

Characterization of epitranscriptome reader proteins experimentally and *in silico*: Current knowledge and future perspectives beyond the YTH domain

Lucas G. Miller,^a Madeline Demny,^b Phanourios Tamamis,^{b,c,*} and Lydia M. Contreras^{a,d,**}

^aMcKetta Department of Chemical Engineering, The University of Texas at Austin, Austin, TX, USA

^bArtie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX, USA

^cDepartment of Materials Science & Engineering, Texas A&M University, College Station, TX, USA

^dInstitute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, TX, USA

Abstract:

To date, over 150 chemical modifications to the four canonical RNA bases have been discovered, known collectively as the epitranscriptome. Many of these modifications have been implicated in a variety of cellular processes and disease states. Additional work has been done to identify proteins known as “readers” that selectively interact with RNAs that contain specific chemical modifications. Protein interactomes with N6-methyladenosine (m⁶A), N1-methyladenosine (m¹A), N5-methylcytosine (m⁵C), and 8-oxo-7,8-dihydroguanosine (8-oxoG) have been determined, mainly through experimental advances in proteomics techniques. However, relatively few proteins have been confirmed to bind directly to RNA containing these modifications. Furthermore, for many of these protein readers, the exact binding mechanisms as well as the exclusivity for recognition of modified RNA species remain elusive, leading to questions regarding their roles within different cellular processes. In the case of the YT-521B homology (YTH) family of proteins, both experimental and *in silico* techniques have been leveraged to provide valuable biophysical insights into the mechanisms of m⁶A recognition at atomic resolution. To date, the YTH family is one of the best characterized classes of readers. Here, we review current knowledge about epitranscriptome recognition of the YTH domain proteins from previously published experimental and computational studies. We additionally outline knowledge gaps for proteins beyond the well-studied human YTH domains and the current *in silico* techniques and resources that can enable investigation of protein interactions with modified RNA outside of the YTH-m⁶A context.

Keywords: Epitranscriptomics, RNA binding proteins, YT-521B Homology (YTH) protein family, Molecular dynamics (MD) simulations, AlphaFold, Protein Structure Database

1. Introduction

Since the initial discovery of chemical modifications to RNA in the form of pseudouridine in 1957 [1], [2], published work has catalogued upwards of 150 modified versions of the four canonical RNA bases that make up what is known as the epitranscriptome; these modifications can be found across all domains of life and many types of RNA [3]. Modified bases have been mostly detected using several techniques that include separation and analysis by RNA chromatography and mass spectrometry methods [4]. More recently, next-generation sequencing has also been used to identify RNA modifications due to differences in their chemical properties from their unmodified base equivalents [3], [5]. Modified RNA species have been found in high abundance in transfer RNA (tRNA) and ribosomal RNA (rRNA) but can also be found in messenger RNA (mRNA) and other long non-coding RNAs (lncRNAs) [3]. The most abundant modified RNA base in eukaryotic mRNA identified thus far, N6-methyladenosine (m6A), has provided a wealth of insight into how RNA modifications might accumulate on transcripts and how their presence impacts cellular processes. For instance, the m6A modification has been shown to affect processes like alternative splicing of pre-mRNAs [5], [6], [7], cell growth and differentiation [8], [9], [10], and RNA localization [11].

Characterization of the enzymes that deposit or remove the m6A mark has led to a more generalized model for dynamic “writing” or “erasing” of modifications to RNA (Fig. 1). In the case of m6A, enzymes known as “writer” proteins (e.g., the METTL3/METTL14 complex [12]) add a methyl group to adenine to form m6A; removal of this chemical adduct is accomplished via enzymes referred to as “eraser” proteins (e.g., FTO [13]) [14]. These proteins act to strike a cellular balance of m6A levels which, when disturbed, have implications in cancer proliferation and dysregulation of cellular processes [15]. In addition to the writers and erasers, the m6A mark is recognized by a class of proteins known as “readers” that bind to RNA containing this modification. The reader proteins for m6A selectively bind to modified transcripts over unmodified ones through either direct binding to the modified base and the bases flanking it, as is the case for the YT521-B homology (YTH) domain proteins [16] and the IGF2 binding proteins [17], or through indirect interactions with regions adjacent to the modified site, particularly in structured RNA regions, as with the HNRNPC protein [18]. These interactions between reader proteins and modified RNA can direct different transcripts for processes such as translation initiation, enhanced degradation, or for RNA localization [11], [19], [20], [21], [22], [23]. Other modifications are regulated by similar classes of proteins such as the ADAR1 writer protein for inosine [24], [25], [26] and the ALKBH eraser protein family for RNA methylations such as N1-methyladenosine (m1A), N3-methylcytosine (m3C), and m6A [27]. Although overall discovery and characterization of the writer and eraser proteins is important for understanding the regulatory networks of these modified RNAs (reviewed in [13], [27]), this minireview focuses on the reader proteins of m6A, m1A, N5-methylcytosine (m5C), and 8-oxo-7,8-dihydroguanosine (8-oxoG) [28], [29], [30]. Additionally, we acknowledge the numerous works to characterize the RNA sequence specificity of these protein readers [16], [31], [32], [33], [34] but limit our discussion to the biophysical investigations of protein-RNA interactions specific to modified RNA. To that end, we showcase the mechanistic understanding of protein interactions with the epitranscriptome that has been generated using both experimental and computational techniques. We discuss the power of the investigations that characterize the interactions between the YTH domain proteins and m6A, but we also outline current gaps in knowledge around the binding mechanisms for non-YTH readers

of m6A and readers for other RNA modifications. Finally, we discuss the challenges and promise of conducting in silico biophysical studies of protein interactions with modified RNA and highlight current work in characterizing these interactions.

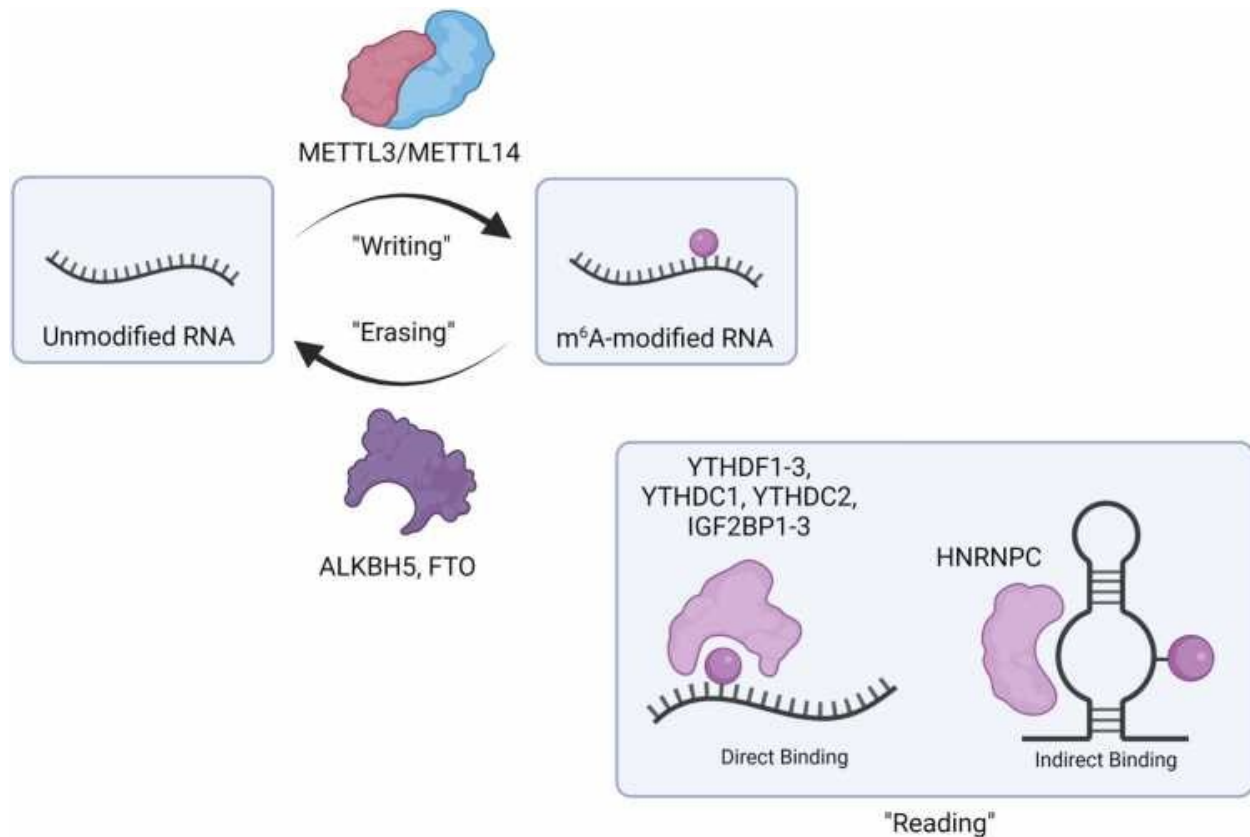


Fig. 1: Classes of proteins that regulate N6-methyladenosine (m6A). The METTL3/METTL14 “writer proteins” acts to deposit a methyl group onto its RNA substrate to form an m6A-modified RNA [12]. This modification can be removed through the oxidative demethylation process catalyzed by the ALKBH family of proteins, including ALKBH5 and FTO [13], [14]. The m6A-modified RNA can be recognized by different “reader proteins” such as the YTH domain family and the IGF2BP family (IGF2BP1–3) through direct interactions with the RNA modification [9], [17]. Additionally, m6A can affect the local RNA structure of its transcript, allowing for indirect binding interactions with proteins such as HNRNPC [18]. Created with BioRender.com.

2. Characterization of the YT-521B homology (YTH) family using in vitro and in silico techniques

2.1. Discovery of the YTH protein family as m6A readers

Initial identification of proteins that recognize modified RNA transcripts has been accomplished through large scale RNA affinity chromatography studies [29], [30], [35], [36], [37]. In methods of this kind, a modified transcript acts as a “bait” for proteins that bind directly and indirectly (i.e., through protein-protein interactions or possibly RNA structure-specific interactions) to the modified RNA. These proteins are then isolated from whole cell lysate through affinity pulldown of the RNA [35], and mass spectrometry is used to determine the identity of the associated proteins. One such study [37] identified two proteins, YTHDF2 and YTHDF3, that interacted preferentially with an m6A-containing RNA bait versus an unmodified control. This was the first study to identify the YTHDF2 and YTHDF3 proteins as m6A interactors. In the case of YTHDF2 and YTHDF3, both proteins contain a YTH domain that is widely conserved across eukaryotes [31], [38]. After the discovery of YTHDF2 and YTHDF3 as putative m6A readers, these proteins as well as the three others found in humans that contain the YTH domain (YTHDF1, YTHDC1, and YTHDC2) were investigated for their recognition of m6A-containing transcripts in vitro [21], [28], [39], [40], [41], [42], [43], [44], [45]. Amino acid sequences corresponding to either the full YTH protein of interest or the isolated YTH domains were expressed recombinantly in *E. coli*, purified, and subjected to in vitro binding experiments (Table 1, [21], [28], [39], [40], [42], [43], [44], [45], [46]). These binding experiments, whose base principles are detailed in [47], [48], [49], [50], allowed for the calculation of protein affinity for a variety of RNA substrates via the dissociation constant (KD). It is difficult to directly compare the reported KD values in some cases because of the effects of different buffer compositions, incubation temperatures, and incubation times on these values [51], [52]. However, common observations emerge from these studies regarding RNA containing m6A or the unmodified adenine. Regardless of RNA length, the YTH domains showed micromolar to sub-micromolar affinities and selectivity for m6A with respect to adenine. Specifically, these proteins showed either no binding to unmodified RNA [28], [45] or at least an order of magnitude difference in the KD value to unmodified RNA for those studies that did show binding [21], [28], [40], [46]. An exception to this selectivity difference can be found in the electrophoretic mobility shift assays (EMSAs) for YTHDC2 [43] which showed a roughly two-fold increase in affinity for m6A-containing RNA, relative to unmodified RNA. Another common observation regarding m6A binding can be found in the effects of RNA sequence and length on KD. The YTH family of proteins has been shown to recognize predominantly a RRACH sequence motif, where R represents a purine base (A or G), A is modified to m6A, and H represents either A, C, or U [9], [53], [54]. Alterations of this sequence motif, particularly in the position preceding the m6A modification (“G-1”), show the preference of YTHDC1 for RGAC [39], [42], [45]. Mutation of the “G-1” position to A lead to an increase in KD from 0.3 μ M to 2.0 μ M for YTHDC1 [39], whereas the same RNA mutation shows only an increase of 1.0–1.1 μ M for binding assays with YTHDF1. Additionally, 5-mer m6A-containing RNA sequences bound at lower levels of affinity to YTHDF1 and YTHDC1, which could be linked to the importance of regions flanking m6A for stabilizing the protein-RNA complex [42], [45]. Although these binding experiments with the wild-type protein sequences demonstrate the selectivity of the human YTH proteins and—more specifically—their domains for m6A, the key regions responsible for this selectivity needed to be determined. To elucidate the protein features responsible for the selective recognition of m6A containing transcripts, crystal structures were resolved in tandem with binding characterization efforts for all human YTH domains [39], [40], [41], [42], [43], [45], [55], [56], [57], [58], [59], [60].

Table 1

In vitro binding affinities for the human YTH domain family of proteins.

