Supplementary Data

Structural basis for RNA recognition by the N-terminal tandem RRM domains of human RBM45

Xiaolei Chen[1,2,#], Zhongmei Yang[1,2,#], Wenfeng Wang[1,2], Kaiyue Qian[1,2,3], Mingjie Liu[1,2,3], Junchao Wang[1,2,3], Mingzhu Wang[1,2,3,*]

[1] Institutes of Physical Science and Information Technology, Anhui University, Hefei 230601, Anhui, China;

[2] School of Life Sciences, Anhui University, Hefei 230601, Anhui, China;

[3] Key Laboratory of Human Microenvironment and Precision Medicine of Anhui Higher Education Institutes, Anhui University, Hefei 230601, Anhui, China

[#] These authors contributed equally to this work.

* To whom correspondence should be addressed, wangmzh@ahu.edu.cn (MW)

Table S1. The FP results for sequence preferences of RBM45^{RRM1-2} and its mutants

Position	Sequence 5'-3'	K_D (μM)		
		WT	F124A/Y165A	F29A/F70A
1	GGACGG	0.477±0.051	1.21±0.10	1.32±0.10
	AGACGG	0.773±0.088	1.35±0.19	1.51±0.15
	CGACGG	0.427±0.051	1.09±0.13	1.06±0.11
	UGACGG	0.506±0.050	1.06±0.11	1.39±0.13
2	GGACGG	0.477±0.051	1.21±0.10	1.32±0.10
	GAACGG	6.13±0.41	21.8±2.9	16.7±4.5
	GCACGG	7.28±0.39	21.4±3.3	26.3±6.0
	GUACGG	4.40±0.29	15.0±1.9	19.1±3.1
3	GGACGG	0.477±0.051	1.21±0.10	1.32±0.10
	GGGCGG	2.72±0.15	15.8±1.6	21.1±2.1
	GGCCGG	4.25±0.17	18.9±1.3	21.6±2.8
	GGUCGG	1.88±0.10	16.6±1.1	18.0±2.0
4	GGACGG	0.477±0.051	1.21±0.10	1.32±0.10
	GGAAGG	3.90±0.18	23.1±2.3	23.7±3.2
	GGAGGG	2.06±0.19	16.0±1.1	22.8±1.6
	GGAUGG	3.04±0.19	19.8±1.2	23.6±1.9
5	GGACGG	0.477±0.051	1.21±0.10	1.32±0.10
	GGACAG	0.879±0.092	1.93±0.13	2.37±0.20
	GGACCG	1.00±0.10	2.19±0.20	2.57±0.20
	GGACUG	0.654±0.065	1.49±0.13	1.78±0.13
6	GGACGG	0.477±0.051	1.21±0.10	1.32±0.10
	GGACGA	0.552±0.061	1.77±0.13	2.06±0.14
	GGACGC	0.683±0.062	2.01±0.15	2.35±0.16
	GGACGU	0.553±0.054	1.56±0.13	2.07±0.13

Table S2. The RNA sequences for 2-GAC-motif RNA binding assays

RNA	Length (nt)	Distance (nt)	Sequence (5'-3')
6-mer	6	1 GAC	CGACGG
9-mer	9	0	CGACGACGC
10-mer	10	1	CGACGGACGC
11-mer	11	2	CGACGGGACGC
12-mer	12	3	CGACGGGGACGC
12-mer (GAC-polyU)	12	1 GAC	CGACGGUUUUUU
14-mer	14	5	CGACGGUUGGACGC
16-mer	16	7	CGACGGUUUUGGACGC
18-mer	18	9	CGACGGUUUUUUGGACGC
20-mer	20	11	CGACGGUUUUUUUUGGACGC
22-mer	22	13	CGACGGUUUUUUUUUUGGACGC
26-mer	26	17	CGACGGUUUUUUUUUUUUUUGGACGC

The GAC motifs are underlined.

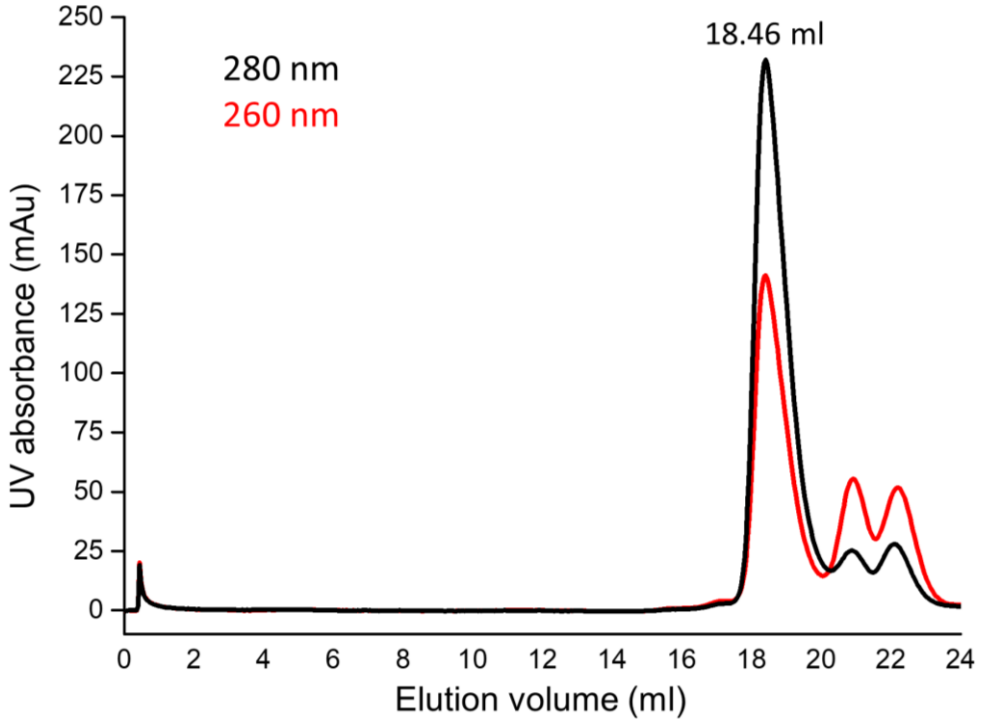

Fig. S1 Gel filtration of RBM45^{RRM1-2}. A Superdex 200 Increase 10/300 GL column was used for this experiment. The 280 nm and 260 nm absorbances are shown as black and red curves, respectively.

