured regions in ribosomal proteins [46]. The flexible extended regions of ribosomal proteins penetrate into the core of the ribosomal subunits, where it is assumed that they undergo disorder to order transition and facilitates the ribosomal RNA (rRNA) folding [47]. This is exemplified in Fig. 2a, which shows the small subunit protein S6 and S11 have defined secondary structures in the periphery of the ribosomal complex but interact with the core region mainly through long disordered extensions [48]. The binding of disordered regions of ribosomal proteins with its partner RNA in the ribosomal assembly is facilitated by the electrostatic interaction between the positively charged regions of the proteins and the negatively charged RNA backbone [46]. High glycine content, one of the characteristic features of IDPs, permits the extension of ribosomal proteins to be flexible and interact with the rRNA. This is evident in almost all the small subunit proteins, and 50 % of the large subunit proteins [49]. Figure 2b shows that the small subunit ribosomal protein, S12, interacts with mRNA with its flexible linkers. Structural analysis of the globular domains
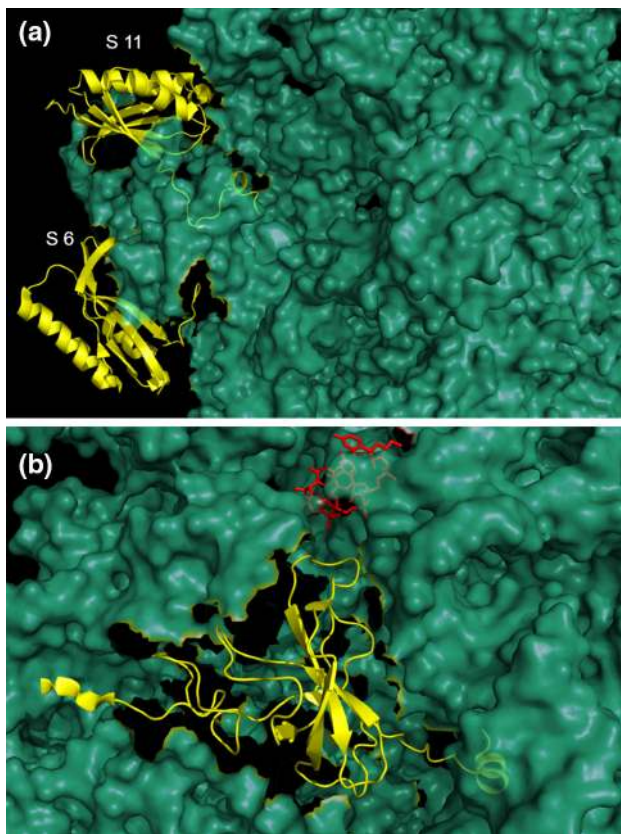
Fig. 2 Ribosomal proteins of 30S subunit (PDB ID: 1N34) [48]. **a** Small subunit proteins S6 and S11 are shown in *yellow cartoon* and other small subunit proteins are shown in *green surface*. **b** Small subunit protein S12 (shown in *yellow cartoon*) is interacting with mRNA (shown as a *red fragment*) through its disordered extension, the other small subunit proteins are shown in *green surface*.

of the assembly indicates the presence of unfamiliar morphologies of the ordered domains that may be attributed to conformational polymorphism. Further, Nussinov's plot analysis probed that many globular domains are formed as a result of binding-induced folding mechanism [46, 50]. Eukaryotic core ribosomal proteins, L4, L22, L23, L29, those make the polypeptide exit tunnel has long extended regions which stretches up to the border of the large subunit. Intruding flexible extensions of L24e crosses the 60S subunit and penetrates into the 40S subunit [51]. Some of the very long extensions are also seen in L2, L3 and S12, which can reach up to the peptidyl transferase domain and participate in some vital interactions with the ribozyme [52]. Apart from taking part in the major events of translation, some of the ribosomal proteins (S1, L1 and L4) are also associated with extra ribosomal functions like maintaining stability among various ribosomal components and preventing apoptosis by reducing the nuclear stress [53].

