Biological and Medical Physics, Biomedical Engineering

Chirlmin Joo
David Rueda *Editors*

Biophysics of RNA-Protein Interactions

A Mechanistic View

Springer

Biological and Medical Physics, Biomedical Engineering

BIOLOGICAL AND MEDICAL PHYSICS, BIOMEDICAL ENGINEERING

This series is intended to be comprehensive, covering a broad range of topics important to the study of the physical, chemical and biological sciences. Its goal is to provide scientists and engineers with textbooks, monographs, and reference works to address the growing need for information. The fields of biological and medical physics and biomedical engineering are broad, multidisciplinary and dynamic. They lie at the crossroads of frontier research in physics, biology, chemistry, and medicine.

Books in the series emphasize established and emergent areas of science including molecular, membrane, and mathematical biophysics; photosynthetic energy harvesting and conversion; information processing; physical principles of genetics; sensory communications; automata networks, neural networks, and cellular automata. Equally important is coverage of applied aspects of biological and medical physics and biomedical engineering such as molecular electronic components and devices, biosensors, medicine, imaging, physical principles of renewable energy production, advanced prostheses, and environmental control and engineering.

More information about this series at http://www.springer.com/series/3740

Chirlmin Joo · David Rueda
Editors

Biophysics of RNA-Protein Interactions

A Mechanistic View

Springer

Editors
Chirlmin Joo
Kavli Institute of NanoScience
Delft University of Technology
Delft, Zuid-Holland, The Netherlands

David Rueda
Department of Medicine
Imperial College London
London, UK

ISSN 1618-7210 ISSN 2197-5647 (electronic)
Biological and Medical Physics, Biomedical Engineering
ISBN 978-1-4939-9724-4 ISBN 978-1-4939-9726-8 (eBook)
https://doi.org/10.1007/978-1-4939-9726-8

Contents

Part I
RNA Binding Proteins

Chapter 1
How Proteins Recognize RNA

Rajan Lamichhane

1.1 Introduction

According to the central dogma of molecular biology, genetic information is transformed from DNA to RNA during a process called transcription [1]. In eukaryotes, after transcription, the pre-mRNA undergoes several processing events including 5′ end capping, splicing, editing, and 3′ end polyadenylation before entering the ribosome for protein synthesis (Fig. 1.1). RNA has various structural, catalytic, and regulatory roles in the cell [2]. Perhaps in the cell, most functional RNAs interact with proteins to carry out functions, such as processing, nuclear export, transport, translation, modification, RNA stabilization, and localization [2–7]. For example, during posttranscriptional regulation of gene expression, RNA interacts directly with proteins to form ribonucleoprotein particles (RNPs) [8–10]. These RNPs are important for recognition of specific sequence elements present in RNA to control the function of the RNA molecule [9, 11]. Since there are many RNAs and a vast number of RNA-binding proteins, the biogenesis of RNPs must be performed with high fidelity. Incorrect formation of RNP complexes or aberrant expression of RNA-binding proteins can cause genetic disorders that may lead to diseases, such as neuromuscular and neurodegenerative disorders and cancers [12–18]. Therefore, understanding the molecular mechanism of protein–RNA interactions and their applications to function is an important aspect of structural and biological research [17, 19].

RNA molecules can adopt different secondary and tertiary structures from standard Watson–Crick base pairs to non-canonical base pairs, creating a platform that

This chapter is based on a dissertation submitted in 2011 (Department of Chemistry, Wayne State University, Detroit MI).

R. Lamichhane (✉)
Department of Biochemistry & Cellular and Molecular Biology, University of Tennessee, Knoxville, TN 37996, USA
e-mail: rajan@utk.edu

C. Joo and D. Rueda (eds.), *Biophysics of RNA-Protein Interactions*,
Biological and Medical Physics, Biomedical Engineering,
https://doi.org/10.1007/978-1-4939-9726-8_1

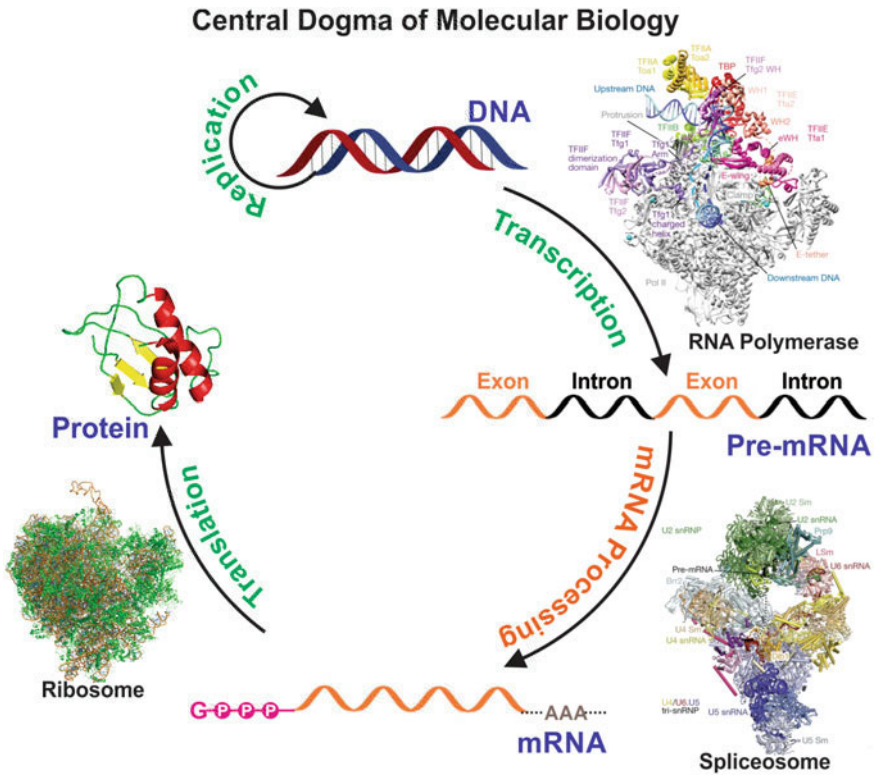

Fig. 1.1 Central dogma of molecular biology representing the general cellular processes in eukaryotic cells. DNA replicates its information and creates new copies of DNA during the process of replication. In eukaryotes, RNA polymerase transcribes DNA information into pre-mRNA, which undergoes RNA processing with the help of spliceosomes. Finally, ribosome translates RNA information into a protein. Structures are reprinted from the following: RNA polymerase II initiation complex is adapted from Plaschka et al. [20], with permission from Springer Nature; structure of a pre-catalytic spliceosome is adapted from Plaschka et al. [21], with permission from Springer Nature; structure of the human 80S ribosome is generated using PyMOL (PDB:4UG0) from Khatter et al. [22]

allows for interaction with a wide variety of ligands. These structures include single-stranded RNA (ssRNA), double-stranded RNA (dsRNA), hairpin loops, bulge loops, internal loops, junction loops, kink turn, and pseudoknots and are recognized by various proteins to form protein–RNA complexes (Fig. 1.2) [23]. These protein–RNA complexes have a wide variety of structural and functional roles in the cell [5, 7].

