Supplemental Information

Rbfox proteins regulate microRNA biogenesis by sequence specific binding to their precursors and target downstream Dicer

Yu Chen, Lorena Zubovic, Fan Yang, Katherine Godin, Tom Pavelitz,
Javier Castellanos, Paolo Macchi, Gabriele Varani

Paramagnetic spin-labeling of RNA and PRE data collection

A previously published protocol[1] was followed to attach the 3-(2-iodoacetamido)-proxyl spin label to a chemically synthesized 46 nucleotide pre-miR-20b containing a single 4-thio Uracyl 43 base (purchased from Dharmacon). This RNA is shorter than the completepre-miR-20b but the lower part of the structure is too distant from the loop to be contacted by the protein. Complete coupling was achieved by overnight reaction at room temperature, as monitored by complete shift of the maximum in UV absorption to ~320 nm. The protein-RNA complex were prepared by adding pre-dyalized paramagnetic spin-labeled 46 nucleotide pre-miR-20b into ^{15}N-labeled Rbfox RRM (amino acids 109-225 of Rbfox1) at a concentration of 0.2 mM in NMR buffer. ^{1}H-^{15}N HSQC spectra were collected at 298 K on Bruker Avance 600 before and after adding 10 mM sodium hydrosulfite for reduction of the paramagnetic label.

NMR resonance assignments

For resonance assignments of the free RNA, 2D NOESY spectra (mixing time t_m of 100 ms and 300 ms), 2D TOCSY (t_m of 80 ms), ^{13}C CT-HSQC and 3D ^{13}C NOESY-HSQC (t_m of 300 ms) were collected on unlabeled or ^{13}C, ^{15}N labeled RNA in D_2O at 298 K to assign the non-exchangeable protons and their attached carbons. 2D NOESY spectra (mixing times t_m = 100 ms and 300 ms) were collected on unlabeled RNA in 90% H_2O, 10% D_2O at 279 K in order to assign the exchangeable protons. A ^{15}N HSQC was collected on ^{15}N labeled RNA to confirm the assignments of exchangeable protons and assign their attached nitrogens. Nearly complete assignments for H8/H6, H2, H5 and H1'-H5'/H5'' were obtained for the free miR20b. For the resonance assignments of the protein in complex with the RNA, 2D ^{15}N/^{13}C HSQCs, 3D HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO, HBHA(CO)NH and HCCH-TOCSY spectra were collected on ^{15}N, ^{13}C labeled protein in complex with unlabeled RNA to assign the backbone and the non-aromatic side-chains. Aromatic side chains were assigned using 2D ^{13}C HSQC and 3D ^{13}C NOESY-HSQC (t_m of 150 ms). A 3D ^{15}N NOESY-HSQC (t_m of 150 ms) spectrum was also used to verify the backbone connectivity. Nearly all the assignments for backbone atoms were obtained except for the three N-terminal residues that are flexible and more than 95% of the assignments for side chain atoms were completed as well. For the assignments of the RNA in complex with the protein, 2D F1-, F2-filtered NOESYs (t_m of 100 ms and 300 ms) and 2D TOCSY (t_m of 80 ms) were collected on unlabeled RNA in complex with ^{13}C, ^{15}N labeled protein in D_2O at 298 K to assign H8/H6, H5 and H1'. In addition, ^{15}N HSQC, ^{13}C CT-HSQC and 3D ^{13}C NOESY-HSQC (t_m of 300 ms) were collected on ^{13}C, ^{15}N labeled RNA in complex with unlabeled protein to assign nitrogens, carbons and the other sugar protons. Nearly complete assignments of were done for the base protons, but only 75% of assignments were obtained for sugar protons with those of A_{22}, C_{35}, C_{37} and U_{38} especially difficult to assign.

Structure determination

For the protein-free RNA, manually assigned NOE distance restraints derived from 3D ^{13}C NOESY-HSQC and 2D NOESYs at 100 ms of mixing time were separated into three ranges based on the cross-peak intensities, strong (1.8 Å -3.5 Å), medium (1.8 Å-4.5 Å) and weak (1.8 Å-5.5 Å). Additional NOEs observed only from 2D NOESY with 300 ms of mixing time were assigned as very weak (1.8 Å-6.5 Å). Hydrogen bonds restraints for the base pairs were added based on strong NOE cross peaks to the imino protons in 2D water NOESY. Dihedral angle restraints for the conformation of sugar rings (C2'-endo or C3'-endo) were also added based on H1'-H2' cross-peak intensities in 2D TOCSY spectra. With all of the restraints above, 500 initial structures were generated in CYANA[2], and the 100 structures with the lowest target function were further refined in implicit solvent using the SANDER module of Amber 8.0 and the ff99 force field[3]. The script for the restrained simulated annealing protocol was modified from Tolbert *et al*.[4] The NOE distance restraints for the complex could be divided into two parts, 'real' NOE derived experimentally and 'virtual' NOE predicted from the free RNA based on chemical shift similarities. The 'real' NOE list was obtained from the combination of three NOE lists: intra-protein NOEs automatically assigned from ^{15}N and ^{13}C NOESY-HSQCs obtained by CYANA; intra-RNA NOEs manually assigned from 2D F1-, F2-filtered NOESYs and 2D TOCSY spectra; and intermolecular NOEs manually assigned from 2D F1-filtered, F2-edited NOESY and 3D ^{13}C F1-filtered, F3-edited NOESY-HSQC. 'Virtual' NOEs include only intra-RNA restraints for the lower stem region that retain the same conformation of free RNA, as judged from small chemical shift changes. Hydrogen bonds and dihedral angle restraints for the RNA in complex with protein were obtained the same way as for the free RNA. Protein torsion angles were obtained by TALOS+[5]. Structure calculation for the complex was carried out the same way as for the free RNA, except that the system was heated to 1500 K instead of 600 K during Amber simulated annealing refinement.

NMR structural analysis of free miR-20b and its complex with Rbfox-RRM

Pre-miR-20b adopts a rigid terminal loop structure - Five Watson-Crick base pairs $G_{19}C_{41}$, $G_{20}C_{40}$, $U_{21}A_{39}$, $A_{22}U_{38}$ and $G_{23}C_{37}$ were initially identified through the observation of slow exchanging NH's and strong NOE cross peaks across the strand in imino-proton spectra, forming a continuous A-form helical stem. Stacking above these base pairs, we observe a non-canonical U_{24}-U_{36} base pair, identified because the two uridines both connect to G_{23} and the imino protons of U_{24} and U_{36} show strong NOE interactions with each other and slow solvent exchange, as expected for a base pair. U_{25} and C_{35} were confirmed to stack above U_{24} and U_{36}, respectively, as supported by strong sequential sugar to base and moderate base to base NOEs. The two bases can also form a base pair with a single hydrogen bond between U-O4 and C-NH4 stabilized at a pH below 5.0. At the very top of the stem, U_{26}-A_{34} and the wobble pair U_{27}-G_{33} were assigned from the left-over imino resonances in the 2D water NOESY, which are weak due to exposure to solvent but clearly observed. Because of the stacking of nine continuous base pairs (including U_{25}-C_{25} at pH 5.0), the A-form helix is highly regular, whereas the conformation of apical loop residues is less regular. The G_{28} base continues stacking on top of U_{27}, while the base of C_{30} faces inwards towards the minor groove, while the bases of G_{29}, A_{31} and U_{32} are splayed out with their Watson-Crick face pointing outwards to the solvent. Numerous moderate to weak NOEs between sugars of G_{29} and C_{30} were observed, confirming their sugar rings stacking above each other, as shown in Figure 3C. In most of the structures, the sugar ring stacking conformation can be extended to A_{31} and U_{32}, forming a backbone 'trunk' from which the bases branch out into the solvent.

