

ADVANCED REVIEW

Deep structural insights into RNA-binding disordered protein regions

András Zeke  | Éva Schád | Tamás Horváth | Rawan Abukhairan |
Beáta Szabó  | Agnes Tantos 

Institute of Enzymology, Research Centre for Natural Sciences, Budapest, Hungary

Correspondence

Agnes Tantos, Institute of Enzymology,
Research Centre for Natural Sciences,
1117 Budapest, Hungary.
Email: tantos.agnes@ttk.hu

Funding information

This work was supported by the National
Research, Development, and Innovation
Office, Hungary grant K-125340.

Edited by: Jeff Wilusz, Editor-in-Chief

Abstract

Recent efforts to identify RNA binding proteins in various organisms and cellular contexts have yielded a large collection of proteins that are capable of RNA binding in the absence of conventional RNA recognition domains. Many of the recently identified RNA interaction motifs fall into intrinsically disordered protein regions (IDRs). While the recognition mode and specificity of globular RNA binding elements have been thoroughly investigated and described, much less is known about the way IDRs can recognize their RNA partners. Our aim was to summarize the current state of structural knowledge on the RNA binding modes of disordered protein regions and to propose a classification system based on their sequential and structural properties. Through a detailed structural analysis of the complexes that contain disordered protein regions binding to RNA, we found two major binding modes that represent different recognition strategies and, most likely, functions. We compared these examples with DNA binding disordered proteins and found key differences stemming from the nucleic acids as well as similar binding strategies, implying a broader substrate acceptance by these proteins. Due to the very limited number of known structures, we integrated molecular dynamics simulations in our study, whose results support the proposed structural preferences of specific RNA-binding IDRs. To broaden the scope of our review, we included a brief analysis of RNA-binding small molecules and compared their structural characteristics and RNA recognition strategies to the RNA-binding IDRs.

This article is categorized under:

RNA Structure and Dynamics > RNA Structure, Dynamics, and Chemistry
RNA Interactions with Proteins and Other Molecules > Protein–RNA
Recognition
RNA Interactions with Proteins and Other Molecules > Small Molecule–
RNA Interactions

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. WIREs RNA published by Wiley Periodicals LLC.

KEYWORDS

intrinsically disordered protein, protein–RNA binding, RNA structure, RNA recognition, IDP–RNA complex

1 | INTRODUCTION

Protein–RNA interactions are indispensable for all living organisms. These molecular complexes not only guide the birth and death of proteins and nucleic acids but are also implicated in a multitude of cellular processes. Proteins administer all events in the life cycle of RNAs, regulating their synthesis, splicing, transport, stability, and degradation (Glisovic et al., 2008). In turn, RNAs also regulate their protein partners through specific targeting, complex assembly, and regulation of activity (Cech & Steitz, 2014), as well as a myriad of other ways. Given the immense importance of RNA binding proteins (RBPs), a lot of scientific effort has been directed towards identifying and characterizing their molecular features, resulting in a significant amount of data (Kilchert et al., 2020; Kwok, 2016; Zhang et al., 2015) accumulated over the recent years. Advances in the identification of RBPs have resulted in the discovery of hundreds of proteins that earlier had not been associated with RNA binding, extending the pool of RNA recognition elements far beyond the traditional RNA binding domains (Beckmann et al., 2015).

As a result of these efforts, several RNA recognition modules have been described, together with their structural characteristics (Corley et al., 2020; Dominguez et al., 2018), such as RNA recognition motif (RRM) domains, K-homology (KH) domains, various zinc fingers such as Cys₃-His₁ (C3H1) domains, as well as many others. While domain-mediated RNA binding has been studied for several years and the available structural information is extensive (Nikulin, 2021), many details of the complex molecular mechanisms driving RNA–protein association remain elusive. Recent advances in the identification of RBPs revealed that a multitude of RNA-binding proteins lack known conventional RNA interaction domains, raising awareness that our knowledge on protein–RNA interaction is far from complete (Beckmann et al., 2015). The most important observation was that protein regions without any fixed 3D structure are also capable of RNA binding either on their own, or in cooperation with structured RNA binding domains (Balcerak et al., 2019; Järvelin et al., 2016). The discovery that intrinsically disordered protein regions (IDRs) binding RNA molecules are often the driving force behind phase transition events within the cells (Chong et al., 2018; Lin et al., 2017), has given a new boost to the research of these types of interactions.

As it is now apparent, proteins use a wide array of different RNA binding elements. In many cases, they combine these modular elements within their sequence (Lunde et al., 2007) to achieve specific, tuneable binding. The versatility of protein–RNA interactions is reflected in the wide range of affinities (from nanomolar to micromolar) and different levels of recognition specificity, from highly specific molecular recognition to promiscuous RNA binding (Corley et al., 2020; Dominguez et al., 2018). As opposed to the folded RNA binding domains, disordered regions do not form classical binding pockets and are generally significantly shorter than the average RNA binding domain. This necessarily results in different binding and recognition strategies of these regions. Protein–protein and protein–DNA binding events are already well-known and thoroughly characterized interactions that can be mediated by disordered regions (Tompa et al., 2015), and it is widely accepted that IDR-mediated binding can have several advantages over globular domain-mediated binding (Habchi et al., 2014). These might include an increased rate of binding, an easy regulation through posttranslational modifications, and a flexibility of the interaction.

While the physiological importance and most details of RNA recognition by the classical RNA binding domains are well established, much less has been discovered about the structural parameters of the IDR–RNA complexes. Disordered RNA binding elements known so far are characterized by very specific amino acid compositions and are most often categorized based on their sequence (Järvelin et al., 2016), rather than their structural properties. A characteristic feature of IDR binding is that they are capable of folding upon binding to their partners (Yang et al., 2019), where IDRs adopt a distinct three-dimensional structure, which can be determined using x-ray crystallography or NMR (Schneider et al., 2019). Analysis of these known IDR/RNA complexes allows us to assess the distinct characteristics of each binding event and understand the molecular details of recognition.

2 | CLASSIFICATION OF DISORDERED RNA BINDING REGIONS

To determine the detailed structural characteristics of IDR-RNA complexes, we conducted a comprehensive search in the Protein Data Bank (rcsb.org; Berman, 2000) for structures of IDRs bound to RNA. After extensive efforts, we managed to collect merely 24 complexes (not always representing different proteins or nucleic acids). Although the number of the available complexes is very limited, we were already able to distinguish two major structural classes, based on the dominant conformation of the IDR in the bound form: either alpha-helical or forming sharp turns or loops. Taking the geometry of the RNA partner into account offers a more refined insight into different possible complexes, giving at least 4–4 subtypes for each group. These structural classes with their most characteristic features are listed in Table 1, while a few representative structures are shown in Figure 1. Most of them are NMR-based structures, with just a few x-ray complexes, sometimes representing the same complex determined with a different method.

2.1 | Protein loops and turns binding RNA

Our structural assessments suggest that superficially similar IDRs can associate with a multitude of different RNA structures. The most fundamental split between RNA–protein complexes—purely from the protein perspective—is between helical and turn/loop type motifs. While both groups contain motifs that are rich in positively charged amino acids (Lys and Arg), motifs from the latter group also incorporate structure-breaking residues (Pro as well as Gly). These regions tend to shun the alpha-helical conformation and fold into turns or loops instead when bound by an RNA partner.

Glycines have an especially important role by allowing the formation of both extended loops and sharp turns that is conducive to RNA–protein interactions. Together with arginines, they form the well-known RGG regions. Deeper analysis of the binding mode of these sequences reveals the driving force behind their sequential characteristics. In an ideal case, the residues with planar architecture, capable of forming pi-stacking interactions (Arg, Tyr, Phe) are also able to form some hydrogen and/or electrostatic interactions, creating stable, strong binding surfaces (Figure 2). Considering the geometry of the nucleic acids, at least two or three smaller amino acids (Gly, Ser, Asp, Asn, etc.) need to separate these larger volume residues, to make stacking against more than one bases possible. This simple model gives rise to the extremely common RGGR or RGGGR repetitive sequences whose physicochemical properties are similar to nucleic acids in many respects, except for their charge. These generic rules are realized also in proteins in more specific versions, as in nucleolins (Masuzawa & Oyoshi, 2020), RBM family proteins (Cai, Cinkornpumin, et al., 2021), DDX4 (Nott et al., 2015), or in FUS (Ozdilek et al., 2017), EWSR1, and TAF15 (Li et al., 2018). FUS contains three RGG domains and its C-terminal RGG box has been shown to recognize G-quadruplex structures in the mRNA of neuronal proteins (Imperatore et al., 2020) with nanomolar affinity. The question whether these specific versions also interact with specific RNA structures has still not been answered (Chau et al., 2016; Ozdilek et al., 2017).

An archetypical RNA structure that is suggested to be bound by RGG regions is the transitional segment between a duplex and a G-quadruplex, as exemplified by the structure presented in Figure 1a. Such regions are widespread in the genome of many Eukaryotes and likely constitute important functional modules (Fay et al., 2017). The key biological role of RGG regions is also indicated by the presence of dedicated secondary modifications. Methyltransferase enzymes can modify arginines in RGG-like context, forming monomethyl or asymmetric dimethylarginines (Fulton et al., 2019). These modifications likely also impact the RNA-binding ability of these loop- or turn-type segments, since the methylated guanidine groups have larger surface, yet less H-bonding opportunities. While no structural data is available at this point regarding arginine methylation and RNA binding, examples of its functional importance abound in the literature (Bhatter et al., 2019; Cai, Yu, et al., 2021; Mersaoui et al., 2019; Tsai et al., 2016).

Several known examples of the turn/loop structural class involve proteins from lentiviruses, such as HIV. These viruses utilize the interaction between the viral Tat protein and the viral mRNA to enhance the transcription of their own proteins. A disordered region in the Tat protein recognizes a region at the 5' end of the RNAs called transactivation response element (TAR). According to the complex structures solved, the peptide binds the RNA in a hairpin-like turn conformation at the stem of a stem-loop structure in the RNA (Davidson et al., 2011; Puglisi et al., 1995; Ye et al., 1995). The sharp, hairpin-like turn structure seems to be important to this type of binding, as the Tat peptides unable to form stable hairpin structures proved to be unsuccessful in binding to the TAR RNA (Davidson et al., 2011). Although the above recognition mechanism is best known for viral protein–RNA complexes, we can find

TABLE 1 Structural classification of intrinsically disordered peptide–RNA complexes found in the Protein Data Bank