Protein	Domain/Full Protein	Technique	Incubation conditions	Oligomer Sequence (5' - 3')	K _d (nM)	Reference
YTHDF2	Full Protein	EMSA	Ice, 30 min	AUGGGCCGUUCAUCUGCUAAAAGG(m ⁶ A)CUGCUUUUGGGCUUGU	179 ± 47	[21]
YTHDF2	Full Protein	EMSA	Ice, 30 min	AUGGGCCGUUCAUCUGCUAAAAGACUGCUUUUGGGCUUGU	2844 ± 656	[21]
YTHDF2	Full Protein	EMSA	Ice, 30 min	AUGGGCCGUUCAUCUGCUAAAACU(m ⁶ A)CUGCUUUUGGGCUUGU	520 ± 155	[21]
YTHDF2	Full Protein	EMSA	Ice, 30 min	AUGGGCCGUUCAUCUGCUAAAAGACUGCUUUUGGGCUUGU	5187 ± 1330	[21]
YTHDF3	Full Protein	EMSA	Ice, 30 min	AUGGGCCGUUCAUCUGCUAAAAGG(m ⁶ A)CUGCUUUUGGGCUUGU	323 ± 119	[21]
YTHDF3	Full Protein	EMSA	Ice, 30 min	AUGGGCCGUUCAUCUGCUAAAAGACUGCUUUUGGGCUUGU	1673 ± 1149	[21]
YTHDF1	Full Protein	EMSA	Ice, 30 min	AUGGGCCGUUCAUCUGCUAAAAGG(m ⁶ A)CUGCUUUUGGGCUUGU	255 ± 46	[21]
YTHDC1	Domain	EMSA	4 °C, 1 hr	biotin-CCGUUCCGCCC(m ⁶ A)GGCCGCGCCCAGCUGGAAUGCA	700 ± 100	[28]
YTHDC1	Domain	EMSA	4 °C, 1 hr	biotin-CCGUUCCGCCC(m ¹ A)GGCCGCGCCCAGCUGGAAUGCA	(23.3 ± 2.1) • 10 ³	[28]
YTHDC1	Domain	EMSA	4 °C, 1 hr	biotin-CCGUUCCGCCCAGGCCGCGCCCAGCUGGAAUGCA	(68.8 ± 13.9) • 10 ³	[28]
YTHDF1	Full protein	EMSA	4 °C, 1 hr	biotin-CCGUUCCGCCC(m ⁶ A)GGCCGCGCCCAGCUGGAAUGCA	1300 ± 100	[28]
YTHDF1	Full protein	EMSA	4 °C, 1 hr	biotin-CCGUUCCGCCC(m ¹ A)GGCCGCGCCCAGCUGGAAUGCA	(16.5 ± 1.5) • 10 ³	[28]
YTHDF1	Full protein	EMSA	4 °C, 1 hr	biotin-CCGUUCCGCCCAGGCCGCGCCCAGCUGGAAUGCA	NB	[28]
YTHDF2	Full protein	EMSA	4 °C, 1 hr	biotin-CCGUUCCGCCC(m ⁶ A)GGCCGCGCCCAGCUGGAAUGCA	1300 ± 100	[28]
YTHDF2	Full protein	EMSA	4 °C, 1 hr	biotin-CCGUUCCGCCC(m ¹ A)GGCCGCGCCCAGCUGGAAUGCA	5800 ± 1700	[28]
YTHDF2	Full protein	EMSA	4 °C, 1 hr	biotin-CCGUUCCGCCCAGGCCGCGCCCAGCUGGAAUGCA	NB	[28]
YTHDF3	Full protein	EMSA	4 °C, 1 hr	biotin-CCGUUCCGCCC(m ⁶ A)GGCCGCGCCCAGCUGGAAUGCA	1900 ± 100	[28]
YTHDF3	Full protein	EMSA	4 °C, 1 hr	biotin-CCGUUCCGCCC(m ¹ A)GGCCGCGCCCAGCUGGAAUGCA	7000 ± 1000	[28]
YTHDF3	Full protein	EMSA	4 °C, 1 hr	biotin-CCGUUCCGCCCAGGCCGCGCCCAGCUGGAAUGCA	NB	[28]
YTHDC1	Domain	ITC	25 °C	GAACCGA(m ⁶ A)CUGUCUUA	2000 ± 400	[39]
YTHDC1	Domain	ITC	25 °C	GAACCGG(m ⁶ A)CUGUCUUA	300 ± 60	[39]
YTHDC1	Domain	ITC	25 °C	GAACCGC(m ⁶ A)CUGUCUUA	500 ± 120	[39]
YTHDC1	Domain	ITC	25 °C	GAACCGU(m ⁶ A)CUGUCUUA	400 ± 70	[39]
YTHDC1	Domain	ITC	25 °C	AAGAACCGG(m ⁶ A)CUGUCUUA	310 ± 70	[39]
YTHDC1	Domain	ITC	25 °C	AG(m ⁶ A)CU	3800 ± 400	[39]
YTHDC1	Domain	ITC	25 °C	GG(m ⁶ A)CU	2000 ± 100	[39]

Protein	Domain/Full Protein	Technique	Incubation conditions	Oligomer Sequence (5' - 3')	K _D (nM)	Reference
YTHDC1	Domain	ITC	25 °C	UG(m ⁶ A)CU	4300 ± 500	[39]
YTHDF2	Domain	FP	4 °C, 30 min	FAM-UUCUUCUGUGGACUGUG	21.39 • 10 ³	[40]
YTHDF3	Domain	FP	4 °C, 30 min	FAM-UUCUUCUGUGG(m ⁶ A)CUGUG	2.54 • 10 ³	[40]
YTHDF1	Domain	ITC	25 °C	CCGA(m ⁶ A)CUGU	1100 ± 200	[45]
YTHDC1	Domain	ITC	25 °C	CCGA(m ⁶ A)CUGU	1000 ± 100	[45]
YTHDC1	Domain	ITC	25 °C	CCGG(m ⁶ A)CUGU	220 ± 30	[45]
YTHDC1	Domain	ITC	25 °C	CCGC(m ⁶ A)CUGU	320 ± 30	[45]
YTHDC1	Domain	ITC	25 °C	CCGU(m ⁶ A)CUGU	300 ± 60	[45]
YTHDF1	Domain	ITC	25 °C	CCGG(m ⁶ A)CUGU	800 ± 300	[45]
YTHDF1	Domain	ITC	25 °C	CCGC(m ⁶ A)CUGU	800 ± 200	[45]
YTHDF1	Domain	ITC	25 °C	CCGU(m ⁶ A)CUGU	900 ± 200	[45]
YTHDF1	Domain	ITC	25 °C	GAACCGA(m ⁶ A)CUGUCUUA	1100 ± 200	[45]
YTHDF1	Domain	ITC	25 °C	GAACCGG(m ⁶ A)CUGUCUUA	1000 ± 300	[45]
YTHDF1	Domain	ITC	25 °C	GAACCGC(m ⁶ A)CUGUCUUA	900 ± 200	[45]
YTHDF1	Domain	ITC	25 °C	GAACCGU(m ⁶ A)CUGUCUUA	1700 ± 400	[45]
YTHDF1	Domain	ITC	25 °C	AAGAACCGG(m ⁶ A)CUGUCUUA	1000 ± 100	[45]
YTHDF1	Domain	ITC	25 °C	AG(m ⁶ A)CU	(30 ± 4) • 10 ³	[45]
YTHDF1	Domain	ITC	25 °C	GG(m ⁶ A)CU	(22 ± 4) • 10 ³	[45]
YTHDF1	Domain	ITC	25 °C	UG(m ⁶ A)CU	(34 ± 4) • 10 ³	[45]
YTHDF1	Domain	ITC	25 °C	GAACCGACUGUCUUA	NB	[45]
YTHDF1	Domain	ITC	25 °C	GGACU	NB	[45]
YTHDF2	Domain	ITC	25 °C	CCGA(m ⁶ A)CUGU	900 ± 200	[45]
YTHDF2	Domain	ITC	25 °C	CCGG(m ⁶ A)CUGU	900 ± 100	[45]
YTHDF2	Domain	ITC	25 °C	CCGC(m ⁶ A)CUGU	700 ± 200	[45]
YTHDF2	Domain	ITC	25 °C	CCGU(m ⁶ A)CUGU	800 ± 200	[45]
YTHDF2	Domain	ITC	25 °C	GAACCGA(m ⁶ A)CUGUCUUA	1500 ± 300	[45]
YTHDF2	Domain	ITC	25 °C	GAACCGG(m ⁶ A)CUGUCUUA	1700 ± 400	[45]
YTHDF2	Domain	ITC	25 °C	GAACCGC(m ⁶ A)CUGUCUUA	1300 ± 200	[45]
YTHDF2	Domain	ITC	25 °C	GAACCGU(m ⁶ A)CUGUCUUA	2000 ± 600	[45]
YTHDF2	Domain	ITC	25 °C	GAACCGACUGUCUUA	NB	[45]

Protein	Domain/Full Protein	Technique	Incubation conditions	Oligomer Sequence (5' - 3')	K _D (nM)	Reference
YTHDF2	Domain	ITC	25 °C	GGACU	NB	[45]
YTHDC2	Full Protein	EMSA	Ice, 30 min	ACCGGACUGUUACCAACACCCACACCCC-FAM	859.3 ± 281.2	[43]
YTHDC2	Full Protein	EMSA	Ice, 30 min	ACCGG(m ⁶ A)CUGUUACCAACACCCACACCCC-FAM	321.6 ± 61.9	[43]
YTHDC1	Domain	MST	NR	CGCGG(m ⁶ A)CTCTG (DNA)	9 ± 1	[44]
YTHDC1	Domain	MST	NR	CGCGG(m ⁶ A)CUCUG (RNA)	50 ± 10	[44]
YTHDC1	Domain	ITC	25 °C	CGCGG(m ⁶ A)CTCTG (DNA)	10 ± 1	[44]
YTHDC1	Domain	ITC	25 °C	CGCGG(m ⁶ A)CUCUG (RNA)	50 ± 10	[44]
YTHDF2	Domain	ITC	25 °C	CGCGG(m ⁶ A)CTCTG (DNA)	110 ± 10	[44]
YTHDF2	Domain	ITC	25 °C	CGCGG(m ⁶ A)CUCUG (RNA)	80 ± 20	[44]
YTHDF3	Domain	ITC	25 °C	CGCGG(m ⁶ A)CTCTG (DNA)	180 ± 20	[44]
YTHDF3	Domain	ITC	25 °C	CGCGG(m ⁶ A)CUCUG (RNA)	80 ± 10	[44]
YTHDF1	Domain	MST	25 °C, 2 hr	CGAGG(m ¹ A)GGUGUAC-fluorescein	280 ± 60	[46]
YTHDF1	Domain	MST	25 °C, 2 hr	CGAGGAGGUGUAC-fluorescein	1250 ± 10	[46]
YTHDF2	Domain	MST	25 °C, 2 hr	CGAGG(m ¹ A)GGUGUAC-fluorescein	620 ± 60	[46]
YTHDF2	Domain	MST	25 °C, 2 hr	CGAGGAGGUGUAC-fluorescein	2530 ± 20	[46]
YTHDF1	Domain	EMSA	4 °C, 1 hr	CGAGG(m ¹ A)GGUGUAC-fluorescein	130 ± 47	[46]
YTHDF1	Domain	EMSA	4 °C, 1 hr	CGAGGAGGUGUAC-fluorescein	640 ± 200	[46]
YTHDF2	Domain	EMSA	4 °C, 1 hr	CGAGG(m ¹ A)GGUGUAC-fluorescein	390 ± 30	[46]
YTHDF2	Domain	EMSA	4 °C, 1 hr	CGAGGAGGUGUAC-fluorescein	1380 ± 60	[46]
YTHDF1	Domain	EMSA	4 °C, 1 hr	CUUUU(m ¹ A)AAGUAC-fluorescein	150 ± 44	[46]
YTHDF1	Domain	EMSA	4 °C, 1 hr	CUUUUAAAGUAC-fluorescein	770 ± 260	[46]
YTHDF2	Domain	EMSA	4 °C, 1 hr	CUUUU(m ¹ A)AAGUAC-fluorescein	350 ± 48	[46]
YTHDF2	Domain	EMSA	4 °C, 1 hr	CUUUUAAAGUAC-fluorescein	≥ 1720 ± 530	[46]
YTHDC1	Domain	EMSA	4 °C, 1 hr	CGAGG(m ¹ A)GGUGUAC-fluorescein	≥ 1857 ± 350	[46]
YTHDC1	Domain	EMSA	4 °C, 1 hr	CGAGGAGGUGUAC-fluorescein	≥ 361.1 ± 31	[46]
YTHDC1	Domain	EMSA	4 °C, 1 hr	CGAGG(m ⁶ A)GGUGUAC-fluorescein	68.4 ± 33	[46]
YTHDC1	Domain	EMSA	4 °C, 1 hr	CUUUU(m ¹ A)AAGUAC-fluorescein	≥ 6185 ± 1180	[46]
YTHDC1	Domain	EMSA	4 °C, 1 hr	CUUUUAAAGUAC-fluorescein	≥ 2614 ± 2240	[46]
YTHDC1	Domain	EMSA	4 °C, 1 hr	CUUUU(m ⁶ A)AAGUAC-fluorescein	119.7 ± 10	[46]

Abbreviations: EMSA – Electrophoretic mobility shift assay; ITC – Isothermal titration calorimetry; FP – Fluorescence polarization; MST – Microscale thermophoresis; NR – Not reported in the referenced study; KD – dissociation constant.

2.2. Crystallography studies identify similarities and differences in YTH domain structure

Studies have investigated human YTH domains in complex with m6A-containing RNA (referred to as bound or “holo” structures) and in the absence of RNA (referred to as unbound or “apo” structures) (Fig. 2, Fig. 3; superposition was performed with [61]). This investigation has enabled identification of similarities and differences of the binding mechanisms across YTH domains, as well as in the framework of apo vs holo states. The YTH domains share a hydrophobic pocket, as can be seen in the unbound (apo) and bound (holo) crystal structures of YTHDF2, that contains Tyr418, Trp432, Trp486, and Trp491 (Fig. 3 A) [40], [41], [60]. Two of these tryptophan residues, Trp432 and Trp486, were shown to be important for binding to m6A, as mutation of these residues to alanine markedly reduced affinity for an m6A-containing RNA [40]. The importance of this “aromatic cage,” a term coined by Xu and colleagues [39], in binding m6A-containing RNA was additionally depicted in the later crystallized YTH domain from YTHDF2 bound to a GG(m6A)CU pentanucleotide [41]. This binding pocket was also shown in the crystal structures of human YTHDC1 (Fig. 3 B) and later in those of YTHDF1 (Fig. 3 C), YTHDF3 (Fig. 3 D), and YTHDC2 (Fig. 3 E, 3 F) in both apo and m6A-bound (holo) forms [39], [45], [56], [59]. In each case, the aromatic cage was shown to be the site of m6A recognition. This was validated for YTHDC1 and YTHDF1 through mutagenesis of the corresponding tryptophan residues and subsequent in vitro binding assays [39], [45].