The terminal loop of pre-miR-20b opens up upon binding of the Rbfox RRM - The complex of pre-miR-20b with Rbfox RRM was determined from a very large set of intermolecular NOE restraints (197, a large number for a complex of this size). Combined with the predicted restraints for the unperturbed A-form helix, based on chemical shift similarity with free pre-miR-20b, we successfully determined the structure of the complex. The RMS deviation of the heavy atoms considering the ordered parts of both RNA and protein (G_{29}-A_{34} and Pro116-Arg194) is 0.90 Å. Overall, in the structure of the complex, the Rbfox RRM adopts the conventional β1α1β2β3α2β4 fold with an additional two-stranded β sheet between α2 and β4, as described for its complex with ssRNA 5'-UGCAUGU-3'. The RMS deviation between Cα traces of Rbfox RRM in both complexes is 1.13 Å, showing great similarity.

The loop residues of pre-miR-20b are wrapped around the protein β2β3 loop in a manner reminiscent of the U1A complex, with U_{32} and G_{33} lying across the β sheet of the RRM. Most of the intermolecular NOE interactions come from G_{28}-A_{34} of pre-miR-20b and are similar to the complex with single stranded RNA, but they also differ in several ways. Because the most significant differences are found near the helical region, these distinctions are very likely to be real and reflect different structural context provided by the stem-loop. Instead of hydrogen bonding to the side chain of R184, the base of G_{29} favors stacking on it, as seen in 70% of the converged structures, making G_{29} the center layer of a sandwich with F126 and R184. C_{30} is not well restrained due to the absence of its hydrogen bond with G_{28}, as observed instead in the Fox-1-UGCAUGU complex. A_{31} and U_{32} are found in the same conformation as in the Fox-1-UGCAUGU complex, with similar intra- and intermolecular interactions. In contrast, even though G_{33} maintains the stacking with F160 and adopts the same *syn* conformation, intra-RNA hydrogen bond G_6-2'OH to U_7-O3' are not found and its hydrogen bond to the side chain of R118 is too far to form in most of the conformers. Moreover, the two immediately flanking nucleotides of the consensus sequence $G_{29}C_{30}A_{31}U_{32}G_{33}$, G_{28} and A_{34}, demonstrate the greatest differences in binding. In the complex of Rbfox1 RRM with single-stranded RNA ($U_1G_2C_3A_4U_5G_6U_7$)[6], U_1, equivalent to G_{28}, is orientated either parallel or perpendicular to F126 in different converged structures. In our complex, the conformation of G_{28} cannot be established due to very few observed NOE constraints. In the RbFox1 RRM-UGCAUGU complex, the last residue (U_7) is outside of the protein β-sheet binding surface and not constrained. In contrast, A_{34} in the current structure has its base ring hanging above the β2β3 loop. Numerous intermolecular NOEs connect A_{34} to the Rbfox RRM, but no intermolecular hydrogen bonds are confidently detected. The remaining intermolecular contacts occur mainly between the phosphate backbone of U_{25}-U_{27} of pre-miR-20b and the β2β3 loop of Rbfox RRM.

Supplemental Table S1. List of human precursor miRNAs containing the Rbfox RRM target sequence GCAUG, as deduced by analysis of miRBase. TSL, refers to miRNAs with Rbfox binding sites found within the terminal stem loop sequences at or above the Dicer cleavage sites; Upper stem, binding sites are found within regions of mature miRNA sequences between the Dicer and Drosha cleavage sites; Lower stem, binding sites below Drosha cleavage sites (**Supplemental Fig. 1A)**.

TSL	Upper Stem	Lower Stem[d]
miR-20b, **miR-107**, miR-134, miR-486-2, miR-507, miR-767[a], miR-1178, miR-1236, miR-1265, miR-3162, miR-3175, miR-3180, miR-3677, miR-4441, miR-4504, miR-5002, miR-5192, miR-5695, miR-6505*, miR-6762, miR-6826, miR-7515, miR-7156, miR-7515, miR-8066, miR-8085,	miR-23b, miR-32*, miR-33a, miR-105, miR-152, miR-188, miR-202, miR-297[b], miR-346, miR-378i, miR-450a-1, miR-466, miR-548q, miR-596, miR-619, miR-640, miR-668, miR-1226, miR-1256, miR-1269a, miR-1282, miR-1912, miR-2117, miR-3174, miR-3607, miR-3622b, miR-3651, miR-3680, miR-3976, miR-3978, miR-4253, miR-4260, miR-4267, miR-4294, miR-4304, miR-4323, miR-4327, miR-4419b, miR-4447, miR-4464, miR-4524a, miR-4525, miR-4540, miR-4666b, miR-4704, miR-4728, miR-4743, miR-4752, miR-4761, miR-4799[c], miR-5011, miR-5087, miR-5706, miR-6508, miR-6722*, miR-6771, miR-6775, miR-6797, miR-6799, miR-6806, miR-6810, miR-6825*, miR-6862, miR-8052, miR-8074	miR-138-1, miR-197*, miR-205, miR-221,miR-509, miR-513a, miR-514a, miR-548h-3, miR-601, miR-634, miR-676, miR-1183*, miR-1266, miR-1972, miR-2681, miR-3907, miR-4274, miR-4486, miR-4642*, miR-4754, miR-4784, miR-4786, miR-5188, miR-6076, miR-6133*, miR-6869*, miR-8062

*: The GCAUG sequence overlaps with the Dicer or Drosha cleavage sites.

[a] miR-767 has two GCAUG sequence (GCAUGCAGCAUG) within its terminal stem-loop.

[b] miR-297 has two UGCAUG sequence (UGCAUG UGCAUG) within its upper stem.

[c] miR-4799 has two GCAUG sequences within its upper stem.

[d] The lengths of the lower stems of miRNA hairpins from miRBase are variable, suggesting that some sequences may not be complete. If so, the number of miRNA hairpins containing targeting sequences that can be identified computationally would be under-estimated.

Supplemental Table S2. List of human precursor miRNA stem-loops from miRBase containing the GCACG sequence element.

TSL	Upper Stem	Lower Stem
miR-33b*, miR-658, miR-3652, miR-3653 miR-3939, miR-4297[a], miR-4747, miR-4758,	miR-16, miR-26a-1*, miR-124, miR-139, miR-196b, miR-363, miR-548-ay, miR-557, miR-564, miR-595[a], miR-1233, miR-3177, miR-3622a, miR-4445*, miR-4465, miR-4481, miR-4536, miR-4636, miR-4672*, miR-6715b, miR-6770,	miR-34a, miR-132, miR-135a-1*, miR-148b, miR-200b, miR-211, miR-596[b], miR-718*, miR-885, miR-3945, miR-4700*, miR-4781,

*: The GCACG sequence overlaps with the Dicer or Drosha cleavage sites.

[a] miR-595 and miR-4297 have two GCACG site;

[b] miR-596 has one GCACG and one GCAUG site.

Supplemental Table S3. NMR structure statistics for the pre-miR-20b RNA in the free form and in complex with the Rbfox RRM.

	Free RNA	Protein/RNA complex
NMR constraints		
Distance constraints		
Total NOEs	596	2359
RNA intra-residue	329	
RNA inter-residue	267	
Protein-RNA intermolecular		197
RNA intramolecular		424
Protein intramolecular		1738
Protein intra-residue		514
Sequential (|i-j|=1)		440
Medium range (1<|i-j|<5)		236
Long range (|i-j|≥5)		548
Hydrogen-bond constraints	38	62
Torsion angle constraints	63	149
Structure statistics (20 structures of lowest energy)		
Violations		
NOE violations > 0.3 Å	0	0
Torsion angle violations > 5°	0	0
Ramachandran plot statistics		
Residues in most favored regions		80.6%
Residues in additional allowed regions		16.3%

Residues in generously allowed regions		2.4%
Residues in disallowed regions		0.7%
RMS deviations from idealized geometry		
Bond lengths (Å)	0.014	0.014
Bond angles (°)	1.7	2.0
RMS deviations from the mean structure		
RNA heavy atoms (G20-U27, G33-C40)	0.69	
Protein backbone (Pro116-Arg194)		0.39
Protein heavy atoms (Pro116-Arg194)		0.93
RNA heavy atoms (G34-A39)		0.76
Complex heavy atoms (G34-A39 and Pro116-Arg194)		0.90

