Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding

Bagdeser A. Ozdilek[a], Valery F. Thompson[b,], Nasiha S. Ahmed[b,c], Connor I. White[b], Robert T. Batey[d*] and Jacob C. Schwartz[b*]

[a] Department of Molecular, Cellular and Developmental Biology, University of Colorado at Boulder, Campus Box 347, Boulder Colorado 80309.

[b] Department of Chemistry and Biochemistry, University of Arizona Tucson, Tucson Arizona

[c] Department of Molecular and Cellular Biology, University of Arizona Tucson, Tucson, Arizona

[d] Department of Chemistry and Biochemistry, University of Colorado at Boulder, Campus Box 596, Boulder Colorado 80309-0596.

*Corresponding Co-author:

Tel: (303)-735-2159

Fax: (303)-492-8425

Email: robert.batey@colorado.edu

Tel: 940-367-8186

Email: jcschwartz@email.arizona.edu

Supplementary Table S1: Sequences of RNA substrates used in this study.

Name	Length (Nucleotides)	Sequence	Structure
DNMT	48	AUUGAGGAGCAGCAGAGAAGUUGGA GUGAAGGCAGAGAGGGGUUAAGG	N.A
Sc1	36	GCUGCGGUGUGGAAGGAGUGGCUG GGUUGCGCAGG	Stem loop[1]
dsGC	36	AUAUACGCGCGUAUAUUUCGAUAUAC GCGCGUAUAU	Hairpin
dsAU	36	AUAUAUAUAUAUAUAUUUCGAUAUAU AUAUAUAUAU	Hairpin
hRRD	152	CAUGGAUCCCUGAGGUCGGUCCCCAAUA CGACAAGACAAUUUGAUAUCAUAAUAGAA CACUGCAGAAACAAUGCUGAGUGAAGAA GAGUAGAAAUGGGAAGACUUGGUUGAGC GGAAACUGAGUUCUUGAAAAGAGGAGAU GCUUGAUGAGG	N.A[a]
mRRD	155	AACUGGCCCCUGGGAUUUGCUGCUC AGAACCUGAGUUCACUGAGACAUCAG GAGCAAGCACUGGAGGCCGGGUGCU GCUGGACCCAGAUGGGAGCCAUGCA GGACUUGACCAUGGCCUGCACACAC UUCUUCCCAGGAGAAGGGGAAUGAG GAAG	N.A
GGUG	25	UUGUAUUUUGAGCUAGUUUGGUGAU	N.A
CRL	36	AUACAACAUACAACAUACAACAUACAA CAUACAACA	Single stranded[2]
Poly-A	40	AAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAA	Single stranded[3]

N.A.[a]: Not available.

Supplementary Table S2: $K_{D,app}$ (μM) values of ZnF, RRM and RRM+3 RGG interaction with different RNA molecules.

	RNA						
	Sc1	DNMT	hRRD	mRRD	GGUG	CRL	Poly-A
Protein							
ZnF	n.d[a]	n.d	n.d	n.d	n.d	n.d	n.d
RRM	48 ± 3	n.d	n.d	n.d	n.d	n.d	45 ± 2
RRM+3 RGG	27 ± 2	4 ± 0.5	16 ± 3	10 ± 1	21 ± 1	63 ± 8	40 ± 5

[a]n.d. : not detectable (>100 μM)

Supplementary Table S3: Mutated arginine residues in SGG mutants.

SGG1	R213S, R216S, R218S, R234S, R242S, R244S, R248S, R251S, R259S
SGG2	R377S, R383S, R386S, R388S, R394S, R407S, R422S
SGG3	R472S, R473S, R476S, R481S, R485S, R487S, R491S, R495S, R498S, R503S
SGG4	R213S, R216S, R218S, R234S, R242S, R244S, R248S, R251S, R259S, R377S, R383S, R386S, R388S, R394S, R407S, R422S, R472S, R473S, R476S, R481S, R485S, R487S, R491S, R495S, R498S, R503S

Supplementary Table S4: Corresponding $K_{D,app}$ (μM) values of heat-map data in Figure 5.