Despite their functional importance in biology, the actual mechanisms of protein–RNA interactions are poorly understood. Over the last decades, much work has been done to understand the structural and functional relationships of different types of protein–RNA interactions [4, 5, 19, 24, 25, 26]. Several biophysical methods have been used to characterize protein–RNA interactions. For example, X-ray

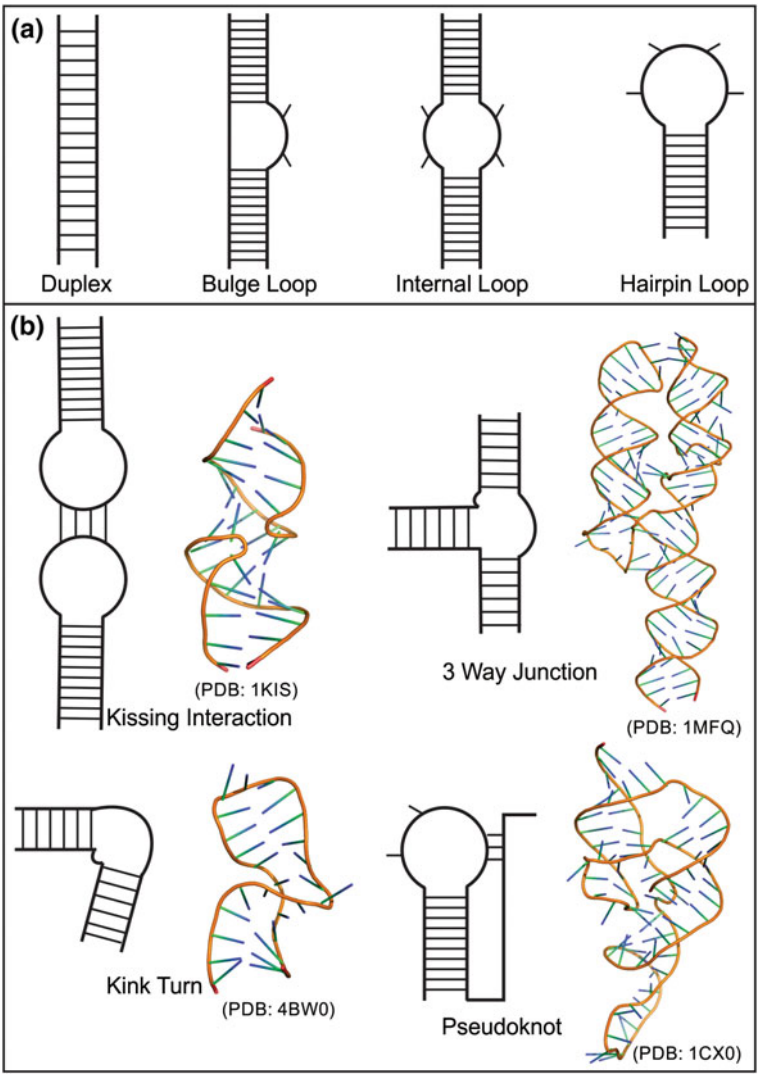

Fig. 1.2 Common RNA secondary structures and tertiary interactions. **a** Two-dimensional representation of common RNA secondary structural motifs (duplex RNA, bulge loop, internal loop, and hairpin loop). **b** Common RNA tertiary structural motifs and interactions with examples. Three-dimensional examples are generated using PyMOL and PDB files as mentioned: [kissing interaction (PDB: 1KIS); three-way junction (PDB: 1MFQ); kink turn (PDB: 4BW0); and pseudoknot (PDB: 1CX0)]

crystallography can be useful to obtain information concerning the detailed molecular interactions of a structured system, while cryo-electron microscopy (cryo-EM) can provide the overall shape of a protein–RNA complex. However, both of these methods have certain restrictions for a system with conformational flexibility and structural heterogeneity [19, 27, 28]. Recent advances have made nuclear magnetic resonance (NMR) one of the best techniques to study protein–RNA interactions in solution by using specific isotope labeling strategies. Coupling of NMR with complimentary small-angle X-ray scattering (SAXS) and electron paramagnetic resonance (EPR) is very helpful to solve larger protein–RNA complexes [27, 29, 30, 31]. Several solution-based protein–RNA structures have been reported in the Protein Database (PDB) [27]. Furthermore, computational modeling has also added insight into the structural analysis of protein–RNA complexes on the basis of different experimental interpretations [27, 32, 33]. The recent advancements of single-molecule spectroscopic techniques have added an effort to understand both the structural and the dynamic behaviors of protein–RNA interactions [34–37].

In this review, a comparison of structural and functional aspects of important known RNA-binding proteins will be discussed. Some important examples of common RNA-binding domains are summarized in Table 1.1 with their PDB entry numbers as an example.