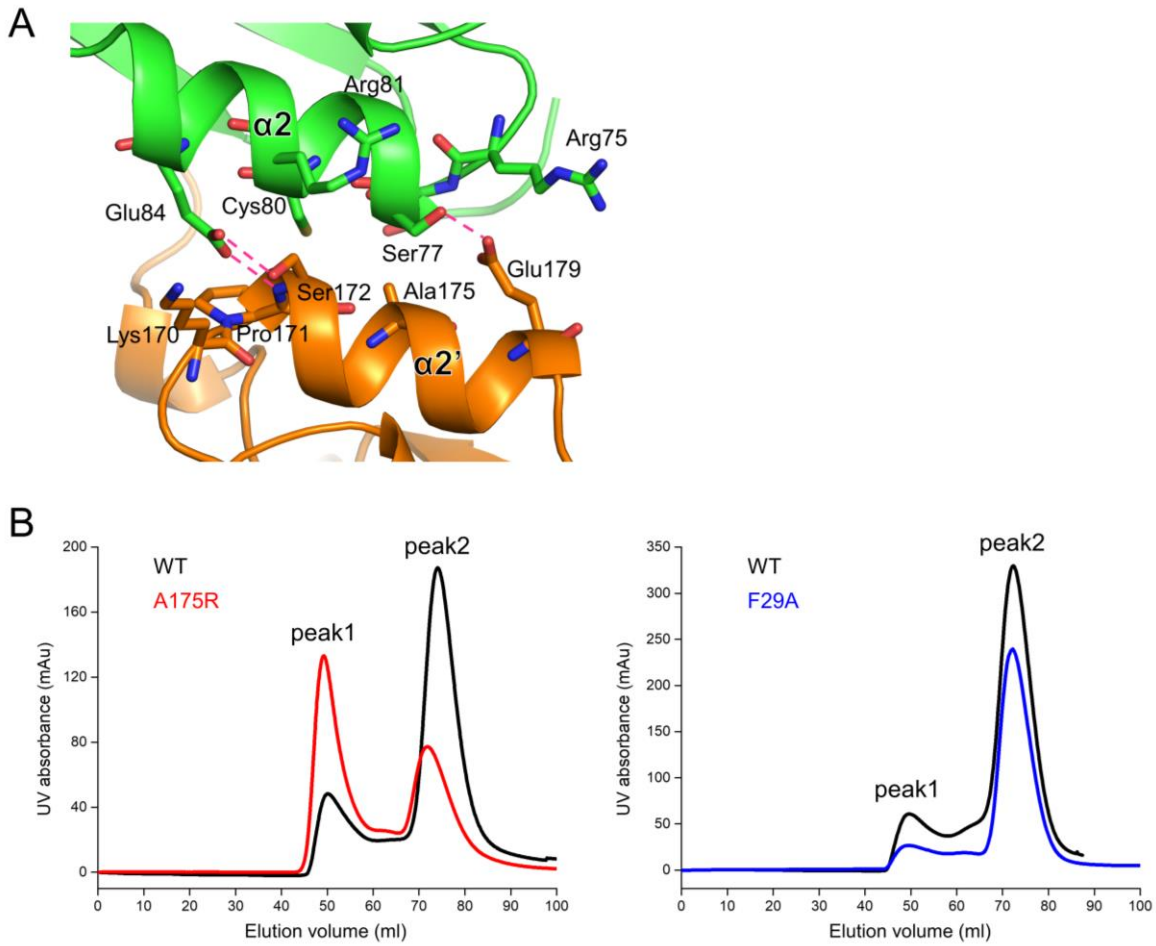

Fig. S2. Interactions between α2 of RRM1 and α2' of RRM2. (A) The residues involved in interdomain interactions are shown as sticks. The RRM1 and RRM2 are colored in green and orange, respectively. The inter-domain hydrogen bonds are indicated as magenta dashed lines. (B) Gel filtration of WT, A175R, and F29A mutant of RBM45^{RRM1-2}. WT and mutants of RBM45^{RRM1-2} were loaded to a Superdex 75 PG 16/600 column (GE) with a solution of 20 mM Tris-HCl pH7.5, 300 mM NaCl, and 1 mM DTT. The 280 nm absorbances of WT and A175R mutant are shown as black and red curves, respectively. The 280 nm absorbances of the F29A mutant, which is not involved in the RRM1-RRM2 interaction, is shown as a blue curve. The peak1 were eluted at the void volume, indicating an aggregation; the peak2 were the RBM45^{RRM1-2} monomer fractions. The A175R mutation led to more aggregation, whereas F29A did not.

