

Molecular basis for the recognition of methylated adenines in RNA by the eukaryotic YTH domain

Shukun Luo and Liang Tong¹

Department of Biological Sciences, Columbia University, New York, NY 10027

Edited by Roy Parker, University of Colorado, Boulder, CO, and approved August 5, 2014 (received for review July 6, 2014)

Methylation of the N6 position of selected internal adenines (m⁶A) in mRNAs and noncoding RNAs is widespread in eukaryotes, and the YTH domain in a collection of proteins recognizes this modification. We report the crystal structure of the splicing factor YT521-B homology (YTH) domain of *Zygosaccharomyces rouxii* MRB1 in complex with a heptaribonucleotide with an m⁶A residue in the center. The m⁶A modification is recognized by an aromatic cage, being sandwiched between a Trp and Tyr residue and with the methyl group pointed toward another Trp residue. Mutations of YTH domain residues in the RNA binding site can abolish the formation of the complex, confirming the structural observations. These residues are conserved in the human YTH proteins that also bind m⁶A RNA, suggesting a conserved mode of recognition. Overall, our structural and biochemical studies have defined the molecular basis for how the YTH domain functions as a reader of methylated adenines.

The methylation of the N6 position of selected internal adenines (m⁶A) modification is widespread in eukaryotic mRNAs and noncoding RNAs (1–3), reviewed in refs. 4–6. Recent studies have linked this modification to the regulation of alternative splicing (2), RNA processing, and mRNA degradation (7). Although the exact cellular functions of this modification are still not completely understood, m⁶A has been linked to the regulation of the circadian clock (8) and m⁶A levels are highest during yeast meiosis (1). The m⁶A methyl group can be removed by the dioxygenases FTO (9) and ALKBH5 (10, 11), suggesting that m⁶A is a reversible modification on the RNA.

A consensus sequence G(m⁶A)C has been identified for this modification based on transcriptome-wide mapping (1, 2), which is consistent with that identified from earlier biochemical studies (4–6). Several YTH domain (12) family members, YTHDF1, YTHDF2, and YTHDF3 in humans (2, 7) and methylated RNA-binding protein 1 (MRB1) in yeast (1), have been shown to bind RNAs with m⁶A modification, and the binding consensus for the YTH domain of YTHDF2 is also G(m⁶A)C (7), consistent with that found by transcriptome-wide mapping.

The YTH domain contains ~160 residues and is found in yeast, plants, and animals (Fig. S1) (12, 13). The domain is located at the C-terminal end of yeast MRB1 and human YTHDF1–3 (Fig. 1A), and its sequence is well conserved among these proteins (Fig. 1B and Fig. S1). The N-terminal regions of these proteins are poorly conserved, although that of YTHDF2 mediates its function in regulating mRNA localization and degradation (7). Yeast MRB1 regulates phosphate metabolism by destabilizing the mRNA of a transcription factor of the pathway and, hence, it is also known as Pho92 (14), although direct evidence of MRB1 regulating m⁶A-containing mRNAs in yeast cells is lacking.

The structures of the YTH domains of two related proteins, human YTH domain containing protein 1 [YTHDC1; Protein Data Bank (PDB) ID code 2YUD] and YTHDC2 (2YU6), have been reported. Human YTHDC1 has 29% sequence identity to human YTHDF1 for the YTH domain (Fig. S1). YTHDC1 binds a degenerate unmethylated RNA sequence (13), which does not have similarity to the G(m⁶A)C consensus. The interaction between its YTH domain and an unmethylated RNA was studied by chemical-shift perturbation (13), but the structure of a complex

is not available. The molecular mechanism for how the YTH domain recognizes the m⁶A modification is not known.

Results and Discussion

We have determined the crystal structure at 2.7 Å resolution of the YTH domain of *Z. rouxii* MRB1 (ZrMRB1), a close homolog of *Saccharomyces cerevisiae* MRB1 (Fig. 1B), in complex with a 7-mer oligoribonucleotide with the sequence A_{−3}G_{−2}G_{−1}(m⁶A)₀C₊₁A₊₂U₊₃. The atomic model has good agreement with the X-ray diffraction data and the expected geometric parameters (Table S1). Roughly 96.5% of the residues are in the favored region of the Ramachandran plot, 3.5% in the allowed region, and no residues in the outlier region.