	RNA								
	Sc1	DMNT	hRRD	mRRD	dsGC	dsAU	GGUG	CRL	poly-A
Protein									
FMRP-RGG	0.09 ± 0.02	2.5 ± 0.4	0.30 ± 0.01	0.35 ± 0.03	8.0 ± 0.1	8.3 ±0.5	27 ± 3	27 ± 1	50±1
hnRNPU-RGG	0.25 ± 0.01	0.5 ± 0.1	0.30 ± 0.02	0.5 ± 0.1	15 ± 1	43 ± 20	8.5 ± 1.0	8.0 ± 0.5	9.0±0.5
FUS-RGG1	2.8 ± 0.1	3.0 ± 0.1	1.6 ± 0.1	1.3 ± 0.1	22 ± 1	29 ± 1	16 ± 1	17 ± 1	25±5
FUS-RGG2	25 ± 1	60 ± 15	14 ± 1	26 ± 4	n.d.[a]	n.d.	n.d.	n.d.	n.d.
FUS-RGG3	3.7 ± 0.2	8.0 ± 0.5	4.5 ± 1.0	4.5 ± 0.5	100 ± 30	n.d.	50±3	65±6	110±30
FUS-LC-RGG1	2.8 ± 0.5	6.2 ± 0.3	3.0 ± 0.3	1.0 ± 0.3	45 ± 2	72 ± 5	12 ± 1	19 ± 1	5 ± 2
FUS-RRM-RGG2	2.8 ± 0.2	2.5 ± 0.1	1.2 ± 0.2	1.8 ± 0.1	15 ± 1	17 ± 2	11 ± 1	9.5 ± 0.5	17 ± 3
FUS (wt)	0.30 ± 0.02	0.7 ± 0.2	0.60 ± 0.05	0.33 ± 0.07	7.3 ± 0.2	8.4 ± 0.5	3.0 ± 0.1	10 ± 1	3.2 ± 0.2

[a]n.d. : not detectable (>100 μM).

Supplementary Table S5: Effects of mutations in Sc1 RNA on RNA binding activity of RGG domains.

	$K_{D,relative}$ ($K_{D,app,\ mutant}$ / $K_{D,app,\ Sc1}$)	
Protein	Mutant C5U-G31A	Mutant G7A-C30U
FMRP-RGG	28	24
hnRNPU-RGG	11	10
FUS-RGG1	4	2
FUS-RGG2	1.5	1.5
FUS-RGG3	4.5	3.8
FUS (WT)	5.5	5

Figure Legends

Supplementary Figure S1. SDS-PAGE analysis of protein expression and purification. Representative SDS-PAGE image of FUS protein purification. FUS was expressed and purified as described in methods. M: protein marker (Invitrogen), lane 1: non-induced (BL21 (DE3) Rosetta Plys) cells, lane 2: IPTG-Induced cells, lane 3: supernatant, lane 4: flow through, lane 5, 6 and 7 are washes of the Ni-NTA with lysis buffer, lane 8: wash of the Ni-NTA with 100 mM imidazole in lysis buffer, lane 9: Elution of FUS protein with 250 mM imidazole in lysis buffer, lane 10: purified FUS protein by size exclusion chromatography.

Supplementary Figure S2. RGG/RG domains and sequence are conserved across diverse metazoan species. The amino acid sequences of RGG domains of FUS, FMRP and hnRNPU from human, mouse, xenopus and zebrafish were aligned. Pink and yellow colors show conserved RGG and RG repeats, respectively. "*" symbol shows the arginine amino acids of FMRP-RGG that directly make hydrogen bonds with guanines in Sc1 RNA.

Supplementary Figure S3. RRM and ZnF domains of FUS correctly folded, but do not bind to RNA. (a) CD spectra of RRM and ZnF domains (no MBP tag) of FUS. The values are 5% alpha-helix, 23% β-strand for ZnF and 28% alpha-helix, 25% β-strand for RRM. Representative binding isotherms and non-linear curve fitting for the titrations of 800 μM DNMT RNA in 200 μM (b) ZnF or (c) RRM and for the titrations of 400 μM DNMT RNA in 100 μM RGG1-RRM-RGG2 and RRM-RGG2. (Top panels) Raw heats of binding obtained by Isothermal titration calorimetry (ITC) when proteins mixed with DNMT RNA. (Bottom panels) Binding isotherms fitted to the raw data using single-state

binding model. Titrations were performed in 150 mM KCl, 50 mM tris buffer. RGG1-RRM-RGG2 binds to DNMT RNA with 3 ± 0.4 μM affinity and RRM-RGG2 shows a 5 ± 0.2 μM binding affinity.