Supplemental Table S4. List of human miRNAs predicted to target the 3´-UTR of PTEN (according to TargetScan and PicTar [7,8]), which also contain sequence elements within their precursor hairpins targeted by Rbfox family proteins.

miRNA	Number of pairing sites on 3´-UTR of PTEN	Predicted targeting site on miRNA hairpin by Rbfox family proteins
miR-20b[9]	1	GCAUG in TSL
miR-23b[10]	2	GCAUG at bulge in upper stem
miR-26a[11]	3	UGCACG at Drosha cleavage site
miR-32[12]	1	UGCAUG in TSL
miR-148b	2	GCACG in lower stem
miR-152	2	UGCAUG at internal loop in upper stem
miR-200b[13]	1	UGCACG in low stem
miR-205[14]	1	GCAUG in low stem
miR-221[15]	1	GCAUG in low stem
miR-363[16]	1	UGCACG at internal loop in upper stem
miR-486[17]	1	GCAUG in TSL
miR-4465	3	GCACG at an internal loop in upper stem

Supplemental Figure S1

(**A**) Secondary structure model of primary miRNA hairpin, which we arbitrarily divide into three regions for the purpose of the present analysis, as indicated in Supplemental Tables 1 and 2. (**B**) Alignment of miR-20b and miR-107 stem-loops from various species using sequences derived from miRBase. The mature miRNA sequences are colored as green and passenger strands are colored as yellow for guiding purpose. The conserved GCAUG sequence element targeted by the Rbfox RRM is colored in red. It is widely conserved in vertebrates for pre-miR-107, but limited only to primates in pre-miR-20b. (**C**) Predicted secondary structures of full-length precursor miR-20b/107 hairpins from miRbase. Mature miRNA sequences (5p and 3p) are colored in blue. (**D**) Truncated pre-miR-20b/107 hairpins used in EMSA binding studies.

Supplemental Figure S2

NMR analysis of the pre-miR-20b stem-loop and its interaction with the Rbfox RRM. (**A**) Overlay of the ^{15}N HSQCs of 23 nucleotides pre-miR-20b displaying base pairing NHs in the free (black) and bound (red) states, with assignments. (**B**) Overlay of the ^{15}N HSQCs of the 46 nucleotides pre-miR-20b displaying base pairing NHs in the free (black) and bound (red) states, with assignments. (**C**) Strips of 2D F1-filtered, F2-edited NOESY of the complex showing intermolecular NOE cross peaks to the base protons of G_{29}, A_{31}, U_{32}, G_{33} and A_{34}.

Supplemental Figure S3

(**A**) Sequence alignment of Rbfox family proteins in human (Fox-1, NP_061193; Fox-2, NP_001026865; Fox-3; NP_001076044), mouse (Fox-1, NP_067452), zebrafish (Fox-1, NP_001005596) and C. elegans (Fox-1, NP_001248445). (**B**) Sequences of the Rbfox RRM (109-194), (109-208) and (109-225) constructs used in the present study.

Supplemental Figure S4

Representative ITC binding isotherms for titration of (**A**) Rbfox RRM (109-194 of Rbfox1), (**B**) Rbfox RRM (109-208 of Rbfox1) and (**C**) Rbfox RRM (109-225 of Rbfox1) into pre-miR-20b. Corresponding thermodynamic parameters determined by ITC are shown as well. Two-site binding models were applied for better curve-fitting. The second weaker binding event is probably due to a weaker non-specific interaction between Rbfox RRM and pre-miR-20b.

Supplemental Figure S5

Analysis of endogenous Rbfox1 and Rbfox2 protein expression and of endogenous levels of mature miR-20b and miR-107. (**A**) Immunoblots analysis of Rbfox1 and Rbfox2 in HEK293, MCF7, NSC34, SHSY-5Y and HeLa cells. As indicated by the arrowhead, Rbfox1 is mainly expressed in the mouse motor neuron cell line NCS-34. In contrast, no expression is observed in the other cell lines analyzed. Beta-tubulin serves as a loading control. In contrast, Rbfox2 is highly expressed by HEK293 and in the neuroblastoma cell line SHSY-5Y. The asterisk indicates a possible Rbfox2 (1F) isoform. (**B**) Quantification of endogenous mature miRNAs in HeLa, HEK293, MCF7 and SHSY5Y cells. The quantification of mature miR-20b and miR-107 was done by qRT-PCR. n =3 biological replicates; average ±s.e.m., * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (OneWay ANOVA statistical test).

Supplemental Figure S6

(**A**) Immunoblot analysis of Rbfox2 dowregulation by RNAi in SHSY-5Y cells. Beta-tubulin serves as a loading control. Cells were collected 48 hours after siRNA transfection. CTRL: cell lysate from cells transfected with scrambled siRNA. (**B**) Immunoblot analysis of FLAG-tagged Rbfox2 over expression. Cell lysates were prepared from MCF7 cells 48 hours after transfected with either an empty expression plasmid (CTRL) or with a plasmid expressing a FLAG-tagged Rbfox2. Actinin serves as a loading control.

Supplemental Figure S7

Analysis of endogenous Rbfox1 and Rbfox2 protein expression and the levels of mature miR-20b and 107. (**A**) Immunoblot analysis of Rbfox2 in MCF7 and mda-mb-231 cells. Rbfox-1A and Rbfox-1F represent two different protein isoforms. (**B**) Quantification of endogenous mature miR-20b and miR-107 levels in MCF7 and mda-mb-231 cells. (**C**) Quantification of endogenous mature miR-20b in mda-mb-231 cells after knockdown of Rbfox2 protein. (**D**) Effect of downregulation of Rbfox2 on mature miR-107 in mouse 4TO7 cells. (**E**) Immunoblot analysis of Dicer and Rbfox2 proteins in 4TO7 cells. Rbfox2-1A and Rbfox2-1F reppresent two different protein isoforms. Actinin serves as a loading control.

Supplemental Figure S8

Structural comparison of the Rbfox RRM (i.e. Fox-1 RRM) in complex with pre-miR-20b and single-stranded UGCAUGU6. Rbfox RRM and pre-miR-20b in the current structure are colored in green and orange; Fox-1 RRM and UGCAUGU in the reported structure are colored in white and blue, respectively. To differentiate residues and nucleotides from each complex, labels for Fox-1-UGCAUGU are provided in brackets. (A) Overall comparison of the two structures indicates high similarity. (B) Close-up view of G29 (G2) and their interactions with the protein. (C) Close-up view of G28 and C30 (U1 and C3) and their interactions with the protein. (D) Close-up view of A31 and U32 (A4 and U5) and their interactions with the protein. (E) Close-up view of G33 and A34 (G6 and U7) and their interactions with the protein.

Supplemental Figure S9

Long-range interactions between the C-terminal tail of Rbfox1 RRM and the lower-stem of pre-miR-20b. (**A**) Overlays of ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled Rbfox1 RRM (109-225) in complex with paramagnetic spin-labeled pre-miR-20b before (red) and after (green) the reduction of the spin label introduced at position U_{43}. Some resonances from the C-terminal tail of Rbfox1 RRM, which are broadened by the paramagnetic spin label of pre-miR-20b, are annotated. (**B**) Intensity ratios of NH cross-peaks from Rbfox1 RRM (109-225) in complex with pre-miR-20b, between paramagnetic and diamagnetic forms. Residues from the $\beta_2\beta_3$ loop (around E152) and the C-terminal region (V195-V215) show significant depressions, indicating long-range contacts between the bottom part of the RNA and the protein. (**C**) Cartoon representation showing how the highly conserved C-terminal tail of RRM can reach the bottom part of the stem-loop to provide additional contacts that are obviously not possible with single stranded RNA.

Supplementary References

1. Ramos, A., Varani, G. A new method to detect long-range protein-RNA contacts: NMR detection of electron-proton relaxation induced by nitroxide spin-labeled RNA. *J. Am. Chem. Soc.* **120**, 10992–10993 (1998).

2. Herrmann, T., Güntert, P., & Wüthrich, K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319** (1), 209–227 (2002).