There are six copies of the YTH–RNA complex in the asymmetric unit. Interestingly, the six RNA molecules form three unusual, parallel dimers through extensive base-stacking, but no base-pairing, interactions (Fig. S2). The three dimers then associate into a hexamer, primarily through base stacking of the G_{−2} nucleotides of neighboring molecules. The m⁶A base is not involved in the formation of this dimer or hexamer. This hexameric assembly of the RNA molecules is, in turn, flanked by three YTH domains on each face (Fig. S2). The YTH–RNA complex has 1:1 stoichiometry in solution based on our gel filtration data and, therefore, the 6:6 complex is likely formed during crystallization. Nonetheless, the unusual assembly mechanism of this dimer and hexamer could have relevance for RNA structures in general.

The structure of the YTH domain of ZrMRB1 has a central, six-stranded β-sheet (Fig. 1C), although the strand at each edge of the sheet makes only two hydrogen bonds with the neighboring strand. There are also two smaller β-sheets, one with three strands and the other with two. Three helices cover some of the surfaces

Significance

Methylation of the N6 position of selected internal adenines (m⁶A) in mRNAs and noncoding RNAs is widespread in eukaryotes, and the YTH domain in a collection of proteins recognizes this modification. We report the crystal structure of the splicing factor YT521-B homology (YTH) domain of *Zygosaccharomyces rouxii* methylated RNA-binding protein 1 in complex with a heptaribonucleotide with an m⁶A residue in the center. The m⁶A modification is recognized by an aromatic cage, and there are also interactions with other regions of the RNA. Mutations in the RNA binding site can abolish the formation of the complex. Overall, our structural and biochemical studies have defined the molecular basis for how the YTH domain functions as a reader of methylated adenines.

Author contributions: S.L. and L.T. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 4U8T).

¹To whom correspondence should be addressed. Email: ltong@columbia.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1412742111/-DCSupplemental.