Multi domain linkers

Many RBPs utilise multiple domains to recognise its partner RNA. This multi-domain binding involves cooperative interactions between the RBDs and dictates the dynamics of conformational polymorphism in RNA binding [54]. These interacting domains of RBPs are often linked by a flexible stretch of polypeptides called linkers or spacers. The NMR relaxation data and missing residues or high B-factor in X-ray crystallographic data of the linkers indicate their intrinsic disorderedness [6]. These linkers play a crucial role in determining the sequence specificity of RBPs. Such interplay has been observed in the complex between the zinc finger domains of TIS11d and the class II AU-rich element (ARE) [55] (Fig. 3). Here, the two finger motifs are spaced with an 18 residue linker region. The highly conserved residues of the linkers, residing between the two finger motifs, make significant contacts with the U-rich motif of the RNA and stabilises the TIS11d-ARE complex. Recent studies by Barik et al. [56] showed that salt-bridges play an important role in protein-RNA recognition by contributing to the binding affinity. Moreover, they also showed that stacking interactions involving protein side chains and nucleotide bases also contribute to the recognition process. The aromatic residues phenylalanine and tyrosine actively interact with the uracil to make stacking interactions [56, 57]. This phenomenon is observed in TIS11d-ARE complex. Here, the C-terminal linker in TIS11d interacts closely with ARE RNA and is stabilised by a slat bridge between the OE1 of Glu195 and 2′ OH of U1 base (Fig. 3). Besides, stacking interaction is also observed between the side chain of Phe214 and the U2 base. In some cases, the length of the linker is expected to be conserved better than the sequence as the length regulates the conformational dynamics of the interactions [6]. Leepar et al. [3] showed that the RRM domains of Hrp1 and Rna15 protein of yeast form a ternary complex with the pre-mRNA segment during mRNA processing. The solution structure of this complex reveals the role of the linker that forms a helical conformation between two RRM domains of Hrp1. This structured linker, arises due to the conformational polymorphism, enhances contacts with the RNA and stabilises the ternary complex. Apart from stabilising the protein-RNA interaction, the linker of dsRBD of ADAR2 is assumed to facilitate its interaction with the stem-loop pre-mRNA by exhibiting fly-casting mechanism as shown in Fig. 4 [4]. Such mechanism contributes to the increased binding affinity of adjacent domains, which, otherwise has weak interaction with the RNA [58].

Fig. 3 The RBD of TIS11d bound to AU-rich motif of RNA (PDB ID: 1RGO) [55]. The linker between the RBDs of TIS11d is closely interacting with the RNA, which is stabilised by stacking interaction between F214 (*yellow sticks*) and U2, and a salt bridge (shown in *yellow dashed lines*) between E195 and U1. The protein is shown in *green cartoon* and the RNA is shown in *red sticks*
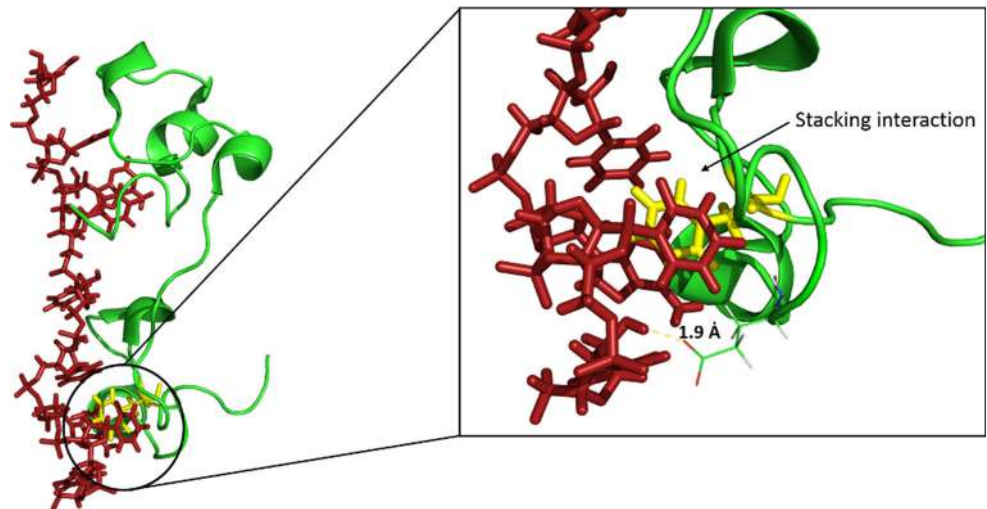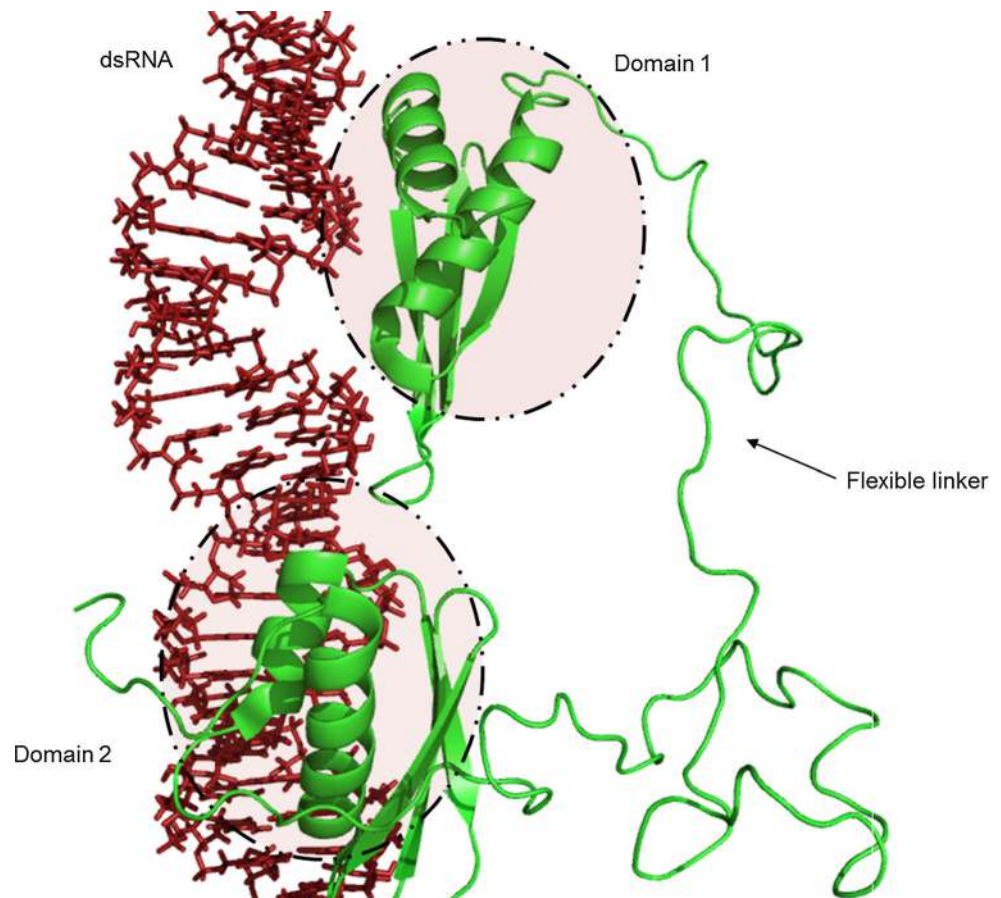
Fig. 4 ADAR2 bound to dsRNA showing multiple-domain interaction regulated by a linker in between (PDB ID: 2L3J) [4]. The protein is shown in *green cartoon* and the RNA is shown in *red stick*

