NMR chemical shift assignments of RNA oligonucleotides to expand the RNA chemical shift database

Yaping Liu[1#], Anita Kotar[1,5#], Tracy L. Hodges[1], Kyrillos Abdallah[1], Mallak H. Taleb[1], Brayden A. Bitterman[1], Sara Jaime[1], Kyle J. Schaubroeck[1], Ethan Mathew[1], Nicholas W. Morgenstern[1], Anthony Lohmeier[1], Jordan L. Page[1], Matt Ratanapanichkich[1], Grace Arhin[1], Breanna L. Johnson[1], Stanislav Cherepanov[1], Stephen C. Moss[1], Gisselle Zuniga[1], Nicholas J. Tilson[1], Zoe C. Yeoh[2], Bruce A. Johnson[3], and Sarah C. Keane*[1,4]

[1] Biophysics Program, University of Michigan, 930 N. University Avenue, Ann Arbor, MI 48109, USA

[2] Department of Biological Chemistry, University of Michigan, 1150 W. Medical Center Drive, Ann Arbor, MI 48109, USA

[3] Structural Biology Initiative, CUNY Advanced Science Research Center, 85 St. Nicholas Terrace, New York, NY, 10031, USA

[4] Department of Chemistry, University of Michigan, 930 N. University Avenue, Ann Arbor, MI 48109, USA

[5] Current Address: Slovenian NMR Centre, National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana, Slovenia

[#] denotes equal contribution

*Corresponding author: sckeane@umich.edu

Abstract – 150-250 words

RNAs play myriad functional and regulatory roles in the cell. Despite their significance, three-dimensional structure elucidation of RNA molecules lags significantly behind that of proteins. NMR-based studies are often rate-limited by the assignment of chemical shifts. Automation of the chemical shift assignment process can greatly facilitate structural studies, however, accurate chemical shift predictions rely on a robust and complete chemical shift database for training. We searched the Biological Magnetic Resonance Data Bank (BMRB) to identify sequences that had no (or limited) chemical shift information. Here, we report the chemical shift assignments for 12 RNA hairpins designed specifically to help populate the BMRB.

Keywords:

RNA, solution NMR spectroscopy, GAGA tetraloop, chemical shift database

Declarations:

Funding.

This work was supported by National Science Foundation grant MCB-1942398 (to S.C.K.) and by National Institute of General Medical Sciences of the National Institutes of Health grant U54 GM 103297 (to B.A.J.). Research reported in this publication was supported by the University of Michigan BioNMR Core Facility (U-M BioNMR). U-M BioNMR Core is grateful for support from U-M including the College of Literature, Sciences and Arts, Life Sciences Institute, College of Pharmacy and the Medical School along with the U-M Biosciences Initiative.

Conflicts of interest/Competing interests.

The authors declare that they have no conflict of interest.

Availability of data and material.

Assigned chemical shifts along with raw NMR data have been deposited in the BMRB. RNA5: 50933, RNA7: 50932, RNA8: 50931, RNA21: 50930, RNA23: 50929, RNA24: 50928, RNA73: 50927, RNA74: 50926, RNA75: 50925, RNA89: 50924, RNA90: 50923, RNA91: 50922.

Biological context

RNA is a multifaceted biomolecule that plays important functional and regulatory roles in the cell. In the simplest case, RNA has a crucial role in the central dogma of molecular biology. Not only do messenger (m) RNAs serve as the template for protein synthesis in the cell, the cellular machinery responsible for protein synthesis, the ribosome, is largely composed of RNA molecules. Outside of its role in the central dogma, RNAs contribute to the overall health of the cell by controlling splicing (Papasaikas and Valcarcel, 2016), translation efficiency (Rodnina, 2016), and genomic stability (Theimer and Feigon, 2006). More recently, RNA molecules have been used as the foundation for vaccines (Corbett et al., 2020; Pardi et al., 2018), investigated as therapeutics (Burnett and Rossi, 2012; Damase et al., 2021; Winkle et al., 2021), and have been identified as potential therapeutic targets themselves (Howe et al., 2015; Meyer et al., 2020; Warner et al., 2018).