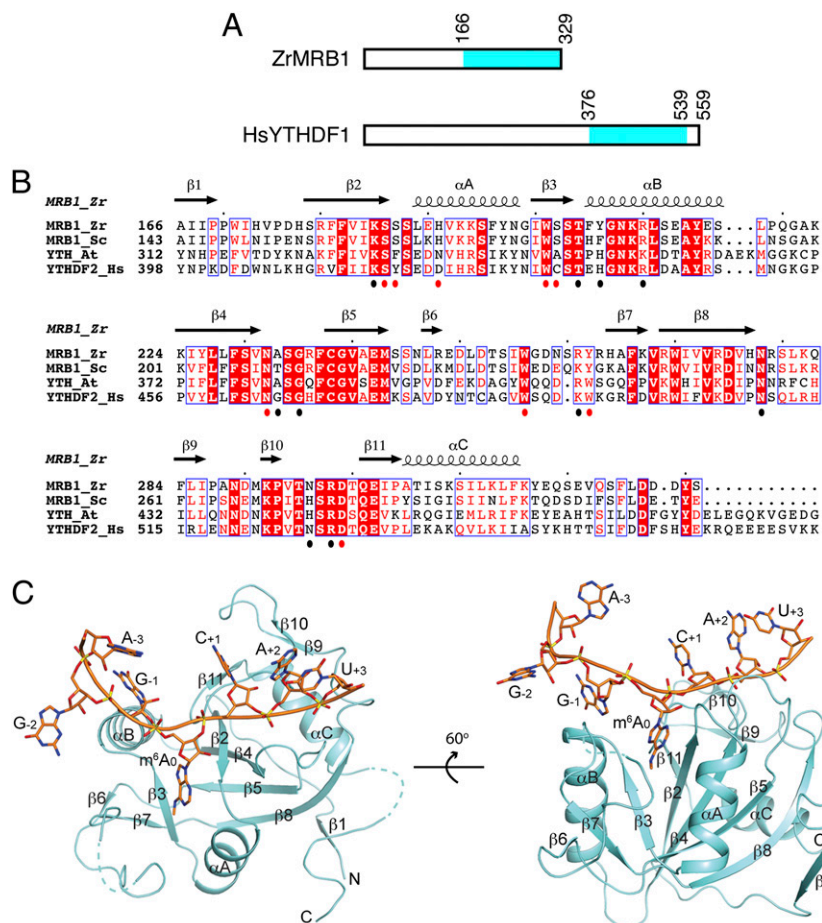


Fig. 1. Crystal structure of the YTH domain of *Z. rouxii* MRB1 in complex with an m⁶A RNA. (A) Domain organization of ZrMRB1 and human YTHDF1. The YTH domain is shown in cyan. (B) Sequence alignment of the YTH domains of the ZrMRB1 YTH domain are indicated. Residues that contact the m⁶A residue are indicated with the red dots below the alignment, and those that contact the rest of the RNA are indicated with the black dots. Produced with ESPrpt (25). (C) Two views of the structure of the YTH domain of ZrMRB1 (in cyan) in complex with a 7-mer m⁶A RNA (in orange). The secondary structure elements in the YTH domain are labeled. The structure figures were produced with PyMOL (www.pymol.org).

of these β-sheets. The overall structures of the six YTH domains in the crystal asymmetric unit are similar, with rms distance of ~0.3 Å among equivalent Cα atoms of any pair of the YTH domains (Fig. S3). Variations in the conformations of several side chains on the surface of the structure are observed. There are also conformational differences for several side chains in the RNA binding site (see below). The overall structures of the six RNA molecules are also similar to each other except for the last nucleotide (Fig. S3), which has weaker electron density in some of the molecules. Two different conformations are observed for the 5' phosphate group of G₋₂, and, in fact, several of them assume both conformations.

The closest structural homologs of ZrMRB1 are the YTH domains of human YTHDC1 and YTHDC2, with rms distance of ~2 Å and sequence identity of ~32% among structurally equivalent residues (Fig. S4). Other structural homologs, as identified in an earlier study based on YTHDC1 (13), include the DUF55 domain of human thymocyte nuclear protein 1 (15) and the EVE domain that is found in a collection of prokaryotic proteins (16) such as *Pyrococcus horikoshii* protein PH1033 (17), *Agrobacterium tumefaciens* Atu2648 (16), and a *Leishmania major* protein (18) (Fig. S4), with Z scores between 9 and 13 from the program DaliLite (19) and sequence identities between 10 and 18%. However, the functions of most of these proteins are not known.

Clear electron density is observed for most of the m⁶A heptanucleotide (Fig. 2A). The RNA is positioned across the top of

the central β-sheet of the YTH domain (Fig. 1C), and the m⁶A base is inserted into a deep pocket in the structure (Fig. 2B), providing the anchoring contacts with the protein. Residues in the RNA binding site are highly conserved among YTHDF1-3 and the MRB1 proteins, especially those involved in m⁶A binding (Figs. 1B and 2C). The 6-methylamino group is recognized by an aromatic cage, being sandwiched between the side chains of Trp200 and Tyr260, and the methyl group is pointed toward the side chain of Trp254 (Fig. 2D and E). Interestingly, the Trp200 and Tyr260 side chains flank the methylated N6 rather than the adenine base (Fig. 2F), indicating that this aromatic cage is organized to recognize the methylated base. The remaining hydrogen atom on the N6 amino group is hydrogen bonded to the main-chain carbonyl oxygen of Ser201, thereby precluding the binding of doubly methylated adenine. The N1 nitrogen atom in the adenine ring is hydrogen bonded to the side chain of His190, and the N7 atom interacts with the side chain of Asp297 through a well-ordered water molecule, which is present in all six copies of the complex in the asymmetric unit. Ser185 is positioned against one face of the adenine base, and the Ser186 side chain is near the N3 atom of the base. Overall, the structure indicates that this YTH domain produces a well-defined pocket that recognizes the m⁶A base.