Spliceosome complex

Among the eukaryotic gene regulation machineries, spliceosome is one major complex that involves extensive protein-RNA interactions. This ribonucleoprotein complex works on precursor mRNA to splice the introns and create the functional open reading frame [59]. Structural investigation on splicing factors has revealed the presence of regions showing polymorphic conformations, which participate in various protein–protein and protein-RNA interactions [60]. A study on human proteome also predicted abundant disorderedness in human spliceosomal

proteins [61]. The splicing factors, SR proteins, are one of the most important components of metazoan gene expressions. They have one or two RRM domain at the N terminal and an arginine/serine rich RS domain at the C terminal [62]. Sequence analysis study involving charge hydropathy classification, cumulative distribution function analysis and disorder prediction on SR proteins showed that their amino acid sequence exhibits similar properties to that of IDPs [63]. While the RRM domain is mainly responsible for recognising the specific RNA [64], RS domain interacts with other splicing factors [65]. Furthermore, protein-RNA crosslinking studies has reported that the RS domain bound to splicing enhancers make significant contacts with the pre-mRNA branch point and promote pre-spliceosome assembly formation [66]. This RS domain is vital for the functioning of the splicing factor, and was found completely disordered according to the sequence based studies [63]. This disorderedness is also confirmed by the biophysical study where the CD spectrum shows random coil characteristics of this domain [67]. SR proteins play a very important role in the large spliceosome complex by regulating the splicing phenomenon. Due to the limited structural information of splieceosome complex [68], the molecular and structural mechanism of RNA binding by the flexible RS domain is still elusive; however, the extensive presence of conformational diversity implies their significant role in binding [63].

Viral RNA

Viruses in their host cells are one of the most dynamic living organisms on earth. They constantly adapt itself to the hosts' environment, come up with the mechanism to evade the hosts' immune system, propagate and evolve. They are also found to harbour many proteins having highly polymorphic conformations, which performs multiple functions including binding viral RNA. The hepatitis C virus (HCV) codes for a core protein (HCV-C), which has two N terminal binding domains and a signal peptide [69]. Domain 1 has a 150 residue long hydrophilic region that is essential for the assembly of nucleocapsid-like particles (NLPs) and RNA binding; while domain 2 plays a role in HCV-C interaction with lipid droplets [70]. Computational analysis as well as biophysical experiments such as far UV CD spectra and NMR spectroscopy revealed the random-coil conformation of the N-terminal of HCV-C [71]. The chaperoning activity was found to be active even after heat denaturation, which further supports the disorderedness of the HCV-C [72]. Thereby, it was assumed that the flexibility in the terminal region is the facilitating factor in binding assisted folding of the viral RNA in the process of viral genome packaging. In the nucleocapsid protein (NC) of