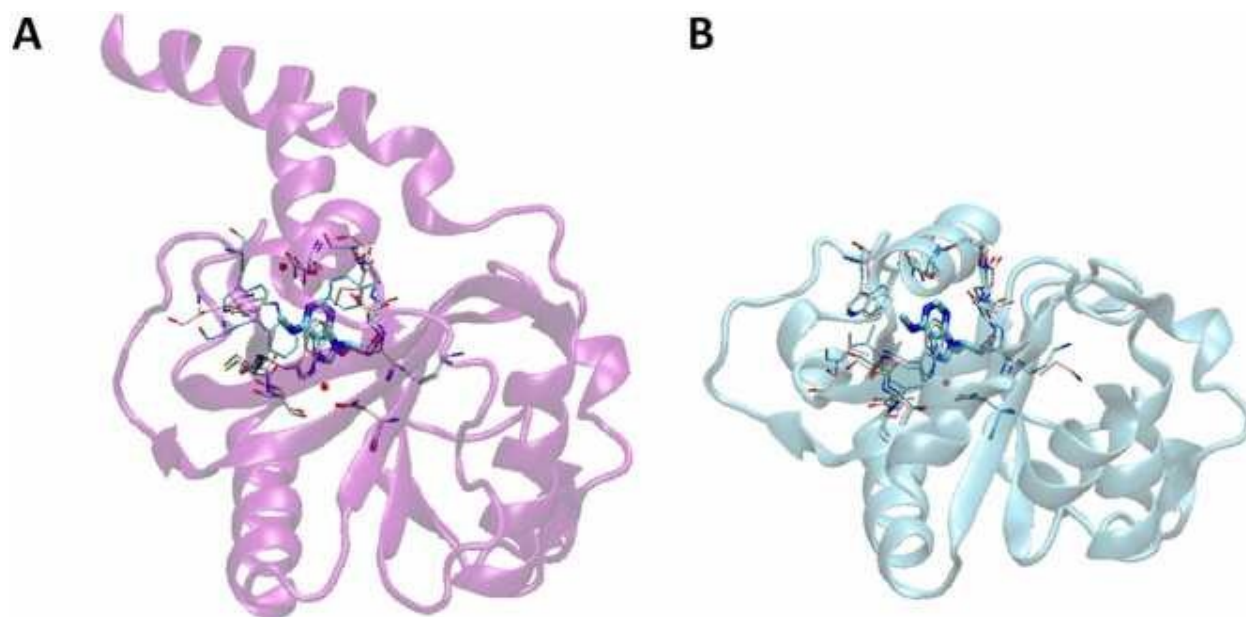


Fig. 2: Experimentally resolved structures of human YTHDF1 and YTHDC1 domains determined by crystallography. Superposition was performed using iPBA web server [61]. In the descriptions below, all apo and holo structures have key interacting residues shown with pink and cyan carbon licorice, respectively. A) YTHDF1 holo structure bound to GG(m6A)CU 5mer RNA (PDB: 4rcj,

YTHDF1 domain shown in purple cartoon) superimposed with YTHDF1 apo structure (PDB: 4rci). B) YTHDC1 holo structure bound to GG(m6A)CU 5mer RNA (PDB: 4r3i, YTHDC1 domain shown in cyan cartoon) superimposed with YTHDC1 apo structure (PDB: 4r3h).

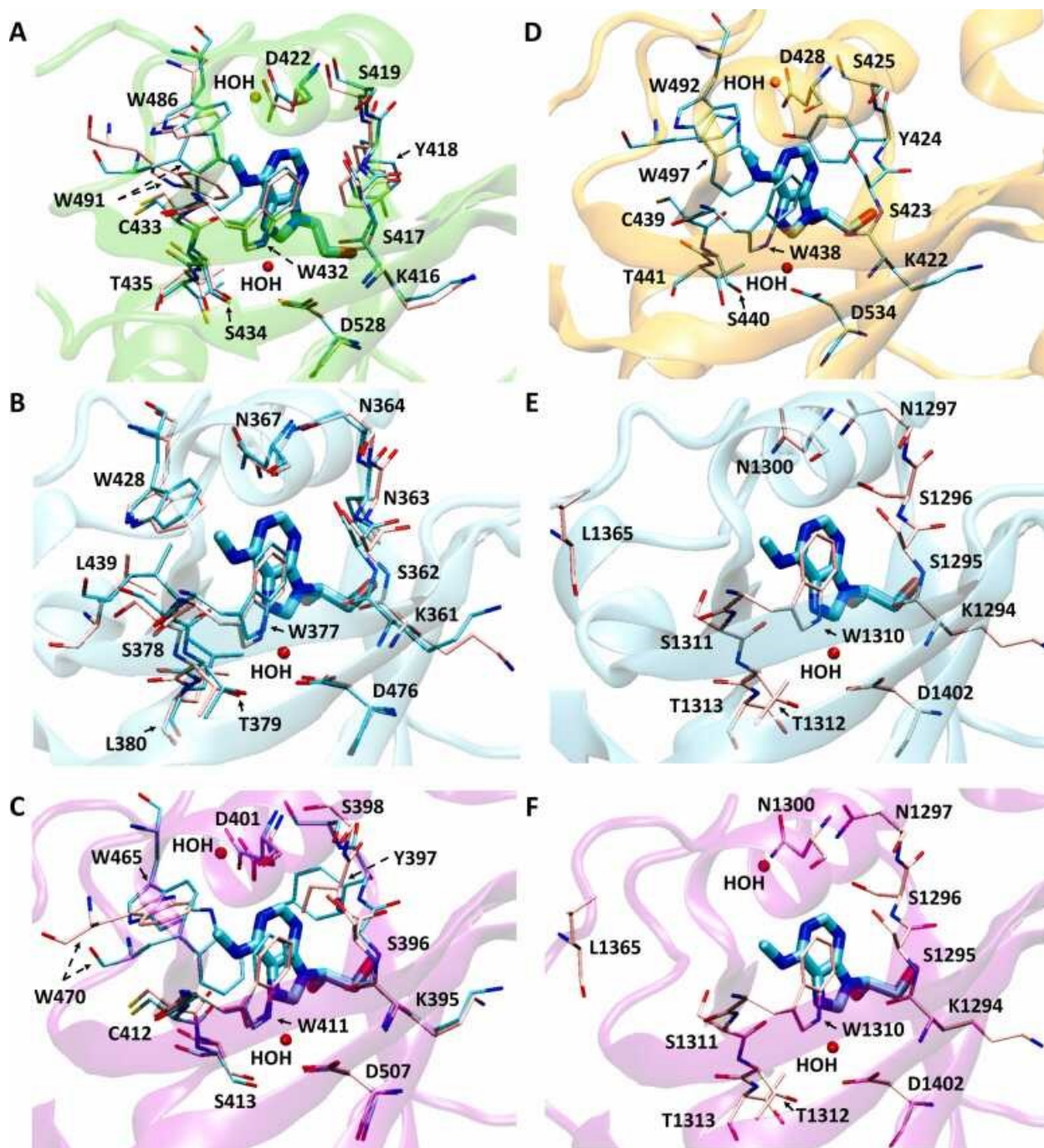


Fig. 3: Experimentally resolved structures of human YTH domains determined by crystallography. Superposition was performed using iPBA web server [61]. In the descriptions below, all apo and holo structures have key interacting residues shown with pink and cyan carbon licorice, respectively. A) YTHDF2 apo structure (PDB: 4rdo) superimposed on YTHDF2 holo structure bound to mononucleotide m6A (PDB: 4rdn, YTHDF2 domain shown with green cartoon). B)

YTHDC1 apo structure (PDB: 4r3h) superimposed on YTHDC1 holo structure bound to GG(m6A)CU 5mer RNA (PDB: 4r3i, YTHDC1 domain shown with cyan cartoon). C) YTHDF1 apo structure (PDB: 4rci) superimposed on YTHDF1 holo structure bound to GG(m6A)CU 5mer RNA (PDB: 4rcj, YTHDF1 domain shown with purple cartoon). D) YTHDF3 holo structure bound to GG(m6A)CU 5mer RNA (PDB: 6zot, YTHDF3 domain shown with orange cartoon). Apo structure not yet experimentally resolved. E) YTHDC2 apo structure (PDB: 6k6u) with m6A nucleotide and water molecule from YTHDC1 holo structure (PDB: 4r3i, m6A nucleotide shown with cyan carbon licorice, YTHDC1 domain shown with cyan cartoon). F) YTHDC2 apo structure (PDB: 6k6u) with m6A nucleotide and water molecules from YTHDF1 holo structure (PDB: 4rcj, m6A nucleotide shown with cyan carbon licorice, YTHDF1 domain shown with purple cartoon).

Although all the YTH domains from human proteins show similarity in their core aromatic binding cage for m6A recognition, the m6A-bound and unbound states show noteworthy differences. When the apo crystal structure of YTHDF2 was first resolved, the same study also resolved a holo crystal structure of YTHDF2 in complex with an m6A mononucleotide for comparison [60]. A local conformational adjustment of the loop between $\beta 4$ and $\beta 5$ (residues 476–492 in PDB: 4rdo) was observed in the presence of m6A. This loop contains Trp486 and Trp491, which accounts for two out of the four residues in the aromatic cage. Trp486 forms the “base” of the aromatic cage in both the apo and holo crystal structures of YTHDF2. The aromatic rings of Trp491, however, “flip” from the apo structure to become parallel to the aromatic rings of Trp432 in the holo structure [60]. These residues form the “walls” of the aromatic cage and the site of m6A recognition [60]. Similarly, in crystallography studies of YTHDC1, the loop between $\beta 4$ and $\beta 5$ (residues 418–440 in PDB 4r3h) was observed to be disordered and unresolved in the apo structure but resolved in the presence of m6A [39]. From the published crystal structures of YTHDC1, this sequentially homologous loop is also completely unresolved in apo structures of YTHDF1 and YTHDC2, including residues 460–469 [45] (PDB: 4rci) and residues 1354–1362 [56] (PDB: 6k6u), respectively. Overall, RNA binding induces some protein conformational stability to the loop. This trend holds for different human YTH domains [39], [41], [45], [56], [60], and certain residues share different orientations and positions in the bound versus unbound structures. These residues include Trp491 in YTHDF2, as well as Trp470 and Tyr397 in YTHDF1 (Fig. 3 A, 3 C). In YTHDF2, Tyr418 experiences a change in orientation compared to its homologues YTHDF1 and YTHDF3; this was attributed to the presence of a smaller ligand since YTHDF2 was crystallized with a m6A mononucleotide instead of a pentanucleotide (Fig. 3 A, 3 C, 3 D) [59]. Interestingly, in a more recent study [41], two orientations of Tyr418 were observed in the resolved YTHDF2 structure bound to a pentanucleotide; one of these orientations is similar to that depicted in Fig. 3 A while the other orientation is similar to that of its corresponding residues in YTHDF1 and YTHDF3 (Fig. 3 C, 3 D) [41]. Overall, the crystallography studies identified commonalities in the YTH domain binding pocket for m6A across these human proteins. Identification of the aromatic cage shed light on the mechanisms for m6A recognition, but the differences in conformational states in the unbound and bound forms of the protein, specifically in the homologous loop from YTHDC1 studies, highlighted the need for further study on the dynamics of the binding mechanisms for these proteins with m6A.

2.3. In silico investigation of the YTH domain from YTHDC1 identifies key conformational changes and solvent interactions responsible for the selectivity of m6A over adenine

To further investigate the interactions between the YTH domain and m6A, YTHDC1 was used as a model case for unbiased molecular dynamics (MD) simulations, alchemical transformations, and metadynamics [62]. Alchemical transformations to convert the N6 methyl group to an amino group were performed using the crystal structure of the YTH domain from YTHDC1 in complex with m6A to examine in detail the overall contribution of the methyl group to YTH domain binding [62]. In three steps, the partial charges were removed from the methyl group, converted to an uncharged amino group, and then recharged as an amino group to fully convert m6A to adenosine. The steps along this unphysical transformation were analyzed using thermodynamic integration (TI) [63] and showed that the methyl group represented a 16-fold difference in affinity with YTH between the m6A containing RNA and its unmodified counterpart [62]. Additionally, long unbiased MD simulations show that free GG(m6A)CU oligomer adopts a more favorable conformational state for association with the YTH domain in solvent relative to the GGACU unmodified RNA. When examining attributes of the binding pocket that could lead to m6A specificity, multiple microsecond-scale simulations of the apo YTH domain showed that conformational shifts of Met438 and Trp428 lead to metastable states that rearrange the structure of the aromatic binding pocket. In addition to the structural rearrangements, Thr379 was shown to be key for maintaining indirect interactions with m6A via a conserved water molecule. Follow-up crystallography experiments with a Thr379Val mutant to disrupt hydrogen bonding with the particular conserved water (referred to as “water 1”) resulted in a 140-fold decrease in affinity for a GG(m6A)CU RNA oligomer as measured by differential scanning fluorimetry [62]. The role of water solvation in the context of YTHDC1 was also investigated separately using unbiased simulations of the human YTH domain from YTHDC1 in the apo form and *Rattus norvegicus* YTH-G(m6A) bound form of this protein [64]. In the holo simulations, the N6 methyl group of m6A was shown to expel invading water molecules that occupied the m6A binding pocket in the apo simulations.

Water molecules involved in mediating interactions between YTH domain of YTHDC1 and m6A were resolved in the crystal structures of m6A-bound human YTH domains. In YTHDF1, a water molecule mediates interactions between the N7 of m6A and sidechains of Trp411 and Asp507 (Fig. 3 C). Similarly, a water molecule was resolved in the YTHDF2 bound structure to mediate interactions between the N7 of m6A and sidechains of Trp432 and Asp528 (Fig. 3 A) and in the YTHDF3 bound structure to mediate interactions with the sidechains of Trp438 and Asp534 (Fig. 3 D) [59]. The same water molecule described above plays a role in the binding of YTHDC1 to m6A by facilitating a hydrogen bond network between the N7 of m6A and Thr379 (Fig. 3 B) [62]. This water molecule was also resolved in the apo structures of YTHDF1, YTHDF2, and YTHDC1 and mediates interactions between Trp411 and Asp507, Trp432 and Asp528, and Tyr379 in each of these domains, respectively [59], [62].

It is also important to note that another water molecule was resolved in the crystal structures for YTHDF2 and YTHDF1 in the bound but not unbound structure (Fig. 3 A, 3 C) [40], [45]. This water molecule was also resolved in the bound structure of YTHDF3 (Fig. 3 D), although an apo structure has not yet been experimentally resolved for a direct comparison [59]. In m6A-bound YTHDF1, the aromatic cage can be further stabilized through this water mediated interaction between Trp470 and Asp401. Similarly, this water molecule mediates the interaction between Trp491 and Asp422 in m6A-bound YTHDF2, where it has been observed that Trp491 changes

orientation in the presence of m6A [40]. This water molecule was resolved in the m6A-bound YTHDF3 structure [59], mediating the interaction between Trp497 and Asp428 but is absent in the bound YTHDC1; interestingly, it was resolved in the unbound structure of YTHDC1, and was referred to as an unstable water molecule which is replaced upon RNA binding [62].

Taken together, the structural insights provided by these studies have allowed for a mechanistic understanding of m6A recognition through not only direct protein-RNA interactions, but also via protein interactions with solvent. This type of investigation of the YTH domain of YTHDC1 would be difficult with experimental techniques alone. These studies demonstrate the importance of combined computational and experimental approaches in uncovering how YTHDC1 interacts dynamically, and uniquely, with m6A-containing RNA through key structural rearrangements and the interaction with a mediating water molecule.