Protein structure upon binding	Target RNA 3D structure	Detailed description	Protein sequence features	Example structures (PDB entries)	References
Turn-forming (including beta-turns), loops or random coil	Distorted major (or minor) groove of a double helix	Either random coils or turns sometimes configured almost as beta-sheets	Charges (mostly Arg) intermixed with structure breaking residues (e.g., Gly)	1MNB, 1ZBN, 2KX5, 2KDQ, 2A9X, 6D2U, 1BIV	(Puglisi et al., 1995; Calabro et al., 2005; Davidson et al., 2011; Davidson et al., 2009; Leeper et al., 2005; Shortridge et al., 2019; Ye et al., 1995)
	Loop capping at the end of a double helix	A subclass of the above, but also with added loop capping	Charges, structure breaking plus an aromatic position	484D	(Ye et al., 1999)
	Structural transitions: Duplex-to quadruplex	Short, sharp turns, probably also found at other structural transitions	RGG regions and similarly flexible motifs	2LA5, 5DE5	(Phan et al., 2011; Vasilyev et al., 2015)
	Quadruplex capping	Use of planar pi-stacking of aromatic side chains	Aromatics: Trp, Tyr, or Phe present with Pro	2RU7	(Hayashi et al., 2014)
Partly or completely alpha-helical	Distorted major (or minor) groove of a double helix	Normally bind to the large groove of the RNA in an alpha-helical conformation	Very high Arg and Lys content (including R/E/S-rich regions)	1ETG, 1ULL, 1G70, 1EXY, 1I9F	(Battiste et al., 1996; Ye et al., 1996; Jiang et al., 1999; Zhang et al., 2001)
	Loop capping at the end of a double helix	Binding the groove with a helix and capping it with pi-stacking	Numerous charges (Lys, Arg) with 1 aromatic (e.g., Trp)	1QFQ, 1A4T, 1NYB, 1HJI	(Schärfp et al., 2000; Cai et al., 1998; Faber et al., 2001)
	Structural transitions: Stem-stem junctions	Complex geometry, with both helical as well as non-helical segments	In addition to charges and pi-stacking amino acids: nonhelical and helix breaking	1XOK	(D'Souza & Summers, 2004; Guogas et al., 2004)
	Quadruplex capping	Helix with a very flat, hydrophobic side contacting the RNA	Small amino acids on one side (e.g., Gly, Ala), hydrophilic on the other	2N21, 6Q6R (crystal with DNA only)	(Heddi et al., 2015; Heddi et al., 2020)

Note: Alpha-helical or turn-type motifs can also bind to at least four different RNA structures in each case.

examples of similar binding strategies in other systems too. In the case of human FMRP (fragile X mental retardation protein) binding to a G-rich RNA motif (Figure 1a), the interaction localizes to the base of the G-quadruplex. This specific positioning of the disordered chain can be explained by the peculiar organization of the RNA at the stem of the quadruplex structure. The helix gets distorted to allow for the formation of the quadruplex, enabling the arginine residues of the RGG peptide to access the guanines at the base of the G-tetrad. The strict shape complementarity requires the glycine residues as spacers between the binding arginines (Phan et al., 2011), but they can also form sequence-specific interactions on their own (Vasilyev et al., 2015), contributing to the recognition specificity. Although no detailed structural information is available, the similarities in the recognized RNA structure and the binding mode

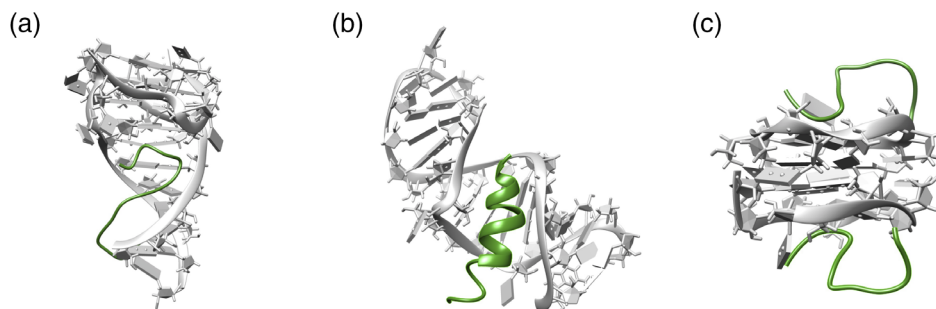


FIGURE 1 Samples structures of a few suggested structural cases: (a) Turns/loops binding to an RNA structural transition: PDB 2LA5. (b) Charged helices within a distorted groove: PDB 1ULL. (c) Quadruplex cappings: PDB 2RSK

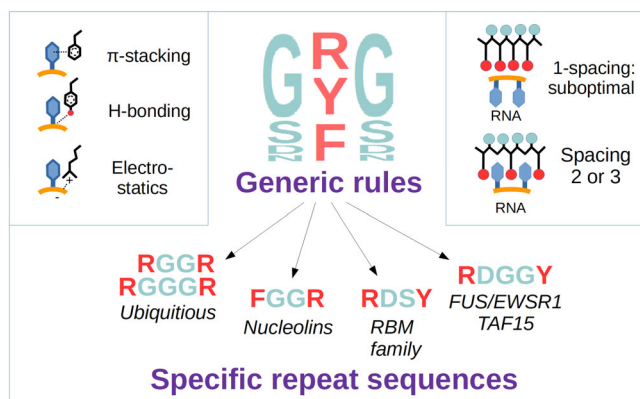


FIGURE 2 Rulesets governing the biochemistry of glycine-rich (loop-like) RNA binding regions. These intrinsically disordered elements are contacting the RNA with an amino acid capable of pi-stacking (Phe, Tyr, Arg), H-bonding (Tyr, Arg), or electrostatic interactions (Arg). To yield the proper side chain geometry, highly flexible residues (preferably Gly, sometimes Ser or other) need to be intercalated at both flanks to the central amino acid. In addition, the physical spacing of nucleobases versus the smaller protein chain calls for more than one such intervening amino acid for optimal RNA–protein contacts

suggest that the C-terminal RGG motif of FUS might apply the same molecular strategy in RNA binding (Imperatore et al., 2020).

High glycine content is not an absolute prerequisite for a protein turn. Peptides with lower Gly content can still form sharp turns or even beta-turns with some beta sheet-like character if their Gly is positioned strategically at the right spot, as indicated by published RNA–IDR structures (Puglisi et al., 1995). Such beta-turns can occupy the distorted grooves of RNA double helices with ease. On the other hand, prolines appear to be much less suited to create RNA-binding segments due to their rigidity but are nevertheless sometimes seen among RNA-capping motifs (Hayashi et al., 2014). Interestingly, the sharp beta-turn binding mode is utilized in some classical RRM domains similarly to IDRs, exemplified by the RMBY–RNA stem–loop interaction (Skrivsovská et al., 2007). Here, a disordered loop within the RRM domain inserts into the major groove of the stem. Mutation analysis showed that this loop is essential for the stem–loop binding capacity of the RMBY RRM.

2.2 | Alpha-helical RNA binding motifs

Helical motifs are mostly represented by short poly-Arg-Lys stretches in published structures, but it is likely that their spectrum is much wider in the whole proteome. Many segments with a high Arg/Glu/Ser content (R/E/S-rich repeats) are likely to adopt this geometry, ideal to bind RNA if their Glu content is much lower than Arg. Detailed analysis of these sequences often shows a marked tendency to adopt a helical conformation when bound. While the physiologically relevant binding modes of SR-repeat sequences have never been uncovered, one possibility is that they also adopt an at least partially helical geometry upon engaging with the RNA. Structure forming tendency in SR-repeat segments

(including nonhelical structures) can be greatly improved by the phosphorylation of the serine residues, which in turn results in a weaker RNA-binding capability (Xiang et al., 2013). These observations might elegantly explain why there is usually little overlap between the serine-rich (preferred geometry: helix) and the glycine-rich IDRs (preferred geometry: sharp turn) in most proteins. It is generally believed that SR regions bind RNA in a sequence-independent manner (Järvelin et al., 2016), and our structural analysis supports this observation. Since the helical region of the protein is inserted in the major groove of the RNA double helix, these interactions are likely to be driven by electrostatic interactions rather than pi-stacking as observed in the RGG motifs.

As for the recognized RNA structures, there are a few examples where short RNA-binding alpha-helices bind an intact RNA double helix (conformation A), although they are only a tiny minority, even when considering folded RNA-binding domains (exemplified by Staufen nucleases; Yadav et al., 2020). The majority of helical RNA-binding regions, especially the longer ones, prefer RNA segments with distorted structures (Bartel et al., 1991; Leclerc et al., 1994). These segments are poor in perfect base-pairing, but those that are formed are especially important to maintain the global conformation of the RNA chain (Peterson et al., 1994). The addition of the protein helix can likely yield further stabilization of these distorted structures, also altering their winding parameters in the process (Tanaka et al., 1999). It has also been observed that the binding of the protein partner can induce rearrangement in the base-pairing pattern of the RNA, facilitating the interaction (Battiste et al., 1994; Zhang et al., 2001). Experimental results suggest that this type of recognition is more reliant on structural determinants than on sequence specificity (Zhang et al., 2001). Arginine-rich motifs have also been shown to bind GNRA tetraloops (Legault et al., 1998), an interaction crucial for several bacteriophages. This type of binding requires the helix to adopt a bent conformation, forming a cap-like structure at the top of the tetraloop (Schärfp et al., 2000). In addition to the above-mentioned binding modes, more complex interaction strategies were also observed for Arg-rich motifs. The genomic RNA of the Alfalfa mosaic virus (AMV) relies on the interaction with the viral coat protein (CP) to properly fold its 3' UTR for RNA polymerase recognition (Houser-Scott et al., 1994). Detailed structural analysis revealed that the RNA adopts a complex structure containing two hairpins and a number of non-canonical base-pairings which are facilitated by the binding of the disordered motif in CP. The RNA binding region of CP, disordered in its free form, adopts an alpha-helical structure with an extended N-terminal part (Guogas et al., 2004).

2.3 | Modeling of RNA–protein complexes complements our knowledge

The relative scarcity of published structures often urges state-of-the-art RNA biologists to resort to molecular modeling, docking, and/or molecular dynamics. This can be done by either homology modeling, de novo assembly, or guided modeling using experimental restraints (e.g., SHAPE; Li, Cao, et al., 2020). Even if experimental data on solvent exposure or crosslinking points are not available, potential 3D models of nucleic acids can still be informative. These computational biology tools are useful to either gain insights into a (possible) geometry of an unknown RNA–protein complex or at least to test the stability of a suggested conformation (Dawson & Bujnicki, 2016). In this article, we briefly illustrate the power of modeling with two theoretical examples. These are by no means meant to be fully representative of the arsenal of modeling approaches used today (Dawson & Bujnicki, 2016). To show that disordered segments can stably adopt either a turn-type or a helical geometry when binding different types of complex RNA folds, we conducted a few short folding and docking simulations of RNA–protein partner pairs.

Our first example involves a short, but complex RNA molecule, which is sometimes misclassified as a simple stem-loop structure if only 2D prediction methods are used. The short RNA segment (WECcore) identified as a binding partner of RBMX protein disordered segment by Kanhoush et al. (Kanhoush et al., 2010) was modeled here. We started from a stem-loop structure, with the RNA gradually refolding to a more stable state using simulated annealing. As shown in Figure 3a, the resulting complex RNA formed kink- and bulge-rich structures with a lot of optimal slots for either peptide turns or short helices. Then the protein was prepared similarly, docked to the RNA, and was subjected to replica-exchange dynamics until the complex stabilized. The most populous cluster shows a predominantly turn-type peptide geometry engaging the RNA, as already suggested by the Gly-rich sequence of the former molecule (Figure 3b). Interestingly, this RGG region appears to be a critical RNA-binding hotspot, and its selective loss has been implicated in the hereditary disease Shashi-XLID syndrome (Cai, Cinkornpumin, et al., 2021).