severe acute respiratory syndrome coronavirus (SARS-CoV), stretches of flexible linkers are found in between the two structured domains, NTD (N-terminal domain) and CTD (C-terminal dimerisation domain) [73]. The NC protein binds to the RNA in cooperative manner. Although the structural mechanism of RNA binding by NC protein is still unknown, it has been assumed that the flexible linker between the domains makes significant contacts on multiple RNA sites [73]. The human immunodeficiency virus type-I (HIV-I) encodes a trans-activator of viral transcription named Tat protein [74]. It interacts with kinases and transcription factors thereby facilitating the viral RNA transcription elongation process. The assembly of the elongation complex B is regulated by the interaction of Tat with transactivation response region TAR located at the 5′ end of the viral transcript. Amino acid sequence analysis showed that Tat has high net positive charge and lacks hydrophobic residues, indicating its disordered nature. This has also been established by CD and NMR studies [75]. Disorder to order transition that can be attributed to the conformational polymorphism is observed in anti-termination N proteins of bacteriophages (like P22 and λ) when it interacts with the boxB RNA motif of the viral mRNA [76]. NMR experiments revealed that the N protein is completely disordered and upon binding to boxB RNA, only the N terminal region (N peptide) gets structured and gains stability [77, 78]. This phenomenon is also supported by the molecular dynamic simulation study on P22 N protein and boxB RNA. Here, Bahadur et al. [79] showed that the electrostatic field of the RNA has a favourable influence on the coil-to-a-helix transition of the N peptide. This is evident in Fig. 5, where the N peptide of P22 N protein is disordered in the unbound form but attains a helical conformation upon binding with the boxB RNA.
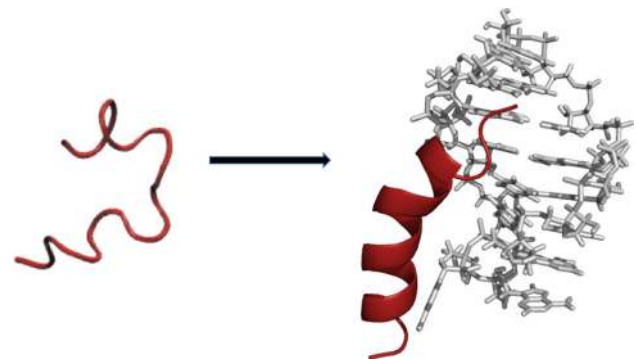

Fig. 5 The disordered *N*-peptide is transformed into helical conformation upon binding with the boxB RNA (PDB ID: 1A4T) [79]. The peptide is shown in *red* and the RNA is shown in *grey*

Human diseases associated with IDP-RNA interactions

Protein-RNA interaction, being one of the most abundant cellular phenomenon, is associated with a number of human diseases [80]. IDPs are also annotated with diseases as they are part of various cell signalling and regulatory pathways [81]. Recent analysis shows that a number of RBPs encoded by *FET* (*FUS/TLS*, *EWS* and *TARF15*) genes have low complexity regions, and are found to play significant role in DNA damage response [82]. Several point mutations were found within the *FET*, which are associated with protein aggregation resulting in neurodegenerative diseases including amyotrophic lateral sclerosis (ALS) and frontotemporal lobe degeneration (FTLD). A computational study on disease causing mutations revealed that a considerable section of such mutations are associated with the disordered regions [83]. These mutations were found to affect the post-translational modifications, macromolecular assembly and other regulatory processes. The study also shows that the mutations are majorly responsible for the disorder to order transition in the IDRs, hence curbing the structural flexibility and limiting its binding ability. This is evident from the frequent mutations found in the Arg-rich RGG/RG motifs in fused sarcoma protein (FUS), a RBP associated with neurodegenerative disorder ALS [84]. Another autosomal recessive disease, spinal muscular atrophy (SMA) is caused by mutations in the SMN1 (survival of motor neuron) protein [85]. The central region of SMN1 forms the Tudor domain, which shows conformational polymorphism in solutions. Mutations found in the Tudor domain of the protein interfere with its binding to the RG rich domains of various snRNPs, which finally causes neuronal apoptosis [86, 87]. FMRP (fragile X mental retardation protein) is a RBP that exhibits chaperoning activity and has disordered regions [88, 89]. Suppression of this protein is the major cause for the fragile X syndrome, a X-linked disorder. The FMRP is found to recognise U-rich RNA sequences, and interact with various mRNAs along with non-coding miRNA and siRNA [90–93]. The nucleocapsid protein of HIV-1 (NCp7) is another disordered protein having RNA chaperoning function [94]. NCp7 interacts with the viral RNA and coats it by forming oligomers [95]. Based on the degree of RNA occupancy, NCp7 perform activities from assembly of virus particle to genome packaging. The polymorphic conformations of NCp7 allow it to perform such array of activities, which makes it a vital factor for proper viral replication and propagation of the disease [95].