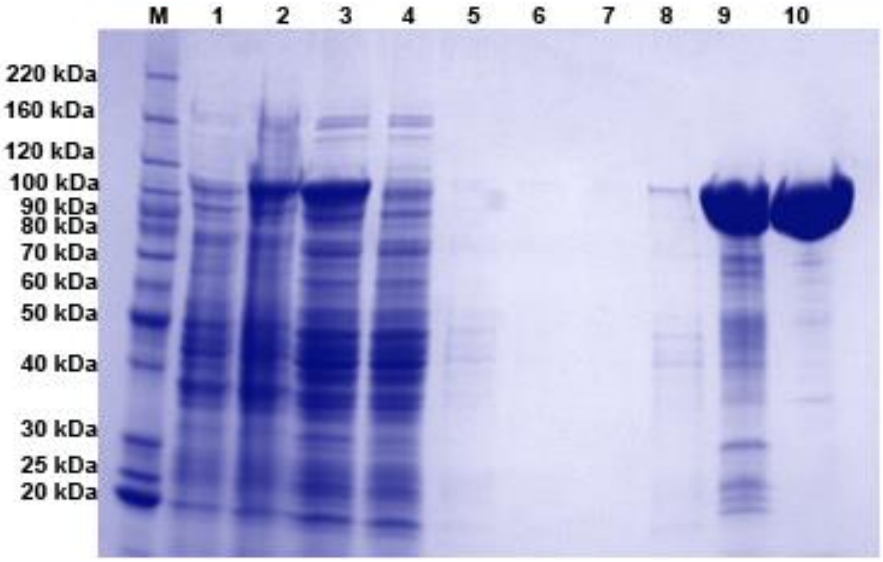
Supplementary Figure S4. Flanking RGG domains impart the RNA binding activity of the ZnF domain. (a) Representative EMSAs and the corresponding binding curves showing binding of DNMT RNA to RGG2-ZnF (372-453), ZnF-RGG3 (423-501) and RGG2-ZnF-RGG3 (372-501). *b* = bound DNMT RNA and *f* = free DNMT RNA. '(-)' shows no protein lane. Error bars represent the S.D. of three independent titrations for each construct.

Supplementary Figure S5. Interaction between RGG1-RRM-RGG2 and DNMT RNA shows a salt dependence. ITC titrations were performed for the interaction of RGG1-RRM-RGG2 and DNMT RNA in different KCl concentrations (50 mM, 100 mM, 150 mM, 200 mM, 250 mM, 300 mM). The graph shows linear correlation between log (Ka) and log (KCl concentration in M) for three independent repeats. Values of each repeat are shown with black circles. Red line indicates mean of the three repeats.