3. Case, D. A., *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26** (16), 1668–1688 (2005).

4. Tolbert, B. S., *et al.* Major groove width variations in RNA structures determined by NMR and impact of 13C residual chemical shift anisotropy and 1H-13C residual dipolar coupling on refinement. *J. Biomol. NMR*, **47** (3), 205–219 (2010).

5. Shen, Y., Delaglio, F., Cornilescu, G., & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR*, **44** (4), 213–223 (2009).

6. Auweter, S. D., Fasan, R., Reymond, L., Underwood, J. G., Black, D. L, Pitsch, S., Allain, F. H-T. Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J*. **25**, 163-173 (2006).

7. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15-20 (2005).

8. Krek, A. *et. al* Combinatorial microRNA target predictions. *Nat. Genetics* **37**, 495-500 (2005).

9. Zhou, W., Shi, G., Zhang, Q., Wu, Q., Li, B. & Zhang, Z. MicroRNA-20b promotes cell growth of breast cancer cells partly via targeting phosphatase and tensin homologue (PTEN). *Cell & Bioscience* **4**, 62-72 (2014).

10. Zaman, M.S., *et al.* Inhibition of PTEN gene expression by oncogenic miR-23b-3p in renal cancer. *PLoS One*, 7(11), e50203 (2012).

11. Huse, J. T. *et al*. The PTEN-regulating microRNA miR-26a is amplified in high-grade glioma and facilitates gliomagenesis *in vivo*. *Genes Dev*. **23**, 1327–1337 (2009).

12. Wu, W., Yang, J., Feng, X., Wang, H., Ye, S., Yang, P., Tan, W., Wei, G. & Zhou, Y. MicroRNA-32 (miR-32) regulates phosphatase and tensin homologue (PTEN) expression and promotes growth, migration, and invasion in colorectal carcinoma cells. *Mol. Cancer* **12**:30 (2013).

13. Soubani, O., Ali, A.S., Logna, F., Ali, S., Philip, P.A. & Sarkar, F.H. Re-expression of miR-200 by novel approaches regulates the expression of PTEN and MT1-MMP in pancreatic cancer. *Carcinogenesis* **33**, 1563-1571 (2012).

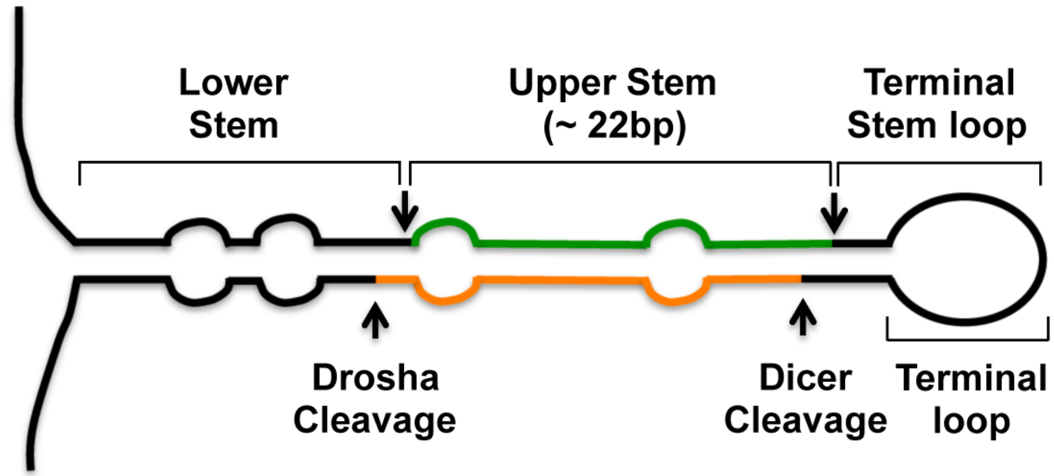
14. Cai, J., *et al.* miR-205 targets PTEN and PHLPP2 to augment AKT signaling and drive malignant phenotypes in non-small cell lung cancer. *Cancer Res*. **73**, 5402-5415 (2013).

15. Chun-Zhi, Z. *et al.* MicroRNA-221 and microRNA-222 regulate gastric carcinoma cell proliferation and radioresistance by targeting PTEN. *BMC Cancer* **10**, 367-376 (2010).

16. Poliseno, L. *et al.* Identification of the miR-106b**-**25 microRNA cluster as a proto-oncogenic PTEN-targeting intron that cooperates with its host gene MCM7 in transformation, *Sci. Signal*. **3**, ra29 (2010).

17. Small, E. M. *et al.* Regulation of PI3-kinase/Akt signaling by muscle-enriched microRNA-486. *Proc. Natl Acad. Sci. USA* **107**, 4218–4223 (2010).

A



B

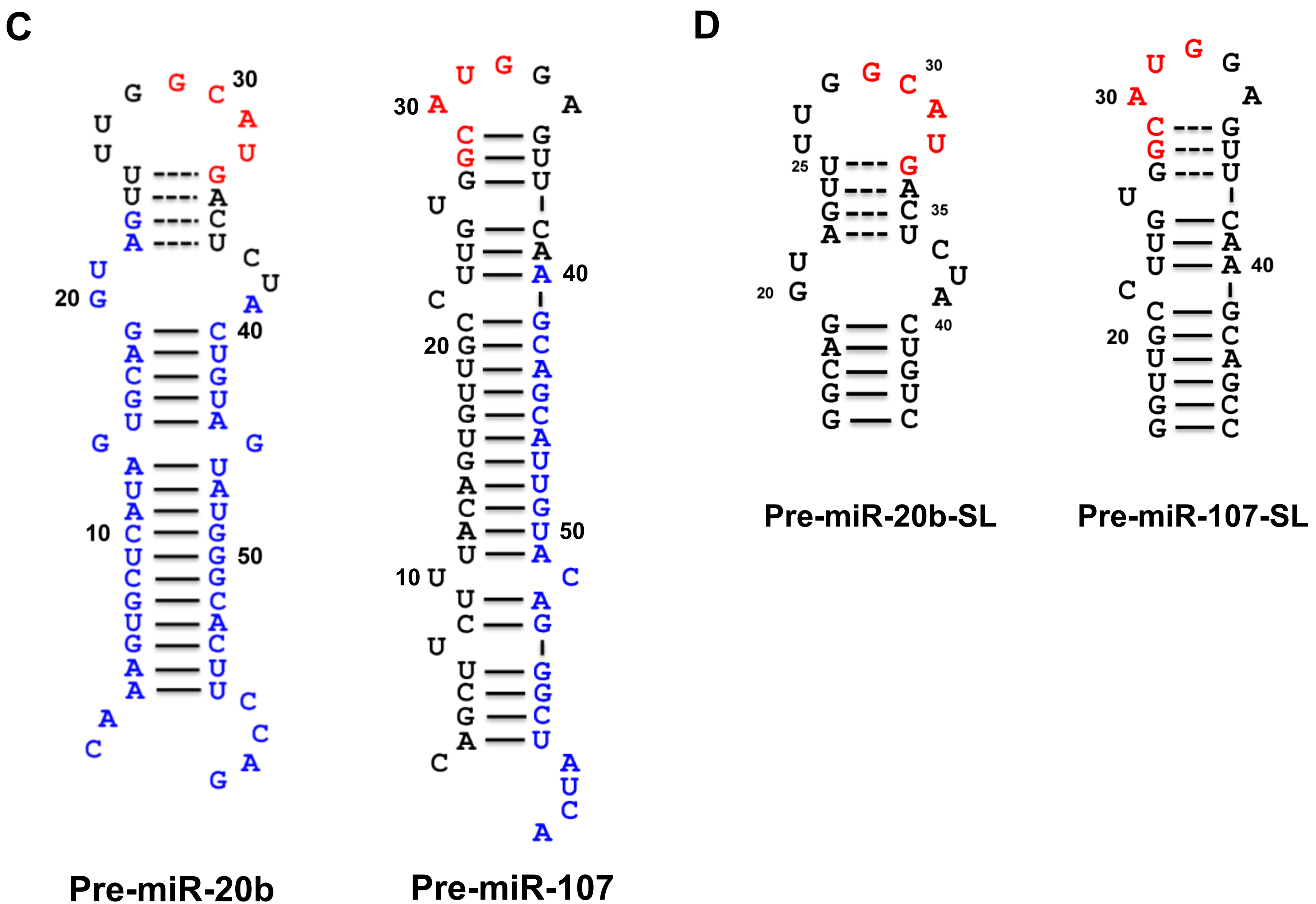
Pre-miR-20b

```
Human        -------------------AGUACCAAAGUGCUCAUAGUGCAGGUAGUUUUGGCAUGACU---------CUACUGUAGUAUGGGCACUUCCAGUACU--------------
Orangutang   -------------------AGUACCAAAGUGCUCAUAGUGCAGGUAGUUUUGGCAUGACU---------CUACUGUAGUAUGGGCACUUCCAGUACU--------------
Gorilla      -AGAUU---GGGUCCUAGUAGUACCAAAGUGCUCAUAGUGCAGGUAGUUUUGGCAUGACU---------CUACUGUAAUAUGGGCACUUCCAGUACUCUUGGAUAACAAAU
Rhesus       -------------------AGUACCAAAGUGCUCAUAGUGCAGGUAGUUUUGGCAUGACU---------CUACUGUAGUGUGGGCACUUCCAGUACU--------------
Mouse        -------------CCUAGUAGUGCCAAAGUGCUCAUAGUGCAGGUAGUUUUUAUACCACU---------CUACUGCAGUGUGAGCACUUCUAGUACUCCUGG---------
Rat          --------GU---------AGUGCCAAAGUGCUCAUAGUGCAGGUAGGUUUUGCUGCACU---------CUACUGCAGUGUGAGCACUUCUGGUACUC-------------
Dog          --------------------GUACCAAAGUGCUCACAGUGCAGGUAGU---------CUGGGCUUUUCUCUACUGUAGUGUGGGCACUUCCGGUAAC--------------
Cow          -------------------AGUACCAAAGUGCUCACAGUGCAGGUAGU----UUUGGCAGCGCU-----CUACUGUAGUGUGGGCACUUCCAGUACU--------------
Opossum      ------------------------CAAAGUGCUCAUAGUGCAGGUAGUUUU--UUGCAAAUGACU------ACUGUACUAUGGGCACUUCAG--C----------------
Platypus     UGGAAAUCG-UAUCCUGACAGUACCAAAGUGCUCAUAGUGCAGGUAGUUUU--UUUCAAUUGAU-U---CUACUGUAAUGUGGGCACUUACAGUACUCCAGGAUAAAGUGC
Frog         -----------------GCAGUUCCAAAGUGCUCAUAGUGCAGGUAGU------UGUAUU-GAUGUU--CUACUGUAAUAUGGGCACUUACAGU-ACUGCU----------
                                    *********** ***********                         **** * *.** ******              .   .
```