Despite the importance of RNAs in biology, structural elucidation of RNA molecules lags significantly behind their protein counterparts. As of May 2021, the Protein Data Bank (PDB) (Berman et al., 2000a; Berman et al., 2000b) has 173,534 protein-containing structures and only 5,364 RNA-containing structures. Similarly, the Biological Magnetic Resonance Data Bank [BMRB (Ulrich et al., 2008)] contains 13,420 protein depositions and only 467 RNA depositions. This imbalance in our RNA structural knowledge has led to a similar disparity in our mechanistic understandings of the functions of these RNAs.

Chemical shifts are a fundamental indicator of molecular structure and are at the heart of all NMR studies (Ebrahimi et al., 2001; Fares et al., 2007; Ghose et al., 1994; Giessner-Prettre and Pullman, 1987; Rossi and Harbison, 2001; Sripakdeevong et al., 2014). However, assignment of chemical shifts remains a laborious and often rate-limiting step in the determination of biomolecular structure and characterization of biomolecular dynamics and ligand interactions. A number of groups are actively working to develop and improve methods to predict chemical shifts of nuclei in proteins (Han et al., 2011; Meiler, 2003), DNA (Kwok and Lam, 2013; Lam, 2007; Lam et al., 2007; Ng and Lam, 2015), and RNAs (Aeschbacher et al., 2013; Bahrami et al., 2012; Barton et al., 2013; Brown et al., 2015; Frank et al., 2013; Frank et al., 2014; Krahenbuhl et al., 2014; Wang et al., 2021). A robust prediction, coupled with a linked, interactive data analysis software (Marchant et al., 2019), can significantly reduce the time required for the assignment of chemical shifts. However, the accuracy of these algorithms requires a robust database of chemical shifts that represent a wide variety of both sequence and structural motifs. At the beginning of this study, we found only 244 entries at the BMRB for RNA molecules that were usable with our chemical shift analysis

tools. Features which excluded some (27) BMRB entries from our analysis included the lack of an associated PDB file, presence of unnatural or modified nucleotides, or lack of relevant chemical shifts. The sparseness of this database limits the accuracy of the available tools.

In a canonical A-form helix, the neighboring base pairs have the most significant impact on the chemical shifts of the central base pair (Barton et al., 2013), forming 64 possible canonical base pair triplets or "core sequences". We aim to identify the effect that bases flanking these core sequences have on the chemical shifts of the central base pair. Since the flanking base pairs have a relatively small effect on the chemical shifts of the central base pair as compared to the internal bases of the core sequence, we considered only nucleotide type (purine or pyrimidine) rather than nucleotide identity at the flanking positions. This is consistent with the attributes used in the current prediction algorithm used in NMRFx Analyst (Marchant et al., 2019). For the purposes of this study, we define a "group" as a core sequence with variable flanking bases. If each of these core sequences is flanked by a purine-pyrimidine base pair or a pyrimidine-purine base pair at the 5′- and 3′-ends, there are 256 possible combinations of 5 base pair sequences (**Table S1**). With these 256 sequences in hand, we examined the completeness of aromatic proton and carbon chemical shifts deposited in the BMRB and identified four "groups" of RNAs whose chemical shifts were incomplete (**Fig. 1**). We prepared 12 unique RNAs, three from each group, for chemical shift assignment. These sequences were chosen from a number of missing sequences (see Methods, below) because they represented groups that were the most incomplete (i.e. three of the four group members had missing chemical shift data).