Table 1.1 General properties and examples of common RNA-binding domains. The table is modified from [4], with permission from Springer Nature

Domain	Topology	RNA recognition motif	Protein interaction	Examples (PDB ID)
RRM	βαββαβ	β sheet makes a flat, solvent-exposed RNA-binding surface	Interacts with ssRNA through stacking, electrostatic interactions, and hydrogen bonding	PTB (2ADC) [38] Fox-1 (2ERR) [39]
KH	βααββα αββααβ	A cleft formed by GXXG loop and variable loop	Recognizes at least four nucleotides of ssRNA through hydrophobic interactions, backbone contacts from the loop, and hydrogen bonding with bases	Nova-1 (1EC6) [40] NusA (2ATW) [41]
TRAP	β-sandwich	Edges of β-strand	Bind GAG triplet through protein–base interactions, stacking, or hydrogen bonding	TRAP (1C9S) [42]

(continued)

Table 1.1 (continued)

Domain	Topology	RNA recognition motif	Protein interaction	Examples (PDB ID)
Sm/LSm proteins	αββββββ	Loops formed by β2-β3 and β4-β5	Recognizes poly U of ssRNA through stacking and hydrogen bonding	Sm core protein (1M8 V) [43], Hfq (1KQ2) [44]
Pumilio homology	α	Helix α2 provides the RNA interacting pocket	Stacking interactions and two amino acids in α2 make hydrogen bonds with Watson–Crick edge of a base	Pumilio 1 (1M8Y) [45], Nop9 (5WTY) [46]
Zinc finger	αβ	Amino acid residues in α helices	Sequence-specific (UAUU-TIS11d [47, 48]), hydrogen bonding to the protein backbone, and shape determine the specificity	TIS11D (1RGO) [48] MBNL (5U6H and 5U6L) [49]
PAZ	αβ (β-barrel)	Hydrophobic pocket formed by β-barrel and inserted αβ motif	Single-stranded RNA (ssRNA), and the 5′-phosphate and 3′-OH contribute to specificity	PAZ (1SI3) [50] Argonaute 2 (4OLA)[51]
dsRBM	αβββα	α1 helix and β1-β2 loop	Shape-specific recognition of RNA minor groove of A-form helix (stem-loop), and sequence-specific (G-X_n-A/G) contact with the 2′-OH of sugar and phosphate backbone	ADAR2 (2L3C) [52] Staufen (1EKZ) [53]
SAM	αααααα	Hydrophobic core packed with electropositive regions	Shape-specific recognition of RNA stem-loop, and interaction with phosphate backbone and a single nucleotide G at position 3 of the pentaloop	Vts1p (2ESE) [54]

1.2 RNA-Binding Proteins Are Modular

Most RNA-binding proteins have a modular structure formed by RNA-binding domains. These RNA-binding domains are encoded by sequences of 70–150 amino acids that are important for RNA recognition and interaction [4, 55, 56]. Most of the RNA-binding proteins (RBPs) consist of one or more RNA-binding domains (Fig. 1.3). These include the RNA-binding domain (RBD), most abundant and often called RNA recognition motif (RRM); K-homology (KH) domain; zinc finger (ZnF); Pumilio/FBF (PUF) domain; Piwi/Argonaute/Zwille (PAZ); sterile alpha motif (SAM) domain; double-stranded RNA-binding domain (dsRBD); DEAD box

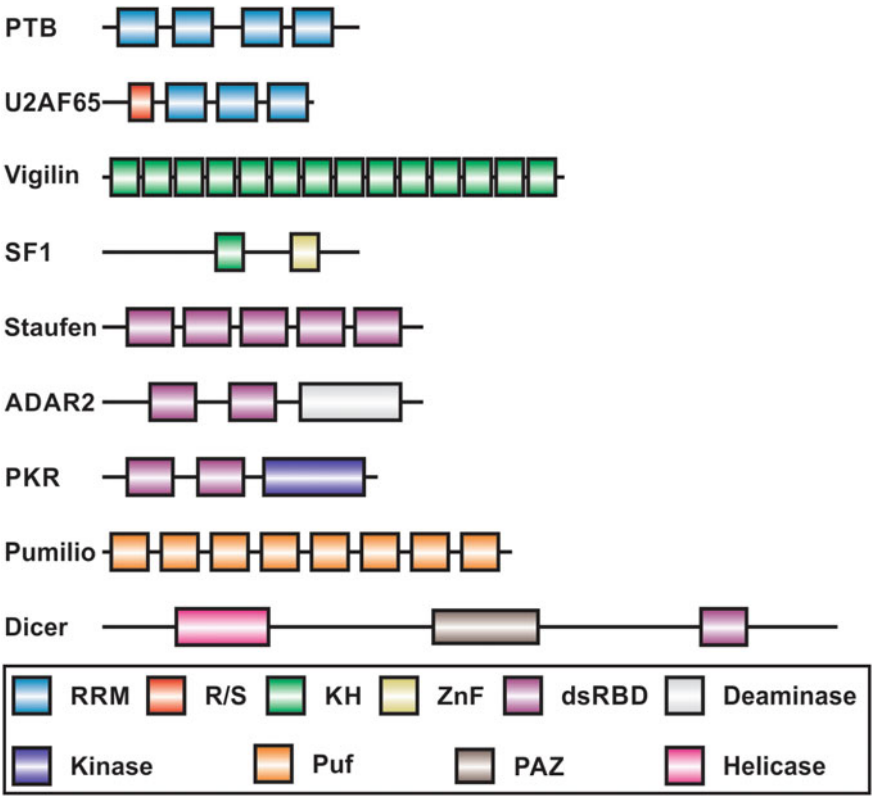

Fig. 1.3 Different modular structures of RNA-binding proteins (RBPs). Examples are taken from the most common RBPs. Each RBP contains many domains as shown by the colored boxes. These include RNA recognition motif (RRM), K-homology (KH) domain, RNA-binding zinc finger (ZnF), double-stranded RNA-binding domain (dsRBD), Puf RNA-binding repeats (Puf), and Piwi/Argonaute/Zwille (PAZ) domain. PTB, polypyrimidine tract binding; R/S, arginine/serine-rich domain; SF1, splicing factor-1; PKR, protein kinase R; U2AF, U2 auxiliary factor; and ADAR, adenosine deaminase. The figure is modified from [4], with permission from Springer Nature (the figure is not drawn to scale)

helicase domain (DDX); and the Sm domain. These modular architectures allow RBPs to recognize RNA with high specificity and affinity, as well as create functional diversity within the RBPs [2, 4, 57]. Proteins with multiple domains can bind long RNA strands or also interact with multiple RNAs; furthermore, modulation of RNA-binding domains with other auxiliary functional domains helps to recognize RNA as well as perform the enzymatic activity. For example, adenosine deaminases that act on RNA 2 (ADAR2) and protein kinase R (PKR) have similar dsRBD but different auxiliary functional domains. ADAR2 converts adenosine to inosine, while PKR has a kinase activity in its target RNA [58, 59].

Frequently, RNA-binding domains are connected with interdomain linkers of variable length. The importance of these linkers is in recognition of the discrete target, and they may act as spacers to regulate the catalytic action of each domain [4]. In some cases, linkers can interact with the RNA-binding domains to allow two domains to function synergistically as observed in polypyrimidine tract-binding protein domains 3 and 4 (PTB34) [35, 38]. Eukaryotic genomes have been shown to have higher numbers of modular RBPs, which might reflect the evolution of highly specific gene expression and modification patterns [2, 7, 9, 60].

