Received: 6 October 2023 | Revised: 19 June 2024 | Accepted: 12 July 2024

DOI: 10.1002/bmb.21854

HISTORICAL REVIEW

Biochemistry and Molecular Biology Education

WILEY

A primer for junior trainees: Recognition of RNA modifications by RNA-binding proteins

Murphy Angelo | **Yash Bhargava** | **Scott Takeo Aoki**

Department of Biochemistry and Molecular Biology, School of Medicine, Indiana University Purdue University Indianapolis, Indianapolis, Indiana, USA

Correspondence
Scott Takeo Aoki, Department of Biochemistry and Molecular Biology, School of Medicine, Indiana University Purdue University Indianapolis, Indianapolis, IN 46202, USA.
Email: staoki@iu.edu

Funding information
School of Medicine, Indiana University, Precision Health Initiative (PIH); National Institute of General Medical Sciences, Grant/Award Number: R35 GM142691

Abstract

The complexity of RNA cannot be fully expressed with the canonical A, C, G, and U alphabet. To date, over 170 distinct chemical modifications to RNA have been discovered in living systems. RNA modifications can profoundly impact the cellular outcomes of messenger RNAs (mRNAs), transfer and ribosomal RNAs, and noncoding RNAs. Additionally, aberrant RNA modifications are associated with human disease. RNA modifications are a rising topic within the fields of biochemistry and molecular biology. The role of RNA modifications in gene regulation, disease pathogenesis, and therapeutic applications increasingly captures the attention of the scientific community. This review aims to provide undergraduates, junior trainees, and educators with an appreciation for the significance of RNA modifications in eukaryotic organisms, alongside the skills required to identify and analyze fundamental RNA–protein interactions. The pumilio RNA-binding protein and YT521-B homology (YTH) family of modified RNA-binding proteins serve as examples to highlight the fundamental biochemical interactions that underlie the specific recognition of both unmodified and modified ribonucleotides, respectively. By instilling these foundational, textbook concepts through practical examples, this review contributes an analytical toolkit that facilitates engagement with RNA modifications research at large.

KEYWORDS

review, RNA, RNA modifications, RNA-binding proteins, RNA–protein interactions

1 | THE CENTRAL DOGMA OF MOLECULAR BIOLOGY

The discovery of RNA modifications has introduced a new layer of complexity to the ever-changing landscape of biochemistry and molecular biology. Biochemistry and molecular biology involve the study of the structure, function, and interactions between biological macromolecules. Molecular biology is based on the interplay between deoxyribonucleic acid (DNA), ribonucleic acid (RNA), and protein. These macromolecules are essential for life, with DNA serving as an information storage, RNA as an intermediary dealer of DNA's information and a jack of all trades, and protein as a versatile, functional building block that makes a cell go. This process, where DNA is made into RNA, which is made into

Murphy Angelo and Yash Bhargava contributed equally to this study.

protein, is the central dogma of molecular biology.[1] Despite exceptions, the central dogma continues to serve as a fundamental framework for comprehending molecular biology (for a more comprehensive review, see Reference 2). This review focuses on the chemical interactions of RNA with RNA-binding proteins, the impacts of RNA modifications on these interactions, and the connection between RNA modifications and the central dogma. Understanding how RNA interacts with protein gives insights into the processes and mechanisms responsible for gene regulation, life, and disease. A greater understanding of RNA biology will lead to new tools to investigate their roles in organisms and new therapies for human and animal diseases.

2 | OVERVIEW OF RNA AND RNA MODIFICATIONS

There are three key structural differences between RNA and DNA.[3] Firstly, RNA has a hydroxyl group at the 2′ position of its ribose sugar (Figure 1a). DNA does not, thus defining its "Deoxyribose" name. Secondly, RNA uses adenine (A), cytosine (C), guanine (G), and uracil (U) as bases. DNA also uses A, C, G, but typically uses thymine (T) instead of U (Figure 1a),[3] with some notable exceptions, such as the use of U instead of T in certain bacteriophage DNA.[4] Both DNA and RNA can form base pairs, transitioning from single-stranded nucleic acids to antiparallel double-stranded helices. They can also form other structural assemblies, such as by folding on each other to form tertiary structures, similar to protein. However, the third difference is that RNA is typically found in cells as single-stranded, double-stranded, or in tertiary structures. DNA is primarily found as double-stranded helices, stabilizing the nucleic acid, protecting it from degradation to permit long-term storage of biological information. Thus, DNA and RNA have structural similarities, but their inherent chemical attributes enable them to be used for different purposes in biology. For a more comprehensive discussion of these similarities and differences, interested readers are encouraged to refer to other published reviews.[3]

RNA modifications can occur on all four bases and encompass a diverse array of chemical changes to the nitrogenous base or ribonucleoside sugar. The study of naturally occurring ribonucleoside modifications began in 1951 with the discovery of pseudouridine (Ψ),[5] an isomer of uridine where a carbon and nitrogen in the uracil ring have switched places. Since this discovery, over 170 additional modifications have been identified (see References 6–10 for comprehensive reviews on RNA modifications). The pace of identifying new modifications

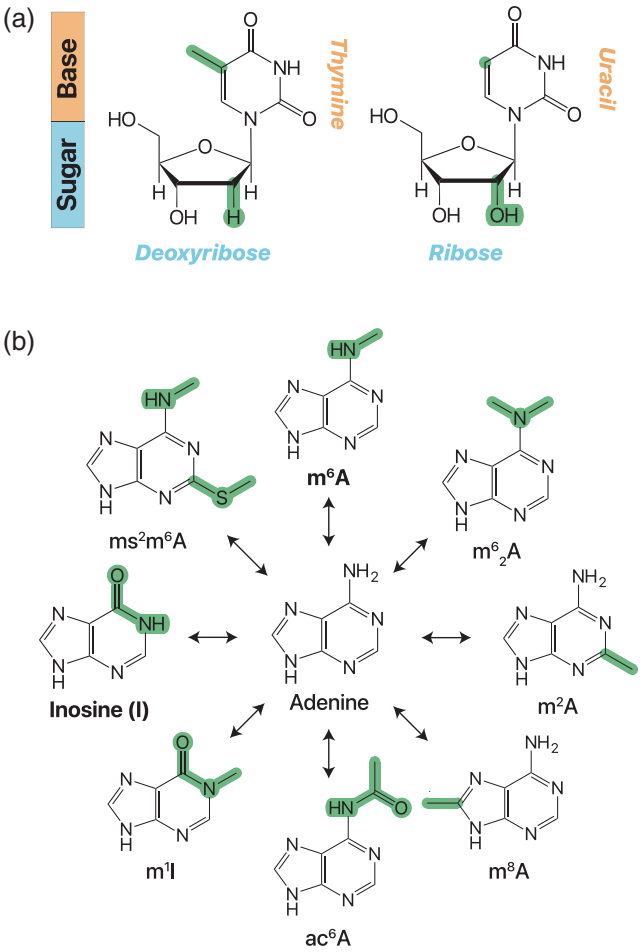

FIGURE 1 DNA, RNA, and RNA modifications. (a) Chemical structures of DNA thymidine and RNA uridine. Chemical differences highlighted in green. (b) Adenine and a selection of natural RNA base modifications. N^6-methyladenosine, m^6A; N^6,N^6-dimethyladenosine, $m^6{}_2A$; 2-methyladenosine, m^2A; 8-methyladenosine, m^8A; N^6-acetyladenosine, ac^6A; 1-methylinosine, m^1I; 2-methylthio-N^6-methyladenosine, ms^2m^6A. Chemical differences again highlighted in green.