Our other example relates to the poorly known SR-repeat segments frequently found in splicing factors as well as R/E/S rich regions in other RNA-binding proteins (Daniels et al., 2021). While their real RNA targets and bound conformations are unknown, we can make a few assumptions based on modeling. Models based on existing, published

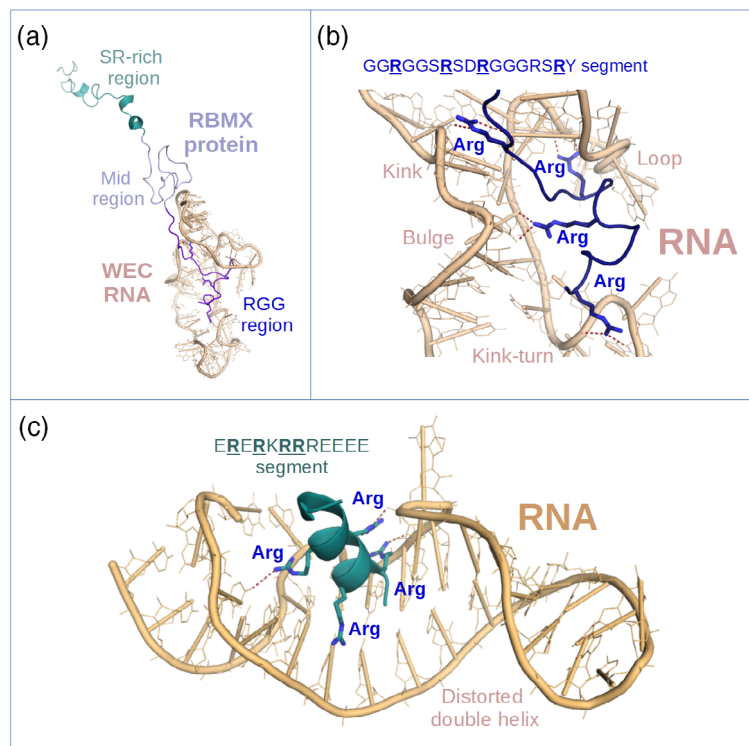


FIGURE 3 Modeling of two RNA–protein structures. The RBMX protein has both SR-rich and RGG-type regions, out of which the latter docked to the model complex RNA (WEC) published as a binding partner (a). Zooming into the RNA–protein contacts show that four arginine residues play a key role through establishing numerous polar contacts to both the sugar-phosphate backbone and the nucleobases of the complex RNA (b). Although its exact RNA partners are unknown, the Arg/Glu/Ser rich segments of the RNA binding disordered segments of LUC7L3 are likely to be helical, and dock stably into a model RNA molecule binding helical peptides (c)

RNA–protein structures suggest that distorted RNA double helices could provide an excellent target for Arg/Glu/Ser rich segments, should the latter become alpha-helical. Thus, if we docked short alpha-helical models of select LUC7L3 protein segments into a suitable RNA model [derived from Ye et al. (1996)], they turned out to be rather stable under molecular dynamics (Figure 3c). These examples are simple, but powerful illustrations of how advanced modeling techniques will help to formulate new structural hypotheses regarding RNA–protein complexes: Especially where experimental structure determination is not an option yet due to the size and heterogeneity of the RNA molecules.

2.4 | RNA–protein contacts at an atomic level

Going deeper into the available structures reveals more interesting details than the secondary structures. It turns out that there are fairly generic physicochemical rules governing how amino acid side chains can bind RNA, that also act the same way in ordered RNA-binding domains. At a physical level, these interactions can be described with electrostatic (charge interaction), pi-stacking (planar delocalized and aromatic systems), and H-bonding contributions (Baulin et al., 2020; Krüger et al., 2018). These interactions operate slightly differently on the sugar-phosphate backbone (with a lot of charge interactions and some H-bondings) and the nucleobases themselves (where pi-stacking and directed H-bonding predominate).

Amino acids can often participate in more than one type of contact. While phenylalanine and tryptophan usually only yield pi-stacking against the nucleobases, tyrosine can sometimes also provide H-bonding opportunities. Asparagine and glutamine side chains are surprisingly good at providing multiple H-bonds (both as acceptor and donor) towards bases, in addition to a weaker pi-stacking option. Lysine mostly provides the charge, but often it also gets H-bonded to the main chain. And finally, arginine is the most versatile of all chemical structures (Krüger et al., 2018). The guanidine group carried by arginine side chains is a veritable “Swiss army knife” when it comes to binding RNA. It is protonated under physiological pH and has numerous hydrogens, yielding an excellent sugar-phosphate backbone

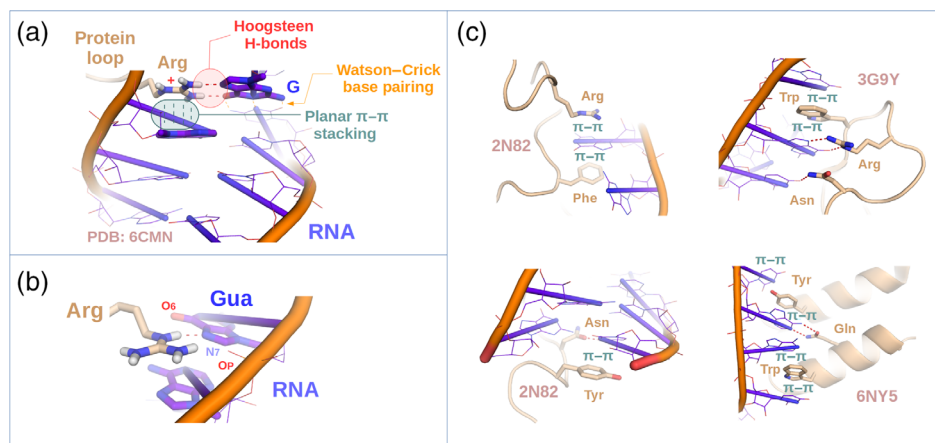


FIGURE 4 Atomic-level details of RNA–protein interactions. The importance of arginine (Arg) lies in the wealth of molecular interactions it can establish: Electrostatic interactions, pi–pi stacking as well as dedicated H-bonds, preferably toward the Hoogsteen edge of guanosine (G) nucleobases (a). It can bind the sugar–phosphate backbone or nucleobases or even both simultaneously (b). Optimal coordination of nucleobases can only be achieved through consecutive amino acids that provide both pi–stacking and H-bonding interactions, as shown by examples from the protein data bank (c)

binder. But the guanidine group is also highly planar due to its delocalized bonds, becoming an outstanding pi-stacking partner, to cap any nucleobases left exposed to the solvent. At the same time, arginine is more than just an intercalator: In many of the known structures, it preferably contacts guanine (G) nucleobases along their Hoogsteen edge, keeping Watson–Crick contacts between nucleobases. The fact that it can also excellently stack on purine rings might explain why Arg-rich disordered segments often have a preference toward complex G-rich RNA structures (Gupta & Gribskov, 2011; Ozdilek et al., 2017). What is more, Arg can also be a simultaneous nucleobase binder, stacking element as well as backbone interactor at the same time (Figure 4a,b).

Amino acids at neighboring positions along with the peptide chain must also cooperate heavily to properly coordinate the winding RNA backbone and its exposed nucleobases (Figure 4c). This local interdependence explains the high glycine and serine content of many RNA-binding disordered motifs, in addition to other amino acids that cannot bind RNA themselves, such as aspartate and glutamate. While the latter two probably contribute to motif helicities by creating internal structures, glycine can only be truly understood in a loop- or turn-like structural context. In the latter geometries, the key amino acids (mostly Arg, but also Tyr or Phe) are typically alternating with the turn-forming glycines or serines. This gives optimal exposure for the large side chain to stack between nucleobases, especially if not just one, but at least two such intervening small amino acids are used. Such a simple algebra can already explain why (RGG)_n-like repeats are suitable RNA binding elements, in addition to their more specialized variants (Figure 4c). Different amino acids have slightly differing nucleobase preferences (e.g., Phe prefers to contact U; Wilson et al., 2016), giving a possible biochemical explanation for the existence of these specialized repeats.

2.5 | DNA versus RNA: Shared and distinct protein partnerships

Because the chemical nature and building blocks of nucleic acids are so similar, it is imperative to compare RNA-associated disordered motifs to their DNA binding counterparts. From a structural point of view, cellular DNA is much simpler than RNA. Unless locally damaged or unzipped by enzymes, it almost exclusively folds into a near-infinite double-helical structure (B helix). Unsurprisingly, DNA associating disordered protein regions are structurally much less variable than those binding the structural imperfections of an RNA molecule: most of them adopt an α -helical conformation. The best examples of helices inserting into the major groove of DNA can be seen among basic leucine zipper (bZIP) motifs (Figure 5a; Vinson et al., 2006). These proteins utilize specific variations of amino acids within their basic regions to achieve sequence-specific binding (Miller, 2009) as opposed to the mainly structure-based recognition in IDR–RNA interactions.

Nevertheless, loop-like motifs were also described among DNA binding elements, notably lacking any base-intercalating interactions. Their best-characterized examples are the adenosine-thymidine-hook (AT-hook) motifs (Susbielle et al., 2005). Rich in Arg and Gly/Ser, the latter do resemble their RNA binding RGG-type counterparts. They

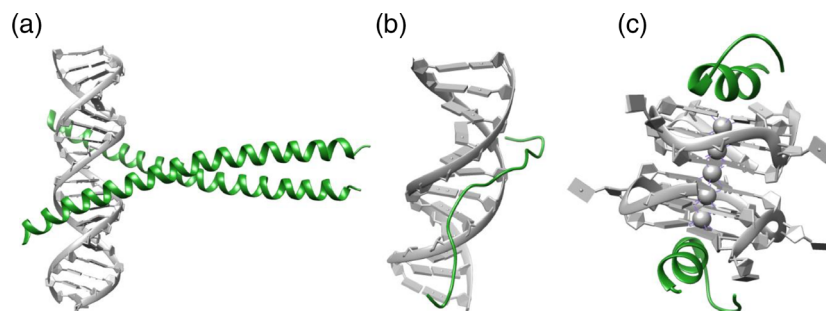


FIGURE 5 Representative structures of structurally different classes of DNA binding disordered protein segments: (a) Double helix binding helical/helix-containing motif (basic leucine zipper (bZIP); PDB 1JNM), (b) Loop-like binding (AT-hook; PDB 2EZD). (c) Capping motif (G-quadruplex capping; PDB 6Q6R) [Correction added on 9 February 2022 after first online publication: Figure 5 has been updated; Figure 1 was incorrectly published as Figure 5.]

preferentially bind to the minor groove of the DNA in a nonspecific manner driven by electrostatic interactions (Brodsky et al., 2021; Rohs et al., 2009), although examples of sequential recognition have been also described (Crane-Robinson et al., 2006; Huber et al., 2012; Tunnicliffe et al., 2017). While purely disordered dsDNA-binding loop- or turn-configured regions are rare, they are likely to be more common as auxiliary elements found alongside well-established folded DNA-binding domains. Such elements, cooperating with neighboring domains are likely important for precise transcription factor binding site specification (Brodsky et al., 2020). In addition, some highly exotic DNA binding proteins, such as the tardigrade DNA damage suppressors (Dsup) might perform dsDNA binding without helical structures (Mínguez-Toral et al., 2020).

Since perfectly complementary, double-stranded DNA admits no stable stem-loops or quadruplexes, many other binding modes seen with RNA are not observable, except for rare proteins interacting with transient cellular DNA configurations. Proteins specifically recognizing damaged DNA (e.g., DDB2 or APE1) often make use of loops that recognize imperfectly paired nucleotides, with loose similarity to their RNA-binding counterparts (Fischer et al., 2011; Mol et al., 2000). Similarly, quadruplex-binding proteins can also recognize G-rich DNA strands if they become separated during transcription (due to R-loop formation; Dettori et al., 2021).

2.6 | Other RNA ligands: Small molecules versus proteins and peptides

RNA-binding proteins are not the only biomolecules that are important for RNA structure and function, as RNA molecules typically require cellular counter-ions to establish stable structures (Kolev et al., 2018). The role of cations (Na^+ , K^+ , Mg^{2+} , etc.) is most obvious at folds where the sugar-phosphate backbones contact, or at quadruplexes (Bhattacharyya et al., 2016). However, inorganic ions can also be exchanged for organic ones, especially polyamines (spermidine, spermine) or even Lys/Arg side chains of peptides can fulfill the same role.

