Molecular basis for the specific and multivariant recognitions of RNA substrates by human hnRNP A2/B1

Baixing Wu¹, Shichen Su¹, Deepak P. Patil², Hehua Liu¹,³, Jianhua Gan³, Samie R. Jaffrey² & Jinbiao Ma¹

Human hnRNP A2/B1 is an RNA-binding protein that plays important roles in many biological processes, including maturation, transport, and metabolism of mRNA, and gene regulation of long noncoding RNAs. hnRNP A2/B1 was reported to control the microRNAs sorting to exosomes and promote primary microRNA processing as a potential m⁶A “reader.” hnRNP A2/B1 contains two RNA recognition motifs that provide sequence-specific recognition of RNA substrates. Here, we determine crystal structures of tandem RRM domains of hnRNP A2/B1 in complex with various RNA substrates, elucidating specific recognitions of AGG and UAG motifs by RRM1 and RRM2 domains, respectively. Further structural and biochemical results demonstrate multivariant binding modes for sequence-diversified RNA substrates, supporting a RNA matchmaker mechanism in hnRNP A2/B1 function. Moreover, our studies in combination with bioinformatic analysis suggest that hnRNP A2/B1 may mediate effects of m⁶A through a “m⁶A switch” mechanism, instead of acting as a direct “reader” of m⁶A modification.
Heterogeneous nuclear ribonucleoproteins (hnRNPs) play a variety of roles in regulating transcriptional and post-transcriptional gene expression, including RNA splicing, polyadenylation, capping, modification, export, localization, translation, and turnover. Each hnRNP contains at least one RNA-binding domain (RBD), such as RNA recognition motif (RRM), K-Homology (KH) domain, or an arginine/glycine-rich box. Sequence-specific association between hnRNPs and their RNA targets are typically mediated by one or more RBDs, which usually bind short, single-stranded RNA, but in some instances, also recognize structured RNAs.

As a core component of the hnRNP complex in mammalian cells, hnRNP A2/B1 is an abundant protein and has been implicated in numerous biological processes. The HNRNPA2B1 gene encodes two protein isoforms, A2 and B1, through alternative splicing. The B1 isoform contains an insertion of 12 amino acids (aa) at its N terminus. Both isoforms have an RNA-binding domain (RBD) composed of tandem RRMs separated by a 15-aa linker, and a C-terminal Gly-rich low complexity (LC) region that includes a prion-like domain (PrLD), an Arg box, and a PY-motif containing a M9 nuclear localization signal (PY-NLS). These domains are represented schematically with residue numbers based on the hnRNP B1 isoform (Fig. 1a). hnRNP A2/B1 is linked to several biological processes and diseases, especially neurodegenerative disorders, e.g., mutations in core PrLD in hnRNP A2/B1 cause multisystem proteinopathy and amyotrophic lateral sclerosis (ALS), through promoting excess incorporation of hnRNP A2/B1 into stress granules and driving the formation of cytoplasmic inclusions in animal models. hnRNP A2/B1 also regulates ESC self-renewal and pluripotency.

hnRNP A2/B1 has multiple effects on RNA processing through binding specific sequence. It can bind to HIV-1 RNA, causing nuclear retention of the vRNA, as well as microRNAs, sorting them into the exosomes through binding “EXO-motifs.” A transcriptome-wide analysis of hnRNP A2/B1 targets in the nervous system identified a clear preference for UAGG(A/G) motifs confirmed by three independent and complementary in vitro and in vivo approaches. This is consistent with previous studies indicating that hnRNP A2/B1 binds specifically to UAGGG, GGUAGUAG, or AGGAUAGA sequences. Another recent study demonstrated that hnRNP A2/B1 recognizes a consensus motif containing UAAUUUAA (S = G or C) in the 3’ UTR of many mRNAs and helps recruiting the CCR4-NOT deadenylase complex. In addition to participating the regulation of mRNAs, hnRNP A2/B1 is also involved in the activities of many other RNA species. For example, hnRNP A2/B1 can promote association of the long noncoding RNA HOTAIR with the nascent transcripts of HOTAIR target genes, thus to mediate HOTAIR-dependent heterochromatin initiation.

Recently, hnRNP A2/B1 was proposed to bind RNA transcripts containing N4-methyladenosine, a widespread nucleotide modification in mRNAs and noncoding RNAs. hnRNP A2/B1 was found to mediate m4A-dependent nuclear RNA processing events by binding G(m4A)C-containing nuclear RNAs in vivo and in vitro, in which hnRNP A2/B1 associates with a subset of primary microRNA transcripts through binding m4A, promoting primary microRNAs processing by recruiting the microprocessor complex Drosha and DGCGR2.

Although a series of studies with different approaches pointed out the multidimensional functions of hnRNP A2/B1 mediating by diverse RNA motifs in vivo, no detailed mechanism for the binding specificities has been determined. Thus, more biochemistry and structural studies are essential to understand the RNA-binding properties of hnRNP A2/B1 at the molecular level. Here, we report the crystal structures of hnRNP A2/B1 in complex with variant RNA targets to unveil the RNA-binding specificities and multivariant characteristics. Moreover, our structural data along with RNA-binding and bioinformatic analysis do not support that hnRNP A2/B1 functions as an m4A “reader,” since no significant preference for m4A modification by either tandem RRMs or full-length protein of hnRNP A2/B1 are observed.

RESULTS
Crystal structure of hnRNP A2/B1 bound to an 8mer RNA. To elucidate the RNA-binding properties of tandem RRMs of hnRNP A1/B1, we purified a number of truncations of the hnRNP A2/B1 protein. Using isothermal titration calorimetry (ITC) method, we characterized the RNA-binding activities of each construct with a set of RNA oligonucleotides of sequences according to previously determined binding motifs. ITC results showed that the construct containing the N-terminal fragment (aa 1–11) and two RRM domains, i.e., aa 1–195, can bind target RNAs with high affinity. In addition, deletion of the N-terminal fragment had no obvious impact on the RNA binding. This suggested that the construct containing the tandem RRMs of hnRNP A2/B1 is sufficient for binding target RNAs.