1.3 Single-Stranded RNA Recognition

In most cases, RNA-binding proteins (RBPs) recognize ssRNA as their target. Many ssRNA-binding domains have been identified and have been shown to recognize RNA by conserved RNA-binding domains (RRM and KH) and by repeats of RNA-binding domains (TRAP and Sm). The oligonucleotide-/oligosaccharide-binding protein (OB-fold) domains recognize structured RNAs [61]. Many of ssRBPs are sequence-specific RNA-binding proteins with a hydrophobic binding surface to maximize intermolecular contacts with the RNA bases. The most common ssRBPs and their structures are discussed in detail.

1.3.1 RNA Recognition Motifs (RRMs)

The RNA recognition motif (RRM) domain is the most abundant and the best-characterized RNA-binding domain in higher eukaryotes. These domains, also known as ribonucleoprotein (RNP) domain or RNA-binding domain (RBD), consist of 80–100 amino acid residues [57, 62] and are often found in multiple copies. Single RRMs recognize a minimum of two to a maximum of eight nucleotides in the RNA [63, 64]. RRM has four antiparallel β-sheets packed against two α-helices with a topology of βαββαβ (Fig. 1.4a, b). An unusual fifth β-strand is present in RRM3 of polypyrimidine tract-binding protein (PTB) (Fig. 1.4c) [38, 65]. Most of the studied structures of RRM protein in complex with RNA have led to two proposed

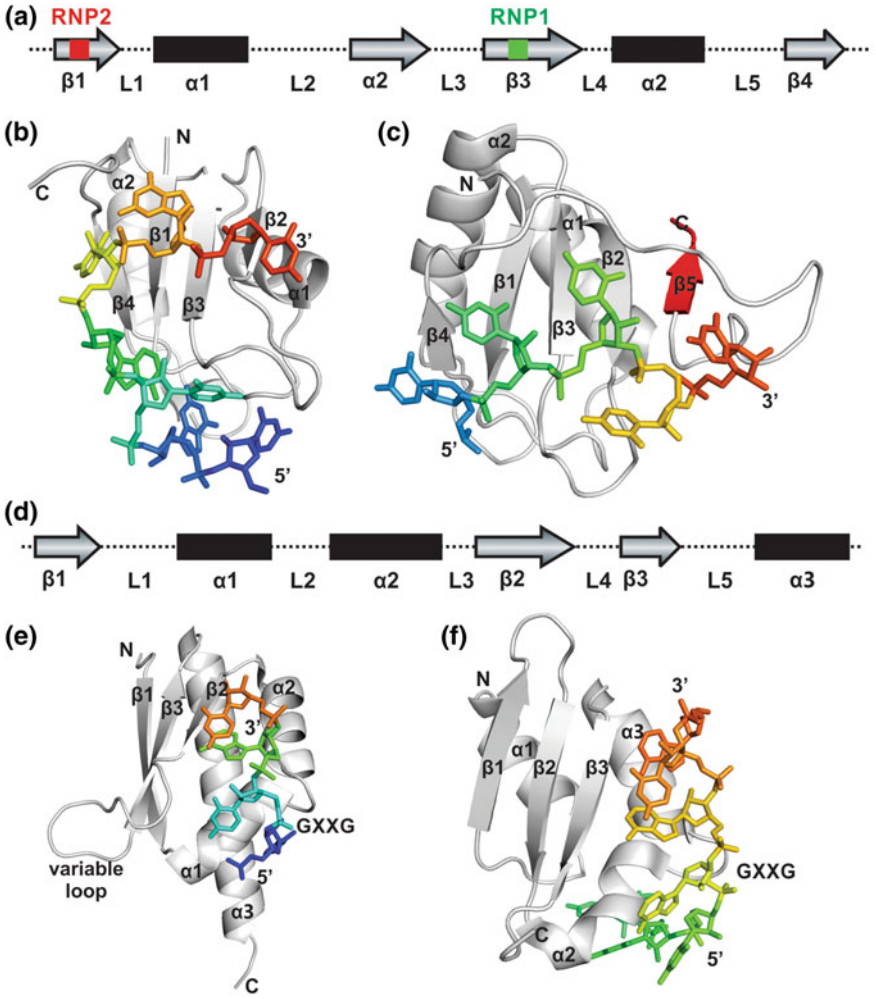

Fig. 1.4 Structures for common single-stranded RNA-binding protein RRM and KH domains. **a** The secondary structure for RRM domain with conserved sequences RNP2 (red) and RNP1 (green). **b** The RRM for Fox-1 domains (PDB: 2ERR). **c** The RRM domain 3 of PTB (PDB: 2ADC) showing the extra β-strand (red). **d** The secondary structure for type I KH domain. **e** Type I KH domain of Nova-1 (PDB: 1EC6) with GXXG conserved loop. **f** Type II KH domain in NusA (PDB: 2ATW). RNA nucleotides are represented in color, and protein secondary structures are shown in gray. The figures are generated with PyMOL

primary conserved sequence stretches that contribute to the RNA binding known as RNP1 ([R/K]-G-[F/Y]-[G/A]-[F/Y]-[I/L/V]-X-[F/Y]) and RNP2 ([I/L/V]-[F/Y]-[I/L/V]-X-N/L) (Fig. 1.4a) [62]. These RNA-binding sequences often rely on the surface of the central β-strands: β1 and β3 [38, 66, 67, 68]. To form these RRM–RNA complexes, solvent-exposed charged residues (Arg or Lys) form a salt bridge to the phosphodiester backbone of the RNA and two aromatic residues can form a ring-stacking interaction or hydrogen bonds with the RNA nucleobases [12, 62]. The wide range of RNA structures and recognition sequence elements has associated RRM proteins with diverse biological functions. These motifs in eukaryotes are implicated in posttranscriptional gene regulation, like pre-mRNA splicing, alternative splicing, capping, mRNA stability and export, RNA editing, and poly(A) recognition [19, 57]. During alternative splicing, many ssRBPs associate with pre-mRNA (RNPA1, $U2AF^{65}$, $U2AF^{35}$, PTB, Fox-1, sex-lethal) to regulate splicing [69]. For example, SR proteins recognize exonic splicing sites to promote alternative splicing whereas Fox-1 does the same activity by interaction with intronic splicing elements [70, 71]. Recent studies have shown that RRMs are also involved in protein–protein interactions for the recognition and interaction with RNA with very distinct mechanisms from protein–RNA interactions [57].