When studying disordered protein segments binding to RNA, one cannot avoid discussing various small molecules targeting nucleic acids (Aboul-ela, 2010; Hargrove, 2020; Warner et al., 2018). In addition to natural ligands (e.g., riboswitches and various antibiotics), the field of synthetic RNA-binding molecules has been rapidly expanding in recent years. Small molecules bind complex RNA in ways that are similar to disordered peptides in many regards. Such ligands now also include a lot of novel, early-stage pharmaceutical candidates. Similar to their peptidic counterparts, many small molecules are fairly unspecific, and will happily bind unrelated RNAs, provided that they possess the required structural elements (e.g., unstacked nucleotides, bugles, etc.; Kelly et al., 2021).

Chemically speaking, small molecules use the same three major features to bind nucleic acids as disordered proteins do (Figure 6b). Positive charges offer a generic, nonspecific attraction to RNA, while hydrogen bond donor or acceptor groups yield more specific contacts. The third important chemical feature is planarity/aromaticity, allowing for pi-stacking against exposed nucleobases. These features also combine in certain functional groups, for example, charged rings, aromatic amides, and so forth, giving rise to the chemotypes often observed among RNA binding synthetic ligands (Figure 6a; Childs-Disney et al., 2018; Hermann, 2003).

Unfortunately, our current knowledge of non-double-helix nucleic acid structures is limited, precluding the structure-guided design of specific RNA binders. What is clear is that small-molecule ligands tend to prefer irregular

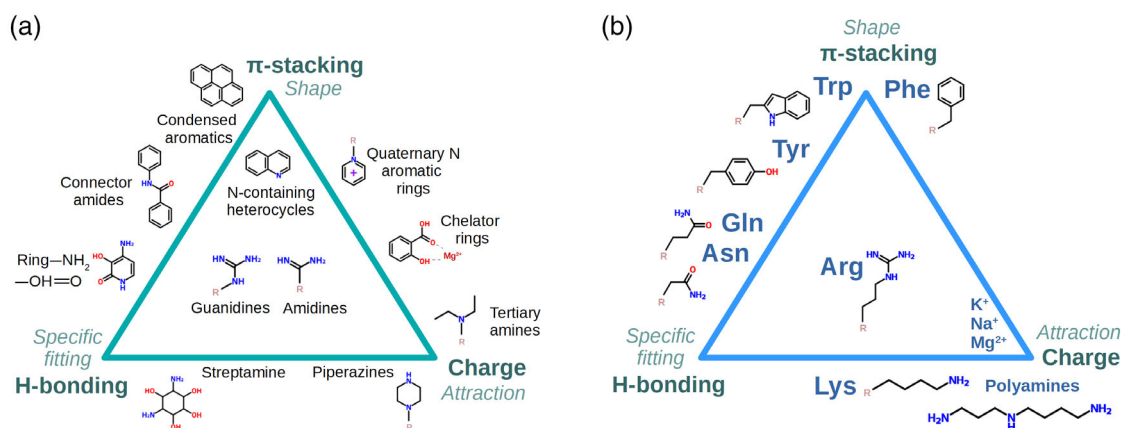


FIGURE 6 Examples of chemical moieties found in synthetic small molecule RNA ligands (a) compared to amino acid side chains in RNA binding proteins and other biomolecules (b). Guanidine and amidine groups fulfill a special role in both categories, capable of π-stacking, charge interactions, and H-bonding to the RNA nucleobases at the same time

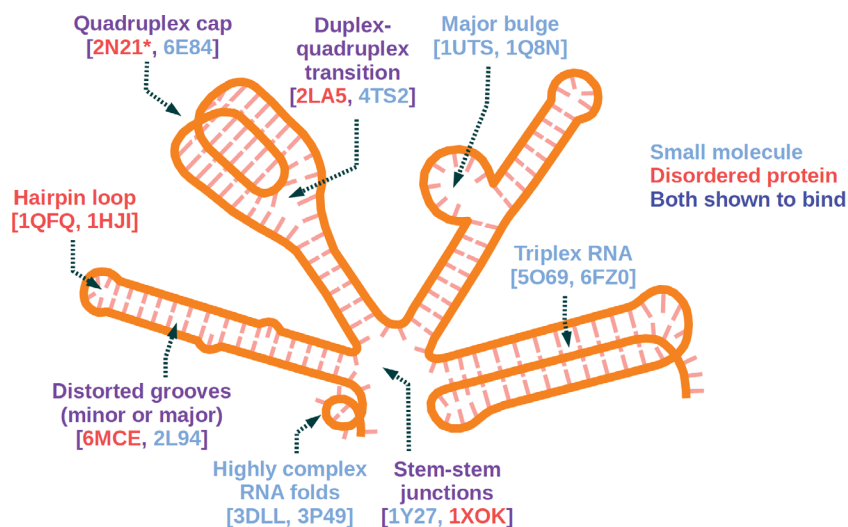


FIGURE 7 Binding sites of small molecules (blue), disordered peptides (red) or both (magenta) observed in published PDB structures (* marks examples that bind RNA and DNA similarly, and were crystallized with the latter)

structures as well as spots where more than two RNA strands meet. The simplest of these structures are single-sided bulges between two loops, while more complex examples include imperfect triplex and quadruplex-like highly complex regions, or junctions and branch points within RNA structures. Overall, the known binding sites of small molecules appear to be more diverse than those of disordered proteins, although protein/peptide binding sites have not been explored very exhaustively (Figure 7).

2.7 | Imperfections make RNA unique and provide the binding sites

Since natural RNA molecules rarely adopt a perfect A-helix geometry and even the “stem” regions most often have a few mismatching bases, bulging nucleotides or some kink-turns, RNA-binding proteins rarely have the luxury to bind a complete double helix. Natural RNA folds contain a lot of structural transitions, including stem–stem-junctions, complex branchings, kissing loops and they can fold into more complex 3D structures, including triple-helical or quadruplex regions. Consequently, their protein partners are best adapted to bind the very structural imperfections that make the RNA so unique. Even if the binding site falls to a roughly double-helical region, it is often distorted: imperfect base-pairings and noncanonical base-backbone contacts are their key characteristics. Bulges and junctions between

multiple loops also offer unique binding sites for interacting protein chains. Triple helical or quadruplex regions show complicated base stackings across both the Watson–Crick and the Hoogsteen interface, as well as other interfaces in some cases, with plenty of opportunities for a planar ligand to stack on top of them, stabilizing their start or end (Figure 7).

Due to the complex structures and high flexibility of both RNA and IDRs, it is not unprecedented that a physiologically relevant RNA structure requires a protein partner to be able to form, through a highly specific co-folding mechanism (Guogas et al., 2004). The exposure of specific bases can also serve as a regulatory step for protein binding, as exemplified by the nucleocapsid (NC) binding to retroviral RNA dimers (D'Souza & Summers, 2004). In this system, the protein-interacting RNA motif remains buried through base-pairings in the monomer and only becomes exposed when dimerization occurs—offering a binding surface for the NC protein.

2.8 | The role of repeats and multivalency

While RNA-binding disordered segments can sometimes be realized as true short linear motifs (<15 aa long), it is more common to see them as repetitive elements stretching over dozens if not hundreds of amino acids. Such regions may contain numerous imperfect repeats of the same basic RNA binding motif, best exemplified by the RGG/RG or SR-type repeats (Lin & Fu, 2007). While the former probably binds in a loop-like conformation to complex RNA segments, the latter is structurally unexplored, although theoretically, SR-type repeats could adopt either random coil/loop or helical conformation. Often grossly conserved as a long region in RNA-binding proteins, these motifs are evolutionarily less constrained due to their tandem nature and behave differently from canonical linear motifs. Imperfect and suboptimal repeats are well tolerated as long as the region or the protein as a whole retains its function.

Having many RNA-binding elements instead of one in the same protein confers numerous advantages. The protein can now engage multiple, large RNA molecules in a fairly stochastic manner (Figure 8a). Having these proteins in high concentration inside various subcellular compartments will inevitably lead to liquid–liquid phase separation and the rise of membraneless organelles (nucleoli, Cajal bodies, P-bodies, stress granules, etc.; Protter et al., 2018). RNA–RNA interactions further aid the formation of these structures (Van Treeck & Parker, 2018). Many, but not all loose

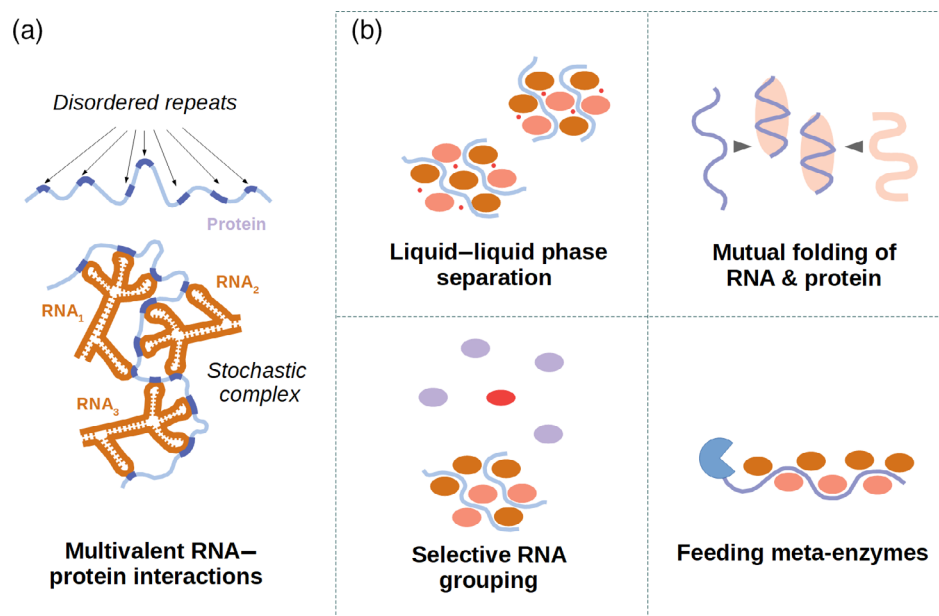


FIGURE 8 The role of multivalency in disordered protein–RNA interactions. Proteins carrying numerous RNA-binding tandem repeats can engage multiple RNA molecules at their structurally matching sites, to form stochastic complexes (a). These complexes endow the cells with numerous advantages (b): They allow the formation of organelles through liquid–liquid phase separation. Specific, repetitive proteins can recruit specific sets of RNA into these complexes, leaving others out. RNAs and disordered proteins can exist in a symbiotic relationship, properly folding only in the presence of their partner. Last but not least, these complexes can contain enzymatic components, processing the recruited nucleic acids

complexes also tend to incorporate certain, specific RNAs only (thanks to both 3D structural and sequential information encoded in RNA), while excluding other nucleic acids. On the other hand, many RNA-binding proteins can also accumulate in nucleoli in a fairly nonspecific manner (Emmott & Hiscox, 2009).