We first determined the crystal structure of RRMs (aa 12–195) of hnRNP A2/B1 bound to the 8-nt RNA oligonucleotide 5′-A1G1G2A1C1G1U1G1C2G3-3′ (termed 8mer RNA), which is derived from a recent individual-nucleotide-resolution CLIP study. ITC analysis showed that binding of the 8mer RNA occurs at a 1:1 ratio with a Kd of 276.2 nM (Fig. 1b). The crystal structure of RRMs (aa 12–195) in complex with the 8mer RNA molecule was determined to 2.60 Å resolutions, details about data collection and structure refinement are summarized in Supplementary Table 1. The tandem RRMs and an RNA molecule are in one asymmetric unit (Fig. 1c); both RRM domains of hnRNP A2/B1 adopt the characteristic RRM fold, which is a typical β α β α β α β α β α β α β α β α structure24 (Fig. 1c). Each RNA molecule is bound by an RRM1 domain from one hnRNP A2/B1 molecule in an asymmetric unit and an RRM2 domain from another hnRNP A2/B1 molecule in adjacent asymmetric unit (Fig. 1d).

RRM1 specifically recognizes AGG motif. The AGG motif of 8mer RNA substrate is specifically recognized mainly by RRM1 (Fig. 1e). For the recognition of the adenine at the first position (A1), the 2′-OH group forms a hydrogen bond and π–π interactions with the side chain of His108. Besides hydrogen bonding interactions, base stacking with the Phe24 on the other side also contributes to the definition of the binding environment (Fig. 1f, g). The 2′-OH of G2 forms a hydrogen bond with the side chain of Arg99, and N2′ groups of G2 form hydrogen bonds with the carboxyl group of the main chain of Val97 while N7 interacts with the side chain of Lys22. The base of this guanine G2 engages in stacking interactions with the base of the benzene ring of Phe66 and guanidyl group of Arg99 (Fig. 1f, g). N2′ and N2 of G3, the last nucleotide in the core recognition AGG motif, are hydrogen bonded to the side chain of Asp49, and O6 and N7 are recognized by the side chain of Arg99 (Fig. 1f, g). However, the RNA substrate from A4 to C8 are not well specifically recognized (Fig. 1f). The N6 of A4 is hydrogen bonded to the main chain of Lys186, and the 2′-OH group is hydrogen bonded to the side chain of Glu192 (Fig. 1f, g). The base of U4 is sandwiched between Phe115 and Ua base via π–π stacking, whereas the O4 of Ua forms hydrogen bonds with the amino group of the main chain of Arg185 (Fig. 1f, g).
Both RRM2 are involved in recognition of the 10mer RNA. The crystal structure of hnRNP A2/B1(aa 12–195) in complex with the 8mer RNA did not provide insight into specific RNA recognition by RRM2. We thus designed another RNA oligonucleotide shown in Supplementary Table 2, based on the speculation that RRM2 might recognize UAG according to previous sequencing results. This RNA contains both the AGG motif and the UAG motif. ITC results confirmed that the 10-nt RNA oligo 5′-A0A1G2G3A4C5U6A7G8C9-3′ (termed 10mer RNA) has a higher affinity (Fig. 1b and Supplementary Table 2). We successfully
obtained the crystal of hnRNP A2/B1 (12–195) in complex with the 10mer RNA, which was determined to 1.85 Å resolution (Supplementary Table 1). Similar to the previous complex structure with 8mer RNA, there is also only one protein molecule and one RNA molecule in the asymmetric unit. However, this time the RNA molecule is recognized by both RRM1 and RRM2 from two hnRNP A2/B1 molecules, of which the other one is from the asymmetric molecule (Fig. 2a). The 10mer RNA molecule adopts a single-stranded conformation accommodated into a positively charged groove comprised by the canonical RNA-binding surface of the RRM1 and RRM2 from two hnRNP A2/B1 proteins, in a 5′–3′ direction from RRM1 to RRM2 (Fig. 2b).

Unlike the complex structure of 8mer RNA in which only RRM1 is involved in specific recognition, the complex structure containing 10mer RNA also shows specific recognition by RRM2 (Fig. 2c). When the structures of RRM1 bound to 8mer and 10mer complex are superimposed, the AGG motifs of the two RNA substrates also superimpose well (Fig. 2d). However, the conformation of the rest of the oligos are dramatically different. Notably, the 10mer RNA is more stretched than the 8mer RNA (Fig. 2d). Moreover, the root-mean-square deviation of protein backbones between the two structures is merely 0.4 Å, suggesting the protein has not changed substantially when binding to the two different RNA targets.

The RRM2 specifically recognizes UAG motif. Due to the higher resolution, more detailed interactions are observed in the complex structure of 10mer RNA than in the 8mer RNA complex. Though the recognition of the AGG motif in the two structures is quite similar, more specific recognitions of A1 and G2 in the 10mer RNA complex structure are observed. N1 of A1 is recognized by the main chain amine group of Val97, whereas N6 and N7 form hydrogen bonds with the Lys94 side chain either directly or mediated by a water molecule (Fig. 3a, b). G2 has the most complicated interacting network in this structure, in which two more G2 base-specific recognitions are seen in the 10mer complex. These are the side chains of Gin19 and Ser102, which cooperatively bond to O6 of G2, respectively (Fig. 3b). In addition to specific recognition of the AGG motif, N6 of the 5′-end extended adenine (A0) is hydrogen bonded to the Glu92, and the N3 and N6 of A4 form hydrogen bonds with the side chain of Lys120 and Asn181 mediated by water, respectively (Fig. 3a, b).