The RNAs examined in this study were designed to fill specific gaps in the BMRB chemical shift database. Additionally, these RNAs represent helical regions from a diverse set of biological RNAs including ribosomes, tRNAs, riboswitches, internal ribosome entry sites, and viral RNAs (**Table S2**) as identified via the RNA fragments search engine and database [RNA FRABASE, (Popenda et al., 2008; Popenda et al., 2010)]. The RNA FRABASE searches for RNA fragments based on known structures, therefore there are likely many more biologically relevant RNAs that are represented by these RNA oligonucleotides. Our efforts focused on the assignment of aromatic and anomeric protons, as these are the chemical shifts that are currently incorporated into the NMRFx pipeline. Furthermore, we report imino proton assignments which inform on secondary structure and continue to build on the database and prediction tool reported earlier this year (Wang et al., 2021). We expect that these additional chemical

shift assignments, which are a step towards a more complete chemical shift database, will benefit the RNA NMR community.

Materials and Methods

Identification of RNA sequences that are underrepresented in the BMRB

The identification of RNA sequences that are underrepresented in the BMRB was done using the tools previously developed for training models for the prediction of RNA chemical shifts (Brown et al., 2015). We used these tools to analyze NMR-STAR files containing RNA data that we downloaded from the BMRB. The downloaded NMR-STAR files were analyzed with the scripts developed for the prediction model training. This process automatically identifies PDB files from which secondary structures are generated and does various validation checks. After this process was complete, 244 files were available for the subsequent analysis. This analysis generated a table of chemical shifts and attributes as used in the prediction software. The tabular data was searched for all 256 five-residue sequences described above and a new table generated containing a line for each of the 256 sequences. Each line contains the number of measured chemical shifts for each of the aromatic proton and carbon atoms for the corresponding sequence. Of these 256 sequences, 182 of them have at least one measurement of all the aromatic proton and carbons (**Table S1**). The remaining 74 entries were missing the measurement of the chemical shift of at least one atom.

Construct design and template preparation

DNA oligonucleotides corresponding to the 12 RNA constructs (**Table 1**) were ordered (Integrated DNA Technologies) with 2′-*O*-methoxy modifications at the two 5′-most positions to reduce non-templated transcription (Kao et al., 1999). To generate the templates for transcription, DNA oligonucleotides (**Table 1**) were annealed with an oligo corresponding to the T7 promoter sequence (5′-TAATACGACTCACTATA-3′). Briefly, 40 µL of each DNA oligonucleotide (200 µM) was individually mixed with 20 µL of the T7 promoter sequence oligonucleotide (600 µM). These samples were placed in boiling water for 3 minutes and left in the water bath to slowly cool to room temperature over at least 2 hours. The annealed template was then diluted with 940 µL H_2O to produce the partially double-stranded DNA templates for transcription.

Table 1. RNA constructs.

RNA	RNA sequence (5′-3′) [a]	DNA template (5′-3′)[b,c]
RNA5	GGGAAUGCUGAGAAGCAUUCCC	mGmGGAATGCTTCTCAGCATTCCC*TATAGTGAGTCGTATTA*
RNA7	GGCAAUGCUGAGAAGCAUUGCC	mGmGCAATGCTTCTCAGCATTGCC*TATAGTGAGTCGTATTA*
RNA8	GGCAAUCCUGAGAAGGAUUGCC	mGmGCAATCCTTCTCAGGATTGCC*TATAGTGAGTCGTATTA*
RNA21	GGGAUUGCUGAGAAGCAAUCCC	mGmGGATTGCTTCTCAGCAATCCC*TATAGTGAGTCGTATTA*
RNA23	GGCAUUGCUGAGAAGCAAUGCC	mGmGCATTGCTTCTCAGCAATGCC*TATAGTGAGTCGTATTA*
RNA24	GGCAUUCCUGAGAAGGAAUGCC	mGmGCATTCCTTCTCAGGAATGCC*TATAGTGAGTCGTATTA*
RNA73	GGGUAGGCUGAGAAGCCUACCC	mGmGGTAGGCTTCTCAGCCTACCC*TATAGTGAGTCGTATTA*
RNA74	GGGUAGCCUGAGAAGGCUACCC	mGmGGTAGCCTTCTCAGGCTACCC*TATAGTGAGTCGTATTA*
RNA75	GGCUAGGCUGAGAAGCCUAGCC	mGmGCTAGGCTTCTCAGCCTAGCC*TATAGTGAGTCGTATTA*
RNA89	GGGUUGGCUGAGAAGCCAACCC	mGmGGTTGGCTTCTCAGCCAACCC*TATAGTGAGTCGTATTA*
RNA90	GGGUUGCCUGAGAAGGCAACCC	mGmGGTTGCCTTCTCAGGCAACCC*TATAGTGAGTCGTATTA*
RNA91	GGCUUGGCUGAGAAGCCAAGCC	mGmGCTTGGCTTCTCAGCCAAGCC*TATAGTGAGTCGTATTA*