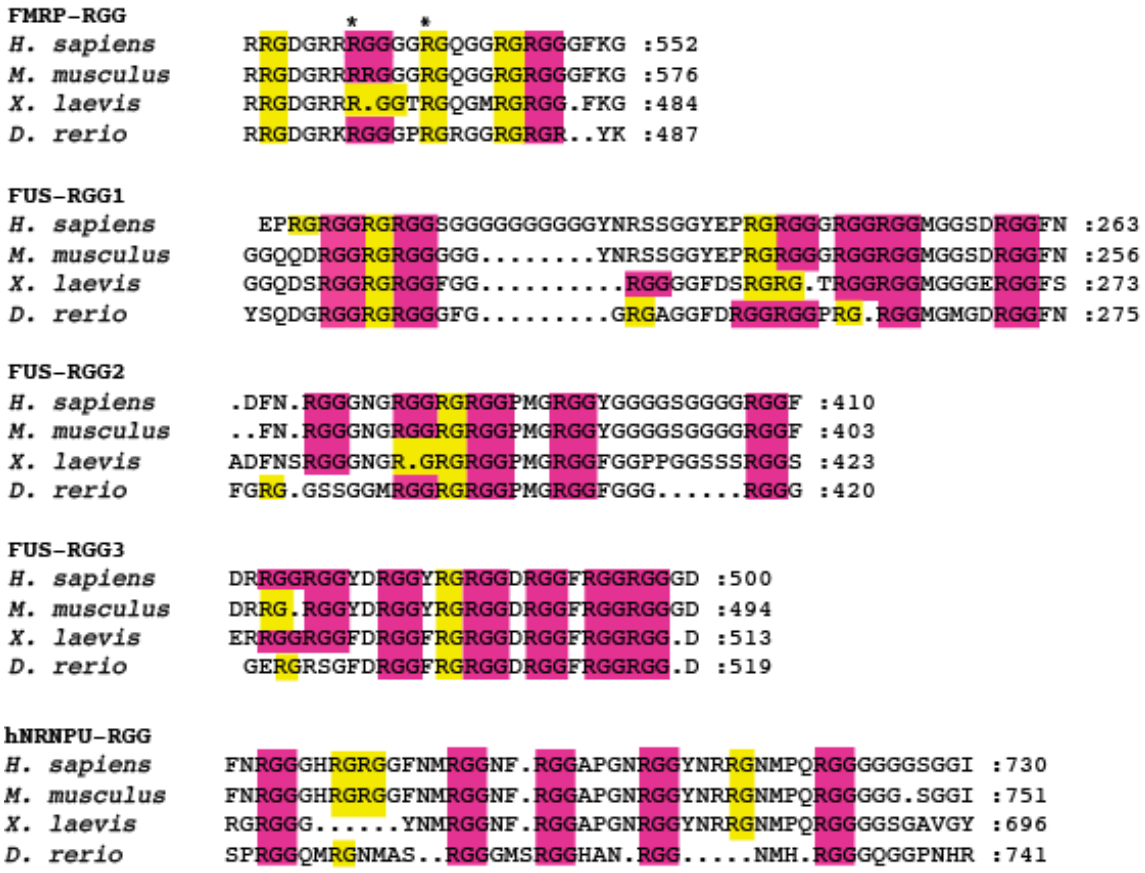
Supplementary Figure S6. G-quartets formation in Sc1, DNMT and GGUG RNAs. (Top) Representative gels with RNAseT1 digestion of (a) Sc1, (b) DNMT and (c) GGUG RNAs in KCl and LiCl buffers. (Bottom) Bands were quantified by Image Quant and the amount of protection for each band (fold change in band intensity, KCl/LiCl) was calculated. The Guanines involved in G-quadruplex formation in Sc1 RNA were shown with '*' symbol. Lane 1 is RNA alone (No RNase T1), lane 2 shows alkaline hydrolysis of corresponding RNA, lane 3 and lane 4 is RNAse T1 reaction in LiCl and KCl, respectively.

Supplementary Figure S7. Effects of G-quartet formation on RNA binding activity of different RGG domains. Binding of RGG domains to Sc1 RNA was measured in the presence of KCl and LiCl by EMSA. Representative binding curves of FUS and FMRP-RGG binding to Sc1 RNA was shown and fold change in the dissociation constant of RGG domains for LiCl and KCl buffers were presented. Error bars represent the S.D. of at least two independent titrations for each construct.