Some important RNAs only fold properly when complexed inside these organelles, as shown by the rather complicated ribosomal and spliceosomal biogenesis: Here, the assembly of particles often demands that the nucleic acids and the proteins fold in cooperation. This mutual stabilization aspect might be widespread, but it is still unexplored for other stable RNA–protein complexes (such as lncRNAs with their partner proteins). But, these protein–RNA complexes are more than just randomly oriented and composed assemblies: they can incorporate catalytic components that use other RNA binding, disordered proteins to constantly feed them with substrates (as envisioned in P-bodies, the site of controlled mRNA degradation; Figure 8b). And while many examples show that protein binding can stabilize RNA structure (Frenkel-Pinter et al., 2020; Guogas et al., 2004), we can find instances where the opposite of this is true. In the bipartite binding of FUS to RNA, the disordered RGG region in FUS destabilizes the structure of the RNA, probably facilitating the binding of other interacting partners (Loughlin et al., 2019).

2.9 | Folded RNA binding domains and IDRs in physiology and disease

Although, with the exception of FMRP, the examples listed above represent mostly disordered proteins without other known RNA binding domains, it is important to note that ordered RNA binding elements and disordered regions are capable to bind their nucleic acid partners independently. However, they are commonly seen to synergize in naturally evolved proteins, with the IDRs preferentially located in proximity of known RNA binding domains (RRM, KH, Zn-fingers, etc.) Their cooperation is likely mutually beneficial due to different modalities of binding: known RNA binding domains tend to show sequence-specific binding to linearly exposed RNA epitopes (e.g., at loops), while IDRs do the opposite, associating in a structure-dependent manner to highly folded, complex RNA regions. Their difference in binding supports a simplified hypothesis that the domains mainly contribute to sequence specificity, while the IDRs greatly increase affinity and add a layer of (looser) structure specificity (Loughlin et al., 2019). Both are indispensable for the complete biological function. Cooperative function of the ordered and disordered RNA binding domain can also be rationalized by the well-known high binding speed of IDRs (Tompa, 2003), which enables the tethering of partner RNAs while the relatively slower folded domains establish the appropriate contacts.

Because RNA-binding IDRs can be involved in recruiting a multitude of different target RNAs, their alterations are most likely to cause complex cellular effects, with the genotype-to-phenotype correlations quite difficult to predict. One important feature that mutations in RNA-binding IDRs can alter, is the liquid–liquid phase separation capability of these proteins. Altered amino acid sequence can result in disarranged RNA interactome, but also in the changed structural behavior of the protein, leading to the formation of solid aggregates (Armaos et al., 2021). Spliceosomal networks are also highly sensitive to mutations in RNA binding regions, ordered and disordered alike. Mutations can lead to the remodeling of the complete spliceosomal network (Lang et al., 2021), or can cause the abnormal localization of the protein (Gaertner et al., 2020), preventing the formation of physiologically relevant interactions. Unsurprisingly, mutations of RNA binding proteins are common cause of various diseases (Gebauer et al., 2021). While the folded domains are preferentially targeted by pathogenic mutations, alteration of RNA-binding IDRs is also increasingly recognized.

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia are complex genetic diseases that can typically be caused by mutations to intrinsically disordered RNA-binding protein regions (Harrison & Shorter, 2017). In FUS and TAF15, such mutations often impact the RGG regions flanking the folded RNA-binding domains (RRM domains and Zn fingers), in addition to the aggregation-prone GYQ-rich regions (Couthouis et al., 2011; Deng et al., 2014). In TDP43, a single region is responsible for liquid condensate formation and potential auxiliary RNA binding, which is frequently hit by ALS-inducing point mutations (Prasad et al., 2019). Although less common, RGG-like region affecting mutations are also seen in EWSR1 and hnRNPA1 (Couthouis et al., 2012; Kim et al., 2013). In all these cases, most typically the glycines are mutated into other amino acids—mutation of arginines or other amino acids is notably less common. Glycine point mutations subtly decrease the RNA binding potency, increase peptide chain rigidity, and facilitate stable beta-sheet formation, especially in the less charged subregions. All these combined effects lead to the formation of insoluble cellular aggregates (Harrison & Shorter, 2017). Mutations or truncations of RNA-binding IDRs also cause other rare genetic diseases: In the case of Shashi-XLID syndrome, the loss of the RBMX protein C-terminus is seen (Cai, Cinkornpumin, et al., 2021). Kabuki syndrome is another congenital disease caused by deleterious alterations

(truncations and mutations) in the MLL4 (KMT2D) gene. One such mutated region lies between exons 31 and 39 in KMT2D, encoding a disordered RNA-binding segment (Liu et al., 2015; Szabó et al., 2018). The specific mutations located in this conserved region cause a multiple malformations disorder that is similar, but not identical to Kabuki syndrome (Cuvertino et al., 2020).

While RNA-binding IDRs were not observed as a primary somatic mutational target in most cancer types, some intriguing cases have already been identified. In myelodysplastic syndrome, the malignant transformation is initiated by mutations to various splicing modulators. Among them, SRSF2 is typically mutated at a single residue (Pro95) lying the IDR directly adjacent to its folded RNA binding domain. This particular substitution is known to alter the RNA binding sequence specificity of SRSF2, with a consequential change of numerous splicing events, and genome-wide mis-splicing (Kim et al., 2015).

2.10 | True RNA structure is almost never two-dimensional

It is now clear that three-dimensional RNA structure is the dominant factor in protein–RNA complexes, even in interactions mediated by folded domains. If a domain requires an ssRNA segment for binding, it can only happen at places where such segments are structurally exposed (e.g., at RNA loops). Due to entropic factors, the importance of a preformed RNA structure is even higher when the association happens through an intrinsically disordered protein segment. For this reason, we need to revisit the RNA structures that are recognized by their protein counterparts. Unlike DNA, ribonucleic acids can never be fully comprehended using 1D or 2D models. In reality, many RNA molecules have a complicated 3D structure that only rivals that of proteins in complexity. But unlike proteins, where the basic structural elements (i.e., helices, beta-sheets, turns, and disordered regions) are well-established and relatively easy to predict by now, we are still struggling to understand folded ribonucleic acids. The current nomenclature of basic RNA secondary structures (such as various turns and loops) is both complicated, cumbersome, and incomplete (Ge et al., 2018). The complexity and heterogeneity of larger RNA structures are mirrored in the fact that up to this day there is no definitive structural classification of RNA structures, comparable to the structural features of proteins. We clearly need better models that can only be obtained by experiments. But we are still struggling to delineate folded and disordered RNA regions. New data indicate that many RNAs can also be modular, with nearly independent structural elements resembling protein domains (Lu et al., 2020). While noncoding RNAs can be evolutionarily optimized for the structure alone, mRNAs have to carefully balance between sequence (that encodes the genetic information) and the structure (providing stability and guiding protein–RNA interactions).

In silico prediction of correct RNA structures still presents a major challenge, as RNA structural predictions are often very weak (only 2D, disregarding topology issues such as pseudoknots), underperforming protein structural predictions and not taking the possible 3D structure into account. Available structural information shows that (i) RNA is rarely ever single-stranded in real structures. Instead, it forms loose pairings or even higher-order multimers to avoid single-stranded topology. (ii) Among the predicted stem-loops, truly unfolded “loops” rarely exist in 3D: instead, they tend to be cappings, K-turns, or bulges, if not forming an entirely different fold. True loops can only stabilize if they are bound by another macromolecule (usually protein). (iii) Moderately complex 3D regions (K-turns, kissing loops, pseudoknots, triple helices, and quadruplexes) are very common in real RNA folds, but cannot be predicted by 2D methods. These conclusions are all supported by a closer examination of larger, experimental RNA structures. It remains to be seen how prediction methods will evolve in the upcoming years, to bridge the considerable gap between top-down and bottom-up approaches (Dawson & Bujnicki, 2016).

3 | CONCLUSION

In recent years, our knowledge regarding nucleic acids and proteins as well as their complexes expanded considerably. Unusual complexes, formed by IDRs with structured RNA, pushed the boundary of our knowledge in structural biology. RNA–IDR complexes challenge many traditionally held views of how we envision ribonucleic acids or proteins. Still, these complexes are an essential part of cellular life and more detailed research on them will inevitably lead to further key revelations.

As we have shown in our review, disordered regions use fundamentally different binding and partner recognition strategies than globular RNA binding domains. Their unique characteristics enable them to preferentially recognize

structural features rather than specific sequences. To add a further layer of complexity, these different binding elements are most often found in combination with conventional RNA binding domains (Järvelin et al., 2016). There are also examples in the literature, where the cooperativity of the different RNA binding elements has been shown experimentally (Loughlin et al., 2019). Together, they can recognize highly structured RNA segments more efficiently than alone. During protein–RNA binding events, both structural and sequential information of the nucleic acid partner is required for proper recognition. One can only speculate about the key role IDRs play in structure- rather than sequence-specific recognition of RNA partners. These partnerships are important for regulation: Highly structured RNA regions tend to form a higher number of intermolecular interactions (Sanchez de Groot et al., 2019), probably linked to the tighter transcriptional regulation of these RNAs.

Despite several decades of advancement in protein science, in-depth structural studies of folded RNA molecules have just begun. Knowing more of the nucleic acid partner will be essential for a better understanding of how other molecules, such as proteins can interact with them. In this regard, the short RNA-binding IDRs are of great importance. If we could better understand the physicochemical requirements behind their selective binding to certain RNAs, that knowledge could be used to engineer new synthetic analogs. It could finally pave the way to synthetically designed RNA-binding pharmaceutical agents, expanding them well beyond natural antibiotics. Thus, studying IDR–RNA complexes will certainly yield many exciting theoretical as well as practical discoveries in the years to come.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

AUTHOR CONTRIBUTIONS

András Zeke: Conceptualization (equal); data curation (equal); formal analysis (lead); investigation (equal); methodology (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal). **Éva Schád:** Data curation (equal); investigation (supporting); methodology (supporting); software (supporting); visualization (equal); writing – review and editing (equal). **Tamás Horváth:** Formal analysis (supporting); investigation (supporting); methodology (supporting); software (lead); writing – review and editing (supporting). **Rawan Abukhairan:** Data curation (equal); investigation (supporting); writing – review and editing (supporting). **Beáta Szabó:** Data curation (equal); investigation (equal); project administration (lead); writing – review and editing (equal). **Agnes Tantos:** Conceptualization (lead); formal analysis (equal); funding acquisition (lead); investigation (equal); methodology (equal); supervision (lead); visualization (equal); writing – original draft (lead); writing – review and editing (lead).

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

András Zeke  <https://orcid.org/0000-0002-5229-8617>

Beáta Szabó  <https://orcid.org/0000-0003-2287-0827>

Agnes Tantos  <https://orcid.org/0000-0003-1273-9841>

RELATED WIREs ARTICLE

[Classification and function of RNA–protein interactions](#)

REFERENCES

- Aboul-ela, F. (2010). Strategies for the design of RNA-binding small molecules. *Future Medicinal Chemistry*, 2(1), 93–119.
- Armaos, A., Zacco, E., Sanchez de Groot, N., & Tartaglia, G. G. (2021). RNA-protein interactions: Central players in coordination of regulatory networks. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 43(2), e2000118.
- Balcerak, A., Trebinska-Stryjewska, A., Konopinski, R., Wakula, M., & Grzybowska, E. A. (2019). RNA-protein interactions: Disorder, moonlighting and junk contribute to eukaryotic complexity. *Open Biology*, 9(6), 190096.
- Bartel, D. P., Zapp, M. L., Green, M. R., & Szostak, J. W. (1991). HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. *Cell*, 67(3), 529–536.
- Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D., & Williamson, J. R. (1996). Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science*, 273(5281), 1547–1551.
- Battiste, J. L., Tan, R., Frankel, A. D., & Williamson, J. R. (1994). Binding of an HIV Rev peptide to Rev responsive element RNA induces formation of purine-purine base pairs. *Biochemistry*, 33(10), 2741–2747.

