

How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms

Meredith Corley,¹ Margaret C. Burns,^{1,2} and Gene W. Yeo^{1,2,3,*}

¹Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA, USA

²Biomedical Sciences Graduate Program, University of California, San Diego, La Jolla, CA, USA

³Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA, USA

*Correspondence: geneyeo@ucsd.edu

<https://doi.org/10.1016/j.molcel.2020.03.011>

RNA-binding proteins (RBPs) comprise a large class of over 2,000 proteins that interact with transcripts in all manner of RNA-driven processes. The structures and mechanisms that RBPs use to bind and regulate RNA are incredibly diverse. In this review, we take a look at the components of protein-RNA interaction, from the molecular level to multi-component interaction. We first summarize what is known about protein-RNA molecular interactions based on analyses of solved structures. We additionally describe software currently available for predicting protein-RNA interaction and other resources useful for the study of RBPs. We then review the structure and function of seventeen known RNA-binding domains and analyze the hydrogen bonds adopted by protein-RNA structures on a domain-by-domain basis. We conclude with a summary of the higher-level mechanisms that regulate protein-RNA interactions.

RNA-binding proteins (RBPs) potently and ubiquitously regulate transcripts throughout their life cycle (Lorkovic, 2012). RBP interactions with RNA range from single-protein-RNA element interaction to the assembly of multiple RBPs and RNA molecules such as the spliceosome. How RBPs selectively bind their targets is not always understood, although there are currently many techniques used to study these interactions. X-ray crystallography and nuclear magnetic resonance (NMR) experiments facilitate precise study of the amino acids and nucleotides that interact in protein-RNA complexes, and numerous such datasets have been generated for RBP domains in complex with RNA (Berman et al., 2000). Analyses of these data have inferred the number and types of intermolecular interactions and preferred amino acids that characterize specific protein-RNA binding (Han and Nepal, 2007; Pérez-Cano and Fernández-Recio, 2010). Furthermore, numerous studies have built on protein-RNA structural data to develop increasingly accurate software that predicts which residues in proteins interact with RNA. We include a description of up-to-date software and data resources for the purpose of predicting and studying how RBPs interact with RNA.

RNA-binding domains in protein are the functional units responsible for binding RNA. Multiple such domains often occur in a single RBP and these modular arrangements can coordinate and enhance binding to RNA (Cléry and Allain, 2012; Lunde et al., 2007). Additionally, RBPs tend to be enriched in intrinsically disordered regions, which themselves act as RNA-binding domains but limit the structural study of RBPs to ordered domains rather than full-length protein (Järvelin et al., 2016). Several ordered domains have been studied for decades, although it is important to note that RNA-binding domains are remarkably heterogeneous and can be difficult to classify (Gerstberger et al., 2014). Additionally, many domains remain to be characterized, where hundreds of RBPs lack known RNA-binding domains (Castello et al., 2016). Here we overview the strategies that

seventeen well-characterized RNA-binding domains use to achieve RNA binding. Furthermore, we present an analysis of the preferences in protein-RNA hydrogen bonds for eight of these domain types.

RBP binding ultimately achieves a range of cellular goals (Gerstberger et al., 2014; Glisovic et al., 2008), but many mechanisms—and many chances for regulation—lie in between binding and biological consequence. These mechanisms we categorize into several layers: protein-RNA assembly, combined action of the ribonucleoprotein (RNP), and modifications and interactions that regulate the previous two (Lovci et al., 2016; Lunde et al., 2007; Thapar, 2015). Here we describe these high-level processes and provide functional examples (Fiorini et al., 2015; Jackson et al., 2010; Śledź and Jinek, 2016), all of which were discovered by intense and detailed biochemical work, including by insights from protein-RNA structures. These summaries intersect the areas of study that enable a mechanistic understanding of RBP regulation and we hope serve as a useful and timely resource.

Protein-RNA Molecular Interaction

To understand RBP regulation of RNA targets, one must understand the biochemical underpinnings that facilitate exact and specific interaction with these sites. RBPs bind their RNA targets through the molecular interactions of chemical moieties between protein residues and RNA nucleotides. At this resolution the distinction between RNA and protein begins to blend, as the same intermolecular forces that shape protein and RNA tertiary structures also stitch the two molecules together. These interactions occur dynamically, with sometimes quite large rearrangements in RNA and protein (Hainzl et al., 2005; Leulliot and Varani, 2001). In this section, we will provide a detailed description of the molecular interactions that occur in protein-RNA structures and overview trends determined by previous research. We will also catalog software that uses molecular-level interaction data to predict protein-RNA binding.



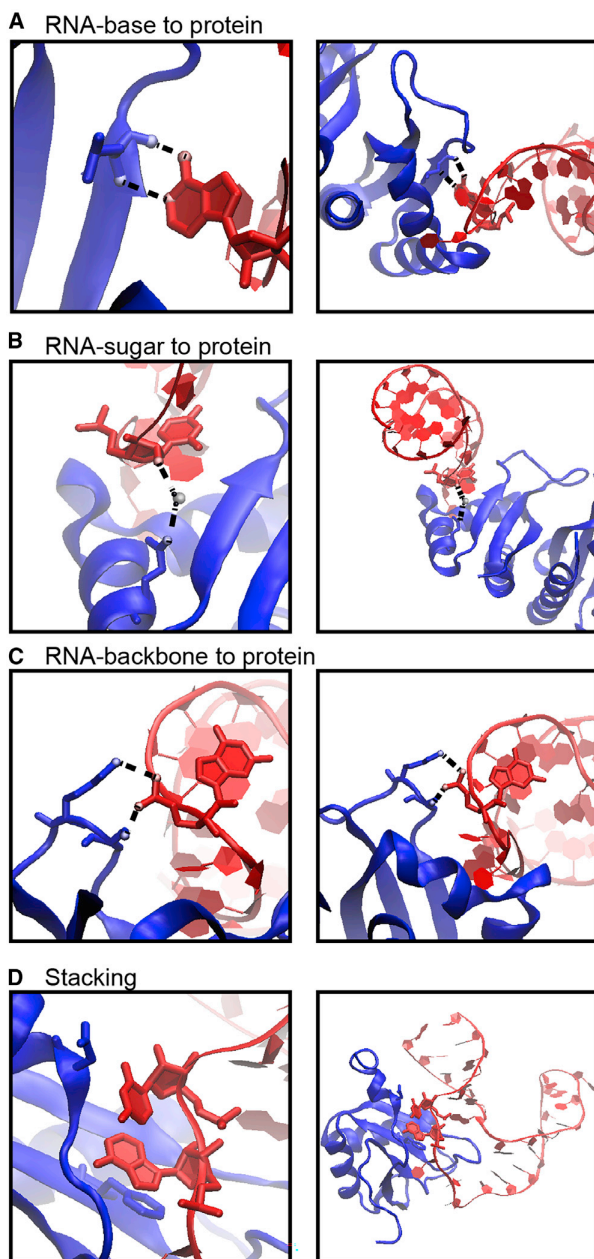


Figure 1. Examples of Protein-RNA Hydrogen Bonds and Stacking Interactions

The KH domain of human NOVA1 (PDB: 2ANR) (Teplova et al., 2011) and human U1A (PDB: 1AUD) (Oubridge et al., 1994) visualized with VMD (Humphrey et al., 1996) in detailed (left) and zoomed-out (right) perspectives. RNA is in red, and protein is in blue.

- (A) Main-chain atoms of a Leu form hydrogen bonds with adenine.
 (B) Hydrogen bonds form between Gln and the 2' OH of a cytosine, bridged by a water molecule.
 (C) Two hydrogen bonds form between the phosphate backbone atoms of guanine and Ser and Lys.
 (D) An adenine and cytosine in an unpaired loop stack between Asp and Phe.

Hydrogen Bonds and Van der Waals Interactions

Hydrogen bonds and Van der Waals (VdW) interactions have been extensively analyzed in protein-RNA interactions (Gupta and Gribskov, 2011; Han and Nepal, 2007; Hoffman et al., 2004; Pérez-Cano and Fernández-Recio, 2010; Treger and Westhof, 2001). Hydrogen bonds form between an electronegative atom bound to a hydrogen atom, whose partial positive charge attracts an electronegative partner. Hydrogen bonds can be formed by both neutral and ionic groups, and can be coordinated by water molecules (Figure 1B). They generally form at distances of 2.4–3.0 Å, contributing 0.5–4.5 kcal/mol per bond (Auweter et al., 2006). The weakest hydrogen bonds are considered to be VdW interactions, which are weak (0.5–1 kcal/mol) electrostatic interactions that occur above ~3.0 Å. All the studies that analyze VdW interactions and hydrogen bonds in protein-RNA structures identify hydrogen bonds with HBPLUS (McDonald and Thornton, 1994) and identify VdW interactions as the hydrogen bonds above a threshold donor-acceptor distance (Allers and Shamoo, 2001; Ellis et al., 2007; Han and Nepal, 2007; Hu et al., 2018; Jones et al., 2001; Morozova et al., 2006; Treger and Westhof, 2001). All RNA bases, the 2' OH, and the phosphodiester backbone can form hydrogen bonds and VdW interactions with protein (Figures 1A–1C) (Teplova et al., 2011). Multiple analyses of hydrogen-bond types in protein-RNA structures have found that hydrogen bonds with base, 2' OH (sugar), and phosphate (RNA backbone) account for an average of 35.5%, 23.5%, and 41% of protein-RNA hydrogen bonds, respectively (Figure 2A) (Gupta and Gribskov, 2011; Han and Nepal, 2007; Hoffman et al., 2004; Treger and Westhof, 2001). Studies of VdW percentages with base, sugar, and phosphate are more variable (Figure 2B), perhaps reflecting inconsistent thresholds in categorizing VdW interactions.

Proteins can interact with RNA using the main chain of any residue and the side chains of most residues. Studies have consistently found that the protein side chain, versus the main chain, is employed in 71.5% of hydrogen bonds and 76% of VdW interactions with RNA (Figures 2A and 2B). Polar amino acids Ser and Asn and positively charged amino acids Lys and Arg, which form strong ionic hydrogen bonds (salt bridges), predominate these interactions (Gupta and Gribskov, 2011; Han and Nepal, 2007; Hoffman et al., 2004; Pérez-Cano and Fernández-Recio, 2010; Treger and Westhof, 2001). VdW interactions generally share the same preferences for amino acids that are observed for hydrogen bonds (Ellis et al., 2007; Han and Nepal, 2007; Jones et al., 2001; Treger and Westhof, 2001). In the overall set of interactions that occur at a protein-RNA interface, VdW interactions are thought to predominate, although estimates of the ratio of VdW-to-hydrogen-bond interactions per protein-RNA complex vary quite a bit (Figure 2C).

Hydrophobic and π Interactions and Stacking

Hydrophobic interactions occur at distances of 3.8–5.0 Å (Morozova et al., 2006; Onofrio et al., 2014) and contribute 1–2 kcal/mol per interaction (Dill et al., 2008). Hydrophobic interactions between RNA bases and hydrophobic side chains can be important stabilizing factors at protein-RNA interfaces by sequestering hydrophobic residues and bases from solvent to form a “hydrophobic core” (Akopian et al., 2013; Allain et al., 1997; Yang et al., 2002; Yu et al., 2014). For example, the SRP54 “M”-binding

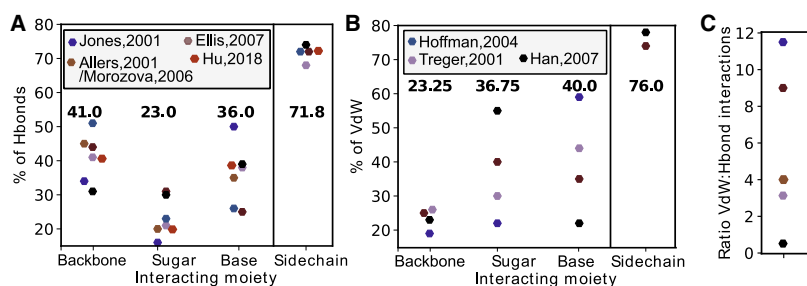


Figure 2. Meta-analysis of Seven Studies Analyzing Hydrogen Bonds and Van der Waals Interactions in Protein-RNA Structures

(A) Reports across studies of the percent of hydrogen bonds in protein-RNA structures that occur with the RNA backbone (phosphate), sugar (2' OH), or base. The percent of hydrogen bonds that occur with the protein side chain (as opposed to the main chain). Averages are shown above each category.

(B) Reports of the percent of VdW interactions in protein-RNA structures that occur with the RNA backbone (phosphate), sugar (2' OH), or base. The

percent of VdW interactions that occur with the protein side chain (as opposed to the main chain). Averages are shown above each category. (C) Reports across studies of the average ratio of VdW interactions to hydrogen bonds per protein-RNA structure.

domain forms a methionine-rich hydrophobic surface with SRP RNA (Akopian et al., 2013). Hydrophobic interactions have been surveyed more sparsely in protein-RNA structures, but may account for up to 50% of the interactions at the protein-RNA interface, depending on the RBP (Hu et al., 2018).

π interactions can form between any nitrogenous base ring and a π -containing amino acid, which include the aromatic residues Trp, His, Phe, and Tyr as well as the charged residues Arg, Glu, and Asp (Wilson et al., 2016). These interactions are relatively strong at ~ 2 – 6 kcal/mol per interaction (Brylinski, 2018; Wilson et al., 2016), and often prefer to be stacked (referred to as π stacking), occurring most frequently with an inter-atom distance of 2.7–4.3 Å (Auweter et al., 2006; Brylinski, 2018; Morozova et al., 2006; Wilson et al., 2016). Analyses of π interactions occurring in protein-RNA crystal structures find multiple such interactions on average per structure (Hu et al., 2018; Wilson et al., 2016). These interactions can contribute considerable stability to protein-RNA binding, where some π interactions are demonstrably crucial to binding function (Auweter et al., 2006; Liao et al., 2018; Oubridge et al., 1994). Such is the case with the extensive stacking interactions cementing human U1A spliceosomal protein with an RNA polyadenylation inhibition element, including two consecutive bases sandwiched between Phe and Asp residues (Figure 1D) (Oubridge et al., 1994). Stacking interactions also occur between bases and hydrophobic residues, and these may be mixed with π - π stacking interactions. For example, a single cytosine in a bulge in bacterial 4.5S SRP RNA stacks with both Phe and Leu in FtsY (Bifsha et al., 2007). More exotic π -stacking configurations include bases that stack on the protein main chain between residues (Auweter et al., 2006) and perpendicular “T stacks” between protein side chains and bases. Stacking interactions with RNA are overall quite crucial and varied, possibly occurring at higher rates than in protein-DNA interactions (Wilson et al., 2016).