Unlike the 8mer RNA complex structure, the UAG motif in the 10mer RNA is specifically recognized by RRM2. The side chain of Arg185 hydrogen bonds to both the 2′-OH and O2′ group of U50, and the N3 and O4 groups are hydrogen bonded to the side chain of Glu183 and Arg99, respectively (Fig. 3a, b). The base of A7 is also clamped by Phe115 and Met193 through hydrophobic interactions; N7 of A7 forms hydrogen bond with Arg185; N3, N6, and the phosphate group are hydrogen bonded to the main chain of Leu188, Lys186, and the side chain of Arg185, respectively (Fig. 3a, b). In addition to the hydrogen bond formed between phosphate group of G8 and the side chain of Arg153, the O6 of G8 base is recognized by the side chain of Lys113, and the base of G8 has another π–π stacking with Phe157 (Fig. 3a, b). The complete structure of C5 and C6 cannot be seen in our structure.

The chains of RRM1 (12–110) and RRM2 (111–195) can be superimposed with a root-mean-square deviation of 0.901 Å with a high sequence identity (Supplementary Fig. 1a). A superimposition of RRM1-AAGG with RRM2-ACUAGC indicated that the recognition of the AG core motif is very similar in RRM1 and RRM2 (Supplementary Fig. 1b, c). We thereafter mutated residues involved in specific recognition of AGG by RRM1 and UAG by RRM2, both of which reduced binding affinities according to ITC (Fig. 3c and Supplementary Fig. 2). Although the results of these amino acid mutations lined with expectations, the nucleotide mutations of 10mer RNA, especially the UAG nucleotides recognized by RRM2, have only moderate effects on the binding affinities (Fig. 3d, Supplementary Fig. 3, and Supplementary Table 2).

Multivariant RNA recognition modes of RRM1 and RRM2. In order to understand the molecular basis for hnRNP A2/B1 recognizing different RNA sequences, we grew the crystals of hnRNP A2/B1 (aa 12–195) in complex with different 10mer RNA mutants. Three complex structures containing RNA mutants A1G, U6G, and A7U were determined at high resolution (Fig. 4a, b, c). For the A1G mutant (5′-A0A1G2G3A4C5U6U7G8C9-3′), the AGG motif is shifted to 5′-end and recognized by RRM1 in a manner almost identical to the wild-type 10mer RNA structure (Fig. 4d). In addition, the G3 is recognized by the side chains of Arg99 and Glu181 of RRM1 through hydrogen bond formation (Fig. 4d, e, f). Therefore, RRM1 of hnRNP A2/B1 can specifically recognize an AGGG motif as demonstrated in the A1G mutant structure. Additionally, RRM2 recognizes UAG in the same manner as was seen in the structure of wild-type 10mer RNA (Fig. 4d, e).

In contrast, hnRNP A2/B1 adopts distinct strategies for binding another two RNAs that contained mutations in the UAG motif recognized by RRM2. When U6 is substituted with G in the U6G complex (5′-A0A1G2G3A4C5G6A7G8C9-3′), G6 lost two hydrogen bonds from Arg99 and Arg185, only keeping hydrogen bonds with Glu182. However, the recognition of the AG core motif by RRM2 is still well maintained (Fig. 4g, h, i). Interestingly, the recognition of AAGG by RRM1 is exactly same as AGGG in the A1G RNA mutant, though their binding modes of AG motif are different, suggesting that RRM1 can accommodate various purine-rich sequences (Fig. 4g, h, i and Supplementary Fig. 4a, b). For the A7U RNA mutant (5′-A0A1G2G3A4C5U6U7G8C9-3′), U7 forms hydrogen bonds with the side chain of Lys113 and Glu42 and forms a π–π stacking interaction with Phe157. More interestingly, U6 adopts a sandwich-like interaction mode with Phe115 and the A4 base, which is exactly same as the U6 in the 8mer RNA substrate (Fig. 5, k, l and Supplementary Fig. 4c, d). This suggests that RRM2 can accommodate the pyrimidine-rich UU sequence.
Meanwhile, the AAGG recognition by RRM1 in A7U RNA mutant is very similar to 8mer and 10mer RNAs, but different from A1G and U6G RNA mutant.

Unlike the effect of mutating the UAG motif recognized by RRM2, which just slightly reduced binding affinity, mutation of the AGG motif recognized by RRM1, such as G2C and G3C RNA mutants, have more obvious effects (Fig. 3d). Although we did not obtain crystal structures of these two mutants, our biochemical and structural studies suggested that RRM1 has more stringent recognition for purine-rich AG motif containing RNA sequences, but RRM2 seems to have more broad compatibility to recognize different RNA sequences, including canonical UAG motif, purine-rich GAG, and pyrimidine-rich UU sequences.