[a] Underlined regions correspond to variable sequences in the RNAs. The GAGA tetraloop sequence is red.
[b] m denotes 2′-*O*-Me modification of the primer.
[c] Italicized nucleotides correspond to the sequence complementary to the T7 promoter.

RNA preparation

RNA transcription conditions were optimized prior to large scale preparation. RNAs were prepared by *in vitro* transcription in transcription buffer (40 mM Tris-base, 5 mM DTT, 1 mM Spermidine, 0.01% Triton-X, pH 8.5) with addition of 2-6 mM unlabeled NTPs, 10-20 mM $MgCl_2$, 8-13 pmol annealed DNA template, 0.2 U/mL yeast inorganic pyrophosphatase (New England Biolabs) (Cunningham and Ofengand, 1990), ~15 µM T7 RNA polymerase, and 10-20% (vol/vol) DMSO (Helmling et al., 2015). Reactions were incubated at 37 °C for 3-4 hours and then quenched using a solution of 7 M urea and 250 mM EDTA (pH 8.5). The quenched transcription mixture was loaded onto 16% preparative-scale denaturing gels for RNA purification. The RNA band was identified by UV shadowing, excised, and electroeluted from the gel for ~24 h using a membrane trap elution system (Elutrap, Whattman). RNAs were spin concentrated, washed with 2 M high-purity sodium chloride, and exchanged into water using Amicon Centrifugal Filter Units (Millipore, Sigma). RNA purity was checked by running RNA on a 16% analytical denaturing gel.

NMR spectroscopy

RNA samples (300-400 µM in 300 µL) were prepared in 50 mM potassium phosphate buffer, pH = 7.4. Samples were lyophilized and rehydrated in an equal volume of D_2O or in 10% D_2O/90% H_2O (99.8%, Cambridge Isotope Laboratories, Inc.). 2D ^{1}H-^{1}H NOESY (in both D_2O and H_2O), ^{1}H-^{1}H TOCSY, and ^{1}H-^{13}C HMQC spectra were recorded for each RNA at 25 °C (see **Table 2** for experimental parameters). NMR spectra were collected on a 600 MHz Bruker AVANCE NEO spectrometer equipped with a 5 mm TCI cryogenic probe (University of Michigan

BioNMR Core). NMR data were processed with NMRFx (Norris et al., 2016) and NMRPipe (Delaglio et al., 1995) and analyzed with NMRViewJ (Johnson and Blevins, 1994). ^{1}H chemical shifts were referenced to water and ^{13}C chemical shifts were indirectly referenced from the ^{1}H chemical shift (Wishart et al., 1995).

Table 2. Experimental parameters for NMR data acquisition.

Experimental parameter	**^{1}H-^{1}H NOESY (100% D_2O)**	**^{1}H-^{1}H NOESY (90% H_2O/10% D_2O)**	**^{1}H-^{1}H TOCSY**	**^{1}H-^{13}C HMQC**
pulse sequence	noesyesgpph (Hwang and Shaka, 1995)	noesyesgpph (Hwang and Shaka, 1995)	mlevphpr.2 (Bax and Davis, 1985)	hmqcphpr (Bax et al., 1983)
ds	16	16	16	16
ns	24	24	24	336
sw(F2)	9.8104	19.8544	9.8104	9.8104
sw(F1)	9.8104	20.0000	9.8104	50
TD(F2)	8129	8192	2048	1058
TD(F1)	512	800	512	80
O1 (ppm)	4.703	4.703	4.703	4.703
O2 (ppm)	-	-	-	148.00
D1 (s)	3	3	2	1.5
mixing time (ms)	300	300	80	-
experiment time (h)	~14	~20	~8	~12