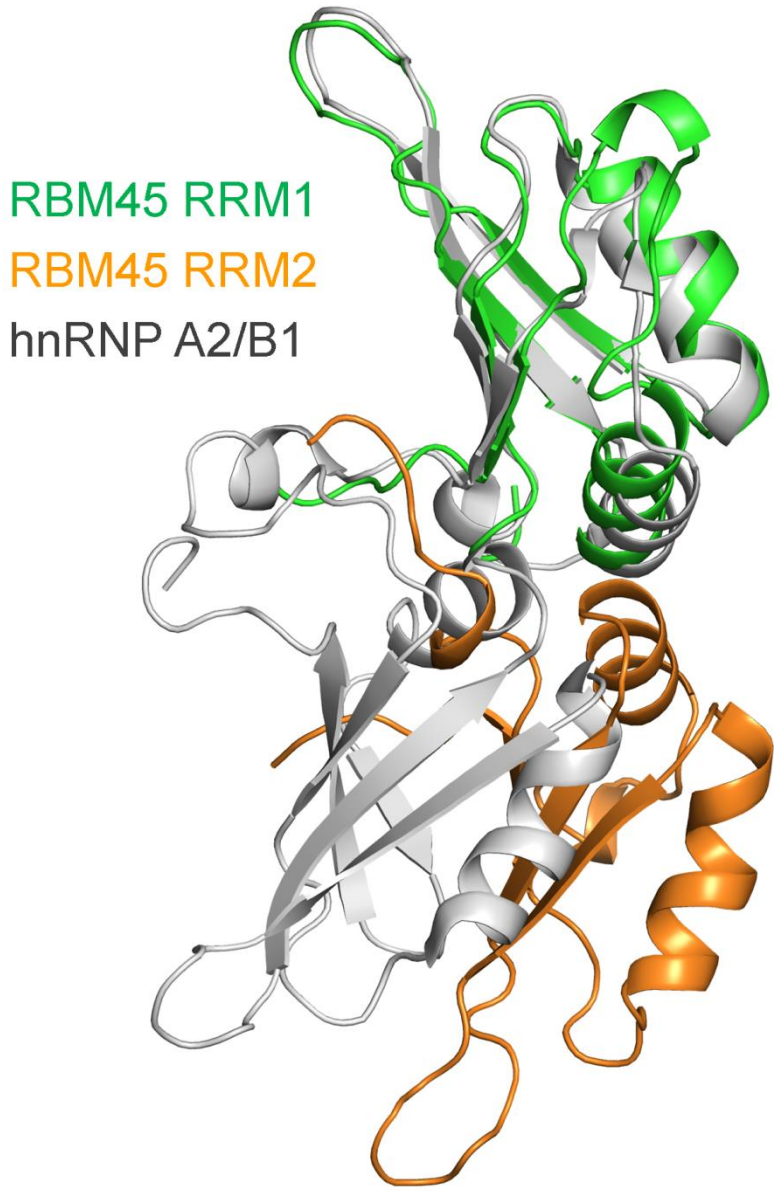

Fig. S3. Superimposition of RBM^{RRM1-2} and hnRNP A2/B1. The RRM1 of RBM45 is superposed with the RRM1 of hnRNP A2/B1 (PDB code: 5HO4). The RRM1 and RRM2 of RBM45 are colored green and orange, respectively, the hnRNP A2/B1 is colored gray.

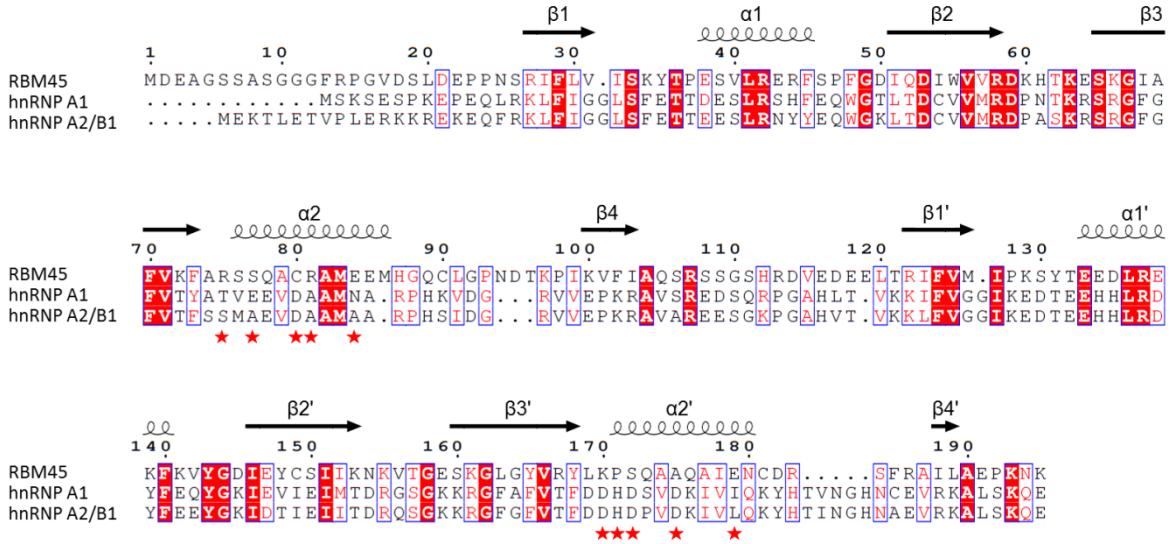

Fig. S4. Sequence alignment of RRM1 and RRM2 of RBM45 with RRM1 and RRM2 of hnRNP A1 and hnRNP A2/B1. The identical and conserved residues are highlighted in red background and red letters, respectively. The residues involved in α2–α2' interaction are denoted by red stars.

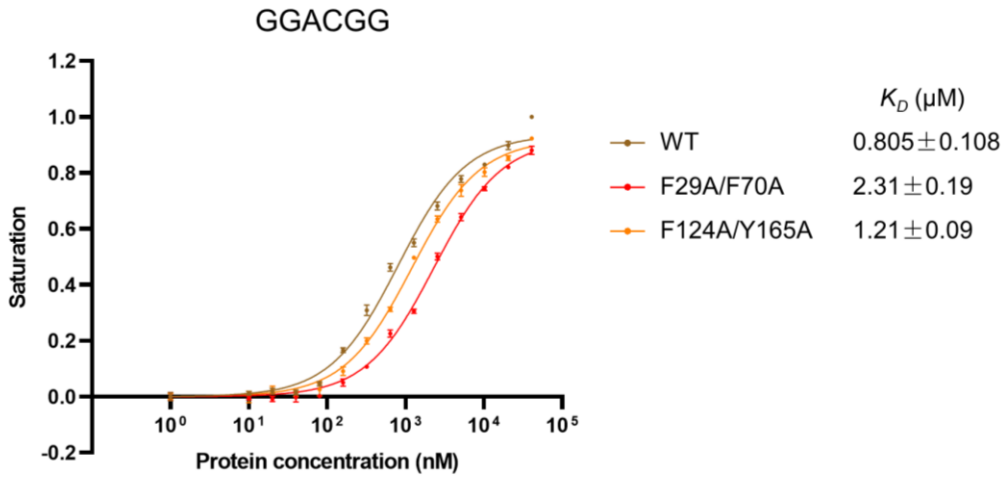

Fig. S5. The FP assays of the ssDNA-binding of RBM45^{RRM1-2} and its mutants. The K_Ds are indicated on the right. The data shown here are the averages of three independent measurements with the same protein batch. The error bars indicate the standard deviations of three replicates.