Differences from Protein-DNA Interactions

Similar analyses of protein-DNA structures allow comparison with protein-RNA interactions (Jones et al., 2001; Luscombe et al., 2001; Wilson et al., 2016). RBPs and DNA-binding proteins show many of the same preferences for interacting residues, that is, positively charged and polar residues (Hoffman et al., 2004; Jones et al., 2001). However, the chemical and structural differences between DNA and RNA molecules result in observable differences in interactions. Approximately 20% of protein interactions with RNA occur with the 2' OH, whereas this is not

available for protein-DNA interactions (Hoffman et al., 2004) (Figure 2). The base-pairing moieties in RNA bases are also much more extensively contacted by protein than in DNA as the Watson-Crick base face is normally base paired in DNA (Allers and Shamoo, 2001; Luscombe et al., 2001). In this same vein, protein-DNA interactions more frequently use the phosphodiester backbone (Hoffman et al., 2004; Jones et al., 2001). DNA-binding proteins tend to surround their target DNA helix, but this mode of binding is not always available to RBPs, which must accommodate a diverse range of stem loops, bulges, and other complex structures (Jones et al., 2001). Additionally, double-stranded RNA (dsRNA) adopts a different helix from the standard DNA B form helix (Bercy and Bockelmann, 2015), explaining why RBPs that interact with dsRNA are best suited to the RNA helix (Vuković et al., 2014).

Despite these overall differences, numerous proteins bind both DNA and RNA (Hudson and Ortlund, 2014). A canonical example is the CCHH zinc-finger (ZnF) protein TFIIIA, which binds both the 5S rRNA gene and 5S rRNA with at least six tandem ZnF domains (Hall, 2005). In binding 5S DNA, TFIIIA ZnFs 1, 3, and 5 interact with the major groove whereas ZnFs 4 and 6 serve as spacers. When binding RNA, ZnF 4 and 6 interact with unpaired 5S rRNA bases and ZnF 5 binds the RNA major groove, albeit by a different mode from how it contacts the DNA major groove. Similarly, human ADAR1 can bind both dsRNA and Z-form DNA, but uses separate domains for each (Barraud and Allain, 2012). Thus, even a dual DNA- and RNA-binding protein may still use unique strategies for contacting each molecule.

Binding Dynamics

Protein-RNA interactions occur through dynamic rearrangements of both molecules (Leulliot and Varani, 2001). NMR and molecular dynamics (MD) simulations as well as crystal structures with and without ligand all shed light on the dynamic process of protein-RNA interaction (Loughlin et al., 2019; Tian et al., 2011; Yu et al., 2014). RNA and protein exhibit mostly local rearrangements during binding, which often entail backbone shifts and bases and residues that “flip out” (Hainzl et al., 2005; Leulliot and Varani, 2001; Matthews et al., 2016; Yang et al., 2002). Upon binding, the site of interaction becomes rigid, locking the molecules together, whereas adjacent elements in the two molecules loosen to balance the decrease in entropy. In this way, nucleotides or residues that do not directly interact can still be instrumental in binding if they direct the necessary

compensatory changes for binding (Leulliot and Varani, 2001; Ravindranathan et al., 2010). Unstructured loops in both protein and RNA are common sites of rearrangement during binding, such as disordered linker regions between well-ordered binding domains in proteins. In fact, a large fraction of residues interacting with RNA tend to be in unstructured loops themselves (Barik et al., 2015; Cléry and Allain, 2012; Han and Nepal, 2007; Treger and Westhof, 2001), and these regions adopt structure upon binding RNA (Balcerak et al., 2019; Leulliot and Varani, 2001). Lastly, the large tertiary flexibility of RNA is a crucial functional feature in protein binding (Flores and Ataíde, 2018; Leulliot and Varani, 2001). Computational modeling of protein-RNA binding found more success with simultaneous RNA folding and docking to a protein interface as opposed to RNA folding and *then* docking (Kappel and Das, 2019), reflecting the importance of the tertiary rearrangements RNA requires to bind protein.

Protein-RNA Prediction and Resources

A great deal of interest lies in predicting RNA binding sites in proteins. For example, the above-mentioned analyses of protein-RNA structures indicated that protein-RNA interfaces prefer positively charged residues, distinguishing them from protein-protein interfaces, which prefer polar residues (Treger and Westhof, 2001). These sorts of metrics and the growing number of solved protein-RNA structures have enabled machine-learning-based attempts at predicting protein-RNA interaction. A meta-analysis of these algorithms finds that the most successful features for the prediction of RNA binding sites in protein are residue composition, conservation, and solvent accessibility (Zhang et al., 2019). Additionally, improved thermodynamic models enable docking simulations helpful for understanding the tertiary dynamics of protein-RNA binding (Huang et al., 2013; Kappel and Das, 2019). We provide an up-to-date list of these algorithms and how to access them, along with a few key RBP database resources (Table 1). Algorithms published for the purpose of RBP prediction whose source code or web server is no longer available are not included. We should note that these algorithms are biased by available structural data, which are dominated by the most abundant and readily crystallized RNA-binding domains such as the RNA recognition motif (RRM). Thus, predictive algorithms should greatly benefit in the future from the characterization of novel RNA-binding domains and data from alternative structural techniques such as cryogenic electron microscopy.

RNA-Binding Domains

RBPs typically contain discrete domains for the purpose of binding RNA. Many RNA-binding domains are quite small (<100 residues) and utilize only a handful of their residues to directly interact with RNA. This begs the question of how specific binding is achieved. The combination of multiple RNA-binding regions engaging in hydrogen bonds, stacking interactions, and additional weaker interactions with all parts of the RNA nucleotide, as described above, cumulatively enables RBPs to bind specific regions in RNAs. Multiple binding domains often co-exist in one RBP, enhancing specific RNA binding (Cléry and Allain, 2012; Lunde et al., 2007). Linkers between domains have been shown to mediate important RNA contacts as well, and the flexibility of linkers can determine whether adjacent RNA-binding domains bind independently or cooperatively (Cléry and Allain, 2012;

Lunde et al., 2007). Recognition of bipartite sequence motifs is a common occurrence, mediated by multiple RNA-binding domains and their linkers and RNA structural arrangements that present bipartite sequences to the RBP (Dominguez et al., 2018; Loughlin et al., 2019; Lunde et al., 2007; Tan et al., 2014; Walden et al., 2012). Additionally, some RNA-binding domains are capable of mediating protein-protein interactions (PPIs), including in concert with RNA binding (Cieniková et al., 2015; Yang et al., 2002). This multi-layered approach allows almost limitless combinations, such that there exist hundreds of different RBPs that conduct a wide and diverse number of functions. Here we describe in detail seventeen structurally characterized domains that have been described to bind RNA in multiple proteins (Table 2).

RNA Recognition Motif

RRMs are the most common and well-studied RNA-binding domain. A search in the Protein Data Bank (PDB) for “RRM” yields over 500 structures (Table 2), and RRM domains are estimated to occur in 1% of all human proteins (Cléry and Allain, 2012). RRM domains average 90 amino acids in size and adopt a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology forming two α helices against an antiparallel β sheet, which houses the conserved RNA-binding RNP1 and RNP2 motifs in the central β_1 and β_3 strands (Cléry and Allain, 2012). RRM domains interact with 2–8 nt in single-stranded RNA (ssRNA) commonly through several sequential stacking interactions and hydrogen bonds with RNP motifs, often with nanomolar affinities (Auweter et al., 2006; Cléry et al., 2008). Each RRM has its own sequence preferences, often for degenerate sequences such as GU-rich tracts (Cléry et al., 2008). The combination of consecutive RRM domains in an RBP dramatically increases binding affinity and specificity (Maris et al., 2005). For example, the dual binding of both RRM domains in heterogeneous nuclear RNPA1 (hnRNPA1) is crucial to its overall binding ability and function in repressing splicing (Beusch et al., 2017). RRM domains have also been observed interacting with other protein domains, such as hnRNPC, whose single RRM domain drives multimerization with other hnRNPC molecules (Cieniková et al., 2015; Safaee et al., 2012).

K Homology

The K homology (KH) domain was first discovered in heterogeneous nuclear ribonucleoprotein K (hnRNPK). At 70 amino acids, the KH domain is even smaller than the RRM domain, and typically recognizes 4 nt in ssRNA or ssDNA (Cléry and Allain, 2012; Valverde et al., 2008). KH domains adopt either a type I $\beta_1\alpha_1\alpha_2\beta_2\beta'\alpha'$ topology (in eukaryotes) or the reverse type II $\alpha'\beta'\beta_1\alpha_1\alpha_2\beta_2$ topology (in prokaryotes), with a conserved “GXXG” RNA-binding motif located between the α_1 and α_2 helices (Valverde et al., 2008). RNA binding occurs in a hydrophobic pocket and includes several hydrogen bonds coordinated by the “GXXG” motif (Auweter et al., 2006). Stacking interactions between protein and RNA in KH domains are scarce, potentially explaining the domain’s weak micromolar affinities (Cléry and Allain, 2012; Valverde et al., 2008). As with RRM domains, multiple KH domains (type I) in an RBP can independently or synergistically increase binding specificity (Valverde et al., 2008). For example, the two KH domains in MEX-3C increase its binding site to a 5-plus-4-nt bipartite motif bound with 0.17 μ M affinity (Yang et al., 2017). KH domains are also commonly found co-occurring

Table 1. Summary of Studies and Software that Catalog or Predict RBPs and Their Targets


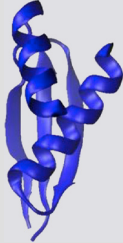
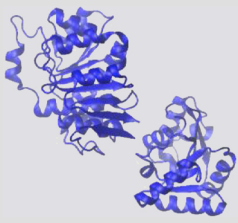
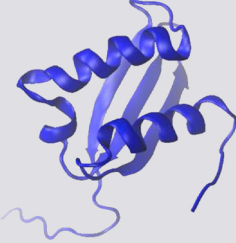
Name	Link	Reference	Description
3dRPC	http://biophy.hust.edu.cn/3dRPC.html	Huang et al., 2013	Command line software accepts PDB structures of a protein and an RNA and docks them.
aaRNA	https://sysimm.ifrec.osaka-u.ac.jp/aarna/	Li et al., 2014	Web server accepts PDB structure or protein sequence to predict residues that bind RNA. Includes graphical output of the binding propensity of each residue.
Arpeggio	http://biosig.unimelb.edu.au/arpeggioweb/	Jubb et al., 2017	Web server and command line software accept PDB file and chain ID and return all interactions that occur with the given chain. Ionic, polar, hydrogen bonds, aromatic ring stacking, etc.
ATTRACT	https://attract.cnice.es/index	Giudice et al., 2016	Database of RBPs with experimental data and their inferred bound sequence motif.
beRBP	http://bioinfo.vanderbilt.edu/beRBP/predict.html	Yu et al., 2019	Web server and command line software, given protein and RNA sequence, predict their interaction.
BioLiP	https://zhanglab.ccmb.med.umich.edu/BioLiP/	Yang et al., 2013	Database of PDB protein structures with ligands (including RNA) that annotates atoms at the binding interface in a given structure.
catRAPID	http://s.tartagliolab.com/page/catrapid_group	Bellucci et al., 2011	Command line software (requires licensing) and web server. Accept protein and RNA sequences and return a heatmap of interaction propensity at each residue-nucleotide pair.
DR_bind1	http://drbind.limlab.ibms.sinica.edu.tw	Chen et al., 2014	Web server accepts a single protein chain in PDB format and predicts which residues bind RNA. Also produces a Jmol image of the protein structure.
DRNAPred	http://biomine.cs.vcu.edu/servers/DRNAPred/	Yan and Kurgan, 2017	Web server accepts (up to 100) a protein sequence(s) and assesses each residue for its RNA (and DNA) interaction probability.
ENTANGLE	On request	Morozova et al., 2006	Software assesses hydrogen bonds and Van der Waals, stacking, and hydrophobic interactions between RNA and protein in the given PDB structure.
HBPLUS	http://www.ebi.ac.uk/thornton-srv/software/HBPLUS/	McDonald and Thornton, 1994	Command line software that lists hydrogen bonds in a given PDB structure.
hybridNAP	http://biomine.cs.vcu.edu/servers/hybridNAP/	Zhang et al., 2019	Web server accepts (up to 10) a protein sequence(s) and calculates each residue's interaction probability with RNA, DNA, and/or protein. Also returns the feature values that determine this probability.
KYG	http://cib.cf.ocha.ac.jp/KYG/	Kim et al., 2006	Web server accepts a single-chain PDB structure and predicts the RNA interface propensity of each residue. Outputs graph, table, and downloadable PDB file with scores.
ndb	http://ndbserver.rutgers.edu	Berman et al., 1992; Coimbatore Narayanan et al., 2014	Database of solved DNA and RNA structures.
NUCPLOT	https://www.ebi.ac.uk/thornton-srv/software/NUCPLOT/	Luscombe et al., 1997	Command line software accepts a protein-RNA/DNA PDB structure and returns a graphic of protein interaction occurring at each nucleotide.