- Baulin, E., Metelev, V., & Bogdanov, A. (2020). Base-intercalated and base-wedged stacking elements in 3D-structure of RNA and RNA-protein complexes. *Nucleic Acids Research*, *48*(15), 8675–8685.
- Beckmann, B. M., Horos, R., Fischer, B., Castello, A., Eichelbaum, K., Alleaume, A. M., Schwarzl, T., Curk, T., Foehr, S., Huber, W., Krijgsveld, J., & Hentze, M. W. (2015). The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nature Communications*, *6*, 10127.
- Berman, H. M. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28*, 235–242. <https://doi.org/10.1093/nar/28.1.235>
- Bhattacharyya, D., Mirihana Arachchilage, G., & Basu, S. (2016). Metal cations in G-quadruplex folding and stability. *Frontiers in Chemistry*, *4*, 38.
- Bhatter, N., Roy, R., Shah, S., Sastry, S. P., Parbin, S., Iyappan, R., Kankaria, S., & Rajyaguru, P. I. (2019). Arginine methylation augments Sbp1 function in translation repression and decapping. *The FEBS Journal*, *286*(23), 4693–4708.
- Brodsky, S., Jana, T., & Barkai, N. (2021). Order through disorder: The role of intrinsically disordered regions in transcription factor binding specificity. *Current Opinion in Structural Biology*, *71*, 110–115.
- Brodsky, S., Jana, T., Mittelman, K., Chapal, M., Kumar, D. K., Carmi, M., & Barkai, N. (2020). Intrinsically disordered regions direct transcription factor in vivo binding specificity. *Molecular Cell*, *79*(3), 459–471.e4.
- Cai, T., Cinkornpumin, J. K., Yu, Z., Villarreal, O. D., Pastor, W. A., & Richard, S. (2021). Deletion of RBMX RGG/RG motif in Shashi-XLID syndrome leads to aberrant p53 activation and neuronal differentiation defects. *Cell Reports*, *36*(2), 109337.
- Cai, T., Yu, Z., Wang, Z., Liang, C., & Richard, S. (2021). Arginine methylation of SARS-Cov-2 nucleocapsid protein regulates RNA binding, its ability to suppress stress granule formation, and viral replication. *The Journal of Biological Chemistry*, *297*(1), 100821.
- Cai, Z., Gorin, A., Frederick, R., Ye, X., Hu, W., Majumdar, A., Kettani, A., & Patel, D. J. (1998). Solution structure of P22 transcriptional antitermination N peptide-boxB RNA complex. *Nature Structural Biology*, *5*(3), 203–212.
- Calabro, V., Daugherty, M. D., & Frankel, A. D. (2005). A single intermolecular contact mediates intramolecular stabilization of both RNA and protein. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(19), 6849–6854.
- Case, D. A., Aktulga, H. M., Belfon, K., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E., III, Cisneros, G. A., Cruzeiro, V. W. D., Darden, T. A., Duke, R. E., Giambasu, G., Gilson, M. K., Gohlke, H., Goetz, A. W., Harris, R., Izadi, S., Izmailov, S. A., Jin, C., ... Kollman, P. A. (2021). *Amber 2021*. University of California.
- Cech, T. R., & Steitz, J. A. (2014). The noncoding RNA revolution—trashing old rules to forge new ones. *Cell*, *157*(1), 77–94.
- Chau, B. L., Ng, K. P., Li, K. K. C., & Lee, K. A. W. (2016). RGG boxes within the TET/FET family of RNA-binding proteins are functionally distinct. *Transcription*, *7*(4), 141–151.
- Childs-Disney, J. L., Tran, T., Vummidi, B. R., Velagapudi, S. P., Haniff, H. S., Matsumoto, Y., Crynen, G., Southern, M. R., Biswas, A., Wang, Z. F., Tellinghuisen, T. L., & Disney, M. D. (2018). A massively parallel selection of small molecule-RNA motif binding partners informs design of an antiviral from sequence. *Chem*, *4*(10), 2384–2404.
- Chong, P. A., Vernon, R. M., & Forman-Kay, J. D. (2018). RGG/RG motif regions in RNA binding and phase separation. *Journal of Molecular Biology*, *430*(23), 4650–4665.
- Corley, M., Burns, M. C., & Yeo, G. W. (2020). How RNA-binding proteins interact with RNA: Molecules and mechanisms. *Molecular Cell*, *78*(1), 9–29.
- Couthouis, J., Hart, M. P., Erion, R., King, O. D., Diaz, Z., Nakaya, T., Ibrahim, F., Kim, H. J., Mojsilovic-Petrovic, J., Panossian, S., Kim, C. E., Frackelton, E. C., Solski, J. A., Williams, K. L., Clay-Falcone, D., Elman, L., McCluskey, L., Greene, R., Hakonarson, H., ... Gitler, A. D. (2012). Evaluating the role of the FUS/TLS-related gene EWSR1 in amyotrophic lateral sclerosis. *Human Molecular Genetics*, *21*(13), 2899–2911.
- Couthouis, J., Hart, M. P., Shorter, J., DeJesus-Hernandez, M., Erion, R., Oristano, R., Liu, A. X., Ramos, D., Jethava, N., Hosangadi, D., Epstein, J., Chiang, A., Diaz, Z., Nakaya, T., Ibrahim, F., Kim, H. J., Solski, J. A., Williams, K. L., Mojsilovic-Petrovic, J., ... Gitler, A. D. (2011). A yeast functional screen predicts new candidate ALS disease genes. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(52), 20881–20890.
- Crane-Robinson, C., Dragan, A. I., & Privalov, P. L. (2006). The extended arms of DNA-binding domains: a tale of tails. *Trends in Biochemical Sciences*, *31*(10), 547–552.
- Cuvertino, S., Hartill, V., Colyer, A., Garner, T., Nair, N., al-Gazali, L., Canham, N., Faundes, V., Flinter, F., Hertecant, J., Holder-Espinasse, M., Jackson, B., Lynch, S. A., Nadat, F., Narasimhan, V. M., Peckham, M., Sellers, R., Seri, M., Montanari, F., ... Banka, S. (2020). A restricted spectrum of missense KMT2D variants cause a multiple malformations disorder distinct from kabuki syndrome. *Genetics in Medicine*, *22*(5), 867–877.
- Daniels, N. J., Hershberger, C. E., Gu, X., Schueger, C., DiPasquale, W. M., Brick, J., Sauntharajah, Y., Maciejewski, J. P., & Padgett, R. A. (2021). Functional analyses of human LUC7-like proteins involved in splicing regulation and myeloid neoplasms. *Cell Reports*, *35*(2), 108989.
- Davidson, A., Leeper, T. C., Athanassiou, Z., Patora-Komisarska, K., Karn, J., Robinson, J. A., & Varani, G. (2009). Simultaneous recognition of HIV-1 TAR RNA bulge and loop sequences by cyclic peptide mimics of tat protein. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(29), 11931–11936.
- Davidson, A., Patora-Komisarska, K., Robinson, J. A., & Varani, G. (2011). Essential structural requirements for specific recognition of HIV TAR RNA by peptide mimetics of tat protein. *Nucleic Acids Research*, *39*(1), 248–256.
- Dawson, W. K., & Bujnicki, J. M. (2016). Computational modeling of RNA 3D structures and interactions. *Current Opinion in Structural Biology*, *37*, 22–28.

- Deng, H., Gao, K., & Jankovic, J. (2014). The role of FUS gene variants in neurodegenerative diseases. *Nature Reviews. Neurology*, *10*(6), 337–348.
- Dettori, L. G., Torrejon, D., Chakraborty, A., Dutta, A., Mohamed, M., Papp, C., Kuznetsov, V. A., Sung, P., Feng, W., & Bah, A. (2021). A tale of loops and tails: the role of intrinsically disordered protein regions in R-loop recognition and phase separation. *Frontiers in Molecular Biosciences*, *8*, 691694.
- Dominguez, D., Freese, P., Alexis, M. S., Su, A., Hochman, M., Palden, T., Bazile, C., Lambert, N. J., van Nostrand, E. L., Pratt, G. A., Yeo, G. W., Graveley, B. R., & Burge, C. B. (2018). Sequence, structure, and context preferences of human RNA binding proteins. *Molecular Cell*, *70*(5), 854–867.e9.
- D'Souza, V., & Summers, M. F. (2004). Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus. *Nature*, *431*(7008), 586–590.
- Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., Wang, L. P., Simmonett, A. C., Harrigan, M. P., Stern, C. D., Wiewiora, R. P., Brooks, B. R., & Pande, V. S. (2017). OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Computational Biology*, *13*(7), e1005659.
- Emmott, E., & Hiscox, J. A. (2009). Nucleolar targeting: The hub of the matter. *EMBO Reports*, *10*(3), 231–238.
- Faber, C., Schärpf, M., Becker, T., Sticht, H., & Rösch, P. (2001). The structure of the coliphage HK022 nun protein-lambda-phage boxB RNA complex. Implications for the mechanism of transcription termination. *The Journal of Biological Chemistry*, *276*(34), 32064–32070.
- Fay, M. M., Lyons, S. M., & Ivanov, P. (2017). RNA G-Quadruplexes in biology: Principles and molecular mechanisms. *Journal of Molecular Biology*, *429*(14), 2127–2147.
- Fischer, E. S., Scrima, A., Böhm, K., Matsumoto, S., Lingaraju, G. M., Faty, M., Yasuda, T., Cavadini, S., Wakasugi, M., Hanaoka, F., Iwai, S., Gut, H., Sugasawa, K., & Thomä, N. H. (2011). The molecular basis of CRL4DDB2/CSA ubiquitin ligase architecture, targeting, and activation. *Cell*, *147*(5), 1024–1039.
- Flores, S. C., & Altman, R. B. (2010). Turning limited experimental information into 3D models of RNA. *RNA*, *16*(9), 1769–1778.
- Frenkel-Pinter, M., Haynes, J. W., Mohyeldin, A. M., C, M., Sargon, A. B., Petrov, A. S., Krishnamurthy, R., Hud, N. V., Williams, L. D., & Leman, L. J. (2020). Mutually stabilizing interactions between proto-peptides and RNA. *Nature Communications*, *11*(1), 3137.
- Fulton, M. D., Brown, T., & Zheng, Y. G. (2019). The biological Axis of protein arginine methylation and asymmetric dimethylarginine. *International Journal of Molecular Sciences*, *20*(13), 3322. <https://doi.org/10.3390/ijms20133322>
- Gaertner, A., Klauke, B., Felski, E., Kassner, A., Brodehl, A., Gerdes, D., Stanasiuk, C., Ebbinghaus, H., Schulz, U., Dubowy, K. O., Tiesmeier, J., Laser, K. T., Bante, H., Bergau, L., Sommer, P., Fox, H., Morshuis, M., Gummert, J., & Milting, H. (2020). Cardiomyopathy-associated mutations in the RS domain affect nuclear localization of RBM20. *Human Mutation*, *41*(11), 1931–1943.
- Ge, P., Islam, S., Zhong, C., & Zhang, S. (2018). De novo discovery of structural motifs in RNA 3D structures through clustering. *Nucleic Acids Research*, *46*(9), 4783–4793.
- Gebauer, F., Schwarzl, T., Valcárcel, J., & Hentze, M. W. (2021). RNA-binding proteins in human genetic disease. *Nature Reviews. Genetics*, *22*(3), 185–198.
- Glisovic, T., Bachorik, J. L., Yong, J., & Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. *FEBS Letters*, *582*(14), 1977–1986.
- Guogas, L. M., Filman, D. J., Hogle, J. M., & Gehrke, L. (2004). Cofolding organizes alfalfa mosaic virus RNA and coat protein for replication. *Science*, *306*(5704), 2108–2111.
- Gupta, A., & Gribskov, M. (2011). The role of RNA sequence and structure in RNA–protein interactions. *Journal of Molecular Biology*, *409*(4), 574–587.
- Habchi, J., Tompa, P., Longhi, S., & Uversky, V. N. (2014). Introducing protein intrinsic disorder. *Chemical Reviews*, *114*(13), 6561–6588.
- Hargrove, A. E. (2020). Small molecule-RNA targeting: Starting with the fundamentals. *Chemical Communications*, *56*(94), 14744–14756.
- Harrison, A. F., & Shorter, J. (2017). RNA-binding proteins with prion-like domains in health and disease. *Biochemical Journal*, *474*(8), 1417–1438.
- Hayashi, T., Oshima, H., Mashima, T., Nagata, T., Katahira, M., & Kinoshita, M. (2014). Binding of an RNA aptamer and a partial peptide of a prion protein: Crucial importance of water entropy in molecular recognition. *Nucleic Acids Research*, *42*(11), 6861–6875.
- Heddi, B., Cheong, V. V., Martadinata, H., & Phan, A. T. (2015). Insights into G-quadruplex specific recognition by the DEAH-box helicase RHAU: Solution structure of a peptide-quadruplex complex. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(31), 9608–9613.
- Heddi, B., Cheong, V. V., Schmitt, E., Mechulam, Y., & Phan, A. T. (2020). Recognition of different base tetrads by RHAU (DHX36): X-ray crystal structure of the G4 recognition motif bound to the 3'-end tetrad of a DNA G-quadruplex. *Journal of Structural Biology*, *209*(1), 107399.
- Hermann, T. (2003). Chemical and functional diversity of small molecule ligands for RNA. *Biopolymers*, *70*(1), 4–18.
- Houser-Scott, F., Baer, M. L., Liem, K. F., Jr., Cai, J. M., & Gehrke, L. (1994). Nucleotide sequence and structural determinants of specific binding of coat protein or coat protein peptides to the 3' untranslated region of alfalfa mosaic virus RNA 4. *Journal of Virology*, *68*(4), 2194–2205.
- Huber, E. M., Scharf, D. H., Hortschansky, P., Groll, M., & Brakhage, A. A. (2012). DNA minor groove sensing and widening by the CCAAT-binding complex. *Structure*, *20*(10), 1757–1768.
- Imperatore, J. A., McAninch, D. S., Valdez-Sinon, A. N., Bassell, G. J., & Mihailescu, M. R. (2020). FUS recognizes G Quadruplex structures within neuronal mRNAs. *Frontiers in Molecular Biosciences*, *7*, 6.