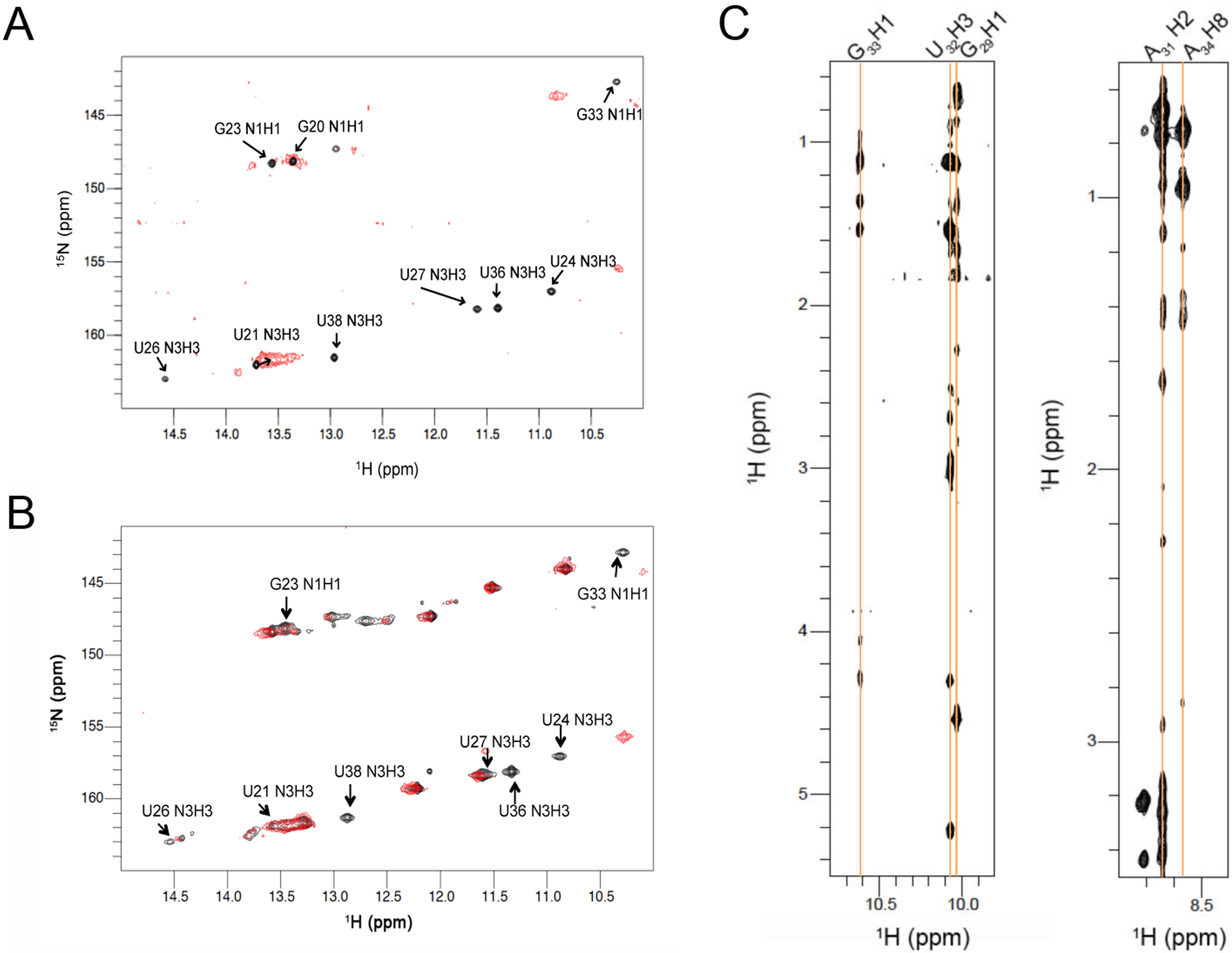
pre-miR-107

```
Human        --CUCUCUGCUUUCAGCUUCUUUACAGUGUUGCCUUGUGGCAUG--GAGUUCAAGCAGCAUUGUACAGGGCUAUCAAAGCACAGA----
Orangutang   --CUCUCUGCUUUCAGCUUCUUUACAGUGUUGCCUUGUGGCAUG--GAGUUCAAGCAGCAUUGUACAGGGCUAUCAAAGCACAGA----
Gorilla      --CUCUUUGCUUUCAGCUUCUUUACAGUGUUGCCUUGUGGCAUG--GAGUUCAAGCAGCAUUGUACAGGGCUAUCAAAGCAUGGA----
Rhesus       --CUCUCUGCUUUCAGCUUCUUUACAGUGUUGCCUUGUGGCAUG--GAGUUCAAGCAGCAUUGUACAGGGCUAUCAAAGCACAGA----
Mouse        UUCUCUGUGCUUUCAGCUUCUUUACAGUGUUGCCUUGUGGCAUG--GAGUUCAAGCAGCAUUGUACAGGGCUAUCAAAGCACAGAGAGC
Rat          UUCUCUCUGCUUUAAGCUUCUUUACAGUGUUGCCUUGUGGCAUG--GAGUUCAAGCAGCAUUGUACAGGGCUAUCAAAGCACAGAGAGC
Dog          --------------AGCUUCUUUACAGUGUUGCCUUGUGGCAUG--GAGUUCAAGCAGCAUUGUACAGGGCUAU---------------
Cow          --CUCUCUGCUUUCAGCUUCUUUACAGUGUUGCCUUGUGGCAUG--GAGUUCAAGCAGCAUUGUACAGGGCUAUCAAAGCACAGA----
Frog         ----CUG--CUUUCAGCUUCUUUACAGUGUUGCCUUGUGGCAUG--GAGUUCAAGCAGCAUUGUACAGGGCUAUCAAAGCA--------
Pufferfish   --CUCCCUGCUCUCAGCCUCUUUACGGUGCUGCCUUGUGGCAUCUUGAU--CAAGCAGCAUGUACAGGGCUAUGAAGUGC---------
                 *    ** *.*** *******.*** ********   **   *********************** *. **
```