(Continued on next page)

Table 1. Continued

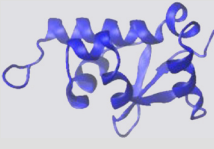
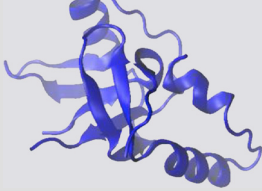
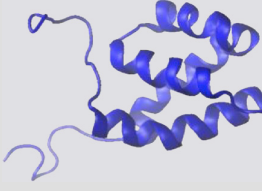
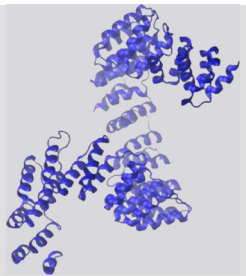
Name	Link	Reference	Description
OPRA	https://life.bsc.es/pid/opra/default/index	Pérez-Cano and Fernández-Recio, 2010	Web server scores residues in PDB structures for interaction probability with RNA. Includes Jmol output of structures with residues colored by predicted value.
PDBsum	http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=index.html	Laskowski et al., 2018	Provides an overview of a given PDB structure, including protein sequence, defined structural regions, sequence of bound RNA/DNA, NUCPLOT depiction of bound DNA/RNA, etc.
PLIP	https://projects.biotec.tu-dresden.de/plip-web/plip/index	Salentin et al., 2015	Web server and command software accept PDB structures of protein-ligand and list each hydrogen bond, salt bridge, π interaction, and hydrophobic interaction with ligand.
PPRInt	https://webs.iitd.edu.in/raghava/pprint/index.html	Kumar et al., 2008	Web server accepts a protein sequence and predicts RNA-binding residues.
PredPRBA	http://PredPRBA.denglab.org/	Deng et al., 2019	Web server accepts a PDB file of protein-RNA structure and predicts the free energy of binding.
PRince	http://www.facweb.iitkgp.ac.in/~rbahadur/prince/home.html	Barik et al., 2012	Web server accepts a PDB structure, given a protein and RNA chain IDs, and will list atoms at the protein-RNA interface.
RAIDv2.0	http://www.rna-society.org/raid2/index.html	Yi et al., 2017	Database of known RNA-RNA and protein-RNA interactions at the full transcript/protein level (not nucleotide/residue detail).
RBP prediction selection tool	https://www.iitm.ac.in:443/bioinfo/RNA-protein/	Nagarajan and Gromiha, 2014	Online tool based on the benchmark of various RBP prediction software. Shows the best software (limited selection) to use for a given RBP/RNA type.
RBPDB	http://rbpdb.ccb.utoronto.ca	Cook et al., 2011	Database of RBPs with available experimental data, categorized by organism or RBP domain.
RBPmap	http://rbpmap.technion.ac.il/index.html	Paz et al., 2014	Web server and command line software search for given RBP-binding motifs in a given RNA sequence.
RCSB PDB	https://www.rcsb.org	Berman et al., 2000	Search parameters for RBPs with RNA: "macromolecule type: contains protein AND contains RNA."
RNAbindPlus	http://ailab1.ist.psu.edu/RNABindRPlus/	Terribilini et al., 2007	Web server predicts residues that bind RNA in a given protein sequence.
RNAbindRv2.0	http://ailab-projects2.ist.psu.edu/RNABindR/	Terribilini et al., 2007	Web server predicts residues that bind RNA in a given protein sequence.
RPISeq	http://pridb.gdcb.iastate.edu/RPISeq/	Muppirla et al., 2011	Web server accepts protein and RNA sequences and predicts their interaction probability.
RsiteDB	http://bioinfo3d.cs.tau.ac.il/RsiteDB/	Shulman-Peleg et al., 2008	Database searches for PDB structure (if published before 2008) and describes protein-RNA interactions: Jmol image, which base, etc.
SPOT-Seq-RNA	https://sparks-lab.org/server/SPOT-RNA/	Yang et al., 2014	Web server and command line software predict whether a given protein sequence is an RBP.
SPOT-Struct-RNA	https://sparks-lab.org/yueyang/server/SPOT-Struct-RNA/	Zhao et al., 2011	Web server and command line software predict whether a given PDB structure is an RBP.
TriPepSVM	https://github.com/marsicoLab/TriPepSVM	Bressin et al., 2019	Command line software accepts a protein sequence and predicts whether it is an RBP.

Table 2. Summary of described RNA binding domains and example structures.

Domain Name	PDB Search Term	PDB Structures		Size (a.a.)	Protein Families Containing Domain	Example Structure PDB ID Code	Example Domain Structure
		Structures with RNA	Structures				
Cold shock domain (CSD)	"Cold shock domain"	46	13	70	Cold shock proteins, Y-box proteins	4A4I	
Double-stranded RNA Binding Domain (dsRBD)	"dsRBD"	59	27	65	RNases, ADARs, Dicer	3LLH	
Helicase	"dead" OR "deah" OR "helicase domain"	701	114	350-400	DEXH/D-box, Ski2-like, RIG-I-like, NS3, UPF1-like RNA binding helicases	2I4I	
Intrinsically disordered region (IDR)	"Intrinsically disordered region"	NA	NA	varies	Most RBPs	NA	NA
K homology (KH)	"KH domain"	117	20	70	hnRNPs, translation regulation proteins, very common	1WH9	

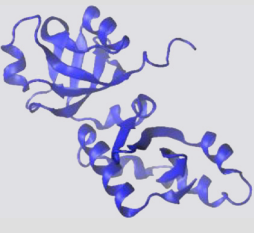

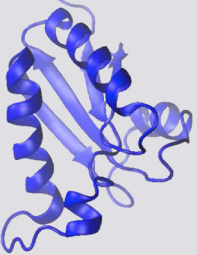
(Continued on next page)

Table 2. Continued

Domain Name	PDB Search Term	PDB Structures	PDB Structures with RNA	Size (a.a.)	Protein Families Containing Domain	Example Structure PDB ID Code	Example Domain Structure
La motif (LAM)	"La motif AND NOT rrm"	54	5	90	La proteins, La-related proteins (LARPs)	1S29	
Piwi-Argonaute-Zwille (PAZ)	"PAZ domain"	71	44	170	Argonaute proteins, Dicer	3O6E	
P-element Induced Wimpy Testis (PIWI)	"PIWI domain"	74	33	290	Argonaute proteins	1X4Q	
Pentatricopeptide repeat (PPR)	"pentatricopeptide repeat"	32	9	35 * n, 1 > n > 30	RNA editing proteins	4M59	

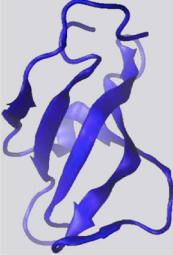
(Continued on next page)

Table 2. Continued

Domain Name	PDB Search Term	PDB Structures	PDB Structures with RNA	Size (a.a.)	Protein Families Containing Domain	Example Structure PDB ID Code	Example Domain Structure
Pseudouridine synthase and archaeosine transglycosylase (PUA)	"Pseudouridine synthase and archaeosine transglycosylase" OR "PUA domain"	119	28	66-98	RNA modifying enzymes, metabolic enzymes	1SQW	
Pumilio-like repeat (PUM)	"Pumilio" OR "PUM domain" OR "Puf protein"	60	49	334	PUF proteins	1M8W	
Ribosomal S1-like (S1)	"S1 RNA binding domain"	25	2	70	Ribosomal proteins, Translation initiation factors, RNase II, PNPase	2EQS	
RNA Recognition Motif (RRM)	"RNA Recognition Motif" OR "RRM"	554	119	90	hnRNPs, splicing factors, very common	2MTG	

(Continued on next page)

Table 2. Continued

Domain Name	PDB Search Term	PDB Structures	PDB Structures with RNA	Size (a.a.)	Protein Families Containing Domain	Example Structure PDB ID Code	Example Domain Structure
Sm and Like-Sm (Sm / Lsm)	"Sm RNA binding domain"	31	23	80	U1 spliceosomal proteins, Hfq	2VC8	
thiouridine synthases, RNA methylases and pseudouridine synthases (THUMP)	"THUMP domain"	12	3	100-110	tRNA modifying enzymes	2DIR	
YTH521-B homology (YTH)	"YTH domain"	28	14	100-150	YTH family m ⁶ A readers	4RCI	
Zinc Finger (ZnF)	"Zinc Finger"	2677	63	30	Transcription factors, METTL enzymes, Very common	5ZC4	

with quaking (QUA) domains as part of the larger signal transduction and activation of RNA (STAR) domain, which greatly extends the binding surface and accommodates 7 or 8 nt with 0.07 μM affinity (Sharma and Anirudh, 2017; Teplova et al., 2013).

Zinc Finger

ZnFs describe a large family of proteins that average 30 amino acids in size and form a simple $\beta\beta\alpha$ topology in which residues in the β hairpin turn and α helix are coordinated by a Zn^{2+} ion (Cléry and Allain, 2012). Most ZnFs bind DNA, but have been additionally shown to bind RNAs, proteins, and small molecules (Lai et al., 2000). ZnF subtypes that interact with RNA include CCHC (zinc knuckle), CCCH, CCCC (RanBP2), and CCHH subtypes, where C and H refer to the interspersed cysteine and histidine residues that coordinate the zinc atom, respectively (Cléry and Allain, 2012). These subtypes display a range of sequence and structural specificities. Zinc knuckles recognize stem-loop elements in RNA (or ssDNA) through contacts with bases in the loop and the phosphate backbone of the stem. CCCH and CCCC subtypes tend to recognize 3-nt repeats through multiple such ZnFs in one RBP (Font and Mackay, 2010; Hall, 2005; Lai et al., 2000). These contacts are formed through hydrogen bonds with bases and the insertion of aromatic side chains that stack between bases. The versatile and abundant CCHH ZnFs interact with both single-stranded and dsRNA as well as DNA (Font and Mackay, 2010; Hall, 2005). Modular arrays of CCHH ZnFs have been successfully engineered to bind desired DNA sequences. Thus, designer ZnFs are thought to have potential for directed binding of RNA sequences, a goal that has been achieved with the much larger Pumilio homology domains (Font and Mackay, 2010).

Pumilio Homology Domain

The Pumilio and FBF (PUF) family of proteins occurs in most eukaryotes and is defined by the Pumilio homology domain (PUM-HD), or the PUF domain. The PUF domain is very large, consisting of eight α -helical repeats of a highly conserved 36-amino acid sequence that forms a concave RNA-binding surface (Wang et al., 2018). Each repeat recognizes one unpaired base through hydrogen bonds and a stabilizing stacking interaction, where the full domain recognizes up to 8 nt in ssRNA with low-nanomolar affinity (Zhao et al., 2018). Wild-type PUF repeats do not specifically recognize cytosine; however, protein engineering has produced repeats that do (Zhao et al., 2018). These advances combined with the PUF domain's predictable base recognition code allow modular design of pumilio proteins that recognize 8- to 10-nt sequences containing all RNA bases (Zhao et al., 2018).

Pentatricopeptide Repeat

Very similar to PUF repeats, eukaryotic pentatricopeptide repeats (PPRs) are each ~ 35 residues in length and form two antiparallel α helices. 2–30 repeats form a solenoid-shaped scaffold that binds specific ssRNA sequences with nanomolar affinity (Ke et al., 2013; Spähr et al., 2018). Two residues in each repeat determine base-specific binding through hydrogen bonds, enabling the development of designer PPRs that bind specified ssRNA or ssDNA sequences (Spähr et al., 2018).

Pseudouridine Synthase and Archaeosine Transglycosylase

The pseudouridine synthase and archaeosine transglycosylase (PUA) domain is found in the aforementioned enzymes as well

as several other RNA-modifying and metabolic enzymes (Pérez-Arellano et al., 2007). PUA domains range from 67 to 94 amino acids in length, with a $\beta_1\alpha_1\beta_2\beta_3\beta_4\beta_5\alpha_2\beta_6$ architecture that forms a pseudobarrel encased by two α helices. PUA domains have been characterized contacting dsRNA and its adjacent loops or overhangs through extensive hydrogen bonds with all parts of the RNA. These contacts are typically formed by a glycine-rich loop between α_1 and β_2 or α_2 and β_6 . Unlike many other domains, PUA domains are not found as tandem repeats (Pérez-Arellano et al., 2007).

THUMP

Named for thiouridine synthase, methyltransferase, and pseudouridine synthase, the THUMP domain is found in numerous tRNA-modifying enzymes. About 100 amino acids long, THUMP domains are always found in proximity to RNA-modifying domains and often in proximity to an N-terminal ferredoxin-like (NFLD) domain (Neumann et al., 2014). THUMP domains display a $\alpha_1\alpha_2\beta_1\alpha_3\beta_2\beta_2$ topology that forms parallel α helices flanking a β sheet (Fislage et al., 2012). The first structure of a THUMP domain bound by RNA, bacterial 4-thiouridine synthetase in complex with tRNA, reveals a 3-dimensional fold that specifically recognizes the 3'-CCA tail and adjoining stem of tRNA. Several hydrogen bonds and VdW contacts correctly position the tRNA for modification by the accompanying pyrophosphate domain (Neumann et al., 2014).

YT521-B Homology

The YT521-B homology (YTH) domain is found in the YTH family of proteins that “read” N⁶-methyladenosine (m^6A) marks in RNA. The YTH domain ranges from 100 to 150 amino acids in length and forms a six-stranded β barrel surrounded by four or five α helices. Three residues in the hydrophobic core of the β barrel trap the methyl group of m^6A in an “aromatic cage” consisting of hydrogen bonds with the adenosine and π interactions between tryptophan rings and the methyl group (Liao et al., 2018; Xu et al., 2014). The YTH domain specifically binds m^6A over unmodified adenosines. Affinity of YTHDC1 for consensus DR $\text{m}^6(\text{A})\text{CH}$ motifs was measured at 0.3 μM , whereas no binding was detected for the unmethylated sequence (Xu et al., 2014). Similarly, YTHDF2 affinity for methylated RNA was measured as 2.54 μM , with 10-fold lower affinity for the unmethylated target (Zhu et al., 2014).