1.3.2 KH-Homology Domain

The heterogeneous nuclear ribonucleoprotein K-homology (KH) domain is highly expressed and most abundant in gene expression and regulatory systems in bacteria, archaea, and eukaryotes [72]. The KH domain consists of nearly 70 amino acid residues with a signature sequence of (I/L/V)IGXXGXX(I/L/V) at the center of the domain [72, 73]. All KH domains are composed of three β-sheets packed against three α-helices. KH domains are divided into two subfamilies: Type I has βααββα topology (Fig. 1.4d, e) (Nova), whereas type II has αββααβ topology (Fig. 1.4f) (NusA) [73]. An important feature of the KH domain is the presence of a variable length loop that connects $\beta 2$ and $\beta 3$ in type I and $\beta 3$ and $\alpha 2$ in type II [74]. In both type I and II, the consensus sequence is formed by a GXXG loop recognized four nucleotides. Hydrophobic interactions between bases and non-aromatic residues, backbone contacts with the GXXG loop, as well as hydrogen bonding with bases are the prevalent interactions observed between protein and RNA [4]. This ssRNA-binding protein domain can also be found in multiple copies (14 copies in chicken vigilin, three KH domains in hnRNP K) that can increase the RNA-binding affinity and cooperativity of this protein [75].

The KH domain is the most abundant RNA-binding domain in eubacteria and eukaryotes, suggesting the evolutionary importance of this ancient RNA-binding domain. Like RRM, KH protein domains are also involved in a myriad of biological processes like splicing (splicing factor 1, SF1) [76], alternative splicing (Nova family protein) [77], transcriptional and translational gene control (hnRNPK) [78],

and mRNA stability, transport, and localization [19]. Unusual expression of this protein has been linked to many diseases, such as human fragile X mental retardation syndrome which is caused by a loss of FMR-1 expression where a mutation on the conserved KH motif has an RNA-binding defect [79].

1.3.3 RNA Recognition by Modular RNA-Binding Repeats

In some cases, RNA-binding domains oligomerize to form modular RNA-binding repeats. The numbers of modular repeats vary; for example, eleven repeats are observed in TRAP proteins, seven in Sm core proteins, and six in Lsm proteins Hfq [42, 43, 44, 80].

The tryptophan RNA-binding attenuation protein (TRAP) is comprised of 70 amino acids in each of the eleven monomers that fold into four antiparallel β-strands to form a β-sandwich-like structure. Tryptophan is inserted between the interfaces of two β-strands. Each monomer oligomerizes into an 11-mer symmetric ring as observed in the crystal structure of *Bacillus subtilis* TRAP bound with a 53-nucleotide ssRNA containing GAG triplets (Fig. 1.5a) [42]. Each monomer contains an RNA-binding pocket created by two β-strands to allow for binding to the GAG triplet through protein–base interactions [42].

The outer edge of the 11-mer oligomeric structure has a symmetric ring with an 80-Å diameter. TRAP regulates the expression of L-tryptophan biosynthesis genes in several bacilli, which is activated by bound L-tryptophan. For regulation, TRAP binds

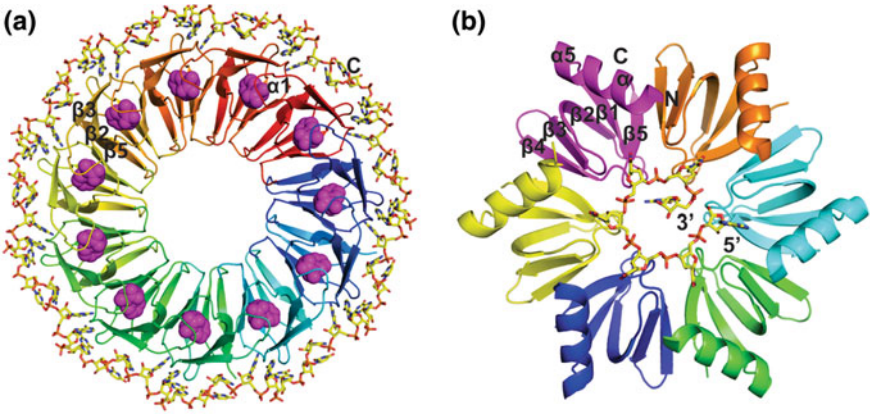

Fig. 1.5 RNA recognition by modular RNA-binding repeats. **a** The crystal structure of the 11-mer TRAP (PDB: 1C9S) protein with GAUGU ssRNA repeats. The surface in magenta is an L-tryptophan inserted in the β-sandwich. **b** Structure of Hfq (PDB: 1KQ2) showing the hexameric ring from *S. aureus*. The central core contains a bound 5′-AU_5G-3′ RNA. For clarity, each protein subunit is colored differently and RNA is in yellow sticks. The figures are generated from PyMOL

to the 5′ ssRNA leader sequence of an mRNA operon and terminates transcription by preventing the formation of the antiterminator stem-loop structure [19, 81].

The classical Sm fold is characterized by an N-terminal α-helix followed by five β-strands with a topology of αββββββ (Fig. 1.5b) [82]. The Sm proteins consist of nearly 80 residues and recognize the uridine-rich site (Sm site) present in small nuclear RNAs (snRNAs). Each Sm protein oligomerizes to form a heptameric ring (~70-Å diameter) structure around the poly(U) RNA [82]. The central hole of this ring can accommodate the U small nuclear RNP (UsnRNP) during pre-mRNA splicing [83, 84]. It has been proposed that the inter-subunit interaction during oligomerization is manifested by hydrophobic contacts between adjacent β-strands and each U-rich RNA is recognized by three conserved residues in the loops of *β2-β3* and *β4-β5* [43]. The interactions between the Sm protein domains and the RNA include stacking and hydrogen bonding. Unlike Sm proteins, LSm proteins, such as bacterial host factor for Q-β bacteriophage (Hfq), form a hexameric doughnut shape with a 12Å central cavity in the absence of RNA [44, 85, 86]. The crystal structure of *S. aureus* Hfq with a short RNA (5′-AU_5G-3′) showed that the RNA is bound around the basic central pore (Fig. 1.5b) [44]. Hfq is known to play a role in posttranscriptional gene regulation where it helps small noncoding RNAs (ncRNAs) to identify its target mRNA [87–90]. Recent studies have shown that an intrinsically disordered C-terminal domain (CTD) of Hfq acts as chaperone that auto-regulates RNA binding in bacteria [91, 92].