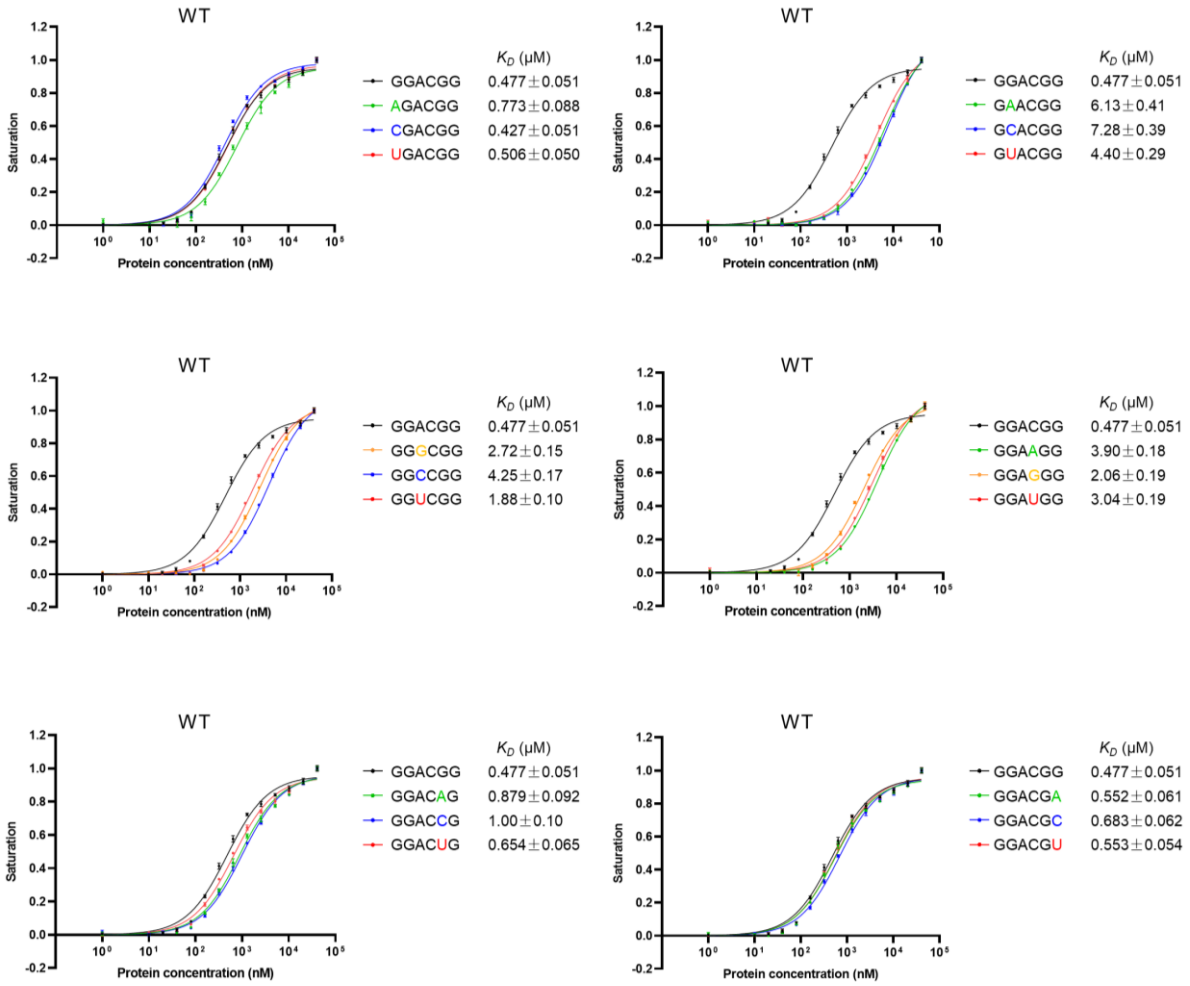

Fig. S6. The RNA-binding preference of RBM^{RRM1-2}. The FP assays of RBM^{RRM1-2} binds RNA of different sequences are shown. The RNA sequence and K_Ds are indicated on the right. For better understanding, the binding curve of GGACGG RNA is showed in each plot as a reference. The data shown here are the averages of three independent measurements with the same protein batch. The error bars indicate the standard deviations of three replicates.