Conclusion

Growing number of atomic structures in PDB facilitate the study of RBP-RNA interactions. Diversity of structural rearrangement of RNA induces the conformational transition to its partner RBP. Mutual folding of RNA and protein in ribonucleoprotein complexes is a common phenomenon and is probably ubiquitous. A recent analysis suggested that the IDPs are proteins waiting for their partner to bind and acquire a particular conformation [96]. Presently, a significant number of atomic structures of RBPs in bound and unbound conformations is available in the PDB [97]. These structural information shows that polymorphic conformation is intrinsic to RBP. In this study, we propose that polymorphic conformation of RBPs promote structural flexibility, which considerably influences the conformational dynamics of RBP-RNA interaction. Indeed, RNA recognition by RBPs is more than a simple handshake and is associated with a variety of conformational flexibility in both the partners. Hence, the lack of structural integrity, camouflaged in conformational polymorphism, is highly favoured in this kind of molecular recognition. Although, the PDB covers only about 15 % of the known protein-RNA complexes of human RBP, many more structural information is required in near future to make a complete repertoire of RBP-RNA interactions, which eventually can lead us to a better understanding of human diseases associated with this recognition process.

References

1. Gerstberger S, Hafner M, Tuschl T (2014) A census of human RNA-binding proteins. Nat Rev Genet 15:829–845
2. Ban T, Zhu JK, Melcher K, Xu HE (2015) Structural mechanisms of RNA recognition: sequence-specific and non-specific RNA-binding proteins and the Cas9-RNA-DNA complex. Cell Mol Life Sci 72:1045–1058
3. Leeper TC, Qu X, Lu C, Moore C, Varani G (2010) Novel protein–protein contacts facilitate mRNA 3′-processing signal recognition by Rna15 and Hrp1. J Mol Biol 401:334–349
4. Stefl R, Oberstrass FC, Hood JL, Jourdan M, Zimmermann M, Skrisovska L, Maris C, Peng L, Hofr C, Emeson RB, Allain FH (2010) The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. Cell 143:225–237
5. Thickman KR, Sickmier EA, Kielkopf CL (2007) Alternative conformations at the RNA-binding surface of the N-terminal U2AF65 RNA recognition motif. J Mol Biol 366:703–710
6. Mackereth CD, Sattler M (2012) Dynamics in multi-domain protein recognition of RNA. Curr Opin Struct Biol 22:287–296
7. Järvelin AI, Noerenberg M, Davis I, Castello A (2016) The new (dis)order in RNA regulation. Cell Commun Signal 14:1–22

8. Williamson JR (2000) Induced fit in RNA-protein recognition. Nat Struct Biol 7:834–837

9. Varadi M, Zsolyomi F, Guharoy M, Tompa P (2015) Functional advantages of conserved intrinsic disorder in RNA-binding proteins. PLoS One 10:e0139731

10. Skupien-Rabian B, Jankowska U, Swiderska B, Lukasiewicz S, Ryszawy D, Dziedzicka-Wasylewska M, Kedracka-Krok S (2016) Proteomic and bioinformatic analysis of a nuclear intrinsically disordered proteome. J Proteomics 130:76–84

11. Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 293:321–331

12. Dunker AK (2001) Intrinsically disordered protein. J Mol Graph Model 19:26–59

13. Bairoch A, Apweiler R (1996) The SWISS-PROT Protein sequence data bank and its new supplement TREMBL. Nucleic Acids Res 24:21–25

14. Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN (2007) Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. J Proteome Res 6:1917–1932

15. Tompa P (2002) Intrinsically unstructured proteins. Trends Biochem Sci 27:527–533

16. Van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, Fuxreiter M, Gough J, Gsponer J, Jones DT, Kim PM, Kriwacki RW, Oldfield CJ, Pappu RV, Tompa P, Uversky VN, Wright PE, Babu MM (2014) Classification of intrinsically disordered regions and proteins. Chem Rev 114:6589–6631

17. Davey NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, Budd A, Diella F, Dinkel H, Gibson TJ (2012) Attributes of short linear motifs. Mol BioSyst 8:268–281

18. Diella F, Haslam N, Chica C, Budd A, Michael S, Brown NP, Trave G, Gibson TJ (2008) Understanding eukaryotic linear motifs and their role in cell signaling and regulation. Front Biosci 13:6580–6603

19. Ren S, Uversky VN, Chen Z, Dunker AK, Obradovic Z (2008) Short linear motifs recognized by SH2, SH3 and Ser/Thr kinase domains are conserved in disordered protein regions. BMC Genom 9(Suppl 2):S26

20. Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, Uversky VN (2006) Analysis of molecular recognition features (MoRFs). J Mol Biol 362:1043–1059

21. Vacic V, Oldfield CJ, Mohan A, Radivojac P, Cortese MS, Uversky VN, Dunker AK (2007) Characterization of molecular recognition features, MoRFs, and their binding partners. J Proteome Res 6:2351–2366

22. Romero P (2001) Sequence complexity of disordered protein. Proteins 42:38–48

23. Fuxreiter M, Simon I, Friedrich P, Tompa P (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J Mol Biol 338:1015–1026

24. Mészáros B, Tompa P, Simon I, Dosztányi Z (2007) Molecular principles of the interactions of disordered proteins. J Mol Biol 372:549–561

25. Dyson HJ, Wright PE (2001) Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states. Methods Enzymol 339:258–270

26. Syme CD, Blanch EW, Holt C, Jakes R, Goedert M, Hecht L, Barron LD (2002) A Raman optical activity study of rheomorphism in caseins, synucleins and tau. New insight into the structure and behaviour of natively unfolded proteins. FEBS J 269:148–156

27. Lobanov MY, Furletova EI, Bogatyreva NS, Roytberg MA, Galzitskaya OV (2010) Library of disordered patterns in 3D protein structures. PLoS Comput Biol 6:e1000958

28. Lee JT (2012) Epigenetic regulation by long noncoding RNAs. Science 338:1435–1439

29. Stefl R, Skrisovska L, Allain FH (2005) RNA sequence-and shape-dependent recognition by proteins in the ribonucleoprotein particle. EMBO Rep 6:33–38

30. Frankel AD, Smith CA (1998) Induced folding in RNA-protein recognition: more than a simple molecular handshake. Cell 92:149–151

31. Dyson HJ (2012) Roles of intrinsic disorder in protein-nucleic acid interactions. Mol BioSyst 8:97–104

32. Pan H, Agarwalla S, Moustakas DT, Finer-Moore J, Stroud RM (2003) Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit. Proc Natl Acad Sci USA 100:12648–12653

33. Font J, Mackay JP (2010) Beyond DNA: zinc finger domains as RNA-binding modules. Methods Mol Biol 649:479–491

34. Clery A, Blatter M, Allain FH (2008) RNA recognition motifs: boring? Not quite. Curr Opin Struct Biol 18:290–298

35. Valverde R, Edwards L, Regan L (2008) Structure and function of KH domains. FEBS J 275:2712–2726

36. Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, Krijgsveld J, Hentze MW (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell 149:1393–1406

37. Phan AT, Kuryavyi V, Darnell JC, Serganov A, Majumdar A, Ilin S, Raslin T, Polonskaia A, Chen C, Clain D, Darnell RB, Patel DJ (2011) Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. Nature Struct Mol Biol 18:796–804

38. Wang C, Uversky VN, Kurgan L (2016) Disordered nucleiome: abundance of intrinsic disorder in the DNA and RNA binding proteins in 1121 species from eukaryota, bacteria and archaea. Proteomics 16:1486–1498

39. Calabretta S, Richard S (2015) Emerging roles of disordered sequences in RNA-binding proteins. Trends Biochem Sci 40:662–672

40. Dan A, Ofran Y, Kliger Y (2010) Large-scale analysis of secondary structure changes in proteins suggests a role for disorder-to-order transitions in nucleotide binding proteins. Proteins 78:236–248

41. Tompa P, Csermely P (2004) The role of structural disorder in the function of RNA and protein chaperones. FASEB J 18:1169–1175

42. DiNitto JP, Huber PW (2003) Mutual induced fit binding of Xenopus ribosomal protein L5 to 5S rRNA. J Mol Biol 330:979–992

43. Coetzee T, Herschlag D, Belfort M (1994) Escherichia coli proteins, including ribosomal protein S12, facilitate in vitro splicing of phage T4 introns by acting as RNA chaperones. Genes Dev 8:1575–1588

44. Gabus C, Derrington E, Leblanc P, Chnaiderman J, Dormont D, Swietnicki W, Morillas M, Surewicz WK, Marc D, Nandi P, Darlix JL (2001) The prion protein has RNA binding and chaperoning properties characteristic of nucleocapsid protein NCP7 of HIV-1. J Biol Chem 276:19301–19309

45. Cumberworth A, Lamour G, Babu MM, Gsponer J (2013) Promiscuity as a functional trait: intrinsically disordered regions as central players of interactomes. Biochem J 454:361–369

46. Peng Z, Oldfield CJ, Xue B, Mizianty MJ, Dunker AK, Kurgan L, Uversky VN (2014) A creature with a hundred waggly tails: intrinsically disordered proteins in the ribosome. Cell Mol Life Sci 71:1477–1504