Supplemental Fig. 1

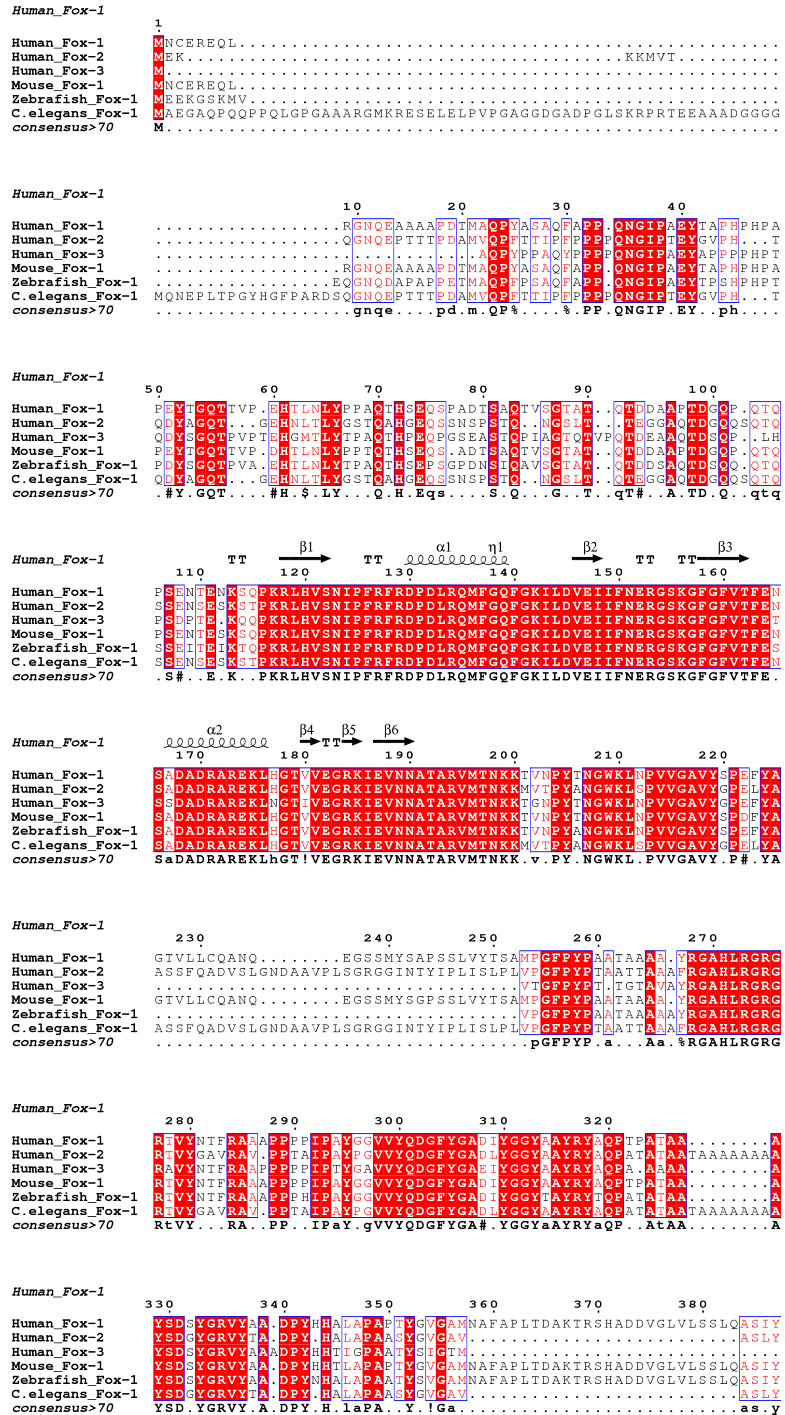


Supplemental Fig. 1



Supplemental Fig. 2

A



Supplemental Fig. 3

B

Rbfox1 RRM (109-194):

NT ENKSQPKRLH VSNIPFRFRD PDLRQMFGQF GKILDVEIIF NERGSKGFGF VTFENSADAD RAREKLHGTV VEGRKIEVNN ATAR

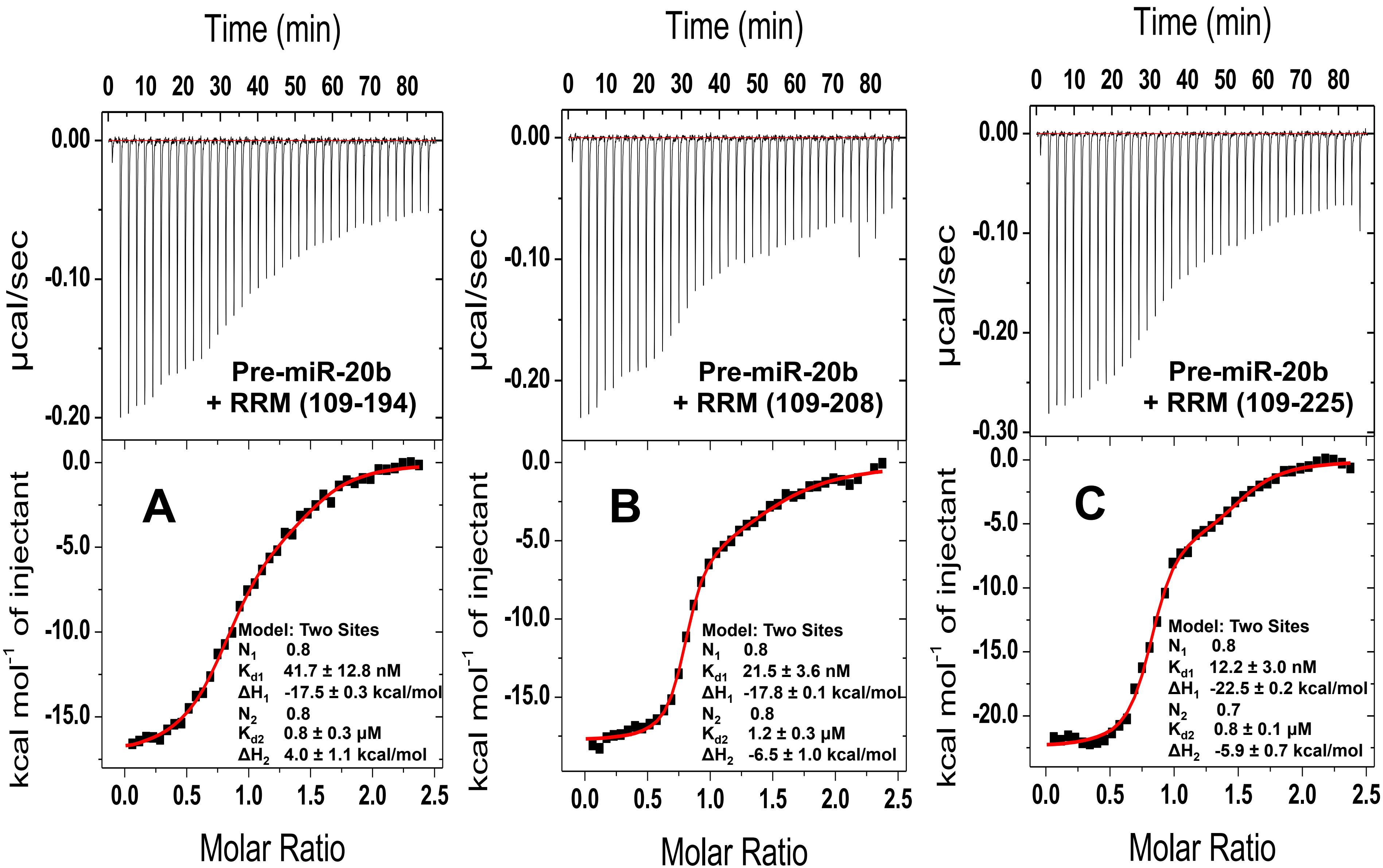
Rbfox1 RRM (109-208):