is rapidly accelerating due to the enhanced precision and accuracy provided by modern molecular biology equipment and techniques, along with the growing appreciation of RNA modifications and their involvement in a myriad of cellular pathways.[11] Other common modification examples include the addition of a hydroxymethyl group on cytidine to form 5-hydroxymethylcytidine (hm^5C), as well as a variety of adenosine modifications such as N^6-methyladenosine (m^6A, Figure 1b). These chemical changes occur via specialized enzymatic pathways unique to the modification and biological context.[12,13] Therefore, RNA modifications are found in differing amounts and RNA sites, depending on the organism, cell type, environment, and other factors.

3 | THE WRITERS AND ERASERS OF m^{6}A

RNA modifications are managed by molecular writers and erasers (Figure 2a) (see References 14,15 for comprehensive m^{6}A reviews). Writers are enzymes that add a specific RNA modification, while erasers are enzymes that remove the modification. Writers and erasers work together to create and manage the epitranscriptome—the dynamic balance of RNA modifications within the cell. This regulation has far-reaching implications on the molecular function and expression of modified RNA targets.[16–19]

The writers and erasers of m^{6}A are particularly well characterized (Figure 2)[15,20] and have a significant impact on gene expression, animal development, and human disease.[16,21–23] As such, this review will use m^{6}A as a prototypical example of the chemistry, biochemistry, and biology of an RNA modification and how it interacts with proteins. Discovered in the 1970s,[24] m^{6}A is prevalent in vertebrate RNA[25] and found on thousands of their messenger RNAs (mRNAs), the RNAs used to code for proteins. These mRNAs have m^{6}A modifications concentrated near stop codons and in their 3′ untranslated regions.[14,25] Methyltransferase complexes modify adenosine into m^{6}A. Although they consist of several proteins, the cores of these complexes involve methyltransferase-like (METTL) enzymes that catalyze the methylation reaction. For example, METTL3 and METTL14 assemble and can modify adenosines in mRNAs but rely on other proteins for enhanced enzymatic activity and site selection.[26,27] METTL3 is the catalytic subunit. METTL14 and other proteins maintain the correct conformation for enzymatic activity. These other proteins can also impart preferences for specific m^{6}A modification sites.[28] The methyltransferase complexes have preferences for specific RNA sequences known as motifs (Figure 2). The targeted RNA sequence for METTL3/METTL14 is the RRACH motif, where R = A or G, and H = A, C, or U.[29,30] The central A of this motif is enzymatically converted to m^{6}A. There are two established m^{6}A erasers: alkylation B homolog 5 (ALKBH5) and fat mass and obesity-associated protein (FTO).[31,32] These demethylases work by modifying the N^6 methyl group further to enable chemistry that can restore the base to unmodified adenosine.[31] Both writers and erasers are associated with human disease. Overexpression of the METTL3/METTL14 m^{6}A writers are associated with liver, gastric, and colon cancer.[33–35] The FTO m^{6}A eraser is associated with obesity.[36] In summary, writers and readers are the enzymatic ying-yang for RNA modifications like m^{6}A. Perturbation of this dynamic balance can lead to disease.

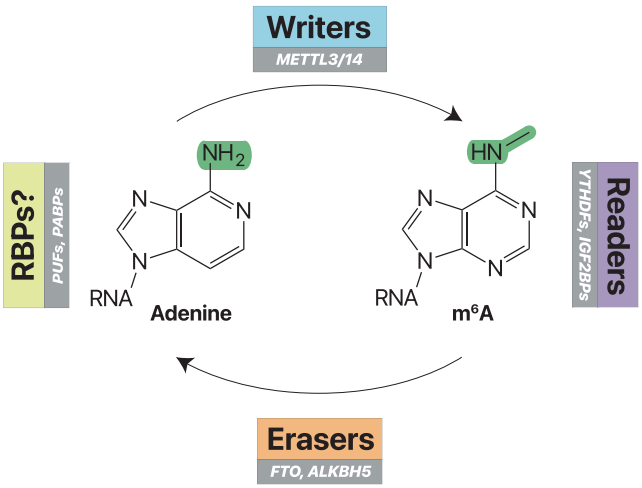

FIGURE 2 Writers, erasers, and readers of m^{6}A. Writer enzymes (e.g., METTL3/14) add a methyl group to the nitrogen-6 position of adenosine to make N^6-methyladenosine (m^{6}A) RNA. Eraser enzymes (e.g., FTO, ALKBH5) can remove these methyl marks. Reader RNA-binding proteins (e.g., YTHDFs, IGF2BPs) specifically recognize and bind m^{6}A RNA. Unmodified RNA may be recognized by other RNA-binding proteins (RBPs; e.g., PUFs, PABPs).

4 | THE READERS OF m^{6}A

While RNA modifications can affect RNA tertiary structure and overall stability for larger biological implications (e.g., transfer RNA (tRNA) modifications, stability, and thermotolerance[37]), this review focuses on protein interactions with modified RNA. Readers are binding proteins that recognize specific RNA modifications. This interaction can lead to regulation of the RNA target. The best characterized readers for m^{6}A are the YT521-B homology (YTH) domain family of proteins (YTHDFs) and YTH domain-containing proteins (YTHDCs). YTHDFs and YTHDCs recognize m^{6}A in the nucleus and cytoplasm, resulting in different biochemical functions contingent upon the specific reader protein and the cellular context.[38] For example, YTHDF2 and other YTH proteins can recruit the CCR4-NOT deadenylation complex for mRNA decay.[39] CCR4-NOT removes the poly-A tail of mRNAs, leading to mRNA turnover. Additional functions of the YTH proteins are still being studied, but a critical aspect is that these proteins must bind to their RNA target to elicit their biochemical function.[15,40] Thus, the molecular recognition of YTH and other RNA-binding proteins depends on their interactions with target RNA.