2.4. In silico investigation of YTH domain flexibility adds to the debate on the redundant function of the YTHDF1, YTHDF2, and YTHDF3 proteins

In addition to uncovering a mechanistic understanding of how the YTH domain from YTHDC1 recognizes m6A, in silico studies have also allowed for comparison of the binding pocket dynamics and recognition mechanisms for the other YTH domains. Although the YTH domains from human proteins have shown similar binding affinities for m6A containing RNA (Table 1), the similarity in function of these proteins in vivo is a topic of debate. While each protein possesses a YTH domain that shares high sequence and structural similarity with the other human YTH proteins, they have been suggested to actuate different functions in vivo [9]. YTHDF1 has been hypothesized to upregulate translation initiation of m6A-containing transcripts in HeLa cells [65]. This upregulation is thought to involve associations of the 40 S ribosomal subunit and translation initiation factors such as eIF3 with the N-terminal region of the protein. Similarly, YTHDF3 associates with the 40S subunit but does not directly associate with subunits of the eIF3 translation initiation complex to upregulate translation [22]. In contrast to YTHDF3 and YTHDF1, YTHDF2 has been shown to promote destabilization and degradation of RNA through direct interaction with m6A containing transcripts and recruitment of the CCR4-NOT complex via a deadenylation mechanism [66]. This recruitment has been shown to be dependent on interactions between the CNOT1 SH domain and the N-terminal region of YTHDF2, rather than the C-terminal YTH domain. Although the YTHDF1 and YTHDF3 proteins seem to perform different functions than YTHDF2, all three proteins overlap considerably in the specific transcripts they interact with [22], [67]; examination of the crystal structures of the YTH domains from human YTHDF1 [45] and YTHDF2 [60] also shows that the m6A-binding residues and adjacent residues are conserved in the sequence of the YTH domain of YTHDF3, suggesting that these proteins may function redundantly to regulate the fate of m6A containing transcripts [67]. To further investigate the possible similarities in the conformational dynamics of these proteins, the unbound structures of the YTH domains of YTHDF1, YTHDF2, and YTHDF3 were subjected to 5 μ s of simulation to compare the flexibility in the domains, particularly in the “recognition loop” containing two of the tryptophan residues responsible for forming the aromatic cage that recognizes m6A [59]. Root mean square fluctuation (RMSF) of regions within this recognition loop were found to be correlated across all three protein domains, with differences between conformational motion of the proteins found far from the binding pocket, suggesting similar structural dynamics of the

domains in the context of m6A recognition [59]. This observation, along with the similarity in the bound structure of YTHDF3 to YTHDF2 [60] and YTHDF1 [45] described in Section 2.2, lends support to the hypothesis that these proteins recognize their m6A substrates in a similar manner. This study [59] investigated these proteins only in the context of their YTH domains, in absence of their N-terminal disordered regions. Further investigation is needed to place this data in context with the dissimilar associations of other protein factors with the N-terminal regions of the human YTH protein family [22], [65], [66], [68].

The ongoing investigation of the five YTH human proteins, experimentally and computationally, has led to a wealth of understanding around their binding properties. Specifically, *in silico* studies of YTHDC1 [62], [64] as well as YTHDF1, YTHDF2, and YTHDF3 [59] have provided significant insights as to how these m6A readers mechanistically recognize their modified substrate. Notably, future work is needed to understand in more detail the recognition properties of bound YTHDC2 in comparison to other YTH domains, and to uncover how these mechanisms affect the functions of the human YTH family of proteins overall.

2.5. Identification of small molecule inhibitors for the YTH domain family of proteins elucidates the druggability of epitranscriptome binding proteins

The regulation of m6A-containing transcripts by the YTH domain family of proteins has been implicated in a large variety of disorders (reviewed in [10], [15]) Due to their implications in human disease, the inhibition of these proteins with their cognate modified RNA substrates has been selected as a druggable target [41], [69], [70], [71]. YTHDC1 was initially selected as a candidate for small molecule drug discovery [69]. Through a fragment-based drug design methodology (reviewed in [72]), 30 small molecule fragments that consisted of m6A nucleobase analogs, uracil scaffold molecules, and other bicyclic compounds were identified to interact with the YTH domain of YTHDC1 [69]. These fragments were identified from computational docking of libraries of small molecules and further validated for their binding affinity using homogeneous time-resolved fluorescence (HTRF) [73] and isothermal titration calorimetry. Additionally, crystal structures of the protein-fragment complexes were generated to examine the binding modes of the small molecules. Of these, four fragments were found to show binding affinities below 1 mM and ligand efficiencies ranging from 0.25 to 0.4 kcal mol⁻¹ nHA-1 [69]. Importantly, all of these fragments form interactions with the tryptophan residues that make up the aromatic cage (Trp377 and Trp428) and with Ser378, suggesting a link between interaction with these regions and metrics amenable to future drug development. In addition to these fragments, 25 small molecules were later identified through a similar computational and experimental pipeline [70]. These molecules included m6A base analogs, molecules containing an N-methyl amide that interacts with the polar residues Asn367 and Ser378 in the binding pocket, molecules containing a morpholine group that disrupts the recognition loop, and uracil derivatives that displace the structural water found in the binding pocket. Of these four classes of small molecules, the m6A analog with an N-methyl amide (referred to in the study as compound 6) was identified as a promising candidate for further design due to its IC₅₀ of 39 μM and its ligand efficiency of 0.6 kcal mol⁻¹ nHA-1. Additionally, this finding further underlines the importance of interactions with polar residues in the binding pocket of YTHDC1 and motivates additional development of a small molecule inhibitor for YTHDC1.

The YTHDF2 and YTHDF1 proteins have also been selected as targets for small molecule drug design. In addition to crystallizing the first structure of the YTH domain of YTHDF2 in complex with a pentanucleotide (PDB: 7z26) [41], Nai and colleagues identified 17 fragments from a combination of the previous YTHDC1 studies [70] and structure-based design that inhibit m6A-YTHDF2 binding activity. It is important to note that due to the structural and sequence similarity of YTHDF2 to YTHDF1 and YTHDF3, these small molecules might act as general inhibitors for all of these proteins [41]. The small molecules tested include m6A nucleobase and uracil analogs; as well as pyrazolopyrimidine, triazine, and pyrimidine derivatives. These molecules were analyzed for their inhibitory effects with HTRF, and their interactions with the YTH domain from YTHDF2 were determined using X-ray crystallography. One of these molecules, 6-cyclopropyluracil (referred to as compound 11) represents a promising candidate for future drug development with an IC₅₀ of 174 μ M and a ligand efficiency of 0.47 kcal mol⁻¹ nHA-1, owing its potency to the interaction of the cyclopropyl group with the aromatic cage of YTHDF2 [41]. Furthermore, this study was the first of its kind to identify small molecule scaffolds for further drug discovery in the context of this YTH domain. In addition to the compounds identified as potential inhibitors of YTHDF2, the small molecule ebselen has also been proposed as a small molecule inhibitor of YTHDF1 [71]. Through a high-throughput tryptophan fluorescence quenching assay, ebselen was shown to directly inhibit binding of RNA both through in vitro binding assays and through immunoprecipitations from PC-3 prostate cancer cells treated with a non-lethal concentration of ebselen for 24 h. Ebselen was found to bind covalently with Cys412 through selenium sulfide bonds or reversibly with the m6A binding pocket depending on the reducing or oxidizing nature of the binding pocket environment as determined by X-ray crystallography [71]. Furthermore, the ebselen scaffold was used to design additional compounds with similar inhibitory characteristics, further demonstrating the use of the compound for further design efforts. Overall, the budding space of small molecule drug design for YTH domains offers exciting opportunities to probe the interaction dynamics of these proteins, both in the context of basic understanding of epitranscriptome regulation and in the context of human disease.

3. Characterization of epitranscriptome reader proteins beyond the YTH-m6A paradigm

3.1. The binding mechanisms for non-YTH m6A readers remain elusive

RNA chromatography studies that identify m6A-interacting proteins have also uncovered readers outside of the YTH family such as the insulin-like growth factor 2 binding proteins (IGFBPs) [17] and the heterogeneous nuclear ribonucleoproteins (HNRNPs), such as HNRNPC [5] and HNRNPA2B1 [7], that bind selectively to m6A-containing transcripts. The IGFBP proteins (IGFBP1/IGFBP2/IGFBP3) have been found to bind m6A-containing transcripts with a 3-to-4-fold higher affinity relative to unmethylated transcripts [17]. Additionally, the RNA binding sites of these proteins overlap with sites of m6A methylation in both single-stranded and structured, hairpin RNA. These proteins contain two RNA recognition motif (RRM) and four K-homology (KH) RNA binding domains. The KH3 and KH4 domains of these proteins were shown to be key for binding a single-stranded m6A oligomer through RNA pulldown experiments followed by Western blotting to visualize the protein-RNA complex [17]. The IGFBPs represent a departure from the canonical m6A binding pocket found in the YTH domain; however, the specifics of the structural similarities (or differences) between the m6A binding pockets of these proteins remains unclear. Furthermore, these proteins are currently being investigated as druggable targets in the context of colorectal cancer proliferation [74], and additional work on their interactions with m6A could elucidate the impact of a small molecule inhibitor on the IGFBP-m6A interaction.

In addition to readers like the YTH proteins and the IGFBPs that bind directly to the modified m6A base, a subclass of proteins recognizes m6A in a more indirect, structurally dependent manner [75]. HNRNPC is one such m6A reader that is thought to bind to methylated transcripts via an indirect mechanism involving RNA structure [5]. One of the common secondary structures of RNA is known as a hairpin or stem loop, in which a single stranded region (or multiple regions) is flanked by double-stranded RNA (dsRNA) regions. This type of RNA structure can be impacted by the presence of an m6A modification, creating a structure distinct from its unmodified counterpart, which is preferably recognized for binding by the HNRNPC reader protein. One such example of this “m6A switch” behavior within the local RNA structure is found in the human metastasis-associated lung adenocarcinoma transcript (MALAT1) [5]; here, a portion of MALAT1 forms a 30-nucleotide stem loop containing the GGACU m6A methylation motif in a dsRNA region. Upon methylation of the adenine in this motif, the base pairing of adenine to uracil in the dsRNA region within the hairpin loop is disrupted, leading to a partial opening of the stem loop. The now single-stranded region of the partially opened stem loop previously bound to the GGACU motif becomes accessible to proteins like HNRNPC, which has been well characterized for its binding to MALAT1 in an m6A-dependent manner [5], [18], [76]. The recognition of MALAT1 by HNRNPC was shown to be structurally dependent through a GG(A→U)CU mutation in the MALAT1 stem loop, which mimics the base pairing disruption associated with an m6A methylation event [76]. Overall, from the aforementioned studies, HNRNPC represents a unique case of m6A “reading” as it appears not to recognize the m6A modification through direct binding, but rather recognizes unmodified regions of RNA that are made accessible for binding through structural rearrangements dependent on the presence of m6A. While other HNRNPs have been identified as m6A interacting proteins [37], many have not been further characterized as specific readers. One of these proteins, HNRNPA2B1, which has shown direct binding to transcripts containing m6A [7]; yet in another study, HNRNPA2B1 was shown to exhibit a 1.6, 1.7 and 11.5-fold decrease in affinity for m6A modified RNA when compared with relative to unmodified for a 5-mer- 8-mer

and 10-mer, respectively [77]. All in all, the mechanisms of HNRNPA2B1 binding to m6A is still an eluding problem. Thus, additional studies are needed to explore the nature of the molecular recognition mechanism.

Overall, non-YTH readers for m6A have only been recently investigated in terms of their mechanisms of binding to methylated transcripts. Excitingly, however, these investigations have not only identified a possibly novel mechanism for direct recognition of m6A, but also a potentially indirect, structurally dependent mechanism that might extend to other m6A reader proteins (besides HNRNPC). Further work is needed to characterize the dynamics of these binding events as well as to uncover alternative binding surfaces for m6A beyond the YTH aromatic cage.

3.2. Identification of proteins interacting with other modifications beyond m6A leads to further avenues of investigation for epitranscriptome recognition

In addition to the work that has been performed to identify interacting proteins with m6A modified RNAs, similar mass spectrometry techniques have been used to identify proteins interacting with m1A [28], m5C [29], and 8-oxoG [30] (Table 2, [5], [7], [11], [17], [21], [28], [29], [30], [37], [39], [40], [45], [56], [58], [59], [62], [76], [77], [78], [79], [80], [81], [82], [83]). Stable isotope labeling by amino acid in cell culture (SILAC) [84] has been used to identify proteins interacting with m1A with a 34-mer RNA probe designed to carry a portion of the SOX18 gene known to be modified in vivo [28], [36], [85]. From subsequent liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis, the YTH domain family proteins YTHDF1, YTHDF2, and YTHDF3 as well as other proteins such as the heterogeneous nuclear ribonucleoprotein hnRNPd and the TAR DNA-binding protein were identified as putative m1A readers. Follow-up in vitro binding characterization was conducted with the YTH domain proteins via EMSA [28]. These assays showed that the YTHDF1, YTHDF2 and YTHDF3 proteins, as well as the YTH domain from YTHDC1, bound to the same SOX18 RNA oligomer containing m1A at a lower affinity relative to the same oligomer containing m6A. Despite this lower affinity, these proteins were selective for both m1A and m6A relative to an unmodified RNA oligomer containing adenine (Table 1). To determine if the binding mechanism of the YTHDF2 with m1A modified RNA is like the binding mechanism of these proteins with m6A modified RNA, mutagenesis was performed for the YTHDF2 protein followed by EMSAs to determine relative changes in affinity for the m1A-containing RNA substrate [28]. Specifically, mutation of one of the key tryptophan residues that make up the “aromatic cage” responsible for m6A modified RNA recognition (Trp432 in YTHDF2) led to an abrogation of m1A binding activity, suggesting that m1A modified RNA may be recognized through a similar mechanism to m6A modified RNA; however, the other proteins identified in the m1A interactome from this study were not investigated further. The large difference in affinity of these YTH proteins for m1A and m6A modified RNA was later investigated using follow-up EMSAs and microscale thermophoresis (MST) experiments [46]. This study observed sub-micromolar KD values with m1A containing oligomers for both YTHDF1 and YTHDF2, as well as a 3-to-5-fold decrease in affinity for an equivalent unmodified oligomer (Table 1). Interestingly, this study also contradicts the initial claim that YTHDC1 recognizes m1A, showing no selectivity for m1A over unmodified RNA, suggesting that “m1A recognition is specific to YTHDF1/2” [46]. YTHDF2 also appears in the list of m5C associated proteins, along with the cleavage stimulation factor proteins CSTF1, CSTF2, and CSTF3 (Table 2). While these proteins were investigated for their direct binding to

m5C, only YTHDF2 was selected for quantitative binding characterization via EMSA. This protein shows a higher binding affinity to m5C than unmodified RNA and shows binding dependence on the same Trp432 within its aromatic cage found to be relevant for m1A and m6A binding [29]. These results have suggested that the recognition capabilities of at least some of the YTH domain proteins could extend beyond m6A but also highlight the unexplored mechanisms of RNA modification recognition. Finally, 8-oxoG reader proteins such as HNRNPD, PCBP1, YB-1, and HNRNPC have been investigated in vitro to varying degrees with respect to their direct binding to modified RNA [30], [80], [81], [82]. HNRNPD has been shown to bind to 8-oxoG containing transcripts with high affinity through both RNA pulldown experiments and subsequent Western blotting [30], [81]. Interestingly, PCBP1 has also shown specific binding 8-oxoG over unmodified RNA but shows preference for two of these modified bases spaced 6 nucleotides apart rather than a single modification as shown for other proteins [82], showcasing the diversity in binding behavior of the studied 8-oxoG readers. The YB-1 protein has been shown to readily form complexes with RNA containing 8-oxoG [80], and the central protein region—along with the C-terminus—is required for modification binding. Interestingly, YB-1 has been shown to associate with the IGF2BPs to regulate MYC and BCL2 RNA transcripts in an m6A dependent manner [86] which could suggest interplay between the two modified RNA pools. However, the binding dynamics for YBX1 and 8-oxoG are not well understood, and more work is needed to understand the 8-oxoG recognition mechanism by these proteins both in silico and in vitro. No such investigation has been performed in silico, despite the availability of numerous crystal structures in complex with an RNA strand (Table 2). Similarly, although a crystal structure for HNRNPC has been resolved, its RRM binding domain has yet to be studied mechanistically for 8-oxoG recognition using in vitro or in silico techniques. In summary, the discovery of YTH proteins and others that recognize modifications beyond the well-studied m6A represent exciting opportunities to not only identify how reader proteins might flexibly recognize the epitranscriptome, but also to identify other characteristic mechanisms of chemically modified RNA recognition by reader proteins.