Double-Stranded RNA-Binding Domain

Double-stranded RNA-binding domains (dsRBDs), or motifs (dsRBMs), consist of 65–70 amino acids and are the third most common RNA-binding domain (Masliah et al., 2013). dsRBDs specifically recognize and bind dsRNA and are found in proteins with roles in viral protection, RNAi, and cellular transport (Masliah et al., 2013). dsRBDs often appear as tandem repeats or in combination with other functional RNA-binding domains, such as RNA-editing or helicase domains (Cléry and Allain, 2012; Ranji et al., 2011). The domain is made up of an $\alpha_1\beta_1\beta_2\beta_3\alpha_2$ fold that forms an antiparallel β sheet flanked by α helices on one face (Cléry and Allain, 2012; Masliah et al., 2013). dsRBDs specifically recognize the structure of an A-form RNA helix, spanning up to 16 bp with hydrogen-bond contacts to the phosphodiester backbone and 2' OH (Cléry and Allain, 2012; Ramos et al., 2000). In some cases, dsRBDs have demonstrated base-specific contacts, such as to bases in adjacent loops (Cléry and Allain, 2012; Masliah

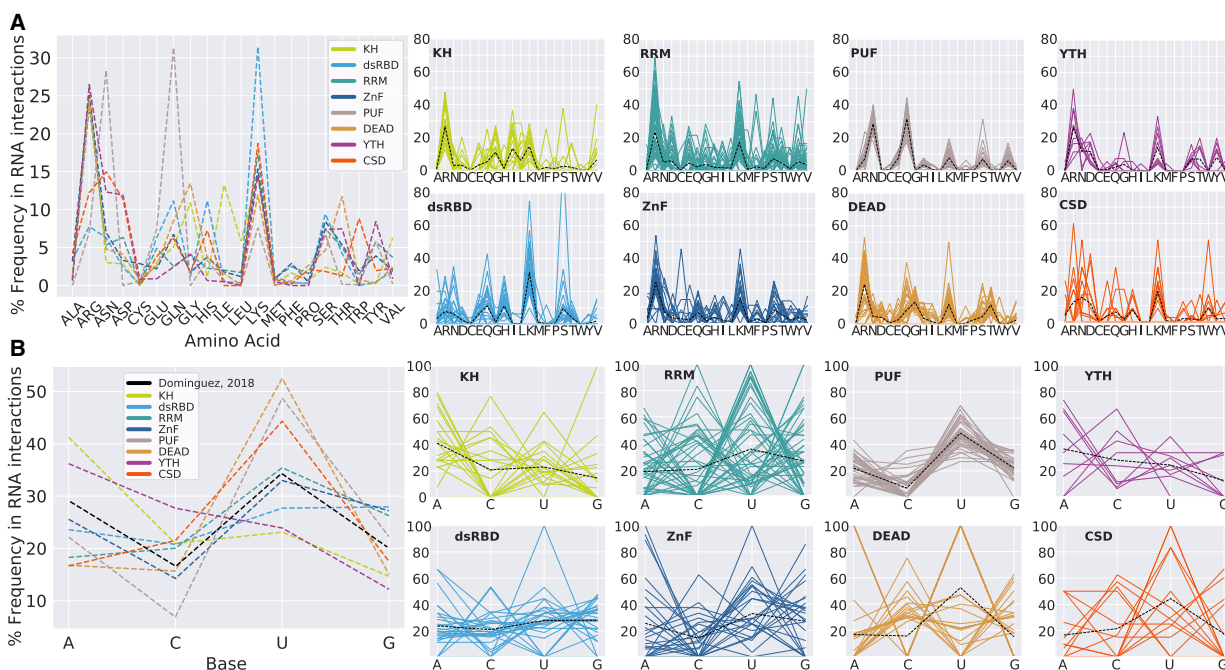


Figure 3. Amino Acid and Base Preferences in Protein-RNA Hydrogen Bonds Observed in over 200 Structures, Organized by Domain

(A) The average frequency of each amino acid in forming hydrogen bonds with RNA across eight RNA-binding domain types (left). The frequency of each amino acid (one-letter abbreviations) in forming hydrogen bonds with RNA in multiple structures, separated by domain type (right, smaller plots), is shown. CSDs use Trp more frequently than other domains ($p = 2.38 \times 10^{-5}$). KH domains use Leu and Ile more frequently than other domains ($p = 7.26 \times 10^{-6}$, 6.16×10^{-9}). (B) The average frequency of each RNA nucleotide in forming hydrogen bonds with protein across eight RNA-binding domain types, as well as the average frequency of each base in sequence motifs from Bind-n-Seq data (Dominguez et al., 2018) (left). The frequency of each RNA nucleotide in forming hydrogen bonds with protein in multiple structures, separated by domain type (right, smaller plots), is shown. PUF domains contact cytosine least frequently ($p = 0.001$). KH domains contact adenosine most frequently ($p = 0.002$).

et al., 2013). Stacking interactions are rare, potentially explaining the low affinities (high-nanomolar-to-micromolar range) of dsRBDs to RNA targets (Steff et al., 2010; Wang et al., 2011).

Helicase

Helicase domains are found in all forms of life in helicase proteins, which unwind both DNA and dsRNA. Helicases comprise six superfamilies (SFs), of which SF1 and SF2 contain all the eukaryotic RNA and DNA helicases. RNA-binding helicases include the Upf1-like family in SF1 and the DEAD-box, DEAH, RIG-I-like, Ski2-like, and NS3 families in SF2. The remaining SFs, 3–6, contain bacterial and viral helicases that form multimeric rings (Jankowsky, 2011). Helicase domains are very large, containing 350–400 amino acids. In SF1 and SF2, the helicase domain is composed of two “recombinase A (recA)-like” subdomains, each of which contains an ATP-catalytic core, a nucleic-acid-binding region, and subdomains that coordinate the two. Within families of helicases these subdomains are quite conserved. Helicase monomers in the ring-forming SFs of helicases are similarly quite large and composed of multiple subdomains (Gai et al., 2004; Kainov et al., 2008). Bound RNA is surrounded by recA-like domains or, in the case of multimeric helicases, RNA is pulled through the center of the ring. Contacts with RNA are dominated by hydrogen bonds to phosphate and sugar moieties, but contacts with bases are occasionally observed (Jankowsky, 2011; Kainov et al., 2008; Linder and Jankowsky, 2011; Weir

et al., 2010). Multiple nucleotides are typically contacted simultaneously; DEAH/DEAD-box helicases, for example, tend to accommodate at least 5 single-stranded or base-paired nucleotides (Jiang et al., 2011; Linder and Jankowsky, 2011; Weir et al., 2010). Affinities to RNA are often in the nanomolar range, although they vary greatly by helicase and are modulated by other subdomains of the helicase. ATP binding generally promotes higher affinity to RNA by causing the helicase RNA-binding regions to “clamp.” ATP hydrolysis subsequently promotes conformational changes that cause the helicase to translocate 1 nt and/or unwind its substrate (lost et al., 1999; Jankowsky, 2011; Jiang et al., 2011; Kainov et al., 2003).

Cold Shock Domain

The cold shock domain (CSD) is found in a large family of proteins associated with cold adaptation found in all domains of life. CSDs are composed of ~70 amino acids (more in eukaryotes) and five antiparallel β strands that form a common β barrel structure known as an oligosaccharide/oligonucleotide-binding (OB) fold. CSDs contain the conserved RNP1 and RNP2 motifs common to RRM, which bind ssRNA and ssDNA (Amir et al., 2018). CSDs contact 3 or 4 nt through sequential stacking interactions and hydrogen bonds with bases, achieving nanomolar affinities (Kljashorny et al., 2015; Sachs et al., 2012). CSD-containing proteins vary greatly in the types of sequences they recognize. Bacterial CspB is reported to bind pyrimidine-rich

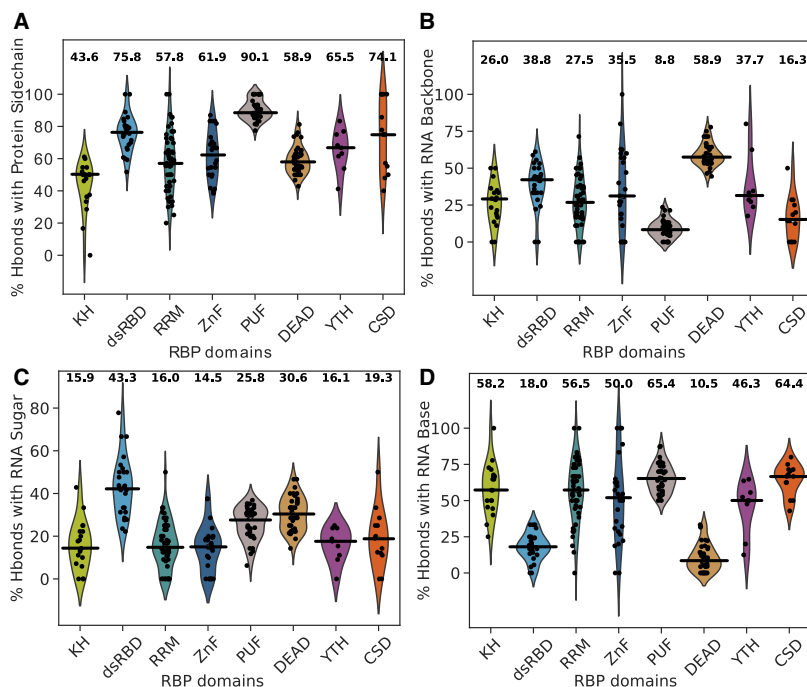


Figure 4. Assessment of Protein-RNA Hydrogen Bonds in over 200 Structures, Organized by RNA-Binding Domain

Averages for each statistic are listed above each domain's violin plot and medians are indicated with black horizontal bars.

(A) The percent of protein-RNA hydrogen bonds that are formed using protein side chains (as opposed to the main chain). KH domains use sidechains to hydrogen bond with RNA significantly less than PUF domains ($p = 2.25 \times 10^{-13}$).

(B) The percent of protein-RNA hydrogen bonds that are formed with RNA backbone atoms. PUF domains hydrogen bond with the RNA backbone the least ($p = 1.60 \times 10^{-13}$) and DEAD domains the most ($p < 1 \times 10^{-307}$).

(C) The percent of protein-RNA hydrogen bonds that are formed with RNA sugar atoms. dsRBDs hydrogen bond most frequently with the 2' OH ($p = 1.90 \times 10^{-10}$).

(D) The percent of protein-RNA hydrogen bonds that are formed with RNA base atoms. PUF domains hydrogen bond most frequently with the RNA base ($p = 1.31 \times 10^{-7}$) and least frequently with dsRBDs and DEAD domains ($p = 5.44 \times 10^{-8}$, $p < 1 \times 10^{-307}$).

sequences and prefer ssDNA to ssRNA by up to 10-fold (Sachs et al., 2012). Y box proteins contain the most well-studied eukaryotic CSDs, showing a preference for G-rich ssRNA sequences over ssDNA (Kljashorny et al., 2015).

S1

The S1 RNA-binding domain was originally discovered in S1 ribosomal protein, which binds both mRNA and rRNA. The ~70-amino acid S1 domain forms a 5-stranded antiparallel β barrel in the same OB-fold family as the CSD (Mihailovich et al., 2010). Despite sharing a common tertiary structure, the two domains show no sequence similarity, suggesting that their shared tertiary structure was achieved through convergent evolution (Mihailovich et al., 2010). S1 domains are additionally found in several exoribonucleases and eukaryotic translation initiation factors and in combination with other RNA-binding domains such as the KH domain or CSDs (Amir et al., 2018; Chekanova et al., 2002; Hossain et al., 2016; Worbs et al., 2001). Despite their abundance, very little structural information is available for S1 domains in complex with RNA. S1 domains interact with both ssRNA and dsRNA in the context of the RNA-binding channel of exoribonucleases (Hossain et al., 2016). Similarly, S1 domains of the ribosomal S1 protein likely interact with mRNA at the entry channel of the ribosome (Loveland and Korostelev, 2018).

Sm

The Sm RNA-binding motif is found in Sm and like-Sm (Lsm) proteins in eukaryotes and archaea and in Hfq protein in prokaryotes (Schumacher et al., 2002; Thore et al., 2003). The Sm motif consists of ~70 residues with an $\alpha_1\beta_1\beta_2\beta_3\beta_4\beta_5$ topology that forms a curved antiparallel β sheet. Sm-containing proteins readily multimerize through interactions between

strands β_4 and β_5 in two Sm motifs. For example, Sm-Sm interactions link the seven human Sm proteins that make up the protein core of small nuclear ribonucleoproteins (snRNPs) in the spliceosome (Thore et al., 2003). The Sm multimers bind RNA with nanomolar affinity. Two Sm motifs form a 6-nt binding surface that binds specific bases, often uridines, through hydrogen bonds and stacking interactions (Schumacher et al., 2002; Thore et al., 2003).

La Motif

The small ~90-residue La motif (LAM) is found in eukaryotic La and La-related proteins (LARPs). The LAM consists of five α helices and three β strands that form a small antiparallel β sheet against a modified "winged-helix" fold (Bousquet-Antonelli and Deragon, 2009). The winged-helix structure itself is common to several other RNA- and DNA-binding proteins (Teichmann et al., 2012). LAMs are always found adjacent to at least one RRM, where the combination of these two domains likely evolved as a unit (Bousquet-Antonelli and Deragon, 2009). In La proteins, the dual LAM-RRM region tightly binds the UUU-OH elements at the 3' ends of polymerase-III-transcribed small RNAs. Binding occurs in a cleft between the LAM and RRM rather than the traditional RNA-binding surfaces of either the RRM or the LAM winged-helix fold. Several uracil bases stack with highly conserved aromatic residues in the LAM, and hydrogen bonds from both the LAM and RRM coordinate bases, phosphates, and the terminating 2' OH. These contacts result in low-nanomolar affinities of the LAM for 3'-terminal UUU-OH elements (Teplova et al., 2006). The other LAM-containing proteins, LARPs, bind a diverse set of RNAs with as-yet uncharacterized structural mechanisms (Schenk et al., 2012).

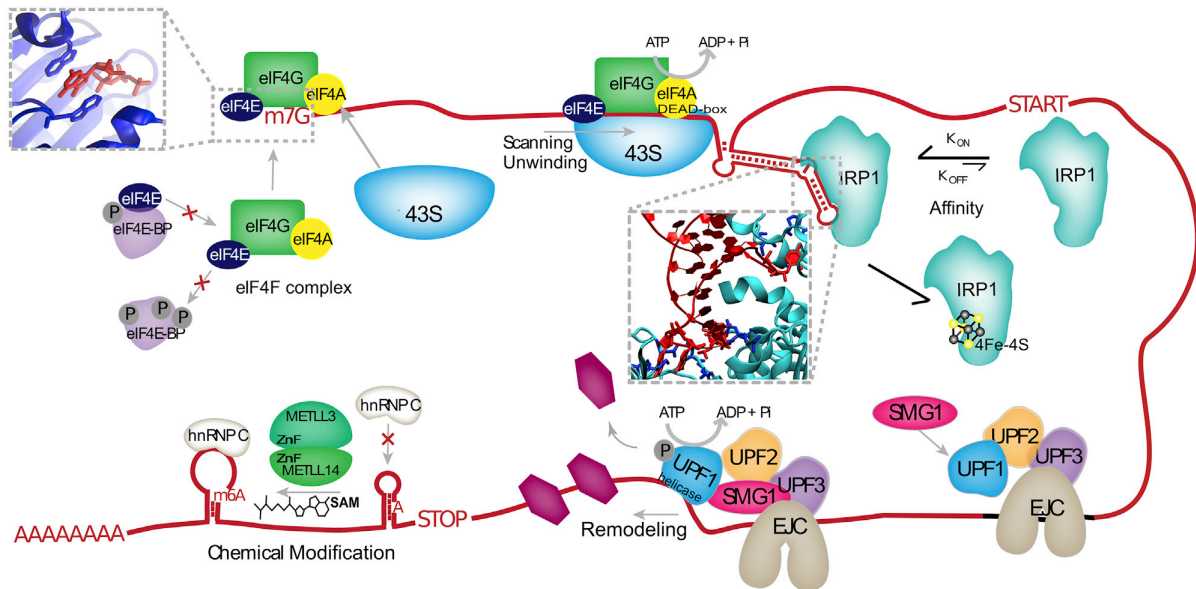


Figure 5. Examples of Mechanisms Controlling RBP Binding, Interactions with RNA, and Their Regulation

eIF4E (dark blue) interacts with the 7-methyl-guanosine cap (m7G), in part through stacking interactions (inset) and binds to RNA as part of eIF4F, which includes the RBPs eIF4G and eIF4A. eIF4E association with eIF4F is prevented by sequestration to hypo-phosphorylated eIF4E-BP. The 43S ribosomal subunit is recruited to the eIF4F complex and processively scans the 5' UTR, aided by ATP-driven helicase activity of the eIF4A DEAD-box domain. The RBP IRP1 specifically binds hairpin elements in the 5' UTR with high affinity through specific residues (inset, dark blue) that hydrogen bond with the bulge and apical loop of the RNA. RNA binding by IRP1 is prevented by 4Fe-4S ligand binding to IRP1. UPF1 is recruited to the exon junction complex (EJC), where its helicase activity is activated by interactions with SMG1 and UPF2. Driven by ATP, UPF1 removes both RNA structures and other bound RBPs in the 5' → 3' direction. The METTL3-METTL14 complex, which contains zinc fingers (ZnF), deposits methyl groups donated by S-adenosyl methionine (SAM) on targeted adenosines (m⁶As). m⁶A modifications reduce base pairing in RNA, such that some locations become available for hnRNPC binding.