**hnRNP A2/B1 binds two antiparallel RNA strands.** A superimposition of all five structures obtained in this study suggested that hnRNP A2/B1 binds two antiparallel RNA strands using RRM1 and RRM2 concurrently (Fig. 5a). The two RRM domains in hnRNP A2/B1, similar to hnRNP A1 both in crystal structure and in solution, are held together in a fixed geometry without flexibility (Supplementary Fig. 5a). It is notable that there are extensive interactions between RRM1 and RRM2 from the same hnRNP A2/B1 molecule, including three salt-bridge interactions of Asp76-Lys168, Arg95-Asp164, Arg82-Asp162, and hydrophobic interactions between Phe20 with Leu171 (Fig. 5b), which are also observed in hnRNP A1 (Supplementary Fig. 5b). In addition, the last β-strand in RRM1 and the first β-strand in RRM2 have the same orientation, which forces the two RRM domains to bind two RNA strands antiparally, because the linker between the two domains blocks the binding of RNA targets from the same strand (Fig. 5c, d). In contrast, RNA substrates in most known structures in complex with tandem RRM proteins are bound as single-stranded RNAs and their orientations are from RRM2 to RRM1, such as HuD, HuR, PABP, U2AF65, and TDP-43 (Fig. 5e, f, g, h).

hnRNP A2/B1 does not specifically recognize m^6^A-modified RNA. In order to assess the hypothesis that hnRNP A2/B1 might be a direct m^6^A “reader” as proposed in a previous study, the m^6^A motif GGACU is included in the 8mer and 10mer RNAs. As shown in the crystal structures, there is no obvious aromatic cage-like surface that can potentially bind the m^6^A nucleotide (Fig. 6a, b), which was shown to be the key m^6^A-specificity element in previous structural studies of YTHDF1, YTHDC1, and MRB1 complexed with GGm^6^ACU (Fig. 6c and Supplementary Fig. 6). The crystal structure of hnRNP A2/B1 in complex with GGm^6^ACU could not be obtained. However, we were able to...
detect binding of hnRNP A2/B1 with the 8mer and 10mer RNA in which an adenosine was replaced with an m^6^A. In both cases, the m^6^A is present within its preferred GGACU sequence context. Notably, the ITC results indicated that the binding affinities of the m^6^A-containing 8mer RNA and 10mer RNA to the tandem RRM (12–195) were reduced onefold and tenfold, respectively, compared to the non-methylated RNA (Fig. 6d).

The N^6^ atoms of A_i in both complex structures of 8mer and 10mer RNAs form hydrogen bonds with hnRNP A2/B1 directly or through a water molecule, which may provide a possible reason why m^6^A modification would reduce the binding affinity (Figs. 1g and 3b). However, these data do not exclude the possibility that full-length hnRNP A2/B1 may form a m^6^A-binding cage through its C-terminal fragment, which contains a RGG box region (Fig. 1a). This domain may also contribute in some way to RNA binding, but was not included in our structural study (Supplementary Fig. 7a). Therefore, we purified the full-length hnRNP A2/B1 with a His_6_ tag or without the tag and a construct containing RGG box (residue 1–249), and used them to measure the binding affinities with various RNA substrates by

---

**Fig. 3** Detailed interactions between RRM and 10mer RNA. a Schematic showing RRM interactions with 10mer RNA sequence. b Close-up view showing the specific recognition from A_1 to G_8. c Mutagenesis study by ITC experiments between protein mutants and 10mer RNA substrate. d ITC experiments between hnRNP A2/B1 RRM and RNA mutants A1G, G2C, G3C, U6G, and A7U.
EMSA and ITC experiment (Supplementary Fig. 7b, c, d). The EMSA analysis indicated that full-length hnRNP A2/B1 has a slightly weaker binding affinity to the RNA with m^6^A modification than the one without m^6^A (Fig. 6e and Supplementary Fig. 7e). The ITC results using the full-length hnRNP A2/B1 or RGG box containing construct (residue 1–249) also showed similar trend as the ITC results using tandem RRMss. These ITC data, along with the EMSA results, suggest that full-length...
hnRNP A2/B1 does not specifically recognize or show enhanced binding to these m6A-modified RNA substrates in vitro (Supplementary Fig. 7f, g).

m6A sites bound by hnRNP A2/B1 in vivo. To better understand the potential binding interactions of m6A and hnRNP A2/B1 in cells, we examined the in vivo binding properties of hnRNP A2/B1. For this analysis, we used a set of 186 nuclear m6A sites comprising m6A sites in XIST, NEAT1, and MALAT1, which have been mapped at single-nucleotide resolution using miCLIP (m6A clusters from miCLIP of poly(A) RNA (Fig. 6g) showed an overlap with YTHDC1 clusters (<P < 0.0001). Thus, unlike hnRNP A2/B1, YTHDC1 appears to function as a general nuclear m6A reader.

Our finding that only a small subset of nuclear m6A sites are positioned near hnRNP A2/B1, and therefore could potentially be involved in a direct m6A-hnRNP A2/B1 interaction, is compatible with the previous analysis by Alarcon et al.33. Alarcon et al. reported that only 17% of their total m6A-seq clusters overlap with the hnRNP A2/B1 tag clusters. To determine if YTHDC1 shows greater overlap with m6A than hnRNP A2/B1 does, we performed a similar cluster overlap analysis. Approximately 43% of the miCLIP clusters from total RNA and 56% clusters from miCLIP of poly(A) RNA (Fig. 6g) showed an overlap with YTHDC1 clusters (<P < 0.0001). Thus, YTHDC1 has a considerably higher overlap with m6A than hnRNP A2/B1. Therefore, this analysis again supports the idea that YTHDC1 is the predominant nuclear m6A reader compared to hnRNP A2/B1.

Discussion

The RNA-binding domain of hnRNP A2/B1 comprises two RNA recognition motifs, RRM1 and RRM2, which is followed by a C-terminal glycine-rich region. hnRNP A2/B1 was previously demonstrated to bind UUAGGG and UAG RNA motifs through various analyses.37, 38. Recent CLIP-Seq data further showed that hnRNP A2/B1 prefers to bind A/G-rich sequences.37 However, there was no molecular basis for the recognition of different RNA substrates of hnRNP A2/B1. Here, we determined the crystal structures of the tandem RRMs of hnRNP A2/B1 in complex with various RNA substrates, revealing the molecular details of specific target RNA recognition and shedding light on the mechanism for hnRNP A2/B1 in various RNA-mediated biological functions.