Table 2

Epitranscriptome-associated proteins with confirmed direct binding to modified RNA transcripts.

Gene Name	Associated Modification	Domains	Discovery method	Direct transcript binding shown?	Notes	Ref.	Relevant crystal structures (PDB)
HNRNPA2B1	N6-Methyladenosine	RRM_1 (1), RRM_6 (1)	RNA affinity pulldown, LC-MS/MS	HITS-CLIP, UV-CLIP, RNA protection assay, RNA pulldown, immunoblotting	Proposed to act in an "m ⁶ A switch" dependent mechanism, rather than direct binding[64]	[7], [37], [77]	5HO4
HNRNPC	N6-Methyladenosine	RRM_1 (1)	RNA affinity pulldown, LC-MS/MS	Filter-binding assay	associates with m ⁶ A-switch constructs by binding the U-tract formed in methylated hairpins	[5]	2MZ1

Gene Name	Associated Modification	Domains	Discovery method	Direct transcript binding shown?	Notes	Ref.	Relevant crystal structures (PDB)
IGF2BP1	N6-Methyladenosine	RRM_1 (1), KH_1 (4), RRM_6 (1)	LC-MS/MS and computational prediction of binding proteins	and m ⁶ A EMSA	Direct binding demonstrated <i>in vitro</i> and <i>in vivo</i> , KH3/4 indispensable in binding to m ⁶ A	[17]	6QEY, 2N8L
IGF2BP2	N6-Methyladenosine	RRM_1 (2), KH_1 (4)	LC-MS/MS and computational prediction of binding proteins	and m ⁶ A EMSA	Direct binding demonstrated <i>in vitro</i> and <i>in vivo</i> , KH3/4 indispensable in binding to m ⁶ A	[17]	6ROL
IGF2BP3	N6-Methyladenosine	RRM_1 (1), KH_1 (4), RRM_6 (1)	LC-MS/MS and computational prediction of binding proteins	and m ⁶ A EMSA	Direct binding demonstrated <i>in vitro</i> and <i>in vivo</i> , KH3/4 indispensable in binding to m ⁶ A	[11], [17]	6FQR
Prrc2a	N6-Methyladenosine	BAT (1)	RNA pulldown, LC-MS	affinity EMSA	direct binding demonstrated <i>in vitro</i> , associated with oligodendroglia proliferation	[79]	
RBMX (aka HNRNPG)	N6-Methyladenosine	RBM1CTR (1), RRM_1 (1)	RNA pulldown, LC-MS/MS	affinity EMSA	binds m ⁶ A through C-terminal low-complexity region	[76]	2MB0
YTHDC1	N6-Methyladenosine	YTH (1)	RNA pulldown, LC-MS/MS	affinity ITC	binds m ⁶ A using an "aromatic cage"	[37], [39], [45], [62]	6ZCN
YTHDC2	N6-Methyladenosine	OB_NTP_bind (1), Ank_2 (1), HA2 (1), Helicase_C (1), YTH (1), R3H (1), DEAD (1)	SILAC-based RNA pulldown, LC-MS/MS	RNA ITC	shown to have a conserved m ⁶ A binding pocket and shares similarities to other YTH domains	[11], [45], [56]	6K6U
YTHDF1	N6-Methyladenosine	YTH (1)	RNA pulldown, LC-MS/MS	affinity EMSA, ITC	binds directly to m ⁶ A, forms conserved aromatic cage to recognize the modification	[21], [45]	4RCJ
YTHDF2	N6-Methyladenosine	YTH (1)	RNA pulldown, LC-MS/MS	affinity FP, EMSA, ITC	YTH domain binds directly to m ⁶ A and shares similarity in structure to YTHDC1; basic residues near the binding cage of YTH domain	[37], [40], [45]	4RDN
YTHDF3	N6-Methyladenosine	YTH (1)	RNA pulldown, LC-MS/MS	affinity EMSA	YTH domain binds directly to m ⁶ A. YTH domain selects for m ⁶ A containing RNA via loop-loop interactions, conformation	[37], [58], [59]	6ZOT

Gene Name	Associated Modification	Domains	Discovery method	Direct transcript binding shown?	Notes	Ref.	Relevant crystal structures (PDB)
ALYREF	N5-Methylcytosine	FoP_duplication (1), RRM_1 (1), FYTT (1)	RNA Immunoprecipitation MS/MS	EMSA	selectivity, and induced fit effects. Lys171 key for recognition of m ⁵ C (found by comparing MBD and YTH domain sequences and performing point mutations)	[78]	1NO8
CSTF1	N5-Methylcytosine		SILAC-based RNA pulldown, LC-MS/MS	RNA pulldowns, Western blotting		[29]	
CSTF2	N5-Methylcytosine	RRM (1)	SILAC-based RNA pulldown, LC-MS/MS	RNA pulldowns, Western blotting		[29]	
CSTF3	N5-Methylcytosine		SILAC-based RNA pulldown, LC-MS/MS	RNA pulldowns, Western blotting		[29]	
YTHDF1	N5-Methylcytosine	YTH (1)	SILAC-based RNA pulldown, LC-MS/MS	RNA pulldowns, Western blotting		[29]	
YTHDF2	N5-Methylcytosine	YTH (1)	SILAC-based RNA pulldown, LC-MS/MS	RNA pulldowns, Western blotting, EMSA	Trp432Ala mutation leads to loss of m ⁶ A affinity	[29]	
YTHDF3	N5-Methylcytosine	YTH (1)	SILAC-based RNA pulldown, LC-MS/MS	RNA pulldowns, Western blotting		[29]	
YTHDF1	N1-Methyladenosine	YTH (1)	RNA affinity pulldown, LC-MS/MS	EMSA		[28]	4RCJ
YTHDF2	N1-Methyladenosine	YTH (1)	RNA affinity pulldown, LC-MS/MS	EMSA		[28]	4RDN
YTHDF3	N1-Methyladenosine	YTH (1)	RNA affinity pulldown, LC-MS/MS	EMSA		[28]	6ZOT
HNRNPC	8-oxo-7, 8-dihydroguanosine	RRM_1 (1)	RNA affinity pulldown, LC-MS/MS	RNA pulldowns, Western blotting, competition experiment		[30]	2MZ1
HNRNPD (AUF1)	8-oxo-7, 8-dihydroguanosine	RRM_1 (2), CBFNT (1)	RNA affinity pulldown, LC-MS/MS	RNA pulldowns, Western blotting		[30], [81]	5IM0, 1X0F
PCBP1	8-oxo-7, 8-dihydroguanosine	KH_1 (3)	RNA affinity chromatography coupled with mass spectrometry	RNA pulldowns, Western blotting	Binds to 2 8-oxoG residues with higher preference than a single mark (positions 9 and 15 in a 30mer RNA oligomer)	[82]	1ZTG, 1WVN
PNPT1	8-oxo-7, 8-dihydroguanosine	S1 (1), PNPase (1), RNase_PH (2), RNase_PH_C (2), KH_1 (1)	RNA protection assays followed by SDS-PAGE analysis	EMSA		[83]	4AM3

Gene Name	Associated Modification	Domains	Discovery method	Direct transcript binding shown?	Notes	Ref.	Relevant crystal structures (PDB)
YBX1	8-oxo-7, 8-dihydroguanosine	S1 (1)	RNase A protection assay, gel shift assay (EMSA)	EMSA	Central and C-terminal protein regions required for 8-oxoG binding activity	[80]	5YTX, 5YTY, 5YTV, 5YTS

Abbreviations: Ref. – reference; LC-MS/MS – Liquid chromatography coupled with mass spectrometry; SDS-PAGE – Sodium dodecyl sulfate (denaturing) polyacrylamide gel electrophoresis; HITS-CLIP - High-throughput sequencing of RNA isolated by crosslinking immunoprecipitation; UV-CLIP – UV cross-linking and immunoprecipitation; EMSA – Electrophoretic mobility shift assay; ITC – Isothermal titration calorimetry; FP – Fluorescence polarization; MST – Microscale thermophoresis.

4. Computational advancements accelerating the study of epitranscriptome reader proteins in silico

Computational advancements have played a key role in the in silico investigation of protein-RNA interactions, including interactions with RNA modifications. In the context of computational studies, MD simulations can be considered a potent tool to study the structure and dynamics, and provide biophysical insights for such interactions starting from an initial protein-RNA structural conformation. The development of molecular mechanics force-fields, programs as well as platforms to build, simulate and analyze such systems has significantly enabled the computational study of RNA-protein interactions using MD simulations [87], [88], [89], [90], [91], [92], [93], [94]. If the structure under investigation has not been experimentally resolved, computational methods can also be used for its initial modeling. Computational methods, when combined with experimental techniques as shown above, represent powerful synergistic approaches for the identification and biophysical characterization of novel protein interactions with modified RNAs.

4.1. Computational methods for protein structure prediction

Historically, the investigation of protein-RNA interactions using MD simulations has been limited partly due to the lack of crystallography data for the protein of interest. The advent of structural modeling tools has allowed the generation of protein structures from a primary amino acid sequence, which can serve as a starting point for further modeling of protein-RNA interactions. One such method of structure prediction involves homology modeling and is leveraged by tools such as I-TASSER [95], Phyre2 [96], HHPred [97] and Modeller [98]. A homology modeling approach builds a protein structure based on fragments of experimentally resolved structures from the Protein Data Bank [99] that share homology with the input primary sequence [100]. These fragment “templates” are then threaded together to generate the structural model of the input protein, which is checked and further refined using a variety of energetic analyses and iteration of the structure building process [95]. For example, I-TASSER and Phyre2 have been used to identify biologically relevant structural and functional features of proteins. I-TASSER was used to predict the structure of caveolin-1 (cav-1), a membrane-associated protein, with HNRNPA2B1, an RNA binding protein responsible for binding microRNA (miRNAs) such as mi-R17/93 present in the resulting microvesicles, which package these miRNA species to control gene expression during periods of oxidative stress [101]. The structural model of the cav-1/HNRNPA2B1 complex correctly predicted the caveolin scaffolding domain (CSD) of cav-1 and the arginine-glycine-glycine (RGG) repeat box of HNRNPA2B1 to be key for complex formation. This predicted protein-protein interaction was confirmed in vitro through inhibition of HNRNPA2B1/cav-1 binding by competition with a CSD peptide and through RGG deletions from HNRNPA2B1 in immunoprecipitation experiments [101]. Another homology modeling tool, Phyre2, was used predict the structure of CcaF1, a previously uncharacterized protein in the archaea *Rhodobacter sphaeroides* [102]. The predicted structure of this protein showed homology to the RNA binding domain in the Smaug protein from *D. melanogaster*, suggesting that CcaF1 might be responsible for binding and regulation of RNA. This prediction was confirmed, as CcaF1 was shown to bind the small RNA CcsR1 in vitro and regulate its stability in vivo [102].

Most recently, the field of protein structure prediction was revolutionized by the deep-learning neural network-based method AlphaFold, which improves upon previous homology-based

approaches (reviewed by [103]). It is worth noting that in the 2020 Critical Assessment of Structure Prediction (CASP14), AlphaFold demonstrated accuracy on par with experimentally resolved structures in a majority of cases and significantly outperformed other computational methods [104]. Additionally, AlphaFold demonstrated the capacity to predict the structure of many difficult protein targets at or near experimental resolution [105]. AlphaFold's success could be attributed to certain key factors, including its methodology and the fact that the single domain protein structure library is basically complete [106]. The advancement of protein structure prediction methods is a key contributing factor to the study of protein interactions with molecules such as RNA, DNA, other proteins, and small molecules. Given that AlphaFold was applied to a wide range of proteins, a significant number of RNA binding protein structures that have not been resolved by experimental approaches have been predicted by AlphaFold [107]. As a result, AlphaFold could largely increase the overall capacity to study protein-RNA recognition, including protein interactions with modified RNAs, due to its ability to predict the overall protein structures with high-accuracy. While AlphaFold can predict a protein structure with accurate conformational packing of the backbone and side chains, the modeled structure could correspond to a particular conformation, as is the case for proteins that show different conformations when in active or inactive states [106]. A protein can exist in one conformation when it is in its bound state, and in another conformation when it is in its unbound state. Therefore, a careful consideration and inspection of AlphaFold predicted models over different known states could be worthy of investigation. In this context, we took the initiative to compare AlphaFold-predicted models provided within the AlphaFold Protein Structure Database [104], [108] for the case of YTH human domains with experimentally resolved bound and unbound structures (Fig. 4). The AlphaFold models predict backbone conformations accompanied by sidechain orientations that more closely resemble those of the experimental structures in the bound state for YTHDF2, YTHDC1, and YTHDF1 (Fig. 4 A, 4 B, B,44 C, [39], [45], [59], [60]). Notably, one particular residue in the AlphaFold model of YTHDF1, Tyr397, has a different orientation than its holo structure (Fig. 4 C). This orientation is reminiscent of its corresponding residue Tyr418 in YTHDF2 holo structure in complex with a mononucleotide (PDB: 4rdn; Fig. 2 A) [60], and one of the two conformations in a more recently resolved YTHDF2 holo structure in complex with a pentanucleotide (PDB: 7z26) [41]. For the AlphaFold model of YTHDF3, which does not have an experimentally resolved apo structure for comparison, similarities are observed in the conformation of most of the aromatic side chains to the bound experimental structure (Fig. 4 D). Similar to the AlphaFold model of YTHDF1, Tyr424 in YTHDF3 adopts an orientation mimicking that of Tyr418 in the same relative position for experimentally resolved structures of YTHDF2 (Fig. 4 A, 4 D). Importantly, in the case of YTHDC2, for which a bound state is not experimentally available, the sidechain positions (especially of YTHDC2 residue Leu1365) of the AlphaFold model shows close resemblance to both bound YTHDF1 and YTHDC1 (Figs. 4 E, 4 F). AlphaFold models predict the YTH domains closer to the bound rather than unbound state, which can possibly be attributed to the fact that unbound YTH domains contain partially unresolved residue moieties; this could potentially be related to the fact that binding sites (such as the YTH RNA binding domain), are in general the most accurately predicted regions of a protein's conformation [106].