NT ENKSQPKRLH VSNIPFRFRD PDLRQMFGQF GKILDVEIIF NERGSKGFGF VTFENSADAD RAREKLHGTV VEGRKIEVNN ATARVMTNKK TVNPYTNG

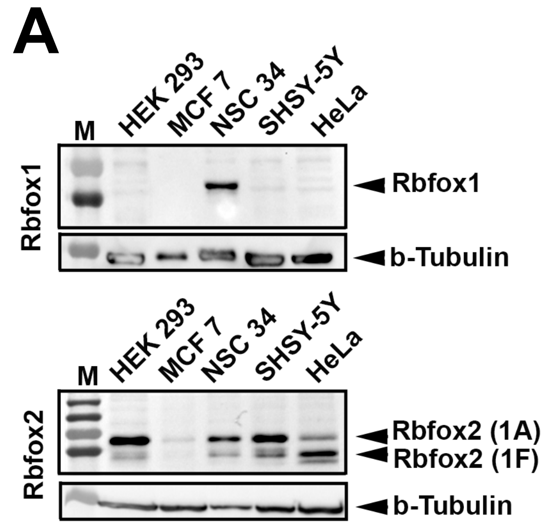
Rbfox1 RRM (109-225):

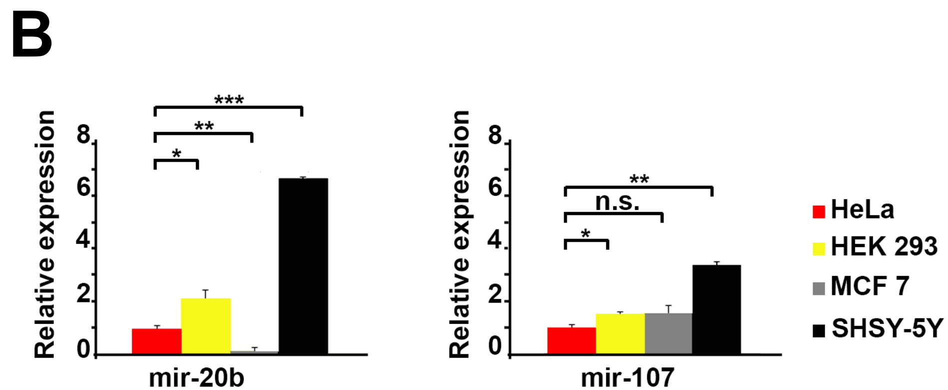
NT ENKSQPKRLH VSNIPFRFRD PDLRQMFGQF GKILDVEIIF NERGSKGFGF VTFENSADAD RAREKLHGTV VEGRKIEVNN ATARVMTNKK TVNPYTNGWK LNPVVGAVYS PEFYA