Interestingly, the aromatic cage observed here for the m⁶A modification has similarity to that seen in chromo and tudor domains for recognizing methylated lysine and arginine residues (20, 21), despite the fact that the YTH domain shares no

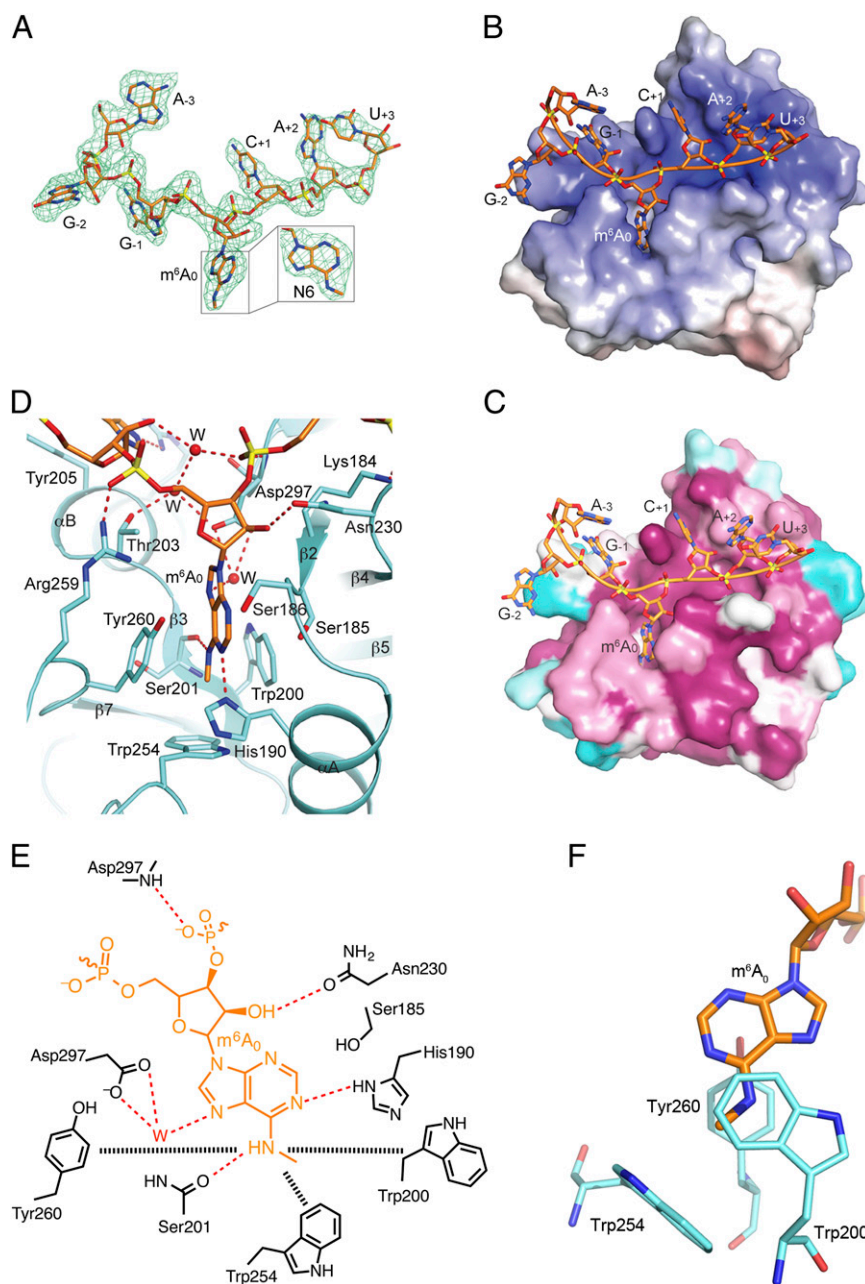


Fig. 2. Recognition of the m^6A RNA by the YTH domain. (A) Omit $F_o - F_c$ electron density for the RNA at 2.7 Å resolution, contoured at 2.5σ . (B) Electrostatic surface of the ZrMRB1 YTH domain in the region of m^6A RNA binding. The RNA is located in a positively charged surface patch (blue) in the protein. (C) Sequence conservation of residues in the RNA binding site, generated based on an alignment of 50 sequences by the program ConSurf (26). Purple, conserved residues; cyan, variable residues. (D) Detailed interactions between the m^6A nucleotide and the YTH domain. Water molecules are shown as red spheres. Hydrogen-bonding interactions are indicated by dashed lines in red. (E) Schematic drawing of the interactions between the m^6A nucleotide and the YTH domain. The interactions between the m^6A modification and the aromatic cage are indicated by the dashed lines in black. (F) Trp200 and Tyr260 flank the methylated N6 rather than the adenine base.

similarity in backbone fold with these other proteins. By manually superposing the methyl-lysine and its aromatic cage in the structures of polycomb chromo domain (22), 53BP tudor domain (23), and JMJD2A tudor domain (24) with the pocket in the YTH domain, the three aromatic residues in each cage are placed at roughly the same position (Fig. S5). This structural similarity demonstrates the remarkable conservation in the recognition of a methylated amino group.