The specific recognition of the AG core motif by both RRM1 and RRM2 is highly consistent with previous studies showing that hnRNP A2/B1 can bind the A2RE sequence39 and provides an explanation for how sumoylated hnRNP A2/B1 directs the loading of specific EXO-miRNAs into exosomes by binding GAGG, the so-called EXO motif40. Furthermore, our results provide the structural basis for hnRNP A2/B1 binding to the U-rich UAAUUAAU motif in the 3' UTR of some mRNAs, which was shown to be necessary for loading the CCR4-NOT complex to mRNAs41. Taken together, our structures illustrate the sequence-specific RNA-binding properties of hnRNP A2/B1 and give support to previously reported diverse binding sites37, 38 (Supplementary Table 3).

The two RRM domains in hnRNP A2/B1 interact with each other extensively in a fixed antiparallel orientation, and the binding surfaces of RRM1 and RRM2 align in the same plane, which forces hnRNP A2/B1 to bind two antiparallel RNA strands or a single-stranded RNA with a long connecting loop. This binding property of hnRNP A2/B1 offers a molecular basis for the previously described "matchmaking" hnRNP A2/B1-HOTAIR interaction, which requires multiple nucleotide recognition motifs within HOTAIR42. hnRNP A2/B1 shares similar antiparallel arrangements of its bound RNA as the polypyrimidine tract-specific splicing regulator PTB, in which the two RRM domains interact each other extensively and bind two antiparallel RNA strands. In the PTB–RNA complex, RRM3 and RRM4 form a heterodimer mediated by a hydrophobic interface, and bring together two remote RNA pyrimidine tracts.40 Moreover, two hnRNP A2/B1 proteins bound to same RNA strands can adopt various orientations, as seen in the structures of the 10mer RNA and the two 10mer RNA mutants A1G and U6G (Fig. 5j, k, l and Supplementary Fig. 5c, d). These diverse orientations are mainly due to the absence of direct interactions between the RRM domains bound to the same RNA strands. This feature may also be involved in RNA-templated aggregation and the formation of hnRNP A2/B1-containing protein–RNA granules in vivo43.

It has recently been demonstrated that hnRNP A2/B1 specifically recognizes m6A-modified RNAs.33 These RNAs share the m6A consensus sequence RGMACH and directly bind to the m6A mark with high affinity in vivo and in vitro.43 Prior to this study, the YTH domain was shown to be a "reader" of m6A. However, in addition to directly binding specific proteins, m6A can affect RNA binding through an indirect mechanism. This has been shown with two proteins, HuR and hnRNPC, both of which contain RRM domains and do not directly bind m6A. In the case of hnRNPC, m6A facilitates hnRNPC binding to a UUUUUU tract in mRNAs and long noncoding RNAs (lncRNAs) by promoting local unfolding of RNA. This unfolding is due to the weaker base pairing of U with m6A compared to A.41. m6A-induced RNA unfolding and increases the accessibility of hnRNPs.
C to single-stranded RNA, and is therefore termed an “m6A-switch.”

HU, also known as ELAVL1, has been found to preferentially bind to the 3′-UTR region of mRNAs that lack m6A. In this case, m6A could impede the formation of a structured RNA motif needed for HuR binding. Our structural study, combined with biochemical and bioinformatic results, suggests that m6A switches may account for the previously seen enhanced HuR binding adjacent to m6A. Instead of direct binding to m6A, m6A may promote accessibility of hnRNPs to certain binding sites, thereby explaining how m6A can facilitate the ability of hnRNPs to enhance nuclear events such as pri-miRNA processing. Further in vitro and in vivo investigations will be required to uncover the details of this mechanism.

Methods

Preparation of protein samples. Plasmids encoding different fragments of hnRNPs were PCR amplified from the human cDNA. PCR products were double digested with restriction endonuclease BamHI and XhoI, then ligated into a modified pET-28a plasmid carrying the Ulp1 cleavage site. Mutations were generated based on the overlap PCR. Recombinant plasmids were confirmed by DNA sequencing and transformed into Escherichia coli BL21 (DE3) to produce target proteins with N-terminal hexahistidine-sumo fusions. E. coli cells were cultured in LB medium at 37 °C with 50 mg/l kanamycin until the OD600 reached 0.6–0.8, then the bacteria were induced with 0.2 mM isopropyl-β-D-thiogalactoside (IPTG) at 18 °C for 16 h. Bacteria were collected by centrifugation, resuspended in buffer containing 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 20 mM imidazole pH 8.0, and lysed by high pressure. Cell extracts were centrifuged at 38,758 × g for 1 h at 4 °C. Supernatants were purified with Ni-NTA (GE), the target protein was washed with lysis buffer and then eluted with a buffer containing 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 20 mM imidazole and 250 mM imidazole. Ulp1 protease was added to remove the N-terminal tag and fusion protein of the recombinant protein and dialyzed with lysis buffer 3 h. The mixture was applied to another Ni-NTA resin to remove the protease and uncleaved proteins. Eluted proteins were concentrated by centrifugal ultrafiltration, loaded onto a pre-equilibrated HiLoad 16/60 Superdex 75-pg column in an Akta-puriﬁer (GE Healthcare), eluted at a flow rate of 1 ml/min with the same buffer containing 10 mM Tris-HCl pH 8.0, 100 mM NaCl. Peak fractions were analyzed by SDS-PAGE (15%, w/v) and stained with Coomassie brilliant blue R-250. Purified fractions were pooled together and concentrated by centrifugal ultraﬁltration. The concentration was determined by A280. The protein was concentrated to 10 mg/ml for crystallization trials.