5 | COMMON RNA–PROTEIN INTERACTIONS

To understand how readers recognize specific RNA modifications, it is essential to have a foundation in the chemical interactions between RNA and protein. RNA–protein interactions commonly encompass: (1) electrostatic interactions, (2) hydrogen bonding, (3) salt bridges, (4) hydrophobic interactions, (5) Van der Waals interactions, and (6) pi stacking interactions (Figure 3).[41]

Electrostatic interactions arise from the attraction or repulsion between charged particles (Figure 3a). The fundamental idea is that opposite charges attract while like charges repel. Protein amino acids have neutral, positive, or negative charges. The propensity of a particular amino acid to carry or lack a charge is governed by the chemical structure of its side chain as well as by the pH of the surrounding environment.[31] While unmodified A, C, G, and U are almost universally neutrally charged within the cell, some RNA modifications carry a charge under physiological conditions.[8,42] Notably, the phosphate group contained in the backbone of nucleic acids carries a negative charge. Consequently, many readers have positively charged residues that nonspecifically attract nucleic acids.[43,44]

Hydrogen bonding occurs when the partial positive charge of a hydrogen atom, bound to an electronegative atom, attracts an electronegative partner (Figure 3b). Electronegativity denotes an atom's propensity to attract electrons. When highly electronegative atoms, like oxygen or nitrogen, are bound to hydrogen atoms, regions of partial negative charge and partial positive charge are formed, indicated as δ− or δ+, respectively. The partial positive charge occurs on the less electronegative hydrogen. Conversely, the partial negative charge occurs on the more electronegative group. When correctly oriented, these oppositely charged regions can establish attractions. Neutral hydrogen bonds at 2.4–3.0 Å distance contribute about 0.5–1.5 kcal/mol per interaction, and charged hydrogen bonds, or "salt bridges," within a 4.0 Å distance contribute 3.5–4.5 kcal/mol per interaction (Figure 3c).[45] Protein and RNA contain chemical moieties capable of hydrogen bonding, and modifications to RNA and protein frequently introduce additional groups that change this capability.[43,44,46] While generally weaker than covalent or ionic interactions, the collective contribution of hydrogen bonds can have considerable impact on the RNA-binding protein selectivity for a given modification. Analyses have estimated the prevalence of hydrogen bonds to the base, the ribose 2′-OH, and the RNA phosphate backbone at 36%, 24%, and 41% of RNA–protein hydrogen bonds, respectively.[47,48]

Hydrophobic interactions (Figure 3d) occur as a result of molecules trying to minimize contact with the surrounding water. The interactions occur between nonpolar regions at distances of 3.8–5.0 Å and contribute approximately 1–2 kcal/mol.[49,50] RNA and protein have hydrophilic and hydrophobic moieties that group with

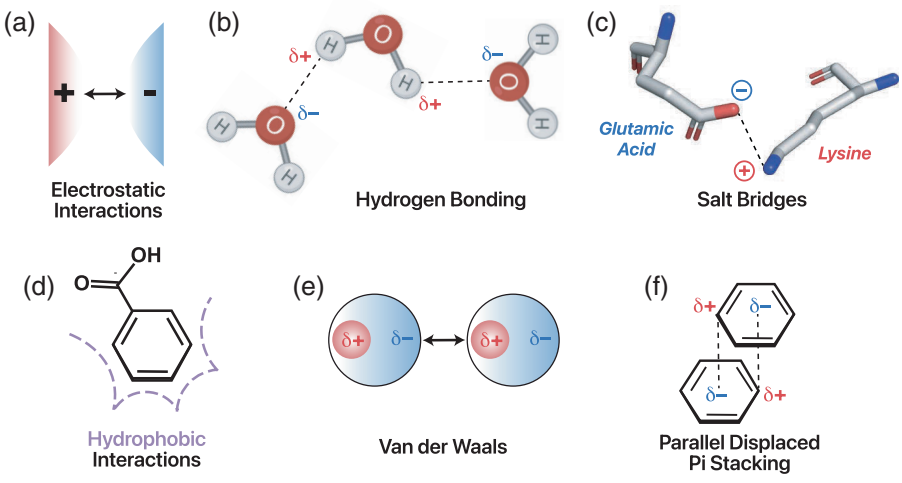

FIGURE 3 Six types of intermolecular forces drive RNA–protein interactions. (a) Opposite charges attract. (b) Hydrogen bonds occur commonly among polar molecules such as water. The difference in electronegativity creates opposite partial charges that attract one another. Images adapted from BioRender (BioRender.com). (c) Salt bridges represent the combination of hydrogen bonding and electrostatic interactions. Example from PDB ID: 5W4A.[82] (d) Hydrophobic interactions form in polar solvents. Nonpolar regions aggregate to minimize disruption of favorable solvent–solvent interactions. (e) London dispersion forces, a subtype of Van der Waals interactions, occur due to temporary induced dipoles. (f) Parallel-displaced pi stacking is favorable due to attractions between the positive hydrogen substituents (δ+) and the negatively-charged (δ−), delocalized pi system.

like elements.[10,42,46] Hydrogen bonding drives hydrophilic interactions directly and hydrophobic interactions indirectly. Amino acids with many nonpolar carbon–carbon bonds, like leucine, isoleucine, phenylalanine, tryptophan, and others, are hydrophobic and fold together to form a "hydrophobic core." This core may also interact with a hydrophobic moiety on RNA.[51–54] Up to 50% of RNA–protein interface interactions may be hydrophobic, depending on the RNA-binding protein.[55]

There are two types of Van der Waals forces: the weaker London dispersion forces and the stronger dipole–dipole forces.[56] London dispersion forces arise due to temporary induced dipoles—imbalances in the charge distribution surrounding molecules (Figure 3e). Stronger Van der Waals interactions may form as a result of permanent dipoles. Hydrogen bonds exceeding a certain threshold distance, typically 3.0 Å, fall into this category.[57,58] Both types of Van der Waals forces are weak electrostatic interactions of about 0.5–1 kcal/mol.[41] They largely play stabilizing roles in the binding of proteins to RNA.[41]

Aromatic rings aligning face-to-face (Figure 3f) or face-to-edge results in pi stacking. These interactions typically form at distances of 2.7–4.3 Å and are relatively strong, contributing about 2–6 kcal/mol per interaction.[59] They are frequently observed in protein and RNA interactions due to the aromaticity present in RNA and many amino acids. In YTH and other RNA-binding proteins, pi stacking interactions play a crucial role in shaping the active site, effectively sandwiching the targeted base in place.[54,60]

6 | RECOGNITION OF RNA BY RNA-BINDING PROTEINS

All RNA-binding proteins follow similar principles when interacting with their targets. Firstly, they have specificity interactions that designate their sequence or secondary structure preferences. Secondly, they use positively charged amino acid side chains to account for the negatively charged phosphate backbone. Thirdly, they often target the 2′ hydroxyl in RNA to differentiate from DNA.