Supplemental Fig. 3



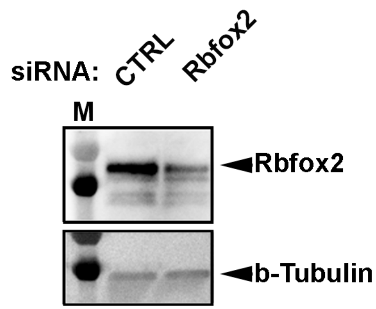
Supplemental Fig. 4



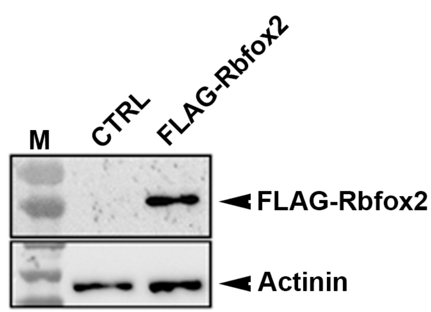


Supplemental Fig. 5

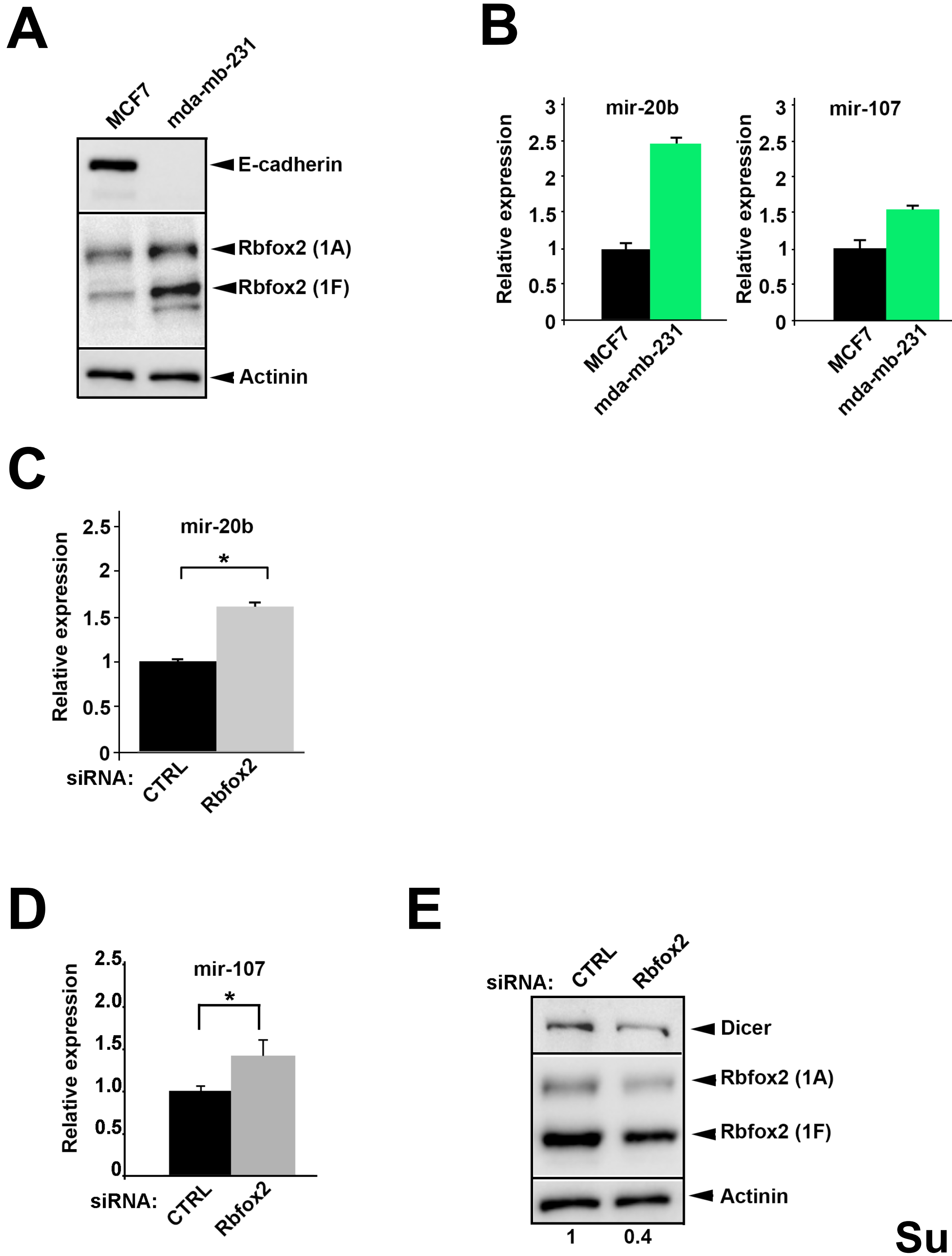
A



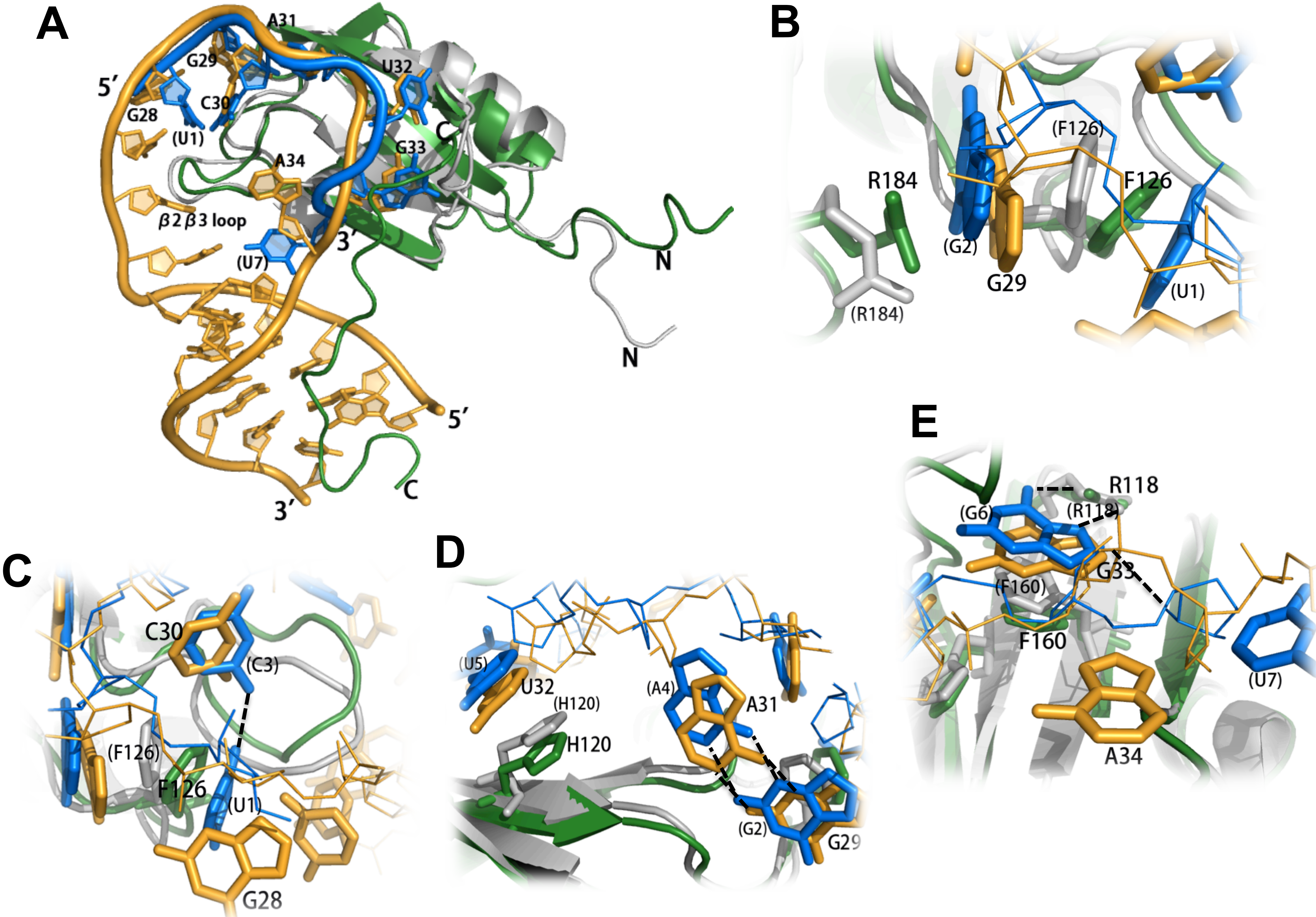
B



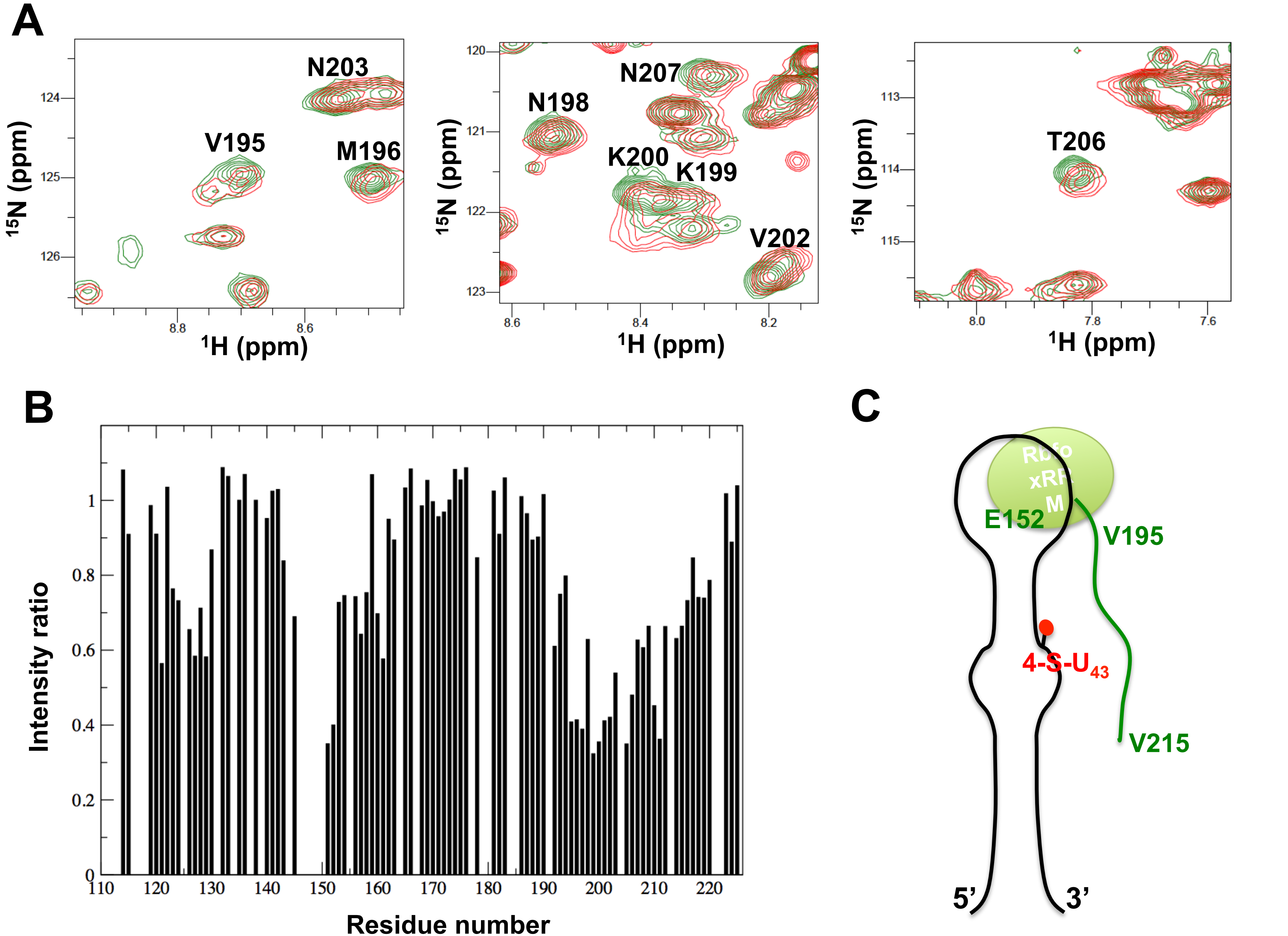
Supplemental Fig. 6



Supplemental Fig. 7



Supplemental Fig. 8



Supplemental Fig. 9