47. Timsit Y, Acosta Z, Allemand F, Chiaruttini C, Springer M (2009) The role of disordered ribosomal protein extensions in the early steps of eubacterial 50S ribosomal subunit assembly. Int J Mol Sci 10:817–834

48. Ogle JM, Murphy FV, Tarry MJ, Ramakrishnan V (2002) Selection of tRNA by the ribosome requires a transition from an open to a closed form. Cell 111:721–732

49. Wilson DN, Nierhaus KH (2005) Ribosomal proteins in the spotlight. Crit Rev Biochem Mol Biol 40:243–267

50. Gunasekaran K, Tsai CJ, Nussinov R (2004) Analysis of ordered and disordered protein complexes reveals structural features discriminating between stable and unstable monomers. J Mol Biol 341:1327–1341

51. Ben-Shem A, Garreau de Loubresse N, Melnikov S, Jenner L, Yusupova G, Yusupov M (2011) The structure of the eukaryotic ribosome at 3.0 Å resolution. Science 334:1524–1529

52. Uversky VN (2011) Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. Chem Soc Rev 40:1623–1634

53. Singh D, Chang SJ, Lin PH, Averina OV, Kaberdin VR, Lin-Chao S (2009) Regulation of ribonuclease E activity by the L4 ribosomal protein of Escherichia coli. Proc Natl Acad Sci USA 106:864–869

54. Hennig J, Sattler M (2015) Deciphering the protein-RNA recognition code: combining large-scale quantitative methods with structural biology. BioEssays 37:899–908

55. Hudson BP, Martinez-Yamout MA, Dyson HJ, Wright PE (2004) Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. Nat Struct Mol Biol 11:257–264

56. Barik A, Pilla SP, Bahadur RP (2015) Molecular architecture of protein-RNA recognition sites. J Biomol Struct Dyn 33:2738–2751

57. Nahalka J (2014) Protein–RNA recognition: cracking the code. J Theor Biol 343:9–15

58. Shoemaker BA, Portman JJ, Wolynes PG (2000) Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc Natl Acad Sci USA 97:8868–8873

59. Chen W, Moore MJ (2014) The spliceosome: disorder and dynamics defined. Curr Opin Struct Biol 24:141–149

60. Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK (2007) Intrinsic disorder and functional proteomics. Biophys J 92:1439–1456

61. Korneta I, Bujnicki JM (2012) Intrinsic disorder in the human spliceosomal proteome. PLoS Comput Biol 8:e1002641

62. Shepard PJ, Hertel KJ (2009) The SR protein family. Genome Biol 10:242

63. Haynes C, Iakoucheva LM (2006) Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. Nucleic Acids Res 34:305–312

64. Perez-Canadillas JM, Varani G (2001) Recent advances in RNA-protein recognition. Curr Opin Struct Biol 11:53–58

65. Wu JY, Maniatis T (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell 75:1061–1070

66. Shen H, Kan JL, Green MR (2004) Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. Mol Cell 13:367–376

67. Labourier E, Rossi F, Gallouzi IE, Allemand E, Divita G, Tazi J (1998) Interaction between the N-terminal domain of human DNA topoisomerase I and the arginine-serine domain of its substrate determines phosphorylation of SF2/ASF splicing factor. Nucleic Acids Res 26:2955–2962

68. Jurica MS, Moore MJ (2003) Pre-mRNA splicing: awash in a sea of proteins. Mol Cell 12:5–14

69. Boulant S, Vanbelle C, Ebel C, Penin F, Lavergne JP (2005) Hepatitis C virus core protein is a dimeric alpha-helical protein exhibiting membrane protein features. J Virol 79:11353–11365

70. Barba G, Harper F, Harada T, Kohara M, Goulinet S, Matsuura Y, Eder G, Schaff Z, Chapman M, Miyamura T, Bréchot C (1997) Hepatitis C virus core protein shows a cytoplasmic localization and associates to cellular lipid storage droplets. Proc Natl Acad Sci USA 94:1200–1205

71. Duvignaud JB, Savard C, Fromentin R, Majeau N, Leclerc D, Gagne SM (2009) Structure and dynamics of the N-terminal half of hepatitis C virus core protein: an intrinsically unstructured protein. Biochem Biophys Res Commun 378:27–31

72. Ivanyi-Nagy R, Lavergne J-P, Gabus C, Ficheux D, Darlix J-L (2008) RNA chaperoning and intrinsic disorder in the core proteins of Flaviviridae. Nucleic Acids Res 36:712–725

73. Chang CK, Hsu YL, Chang YH, Chao FA, Wu MC, Huang YS, Hu CK, Huang TH (2009) Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications for ribonucleocapsid protein packaging. J Virol 83:2255–2264

74. Campbell GR, Loret EP (2009) What does the structure-function relationship of the HIV-1 Tat protein teach us about developing an AIDS vaccine? Retrovirology 6:50