The pumilio and FBF protein family (PUFs) of RNA-binding proteins serve as a good example of sequence specific RNA interactors that use these three concepts. PUFs contain a conserved RNA-binding domain known as the pumilio homology domain (PUM-HD) (see References 61–63 for comprehensive reviews on Pumilio), of which there are many atomic-resolution crystal structures determined with and without RNA.[64–66] The structure of human pumilio 2 homology domain (hPUM2-HD) bound to RNA shows how the canonical PUM-HD has eight α-helical repeats that bind to a conserved RNA sequence, UGUANAUA, with N being A, C, G, or U (Figure 4a).[67–69] Each α-helical repeat recognizes one unpaired base via three amino acid side chains.[70] Two side chains interact with an edge of the base, while the third residue forms pi stacking interactions in the plane between two bases.[63] Thus, in following with the first principle, amino acid side chains form a tripartite code for sequence binding specificity. PUM-HD also has arginine, lysine, and histidine side chains surrounding the RNA-binding surface, following the second principle of positively charged groups attracting negatively charged RNA. hPUM2-HD does not have amino acids interacting with the 2′ hydroxyl groups of the RNA. Notably, this PUM-HD can bind to both RNA and DNA.[66] To summarize, the RNA-bound hPUM2-HD structure shows many of the basic characteristics observed in other RNA-binding proteins. Deviations from the basic principles, such as the lack of specificity for the 2′ hydroxyl, allow the protein to bind to a broader range of substrates.

PUM-HD recognition of adenosine at the fourth RNA position is specific and occurs almost entirely through interactions with the nucleobase (Figure 4b). The ringed tyrosine and positively charged, flat arginine contributes favorable pi stacking interactions, while glutamine forms a hydrogen bond with the adenosine nitrogen (Figure 4b). Uridine at the 3rd RNA position has its base similarly sandwiched between amino acid side chains, but specificity is dictated by a different set of protein residues (Figure 4c). Uridine and adenosine are very different bases. The uracil base of uridine is a pyrimidine with a single, six-membered ring, while the adenine base of adenosine is a purine with two rings—a matching pyrimidine ring fused to an additional five-membered imidazole ring. Uracil has two carbonyl moieties situated at the 2 and 4 positions of its pyrimidine ring. In contrast, the pyrimidine ring of adenine has an amino group situated at its 6 position. PUM-HD uses these moieties as chemical signatures to differentiate uridine from adenosine. The carbonyls on the uracil base form hydrogen bonds with the amide moieties of glutamine and asparagine in PUM-HD (Figure 4c). This binding interaction is incompatible with an adenine base. At the remaining six recognition sites, PUM-HD uses specific combinations of amino acids in each α-helical repeat to target specific nucleobases.[65] Thus, RNA-binding proteins like PUF target specific RNAs by using amino acids that account for the particular chemical signature of their desired targets.

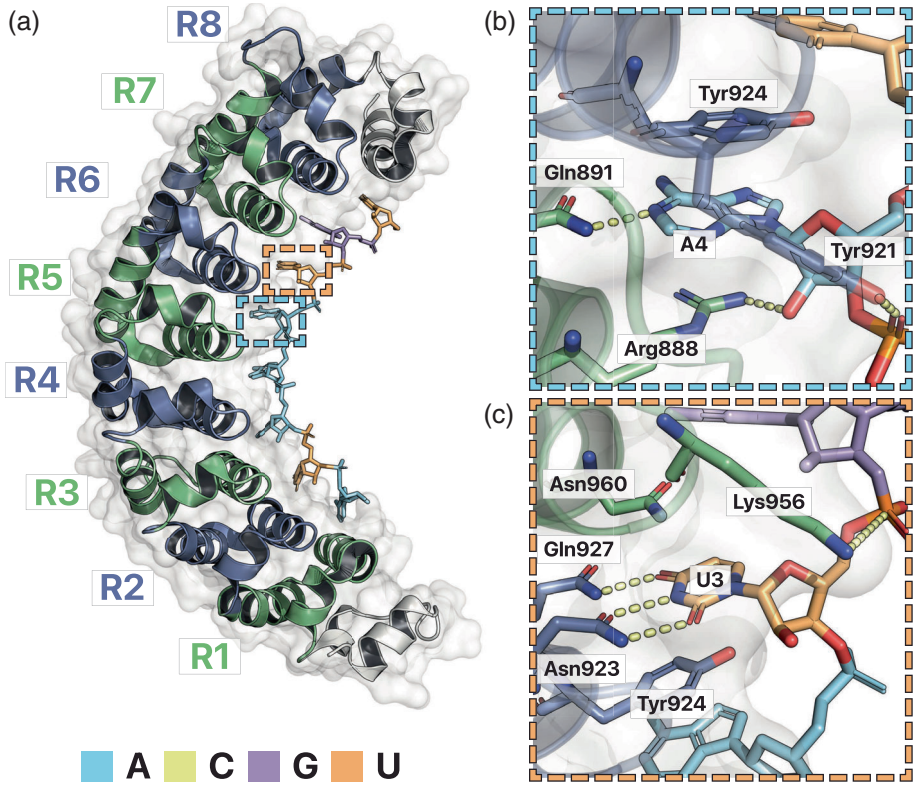

FIGURE 4 Structural features of pumilio, an RNA-binding protein. (a) Crystal structure of human pumilio 2 homology domain (hPUM2-HD) in complex with RNA (PDB ID: 3Q0Q[65]). The RNA-binding domain is composed of eight α-helical PUF repeats (green and blue), each binding a single RNA base (color coded). Dashed boxes are enlarged in B and C. (b) Coordination of adenosine RNA. Recognition of adenosine at site 4 ("A4") is mediated by pi stacking with arginine^{888} (Arg888) and tyrosine^{924} (Tyr924) and hydrogen bonds (dashed lines) with glutamine^{891} (Gln891), arginine^{888} (Arg888), and tyrosine^{921} (Tyr921). (c) Recognition of uridine at site 3 ("U3") is mediated by pi stacking with tyrosine^{924} (Tyr924) and asparagine^{960} (Asn960), hydrogen bonds with glutamine^{927} (Gln927) and asparagine^{923} (Asn923), and electrostatic attractions (dashed lines) between lysine^{956} (Lys956) and the RNA phosphate backbone. Images by PyMOL (Schrödinger, LLC.).