1.3.4 Other SsRNA-Binding Proteins

Several recent studies have shown other proteins that can bind RNA through different structural arrangements than the traditional RRM and KH domains. These protein domains include zinc fingers, Pumilio homology domain (PUF), PAZ domain, and OB-fold. Their structures, RNA recognition motifs, and protein interactions are summarized in Table 1.1 and are mentioned in many research and review articles [61, 62, 93, 94, 95].

1.4 Double-Stranded RNA Recognition

Double-stranded RNA-binding motifs (dsRBMs) recognize perfectly duplexed RNA and are distributed in eukaryotes, and bacterial and viral proteins [96]. This motif adopts an α/β sandwich global fold with an αβββα topology that contains 70–90 amino acid residues (Fig. 1.6a) [4, 23, 97, 98, 99]. Previous structural studies of dsRBM protein–RNA complexes proposed that these proteins bind in a shape-specific rather than sequence-specific [96, 99]. Many of the solved structures suggested that dsRBM recognizes the A-form helix of dsRNA, and intermolecular interactions involve the direct contact with the 2′-OH sugar and phosphate backbone [4, 53, 100, 101, 102]. But the recent solution NMR structure of an adenosine deaminase

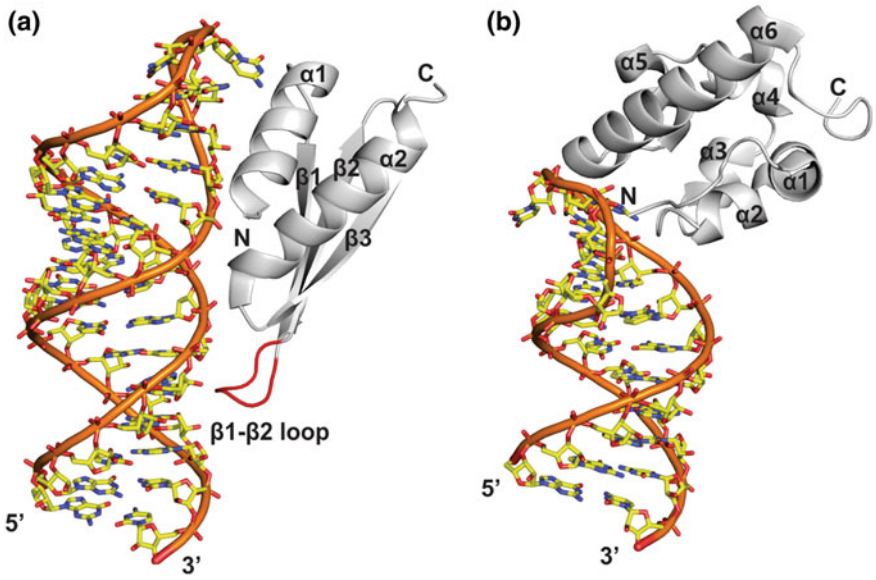

Fig. 1.6 Structure of RNA (yellow sticks) bound with dsRBM and SAM proteins (gray). **a** Upper stem-loop (USL of GluR-2 R/G) RNA recognition by dsRBM1 of ADAR2 (PDB: 2L3C). Shown in red is a β1-β2 loop that is important for sequence-specific recognition of RNA [52]. **b** The structure of Vts1p-SAM (PDB: 2ESE) domain in complex with SRE RNA. The figures are generated from PyMOL

(ADAR2) in complex with a stem-loop pre-mRNA encoding the R/G editing site of GluR-2 has revealed that dsRBM recognizes the shape as well as the sequence of the RNA [52]. The minor groove of the A-form helix in the stem-loop is specifically recognized by the N-terminal helix (α1) and β1-β2 loop of ADAR2 (Fig. 1.6a). The two domains of ADAR2, dsRBM1, and dsRBM2 preferentially recognize G-X_9-A and G-X_8-A RNA sequences, respectively, in a long stem-loop pre-mRNA. The sequence specificity of ADAR2 dsRBM is important for the proper editing function of the enzyme [52].

The double-stranded RBM is involved in several biological processes from RNA editing to protein phosphorylation in translational control [96]. For example, the RNase III domain is involved in RNA processing in the RNA interference (RNAi)/microRNA (miRNA) pathway [103–105]. *Drosophila melanogaster* staufen contains multiple copies of dsRBM domains that control RNP localization [105]. Furthermore, ADAR1 and ADAR2 are RNA editing proteins that regulate gene expression at the RNA level [106] by converting adenosine to inosine (A to I) by hydrolytic deamination in many mRNA and pre-mRNA transcripts [52, 107].

1.5 SAM-Binding Domain

The sterile alpha motif (SAM) domain is the most copious of the eukaryotic protein motifs, initially identified as a protein–protein interaction module involved in transcription regulation and signal transduction [54, 108]. Later, it was reported that the SAM domain also interacts with RNA to control posttranscriptional gene expression [109]. The SAM domain from *Saccharomyces cerevisiae* (Vts1p) and its homolog from *Drosophila melanogaster* (Smaug) specifically interact with the RNA stem-loop [109]. The RNA stem-loop recognized by Smaug contains a CNGGN pentaloop in the Smaug recognition element (SRE) present at the 3′ untranslated region (UTR) of the *nos* transcript [109, 110]. The solution NMR structure of Vts1p-SAM in complex with a 23-nucleotide SRE stem-loop RNA with a CUGGC pentaloop was recently solved (Fig. 1.6b). This study revealed that the SAM domain recognizes RNA in a shape-specific rather than sequence-specific manner specifically recognizing the G in position three of the pentaloop [54]. Two intermolecular hydrogen bonds specifically recognize the identity of the third G in the pentaloop, which also occupies the hydrophobic cavity formed by Leu465 and Ala495 [54]. This protein consists of six α-helices that adopt a globular protein fold and recognize the major groove of the RNA pentaloop through contacts with the RNA sugar phosphate backbone [54].

1.6 Protein–RNA Interactions in the Ribosome

The ribosome is a protein–RNA complex with a catalytic role in protein synthesis. This complex macromolecule consists of more than 50 different ribosomal proteins that interact with RNA. How all of these proteins interact with RNA to form an active structure of the ribosome was a question that proved elusive. The recent X-ray crystal structures of the ribosomal subunits offered a clear picture to explain the interactions between the ribosomal proteins and the RNA [111, 112]. The majority of the ribosomal proteins recognize ribosomal RNA by shape rather than by sequence. Hydrogen bonding, stacking, hydrophobic interactions, as well as interactions with the phosphate backbone were also observed among the characterized protein–RNA interactions.