Assignments of helical regions

For all 12 RNA oligonucleotides, the nonexchangeable proton assignments were unambiguously assigned using 2D ^{1}H-^{1}H NOESY (Hwang and Shaka, 1995), 2D ^{1}H-^{1}H TOCSY (Bax and Davis, 1985), and ^{1}H-^{13}C HMQC (Bax et al., 1983) experiments. The 2D ^{1}H-^{1}H TOCSY spectrum yields strong H5-H6 cross-peaks, reporting on the number of pyrimidines (cytosine and uracil) in an RNA molecule, and was used to identify the H5-H6 correlations in the corresponding ^{1}H-^{1}H NOESY spectrum. The ^{1}H-^{1}H NOESY spectrum was used to make assignments of sequential nucleotides in A-helical regions. Near the diagonal, sequential aromatic-aromatic proton correlations were observable for many regions of each RNA. In the aromatic-anomeric region, sequential assignments were possible by following the NOE pattern for A-helical regions. Briefly, the H6/H8 of each residue (i) has NOE with its own H1′ and the H1′ of the preceding nucleotide (i-1). Assignments could be confirmed by examining additional NOEs, for example the aromatic H6/H8 proton of a residue (i) has a weak, but detectable NOE to the H5 of a following (i+1) pyrimidine. Additionally, the position of the adenosine C2 proton (H2) in the interior of the helix provides rich NOE data informing on both intra-strand (sequential) and cross-strand connectivities. When part of an A-form helix, the adenosine H2 (i) has NOEs to its own H1′ (i), to the H1′ of its 3′ residue (i+1), and an inter-strand

cross-peak to H1′ of the residue (j+1) 3′ of the base to which it is paired (j). The 2D ^{1}H-^{13}C HMQC was used to help distinguish C2-H2 (^{13}C δ ~152 ppm) from C6-H6 and C8-H8 resonances (^{13}C δ ~139 ppm). Assigned ^{1}H-^{1}H NOESY (aromatic) and ^{1}H-^{13}C HMQC spectra for the 12 RNAs are presented in **Figs. 3-7** and **Supplemental Figs. S1-S9**.

Assignments of imino protons

The exchangeable imino protons were unambiguously assigned using the 2D ^{1}H-^{1}H NOESY recorded in 10% D_2O/90% H_2O and confirm the secondary structure of the RNAs. A:U base pairs are easily identifiable via a strong NOE cross peak between the uracil H3 imino proton and the adenosine H2 proton, which serves as a good starting point for the assignment of the imino proton region of the NOESY experiment. For G:C base pairs, the guanosine H1 imino proton shows a strong NOE correlation to the amino protons (H41 and H42) of the base-pairing cytosine. For this reason, the amino protons of cytosines involved in base pairing are easy to identify in comparison to the guanosine and adenosine amino protons due to the strong NOE with the H5 of the cytosine. The imino-imino NOEs can be traced through each RNA secondary structure (**Fig. 7** and **Supplemental Fig. S9**).

GAGA tetraloop

Hairpin or stem-loop structures are pervasive in RNAs. RNA hairpins with a four nucleotide loop are known as tetraloops and are the most common size of loop (Antao and Tinoco, 1992). All RNA constructs were designed to contain a GAGA tetraloop with a U-A closing base pair (**Fig. 2**). The GAGA tetraloop was chosen due to its structural stability (Dale et al., 2000; Sheehy et al., 2010) and characteristic signals in the ^{1}H-^{1}H NOESY spectrum (**Fig. 8**). While GAGA, and more generally, GNRA (where N represents any nucleotide and R represents a purine) are commonly closed with C-G base pairs,(Brown et al., 2020; Legault et al., 1998; Pham et al., 2018; Vallurupalli and Moore, 2003) we chose to close our tetraloops with U-A base pairs to further broaden the chemical shift database. Currently, there are only two deposition in the BMRB with a GAGA tetraloop closed with a U-A base pair (D'Souza et al., 2004; Imai et al., 2016) and only two GNRA tetraloops closed with a U-A base pair (Cornilescu et al., 2016; D'Souza et al., 2004). The NOE walk of the GAGA tetraloop presents a unique pattern which was helpful to both confirm proper folding of the RNA and facilitate resonance assignments. The three-dimensional structure of the nucleotides in the GAGA tetraloop cause a dramatic change in the H1′ frequency of the