RNA oligonucleotides. The RNA oligonucleotides (8mer-m6A: 5′-AGGAm6A-CUGC-3′, 10mer-m6A: 5′-AAGGAm6A-CUGC-3′) with m6A modification were ordered from Dharmacon (Thermo Scientiﬁc), and the other unmodiﬁed or
hnRNP A2/B1 does not specifically recognize m^6^A-modified RNA. 

**a** Surface representation of the environment around A2 in 8mer RNA complex. 
**b** Surface representation of the environment around A2 in 10mer RNA complex. 
**c** Crystallization and data collection. 

8mer-hnRNP A2/B1

Fig. 6 hnRNP A2/B1 does not specifically recognize m^6^A-modified RNA. **a** Surface representation of the environment around A2 in 8mer RNA complex. 
**b** Surface representation of the environment around A2 in 10mer RNA complex. 
**c** Crystallization and data collection. 

8mer-hnRNP A2/B1

hnRNP A2/B1 RRMs (12-195) were ordered from Bioneer Corporation. All the RNA oligonucleotides used for crystallization and biochemical experiments in this study are summarized in Supplementary Table 2.

### Crystallization and data collection. 

hnRNP A2/B1 RRMs (12-195) in complex with 8-nt-RNA 5′-AGAGGACUAGC-3′ and 5′-FAM-GGm6ACU-3′ were ordered from Bioneer Corporation. All the RNA oligonucleotides used for crystallization and biochemical experiments in this study are summarized in Supplementary Table 2.

**Crystallization and data collection.** 

hnRNP A2/B1 RRMs (12-195) in complex with 8-nt-RNA 5′-AGAGGACUAGC-3′ and 5′-FAM-GGm6ACU-3′ were ordered from Bioneer Corporation. All the RNA oligonucleotides used for crystallization and biochemical experiments in this study are summarized in Supplementary Table 2.

hnRNP A2/B1 RRMs (12-195) in complex with 8-nt-RNA 5′-AGAGGACUAGC-3′ and 5′-FAM-GGm6ACU-3′ were ordered from Bioneer Corporation. All the RNA oligonucleotides used for crystallization and biochemical experiments in this study are summarized in Supplementary Table 2.

### Crystallization and data collection. 

hnRNP A2/B1 RRMs (12-195) in complex with 8-nt-RNA 5′-AGAGGACUAGC-3′ and 5′-FAM-GGm6ACU-3′ were ordered from Bioneer Corporation. All the RNA oligonucleotides used for crystallization and biochemical experiments in this study are summarized in Supplementary Table 2.

hnRNP A2/B1 RRMs (12-195) in complex with 8-nt-RNA 5′-AGAGGACUAGC-3′ and 5′-FAM-GGm6ACU-3′ were ordered from Bioneer Corporation. All the RNA oligonucleotides used for crystallization and biochemical experiments in this study are summarized in Supplementary Table 2.

hnRNP A2/B1 RRMs (12-195) in complex with 8-nt-RNA 5′-AGAGGACUAGC-3′ and 5′-FAM-GGm6ACU-3′ were ordered from Bioneer Corporation. All the RNA oligonucleotides used for crystallization and biochemical experiments in this study are summarized in Supplementary Table 2.

hnRNP A2/B1 RRMs (12-195) in complex with 8-nt-RNA 5′-AGAGGACUAGC-3′ and 5′-FAM-GGm6ACU-3′ were ordered from Bioneer Corporation. All the RNA oligonucleotides used for crystallization and biochemical experiments in this study are summarized in Supplementary Table 2.

hnRNP A2/B1 RRMs (12-195) in complex with 8-nt-RNA 5′-AGAGGACUAGC-3′ and 5′-FAM-GGm6ACU-3′ were ordered from Bioneer Corporation. All the RNA oligonucleotides used for crystallization and biochemical experiments in this study are summarized in Supplementary Table 2.

hnRNP A2/B1 RRMs (12-195) in complex with 8-nt-RNA 5′-AGAGGACUAGC-3′ and 5′-FAM-GGm6ACU-3′ were ordered from Bioneer Corporation. All the RNA oligonucleotides used for crystallization and biochemical experiments in this study are summarized in Supplementary Table 2.

hnRNP A2/B1 RRMs (12-195) in complex with 8-nt-RNA 5′-AGAGGACUAGC-3′ and 5′-FAM-GGm6ACU-3′ were ordered from Bioneer Corporation. All the RNA oligonucleotides used for crystallization and biochemical experiments in this study are summarized in Supplementary Table 2.

hnRNP A2/B1 RRMs (12-195) in complex with 8-nt-RNA 5′-AGAGGACUAGC-3′ and 5′-FAM-GGm6ACU-3′ were ordered from Bioneer Corporation. All the RNA oligonucleotides used for crystallization and biochemical experiments in this study are summarized in Supplementary Table 2.
Structure determination and refinement. For hnRNP A2/B1 RRMs (12–195)-nt complex, the diffraction data set was processed and scaled using HKL3000. The phase was determined by molecular replacement using the program Phaser with the structure of UP1 (PDB code: 1U1Q) as the search model. Cycles of refinement and model building were carried out using REFMAC5 and COOT until the crystallography Rfree and Rfree converged to 19.16% and 23.62%, respectively. Ramachandran analysis showed that 96.0% of the residues were in the most favored region, with 2.6% in the additionally allowed region. For hnRNP A2/B1-18-nt complex, the diffraction data set was processed and scaled using the HKL3000 package. The phase was determined by molecular replacement using the program Phaser with the hnRNP A2/B1(12–195) model collected before as the search model. Curve fitting to a single binding site model was performed by the ITC data analysis module of Origin 7.0 (MicroCal) provided by the manufacturer. A$^2$ of protein–RNA binding was computed as R2(1/Kd), where R, T, and Kd are the gas constant, temperature and dissociation constant, respectively.