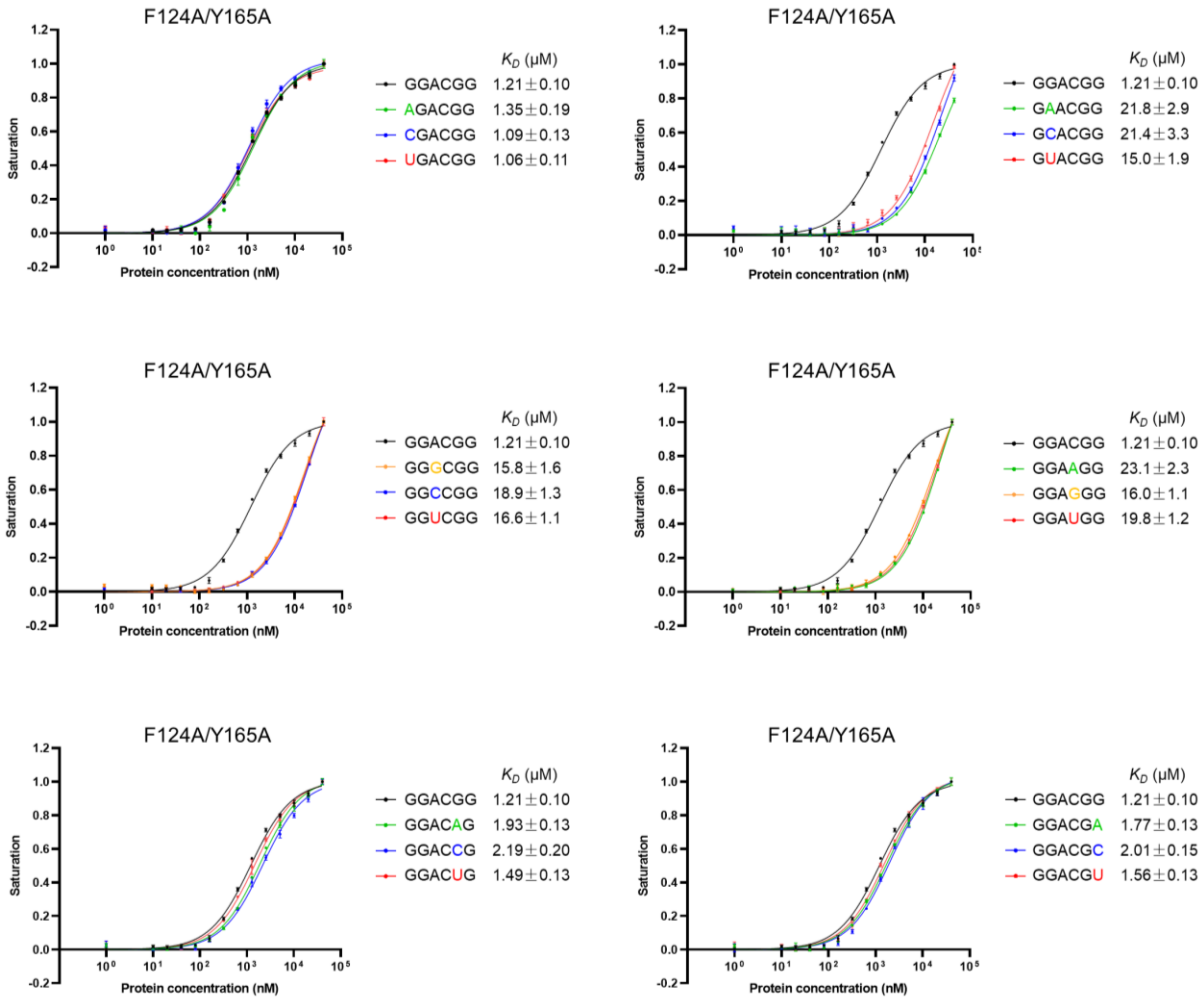

Fig. S7. The RNA-binding preference of F124A/Y165A mutant of RBM^{RRM1-2}. The FP assays of F124A/Y165A mutant of RBM^{RRM1-2} (RRM1 analog) binds RNA of different sequences are shown. The RNA sequence and K_Ds are indicated on the right. For better understanding, the binding curve of GGACGG RNA is showed in each plot as a reference. The data shown here are the averages of three independent measurements with the same protein batch. The error bars indicate the standard deviations of three replicates.

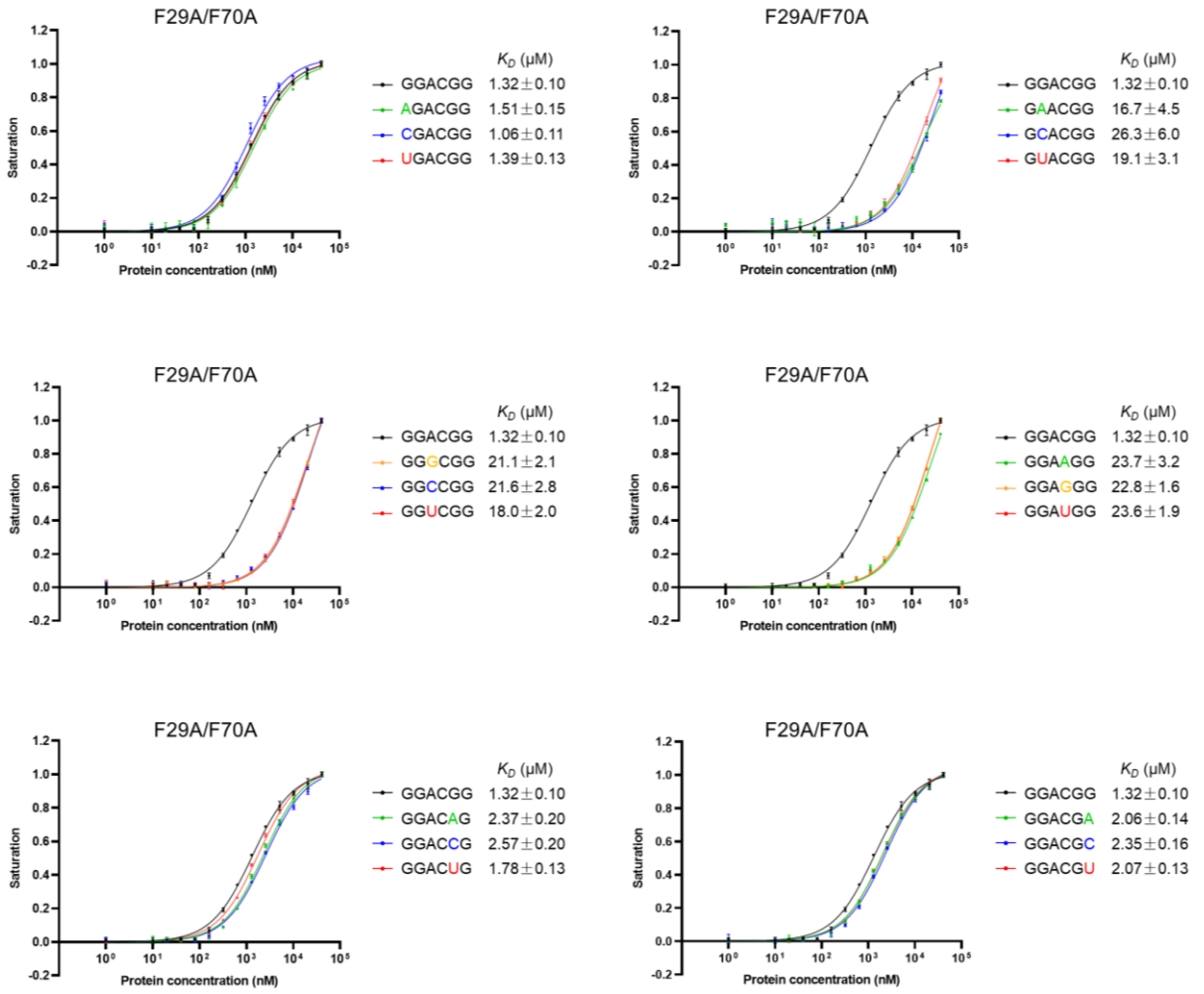

Fig. S8. The RNA-binding preference of F29A/F70A mutant of RBM^{RRM1-2}. The FP assays of F29A/F70A mutant of RBM^{RRM1-2} (RRM2 analog) binds RNA of different sequences are shown. The RNA sequence and K_Ds are indicated on the right. For better understanding, the binding curve of GGACGG RNA is showed in each plot as a reference. The data shown here are the averages of three independent measurements with the same protein batch. The error bars indicate the standard deviations of three replicates.