75. Vendel AC, Lumb KJ (2004) NMR mapping of the HIV-1 Tat interaction surface of the KIX domain of the human coactivator CBP. Biochemistry 43:904–908

76. Cai Z, Gorin A, Frederick R, Ye X, Hu W, Majumdar A, Kettani A, Patel DJ (1998) Solution structure of P22 transcriptional antitermination N peptide-boxB RNA complex. Nat Struct Biol 5(3):203–212

77. Mogridge J, Legault P, Li J, Van Oene MD, Kay LE, Greenblatt J (1998) Independent ligand-induced folding of the RNA-binding domain and two functionally distinct antitermination regions in the phage lambda N protein. Mol Cell 1:265–275

78. Van Gilst MR, Rees WA, Das A, von Hippel PH (1997) Complexes of N antitermination protein of phage λ with specific and nonspecific RNA target sites on the nascent transcript. Biochemistry 36:1514–1524

79. Bahadur RP, Kannan S, Zacharias M (2009) Binding of the bacteriophage P22 N-peptide to the boxB RNA motif studied by molecular dynamics simulations. Biophys J 97:3139–3149

80. Cooper TA, Wan L, Dreyfuss G (2009) RNA and disease. Cell 136:777–793

81. Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Bio 37:215–246

82. Kai M (2016) Roles of RNA-binding proteins in DNA damage response. Int J Mol Sci 17

83. Vacic V, Markwick PR, Oldfield CJ, Zhao X, Haynes C, Uversky VN, Iakoucheva LM (2012) Disease-associated mutations disrupt functionally important regions of intrinsic protein disorder. PLoS Comput Biol 8:e1002709

84. Vance C, Rogelj B, Hortobagyi T, De Vos KJ, Nishimura AL, Sreedharan J, Hu X, Smith B, Ruddy D, Wright P, Ganesalingam J, Williams KL, Tripathi V, Al-Saraj S, Al-Chalabi A, Leigh PN, Blair IP, Nicholson G, de Belleroche J, Gallo JM, Miller CC, Shaw CE (2009) Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. Science 323:1208–1211

85. Sprangers R, Groves MR, Sinning I, Sattler M (2003) High-resolution X-ray and NMR structures of the SMN tudor domain: conformational variation in the binding site for symmetrically dimethylated arginine residues. J Mol Biol 327:507–520

86. Fischer U, Liu Q, Dreyfuss G (1997) The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis. Cell 90:1023–1029

87. Selenko P, Sprangers R, Stier G, Buhler D, Fischer U, Sattler M (2001) SMN tudor domain structure and its interaction with the Sm proteins. Nat Struct Biol 8:27–31

88. Gabus C, Mazroui R, Tremblay S, Khandjian EW, Darlix J-L (2004) The fragile X mental retardation protein has nucleic acid chaperone properties. Nucleic Acids Res 32:2129–2137

89. Adinolfi S, Bagni C, Musco G, Gibson T, Mazzarella L, Pastore A (1999) Dissecting FMR1, the protein responsible for fragile X syndrome, in its structural and functional domains. RNA 5:1248–1258

90. Sung YJ, Conti J, Currie JR, Brown WT, Denman RB (2000) RNAs that interact with the fragile X syndrome RNA binding protein FMRP. Biochem Biophys Res Commun 275:973–980

91. Chen L, Yun SW, Seto J, Liu W, Toth M (2003) The fragile X mental retardation protein binds and regulates a novel class of mRNAs containing U rich target sequences. Neuroscience 120:1005–1017

92. Brown V, Jin P, Ceman S, Darnell JC, O'Donnell WT, Tenenbaum SA, Jin X, Feng Y, Wilkinson KD, Keene JD, Darnell RB, Warren ST (2001) Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. Cell 107:477–487

93. Jin P, Alisch RS, Warren ST (2004) RNA and microRNAs in fragile X mental retardation. Nature Cell Biol 6:1048–1053

94. Summers MF, Henderson LE, Chance MR, Bess JW Jr, South TL, Blake PR, Sagi I, Perez-Alvarado G, Sowder RC 3rd, Hare DR, Arthur LO (1992) Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of intact viruses and the solution-state structure of the nucleocapsid protein from HIV-1. Protein Sci 1:563–574

95. Ivanyi-Nagy R, Davidovic L, Khandjian EW, Darlix JL (2005) Disordered RNA chaperone proteins: from functions to disease. Cell Mol Life Sci 62:1409–1417

96. Janin J, Sternberg MJ (2013) Protein flexibility, not disorder, is intrinsic to molecular recognition. F1000 Biol Rep 5:2

97. Barik A, C N, P M, Bahadur RP (2012) A protein-RNA docking benchmark (I): nonredundant cases. Proteins 80:1866–1871