7 | RECOGNITION OF m^{6}A RNA BY RNA-BINDING PROTEINS

RNA-binding proteins that can bind modified RNAs like m^{6}A use the same binding strategies as RNA-binding proteins that target unmodified RNA. m^{6}A has a methyl group added to the N^6 position of adenosine (Figure 1b). As a result, a nonpolar, bulky methyl group replaces a hydrogen, which could have formed hydrogen bonds. Similar to PUM-HD, RNA-binding proteins that target modified RNAs chemically accommodate their target to differentiate between other modified and unmodified RNAs. For example, the specificity of the YTH domain for m^{6}A is explained in the atomic-resolution crystal structure of YTHDF1 bound with m^{6}A-containing RNA (Figure 5a).[71] Like PUM-HD, the YTH domain has ringed amino acids that pi stack with the adenine base of m^{6}A. Positively charged amino acids such as lysine also form salt bridges with the RNA backbone, serving to nonspecifically attract RNA substrates. Differences are observed at the YTH specificity pocket for m^{6}A versus the PUM-HD pocket for unmodified adenosine. YTH forms a hydrophobic cage of three tryptophans surrounding the modification (Figure 5b). These interactions allow the protein to differentiate m^{6}A from an unmodified adenosine with a hydrophilic N^6 amino group. Additionally, the backbone of the YTH peptide chain hydrogen bonds with a nitrogen of m^{6}A to hold the modified base in place (Figure 5b). Lastly, an asparagine hydrogen bonds with the 2′ hydroxyl of the m^{6}A ribose, allowing YTH to differentiate RNA from DNA (Figure 5b). Thus, the specificity pocket of YTH is designed to accommodate a hydrophobic chemical modification to specify m^{6}A and differentiate from unmodified RNA or DNA.

To summarize, the YTHDF1 and hPUM2-HD have similar strategies to target specific RNA substrates. Firstly, both YTHDF1 and hPUM2-HD use amino acid side chains and a medley of chemical interactions to form a specificity pocket designed to accommodate the

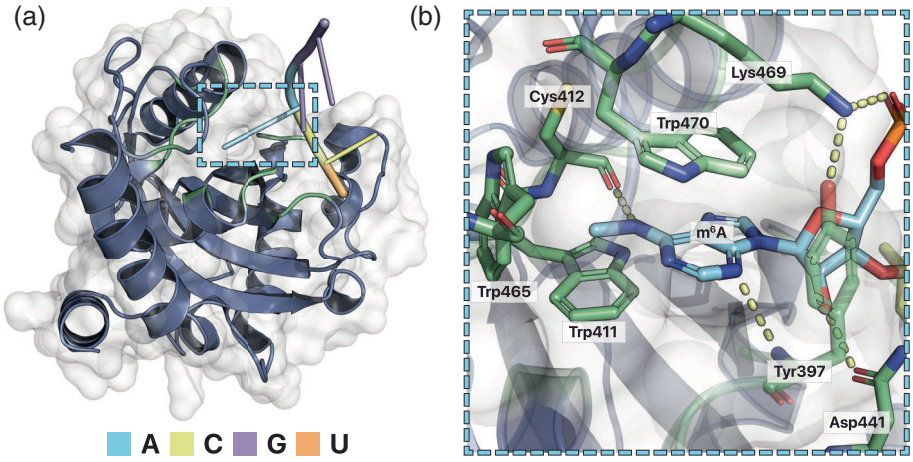

FIGURE 5 Structural features of YTH, a modified RNA-binding protein. (a) Crystal structure of human YTHDF1 in complex with N^6-methyladenosine (m^6A) RNA (PDB ID: 4RCJ[71]). The modified RNA-binding domain has a globular fold that forms a specificity pocket for the m^6A modified RNA base. RNA color coded. Dashed box enlarged. (b) The m^6A binding pocket. Tryptophan[411] (Trp411), tryptophan[465] (Trp465), and tryptophan[470] (Trp470) form a hydrophobic cage that envelopes m^6A. Pi stacking with tryptophan[470] and hydrogen bonds (dashed lines) with cysteine[412] (Cys412) and tyrosine[397] (Tyr397) help stabilize the m^6A nucleotide within the binding pocket. Salt bridges form between lysine[469] (Lys469) and the RNA phosphate backbone. The 2′ RNA hydroxyl group is recognized by asparagine[441] (Asp441). Images by PyMOL (Schrödinger, LLC.).

chemistry of the RNA targeted.[65,71] YTH predominantly employs hydrophobic interactions to form a pocket which accommodates the hydrophobic character of the m^6A methyl group. PUM-HD utilizes hydrophilic interactions to drive its pocket specificity for the unmodified adenosine nitrogen. Secondly, amino acids pi stack to present the base in a proper position for the binding pocket. And thirdly, positively charged residues on the protein's surface attract the RNA phosphate backbone to nonspecifically enhance its affinity for all RNA substrates. Other RNA-binding proteins follow the same principles that can also be appreciated in high resolution, RNA–protein structures.

8 | NEW FRONTIERS IN RNA-BINDING PROTEINS

The central dogma of molecular biology outlines the flow of genetic information from DNA to RNA to protein. The pivotal position of RNA, situated in between the DNA responsible for heredity and the proteins, which represent functional products, renders it a key point for further research in the field of molecular biology. RNA modifications expand the RNA alphabet beyond the four standard ribonucleotides by introducing diverse alterations to their chemical structure. The modifications are created or removed by enzymes, categorized either as writers or erasers. These enzymes play a necessary role in biology for gene regulation, development, obesity, and cancer.[16–18,21–23,32,36,72–74] Readers have evolved to selectively bind distinct unmodified and modified RNA. While differing in structure and sequence, these binding proteins use a conserved set of principles to recognize target RNA. The differences and similarities of RNA-binding proteins are on full display in atomic-resolution RNA–protein structural models.

The study of RNA modifications is in a renaissance and undergoing exponential growth. Only a handful of modifications have been fully characterized, in part because of the lack of methods to identify their RNA targets and sites. Some methods use chemistry or RNA-binding proteins[75] to identify the sites, but these methods must be specifically tailored to each RNA modification. Universal methods to identify any type of RNA modification are challenging but also in the infant stages of development.[76,77] Identifying RNA modification sites provides a starting point for understanding how RNA modifications affect RNA stability, folding, and function. Thus, the development of new, accurate identification methods will be key to investigate the link between currently uncharacterized RNA modifications, biology, and disease.[76,78–81] As discussed in this review, one key mechanism of m^6A is the recruitment of RNA-binding proteins for RNA regulation. A safe prediction is that other RNA modifications will also recruit or prevent interactions with RNA-binding proteins as their biological mechanism. Similar to YTH, PUM-HD, and others, these RNA-binding proteins will undoubtedly follow similar strategies to recognize subtle chemical differences of

modified RNA to deliver a profound impact on RNA form and function.