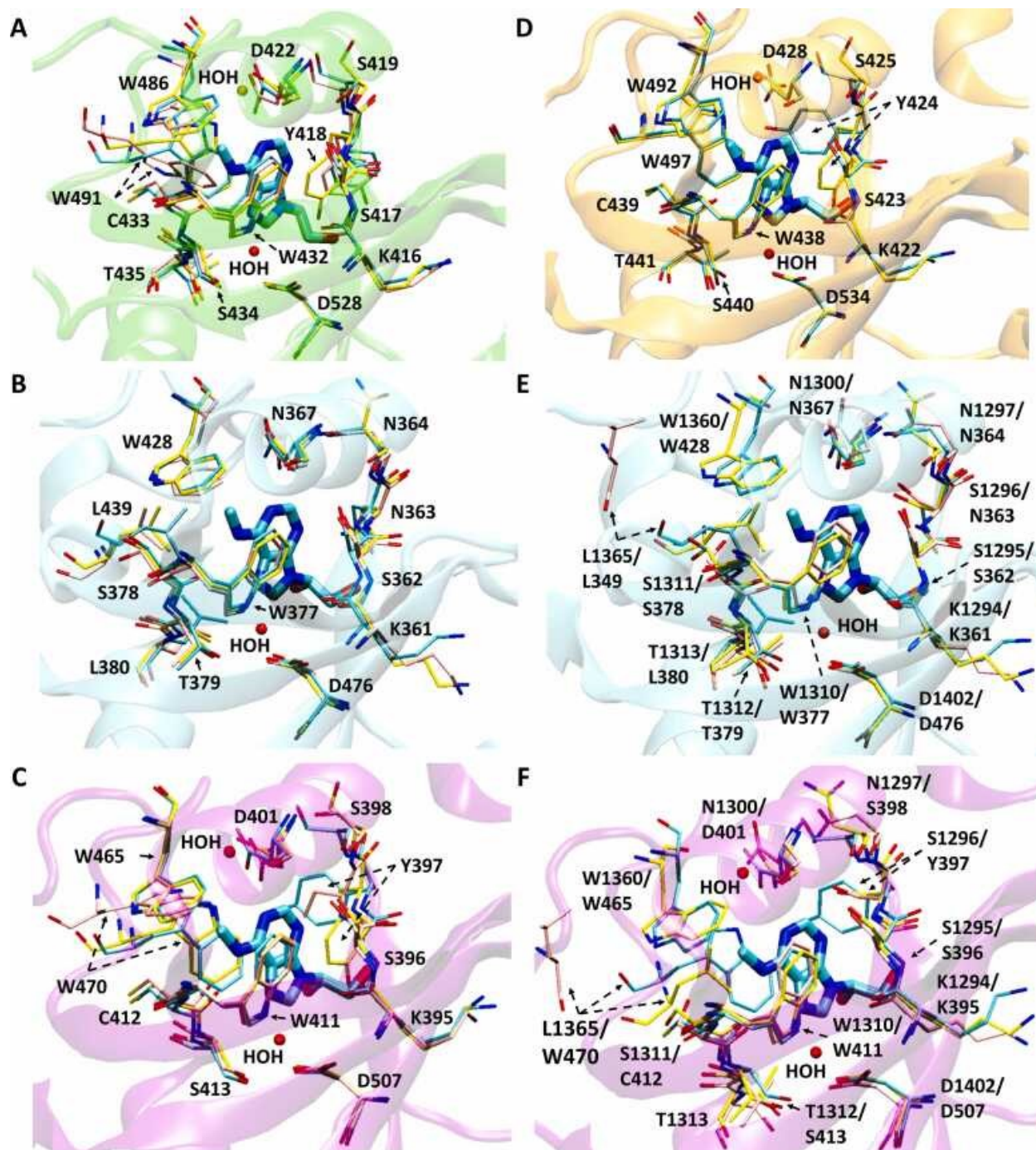


Fig. 4: Experimentally resolved structures of human YTH domains determined by crystallography and their corresponding AlphaFold models. Superposition was performed using iPBA web server [61]. In the descriptions below, all apo, holo, and AlphaFold structures have key interacting residues shown with pink, cyan, and yellow carbon licorice, respectively. A) AlphaFold model of YTHDF2 (UniProt Q9Y5A9) superimposed on YTHDF2 apo structure (PDB: 4rdo) and YTHDF2 holo structure bound to mononucleotide m6A (PDB: 4rdn, YTHDF2 domain shown with green cartoon). B) AlphaFold model of YTHDC1 (UniProt Q96MU7) superimposed on YTHDC1 apo structure (PDB: 4r3h) and YTHDC1 holo structure bound to GG(m6A)CU 5mer RNA (PDB: 4r3i, YTHDC1 domain shown with cyan cartoon). C) AlphaFold model of YTHDF1 (UniProt Q9BYJ9)

superimposed on YTHDF1 apo structure (PDB: 4rci) and YTHDF1 holo structure bound to GG(m6A)CU 5mer RNA (PDB: 4rcj, YTHDF1 domain shown with purple cartoon). D) AlphaFold model of YTHDF3 (UniProt Q7Z739) superimposed on YTHDF3 holo structure bound to GG(m6A)CU 5mer RNA (PDB: 6zot, YTHDF3 domain shown with orange cartoon). Apo structure not yet experimentally resolved. E) AlphaFold model of YTHDC2 (UniProt Q9H6S0) superimposed on YTHDC2 apo structure (PDB: 6k6u) and YTHDC1 holo structure bound to GG(m6A)CU 5mer RNA (PDB: 4r3i, YTHDC1 domain shown with cyan cartoon). F) AlphaFold model of YTHDC2 (UniProt Q9H6S0) superimposed on YTHDC2 apo structure (PDB: 6k6u) and YTHDF1 holo structure bound to GG(m6A)CU 5mer RNA (PDB: 4rcj, YTHDF1 domain shown with purple cartoon).

4.2. Methods for generating a biomolecular interaction model

An “appropriate” protein structure (i.e., either experimentally resolved in the bound state or a computationally predicted structure) can serve as a stepping stone to study and identify the protein-RNA interface, and subsequently build an initial interaction model between protein and RNA [107]. This model can be generated using a variety of methods—analogue to methods developed from protein-protein interactions—including rigid-body docking [109], template-based docking [107], [109], [110], and other machine-learning based methods [110], [111], [112]. Rigid body docking, which searches and superimposes static input structures based on favorable energetics of the resulting complex, is useful for determining an initial interaction model for a protein-RNA complex; however, the highly flexible and dynamic nature of RNA molecules may lead to biased docking based on the input conformation [109]. In template-based (or comparative) docking, the structural similarity between the complex to be modeled and an experimentally resolved complex is assessed. This modeled complex can be constructed by superposition (e.g., of the monomer models with respect to the experimentally resolved complex) and then evaluated through scoring functions measuring structural similarities between the monomer models, as well as the complex template components [109], [110]. In such comparative docking approaches, the choice of experimental template by a local alignment to the complex interfaces over alignment to the entire complex can slightly improve the quality of the modeled complex, as indicated for protein-protein interactions involving a binding-induced conformational change [113]. This approach could also hold when modeling interactions between proteins with nucleic acids as well, given their dynamic nature. However, a caveat to such alignment methods comes from what is known as the “twilight zone” (roughly 25%) of overall sequence similarity found when studying protein-protein interactions [109], [114]. Nevertheless, when similarity between the template and modeled complexes is reasonably acceptable, template-based docking with critical evaluation of the interfacial residue interactions could be considered a valuable tool to initially model protein interactions with other biomolecules, such as RNA. Therefore, evaluating or scoring these interfacial residue interactions should be carefully considered following the generation of this modeled complex (reviewed in [109]).

Apart from the challenges in modeling protein-RNA interactions, additional challenges remain in studying protein-RNA complex structures, including the fact that the interaction involves dynamics of biomolecules involved, both with respect to the RNA (which may also include modified RNAs), as well as with respect to the protein [107]. Nevertheless, the predicted complex structures from

these computational modeling methods (such as template-based docking) can serve as a starting point for simulations to provide critical insights on modeled protein-RNA interactions with respect to refining such complexes, as well as to study their dynamics and provide an in-depth biophysical investigation of the complex with structural and energetic analysis [115], [116], [117], [118], [119], [120], [121].

4.3. Advances in computational methods for investigating protein interactions with modified RNA

Simulations have been widely employed to study protein-RNA interactions [122], [123]; advances in the development of force fields of RNA modifications [124], [125], and in the ability to parametrize chemical groups [126], [127] have laid the foundation for the computational study of the interface between proteins and the epitranscriptome using MD simulations. One such application of these advancements can be found in a high-throughput computational platform for screening protein targets for modified RNA recognition [128]. This protocol employed trees of chemical modifications to the four canonical nucleosides, with the complexity of the chemical modifications increasing along the branch points. Through short implicit solvent simulations, chemical modifications that led to favorable interactions with the protein of interest when compared to the simpler “parent” modification were selected and validated with longer, explicit solvent methods. This computational protocol was applied to the polynucleotide phosphorylase (PNPase) protein from *E. coli*, which has been previously investigated in human cells for its selective recognition of 8-oxoG [83]. Following the screening of the homology modeled PNPase structure, modifications predicted by the pipeline to show increased affinity with PNPase were tested in vitro alongside m5C, a modification screened out at the explicit solvent phase [128]. The experimentally determined binding affinities showed high correlation to the association free energy data from the explicit solvent MD simulations, showcasing the method’s ability to predict possible binding targets in a high-throughput manner. This synergistic experimental and computational approach, along with current techniques to generate the necessary protein and RNA structures, represents a starting point for further investigation of protein interactions with RNA modifications at an atomistic scale. Importantly, this platform served as a steppingstone for solving the “inverse problem” of examining the interaction of PNPase with 8-oxoG in atomic detail to provide insights into the mechanism of 8-oxoG discrimination [129]. Particularly, computations were employed to evolve PNPase for higher 8-oxoG affinity by screening mutants from a library of beneficial mutations and assessed their interactions using MD simulations [128]. Perhaps most importantly, improvements in 8-oxoG binding led to increased cell tolerance to oxidative stress, providing a clear link between molecular discrimination of RNA oxidation and cell survival. Overall, this methodology provided a framework for the rational engineering of modified RNA protein readers that could be applied to other systems outside of the studied PNPase.

5. Summary and outlook

The budding field of epitranscriptomics offers new and exciting opportunities for investigation of novel protein-RNA interactions. Large-scale studies to identify proteins that interact with a handful of RNA modifications have been conducted [28], [29], [30], [37], but relatively few of these proteins have been investigated on a mechanistic level. The current characterization of epitranscriptome reader proteins on this level is limited to proteins recognizing m6A, with the *in silico* investigations focusing on the YTH family of protein readers [59], [62], [64]. These studies demonstrate the power of atomistic simulations to reveal both the molecular basis for the YTH domain selectivity for m6A modified RNA and the possible redundancy in binding mechanisms for the YTHDF proteins. However, the YTH domain family represents only a single model of recognition for m6A, and the binding mechanisms for non-YTH domain readers such as the IGF2BPs [17] and HNRNPA2B1 [7], [77] require further investigation. Relatively few of the proteins identified have been investigated in terms of direct binding to the m6A modification [5], [7], [17], [76], [77] and fewer still for those proteins shown to interact with other modified RNAs such as m1A, m5C, and 8-oxoG. The *in silico* investigation of the YTH family of proteins represents a proof-of-concept for the power of atomistic MD simulations for mechanistic understanding of epitranscriptome reader proteins. The development of both protein structure prediction tools, molecular docking, and molecular mechanics parametrization of more than 100 different modified RNA species [124], [125] have offered support for conducting insightful synergistic computational and experimental studies into how these proteins recognize their modified RNA substrates. Further investigation of these intermolecular interactions holds great promise for uncovering new mechanisms of binding and molecular recognition of RNA modifications by proteins, leading to a rich understanding of how proteins recognize the epitranscriptome.

CRedit authorship contribution statement

Lucas G. Miller: Conceptualization, Writing – original draft preparation, Reviewing & editing. Madeline Demny: Methodology, Writing – original draft preparation, Reviewing & editing, Visualization. Phanourios Tamamis: Conceptualization, Writing – reviewing & editing, Supervision, Funding acquisition. Lydia M. Contreras: Conceptualization, Writing – reviewing & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare no conflicts of interest.

Acknowledgements

Authors acknowledge discussions with members of Contreras' and Tamamis' labs. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. (NSF grant 2137420). Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation. LMC and PT acknowledge funding from NSF: Award Abstracts 2022146 and 2022124 URoL: Epigenetics 1: Collaborative Research: Novel epitranscriptomics tools to understand and modulate interactions of modified RNAs with protein readers and erasers. LMC acknowledges funding from NSF: Award Abstract 2218477 Engineering autoregulation of the epitranscriptome to track and control stress responses. Portions of figures created using BioRender.com and Visual Molecular Dynamics (VMD) [130].