nucleotide following the loop (A14 in these constructs), out of the typical anomeric proton chemical shift range and to a smaller (upfield) chemical shift (Jucker et al., 1996).

Assignments and data deposition

In this work, we have unambiguously assigned all 330 aromatic C6-H6, C8-H8, and C2-H2 correlations. For the 66 adenosine residues, we have assigned 100% of the aromatic and anomeric protons (A-H2, A-H8, and A-H1′). We were able to assign 86% of the adenosine H2′ and 74% of adenosine H3′ protons (57/66 and 49/66, respectively). For the 90 guanosine residues, we have assigned 100% of the aromatic and anomeric protons (G-H8, and G-H1′). We were able to assign 89% of the guanosine H2′ and 86% of guanosine H3′ protons (80/90 and 77/90, respectively). For the 66 cytosine residues, we have assigned 100% of the aromatic and anomeric protons (C-H5, C-H6, and C-H1′). We were able to assign 89% of the cytosine H2′ and 77% of cytosine H3′ protons (59/66 and 51/66, respectively). Finally, for the 42 uracil residues, we have assigned 100% of the aromatic and anomeric protons (U-H5, U-H6, and U-H1′). We were able to assign 90% of the uracil H2′ and 90% of uracil H3′ protons (38/42 and 38/42, respectively). We assigned 100% of the imino protons (66 base paired guanosine G-H1 and 42 base paired uracil U-H3). Additionally, for the 66 base paired cytosines, we assigned 92.4% of the amino C-H41 and C-H42 protons.

Assigned chemical shifts along with raw NMR data have been deposited in the BMRB. RNA5: 50933, RNA7: 50932, RNA8: 50931, RNA21: 50930, RNA23: 50929, RNA24: 50928, RNA73: 50927, RNA74: 50926, RNA75: 50925, RNA89: 50924, RNA90: 50923, RNA91: 50922. The new data has also been used to update the training of chemical shift predictions in NMRFx Analyst (Marchant et al., 2019). The more complete database will immediately allow for better quality predictions in the use of the molecular network assignment tool integrated in NMRFx Analyst (Marchant et al., 2019). Similarly, these additional data will aid in machine learning approaches that predict RNA secondary structure (Zhang and Frank, 2020) and other RNA structural features from the assigned chemical shift data (Zhang et al., 2021).

Acknowledgements

This work was supported by National Science Foundation grant MCB-1942398 (to S.C.K.) and by National Institute of General Medical Sciences of the National Institutes of Health grant U54 GM 103297 (to B.A.J.).

Research reported in this publication was supported by the University of Michigan BioNMR Core Facility (U-M BioNMR). U-M BioNMR Core is grateful for support from U-M including the College of Literature, Sciences and Arts, Life Sciences Institute, College of Pharmacy and the Medical School along with the U-M Biosciences Initiative.

References

Aeschbacher, T., Schmidt, E., Blatter, M., Maris, C., Duss, O., Allain, F. H., Guntert, P., Schubert, M. (2013) Automated and assisted RNA resonance assignment using NMR chemical shift statistics. Nucleic Acids Res 41:e172. 10.1093/nar/gkt665

Antao, V. P., Tinoco, I., Jr. (1992) Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. Nucleic Acids Res 20:819-824. 10.1093/nar/20.4.819

Bahrami, A., Clos, L. J., 2nd, Markley, J. L., Butcher, S. E., Eghbalnia, H. R. (2012) RNA-PAIRS: RNA probabilistic assignment of imino resonance shifts. J Biomol NMR 52:289-302. 10.1007/s10858-012-9603-z