Piwi-Argonaute-Zwille and PIWI

Piwi-Argonaute-Zwille (PAZ) and PIWI RNA-binding domains define the Argonaute family of proteins found in eukaryotes (Höck and Meister, 2008). Found on opposite sides of the Argonaute protein, both domains facilitate binding of small interfering RNA and microRNA guides to mRNA targets (Höck and Meister, 2008).

PAZ domains occur in Dicer proteins in addition to Argonaute proteins (Höck and Meister, 2008). Crystal structures of the PAZ domain display a six-stranded β barrel topped with two α helices and flanked on the opposite side by a special appendage containing a β hairpin and short α helix (Ma et al., 2004; Tian et al., 2011; Yan et al., 2003). A binding pocket formed between this appendage and the β barrel binds the 2-nt 3' overhang in guide RNAs (gRNAs) with low-micromolar affinity (Ma et al., 2004; Tian et al., 2011). Binding is coordinated mostly by conserved tyrosine residues that form hydrogen bonds with the phosphate backbone and sugar hydroxyls of the two terminal nucleotides (Ma et al., 2004; Tian et al., 2011).

The PIWI domain tertiary structure forms an RNase H-like fold consisting of a five-stranded β sheet flanked by α helices on both faces (Boland et al., 2011). The PIWI domain has endonucleolytic activity in some cases, but primarily stabilizes the gRNA-mRNA duplex seed region through hydrogen bonds with the gRNA backbone of nucleotides 3–5 and the 5' overhang base (Boland et al., 2011; Ma et al., 2005; Miyoshi et al., 2016). The PIWI domain also contacts the PAZ domain in certain conformations,

suggesting that its activity may be modulated by the conformational state of the PAZ domain (Boland et al., 2011).

Intrinsically Disordered Region

Intrinsically disordered regions (IDRs) are unstructured and often consist of repeats of arginine/serine (RS repeat), arginine/glycine (RGG box), arginine- or lysine-rich patches (R/K basic patches), or short linear motifs of amino acids (Balcerak et al., 2019; Järvelin et al., 2016). Despite their lack of structure, IDRs have been found to dominate the composition in over 20% of RBPs (Järvelin et al., 2016). It is increasingly observed that IDRs can be the sole RNA-binding domain in an RBP and may actually drive the majority of protein-RNA interactions in the cell (Hentze et al., 2018). Like globular RNA-binding domains, IDRs are conserved, often occur multiple times in one RBP, and can coordinate RNA binding in concert with other domains (Balcerak et al., 2019; Järvelin et al., 2016; Loughlin et al., 2019). IDRs have been shown to drive higher affinity to RNA in RBPs that contain ordered RNA-binding domains and can themselves transition to an ordered state once bound to RNA (Balcerak et al., 2019; Cruz-Gallardo et al., 2019; Järvelin et al., 2016; Leulliot and Varani, 2001). IDRs show little RNA sequence dependence, however, suggesting that these regions' high affinity for RNA is predominantly driven by electrostatic attraction to the phosphodiester backbone (Balcerak et al., 2019; Järvelin et al., 2016).

Other RNA-Binding Domains

The domains detailed above represent a mere fraction of the RNA-binding domains in existence, 23% of the 2,685 RNA-bound

structures in the PDB, whereas many hundreds more RNA-binding domains await characterization (Hentze et al., 2018). Several domains, such as the Brix domain, sterile alpha motif (SAM), and SAF-A/B, Acinus, and PIAS (SAP) domain, are mostly known as protein- or DNA-binding domains but have one or two protein members shown to bind RNA. For example, the SAM domain is a well-known α -helical PPI domain, but the SAM-containing proteins Smaug and its homolog VTS1p bind the pentaloop of an RNA hairpin element called the Smaug recognition element (SRE) with nanomolar affinity (Aviv et al., 2006; Ravindranathan et al., 2010). Conservation of the RNA-interacting residues among Smaug homologs suggests RNA-binding function exists in other SAM-containing proteins (Aviv et al., 2006). Many other RNA-binding domains are utterly unique, i.e., not yet found to resemble any other domain (Gerstberger et al., 2014; Hentze et al., 2018; Tan et al., 2013; Walden et al., 2012). The stem-loop-binding protein domain is found only in the protein of the same name (SLBP) that exclusively binds a conserved stem loop at the ends of histone mRNAs (Tan et al., 2013). IRP1, also known as ACO1, is an aconitase with a unique fold that tightly binds iron response elements in the UTRs of iron-metabolism-related transcripts (Walden et al., 2012). Most ribosomal proteins each contain a unique RNA-binding domain, the S1 domain excepted (Gerstberger et al., 2014). Structural data of viral RBPs reveal incredibly diverse and unique structures specialized for binding highly structured viral RNA elements. Recent large-scale studies have discovered hundreds of novel RNA-binding regions in proteins, including in many well-characterized enzymatic proteins such as GAPDH, that surprisingly “moonlight” as RBPs (Hentze et al., 2018; Hudson and Ortlund, 2014). Overall, the diversity of RBPs is an astounding testament to their all-encompassing cellular roles in many domains of life.

Domain Differences in Hydrogen-Bond Formation with RNA

We additionally assessed the type and number of hydrogen bonds that proteins form with RNA in over 200 structures from eight common domain types (KH, dsRBD, RRM, ZnF, PUF/PUM-HD, DEAD helicase, YTH, and CSD). This includes the frequency of each amino acid used for protein-RNA hydrogen bonds, the frequency of each RNA base that contacts are formed with, and the moieties used to facilitate those bonds (Figures 3 and 4). This hydrogen-bond analysis of protein-RNA structures has been conducted many times before (Figure 2) (Allers and Shamoo, 2001; Ellis et al., 2007; Han and Nepal, 2007; Hu et al., 2018; Jones et al., 2001; Morozova et al., 2006; Treger and Westhof, 2001), but without assessing domains separately. We immediately observe that across all domain types the positively charged amino acids Lys and Arg most frequently facilitate hydrogen bonds with RNA (Figure 3A), which directly agrees with previous analyses (Pérez-Cano and Fernández-Recio, 2010). Asp, Gln, His, and Ser are also frequently used, but are more dependent on domain type. Non-polar amino acids Ala, Cys, Met, and Pro are universally avoided. Trp is strongly avoided in hydrogen bonds with RNA by all domains except CSDs ($p = 2.38 \times 10^{-5}$). Our analysis of amino acid frequencies considers the main-chain atoms of each residue as belonging to that particular amino acid. Thus, we observe that Leu and Ile, whose side chains are not capable of forming hydrogen bonds, are

repeatedly involved in forming hydrogen bonds with RNA among KH domains but no other domain ($p = 7.26 \times 10^{-6}$, 6.16×10^{-9}). This agrees with previous descriptions of salient RNA contacts in crystal structures of KH domains being formed via the main chain of Ile residues (Cléry and Allain, 2012). Interestingly, a single-point mutation replacing an Ile in the KH domain of FMRP is known to cause fragile X syndrome (De Boule et al., 1993).

Preferences for RNA nucleotides in protein-RNA hydrogen bonds were assessed as well. Note that even with small RNAs, not necessarily all of the nucleotides interact with the RBP in a hydrogen bond. Previous analyses have varied in whether they report RBP preferences for interaction with specific nucleotides (Pérez-Cano and Fernández-Recio, 2010), and the types of RNA sequences that have been successfully co-crystallized with RBPs could be biased by technical reasons. Nevertheless, assessing our data, we observe a preference for interactions with uracil and an under-enrichment for cytosine (Figure 3B). Sequence motifs derived from RNA Bind-n-Seq experiments (Dominguez et al., 2018) also follow this pattern of base frequencies ($r^2 = 0.89$, Pearson), providing orthogonal agreement for the observed base preferences. PUF domain structures exhibit the lowest percentage of hydrogen bonds with cytosine ($p = 0.001$), reflecting the lack of cytosine recognition in wild-type PUF repeats (Zhao et al., 2018). The KH domain re-asserts its status as an oddball, as it forms hydrogen bonds with adenines more frequently than any other domain ($p = 0.002$).

We were also interested in statistics summarizing the frequency of hydrogen bonds with the side chain of amino acids and with the base, backbone, or sugar of RNA (Figure 4). We observe pronounced domain differences with the percent usage of amino acid side chains (versus the main chain). For example, KH domains use side chains in 43.6% of hydrogen bonds with RNA, whereas PUF domains use side chains in 90.1% of interactions (p value of difference = 2.25×10^{-13}) (Figure 4A). In fact, side-chain rather than main-chain hydrogen bonds typically dominate protein-RNA interactions (Figure 2A), but KH domains are the only domain analyzed here that violate this trend (Figure 4A). The percentages of hydrogen bonds forming with either the backbone, sugar, or base of RNA nucleotides were also calculated for each protein structure (Figures 4B–4D). PUF domains predictably hydrogen bond with the RNA backbone with the lowest frequency on average ($p = 1.60 \times 10^{-13}$) (Figure 4B), instead interacting with the RNA base in 65.4% of its hydrogen bonds—the highest average of all domains ($p = 1.31 \times 10^{-7}$) (Figure 4D). Also, somewhat predictably, dsRBDs and DEAD-helicase domains hydrogen bond least frequently with the RNA base ($p = 5.44 \times 10^{-8}$, $p < 1 \times 10^{-307}$), and DEAD domains most frequently with the RNA backbone ($p < 1 \times 10^{-307}$). This reflects the known ability of PUF domains to recognize sequences hyper-specifically, and of dsRBDs and DEAD domains to generally bind RNA without sequence preferences. dsRBDs also hydrogen bond most frequently—43.3% of the time—with the 2'-OH moiety ($p = 1.90 \times 10^{-10}$) (Figure 4C), a testament to these domains' specific recognition of dsRNA rather than DNA (Vuković et al., 2014).

Overall, our analysis highlights how domains differ and reinforces what is known about how certain domains form hydrogen bonds with RNA in service of their specific biology. In the future,

assessing stacking, hydrophobic, and VdW interactions by domain would additionally contribute to defining domains' specific binding strategies. Quantifying binding strategies in this way will serve prediction and design efforts aimed at controlling the biology of RBPs.

Regulation Mechanisms

Although RNA-binding domains are directly responsible for interacting with RNA, we must consider the dynamic cellular context that regulates this interaction. In this section, we will describe the factors that determine how RBPs find their targets, what combined function they perform with their targets, and how the first two are regulated.

Protein-RNA Assembly

How does a given RBP find its target? The intermolecular interactions between protein and RNA are the raw starting material for determining their affinity, where interactions with specific moieties and/or binding pockets that “fit” an RNA substrate are common strategies for highly effective binding. In the case of IRP1, its “L-shaped” binding pocket and contacts with select bases yield an incredibly specific and strong picomolar affinity for the iron response element (Figure 5) (Walden et al., 2012). eIF4E, on the other hand, lacks specificity for select RNA elements such as IRP1, instead generally binding all mRNAs through its recognition of their 5' caps. This interaction is strong (nanomolar affinity), facilitated by stable stacking interactions with the m⁷Gppp structure (Figure 5) (Jackson et al., 2010; Niedzwiecka et al., 2002). Lastly, many RBPs exhibit neither specific nor high-affinity binding for the purpose of functionally transient associations with RNA (Auweter et al., 2007; Linder and Jankowsky, 2011). RBP (or RNA) abundance also affects the free energy of binding, where high abundance pushes the equation in favor of association. Thus, it is no wonder that IRP1 interaction with the iron response element in the highly abundant *FTL* transcript is one of the most reliable and well-studied protein-RNA interactions. Recruitment by other proteins is the primary mode by which many RBPs find their targets, especially in the assembly of multi-component RNP complexes. For example, eIF4E binds mRNA 5' caps as a subunit of the eIF4F translation initiation complex, which delivers other members of the complex, such as helicase eIF4A, for action on the transcript (Figure 5). The eIF4F complex additionally recruits the 43S ribosomal complex to begin scanning the 5' UTR (Jackson et al., 2010). In the same way that RBPs are recruited, they may also be prevented from binding by sequestration by other proteins or modifications to their binding sites (described below).

Combined Function of Protein and RNA

The individual components of protein and RNA have a different function once associated. This function can be guided by RNA-binding domains themselves or cooperative enzymatic domains in the same RBP, or may require multiple protein/RNA components. We may separate protein-RNA functions into several broad categories: static binding, scanning/translocation, remodeling, and modification. Many RBPs statically bind specific RNA elements (often small hairpins) (Ravindranathan et al., 2010; Tan et al., 2013; Walden et al., 2006); that is to say, they do not alter the RNA further once bound. This mode of interaction often blocks

access to the RNA, such as with IRP1, which impedes translation initiation by blocking the scanning action of eIF4A (Figure 5) (Walden et al., 2006). Translocation along RNA substrates is common among RNA helicases, which include eIF4A and UPF1 (Figure 5). UPF1, as well as a few other helicases, additionally remodels the RNP landscape of its substrate by processively removing other RBPs in its path (Fiorini et al., 2015; Jankowsky, 2011). Most frequently, helicases serve the purpose of removing RNA secondary structure, where scanning and structure unwinding can occur either together or independently (Fiorini et al., 2015; Jackson et al., 2010).