References

1. Cohn W.E. Minor constituents of ribonucleic acids. *Fed Proc.* 1957;16.
2. Davis F.F., Allen F.W. Ribonucleic acids from yeast which contain a fifth nucleotide. *J Biol Chem.* 1957;227:907–915. doi: 10.1016/s0021-9258(18)70770-9.
3. 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.* 2021;50:D231–D235. doi: 10.1093/nar/gkab1083.
4. Kellner S., Burhenne J., Helm M. Detection of RNA modifications. *Rna Biol.* 2010;7:237–247. doi: 10.4161/rna.7.2.11468.
5. Liu N., Dai Q., Zheng G., He C., Parisien M., Pan T. N6-methyladenosine-dependent RNA structural switches regulate RNA–protein interactions. *Nature.* 2015;518. doi: 10.1038/nature14234.
6. Zhou K.I., Shi H., Lyu R., Wylder A.C., Matuszek Ż., Pan J.N., et al. Regulation of Co-transcriptional Pre-mRNA Splicing by m6A through the low-complexity protein hnRNPG. *Mol Cell.* 2019;76(70–81) doi: 10.1016/j.molcel.2019.07.005.
7. Alarcón C.R., Goodarzi H., Lee H., Liu X., Tavazoie S., Tavazoie S.F. HNRNPA2B1 is a mediator of m6A-dependent nuclear RNA processing events. *Cell.* 2015;162:1299–1308. doi: 10.1016/j.cell.2015.08.011.
8. Cun Y., An S., Zheng H., Lan J., Chen W., Luo W., et al. Specific regulation of m6A by SRSF7 promotes the progression of glioblastoma. *Genom Prote Bioinform.* 2021 doi: 10.1016/j.gpb.2021.11.001.
9. Shi H., Wei J., He C. Where, when, and how: context-dependent functions of RNA methylation writers, readers, and erasers. *Mol Cell.* 2019;74:640–650. doi: 10.1016/j.molcel.2019.04.025.
10. Yan H., Zhang L., Cui X., Zheng S., Li R. Roles and mechanisms of the m6A reader YTHDC1 in biological processes and diseases. *Cell Death Discov.* 2022;8:237. doi: 10.1038/s41420-022-01040-2.
11. Edupuganti R.R., Geiger S., Lindeboom R.G.H., Shi H., Hsu P.J., Lu Z., et al. N6-methyladenosine (m6A) recruits and repels proteins to regulate mRNA homeostasis. *Nat Struct Mol Biol.* 2017;24:870–878. doi: 10.1038/nsmb.3462.
12. 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.* 2013;10:93–95. doi: 10.1038/nchembio.1432.
13. Shen D., Wang B., Gao Y., Zhao L., Bi Y., Zhang J., et al. Detailed resume of RNA m6A demethylases. *Acta Pharm Sin B.* 2022;12:2193–2205. doi: 10.1016/j.apsb.2022.01.003.
14. Dominissini D. Roadmap to the epitranscriptome. *Science.* 2014;346 doi: 10.1126/science.aaa1807. 1192–1192.
15. Lan Q., Liu P.Y., Haase J., Bell J.L., Hüttelmaier S., Liu T. The critical role of RNA m6A methylation in cancer. *Cancer Res.* 2019;79:1285–1292. doi: 10.1158/0008-5472.can-18-2965.

16. Arguello A.E., Leach R.W., Kleiner R.E. In vitro selection with a site-specifically modified RNA library reveals the binding preferences of N6-methyladenosine reader proteins. *Biochem-US*. 2019;58:3386–3395. doi: 10.1021/acs.biochem.9b00485.
17. Huang H., Weng H., Sun W., Qin X., Shi H., Wu H., et al. Recognition of RNA N6-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nat Cell Biol*. 2018;20:285–295. doi: 10.1038/s41556-018-0045-z.
18. Zhou K.I., Parisien M., Dai Q., Liu N., Diatchenko L., Sachleben J.R., et al. N 6 -methyladenosine modification in a long noncoding RNA hairpin predisposes its conformation to protein binding. *J Mol Biol*. 2016;428:822–833. doi: 10.1016/j.jmb.2015.08.021.
19. Zhang F., Kang Y., Wang M., Li Y., Xu T., Yang W., et al. Fragile X mental retardation protein modulates the stability of its m6A-marked messenger RNA targets. *Hum Mol Genet*. 2018;27:3936–3950. doi: 10.1093/hmg/ddy292.
20. Abakir A., Giles T.C., Cristini A., Foster J.M., Dai N., Starczak M., et al. N6-methyladenosine regulates the stability of RNA:DNA hybrids in human cells. *Nat Genet*. 2019;52:48–55. doi: 10.1038/s41588-019-0549-x.
21. Wang X., Lu Z., Gomez A., Hon G.C., Yue Y., Han D., et al. N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature*. 2013;505:117–120. doi: 10.1038/nature12730.
22. Shi H., Wang X., Lu Z., Zhao B.S., Ma H., Hsu P.J., et al. YTHDF3 facilitates translation and decay of N6-methyladenosine-modified RNA. *Cell Res*. 2017;27:315–328. doi: 10.1038/cr.2017.15.
23. Anders M., Chelysheva I., Goebel I., Trenkner T., Zhou J., Mao Y., et al. Dynamic m6A methylation facilitates mRNA triaging to stress granules. *Life Sci Alliance*. 2018;1 doi: 10.26508/lsa.201800113.
24. Bass B.L. RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem*. 2002;71:817–846. doi: 10.1146/annurev.biochem.71.110601.135501.
25. Nishikura K. A-to-I editing of coding and non-coding RNAs by ADARs. *Nat Rev Mol Cell Bio*. 2016;17:83–96. doi: 10.1038/nrm.2015.4.
26. Roth S.H., Levanon E.Y., Eisenberg E. Genome-wide quantification of ADAR adenosine-to-inosine RNA editing activity. *Nat Methods*. 2019;16:1131–1138. doi: 10.1038/s41592-019-0610-9.
27. Esteve-Puig R., Bueno-Costa A., Esteller M. Writers, readers and erasers of RNA modifications in cancer. *Cancer Lett*. 2020;474:127–137. doi: 10.1016/j.canlet.2020.01.021.
28. Dai X., Wang T., Gonzalez G., Wang Y. Identification of YTH domain-containing proteins as the readers for N1-methyladenosine in RNA. *Anal Chem*. 2018;90:6380–6384. doi: 10.1021/acs.analchem.8b01703.
29. Dai X., Gonzalez G., Li L., Li J., You C., Miao W., et al. YTHDF2 binds to 5-methylcytosine in RNA and modulates the maturation of ribosomal RNA. *Anal Chem*. 2019;92:1346–1354. doi: 10.1021/acs.analchem.9b04505.

30. Hayakawa H., Fujikane A., Ito R., Matsumoto M., Nakayama K.I., Sekiguchi M. Human proteins that specifically bind to 8-oxoguanine-containing RNA and their responses to oxidative stress. *Biochem Biophys Res Commun.* 2010;403:220–224. doi: 10.1016/j.bbrc.2010.11.011.
31. Zhang Z., Theler D., Kaminska K.H., Hiller M., Grange P., de la, Pudimat R., et al. The YTH domain is a novel RNA binding domain. *J Biol Chem.* 2010;285:14701–14710. doi: 10.1074/jbc.m110.104711.
32. Zhen D., Wu Y., Zhang Y., Chen K., Song B., Xu H., et al. m6A Reader: Epitranscriptome target prediction and functional characterization of N6-methyladenosine (m6A) readers. *Front Cell Dev Biol.* 2020;8:741. doi: 10.3389/fcell.2020.00741.
33. Yao J.-H., Lin M.-X., Liao W.-J., Fan W.-J., Xu X.-X., Shi H., et al. M1ARegpred: epitranscriptome target prediction of N1-methyladenosine (m1A) regulators based on sequencing features and genomic features. *Front Biosci-Land.* 2022;27:269. doi: 10.31083/j.fbl2709269.
34. He Z., Xu J., Shi H., Wu S. m5CRegpred: epitranscriptome target prediction of 5-methylcytosine (m5C) regulators based on sequencing features. *Genes-Basel.* 2022;13:677. doi: 10.3390/genes13040677.
35. Walker S.C., Scott F.H., Srisawat C., Engelke D.R. RNA-protein interaction protocols. *Methods Mol Biol.* 2008;488:23–40. doi: 10.1007/978-1-60327-475-3_3.
36. Dominissini D., Nachtergaele S., Moshitch-Moshkovitz S., Peer E., Kol N., Ben-Haim M.S., et al. The dynamic N1-methyladenosine methylome in eukaryotic messenger RNA. *Nature.* 2016;530:441–446. doi: 10.1038/nature16998.
37. 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;485:201–206. doi: 10.1038/nature11112.
38. Stoilov P., Rafalska I., Stamm S. YTH: a new domain in nuclear proteins. *Trends Biochem Sci.* 2002;27:495–497. doi: 10.1016/s0968-0004(02)02189-8.
39. Xu C., Wang X., Liu K., Roundtree I.A., Tempel W., Li Y., et al. Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nat Chem Biol.* 2014;10:927–929. doi: 10.1038/nchembio.1654.
40. Zhu T., Roundtree I.A., 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;24:1493–1496. doi: 10.1038/cr.2014.152.
41. Nai F., Nachawati R., Zálesák F., Wang X., Li Y., Caflisch A. Fragment Ligands of the m6A-RNA Reader YTHDF2. *Acs Med Chem Lett.* 2022;13:1500–1509. doi: 10.1021/acsmchemlett.2c00303.
42. Li Y., Bedi R.K., Wiedmer L., Huang D., Śledź P., Caflisch A. Flexible binding of m6a reader protein YTHDC1 to its preferred RNA Motif. *J Chem Theory Comput.* 2019;15:7004–7014. doi: 10.1021/acs.jctc.9b00987.

43. Hsu P.J., Zhu Y., Ma H., Guo Y., Shi X., Liu Y., et al. Ythdc2 is an N6-methyladenosine binding protein that regulates mammalian spermatogenesis. *Cell Res.* 2017;27:1115–1127. doi: 10.1038/cr.2017.99.
44. Woodcock C.B., Horton J.R., Zhou J., Bedford M.T., Blumenthal R.M., Zhang X., et al. Biochemical and structural basis for YTH domain of human YTHDC1 binding to methylated adenine in DNA. *Nucleic Acids Res.* 2020;48:gkaa604. doi: 10.1093/nar/gkaa604.
45. 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;290:24902–24913. doi: 10.1074/jbc.m115.680389.
46. Seo K.W., Kleiner R.E. YTHDF2 recognition of N1-methyladenosine (m1A)-modified RNA is associated with transcript destabilization. *Acs Chem Biol.* 2020;15:132–139. doi: 10.1021/acscchembio.9b00655.
47. Rio D.C. Filter-binding assay for analysis of RNA–protein interactions. *Cold Spring Harb Protoc.* 2012;2012 doi: 10.1101/pdb.prot071449. pdb.prot071449.
48. Hellman L.M., Fried M.G. Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nat Protoc.* 2007;2:1849–1861. doi: 10.1038/nprot.2007.249.
49. Feig A.L. Chapter 19 Studying RNA–RNA and RNA–Protein Interactions by Isothermal Titration Calorimetry. *Methods Enzym.* 2009;468:409–422. doi: 10.1016/s0076-6879(09)68019-8.
50. Moon M.H., Hilimire T.A., Sanders A.M., Schneekloth J.S. Measuring RNA–Ligand Interactions with Microscale Thermophoresis. *Biochem-US.* 2018;57:4638–4643. doi: 10.1021/acs.biochem.7b01141.
51. Jarmoskaite I., AlSadhan I., Vaidyanathan P.P., Herschlag D. How to measure and evaluate binding affinities. *Elife.* 2020;9 doi: 10.7554/elife.57264.
52. Altschuler S.E., Lewis K.A., Wuttke D.S. Practical strategies for the evaluation of high-affinity protein/nucleic acid interactions. *J Nucleic Acids Investig.* 2012;4:3. doi: 10.4081/jnai.2013.e3.
53. Harper J.E., Miceli S.M., Roberts R.J., Manley J.L. Sequence specificity of the human mRNA N6-adenosine methylase in vitro. *Nucleic Acids Res.* 1990;18:5735–5741. doi: 10.1093/nar/18.19.5735.
54. Wei C.-M., Moss B. Nucleotide sequences at the N 6 -methyladenosine sites of HeLa cell messenger ribonucleic acid. *Biochemistry.* 1977;16:1672–1676. doi: 10.1021/bi00627a023.
55. Wu B., Xu J., Su S., Liu H., Gan J., Ma J. Structural insights into the specific recognition of DSR by the YTH domain containing protein Mmi1. *Biochem Bioph Res Co.* 2017;491:310–316. doi: 10.1016/j.bbrc.2017.07.104.
56. Ma C., Liao S., Zhu Z. Crystal structure of human YTHDC2 YTH domain. *Biochem Bioph Res Co.* 2019;518:678–684. doi: 10.1016/j.bbrc.2019.08.107.
57. Luo S., Tong L. Molecular basis for the recognition of methylated adenines in RNA by the eukaryotic YTH domain. *Proc Natl Acad Sci.* 2014;111:13834–13839. doi: 10.1073/pnas.1412742111.

58. Zhou W., Han Z., Wu Z., Gong W., Yang S., Chen L., et al. Specific recognition between YTHDF3 and m6A-modified RNA: An all-atom molecular dynamics simulation study. *Proteins Struct Funct Bioinform.* 2022 doi: 10.1002/prot.26389.
59. Li Y., Bedi R.K., Moroz-Omori E.V., Caflich A. Structural and dynamic insights into redundant function of YTHDF proteins. *J Chem Inf Model.* 2020;60:5932–5935. doi: 10.1021/acs.jcim.0c01029.
60. Li F., Zhao D., Wu J., Shi Y. Structure of the YTH domain of human YTHDF2 in complex with an m6A mononucleotide reveals an aromatic cage for m6A recognition. *Cell Res.* 2014;24:1490. doi: 10.1038/cr.2014.153.
61. Gelly J.-C., Joseph A.P., Srinivasan N., Brevern A.G. de. iPBA: a tool for protein structure comparison using sequence alignment strategies. *Nucleic Acids Res.* 2011;39:W18–W23. doi: 10.1093/nar/gkr333.
62. Li Y., Bedi R.K., Wiedmer L., Sun X., Huang D., Caflich A. Atomistic and thermodynamic analysis of N6-methyladenosine (m6A) recognition by the reader domain of YTHDC1. *J Chem Theory Comput.* 2021;17:1240–1249. doi: 10.1021/acs.jctc.0c01136.
63. Darve E. Free energy calculations, theory and applications in chemistry and biology. *Springer Ser Chem.* 2007:119–170. doi: 10.1007/978-3-540-38448-9_4.
64. Krepl M., Damberger F.F., Schroetter C., von, Theler D., Pokorná P., Allain F.H.-T., et al. Recognition of N6-methyladenosine by the YTHDC1 YTH domain studied by molecular dynamics and NMR spectroscopy: the role of hydration. *J Phys Chem B.* 2021;125:7691–7705. doi: 10.1021/acs.jpcb.1c03541.
65. Wang X., Zhao B.S., Roundtree I.A., Lu Z., Han D., Ma H., et al. N6-methyladenosine modulates messenger RNA translation efficiency. *Cell.* 2015;161:1388–1399. doi: 10.1016/j.cell.2015.05.014.
66. 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;7:12626. doi: 10.1038/ncomms12626.
67. Zaccara S., Jaffrey S.R. A unified model for the function of YTHDF proteins in regulating m6A-modified mRNA. *Cell.* 2020;181:1582–1595. doi: 10.1016/j.cell.2020.05.012. e18.
68. Xiao W., Adhikari S., Dahal U., Chen Y.-S., Hao Y.-J., Sun B.-F., et al. Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. *Mol Cell.* 2016;61:507–519. doi: 10.1016/j.molcel.2016.01.012.
69. Bedi R.K., Huang D., Wiedmer L., Li Y., Dolbois A., Wojdyla J.A., et al. Selectively disrupting m6A-dependent protein–RNA interactions with fragments. *Acs Chem Biol.* 2020;15:618–625. doi: 10.1021/acscchembio.9b00894.
70. Li Y., Bedi R.K., Nai F., Roten V., von, Dolbois A., Zálesák F., et al. Structure-based design of ligands of the m6A-RNA reader YTHDC1. *Eur J Med Chem Rep.* 2022;5 doi: 10.1016/j.ejmcr.2022.100057.