Comparison of hnRNP A2B1 and YTHDC1 binding at m6A sites. hnRNP A2/B1 or YTHDC1 binding and m6A stoichiometry at 10 bp flanking miCLIP sites were on mRNA RNAs such as MALAT1, NEAT1; and XIST was compared using an XY-scatter plot in R. Only m6A sites conforming a non-BCANN consensus were considered for this analysis. These represent unique sites obtained from merging (mergeBed -s -d 2) of CIMS- and CITS-based m6A sites call from ref. 34. All the RNA, tRNA, and mitochondrial genomic miCLIP sites were removed. Tag counting was performed using the bedtools suite. Tag counts (snpPM + 1) were compared using scatter plots and Pearson correlation coefficients (r) were determined in R. Cluster overlap analysis was carried out using bedtools intersect tool (intersectBed -s -u).

Data availability. The coordinates that support the findings of this study have been deposited in the Protein Data Bank with accession codes 5EN1 for RRMs-8mer-RNA complex and SH04 for RRMs–10mer RNA structures. A1G, U6G, and A7U are 5WWE, 5WFW, and 5WGW, respectively. Other data in this study are available from the corresponding author on reasonable request.

Received: 30 June 2017 Accepted: 22 December 2017
Published online: 29 January 2018

References

1. Keene, J. D. RNA regulons: coordination of post-transcriptional events. Nat. Rev. Genet. 8, 533–543 (2007).
2. Glisovic, T., Bachorik, J. L., Yong, J. & Dreyfuss, G. RNA-binding proteins and post-transcriptional gene regulation. FEBs Lett. 582, 1977–1986 (2008).
3. He, Y. & Smith, R. Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B. Cell. Mol. Life Sci. 66, 1239–1256 (2009).
4. Gabut, M., Chaudhry, S. & Blencowe, B. J. SnapShot: the splicing regulatory machinery. Cell 133, 192 e1 (2008).
5. Cook, K. B., Kazhan, H., Zuberi, K., Morris, Q. & Hughes, T. R. RBPD: a database of RNA-binding specificities. Nucleic Acids Res. 39, D301–D308 (2011).
6. Auweter, S. D., Oberstrass, F. C. & Allain, F. H. Sequence-specific binding of single-stranded RNA: is there a code for recognition? Nucleic Acids Res. 34, 4943–4959 (2006).
7. Kozu, T., Henrich, B. & Schafer, K. P. Structure and expression of the gene (HNRPA2B1) encoding the human hnRNP protein A2/B1. Genomics 25, 365–371 (1995).
8. Burd, C. G., Swanson, M. S., Gorlach, M. & Dreyfuss, G. Primary structures of the heterogeneous nuclear ribonucleoprotein A2, B1, and C2 proteins: a diversity of RNA binding proteins is generated by small peptide inserts. Proc. Natl Acad. Sci. USA 86, 9788–9792 (1989).
9. Kim, H. J. et al. Mutations in prion-like domains in hnRNPA2B1 and hnRNP A1 cause multisystem proteinopathies and ALS. Nature 495, 467–473 (2013).
10. Liu, B. J. et al. Rules for nuclear RNA sequence recognition by karyopherin beta 2. Cell 126, 543–558 (2006).
11. Harrison, A. F. & Shorter, J. RNA-binding proteins with prion-like domains in health and disease. Biochem. J. 474, 1417–1438 (2017).
12. Choi, H. S., Lee, H. M., Jang, Y. J., Kim, C. H. & Ryu, C. J. Heterogeneous nuclear ribonucleoprotein A2/B1 regulates the self-renewal and pluripotency of human embryonic stem cells via the control of the G1/S transition. Stem Cells 31, 2647–2658 (2013).
13. Beraud, V. et al. A late role for the association of hnRNP A2 with the HIV-1 hnRNP A2 response elements in genomic RNA. Gag, and Vpr localization. J. Biol. Chem. 279, 44141–44153 (2004).
14. Villarrota-Beltri, C. et al. Sumoylated hnRNP A2B1 controls the sorting of mRNA into exosomes through binding to specific motifs. Nat. Commun. 4, 2980 (2013).
15. Hutten, S. & Dorrnann, D. hnRNP A2/B1 function in neurodegeneration: it’s a gain, not a loss. Neuron 92, 672–674 (2017).
16. Martinez, F. J. et al. Protein-RNA networks regulated by normal and ALS-associating point mutations in MALAT1. Neuron 86, 636–652 (2017).
17. Geissler, R. et al. A widespread location-specific mRNA decay pathway mediated by hnRNP A1 and A2/B1. Genes Dev. 30, 1070–1085 (2016).
18. Meredith, E. K., Balas, M. M., Sindy, K., Haislop, K. & Johnson, A. M. An RNA matchmaker protein regulates the activity of the long noncoding RNA HOTAIR. RNA 22, 995–1010 (2016).
19. Meyer, K. D. & Jeffrey, S. R. The dynamic epitranscriptome: N6-methyladenosine and gene expression control. Nat. Rev. Mol. Cell Biol. 15, 313–326 (2014).
20. Meyer, K. D. et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. Cell 149, 1635–1646 (2012).
23. Alarcon, C. R. et al. HNRNPA2B1 is a mediator of m(6)A-dependent nuclear RNA processing events. Cell 162, 1295–1309 (2015).

24. Daubner, G. M., Chery, A. & Allain, F. H. RRM-RNA recognition: NMR or crystallography… and new findings. Curr. Opin. Struct. Biol. 23, 100–108 (2013).

25. Ding, J. et al. Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. Genes Dev. 13, 1102–1115 (1999).

26. Linder, B. et al. Single-nucleotide-resolution mapping of m6A and m6Am methylated adenosine modifications. Nat. Methods 15, 274–277 (2018).

27. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132 (2004).