Performed by more than just helicases, manipulation of RNA structure can be considered a non-covalent form of RNA modification. Remodeling RNA includes chaperoning RNA structure formation in addition to removing structures. Chaperone activity on RNA has been observed among cold shock proteins and helicases, as well as the pseudouridine synthase TruB, which modifies tRNA structure in addition to chemically modifying the RNA (Keffer-Wilkes et al., 2016; Rajkowsch et al., 2007). RNA is chemically modified a number of ways by enzymes, known as writers, coupled with RNA-binding domains. More than 100 chemical modifications have been identified in RNA involving all bases and the 2' OH (Motorin and Helm, 2011). Among the most well-known writers are the methyltransferase-like (METTL) proteins, such as the dual METTL3-METTL14 complex, which uses a donor methyl group from the co-substrate S-adenosyl methionine (SAM) to methylate the sixth carbon of adenine (m⁶A; Figure 5) (Śledź and Jinek, 2016). Some chemical modifications warp the local secondary structure of their resident RNAs, with reports of base-pair stabilization from pseudouridine modifications (Ge and Yu, 2013) and de-stabilization from inosine and m⁶A (Tanzer et al., 2019). Lastly, RNA is modified in a more extreme fashion by RBPs with nucleolytic activity, such as Argonaute proteins and the RNases Dicer and RNase II (Frazão et al., 2006; Jiang et al., 2011).

Regulation of Binding and Function

PPIs enable cooperative and competitive control over RBP binding. Such is the case for the translation initiation complex: eIF4E binding to the mRNA 5' cap is enhanced by its association with the subunit eIF4G; however, eIF4E interaction with eIF4E-BP prevents its binding to the eIF4F complex (Jackson et al., 2010). PPIs can also modulate RBP function or its efficiency thereof. Exon junction complex (EJC) interactors UPF2 and SMG1 induce conformational changes upon binding to UPF1 that de-repress its helicase activity (Figure 5) (Fiorini et al., 2015). m⁶A writers METTL3 and METTL14 are both able to modify RNA, but their activity is significantly enhanced by their mutual interaction (Śledź and Jinek, 2016). Non-protein ligands can also affect RBP binding, such as 4Fe-4S binding by IRP1, which activates its aconitase activity and mutually excludes iron response element binding (Figure 5) (Hentze et al., 2004). Similarly, ATP binding by helicases commonly induces “clamping” of the helicase domain, which increases its affinity for RNA (Gai et al., 2004; Linder and Jankowsky, 2011).

Post-translational modifications (PTMs) of residues in RBPs direct sophisticated regulation of their interaction sites and functions. RBPs often contain multiple sites for PTMs, the most common of which are phosphorylation (often serine), acetylation

(often lysine), and arginine methylation (Hofweber and Dormann, 2019; Lovci et al., 2016). At the molecular level, PTMs introduce electrostatic charges that can affect the structural stability of binding regions or their ability to interact with RNA or other proteins (Drazic et al., 2016; Law et al., 2003; Lovci et al., 2016). Phosphorylation and acetylation modifications introduce negative charge or neutralize positive charge, respectively. For example, acetylation of at least two lysine residues in the RRM of PTBP1/2 disrupts their ability to hydrogen bond with the RNA backbone, likely by eliminating electrostatic attraction between the positive lysine residues and the negatively charged RNA backbone (Pina et al., 2018). Arginine methylation, on the other hand, has been shown to decrease the favorability for cation- π interactions, such as the RGG-mediated interactions between IDRs that drive phase separation (Hofweber and Dormann, 2019). PTMs are most commonly observed modulating PPIs, such as the association between eIF4E and eIF4E-BP, which is dependent on hypo-phosphorylation of eIF4E-BP. Phosphorylation of at least two threonine residues in eIF4E-BP induces a structural change that buries its eIF4E binding site (Thapar, 2015), allowing eIF4E induction into the eIF4F complex instead (Figure 5).

Taking a less protein-centric view, RNA interactions and RNA modifications regulate RBP binding as well. m⁶A modifications were shown to ultimately determine whether binding sites are available to hnRNPC (Figure 5) (Liu et al., 2015). The small non-coding RNA BC1 inhibits eIF4A helicase activity, and long non-coding RNAs have been shown to act as RBP “sponges,” effectively reducing the abundance of an RBP for functional binding (HafezQorani et al., 2019; Linder and Jankowsky, 2011).

Conclusions

The current state of knowledge on RBPs is rapidly growing and includes many areas of study beyond the scope of this review: RNA binding site detection techniques, RBP synthetic design, and the role of RBPs in stress granules and neurodegenerative diseases, to name a few. All of these areas benefit from a solid mechanistic understanding of protein-RNA target interactions. In this review, we took a mechanistic look at how RBPs interact with RNA, both at the molecular level and bird’s eye views. We repeatedly observe that detailed molecular structures explain the binding behavior and function of RBPs, such as helicase domains preferring interactions with the RNA backbone or specific residues in a YTH m⁶A reader protein “locking in” a methylated base. Similarly, our analysis of hydrogen bonds formed with RNA for different RNA-binding domains reinforces the known structural function of several domains, but also shows how some domains are different from their peers. As the structures of more protein-RNA complexes are determined this analysis can be expanded to include more domain types, determining features that set domains apart or define their mechanism of binding. Additionally, studying RBP-substrate dynamics as well as larger multi-protein complexes is key for understanding the interactive process of protein-RNA regulation. We should note that the structural techniques used to study RNA-binding domains are less capable of capturing the full RNA substrates that RBPs bind. Thus, although we are learning a great deal about what

makes RBPs bind RNA, these techniques tell us much less about the reverse. Additional methods for the study of RNA molecules in complex with RBPs are needed. Finally, computational tools benefit greatly from our structural understanding of protein-RNA interaction mechanisms and, in turn, enable rapid insights for any RBP of interest.

Methods

PDB format files for structures of RNA-interacting proteins were downloaded from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) (Berman et al., 2000) according to protein domain type, using the search terms “KH domain,” “RRM,” “dsRBD,” “zinc finger,” “pumilio,” “DEAD,” “YTH domain,” and “cold shock domain” and further narrowed by selecting X-ray crystallography structures and NMR structures consisting of both protein and RNA. PDB files were processed with HBPLUS (McDonald and Thornton, 1994), which infers hydrogen-bond interactions between any two moieties, command:

```
echo file.pdb | clean #outputs to file.new  
hbplus -d 3.35 -h 2.7 file.new file.pdb #outputs to file.hb2.
```

All code used to assess protein-RNA hydrogen-bond information in hb2 files is available at https://github.com/meracorley/hbplus_tools. Briefly, for each hb2 file, hydrogen bonds occurring between protein and RNA, including those coordinated by water molecules, were stored and organized by moiety type (base, sugar, backbone, or side chain, main chain) as well as amino acid or base identity. To assess the frequency of each amino acid used in hydrogen-bond interactions with RNA for a given structure, the number of interactions involving each amino acid was divided by the total number of potential protein-RNA interactions. To assess the frequency of each RNA nucleotide used in hydrogen-bond interactions with protein, for a given structure the number of interactions involving each nucleotide was divided by the total number of potential protein-RNA interactions. The percentage of base|sugar|backbone|side-chain interactions was similarly calculated as the number of interactions of interest divided by the total number of potential interactions. The total number of hydrogen bonds per protein was counted per protein chain in a given structure and averaged over all chains. The total number of protein residues interacting with RNA was calculated as the count of unique interacting residues per protein chain in the given structure and averaged over all chains. p values were calculated by two-sided t tests.

ACKNOWLEDGMENTS

This work was supported by grants from the NIH (HG004659 and HD085902) (to G.W.Y.). M.C. was supported by an ALS Association Milton Safenowitz post-doctoral fellowship. The authors would like to thank Dr. Aaron Smargon for his input on protein-RNA structure analysis.

AUTHOR CONTRIBUTIONS

Conceptualization, M.C. and G.W.Y.; Formal Analysis, M.C.; Writing – Original Draft, M.C. and M.C.B.; Writing – Review & Editing, M.C. and G.W.Y.; Supervision, G.W.Y.

DECLARATION OF INTERESTS

G.W.Y. is co-founder, member of the board of directors, equity holder, and paid consultant for Locana and Eclipse BioInnovations. G.W.Y. is a Distinguished Visiting Professor at the National University of Singapore. The terms of this arrangement have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies. All other authors declare no competing interests.
Published: March 27, 2020

REFERENCES

- Akopian, D., Shen, K., Zhang, X., and Shan, S.O. (2013). Signal recognition particle: an essential protein-targeting machine. *Annu. Rev. Biochem.* **82**, 693–721.
- Allain, F.H., Howe, P.W., Neuhaus, D., and Varani, G. (1997). Structural basis of the RNA-binding specificity of human U1A protein. *EMBO J.* **16**, 5764–5772.
- Allers, J., and Shammoo, Y. (2001). Structure-based analysis of protein-RNA interactions using the program ENTANGLE. *J. Mol. Biol.* **317**, 75–86.
- Amir, M., Kumar, V., Dohare, R., Islam, A., Ahmad, F., and Hassan, M.I. (2018). Sequence, structure and evolutionary analysis of cold shock domain proteins, a member of OB fold family. *J. Evol. Biol.* **31**, 1903–1917.
- Auweter, S.D., Oberstrass, F.C., and Allain, F.H. (2006). Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res.* **34**, 4943–4959.
- Auweter, S.D., Oberstrass, F.C., and Allain, F.H.T. (2007). Solving the structure of PTB in complex with pyrimidine tracts: an NMR study of protein-RNA complexes of weak affinities. *J. Mol. Biol.* **367**, 174–186.
- Aviv, T., Lin, Z., Ben-Ari, G., Smibert, C.A., and Sicheri, F. (2006). Sequence-specific recognition of RNA hairpins by the SAM domain of Vts1p. *Nat. Struct. Mol. Biol.* **13**, 168–176.
- Balcerak, A., Trebinska-Stryjewska, A., Konopinski, R., Wakula, M., and Grzybowska, E.A. (2019). RNA-protein interactions: disorder, moonlighting and junk contribute to eukaryotic complexity. *Open Biol.* **9**, 190096.
- Barik, A., Mishra, A., and Bahadur, R.P. (2012). PRince: a web server for structural and physicochemical analysis of protein-RNA interface. *Nucleic Acids Res.* **40**, W440–W444.
- Barik, A., C. N., Pilla, S.P., and Bahadur, R.P. (2015). Molecular architecture of protein-RNA recognition sites. *J. Biomol. Struct. Dyn.* **33**, 2738–2751.
- Barraud, P., and Allain, F.H.T. (2012). ADAR proteins: double-stranded RNA and Z-DNA binding domains. *Curr. Top. Microbiol. Immunol.* **353**, 35–60.
- Bellucci, M., Agostini, F., Masin, M., and Tartaglia, G.G. (2011). Predicting protein associations with long noncoding RNAs. *Nat. Methods* **8**, 444–445.
- Bercy, M., and Bockelmann, U. (2015). Hairpins under tension: RNA versus DNA. *Nucleic Acids Res.* **43**, 9928–9936.
- Berman, H.M., Olson, W.K., Beveridge, D.L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.H., Srinivasan, A.R., and Schneider, B. (1992). The Nucleic Acid Database: a comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* **63**, 751–759.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242.
- Beusch, I., Barraud, P., Moursy, A., Cléry, A., and Allain, F.H. (2017). Tandem hnRNP A1 RNA recognition motifs act in concert to repress the splicing of survival motor neuron exon 7. *eLife* **6**, e25736.
- Bifsha, P., Landry, K., Ashmarina, L., Durand, S., Seyrantepe, V., Trudel, S., Quiniou, C., Chemtob, S., Xu, Y., Gravel, R.A., et al. (2007). Altered gene expression in cells from patients with lysosomal storage disorders suggests impairment of the ubiquitin pathway. *Cell Death Differ.* **14**, 511–523.
- Boland, A., Huntzinger, E., Schmidt, S., Izaurralde, E., and Weichenrieder, O. (2011). Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein. *Proc. Natl. Acad. Sci. USA* **108**, 10466–10471.
- Bousquet-Antonelli, C., and Deragon, J.M. (2009). A comprehensive analysis of the La-motif protein superfamily. *RNA* **15**, 750–764.
- Bressin, A., Schulte-Sasse, R., Figini, D., Urdaneta, E.C., Beckmann, B.M., and Marsico, A. (2019). TriPepSVM: de novo prediction of RNA-binding proteins based on short amino acid motifs. *Nucleic Acids Res.* **47**, 4406–4417.
- Brylinski, M. (2018). Aromatic interactions at the ligand-protein interface: implications for the development of docking scoring functions. *Chem. Biol. Drug Des.* **91**, 380–390.
- Castello, A., Fischer, B., Frese, C.K., Horos, R., Alleaume, A.M., Foehr, S., Curk, T., Krijgsveld, J., and Hentze, M.W. (2016). Comprehensive identification of RNA-binding domains in human cells. *Mol. Cell* **63**, 696–710.
- Chekanova, J.A., Dutko, J.A., Mian, I.S., and Belostotsky, D.A. (2002). *Arabidopsis thaliana* exosome subunit AtRrp4p is a hydrolytic 3′ → 5′ exonuclease containing S1 and KH RNA-binding domains. *Nucleic Acids Res.* **30**, 695–700.
- Chen, Y.C., Sargsyan, K., Wright, J.D., Huang, Y.S., and Lim, C. (2014). Identifying RNA-binding residues based on evolutionary conserved structural and energetic features. *Nucleic Acids Res.* **42**, e15.
- Cieniková, Z., Jayne, S., Damberger, F.F., Allain, F.H.T., and Maris, C. (2015). Evidence for cooperative tandem binding of hnRNP C RRM in mRNA processing. *RNA* **21**, 1931–1942.
- Cléry, A., and Allain, F.H.T. (2012). From structure to function of RNA binding domains. In *RNA Binding Proteins*, Z.J. Lorković, ed. (Landes Bioscience), pp. 137–158.
- Cléry, A., Blatter, M., and Allain, F.H. (2008). RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.* **18**, 290–298.
- Coimbatore Narayanan, B., Westbrook, J., Ghosh, S., Petrov, A.I., Sweeney, B., Zirbel, C.L., Leontis, N.B., and Berman, H.M. (2014). The Nucleic Acid Database: new features and capabilities. *Nucleic Acids Res.* **42**, D114–D122.
- Cook, K.B., Kazan, H., Zuberi, K., Morris, Q., and Hughes, T.R. (2011). RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res.* **39**, D301–D308.
- Cruz-Gallardo, I., Martino, L., Kelly, G., Atkinson, R.A., Trotta, R., De Tito, S., Coleman, P., Ahdash, Z., Gu, Y., Bui, T.T.T., and Conte, M.R. (2019). LARP4A recognizes polyA RNA via a novel binding mechanism mediated by disordered regions and involving the PAM2w motif, revealing interplay between PABP, LARP4A and mRNA. *Nucleic Acids Res.* **47**, 4272–4291.
- De Boulle, K., Verkerk, A.J., Reyniers, E., Vits, L., Hendrickx, J., Van Roy, B., Van den Bos, F., de Graaff, E., Oostra, B.A., and Willems, P.J. (1993). A point mutation in the FMR-1 gene associated with fragile X mental retardation. *Nat. Genet.* **3**, 31–35.
- Deng, L., Yang, W., and Liu, H. (2019). PredPRBA: prediction of protein-RNA binding affinity using gradient boosted regression trees. *Front. Genet.* **10**, 637.
- Dill, K.A., Ozkan, S.B., Shell, M.S., and Weikel, T.R. (2008). The protein folding problem. *Annu. Rev. Biophys.* **37**, 289–316.
- 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., et al. (2018). Sequence, structure, and context preferences of human RNA binding proteins. *Mol. Cell* **70**, 854–867.e9.
- Dzanic, A., Myklebust, L.M., Ree, R., and Arnesen, T. (2016). The world of protein acetylation. *Biochim. Biophys. Acta* **1864**, 1372–1401.
- Ellis, J.J., Broom, M., and Jones, S. (2007). Protein-RNA interactions: structural analysis and functional classes. *Proteins* **66**, 903–911.
- Fiorini, F., Bagchi, D., Le Hir, H., and Croquette, V. (2015). Human Upf1 is a highly processive RNA helicase and translocase with RNP remodelling activities. *Nat. Commun.* **6**, 7581.
- Fislage, M., Roovers, M., Tuszyńska, I., Bujnicki, J.M., Droogmans, L., and Versées, W. (2012). Crystal structures of the tRNA:m2G6 methyltransferase Trm14/TrmN from two domains of life. *Nucleic Acids Res.* **40**, 5149–5161.
- Flores, J.K., and Ataíde, S.F. (2018). Structural changes of RNA in complex with proteins in the SRP. *Front. Mol. Biosci.* **5**, 7.
- Font, J., and Mackay, J.P. (2010). Beyond DNA: zinc finger domains as RNA-binding modules. *Methods Mol. Biol.* **649**, 479–491.