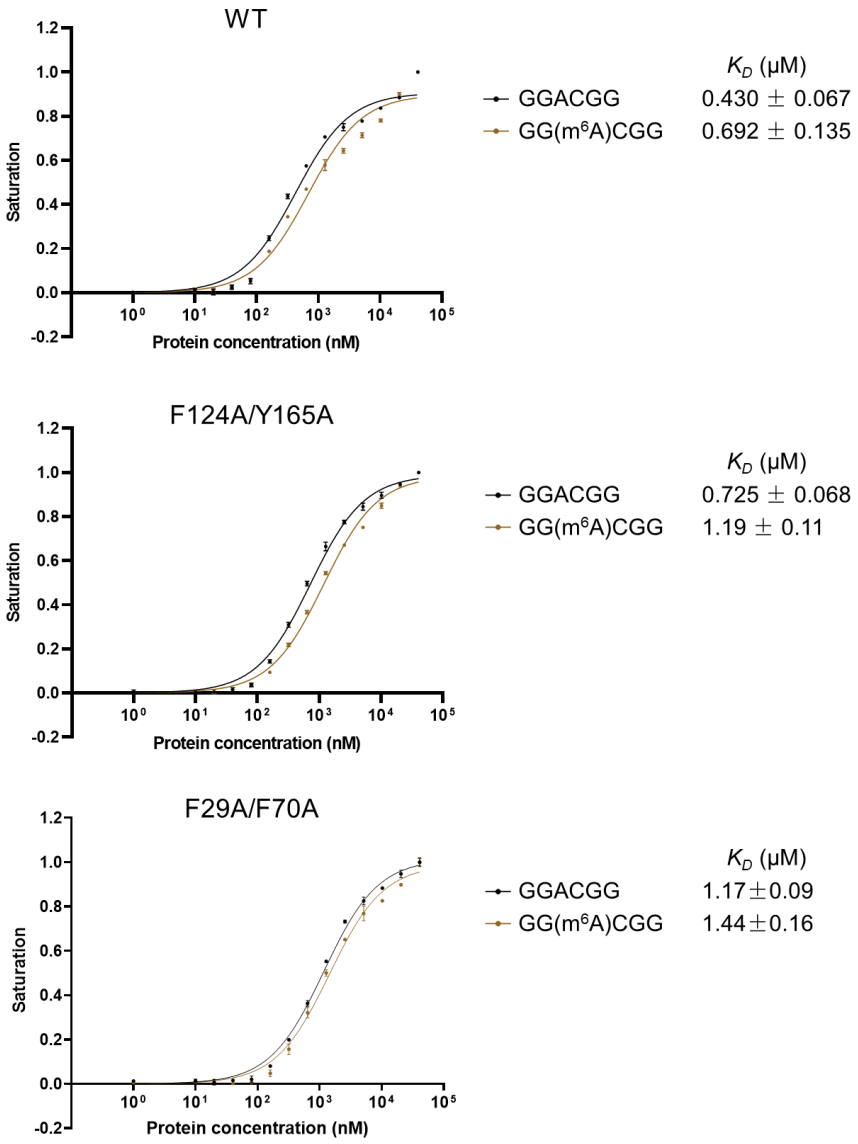

Fig. S9. The N6-methyladenosine (m6A) modified RNA-binding affinities of RBM45^{RRM1-2} and its mutants. The FP assays of WT, F124A/Y165A, and F29A/F70A of RBM45^{RRM1-2} bound GGACGG and GG(m6A)CGG are shown. The K_Ds are indicated on the right. The data shown here are the averages of three independent measurements with the same protein batch. The error bars indicate the standard deviations of three replicates.

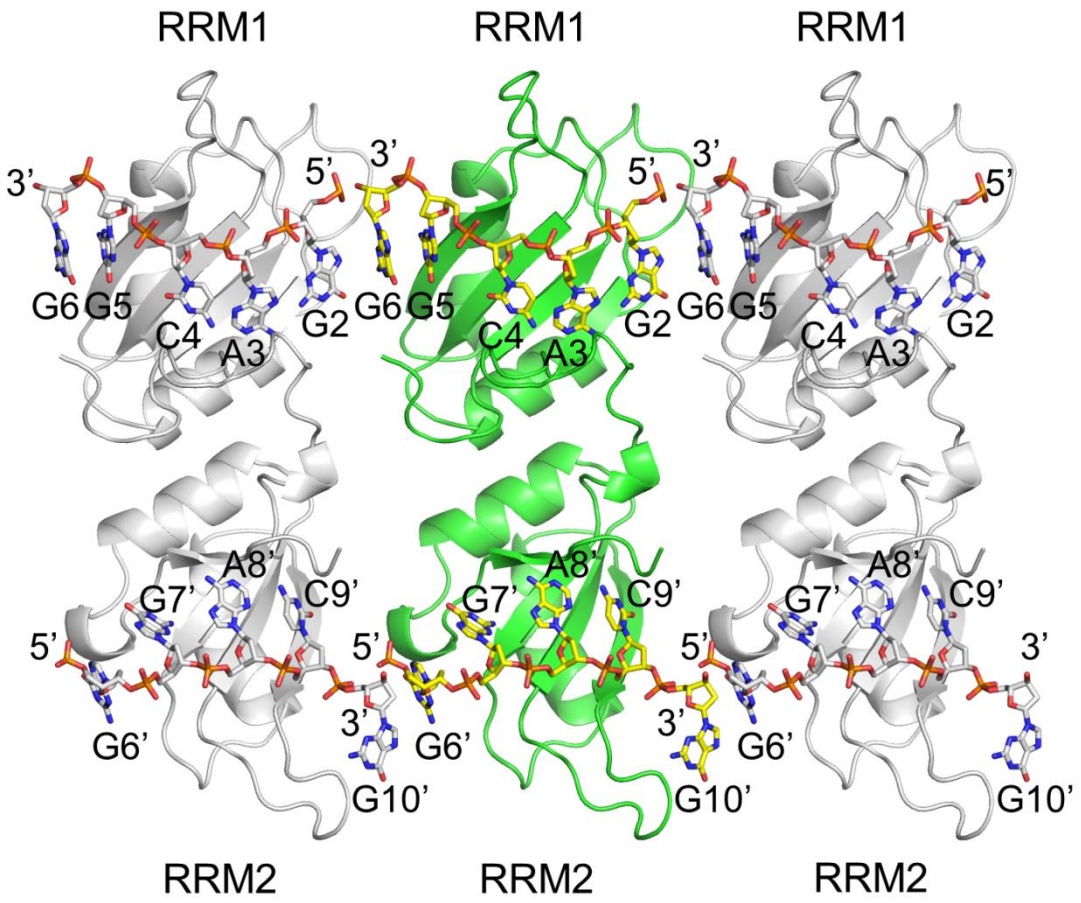

Fig. S10. The crystal packing of the RBM45^{RRM1-2}-ssDNA complex. The complex (green and yellow) in the asymmetric unit and two symmetry-related complexes (gray) are shown. The proteins and DNA are shown as cartoons and sticks, respectively.