However, this aromatic cage is not well conserved in the other structural homologs of the YTH domain (Fig. S4). Although the EVE domain in PH1033 appears to have a complete aromatic

cage, Tyr260 is replaced by a Leu residue in YTHDC1 and YTHDC2, and only Trp200 is conserved in most of the other homologs. Therefore, these proteins may not be able to recognize m^6A with good affinity. This is supported by our mutagenesis studies (see below) and the fact that YTHDC1 failed to complement the function of MRB1 in yeast, whereas YTHDF2 was able to complement (14).

The overall conformations of the three residues in the aromatic cage of MRB1 are similar to the residues in YTHDC1 and YTHDC2 (Fig. S4), which are in the absence of any bound RNA. This structural similarity suggests that the aromatic cage in

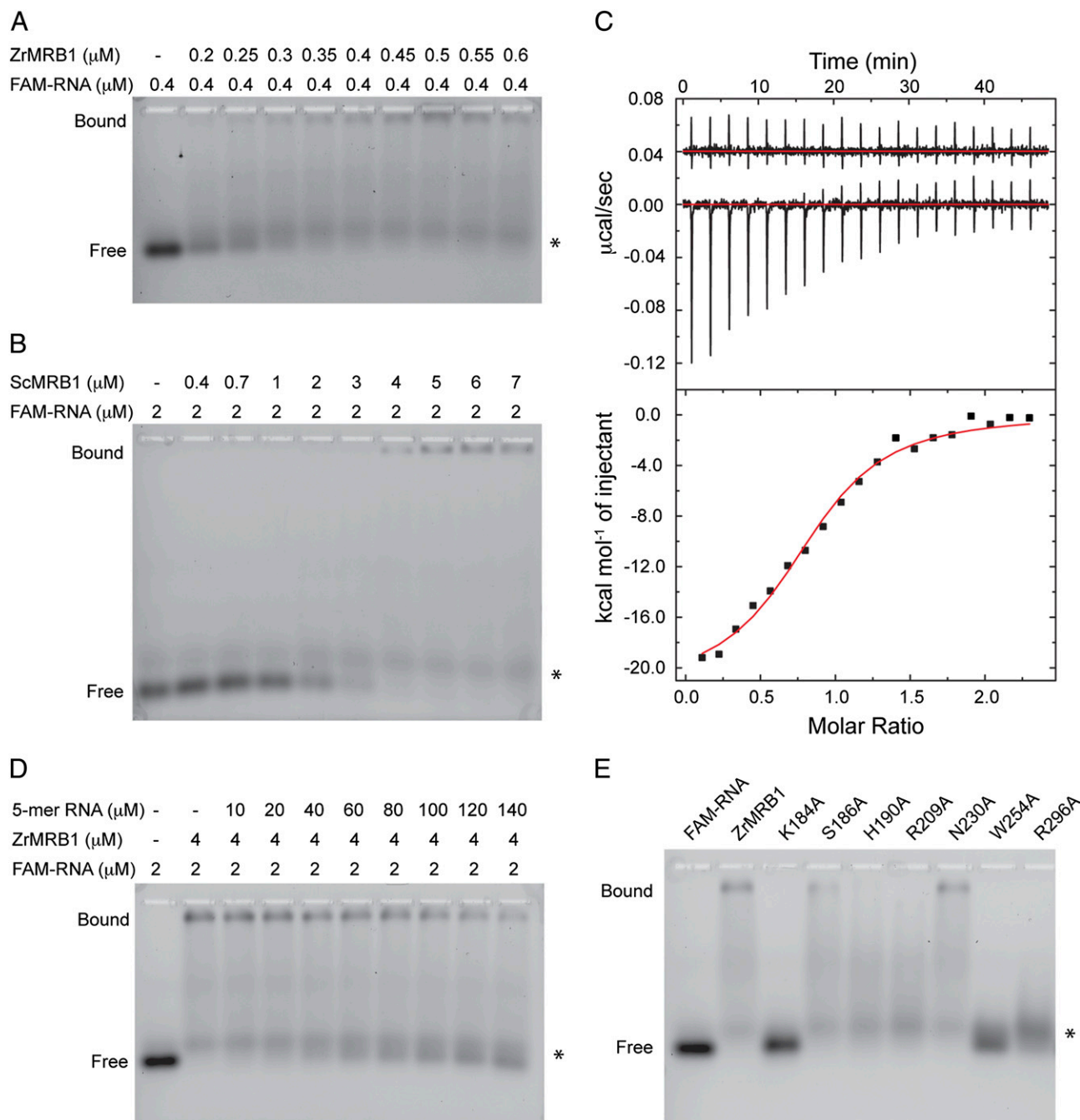


Fig. 4. Characterization of the interactions between the YTH domain and the m^6A RNA. (A) A titration experiment assessing the affinity between the ZrMRB1 YTH domain and the 7-mer m^6A RNA with a 5' FAM fluorophore. The asterisk indicates a minor contaminating species in the RNA sample that does not interact with the YTH domain. (B) A titration experiment for the ScMRB1 YTH domain. (C) ITC data for the interaction between ZrMRB1 YTH domain and the labeled 7-mer RNA. *Inset* shows the heat release from the titration into buffer alone. (D) A competition binding assay testing the affinity of an unlabeled 5-mer m^6A RNA for the ZrMRB1 YTH domain. (E) Effects of YTH domain mutations in the RNA interface on the formation of the complex. The FAM-labeled RNA is at 2 μM concentration, and the protein samples are at 4 μM concentration.

We also assessed the interaction between the ZrMRB1 YTH domain and an RNA of the same sequence but without the m^6A modification. The EMSA produced a smear of bands, and much higher concentrations of the protein were needed to shift most of the RNA molecule (Fig. S6), suggesting that the affinity of the YTH domain for this RNA is substantially lower. Mutations in the YTH domain that disrupt binding to the m^6A RNA (Fig. 4E) also interfere with the binding to this unmethylated RNA, with the interesting exception of the W254A mutation in the aromatic cage (Fig. S6). These observations indicate that the unmethylated