Ribosomal proteins contain globular domains with similar α/β sandwich folds [111, 113]. The topologies of some of the ribosomal proteins are similar to other RNA-binding proteins as described before, reflecting the similar RNA-binding properties among them. Most of these proteins have extended structures like extended α-hairpin (S2), β-hairpin (S5, S10), N-terminal extension (S3), and C-terminal tail (S6) [112, 113]. These extensions are associated with basic amino acid side chains and have extensive contacts with ribosomal RNA that stabilize the tertiary structure of the ribosome and also participate in protein–protein interactions [113]. In the crystal structure, most of the primary binders are globular and surface-oriented and have direct interaction with RNA helices during assembly. For example, S15 is a primary

binder with four α-helices and without any extensions that recognizes the junction of helices h20, h21, and h22 as well as helix h23a in the 16S ribosomal RNA [114]. Proteins with multiple extensions are buried in the RNA and are secondary or tertiary binders. Except for very few (h10, h14, and h33a), most of the RNA helices in the 16S RNA contact proteins and many proteins can recognize a single RNA helix. Most of the proteins in the large subunit, except L12, have direct interaction with RNA [111]. Therefore, it can be theorized that RNA-binding proteins may function in the proper folding of RNA. But some of the ribosomal proteins from large subunit (L1, L10, and L11) are directly involved in protein synthesis. Ribosomal proteins also have significant protein–protein interactions that influence the proper assembly of the ribosomal subunits [113].

1.7 Conclusions

RNA molecules can adopt different secondary and tertiary structures that not only allow it to perform structural, catalytic, and regulatory roles but also create a platform to interact with many proteins to form protein–RNA complexes. These protein–RNA complexes have a wide variety of structural and functional roles in the cell. Most of the RNA-binding proteins are modular, and their mode of RNA recognition is also different. We have discussed the common RNA-binding proteins and how they recognize target RNA based on available information from structural biology. Future works need to focus more on exploring the dynamics and mechanistic importance of protein–RNA interactions and their roles in cellular functions. The experimental approaches like single-molecule techniques in combination with computational biology might help to gain insight into the molecular mechanism and dynamics of protein–RNA interactions and their function.

Acknowledgements RL would like to thank Dr. Amanda Solem and Dr. Sharla Wood for their help and Dr. David Rueda for comments and suggestions on the manuscripts. RL is supported by the start-up funds from the University of Tennessee, Knoxville.

References

1. Crick, F. H. (1958). On protein synthesis. *Symposia of the Society for Experimental Biology, 12,* 138–163.
2. Glisovic, T., Bachorik, J. L., Yong, J., & Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. *FEBS Letters, 582,* 1977–1986.
3. Dreyfuss, G., Kim, V. N., & Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol, 3,* 195–205.
4. Lunde, B. M., Moore, C., & Varani, G. (2007). RNA-binding proteins: Modular design for efficient function. *Nature Reviews Molecular Cell Biology, 8,* 479–490.
5. Foley SW, Kramer MC, Gregory BD (2017) RNA structure, binding, and coordination in Arabidopsis. Wiley Interdiscip Rev RNA.

6. Keene, J. D. (2007). RNA regulons: Coordination of post-transcriptional events. *Nature Reviews Genetics, 8,* 533–543.
7. Gerstberger, S., Hafner, M., & Tuschl, T. (2014). A census of human RNA-binding proteins. *Nature Reviews Genetics, 15,* 829–845.
8. Jones, S., Daley, D. T., Luscombe, N. M., Berman, H. M., & Thornton, J. M. (2001). Protein-RNA interactions: A structural analysis. *Nucleic Acids Research, 29,* 943–954.
9. Ray, D., Kazan, H., Cook, K. B., Weirauch, M. T., Najafabadi, H. S., Li, X., et al. (2013). A compendium of RNA-binding motifs for decoding gene regulation. *Nature, 499,* 172–177.
10. Singh, G., Pratt, G., Yeo, G. W., & Moore, M. J. (2015). The clothes make the mRNA: Past and present trends in mRNP fashion. *Annual Review of Biochemistry, 84,* 325–354.
11. Siomi, H., & Dreyfuss, G. (1997). RNA-binding proteins as regulators of gene expression. *Current Opinion in Genetics & Development, 7,* 345–353.
12. Burd, C. G., & Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. *Science, 265,* 615–621.
13. Cooper, T. A., Wan, L., & Dreyfuss, G. (2009). RNA and disease. *Cell, 136,* 777–793.
14. Bekenstein, U., & Soreq, H. (2013). Heterogeneous nuclear ribonucleoprotein A1 in health and neurodegenerative disease: From structural insights to post-transcriptional regulatory roles. *Molecular and Cellular Neuroscience, 56,* 436–446.
15. Mannoor, K., Liao, J. P., & Jiang, F. (2012). Small nucleolar RNAs in cancer. *Biochimica et Biophysica Acta—Reviews on Cancer, 1826,* 121–128.
16. Ferreira, H. J., Heyn, H., Moutinho, C., & Esteller, M. (2012). CpG island hypermethylation-associated silencing of small nucleolar RNAs in human cancer. *RNA Biology, 9,* 881–890.
17. Mihailovic, M. K., Chen, A., Gonzalez-Rivera, J. C., & Contreras, L. M. (2017). Defective ribonucleoproteins, mistakes in RNA processing, and diseases. *Biochemistry-Us, 56,* 1367–1382.
18. Ramaswami, M., Taylor, J. P., & Parker, R. (2013). Altered ribostasis: RNA-protein granules in degenerative disorders. *Cell, 154,* 727–736.
19. Messias, A. C., & Sattler, M. (2004). Structural basis of single-stranded RNA recognition. *Accounts of Chemical Research, 37,* 279–287.
20. Plaschka, C., Hantsche, M., Dienemann, C., Burzinski, C., Plitzko, J., & Cramer, P. (2016). Transcription initiation complex structures elucidate DNA opening. *Nature, 533,* 353–358.
21. Plaschka, C., Lin, P. C., & Nagai, K. (2017). Structure of a pre-catalytic spliceosome. *Nature, 546,* 617–621.
22. Khatter, H., Myasnikov, A. G., Natchiar, S. K., & Klaholz, B. P. (2015). Structure of the human 80S ribosome. *Nature, 520,* 640–645.
23. Tian, B., Bevilacqua, P. C., Diegelman-Parente, A., & Mathews, M. B. (2004). The double-stranded-RNA-binding motif: Interference and much more. *Nature Reviews Molecular Cell Biology, 5,* 1013–1023.
24. Auweter, S. D., Oberstrass, F. C., & Allain, F. H. (2007). Solving the structure of PTB in complex with pyrimidine tracts: An NMR study of protein-RNA complexes of weak affinities. *Journal of Molecular Biology, 367,* 174–186.
25. Clery, A., Blatter, M., & Allain, F. H. (2008). RNA recognition motifs: Boring? Not quite. *Current Opinion in Structural Biology, 18,* 290–298.
26. Auweter, S. D., & Allain, F. H. (2008). Structure-function relationships of the polypyrimidine tract binding protein. *Cellular and Molecular Life Sciences, 65,* 516–527.
27. Jones, S. (2016). Protein-RNA interactions: Structural biology and computational modeling techniques. *Biophysical Reviews, 8,* 359–367.
28. Mackereth, C. D., Simon, B., & Sattler, M. (2005). Extending the size of protein-RNA complexes studied by nuclear magnetic resonance spectroscopy. *ChemBioChem, 6,* 1578–1584.
29. Carlomagno, T. (2014). Present and future of NMR for RNA-protein complexes: A perspective of integrated structural biology. *Journal of Magnetic Resonance, 241,* 126–136.
30. Lapinaite, A., Simon, B., Skjaerven, L., Rakwalska-Bange, M., Gabel, F., & Carlomagno, T. (2013). The structure of the box C/D enzyme reveals regulation of RNA methylation. *Nature, 502,* 519.