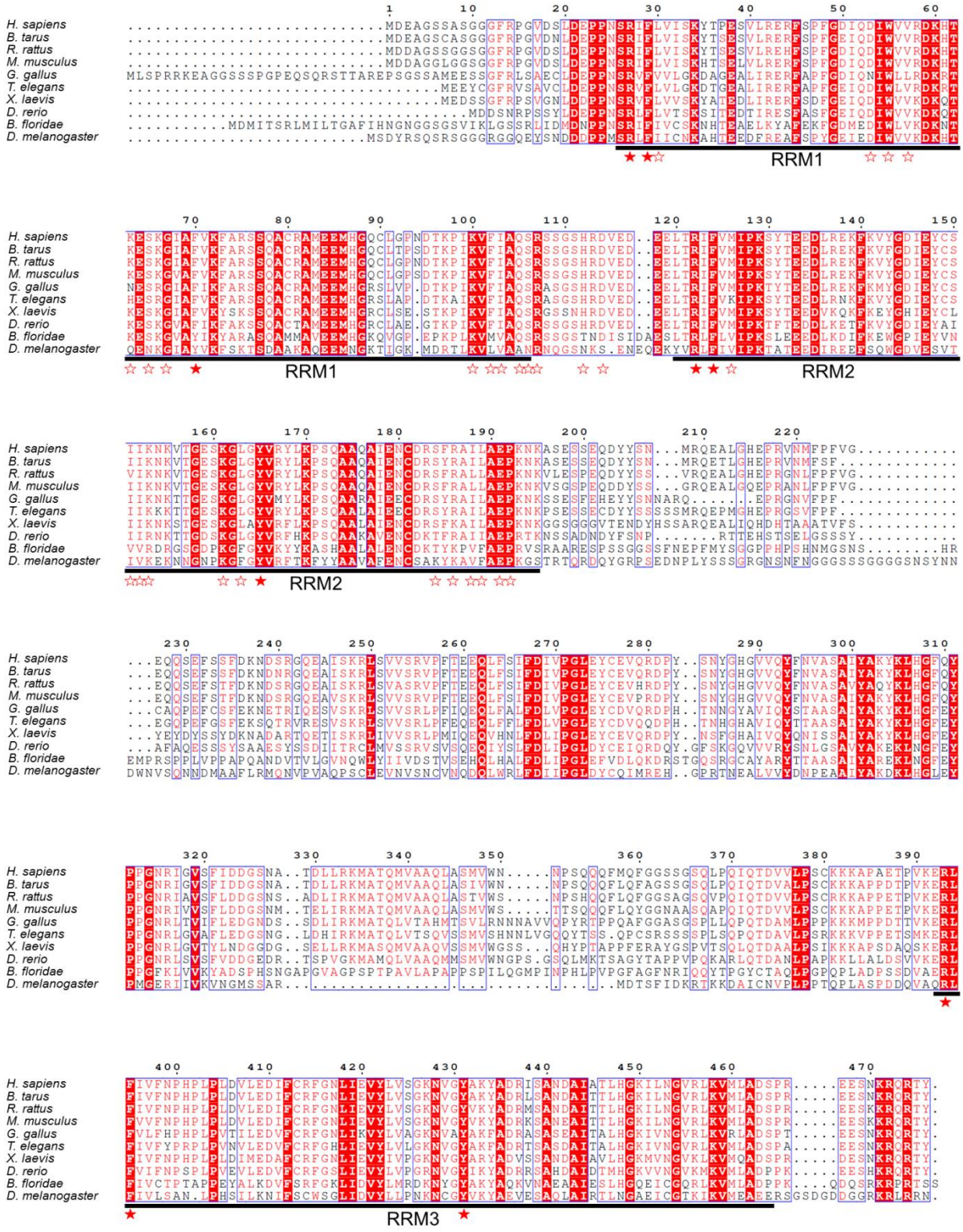

Fig. S11. Sequence alignment of RBM45s from different species. RBM45 from *Homo sapiens*, *Bos taurus*, *Rattus rattus*, *Mus musculus*, *Gallus gallus*, *Trachemys scripta elegans*, *Xenopus laevis*, *Danio rerio*, *Branchiostoma floridae*, and *Drosophila melanogaster* are aligned. The identical and conserved residues are highlighted in red background and red letters, respectively. The RRM domains are indicated by black lines at the bottom of

sequences. The conserved arginine residues and the two conserved aromatic residues of each RRM are denoted by red stars. Other residues that contact DNAs in the RBM45^{RRM1-2}-DNA structure are denoted by blank red stars.