RNA likely assumes a similar binding mode in the YTH domain, although the aromatic cage is not crucial for this interaction. Therefore, the YTH domain in MRB1 and YTHDF1-3 may also bind unmethylated RNAs, with lower affinity, consistent with the observation that some targets of YTHDF2 do not appear to contain m^6A sites (7).

ZrMRB1 residues in the RNA interface are highly conserved among MRB1 and YTHDF1-3 proteins (Fig. S1). Especially, the aromatic cage of YTHDF1-3 contains three Trp residues, with Tyr260 of ZrMRB1 replaced by a Trp residue.

Therefore, our observations with ZrMRB1 should be directly relevant to how these other YTH domains bind the m⁶A RNA. Overall, our structural and biochemical studies have defined the molecular basis for the recognition of methylated adenines by the YTH domain.

Materials and Methods

The YTH domains of MRB1 proteins from *Zygosaccharomyces rouxii*, *S. cerevisiae*, and *Kluyveromyces lactis* were overexpressed in *Escherichia coli* and purified by nickel agarose and gel filtration chromatography. The RNAs were chemically synthesized by Dharmacon (GE Healthcare). Crystals of the YTH domain in complex with the m⁶A RNA were obtained by the

sitting-drop vapor-diffusion method at 4 °C. The crystals belong to space group *P*6₁22, and there are six complexes in the asymmetric unit. The structure was determined by the selenomethionyl single-wavelength anomalous diffraction method. The interactions between the YTH domain and the RNA were assessed with the EMSA and ITC, using an RNA with a 5' 6-FAM fluorophore label. Full experimental details are provided in the *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Neil Whalen and Annie Heroux for access to the X25 beamline. The in-house instrument for X-ray diffraction screening was purchased with a National Institutes of Health (NIH) Grant S10OD012018 (to L.T.). This research is supported by NIH Grant R01GM077175 (to L.T.).