FUNDING INFORMATION

S.T.A. is funded by the NIH/NIGMS (R35 GM142691) and received start-up funds from the Indiana University School of Medicine and its Precision Health Initiative (PHI).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings in this study is available on request from the corresponding author.

ORCID

Scott Takeo Aoki https://orcid.org/0000-0002-1100-4600

REFERENCES

1. Crick FH. On protein synthesis. Symp Soc Exp Biol. 1958;12:138–63.
2. Ille AM, Lamont H, Mathews MB. The central dogma revisited: insights from protein synthesis, CRISPR, and beyond. Wiley Interdiscip Rev RNA. 2022;13(5):e1718. https://doi.org/10.1002/wrna.1718
3. Minchin S, Lodge J. Understanding biochemistry: structure and function of nucleic acids. Essays Biochem. 2019;63(4):433–56. https://doi.org/10.1042/EBC20180038
4. Takahashi I, Marmur J. Replacement of thymidylic acid by deoxyuridylic acid in the deoxyribonucleic acid of a transducing phage for *Bacillus subtilis*. Nature. 1963;197:794–5. https://doi.org/10.1038/197794a0
5. Cohn WE, Volkin E. Nucleoside-5′-phosphates from ribonucleic acid. Nature. 1951;167(4247):483–4. https://doi.org/10.1038/167483a0
6. Boccaletto P, Stefaniak F, Ray A, Cappannini A, Mukherjee S, Purta E, et al. MODOMICS: a database of RNA modification pathways. 2021 update. Nucleic Acids Res. 2022;50(D1):D231–5. https://doi.org/10.1093/nar/gkab1083
7. Cantara WA, Crain PF, Rozenski J, McCloskey JA, Harris KA, Zhang X, et al. The RNA modification database, RNAMDB: 2011 update. Nucleic Acids Res. 2011;39:D195–201. https://doi.org/10.1093/nar/gkq1028
8. Lorenz C, Lünse C, Mörl M. tRNA modifications: impact on structure and thermal adaptation. Biomolecules. 2017;7(4):35. https://doi.org/10.3390/biom7020035
9. Cappannini A, Ray A, Purta E, Mukherjee S, Boccaletto P, Moafinejad SN, et al. MODOMICS: a database of RNA modifications and related information. 2023 update. Nucleic Acids Res. 2024;52(D1):D239–44. https://doi.org/10.1093/nar/gkad1083
10. McCown PJ, Ruszkowska A, Kunkler CN, Breger K, Hulewicz JP, Wang MC, et al. Naturally occurring modified ribonucleosides. Wiley Interdiscip Rev RNA. 2020;11(5):e1595. https://doi.org/10.1002/wrna.1595
11. Hong M, Tao S, Zhang L, Diao LT, Huang X, Huang S, et al. RNA sequencing: new technologies and applications in cancer research. J Hematol Oncol. 2020;13:166. https://doi.org/10.1186/s13045-020-01005-x
12. Alseth I, Dalhus B, Bjoras M. Inosine in DNA and RNA. Curr Opin Genet Dev. 2014;26:116–23. https://doi.org/10.1016/j.gde.2014.07.008
13. Fu Y, Jia G, Pang X, Wang RN, Wang X, Li CJ, et al. FTO-mediated formation of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA. Nat Commun. 2013;4:1798. https://doi.org/10.1038/ncomms2822
14. Patil DP, Pickering BF, Jaffrey SR. Reading m(6) A in the transcriptome: m(6)A-binding proteins. Trends Cell Biol. 2018;28(2):113–27. https://doi.org/10.1016/j.tcb.2017.10.001
15. Shi H, Wei J, He C. Where, when, and how: context-dependent functions of RNA methylation writers, readers, and erasers. Mol Cell. 2019;74(4):640–50. https://doi.org/10.1016/j.molcel.2019.04.025
16. Lin S, Choe J, Du P, Triboulet R, Gregory IR. The m 6 A methyltransferase METTL3 promotes translation in human cancer cells. Mol Cell. 2016;62(3):335–45. https://doi.org/10.1016/j.molcel.2016.03.021
17. McCown PJ, Wang MC, Jaeger L, Brown JA. Secondary structural model of human MALAT1 reveals multiple structure-function relationships. Int J Mol Sci. 2019;20(22):5610. https://doi.org/10.3390/ijms20225610
18. Oakes E, Anderson A, Cohen-Gadol A, Hundley HA. Adenosine deaminase that acts on RNA 3 (ADAR3) binding to glutamate receptor subunit B pre-mRNA inhibits RNA editing in glioblastoma. J Biol Chem. 2017;292(10):4326–35. https://doi.org/10.1074/jbc.m117.779868
19. Zhao Y, Zhao Q, Kaboli PJ, Shen J, Li M, Wu X, et al. m1A regulated genes modulate PI3K/AKT/mTOR and ErbB pathways in gastrointestinal cancer. Transl Oncol. 2019;12(10):1323–33. https://doi.org/10.1016/j.tranon.2019.06.007
20. Jiang X, Liu B, Nie Z, Duan L, Xiong Q, Jin Z, et al. The role of m6A modification in the biological functions and diseases. Signal Transduct Target Ther. 2021;6(1):74. https://doi.org/10.1038/s41392-020-00450-x
21. Choe J, Lin S, Zhang W, Liu Q, Wang L, Ramirez-Moya J, et al. mRNA circularization by METTL3–eIF3h enhances translation and promotes oncogenesis. Nature. 2018;561(7724):556–60. https://doi.org/10.1038/s41586-018-0538-8
22. Paris J, Morgan M, Campos J, Spencer GJ, Shmakova A, Ivanova I, et al. Targeting the RNA m6A reader YTHDF2 selectively compromises cancer stem cells in acute myeloid leukemia. Cell Stem Cell. 2019;25(1):137–148.e6. https://doi.org/10.1016/j.stem.2019.03.021
23. Yoon K-J, Ringeling FR, Vissers C, Jacob F, Pokrass M, Jimenez-Cyrus D, et al. Temporal control of mammalian cortical neurogenesis by m6A methylation. Cell. 2017;171(4):877–889.e17. https://doi.org/10.1016/j.cell.2017.09.003
24. Desrosiers R, Friderici K, Rottman F. Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. Proc Natl Acad Sci. 1974;71(10):3971–5.
25. Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature. 2012;485(7397):201–6. https://doi.org/10.1038/nature11112
26. Huang W, Chen T-Q, Fang K, Zeng Z-C, Ye H, Chen Y-Q. N6-methyladenosine methyltransferases: functions, regulation, and clinical potential. J Hematol Oncol. 2021;14(1):117. https://doi.org/10.1186/s13045-021-01129-8