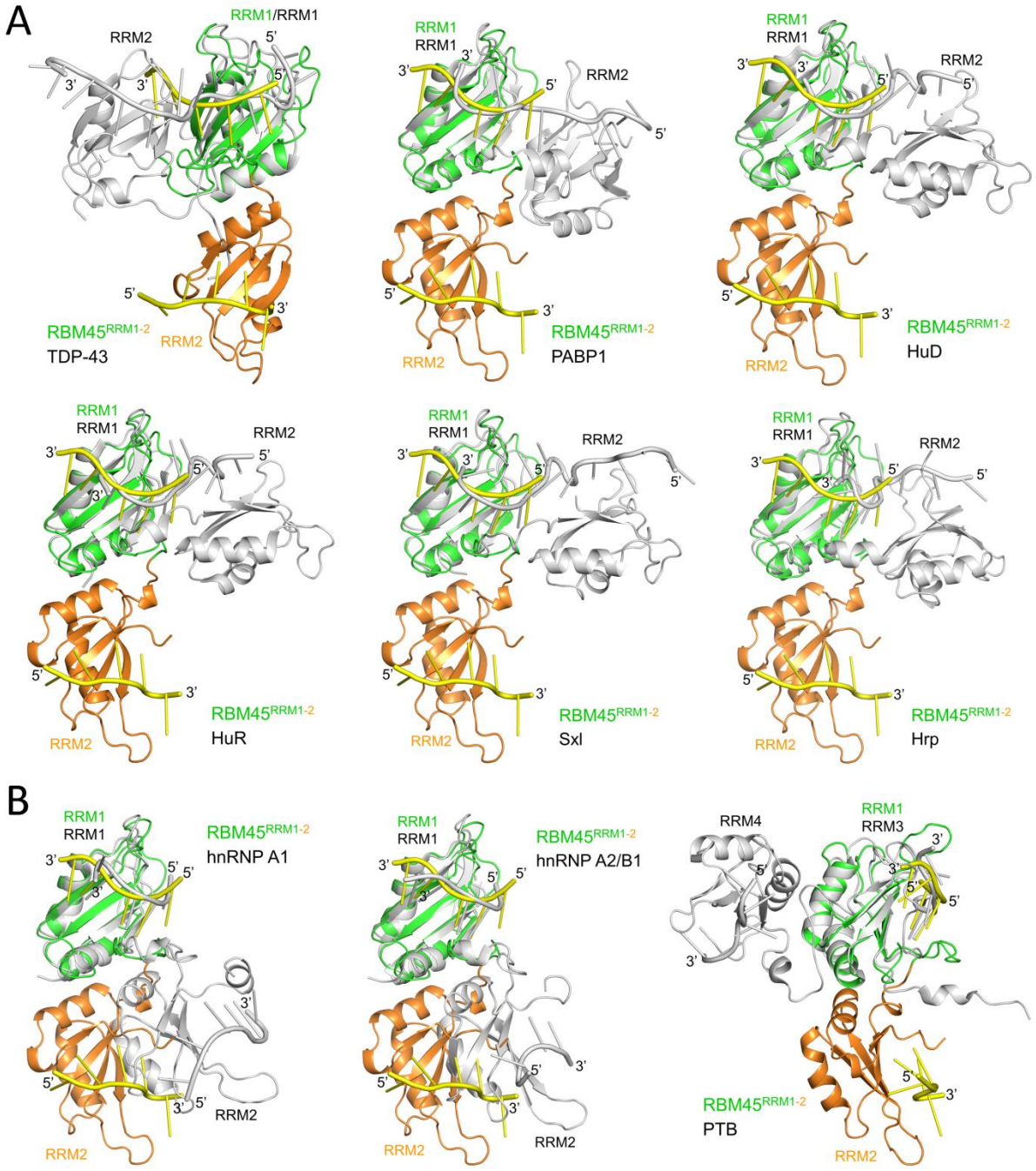

Fig. S12. Comparison of RBM45^{RRM1-2} with other tandem RRMs structures. (A) Comparison of RBM45^{RRM1-2} with tandem RRMs structures with continuous RNA-binding surfaces, including TDP-43 (PDB code: 4BS2), PABP1 (PDB code: 1CVJ), HuD (PDB code: 1FXL), HuR (PDB code: 4ED5), Sxl (PDB code: 1B7F), and Hrp1 (PDB code: 2CJK). The arrangement of RRM domains in TDP-43 significantly differs from other structures, with 5'-end of RNA bound to RRM1 and 3'-end bound to RRM2. The arrangements of RRMs in

other structures are similar, with 5'-end of RNA bound to RRM2 and 3'-end bound to RRM1. (B) Comparison of RBM45^{RRM1-2} with tandem RRMs structures with separate RNA-binding sites, including hnRNP A1 (PDB code: 2UP1), hnRNP A2/B1 (PDB code: 5HO4), and PTB1 (PDB code: 2ADC). The arrangements of RRMs in hmRNP A1 and hnRNP A2/B1 are similar, whereas that of PTB1 is significantly different. The RRM45^{RRM1-2}–ssDNA complex structure in each comparison is generated by superposing its RRM1 with the first RRM of the reference structure. The RRM1, RRM2, and ssDNA of the RRM45^{RRM1-2}–ssDNA complex are represented as green, orange, and yellow cartoons, respectively. The protein and RNA/DNA in other structures are represented as gray cartoons.

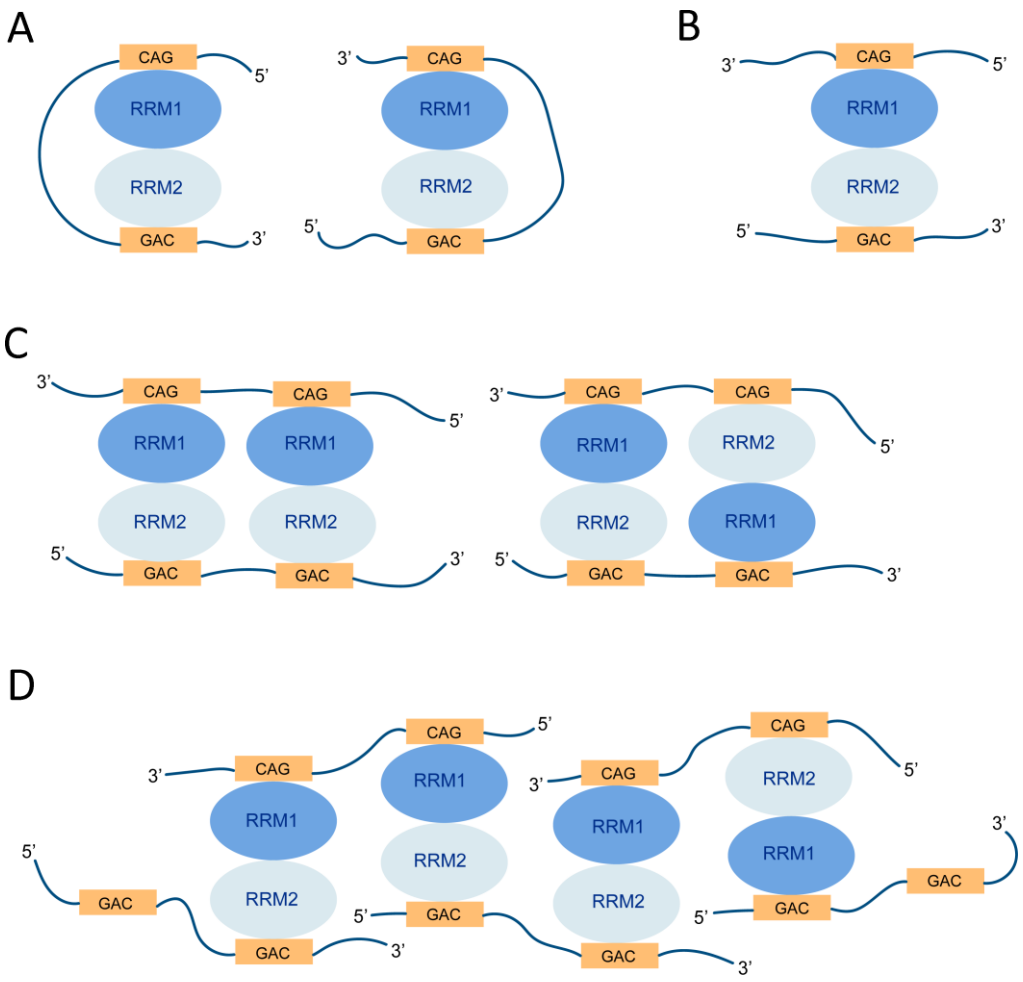

Fig. S13. The RNA-binding model of RBM45^{RRM1-2}. (A) One RBM45^{RRM1-2} binds one RNA strand with two separate GAC motifs. (B) One RBM45^{RRM1-2} binds two GAC motif-containing RNA strands. (C) Two RBM45^{RRM1-2} bind two RNA strands with two separate GAC motifs. (D) The RBM45^{RRM1-2}-RNA network. The RRM1 and RRM2 are represented as blue and slight blue ellipses, respectively. The RNAs are represented as black lines, with the GAC motif highlighted in orange boxes.