71. Micaelli M., Vedove A.D., Cerofolini L., Vigna J., Sighel D., Zaccara S., et al. Small-molecule ebselen binds to YTHDF proteins interfering with the recognition of N 6-methyladenosine-modified RNAs. *Acs Pharm Transl Sci.* 2022;5:872–891. doi: 10.1021/acsptsci.2c00008.
72. Li Q. Application of fragment-based drug discovery to versatile targets. *Front Mol Biosci.* 2020;7:180. doi: 10.3389/fmolb.2020.00180.
73. Wiedmer L., Eberle S.A., Bedi R.K., Śledź P., Caflisch A. A reader-based assay for m6a writers and erasers. *Anal Chem.* 2019;91:3078–3084. doi: 10.1021/acs.analchem.8b05500.
74. Dahlem C., Abuhaliema A., Kessler S.M., Kröhler T., Zoller B.G.E., Chanda S., et al. First small-molecule inhibitors targeting the RNA-binding protein IGF2BP2/IMP2 for cancer therapy. *Acs Chem Biol.* 2022;17:361–375. doi: 10.1021/acscchembio.1c00833.
75. Zhou K.I., Pan T. An additional class of m6A readers. *Nat Cell Biol.* 2018;20:230–232. doi: 10.1038/s41556-018-0046-y.
76. Liu N., Zhou K.I., Parisien M., Dai Q., Diatchenko L., Pan T. N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. *Nucleic Acids Res.* 2017;45 doi: 10.1093/nar/gkx141.
77. Wu B., Su S., Patil D.P., Liu H., Gan J., Jaffrey S.R., et al. Molecular basis for the specific and multivariant recognitions of RNA substrates by human hnRNP A2/B1. *Nat Commun.* 2018;9:420. doi: 10.1038/s41467-017-02770-z.
78. Yang X., Yang Y., Sun B.-F., Chen Y.-S., Xu J.-W., Lai W.-Y., et al. 5-methylcytosine promotes mRNA export — NSUN2 as the methyltransferase and ALYREF as an m5C reader. *Cell Res.* 2017;27:606–625. doi: 10.1038/cr.2017.55.
79. Wu R., Li A., Sun B., Sun J.-G., Zhang J., Zhang T., et al. A novel m6A reader Prrc2a controls oligodendroglial specification and myelination. *Cell Res.* 2019;29:23–41. doi: 10.1038/s41422-018-0113-8.
80. Hayakawa H., Uchiumi T., Fukuda T., Ashizuka M., Kohno K., Kuwano M., et al. Binding capacity of human YB-1 protein for RNA containing 8-oxoguanine. *Biochem-Us.* 2002;41:12739–12744. doi: 10.1021/bi0201872.
81. Ishii T., Hayakawa H., Sekiguchi T., Adachi N., Sekiguchi M. Role of Auf1 in elimination of oxidatively damaged messenger RNA in human cells. *Free Radic Bio Med.* 2015;79:109–116. doi: 10.1016/j.freeradbiomed.2014.11.018.
82. Ishii T., Hayakawa H., Igawa T., Sekiguchi T., Sekiguchi M. Specific binding of PCBP1 to heavily oxidized RNA to induce cell death. *Proc Natl Acad Sci.* 2018;115:201806912. doi: 10.1073/pnas.1806912115.
83. Hayakawa H., Kuwano M., Sekiguchi M. Specific binding of 8-oxoguanine-containing RNA to polynucleotide phosphorylase protein. *Biochem-Us.* 2001;40:9977–9982. doi: 10.1021/bi010595q.
84. Ong S.-E., Blagoev B., Kratchmarova I., Kristensen D.B., Steen H., Pandey A., et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to

expression proteomics*. *Mol Cell Proteom.* 2002;1:376–386. doi: 10.1074/mcp.m200025-mcp200.

85. Li X., Xiong X., Wang K., Wang L., Shu X., Ma S., et al. Transcriptome-wide mapping reveals reversible and dynamic N1-methyladenosine methylome. *Nat Chem Biol.* 2016;12:311–316. doi: 10.1038/nchembio.2040.

86. Feng M., Xie X., Han G., Zhang T., Li Y., Li Y., et al. YBX1 is required for maintaining myeloid leukemia cell survival by regulating BCL2 stability in an m6A-dependent manner. *Blood.* 2021;138:71–85. doi: 10.1182/blood.2020009676.

87. Case D.A., Cheatham T.E., Darden T., Gohlke H., Luo R., Merz K.M., et al. The Amber biomolecular simulation programs. *J Comput Chem.* 2005;26:1668–1688. doi: 10.1002/jcc.20290.

88. Eastman P., Swails J., Chodera J.D., McGibbon R.T., Zhao Y., Beauchamp K.A., et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *Plos Comput Biol.* 2017;13 doi: 10.1371/journal.pcbi.1005659.

89. Brooks B.R., Brooks C.L., Mackerell A.D., Nilsson L., Petrella R.J., Roux B., et al. CHARMM: the biomolecular simulation program. *J Comput Chem.* 2009;30:1545–1614. doi: 10.1002/jcc.21287.

90. Pronk S., Páll S., Schulz R., Larsson P., Bjelkmar P., Apostolov R., et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics.* 2013;29:845–854. doi: 10.1093/bioinformatics/btt055.

91. Huang J., MacKerell A.D. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem.* 2013;34:2135–2145. doi: 10.1002/jcc.23354.

92. Tan D., Piana S., Dirks R.M., Shaw D.E. RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc Natl Acad Sci.* 2018;115:E1346–E1355. doi: 10.1073/pnas.1713027115.

93. Huang J., Rauscher S., Nawrocki G., Ran T., Feig M., Groot B.L. de, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods.* 2017;14:71–73. doi: 10.1038/nmeth.4067.

94. Jo S., Cheng X., Lee J., Kim S., Park S., Patel D.S., et al. CHARMM-GUI 10 years for biomolecular modeling and simulation. *J Comput Chem.* 2017;38:1114–1124. doi: 10.1002/jcc.24660.

95. Yang J., Zhang Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* 2015;43:W174–W181. doi: 10.1093/nar/gkv342.

96. Kelley L.A., Mezulis S., Yates C.M., Wass M.N., Sternberg M.J.E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc.* 2015;10:845–858. doi: 10.1038/nprot.2015.053.

97. Zimmermann L., Stephens A., Nam S.-Z., Rau D., Kübler J., Lozajic M., et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol.* 2018;430:2237–2243. doi: 10.1016/j.jmb.2017.12.007.

98. Söding J., Biegert A., Lupas A.N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 2005;33:W244–W248. doi: 10.1093/nar/gki408.
99. Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., et al. The protein data bank. *Nucleic Acids Res.* 2000;28:235–242. doi: 10.1093/nar/28.1.235.
100. David A., Islam S., Tankhilevich E., Sternberg M.J.E. The AlphaFold database of protein structures: a biologist's guide. *J Mol Biol.* 2022;434 doi: 10.1016/j.jmb.2021.167336.
101. Lee H., Li C., Zhang Y., Zhang D., Otterbein L.E., Jin Y. Caveolin-1 selectively regulates microRNA sorting into microvesicles after noxious stimuli. *J Exp Med.* 2019;216:2202–2220. doi: 10.1084/jem.20182313.
102. Grützner J., Billenkamp F., Spanka D.-T., Rick T., Monzon V., Förstner K.U., et al. The small DUF1127 protein CcaF1 from rhodobacter sphaeroides is an RNA-binding protein involved in sRNA maturation and RNA turnover. *Nucleic Acids Res.* 2021;49:gkab146. doi: 10.1093/nar/gkab146.
103. Lee C., Su B.-H., Tseng Y.J. Comparative studies of AlphaFold, RoseTTAFold and Modeller: a case study involving the use of G-protein-coupled receptors. *Brief Bioinform.* 2022;23:bbac308. doi: 10.1093/bib/bbac308.
104. Jumper J., Evans R., Pritzel A., Green T., Figurnov M., Ronneberger O., et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596:583–589. doi: 10.1038/s41586-021-03819-2.
105. Center P.S.P. 14th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction 2020. <https://predictioncenter.org/casp14/> (accessed March 31, 2023).
106. Skolnick J., Gao M., Zhou H., Singh S. AlphaFold 2: why it works and its implications for understanding the relationships of protein sequence, structure, and function. *J Chem Inf Model.* 2021;61:4827–4831. doi: 10.1021/acs.jcim.1c01114.
107. Wei J., Chen S., Zong L., Gao X., Li Y. Protein–RNA interaction prediction with deep learning: structure matters. *Brief Bioinform.* 2021;23:bbab540. doi: 10.1093/bib/bbab540.
108. Varadi M., Anyango S., Deshpande M., Nair S., Natassia C., Yordanova G., et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 2021;50:D439–D444. doi: 10.1093/nar/gkab1061.
109. Szilagyi A., Zhang Y. Template-based structure modeling of protein–protein interactions. *Curr Opin Struc Biol.* 2014;24:10–23. doi: 10.1016/j.sbi.2013.11.005.
110. Moreira I.S., Fernandes P.A., Ramos M.J. Protein–protein docking dealing with the unknown. *J Comput Chem.* 2010;31:317–342. doi: 10.1002/jcc.21276.
111. Alipanahi B., DeLong A., Weirauch M.T., Frey B.J. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat Biotechnol.* 2015;33:831–838. doi: 10.1038/nbt.3300.

112. Trabelsi A., Chaabane M., Ben-Hur A. Comprehensive evaluation of deep learning architectures for prediction of DNA/RNA sequence binding specificities. *Bioinformatics*. 2019;35:i269–i277. doi: 10.1093/bioinformatics/btz339.
113. Chakravarty D., McElfresh G., Kundrotas P.J., Vakser I.A. How to choose templates for modeling of protein complexes: Insights from benchmarking template-based docking. *Proteins Struct Funct Bioinform*. 2020;88:1070–1081. doi: 10.1002/prot.25875.
114. Aloy P., Ceulemans H., Stark A., Russell R.B. The relationship between sequence and interaction divergence in proteins. *J Mol Biol*. 2003;332:989–998. doi: 10.1016/j.jmb.2003.07.006.
115. Pokorná P., Krepl M., Campagne S., Šponer J. Conformational heterogeneity of RNA stem-loop hairpins bound to FUS-RNA recognition motif with disordered RGG tail revealed by unbiased molecular dynamics simulations. *J Phys Chem B*. 2022;126:9207–9221. doi: 10.1021/acs.jpcc.2c06168.
116. Krepl M., Dendooven T., Luisi B.F., Sponer J. MD simulations reveal the basis for dynamic assembly of Hfq–RNA complexes. *J Biol Chem*. 2021;296 doi: 10.1016/j.jbc.2021.100656.
117. Krepl M., Havrila M., Stadlbauer P., Banas P., Otyepka M., Pasulka J., et al. Can we execute stable microsecond-scale atomistic simulations of protein–RNA complexes. *J Chem Theory Comput*. 2015;11:1220–1243. doi: 10.1021/ct5008108.
118. Zhao Y., Kormos B.L., Beveridge D.L., Baranger A.M. Molecular dynamics simulation studies of a protein–RNA complex with a selectively modified binding interface. *Biopolymers*. 2006;81:256–269. doi: 10.1002/bip.20408.
119. Ormazábal A., Palma J., Pierdominici-Sottile G. Molecular dynamics simulations unveil the basis of the sequential binding of RsmE to the noncoding RNA RsmZ. *J Phys Chem B*. 2021;125:3045–3056. doi: 10.1021/acs.jpcc.0c09770.
120. Tarus B., Chevalier C., Richard C.-A., Delmas B., Primo C.D., Slama-Schwok A. Molecular dynamics studies of the nucleoprotein of influenza A virus: role of the protein flexibility in RNA binding. *Plos One*. 2012;7 doi: 10.1371/journal.pone.0030038.
121. Xu D., Zheng Q.-C. Theoretical investigations on the effects of mutations in important residues of NS1B on its RNA-binding using molecular dynamics simulations. *Comput Biol Med*. 2022;145 doi: 10.1016/j.combiomed.2022.105412.
122. Šponer J., Krepl M., Banáš P., Kührová P., Zgarbová M., Jurečka P., et al. How to understand atomistic molecular dynamics simulations of RNA and protein–RNA complexes? *Wiley Inter Rev Rna*. 2017;8 doi: 10.1002/wrna.1405.
123. MacKerell A.D., Nilsson L. Molecular dynamics simulations of nucleic acid–protein complexes. *Curr Opin Struc Biol*. 2008;18:194–199. doi: 10.1016/j.sbi.2007.12.012.
124. Xu Y., Vanommeslaeghe K., Aleksandrov A., MacKerell A.D., Nilsson L. Additive CHARMM force field for naturally occurring modified ribonucleotides. *J Comput Chem*. 2016;37:896–912. doi: 10.1002/jcc.24307.

125. Aduri R., Psciuk B.T., Saro P., Taniga H., Schlegel H.B., SantaLucia J. AMBER force field parameters for the naturally occurring modified nucleosides in RNA. *J Chem Theory Comput.* 2007;3:1464–1475. doi: 10.1021/ct600329w.
126. Vanommeslaeghe K., Hatcher E., Acharya C., Kundu S., Zhong S., Shim J., et al. CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem.* 2010;31:671–690. doi: 10.1002/jcc.21367.
127. Vanommeslaeghe K., MacKerell A.D. Automation of the CHARMM general force field (CGenFF) I: bond perception and atom typing. *J Chem Inf Model.* 2012;52:3144–3154. doi: 10.1021/ci300363c.
128. Orr A.A., Gonzalez-Rivera J.C., Wilson M., Bhikha P.R., Wang D., Contreras L.M., et al. A high-throughput and rapid computational method for screening of RNA post-transcriptional modifications that can be recognized by target proteins. *Methods.* 2018;143:34–47. doi: 10.1016/j.ymeth.2018.01.015.
129. Gonzalez-Rivera J.C., Orr A.A., Engels S.M., Jakubowski J.M., Sherman M.W., O'Connor K.N., et al. Computational evolution of an RNA-binding protein towards enhanced oxidized-RNA binding. *Comput Struct Biotechnol J.* 2020;18:137–152. doi: 10.1016/j.csbj.2019.12.003.
130. Humphrey W., Dalke A., Schulten K. VMD: visual molecular dynamics. *J Mol Graph.* 1996;14(33–8):27–28. doi: 10.1016/0263-7855(96)00018-5.