28. Xiao, W. et al. Nuclear m(6)A reader YTHDC1 regulates mRNA splicing. Nature 519, 482–485 (2015).

29. Hutchison, S., LeBel, C., Blanchette, M. & Chabot, B. Distinct sets of adjacent polyadenylate RNA by the poly(A)-binding protein. J. Biol. Chem. 281, 2072–2083 (2006).

30. Barraud, P. & Allain, F. H. Solution structure of the two RNA recognition motifs of hnRNP A1 using segmental isotope labeling: how the relative orientation between RRMs influences the nucleic acid binding topology. J. Biomol. NMR 55, 119–138 (2013).

31. Mackereth, C. D. et al. Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. Nature 475, 408–411 (2011).

32. Mukhamedshin, F., LeBel, C. & Chabot, B. Distinct sets of adjacent heterogeneous nuclear ribonucleoprotein A2/B1 response element for RNA translation. Mol. Biol. Cell 22, 2126–2132 (2011).

33. Linder, B. et al. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. Nat. Methods 12, 767–772 (2015).

34. Linder, B. et al. m6A RNA methylation promotes XIST-mediated transcriptional repression. Nature 537, 369–373 (2016).

35. Patil, D. P. et al. m6A RNA methylation promotes XIPI-mediated transcriptional repression. Nature 537, 369–373 (2016).

36. Xiao, W. et al. Nuclear m(6)A reader YTHDC1 regulates mRNA splicing. Mol. Cell 61, 507–519 (2016).

37. McKay, S. J. & Cooke, H. hnRNP A2/B1 binds specifically to single stranded vertebrate telomeric repeat TTAGGG. Nucl. Acids Res. 20, 6461–6464 (1992).

38. Roost, C. et al. Structure and thermodynamics of N6-methyladenosine in RNA: a spring-loaded base modification. J. Am. Chem. Soc. 137, 2107–2115 (2015).

39. Alarcon, C. R. et al. HNRNPA2B1 is a mediator of m(6)A-dependent nuclear RNA processing events. Cell 162, 1295–1309 (2015).

40. Roost, C. et al. Structure and thermodynamics of N6-methyladenosine in RNA: a spring-loaded base modification. J. Am. Chem. Soc. 137, 2107–2115 (2015).

41. Moore, M. J. et al. Mapping argonauta and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CEMS analysis. Nat. Protoc. 9, 263–293 (2014).

42. Zhang, C. & Darnell, R. B. Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nat. Biotechnol. 29, 607–614 (2011).

43. Wang, X. & Tanaka Hall, T. M. Structural basis for recognition of AU-rich element RNA by the HuD protein. Nat. Struct. Biol. 8, 141–145 (2001).

44. Wang, H. et al. The structure of the ARE-binding domains of Hu antigen R (HuR) undergoes conformational changes during RNA binding. Acta Crystallogr. D Biol. Crystallogr. 69, 373–380 (2013).

45. Barraud, P. & Allain, F. H. Solution structure of the two RNA recognition motifs of hnRNP A1 using segmental isotope labeling: how the relative orientation between RRMs influences the nucleic acid binding topology. J. Biomol. NMR 55, 119–138 (2013).

46. Morgan, C. E. et al. The first crystal structure of the UP1 domain of hnRNP A1 bound to RNA reveals a new look for an old RNA binding protein. J. Mol. Biol. 427, 3241–3251 (2015).

47. Daubner, G. M., Clery, A. & Allain, F. H. RRM-RNA recognition: NMR or crystallography… and new findings. Curr. Opin. Struct. Biol. 23, 100–108 (2013).

48. Wang, X. & Tanaka Hall, T. M. Structural basis for recognition of AU-rich element RNA by the HuD protein. Nat. Struct. Biol. 8, 141–145 (2001).

49. Wang, H. et al. The structure of the ARE-binding domains of Hu antigen R (HuR) undergoes conformational changes during RNA binding. Acta Crystallogr. D Biol. Crystallogr. 69, 373–380 (2013).

50. Linder, B. et al. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. Nat. Methods 12, 767–772 (2015).

51. Patil, D. P. et al. m6A RNA methylation promotes XIPI-mediated transcriptional repression. Nature 537, 369–373 (2016).

52. Xiao, W. et al. Nuclear m(6)A reader YTHDC1 regulates mRNA splicing. Mol. Cell 61, 507–519 (2016).

53. McKay, S. J. & Cooke, H. hnRNP A2/B1 binds specifically to single stranded vertebrate telomeric repeat TTAGGG. Nucl. Acids Res. 20, 6461–6464 (1992).

54. Hutchison, S., LeBel, C., Blanchette, M. & Chabot, B. Distinct sets of adjacent heterogeneous nuclear ribonucleoprotein (hnRNP) A1/A2 binding sites control 5’ splice site selection in the hnRNP A1 mRNA precursor. J. Biol. Chem. 277, 29745–29752 (2002).

55. Munro, T. P. et al. Mutational analysis of a heterogenous nuclear ribonucleoprotein A2 response element for RNA translation. J. Biol. Chem. 274, 34389–34395 (1999).

56. Oberstrass, F. C. et al. Structure of PTB bound to RNA: specific binding and implications for splicing regulation. Science 309, 2054–2057 (2005).

57. Roost, C. et al. Structure and thermodynamics of N6-methyladenosine in RNA: a spring-loaded base modification. J. Am. Chem. Soc. 137, 2107–2115 (2015).

58. Liu, N. et al. N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. Nature 518, 560–564 (2015).

59. Wang, Y. et al. N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. Nat. Cell Biol. 16, 191–198 (2014).

60. McCoy, A. J. et al. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674 (2007).

61. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132 (2004).

62. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. D Biol. Crystallogr. 67, 355–367 (2011).