- Izadi, S., Anandakrishnan, R., & Onufriev, A. V. (2014). Building water models: A different approach. *Journal of Physical Chemistry Letters*, 5(21), 3863–3871.
- Järvelin, A. I., Noerenberg, M., Davis, I., & Castello, A. (2016). The new (dis)order in RNA regulation. *Cell Communication and Signaling*, 14, 9.
- Jiang, F., Gorin, A., Hu, W., Majumdar, A., Baskerville, S., Xu, W., Ellington, A., & Patel, D. J. (1999). Anchoring an extended HTLV-1 Rex peptide within an RNA major groove containing junctional base triples. *Structure*, 7(12), 1461–1472.
- Kanhoush, R., Beenders, B., Perrin, C., Moreau, J., Bellini, M., & Penrad-Mobayed, M. (2010). Novel domains in the hnRNP G/RBMX protein with distinct roles in RNA binding and targeting nascent transcripts. *Nucleus*, 1(1), 109–122.
- Kelly, M. L., Chu, C. C., Shi, H., Ganser, L. R., Bogerd, H. P., Huynh, K., Hou, Y., Cullen, B. R., & al-Hashimi, H. M. (2021). Understanding the characteristics of nonspecific binding of drug-like compounds to canonical stem-loop RNAs and their implications for functional cellular assays. *RNA*, 27(1), 12–26.
- Kilchert, C., Sträßer, K., Kunetsky, V., & Änkö, M. L. (2020). From parts lists to functional significance-RNA-protein interactions in gene regulation. *WIREs RNA*, 11(3), e1582.
- Kim, E., Ilagan, J. O., Liang, Y., Daubner, G. M., Lee, S. C. W., Ramakrishnan, A., Li, Y., Chung, Y. R., Micol, J. B., Murphy, M. E., Cho, H., Kim, M. K., Zebari, A. S., Aumann, S., Park, C. Y., Buonamici, S., Smith, P. G., Deeg, H. J., Lobry, C., ... Abdel-Wahab, O. (2015). SRSF2 mutations contribute to Myelodysplasia by mutant-specific effects on exon recognition. *Cancer Cell*, 27(5), 617–630.
- Kim, H. J., Kim, N. C., Wang, Y. D., Scarborough, E. A., Moore, J., Diaz, Z., MacLea, K. S., Freibaum, B., Li, S., Molliex, A., Kanagaraj, A. P., Carter, R., Boylan, K. B., Wojtas, A. M., Rademakers, R., Pinkus, J. L., Greenberg, S. A., Trojanowski, J. Q., Traynor, B. J., ... Taylor, J. P. (2013). Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature*, 495(7442), 467–473.
- Kolev, S. K., Petkov, P. S., Rangelov, M. A., Trifonov, D. V., Milenov, T. I., & Vayssilov, G. N. (2018). Interaction of Na, K, Mg and Ca counter cations with RNA. *Metallomics: Integrated Biometal Science*, 10(5), 659–678.
- Krüger, D. M., Neubacher, S., & Grossmann, T. N. (2018). Protein-RNA interactions: Structural characteristics and hotspot amino acids. *RNA*, 24(11), 1457–1465.
- Kwok, C. K. (2016). Dawn of the in vivo RNA structure and interactome. *Biochemical Society Transactions*, 44(5), 1395–1410.
- Lang, B., Yang, J. S., Garriga-Canut, M., Speroni, S., Aschern, M., Gili, M., Hoffmann, T., Tartaglia, G. G., & Maurer, S. P. (2021). Matrix-screening reveals a vast potential for direct protein-protein interactions among RNA binding proteins. *Nucleic Acids Research*, 49(12), 6702–6721.
- Leclerc, F., Cedergren, R., & Ellington, A. D. (1994). A three-dimensional model of the rev-binding element of HIV-1 derived from analyses of aptamers. *Nature Structural Biology*, 1(5), 293–300.
- Leeper, T. C., Athanassiou, Z., Dias, R. L. A., Robinson, J. A., & Varani, G. (2005). TAR RNA recognition by a cyclic peptidomimetic of tat protein. *Biochemistry*, 44(37), 12362–12372.
- Legault, P., Li, J., Mogridge, J., Kay, L. E., & Greenblatt, J. (1998). NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: Recognition of a GNRA fold by an arginine-rich motif. *Cell*, 93(2), 289–299.
- Li, B., Cao, Y., Westhof, E., & Miao, Z. (2020). Advances in RNA 3D structure modeling using experimental data. *Frontiers in Genetics*, 11, 574485.
- Li, K. K. C., Chau, B. L., & Lee, K. A. W. (2018). Differential interaction of PRMT1 with RGG-boxes of the FET family proteins EWS and TAF15. *Protein Science*, 27(3), 633–642.
- Li, Z., Song, L. F., Li, P., & Merz, K. M., Jr. (2020). Systematic Parametrization of divalent metal ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB water models. *Journal of Chemical Theory and Computation*, 16(7), 4429–4442.
- Lin, S., & Fu, X.-D. (2007). SR proteins and related factors in alternative splicing. *Advances in Experimental Medicine and Biology*, 623, 107–122.
- Lin, Y., Currie, S. L., & Rosen, M. K. (2017). Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. *The Journal of Biological Chemistry*, 292(46), 19110–19120.
- Liu, S., Hong, X., Shen, C., Shi, Q., Wang, J., Xiong, F., & Qiu, Z. (2015). Kabuki syndrome: A Chinese case series and systematic review of the spectrum of mutations. *BMC Medical Genetics*, 16, 26.
- Loughlin, F. E., Lukavsky, P. J., Kazeeva, T., Reber, S., Hock, E. M., Colombo, M., van Schroetter, C., Pauli, P., Cléry, A., Mühlemann, O., Polymenidou, M., Ruepp, M. D., & Allain, F. H. T. (2019). The solution structure of FUS bound to RNA reveals a bipartite mode of RNA recognition with both sequence and shape specificity. *Molecular Cell*, 73(3), 490–504.e6.
- Lu, Z., Guo, J. K., Wei, Y., Dou, D. R., Zarnegar, B., Ma, Q., Li, R., Zhao, Y., Liu, F., Choudhry, H., Khavari, P. A., & Chang, H. Y. (2020). Structural modularity of the XIST ribonucleoprotein complex. *Nature Communications*, 11(1), 6163.
- Lunde, B. M., Moore, C., & Varani, G. (2007). RNA-binding proteins: Modular design for efficient function. *Nature Reviews. Molecular Cell Biology*, 8(6), 479–490.
- Masuzawa, T., & Oyoshi, T. (2020). Roles of the RGG domain and RNA recognition motif of Nucleolin in G-Quadruplex stabilization. *ACS Omega*, 5(10), 5202–5208.
- McGibbon, R. T., Beauchamp, K. A., Harrigan, M. P., Klein, C., Swails, J. M., Hernández, C. X., Schwantes, C. R., Wang, L. P., Lane, T. J., & Pande, V. S. (2015). MDTraj: A modern open library for the analysis of molecular dynamics trajectories. *Biophysical Journal*, 109(8), 1528–1532.
- Mersaoui, S. Y., Yu, Z., Coulombe, Y., Karam, M., Busatto, F. F., Masson, J. Y., & Richard, S. (2019). Arginine methylation of the DDX5 helix-case RGG/RG motif by PRMT5 regulates resolution of RNA:DNA hybrids. *The EMBO Journal*, 38(15), e100986.