- Frazão, C., McVey, C.E., Amblar, M., Barbas, A., Vornrhein, C., Arraiano, C.M., and Carrondo, M.A. (2006). Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex. *Nature* *443*, 110–114.
- Gai, D., Zhao, R., Li, D., Finkielstein, C.V., and Chen, X.S. (2004). Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. *Cell* *119*, 47–60.
- Ge, J., and Yu, Y.T. (2013). RNA pseudouridylation: new insights into an old modification. *Trends Biochem. Sci.* *38*, 210–218.
- Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. *Nat. Rev. Genet.* *15*, 829–845.
- Giudice, G., Sanchez-Cabo, F., Torroja, C., and Lara-Pezzi, E. (2016). ATTRACT—a database of RNA-binding proteins and associated motifs. *Database* *2016*, baw035.
- Glisovic, T., Bachorik, J.L., Yong, J., and Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.* *582*, 1977–1986.
- Gupta, A., and Gribskov, M. (2011). The role of RNA sequence and structure in RNA-protein interactions. *J. Mol. Biol.* *409*, 574–587.
- HafezQorani, S., Houdjedj, A., Arici, M., Said, A., and Kazan, H. (2019). RBP Sponge: genome-wide identification of lncRNAs that sponge RBPs. *Bioinformatics* *35*, 4760–4763.
- Hainzl, T., Huang, S., and Sauer-Eriksson, A.E. (2005). Structural insights into SRP RNA: an induced fit mechanism for SRP assembly. *RNA* *11*, 1043–1050.
- Hall, T.M. (2005). Multiple modes of RNA recognition by zinc finger proteins. *Curr. Opin. Struct. Biol.* *15*, 367–373.
- Han, K., and Nepal, C. (2007). PRI-Modeler: extracting RNA structural elements from PDB files of protein-RNA complexes. *FEBS Lett.* *581*, 1881–1890.
- Hentze, M.W., Muckenthaler, M.U., and Andrews, N.C. (2004). Balancing acts: molecular control of mammalian iron metabolism. *Cell* *117*, 285–297.
- Hentze, M.W., Castello, A., Schwarzl, T., and Preiss, T. (2018). A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* *19*, 327–341.
- Höck, J., and Meister, G. (2008). The Argonaute protein family. *Genome Biol.* *9*, 210.
- Hoffman, M.M., Khrapov, M.A., Cox, J.C., Yao, J., Tong, L., and Ellington, A.D. (2004). AANT: the Amino Acid-Nucleotide Interaction Database. *Nucleic Acids Res.* *32*, D174–D181.
- Hofweber, M., and Dormann, D. (2019). Friend or foe—post-translational modifications as regulators of phase separation and RNP granule dynamics. *J. Biol. Chem.* *294*, 7137–7150.
- Hossain, S.T., Malhotra, A., and Deutscher, M.P. (2016). How RNase R degrades structured RNA: role of the helicase activity and the S1 domain. *J. Biol. Chem.* *291*, 7877–7887.
- Hu, W., Qin, L., Li, M.L., Pu, X.M., and Guo, Y.Z. (2018). A structural dissection of protein-RNA interactions based on different RNA base areas of interfaces. *RSC Adv.* *8*, 10582–10592.
- Huang, Y., Liu, S., Guo, D., Li, L., and Xiao, Y. (2013). A novel protocol for three-dimensional structure prediction of RNA-protein complexes. *Sci. Rep.* *3*, 1887.
- Hudson, W.H., and Ortlund, E.A. (2014). The structure, function and evolution of proteins that bind DNA and RNA. *Nat. Rev. Mol. Cell Biol.* *15*, 749–760.
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* *14*, 33–38, 27–38.
- Iost, I., Dreyfus, M., and Linder, P. (1999). Ded1p, a DEAD-box protein required for translation initiation in *Saccharomyces cerevisiae*, is an RNA helicase. *J. Biol. Chem.* *274*, 17677–17683.
- Jackson, R.J., Hellen, C.U.T., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* *11*, 113–127.
- Jankowsky, E. (2011). RNA helicases at work: binding and rearranging. *Trends Biochem. Sci.* *36*, 19–29.
- Järvelin, A.I., Noerenberg, M., Davis, I., and Castello, A. (2016). The new (dis) order in RNA regulation. *Cell Commun. Signal.* *14*, 9.
- Jiang, F., Ramanathan, A., Miller, M.T., Tang, G.Q., Gale, M., Jr., Patel, S.S., and Marcotrigiano, J. (2011). Structural basis of RNA recognition and activation by innate immune receptor RIG-I. *Nature* *479*, 423–427.
- Jones, S., Daley, D.T.A., Luscombe, N.M., Berman, H.M., and Thornton, J.M. (2001). Protein-RNA interactions: a structural analysis. *Nucleic Acids Res.* *29*, 943–954.
- Jubb, H.C., Higuero, A.P., Ochoa-Montaño, B., Pitt, W.R., Ascher, D.B., and Blundell, T.L. (2017). Arpeggio: a web server for calculating and visualising interatomic interactions in protein structures. *J. Mol. Biol.* *429*, 365–371.
- Kainov, D.E., Pirttimaa, M., Tuma, R., Butcher, S.J., Thomas, G.J., Jr., Bamford, D.H., and Makeyev, E.V. (2003). RNA packaging device of double-stranded RNA bacteriophages, possibly as simple as hexamer of P4 protein. *J. Biol. Chem.* *278*, 48084–48091.
- Kainov, D.E., Mancini, E.J., Telenius, J., Lisal, J., Grimes, J.M., Bamford, D.H., Stuart, D.I., and Tuma, R. (2008). Structural basis of mechanochemical coupling in a hexameric molecular motor. *J. Biol. Chem.* *283*, 3607–3617.
- Kappel, K., and Das, R. (2019). Sampling native-like structures of RNA-protein complexes through Rosetta folding and docking. *Structure* *27*, 140–151.e5.
- Ke, J., Chen, R.Z., Ban, T., Zhou, X.E., Gu, X., Tan, M.H., Chen, C., Kang, Y., Brunzelle, J.S., Zhu, J.K., et al. (2013). Structural basis for RNA recognition by a dimeric PPR-protein complex. *Nat. Struct. Mol. Biol.* *20*, 1377–1382.
- Keffer-Wilkes, L.C., Veerareddygar, G.R., and Kothe, U. (2016). RNA modification enzyme TruB is a tRNA chaperone. *Proc. Natl. Acad. Sci. USA* *113*, 14306–14311.
- Kim, O.T.P., Yura, K., and Go, N. (2006). Amino acid residue doublet propensity in the protein-RNA interface and its application to RNA interface prediction. *Nucleic Acids Res.* *34*, 6450–6460.
- Kjashstomy, V., Nikonov, S., Ovchinnikov, L., Lyabin, D., Vodovar, N., Curmi, P., and Manivet, P. (2015). The cold shock domain of YB-1 segregates RNA from DNA by non-bonded interactions. *PLoS ONE* *10*, e0130318.
- Kumar, M., Gromiha, M.M., and Raghava, G.P.S. (2008). Prediction of RNA binding sites in a protein using SVM and PSSM profile. *Proteins* *71*, 189–194.
- Lai, W.S., Carballo, E., Thorn, J.M., Kennington, E.A., and Blackshear, P.J. (2000). Interactions of CCCH zinc finger proteins with mRNA. Binding of tristetraprolin-related zinc finger proteins to AU-rich elements and destabilization of mRNA. *J. Biol. Chem.* *275*, 17827–17837.
- Laskowski, R.A., Jablonska, J., Pravda, L., Vařeková, R.S., and Thornton, J.M. (2018). PDBsum: structural summaries of PDB entries. *Protein Sci.* *27*, 129–134.
- Law, L.M.J., Everitt, J.C., Beatch, M.D., Holmes, C.F.B., and Hobman, T.C. (2003). Phosphorylation of rubella virus capsid regulates its RNA binding activity and virus replication. *J. Virol.* *77*, 1764–1771.
- Leulliot, N., and Varani, G. (2001). Current topics in RNA-protein recognition: control of specificity and biological function through induced fit and conformational capture. *Biochemistry* *40*, 7947–7956.
- Li, S., Yamashita, K., Amada, K.M., and Standley, D.M. (2014). Quantifying sequence and structural features of protein-RNA interactions. *Nucleic Acids Res.* *42*, 10086–10098.
- Liao, S., Sun, H., and Xu, C. (2018). YTH domain: a family of N⁶-methyladenosine (m⁶A) readers. *Genomics Proteomics Bioinformatics* *16*, 99–107.
- Linder, P., and Jankowsky, E. (2011). From unwinding to clamping—the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.* *12*, 505–516.
- Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M., and Pan, T. (2015). N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* *518*, 560–564.
- Lorković, Z.J. (2012). RNA Binding Proteins (Landes Bioscience).
- Loughlin, F.E., Lukavsky, P.J., Kazeeva, T., Reber, S., Hock, E.M., Colombo, M., Von Schroetter, C., Pauli, P., Cléry, A., Mühlemann, O., et al. (2019). The