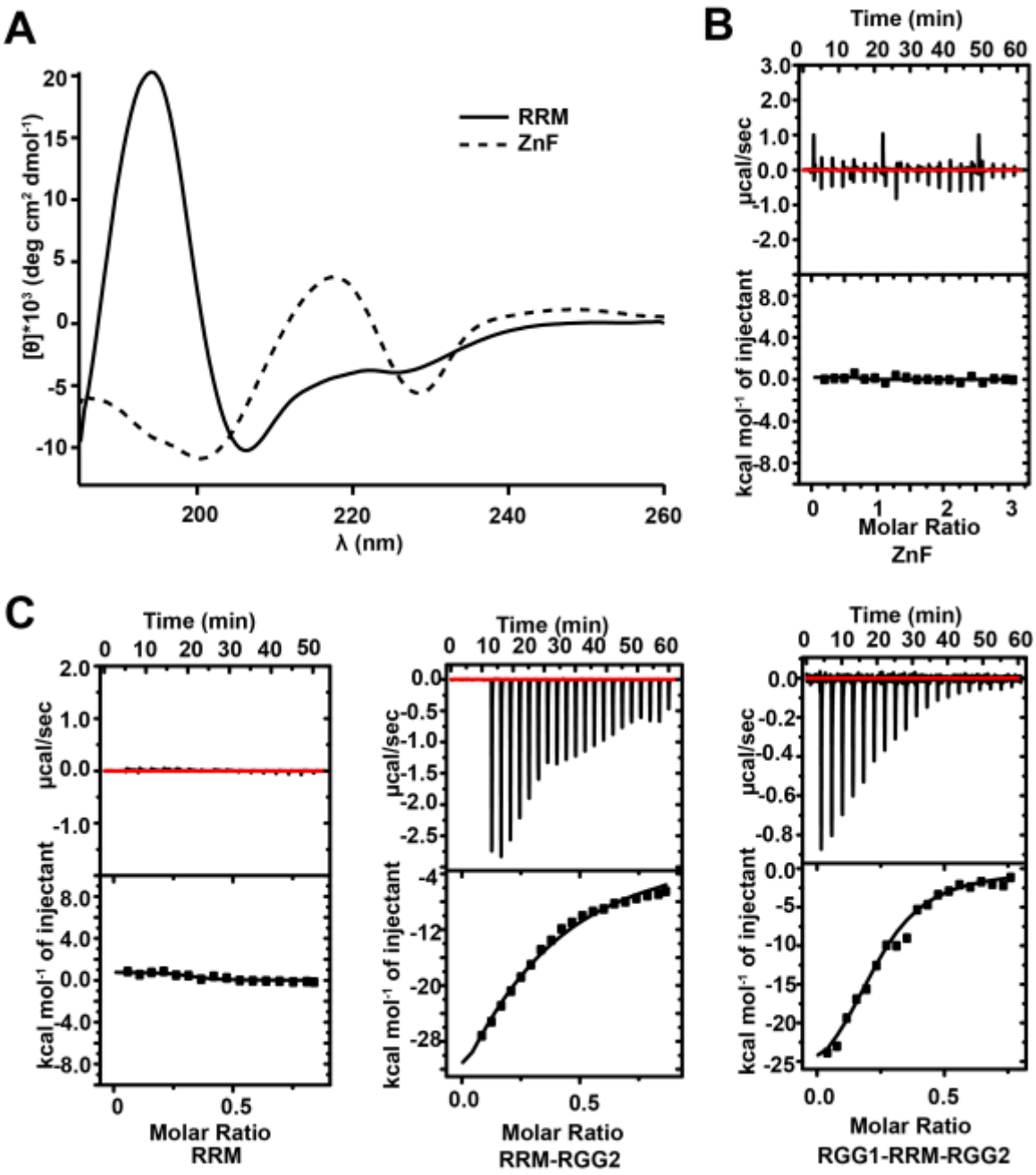
Supplementary Figure S1



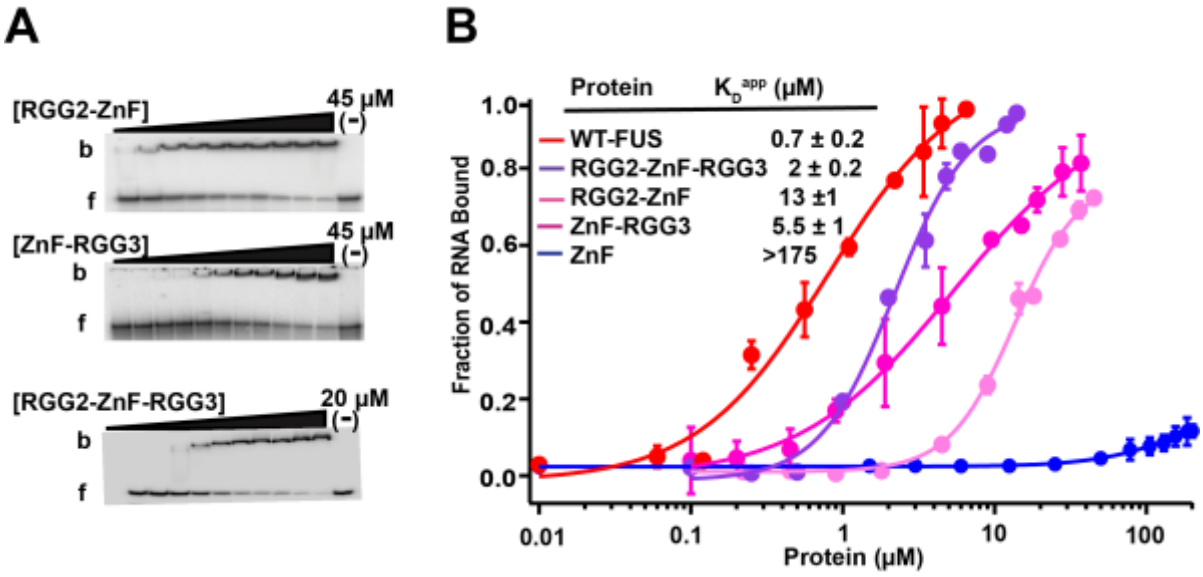
Supplementary Figure S2



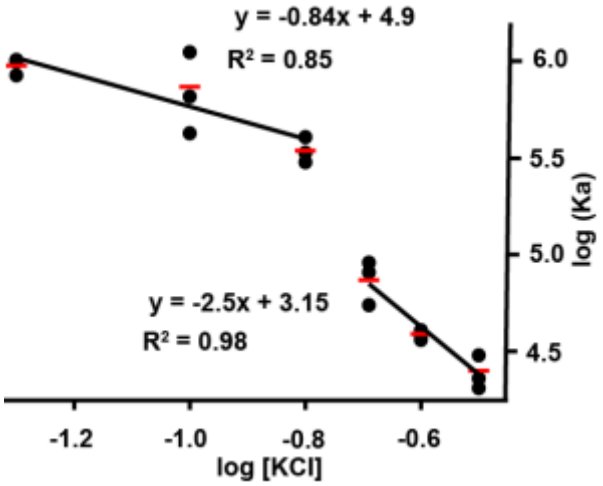
Supplementary Figure S3



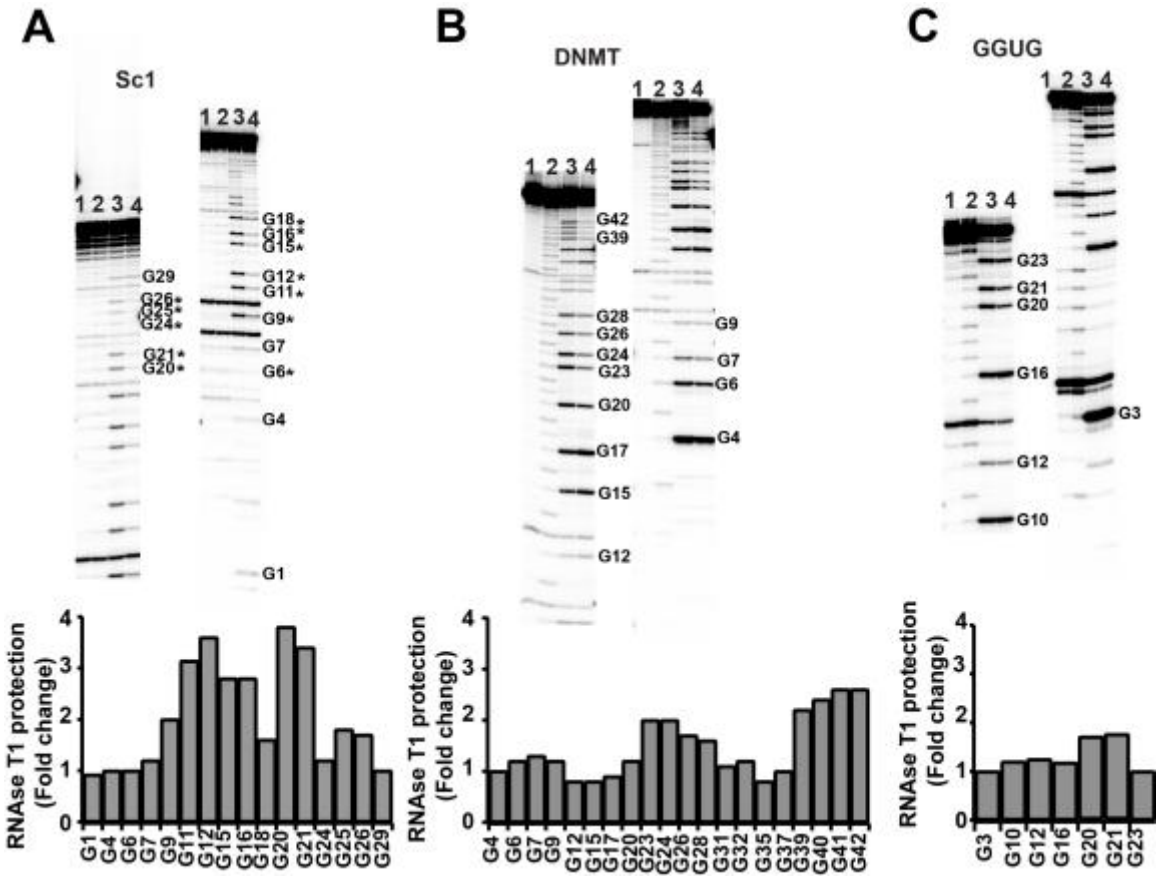
Supplementary Figure S4



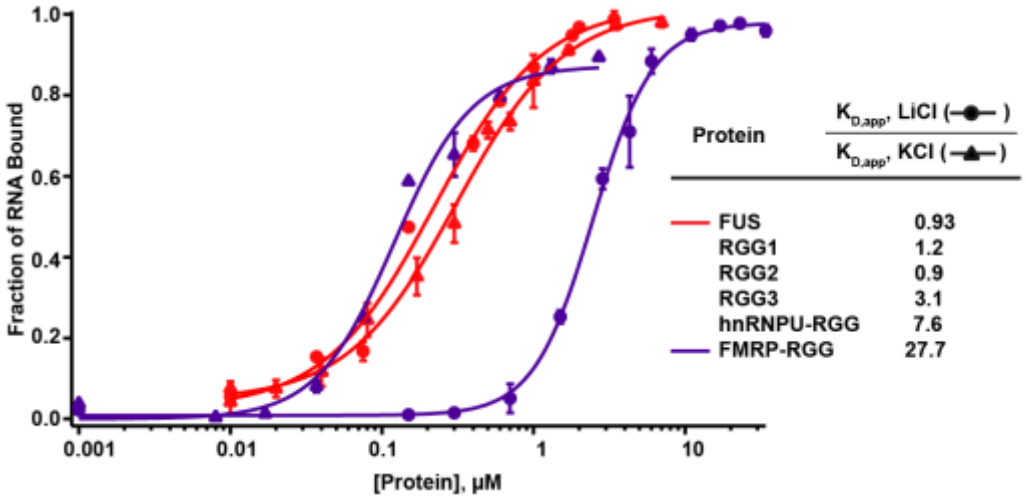
Supplementary Figure S5



Supplementary Figure S6



Supplementary Figure S7



REFERENCES

1. Vasilyev, N. et al. Crystal structure reveals specific recognition of a G-quadruplex RNA by a beta-turn in the RGG motif of FMRP. *Proc Natl Acad Sci U S A* **112**, E5391-400 (2015).
2. Polaski, J.T., Holmstrom, E.D., Nesbitt, D.J. & Batey, R.T. Mechanistic Insights into Cofactor-Dependent Coupling of RNA Folding and mRNA Transcription/Translation by a Cobalamin Riboswitch. *Cell Rep* **15**, 1100-10 (2016).
3. Seol, Y., Skinner, G.M., Visscher, K., Buhot, A. & Halperin, A. Stretching of homopolymeric RNA reveals single-stranded helices and base-stacking. *Phys Rev Lett* **98**, 158103 (2007).