- Miller, M. (2009). The importance of being flexible: The case of basic region leucine zipper transcriptional regulators. *Current Protein & Peptide Science*, 10(3), 244–269.
- Mínguez-Toral, M., Cuevas-Zuñiga, B., Garrido-Arandia, M., & Pacios, L. F. (2020). A computational structural study on the DNA-protecting role of the tardigrade-unique Dsup protein. *Scientific Reports*, 10(1), 13424.
- Mol, C. D., Izumi, T., Mitra, S., & Tainer, J. A. (2000). DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination [corrected]. *Nature*, 403(6768), 451–456.
- Nguyen, H., Roe, D. R., & Simmerling, C. (2013). Improved generalized born solvent model parameters for protein simulations. *Journal of Chemical Theory and Computation*, 9(4), 2020–2034.
- Nikulin, A. D. (2021). Characteristic features of protein interaction with single- and double-stranded RNA. *Biochemistry*, 86(8), 1025–1040.
- Nott, T. J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowitz, A., Craggs, T. D., Bazett-Jones, D. P., Pawson, T., Forman-Kay, J. D., & Baldwin, A. J. (2015). Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Molecular Cell*, 57(5), 936–947.
- Ozdilek, B. A., Thompson, V. F., Ahmed, N. S., White, C. I., Batey, R. T., & Schwartz, J. C. (2017). Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding. *Nucleic Acids Research*, 45(13), 7984–7996.
- Peterson, R. D., Bartel, D. P., Szostak, J. W., Horvath, S. J., & Feigon, J. (1994). 1H NMR studies of the high-affinity rev binding site of the rev responsive element of HIV-1 mRNA: Base pairing in the core binding element. *Biochemistry*, 33(18), 5357–5366.
- Phan, A. T., Kuryavii, V., Darnell, J. C., Serganov, A., Majumdar, A., Ilin, S., Raslin, T., Polonskaia, A., Chen, C., Clain, D., Darnell, R. B., & Patel, D. J. (2011). Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nature Structural & Molecular Biology*, 18(7), 796–804.
- Prasad, A., Bharathi, V., Sivalingam, V., Girdhar, A., & Patel, B. K. (2019). Molecular mechanisms of TDP-43 Misfolding and pathology in amyotrophic lateral sclerosis. *Frontiers in Molecular Neuroscience*, 12, 25.
- Protter, D. S. W., Rao, B. S., van Treeck, B., Lin, Y., Mizoue, L., Rosen, M. K., & Parker, R. (2018). Intrinsically disordered regions can contribute promiscuous interactions to RNP granule assembly. *Cell Reports*, 22(6), 1401–1412.
- Puglisi, J. D., Chen, L., Blanchard, S., & Frankel, A. D. (1995). Solution structure of a bovine immunodeficiency virus tat-TAR peptide-RNA complex. *Science*, 270(5239), 1200–1203.
- Rohs, R., West, S. M., Sosinsky, A., Liu, P., Mann, R. S., & Honig, B. (2009). The role of DNA shape in protein-DNA recognition. *Nature*, 461(7268), 1248–1253.
- Sanchez de Groot, N., Armaos, A., Graña-Montes, R., Alriquet, M., Calloni, G., Vabulas, R. M., & Tartaglia, G. G. (2019). RNA structure drives interaction with proteins. *Nature Communications*, 10(1), 3246.
- Schärf, M., Sticht, H., Schweimer, K., Boehm, M., Hoffmann, S., & Rösch, P. (2000). Antitermination in bacteriophage lambda. The structure of the N36 peptide-boxB RNA complex. *European Journal of Biochemistry*, 267(8), 2397–2408.
- Schneider, R., Blackledge, M., & Jensen, M. R. (2019). Elucidating binding mechanisms and dynamics of intrinsically disordered protein complexes using NMR spectroscopy. *Current Opinion in Structural Biology*, 54, 10–18.
- Sengupta, A., Li, Z., Song, L. F., Li, P., & Merz, K. M., Jr. (2021). Parameterization of monovalent ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB water models. *Journal of Chemical Information and Modeling*, 61(2), 869–880.
- Shortridge, M. D., Wille, P. T., Jones, A. N., Davidson, A., Bogdanovic, J., Arts, E., Karn, J., Robinson, J. A., & Varani, G. (2019). An ultra-high affinity ligand of HIV-1 TAR reveals the RNA structure recognized by P-TEFb. *Nucleic Acids Research*, 47(3), 1523–1531.
- Skrisovska, L., Bourgeois, C. F., Stefl, R., Grellscheid, S. N., Kister, L., Wenter, P., Elliott, D. J., Stevenin, J., & Allain, F. H. T. (2007). The testis-specific human protein RBMY recognizes RNA through a novel mode of interaction. *EMBO Reports*, 8(4), 372–379.
- Susbielle, G., Blattes, R., Brevet, V., Monod, C., & Kas, E. (2005). Target practice: Aiming at satellite repeats with DNA minor groove binders. *Current Medicinal Chemistry. Anti-Cancer Agents*, 5(4), 409–420.
- Szabó, B., Murvai, N., Abukhairan, R., Schád, É., Kardos, J., Szeder, B., Buday, L., & Tantos, Á. (2018). Disordered regions of mixed lineage leukemia 4 (MLL4) protein are capable of RNA binding. *International Journal of Molecular Sciences*, 19(11), 3478. <https://doi.org/10.3390/ijms19113478>
- Tanaka, Y., Fujii, S., Hiroaki, H., Sakata, T., Tanaka, T., Uesugi, S., Tomita, K., & Kyogoku, Y. (1999). A'-form RNA double helix in the single crystal structure of r(UGAGCUUCGGCUC). *Nucleic Acids Research*, 27(4), 949–955.
- Tian, C., Kasavajhala, K., Belfon, K. A. A., Raguette, L., Huang, H., Miguez, A. N., Bickel, J., Wang, Y., Pincay, J., Wu, Q., & Simmerling, C. (2020). ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *Journal of Chemical Theory and Computation*, 16(1), 528–552.
- Tompa, P. (2003). The functional benefits of protein disorder. *Journal of Molecular Structure: THEOCHEM*, 666–667, 361–371.
- Tompa, P., Schad, E., Tantos, A., & Kalmar, L. (2015). Intrinsically disordered proteins: Emerging interaction specialists. *Current Opinion in Structural Biology*, 35, 49–59.
- Tsai, W.-C., Gayatri, S., Reineke, L. C., Sbardella, G., Bedford, M. T., & Lloyd, R. E. (2016). Arginine Demethylation of G3BP1 promotes stress granule assembly. *The Journal of Biological Chemistry*, 291(43), 22671–22685.
- Tunnicliffe, R. B., Lockhart-Cairns, M. P., Levy, C., Mould, A. P., Jowitt, T. A., Sito, H., Baldock, C., Sandri-Goldin, R. M., & Golovanov, A. P. (2017). The herpes viral transcription factor ICP4 forms a novel DNA recognition complex. *Nucleic Acids Research*, 45(13), 8064–8078.
- Van Treeck, B., & Parker, R. (2018). Emerging roles for intermolecular RNA-RNA interactions in RNP assemblies. *Cell*, 174(4), 791–802.

- Vasilyev, N., Polonskaia, A., Darnell, J. C., Darnell, R. B., Patel, D. J., & Serganov, A. (2015). Crystal structure reveals specific recognition of a G-quadruplex RNA by a β -turn in the RGG motif of FMRP. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(39), E5391–E5400.
- Vinson, C., Acharya, A., & Taparowsky, E. J. (2006). Deciphering B-ZIP transcription factor interactions in vitro and in vivo. *Biochimica et Biophysica Acta*, *1759*(1–2), 4–12.
- Warner, K. D., Hajdin, C. E., & Weeks, K. M. (2018). Principles for targeting RNA with drug-like small molecules. *Nature Reviews Drug Discovery*, *17*(8), 547–558.
- Wilson, K. A., Holland, D. J., & Wetmore, S. D. (2016). Topology of RNA-protein nucleobase-amino acid π - π interactions and comparison to analogous DNA-protein π - π contacts. *RNA*, *22*(5), 696–708.
- Xiang, S., Gapsys, V., Kim, H. Y., Bessonov, S., Hsiao, H. H., Möhlmann, S., Klaukien, V., Ficner, R., Becker, S., Urlaub, H., Lührmann, R., de Groot, B., & Zweckstetter, M. (2013). Phosphorylation drives a dynamic switch in serine/arginine-rich proteins. *Structure*, *21*(12), 2162–2174.
- Yadav, D. K., Zigáčková, D., Zlobina, M., Klumpler, T., Beaumont, C., Kubičková, M., Vaňáčková, Š., & Lukavsky, P. J. (2020). Staufen1 reads out structure and sequence features in ARF1 dsRNA for target recognition. *Nucleic Acids Research*, *48*(4), 2091–2106.
- Yang, J., Gao, M., Xiong, J., Su, Z., & Huang, Y. (2019). Features of molecular recognition of intrinsically disordered proteins via coupled folding and binding. *Protein Science*, *28*(11), 1952–1965.
- Ye, X., Gorin, A., Ellington, A. D., & Patel, D. J. (1996). Deep penetration of an alpha-helix into a widened RNA major groove in the HIV-1 rev peptide-RNA aptamer complex. *Nature Structural Biology*, *3*(12), 1026–1033.
- Ye, X., Gorin, A., Frederick, R., Hu, W., Majumdar, A., Xu, W., McLendon, G., Ellington, A., & Patel, D. J. (1999). RNA architecture dictates the conformations of a bound peptide. *Chemistry & Biology*, *6*(9), 657–669.
- Ye, X., Kumar, R. A., & Patel, D. J. (1995). Molecular recognition in the bovine immunodeficiency virus Tat peptide-TAR RNA complex. *Chemistry & Biology*, *2*(12), 827–840.
- Zhang, Q., Harada, K., Cho, H. S., Frankel, A. D., & Wemmer, D. E. (2001). Structural characterization of the complex of the Rev response element RNA with a selected peptide. *Chemistry & Biology*, *8*(5), 511–520.
- Zhang, Y., Xie, S., Xu, H., & Qu, L. (2015). CLIP: Viewing the RNA world from an RNA-protein interactome perspective. *Science China Life Sciences*, *58*(1), 75–88.

How to cite this article: Zeke, A., Schád, É., Horváth, T., Abukhairan, R., Szabó, B., & Tantos, A. (2022). Deep structural insights into RNA-binding disordered protein regions. *Wiley Interdisciplinary Reviews: RNA*, *13*(5), e1714. <https://doi.org/10.1002/wrna.1714>

APPENDIX

Description of the molecular dynamics simulations

Overview and general procedures

The simulations were performed using the OpenMM package (Eastman et al., 2017), with an in-house python code. The input parameter files were prepared with Ambertools (Case et al., 2021). FF19SB forcefield (Tian et al., 2020) with OPC water (Izadi et al., 2014) and 12–6–4 ion parameters. The integrator used was Langevin Middle Integrator with 7.0/ps collision rate, 5 fs timestep, all-bond constraints, 1.2 nm nonbonded cutoff distance, 1.0 nm switch distance. All simulations were of constant volume. In all explicit solvent simulations the following ion concentrations were used: 10.0 mM Na⁺, 137.0 mM K⁺, 25.0 mM Mg²⁺, 197.2 mM Cl⁻, ensuring that at least 1 Mg²⁺ ion was present in the water box, and that the total charge was 0 (Li, Song, et al., 2020; Sengupta et al., 2021). The solvent and ions were equilibrated for 20 ns around the fixed macromolecules. In all cases where replica-exchange was run in explicit water, the exchange probabilities were calculated based on the same structure in GBNeck2 (Nguyen et al., 2013) implicit solvent.

RBMX (RGG) structure modeling

The preliminary RNA structure for the WEC core was folded with MMB (Flores & Altman, 2010) based on the predicted 2D structure by Kannush et al. (Kanhoush et al., 2010), and a preliminary protein structure of RBMX_C was generated with MMB. Both of them were then processed with a replica-exchange simulation. The RNA in OPC explicit water, the protein in GBNeck2 implicit solvent, 12 temperatures between 265–375 K, for 50 ns. The structures were clustered with MDtraj (McGibbon et al., 2015) and scipy's hierarchical clustering linkage, with average distances, and further processed with dendrogram. In all cases, the default settings were used, except when noted. The centroid structures were then extracted from the clusters. A representative structure from both were chosen for docking.

R/E/S-rich structure modeling

The RNA structure was taken from the pdb 1ULL structure, and a default helix generated for LC7L3-pep1. The prepared RNA and protein structures were docked using ClusPro 2. These were simulated using replica exchange, with 12 temperatures between 275 and 345 K, for 50 ns. The structures from the middle four temperature boxes (298.6, 304.9, 311.2, and 317.7 K) were clustered, and centroids extracted the same way as before.