- Schwartz S, et al. (2013) High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. *Cell* 155(6):1409–1421.
- Dominissini D, et al. (2012) Topology of the human and mouse m⁶A RNA methylomes revealed by m⁶A-seq. *Nature* 485(7397):201–206.
- Meyer KD, et al. (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* 149(7):1635–1646.
- Jia G, Fu Y, He C (2013) Reversible RNA adenosine methylation in biological regulation. *Trends Genet* 29(2):108–115.
- Niu Y, et al. (2013) N⁶-methyl-adenosine (m⁶A) in RNA: An old modification with a novel epigenetic function. *Genomics Proteomics Bioinformatics* 11(1):8–17.
- Meyer KD, Jaffrey SR (2014) The dynamic epitranscriptome: N⁶-methyladenosine and gene expression control. *Nat Rev Mol Cell Biol* 15(5):313–326.
- Wang X, et al. (2014) N⁶-methyladenosine-dependent regulation of messenger RNA stability. *Nature* 505(7481):117–120.
- Fustin JM, et al. (2013) RNA-methylation-dependent RNA processing controls the speed of the circadian clock. *Cell* 155(4):793–806.
- Jia G, et al. (2011) N⁶-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol* 7(12):885–887.
- Zheng G, et al. (2013) ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol Cell* 49(1):18–29.
- Xu C, et al. (2014) Structures of human ALKBH5 demethylase reveal a unique binding mode for specific single-stranded N⁶-methyladenosine RNA demethylation. *J Biol Chem* 289(25):17299–17311.
- Stoilov P, Rafalska I, Stamm S (2002) YTH: A new domain in nuclear proteins. *Trends Biochem Sci* 27(10):495–497.
- Zhang Z, et al. (2010) The YTH domain is a novel RNA binding domain. *J Biol Chem* 285(19):14701–14710.
- Kang HJ, et al. (2014) A novel protein, Pho92, has a conserved YTH domain and regulates phosphate metabolism by decreasing the mRNA stability of PHO4 in *Saccharomyces cerevisiae*. *Biochem J* 457(3):391–400.
- Yu F, et al. (2009) Determining the DUF55-domain structure of human thymocyte nuclear protein 1 from crystals partially twinned by tetartohedry. *Acta Crystallogr D Biol Crystallogr* 65(Pt 3):212–219.
- Bertonati C, et al. (2009) Structural genomics reveals EVE as a new ASCH/PUA-related domain. *Proteins* 75(3):760–773.
- Sugahara M, Asada Y, Morikawa Y, Kageyama Y, Kunishima N (2008) Nucleant-mediated protein crystallization with the application of microporous synthetic zeolites. *Acta Crystallogr D Biol Crystallogr* 64(Pt 6):686–695.
- Arakaki T, et al. (2006) Structure of Lmaj006129AAA, a hypothetical protein from *Leishmania major*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 62(Pt 3):175–179.
- Holm L, Kääriäinen S, Rosenström P, Schenkel A (2008) Searching protein structure databases with DALI-Lite v.3. *Bioinformatics* 24(23):2780–2781.
- Khorasanizadeh S (2011) Recognition of methylated histones: New twists and variations. *Curr Opin Struct Biol* 21(6):744–749.
- Yap KL, Zhou MM (2011) Structure and mechanisms of lysine methylation recognition by the chromodomain in gene transcription. *Biochemistry* 50(12):1966–1980.
- Fischle W, et al. (2003) Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev* 17(15):1870–1881.
- Botuyan MV, et al. (2006) Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell* 127(7):1361–1373.
- Lee J, Thompson JR, Botuyan MV, Mer G (2008) Distinct binding modes specify the recognition of methylated histones H3K4 and H4K20 by JMJD2A-tudor. *Nat Struct Mol Biol* 15(1):109–111.
- Gouet P, Courcelle E, Stuart DI, Métoz F (1999) ESPript: Analysis of multiple sequence alignments in PostScript. *Bioinformatics* 15(4):305–308.
- Armon A, Graur D, Ben-Tal N (2001) ConSurf: An algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. *J Mol Biol* 307(1):447–463.