27. Liu J, Yue Y, Han D, Wang X, Fu Y, Zhang L, et al. A METTL3–METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat Chem Biol. 2014;10(2):93–5. https://doi.org/10.1038/nchembio.1432
28. Bokar JA, Rath-Shambaugh ME, Ludwiczak R, Narayan P, Rottman F. Characterization and partial purification of mRNA N6-adenosine methyltransferase from HeLa cell nuclei. Internal mRNA methylation requires a multisubunit complex. J Biol Chem. 1994;269(26):17697–704.
29. Harper JE, Miceli SM, Roberts RJ, Manley JL. Sequence specificity of the human mRNA N6-adenosine methylase *in vitro*. Nucleic Acids Res. 1990;18(19):5735–41. https://doi.org/10.1093/nar/18.19.5735
30. Wei C-M, Moss B. Nucleotide sequences at the N^6-methyladenosine sites of HeLa cell messenger ribonucleic acid. Biochemistry. 1977;16(8):1672–6. https://doi.org/10.1021/bi00627a023
31. Zhao X, Yang Y, Sun B-F, Shi Y, Yang X, Xiao W, et al. FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis. Cell Res. 2014;24(12):1403–19. https://doi.org/10.1038/cr.2014.151
32. Zheng G, Dahl AJ, Niu Y, Fedorcsak P, Huang C-M, Li JC, et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol Cell. 2013;49(1):18–29. https://doi.org/10.1016/j.molcel.2012.10.015
33. Zhang B, Wu Q, Li B, Wang D, Wang L, Zhou YL. m6A regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in gastric cancer. Mol Cancer. 2020;19(1):53. https://doi.org/10.1186/s12943-020-01170-0
34. Shen C, Xuan B, Yan T, Ma Y, Xu P, Tian X, et al. m6A-dependent glycolysis enhances colorectal cancer progression. Mol Cancer. 2020;19(1):72. https://doi.org/10.1186/s12943-020-01190-w
35. Chen M, Wei L, Law CT, Tsang FHC, Shen J, Cheng CLH, et al. RNA N6-methyladenosine methyltransferase-like 3 promotes liver cancer progression through YTHDF2-dependent posttranscriptional silencing of SOCS2. Hepatology. 2018;67(6):2254–70. https://doi.org/10.1002/hep.29683
36. Fawcett KA, Barroso I. The genetics of obesity: FTO leads the way. Trends Genet. 2010;26(6):266–74. https://doi.org/10.1016/j.tig.2010.02.006
37. Ohira T, Suzuki T. Transfer RNA modifications and cellular thermotolerance. Mol Cell. 2024;84(1):94–106. https://doi.org/10.1016/j.molcel.2023.11.041
38. Liao S, Sun H, Xu C. YTH domain: a family of N(6)-methyladenosine (m(6)a) readers. Genomics Proteomics Bioinformatics. 2018;16(2):99–107. https://doi.org/10.1016/j.gpb.2018.04.002
39. Du H, Zhao Y, He J, Zhang Y, Xi H, Liu M, et al. YTHDF2 destabilizes m6A-containing RNA through direct recruitment of the CCR4–NOT deadenylase complex. Nat Commun. 2016;7(1):12626. https://doi.org/10.1038/ncomms12626
40. Stoilov P, Rafalska I, Stamm S. YTH: a new domain in nuclear proteins. Trends Biochem Sci. 2002;27(10):495–7. https://doi.org/10.1016/S0968-0004(02)02189-8
41. Corley M, Burns MC, Yeo GW. How RNA-binding proteins interact with RNA: molecules and mechanisms. Mol Cell. 2020;78(1):9–29. https://doi.org/10.1016/j.molcel.2020.03.011
42. Anderson JT, Droogmans L, Grosjean H. Fine-tuning of RNA functions by modification and editing. Germany: Springer. Vol 12; 2005. p. 121–39.
43. Balcerak A, Trebinska-Stryjewska A, Konopinski R, Wakula M, Grzybowska EA. RNA–protein interactions: disorder, moonlighting and junk contribute to eukaryotic complexity. Open Biol. 2019;9(6):190096. https://doi.org/10.1098/rsob.190096
44. Järvelin AI, Noerenberg M, Davis I, Castello A. The new (dis)order in RNA regulation. Cell Commun Signal. 2016;14(1):9. https://doi.org/10.1186/s12964-016-0132-3
45. Herschlag D, Pinney MM. Hydrogen Bonds: Simple after All? Biochemistry. 2018;57(24):3338–52. https://doi.org/10.1021/acs.biochem.8b00217
46. Hofweber M, Dormann D. Friend or foe—post-translational modifications as regulators of phase separation and RNP granule dynamics. J Biol Chem. 2019;294(18):7137–50. https://doi.org/10.1074/jbc.TM118.001189
47. Gupta A, Gribskov M. The role of RNA sequence and structure in RNA—protein interactions. J Mol Biol. 2011;409(4):574–87. https://doi.org/10.1016/j.jmb.2011.04.007
48. Treger M, Westhof E. Statistical analysis of atomic contacts at RNA-protein interfaces. J Mol Recognit. 2001;14(4):199–214. https://doi.org/10.1002/jmr.534
49. Dill KA, Ozkan SB, Shell MS, Weikl TR. The protein folding problem. Annu Rev Biophys. 2008;37(1):289–316. https://doi.org/10.1146/annurev.biophys.37.092707.153558
50. Onofrio A, Parisi G, Punzi G, Todisco S, Di Noia MA, Bossis F, et al. Distance-dependent hydrophobic–hydrophobic contacts in protein folding simulations. Phys Chem Chem Phys. 2014;16(35):18907–17. https://doi.org/10.1039/c4cp01131g
51. Allain FHT. Structural basis of the RNA-binding specificity of human U1A protein. EMBO J. 1997;16(18):5764–72. https://doi.org/10.1093/emboj/16.18.5764
52. Yang Y. Solution structure of the LicT-RNA antitermination complex: CAT clamping RAT. EMBO J. 2002;21(8):1987–97. https://doi.org/10.1093/emboj/21.8.1987
53. Yu Q, Ye W, Jiang C, Luo R, Chen H-F. Specific recognition mechanism between RNA and the KH3 domain of Nova-2 protein. J Phys Chem B. 2014;118(43):12426–34. https://doi.org/10.1021/jp5079289
54. Zhu T, Roundtree IA, Wang P, Wang X, Wang L, Sun C, et al. Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N6-methyladenosine. Cell Res. 2014;24(12):1493–6. https://doi.org/10.1038/cr.2014.152
55. Hu W, Qin L, Li M, Pu X, Guo Y. A structural dissection of protein–RNA interactions based on different RNA base areas of interfaces. RSC Adv. 2018;8(19):10582–92. https://doi.org/10.1039/c8ra00598b
56. Petrucci RH, Herring FG, Madura JD, Bissonnette C. 1997 General Chemistry: Principles and Modern Applications. United Kingdom: Pearson Education.
57. Allers J, Shamoo Y. Structure-based analysis of protein-RNA interactions using the program ENTANGLE. J Mol Biol. 2001;311(1):75–86. https://doi.org/10.1006/jmbi.2001.4857
58. Jones S, Daley DT, Luscombe NM, Berman HM, Thornton JM. Protein-RNA interactions: a structural analysis. Nucleic Acids Res. 2001;29(4):943–54. https://doi.org/10.1093/nar/29.4.943
59. Wilson KA, Holland DJ, Wetmore SD. Topology of RNA–protein nucleobase–amino acid π–π interactions and