31. Duss, O., Michel, E., Yulikov, M., Schubert, M., Jeschke, G., & Allain, F. H. T. (2014). Structural basis of the non-coding RNA RsmZ acting as a protein sponge. *Nature, 509,* 588.
32. Ren, H., & Shen, Y. (2015). RNA-binding residues prediction using structural features. *BMC Bioinformatics, 16,* 249.
33. Mackereth, C. D., & Sattler, M. (2012). Dynamics in multi-domain protein recognition of RNA. *Current Opinion in Structural Biology, 22,* 287–296.
34. Lamichhane, R., Hammond, J. A., Pauszek, R. F., Anderson, R. M., Pedron, I., van der Schans, E., et al. (2017). A DEAD-box protein acts through RNA to promote HIV-1 Rev-RRE assembly. *Nucleic Acids Research, 45,* 4632–4641.
35. Lamichhane, R., Daubner, G. M., Thomas-Crusells, J., Auweter, S. D., Manatschal, C., Austin, K. S., et al. (2010). RNA looping by PTB: Evidence using FRET and NMR spectroscopy for a role in splicing repression. *Proceedings of the National Academy of Sciences of the United States of America, 107,* 4105–4110.
36. Karunatilaka, K. S., Solem, A., Pyle, A. M., & Rueda, D. (2010). Single-molecule analysis of Mss116-mediated group II intron folding. *Nature, 467,* 935–U975.
37. Bonilla, S., Limouse, C., Bisaria, N., Gebala, M., Mabuchi, H., & Herschlag, D. (2017). Single-molecule fluorescence reveals commonalities and distinctions among natural and in vitro-selected RNA tertiary motifs in a multistep folding pathway. *Journal of the American Chemical Society, 139,* 18576–18589.
38. Oberstrass, F. C., Auweter, S. D., Erat, M., Hargous, Y., Henning, A., Wenter, P., et al. (2005). Structure of PTB bound to RNA: Specific binding and implications for splicing regulation. *Science, 309,* 2054–2057.
39. Auweter SD (2006) Structure and function of PTB and fox, two regulators of alternative splicing. Swiss Federal Institute of Technology, Zurich.
40. Lewis, H. A., Musunuru, K., Jensen, K. B., Edo, C., Chen, H., Darnell, R. B., et al. (2000). Sequence-specific RNA binding by a Nova KH domain: Implications for paraneoplastic disease and the fragile X syndrome. *Cell, 100,* 323–332.
41. Beuth, B., Pennell, S., Arnvig, K. B., Martin, S. R., & Taylor, I. A. (2005). Structure of a Mycobacterium tuberculosis NusA-RNA complex. *EMBO Journal, 24,* 3576–3587.
42. Antson, A. A., Dodson, E. J., Dodson, G., Greaves, R. B., Chen, X., & Gollnick, P. (1999). Structure of the trp RNA-binding attenuation protein, TRAP, bound to RNA. *Nature, 401,* 235–242.
43. Thore, S., Mayer, C., Sauter, C., Weeks, S., & 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. *Journal of Biological Chemistry, 278,* 1239–1247.
44. Schumacher, M. A., Pearson, R. F., Moller, T., Valentin-Hansen, P., & Brennan, R. G. (2002). Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: A bacterial Sm-like protein. *EMBO Journal, 21,* 3546–3556.
45. Wang, X., McLachlan, J., Zamore, P. D., & Hall, T. M. (2002). Modular recognition of RNA by a human pumilio-homology domain. *Cell, 110,* 501–512.
46. Wang, B., & Ye, K. (2017). Nop9 binds the central pseudoknot region of 18S rRNA. *Nucleic Acids Research, 45,* 3559–3567.
47. Wang, X., Zamore, P. D., & Hall, T. M. (2001). Crystal structure of a Pumilio homology domain. *Molecular Cell, 7,* 855–865.
48. Hudson, B. P., Martinez-Yamout, M. A., Dyson, H. J., & Wright, P. E. (2004). Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. *Nature Structural and Molecular Biology, 11,* 257–264.
49. Park, S., Phukan, P. D., Zeeb, M., Martinez-Yamout, M. A., Dyson, H. J., & Wright, P. E. (2017). Structural basis for interaction of the tandem zinc finger domains of human muscle-blind with cognate RNA from human cardiac troponin T. *Biochemistry-Us, 56,* 4154–4168.
50. Ma, J. B., Ye, K., & Patel, D. J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature, 429,* 318–322.
51. Schirle, N. T., & MacRae, I. J. (2012). The crystal structure of human Argonaute2. *Science, 336,* 1037–1040.