- solution structure of FUS bound to RNA reveals a bipartite mode of RNA recognition with both sequence and shape specificity. *Mol. Cell* 73, 490–504.e6.
- Lovci, M.T., Bengtson, M.H., and Massire, K.B. (2016). Post-translational modifications and RNA-binding proteins. *Adv. Exp. Med. Biol.* 907, 297–317.
- Loveland, A.B., and Korostelev, A.A. (2018). Structural dynamics of protein S1 on the 70S ribosome visualized by ensemble cryo-EM. *Methods* 137, 55–66.
- Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell Biol.* 8, 479–490.
- Luscombe, N.M., Laskowski, R.A., and Thornton, J.M. (1997). NUCPLOT: a program to generate schematic diagrams of protein-nucleic acid interactions. *Nucleic Acids Res.* 25, 4940–4945.
- Luscombe, N.M., Laskowski, R.A., and Thornton, J.M. (2001). Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.* 29, 2860–2874.
- Ma, J.B., Ye, K., and Patel, D.J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* 429, 318–322.
- Ma, J.B., Yuan, Y.R., Meister, G., Pei, Y., Tuschl, T., and Patel, D.J. (2005). Structural basis for 5'-end-specific recognition of guide RNA by the A. *fulgidus* Piwi protein. *Nature* 434, 666–670.
- Maris, C., Dominguez, C., and Allain, F.H. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* 272, 2118–2131.
- Masliah, G., Barraud, P., and Allain, F.H. (2013). RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cell. Mol. Life Sci.* 70, 1875–1895.
- Matthews, M.M., Thomas, J.M., Zheng, Y., Tran, K., Phelps, K.J., Scott, A.I., Havel, J., Fisher, A.J., and Beal, P.A. (2016). Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nat. Struct. Mol. Biol.* 23, 426–433.
- McDonald, I.K., and Thornton, J.M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* 238, 777–793.
- Mihailovich, M., Miliuti, C., Gabaldón, T., and Gebauer, F. (2010). Eukaryotic cold shock domain proteins: highly versatile regulators of gene expression. *BioEssays* 32, 109–118.
- Miyoshi, T., Ito, K., Murakami, R., and Uchiumi, T. (2016). Structural basis for the recognition of guide RNA and target DNA heteroduplex by Argonaute. *Nat. Commun.* 7, 11846.
- Morozova, N., Allers, J., Myers, J., and Shamoo, Y. (2006). Protein-RNA interactions: exploring binding patterns with a three-dimensional superposition analysis of high resolution structures. *Bioinformatics* 22, 2746–2752.
- Motorin, Y., and Helm, M. (2011). RNA nucleotide methylation. *Wiley Interdiscip. Rev. RNA* 2, 611–631.
- Muppurala, U.K., Honavar, V.G., and Dobbs, D. (2011). Predicting RNA-protein interactions using only sequence information. *BMC Bioinformatics* 12, 489.
- Nagarajan, R., and Gromiha, M.M. (2014). Prediction of RNA binding residues: an extensive analysis based on structure and function to select the best predictor. *PLoS ONE* 9, e91140.
- Neumann, P., Lakomek, K., Naumann, P.T., Erwin, W.M., Lauhon, C.T., and Ficner, R. (2014). Crystal structure of a 4-thiouridine synthetase-RNA complex reveals specificity of tRNA U8 modification. *Nucleic Acids Res.* 42, 6673–6685.
- Niedzwiecka, A., Marcotrigiano, J., Stepinski, J., Jankowska-Anyszka, M., Wyslouch-Cieszyńska, A., Dadlez, M., Gingras, A.C., Mak, P., Darzynkiewicz, E., Sonenberg, N., et al. (2002). Biophysical studies of eIF4E cap-binding protein: recognition of mRNA 5' cap structure and synthetic fragments of eIF4G and 4E-BP1 proteins. *J. Mol. Biol.* 319, 615–635.
- Onofrio, A., Parisi, G., Punzi, G., Todisco, S., Di Noia, M.A., Bossi, F., Turi, A., De Grassi, A., and Pierrì, C.L. (2014). Distance-dependent hydrophobic-hydrophobic contacts in protein folding simulations. *Phys. Chem. Chem. Phys.* 16, 18907–18917.
- Oubridge, C., Ito, N., Evans, P.R., Teo, C.H., and Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* 372, 432–438.
- Paz, I., Kosti, I., Ares, M., Jr., Ciine, M., and Mandel-Gutfreund, Y. (2014). RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res.* 42, W361–W367.
- Pérez-Arellano, I., Gallego, J., and Cervera, J. (2007). The PUA domain—a structural and functional overview. *FEBS J.* 274, 4972–4984.
- Pérez-Cano, L., and Fernández-Recio, J. (2010). Optimal protein-RNA area, OPRA: a propensity-based method to identify RNA-binding sites on proteins. *Proteins* 78, 25–35.
- Pina, J.M., Reynaga, J.M., Truong, A.A.M., and Keppetipola, N.M. (2018). Post-translational modifications in polypyrimidine tract binding proteins PTBP1 and PTBP2. *Biochemistry* 57, 3873–3882.
- Rajkowitsch, L., Chen, D., Stampfl, S., Semrad, K., Waldsich, C., Mayer, O., Jantsch, M.F., Konrat, R., Bläsi, U., and Schroeder, R. (2007). RNA chaperones, RNA annealers and RNA helicases. *RNA Biol.* 4, 118–130.
- Ramos, A., Grünert, S., Adams, J., Micklem, D.R., Proctor, M.R., Freund, S., Bycroft, M., St Johnston, D., and Varani, G. (2000). RNA recognition by a Staufen double-stranded RNA-binding domain. *EMBO J.* 19, 997–1009.
- Ranji, A., Shkriabai, N., Kvaratskhelia, M., Musier-Forsyth, K., and Boris-Lawrie, K. (2011). Features of double-stranded RNA-binding domains of RNA helicase A are necessary for selective recognition and translation of complex mRNAs. *J. Biol. Chem.* 286, 5328–5337.
- Ravindranathan, S., Oberstrass, F.C., and Allain, F.H.T. (2010). Increase in backbone mobility of the VTS1p-SAM domain on binding to SRE-RNA. *J. Mol. Biol.* 396, 732–746.
- Sachs, R., Max, K.E.A., Heinemann, U., and Balbach, J. (2012). RNA single strands bind to a conserved surface of the major cold shock protein in crystals and solution. *RNA* 18, 65–76.
- Safaei, N., Kozlov, G., Noronha, A.M., Xie, J., Wilds, C.J., and Gehring, K. (2012). Interdomain allostery promotes assembly of the poly(A) mRNA complex with PABP and eIF4G. *Mol. Cell* 48, 375–386.
- Salentin, S., Schreiber, S., Haupt, V.J., Adasme, M.F., and Schroeder, M. (2015). PLIP: fully automated protein-ligand interaction profiler. *Nucleic Acids Res.* 43, W443–W447.
- Schenk, L., Meinel, D.M., Strässer, K., and Gerber, A.P. (2012). La-motif-dependent mRNA association with Sif1 promotes copper detoxification in yeast. *RNA* 18, 449–461.
- Schumacher, M.A., Pearson, R.F., Möller, T., Valentin-Hansen, P., and Brennan, R.G. (2002). Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J.* 21, 3546–3556.
- Sharma, M., and Anirudh, C.R. (2017). Mechanism of mRNA-STAR domain interaction: molecular dynamics simulations of mammalian quaking STAR protein. *Sci. Rep.* 7, 12567.
- Shulman-Peleg, A., Shatsky, M., Nussinov, R., and Wolfson, H.J. (2008). Prediction of interacting single-stranded RNA bases by protein-binding patterns. *J. Mol. Biol.* 379, 299–316.
- Śledź, P., and Jinek, M. (2016). Structural insights into the molecular mechanism of the m(6)A writer complex. *eLife* 5, e18434.
- Späth, H., Chia, T., Lingford, J.P., Siira, S.J., Cohen, S.B., Filipovska, A., and Rackham, O. (2018). Modular ssDNA binding and inhibition of telomerase activity by designer PPR proteins. *Nat. Commun.* 9, 2212.
- Stefl, R., Oberstrass, F.C., Hood, J.L., Jourdan, M., Zimmermann, M., Skrisovska, L., Maris, C., Peng, L., Hofr, C., Emeson, R.B., and Allain, F.H. (2010). The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. *Cell* 143, 225–237.
- Tan, D., Marzluff, W.F., Dominski, Z., and Tong, L. (2013). Structure of histone mRNA stem-loop, human stem-loop binding protein, and 3'hExo ternary complex. *Science* 339, 318–321.

- Tan, D., Zhou, M., Kiledjian, M., and Tong, L. (2014). The ROQ domain of Roquin recognizes mRNA constitutive-decay element and double-stranded RNA. *Nat. Struct. Mol. Biol.* *21*, 679–685.
- Tanzer, A., Hofacker, I.L., and Lorenz, R. (2019). RNA modifications in structure prediction—status quo and future challenges. *Methods* *156*, 32–39.
- Teichmann, M., Dumay-Odelot, H., and Fribourg, S. (2012). Structural and functional aspects of winged-helix domains at the core of transcription initiation complexes. *Transcription* *3*, 2–7.
- Teplova, M., Yuan, Y.R., Phan, A.T., Malinina, L., Ilin, S., Teplov, A., and Patel, D.J. (2006). Structural basis for recognition and sequestration of UUU(OH) 3' termini of nascent RNA polymerase III transcripts by La, a rheumatic disease autoantigen. *Mol. Cell* *21*, 75–85.
- Teplova, M., Malinina, L., Darnell, J.C., Song, J., Lu, M., Abagyan, R., Musunuru, K., Teplov, A., Burley, S.K., Darnell, R.B., and Patel, D.J. (2011). Protein-RNA and protein-protein recognition by dual KH1/2 domains of the neuronal splicing factor Nova-1. *Structure* *19*, 930–944.
- Teplova, M., Hafner, M., Teplov, D., Essig, K., Tuschl, T., and Patel, D.J. (2013). Structure-function studies of STAR family quaking proteins bound to their in vivo RNA target sites. *Genes Dev.* *27*, 928–940.
- Terribilini, M., Sander, J.D., Lee, J.H., Zaback, P., Jernigan, R.L., Honavar, V., and Dobbs, D. (2007). RNABindR: a server for analyzing and predicting RNA-binding sites in proteins. *Nucleic Acids Res.* *35*, W578–W584.
- Thapar, R. (2015). Structural basis for regulation of RNA-binding proteins by phosphorylation. *ACS Chem. Biol.* *10*, 652–666.
- Thore, S., Mayer, C., Sauter, C., Weeks, S., and Suck, D. (2003). Crystal structures of the *Pyrococcus abyssi* Sm core and its complex with RNA. Common features of RNA binding in archaea and eukarya. *J. Biol. Chem.* *278*, 1239–1247.
- Tian, Y., Simanshu, D.K., Ma, J.B., and Patel, D.J. (2011). Structural basis for piRNA 2'-O-methylated 3'-end recognition by Piwi PAZ (Piwi/Argonaute/Zwille) domains. *Proc. Natl. Acad. Sci. USA* *108*, 903–910.
- Treger, M., and Westhof, E. (2001). Statistical analysis of atomic contacts at RNA-protein interfaces. *J. Mol. Recognit.* *14*, 199–214.
- Valverde, R., Edwards, L., and Regan, L. (2008). Structure and function of KH domains. *FEBS J.* *275*, 2712–2726.
- Vuković, L., Koh, H.R., Myong, S., and Schulten, K. (2014). Substrate recognition and specificity of double-stranded RNA binding proteins. *Biochemistry* *53*, 3457–3466.
- Walden, W.E., Selezneva, A.I., Dupuy, J., Volbeda, A., Fontecilla-Camps, J.C., Theil, E.C., and Volz, K. (2006). Structure of dual function iron regulatory protein 1 complexed with ferritin IRE-RNA. *Science* *314*, 1903–1908.
- Walden, W.E., Selezneva, A., and Volz, K. (2012). Accommodating variety in iron-responsive elements: crystal structure of transferrin receptor 1 B IRE bound to iron regulatory protein 1. *FEBS Lett.* *586*, 32–35.
- Wang, Z., Hartman, E., Roy, K., Chanfreau, G., and Feigon, J. (2011). Structure of a yeast RNase III dsRBD complex with a noncanonical RNA substrate provides new insights into binding specificity of dsRBDs. *Structure* *19*, 999–1010.
- Wang, M., Ogé, L., Perez-Garcia, M.D., Hamama, L., and Sakr, S. (2018). The PUF protein family: overview on PUF RNA targets, biological functions, and post transcriptional regulation. *Int. J. Mol. Sci.* *19*, 410.
- Weir, J.R., Bonneau, F., Hentschel, J., and Conti, E. (2010). Structural analysis reveals the characteristic features of Mtr4, a DExH helicase involved in nuclear RNA processing and surveillance. *Proc. Natl. Acad. Sci. USA* *107*, 12139–12144.
- Wilson, K.A., Holland, D.J., and 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*, 696–708.
- Worbs, M., Bourenkov, G.P., Bartunik, H.D., Huber, R., and Wahl, M.C. (2001). An extended RNA binding surface through arrayed S1 and KH domains in transcription factor NusA. *Mol. Cell* *7*, 1177–1189.
- Xu, C., Wang, X., Liu, K., Roundtree, I.A., Tempel, W., Li, Y., Lu, Z., He, C., and Min, J. (2014). Structural basis for selective binding of m⁶A RNA by the YTHDC1 YTH domain. *Nat. Chem. Biol.* *10*, 927–929.
- Yan, J., and Kurgan, L. (2017). DRNAPred, fast sequence-based method that accurately predicts and discriminates DNA- and RNA-binding residues. *Nucleic Acids Res.* *45*, e84.
- Yan, K.S., Yan, S., Farooq, A., Han, A., Zeng, L., and Zhou, M.M. (2003). Structure and conserved RNA binding of the PAZ domain. *Nature* *426*, 468–474.
- Yang, Y., Declerck, N., Manival, X., Aymerich, S., and Kochoyan, M. (2002). Solution structure of the LicT-RNA antitermination complex: CAT clamping RAT. *EMBO J.* *21*, 1987–1997.
- Yang, J., Roy, A., and Zhang, Y. (2013). BioLiP: a semi-manually curated database for biologically relevant ligand-protein interactions. *Nucleic Acids Res.* *41*, D1096–D1103.
- Yang, Y.D., Zhao, H.Y., Wang, J.H., and Zhou, Y.Q. (2014). SPOT-Seq-RNA: predicting protein-RNA complex structure and RNA-binding function by fold recognition and binding affinity prediction. In *Protein Structure Prediction, Third Edition, Methods in Molecular Biology (Methods and Protocols)*, 1137, D. Kihara, ed. (Humana Press), pp. 119–130.
- Yang, L., Wang, C., Li, F., Zhang, J., Nayab, A., Wu, J., Shi, Y., and Gong, Q. (2017). The human RNA-binding protein and E3 ligase MEX-3C binds the MEX-3-recognition element (MRE) motif with high affinity. *J. Biol. Chem.* *292*, 16221–16234.
- Yi, Y., Zhao, Y., Li, C., Zhang, L., Huang, H., Li, Y., Liu, L., Hou, P., Cui, T., Tan, P., et al. (2017). RAID v2.0: an updated resource of RNA-associated interactions across organisms. *Nucleic Acids Res.* *45*, D115–D118.
- Yu, Q., Ye, W., Jiang, C., Luo, R., and Chen, H.F. (2014). Specific recognition mechanism between RNA and the KH3 domain of Nova-2 protein. *J. Phys. Chem. B* *118*, 12426–12434.
- Yu, H., Wang, J., Sheng, Q., Liu, Q., and Shyr, Y. (2019). beRBP: binding estimation for human RNA-binding proteins. *Nucleic Acids Res.* *47*, e26.
- Zhang, J., Ma, Z., and Kurgan, L. (2019). Comprehensive review and empirical analysis of hallmarks of DNA-, RNA- and protein-binding residues in protein chains. *Brief. Bioinform.* *20*, 1250–1268.
- Zhao, H., Yang, Y., and Zhou, Y. (2011). Structure-based prediction of RNA-binding domains and RNA-binding sites and application to structural genomics targets. *Nucleic Acids Res.* *39*, 3017–3025.
- Zhao, Y.Y., Mao, M.W., Zhang, W.J., Wang, J., Li, H.T., Yang, Y., Wang, Z., and Wu, J.W. (2018). Expanding RNA binding specificity and affinity of engineered PUF domains. *Nucleic Acids Res.* *46*, 4771–4782.
- Zhu, T., Roundtree, I.A., Wang, P., Wang, X., Wang, L., Sun, C., Tian, Y., Li, J., He, C., and Xu, Y. (2014). Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N⁶-methyladenosine. *Cell Res.* *24*, 1493–1496.