comparison to analogous DNA–protein π–π contacts. RNA. 2016;22(5):696–708. https://doi.org/10.1261/rna.054924.115
60. Oubridge C, Ito N, Evans PR, Teo CH, Nagai K. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. Nature. 1994; 372(6505):432–8. https://doi.org/10.1038/372432a0
61. Quenault T, Lithgow T, Traven A. PUF proteins: repression, activation and mRNA localization. Trends Cell Biol. 2011; 21(2):104–12. https://doi.org/10.1016/j.tcb.2010.09.013
62. Wickens M, Bernstein DS, Kimble J, Parker R. A PUF family portrait: 3'UTR regulation as a way of life. Trends Genet. 2002; 18(3):150–7. https://doi.org/10.1016/s0168-9525(01)02616-6
63. Goldstrohm AC, Hall TMT, McKenney KM. Post-transcriptional regulatory functions of mammalian Pumilio proteins. Trends Genet. 2018;34(12):972–90. https://doi.org/10.1016/j.tig.2018.09.006
64. Wang X, Zamore PD, Hall TMT. Crystal structure of a Pumilio homology domain. Mol Cell. 2001;7(4):855–65. https://doi.org/10.1016/s1097-2765(01)00229-5
65. Lu G, Hall TMT. Alternate modes of cognate RNA recognition by human PUMILIO proteins. Structure. 2011;19(3):361–7. https://doi.org/10.1016/j.str.2010.12.019
66. Wang X, McLachlan J, Zamore PD, Hall TM. Modular recognition of RNA by a human pumilio-homology domain. Cell. 2002; 110(4):501–12. https://doi.org/10.1016/s0092-8674(02)00873-5
67. Gerber AP, Luschnig S, Krasnow MA, Brown PO, Herschlag D. Genome-wide identification of mRNAs associated with the translational regulator PUMILIO in *Drosophila melanogaster*. Proc Natl Acad Sci. 2006;103(12):4487–92. https://doi.org/10.1073/pnas.0509260103
68. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, et al. Transcriptome-wide identification of RNA-binding protein and MicroRNA target sites by PAR-CLIP. Cell. 2010;141(1):129–41. https://doi.org/10.1016/j.cell.2010.03.009
69. Morris AR, Mukherjee N, Keene JD. Ribonomic analysis of human Pum1 reveals *cis-trans* conservation across species despite evolution of diverse mRNA target sets. Mol Cell Biol. 2008;28(12):4093–103. https://doi.org/10.1128/mcb.00155-08
70. Campbell ZT, Valley CT, Wickens M. A protein-RNA specificity code enables targeted activation of an endogenous human transcript. Nat Struct Mol Biol. 2014;21(8):732–8. https://doi.org/10.1038/nsmb.2847
71. Xu C, Liu K, Ahmed H, Loppnau P, Schapira M, Min J. Structural basis for the discriminative recognition of N6-Methyladenosine RNA by the human YT521-B homology domain family of proteins. J Biol Chem. 2015;290(41):24902–13. https://doi.org/10.1074/jbc.m115.680389
72. Zhou H-X, Pang X. Electrostatic interactions in protein structure, folding, binding, and condensation. Chem Rev. 2018; 118(4):1691–741. https://doi.org/10.1021/acs.chemrev.7b00305
73. Chen M, Wong C-M. The emerging roles of N6-methyladenosine (m6A) deregulation in liver carcinogenesis. Mol Cancer. 2020;19(1):44. https://doi.org/10.1186/s12943-020-01172-y
74. Huang Q, Mo J, Liao Z, Chen X, Zhang B. The RNA m6A writer WTAP in diseases: structure, roles, and mechanisms. Cell Death Dis. 2022;13(10):852. https://doi.org/10.1038/s41419-022-05268-9
75. Wang Y, Xiao Y, Dong S, Yu Q, Jia G. Antibody-free enzyme-assisted chemical approach for detection of N6-methyladenosine. Nat Chem Biol. 2020;16(8):896–903. https://doi.org/10.1038/s41589-020-0525-x
76. Garalde DR, Snell EA, Jachimowicz D, Sipos B, Lloyd JH, Bruce M, et al. Highly parallel direct RNA sequencing on an array of nanopores. Nat Methods. 2018;15(3):201–6. https://doi.org/10.1038/nmeth.4577
77. Zhang Y, Lu L, Li X. Detection technologies for RNA modifications. Exp Mol Med. 2022;54(10):1601–16. https://doi.org/10.1038/s12276-022-00821-0
78. Acera Mateos P, Zhou Y, Zarnack K, Eyras E. Concepts and methods for transcriptome-wide prediction of chemical messenger RNA modifications with machine learning. Brief Bioinform. 2023;24(3):1–14. https://doi.org/10.1093/bib/bbad163
79. Carlile TM, Rojas-Duran MF, Zinshteyn B, Shin H, Bartoli KM, Gilbert WV. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature. 2014; 515(7525):143–6. https://doi.org/10.1038/nature13802
80. Delatte B, Wang F, Ngoc LV, Collignon E, Bonvin E, Deplus R, et al. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. Science. 2016;351(6270):282–5. https://doi.org/10.1126/science.aac5253
81. Khoddami V, Yerra A, Mosbruger TL, Fleming AM, Burrows CJ, Cairns BR. Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution. Proc Natl Acad Sci. 2019;116(14):6784–9. https://doi.org/10.1073/pnas.1817334116
82. Aoki ST, Lynch TR, Crittenden SL, Bingman CA, Wickens M, Kimble JC. Elegans germ granules require both assembly and localized regulators for mRNA repression. Nat Commun. 2021; 12(1):996. https://doi.org/10.1038/s41467-021-21278-1

How to cite this article: Angelo M, Bhargava Y, Aoki ST. A primer for junior trainees: Recognition of RNA modifications by RNA-binding proteins. Biochem Mol Biol Educ. 2024. https://doi.org/10.1002/bmb.21854