Barton, S., Heng, X., Johnson, B. A., Summers, M. F. (2013) Database proton NMR chemical shifts for RNA signal assignment and validation. J Biomol NMR 55:33-46. 10.1007/s10858-012-9683-9

Bax, A., Davis, D. G. (1985) MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. J Magn Reson 65:355-360. 10.1016/0022-2364(85)90018-6

Bax, A., Griffey, R. H., Hawking, B. L. (1983) Correlation of Proton and Nitrogen-15 Chemical Shifts by Multiple Quantum NMR. J Magn Reson 55:301-315. 10.1016/0022-2364(83)90241-X

Berman, H. M., Bhat, T. N., Bourne, P. E., Feng, Z., Gilliland, G., Weissig, H., Westbrook, J. (2000a) The Protein Data Bank and the challenge of structural genomics. Nat Struct Biol 7 Suppl:957-959. 10.1038/80734

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E. (2000b) The Protein Data Bank. Nucleic Acids Res 28:235-242. 10.1093/nar/28.1.235

Brown, J. D., Kharytonchyk, S., Chaudry, I., Iyer, A. S., Carter, H., Becker, G., Desai, Y., Glang, L., Choi, S. H., Singh, K., Lopresti, M. W., Orellana, M., Rodriguez, T., Oboh, U., Hijji, J., Ghinger, F. G., Stewart, K., Francis, D., Edwards, B., Chen, P., Case, D. A., Telesnitsky, A., Summers, M. F. (2020) Structural basis for transcriptional start site control of HIV-1 RNA fate. Science 368:413-417. 10.1126/science.aaz7959

Brown, J. D., Summers, M. F., Johnson, B. A. (2015) Prediction of hydrogen and carbon chemical shifts from RNA using database mining and support vector regression. J Biomol NMR 63:39-52. 10.1007/s10858-015-9961-4

Burnett, J. C., Rossi, J. J. (2012) RNA-based therapeutics: current progress and future prospects. Chem Biol 19:60-71. 10.1016/j.chembiol.2011.12.008

Corbett, K. S., Edwards, D. K., Leist, S. R., Abiona, O. M., Boyoglu-Barnum, S., Gillespie, R. A., Himansu, S., Schafer, A., Ziwawo, C. T., DiPiazza, A. T., Dinnon, K. H., Elbashir, S. M., Shaw, C. A., Woods, A., Fritch, E. J., Martinez, D. R., Bock, K. W., Minai, M., Nagata, B. M., Hutchinson, G. B., Wu, K., Henry, C., Bahl, K., Garcia-Dominguez, D., Ma, L., Renzi, I., Kong, W. P., Schmidt, S. D., Wang, L., Zhang, Y., Phung, E., Chang, L. A., Loomis, R. J., Altaras, N. E., Narayanan, E., Metkar, M., Presnyak, V., Liu, C., Louder, M. K., Shi, W., Leung, K., Yang, E. S., West, A., Gully, K. L., Stevens, L. J., Wang, N., Wrapp, D., Doria-Rose, N. A., Stewart-Jones, G., Bennett, H., Alvarado, G. S., Nason, M. C., Ruckwardt, T. J., McLellan, J. S., Denison, M. R., Chappell, J. D., Moore, I. N., Morabito, K. M., Mascola, J. R., Baric, R. S., Carfi, A., Graham, B. S. (2020) SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. Nature 586:567-571. 10.1038/s41586-020-2622-0

Cornilescu, G., Didychuk, A. L., Rodgers, M. L., Michael, L. A., Burke, J. E., Montemayor, E. J., Hoskins, A. A., Butcher, S. E. (2016) Structural Analysis of Multi-Helical RNAs by NMR-SAXS/WAXS: Application to the U4/U6 di-snRNA. J Mol Biol 428:777-789. 10.1016/j.jmb.2015.11.026

Cunningham, P. R., Ofengand, J. (1990) Use of inorganic pyrophosphatase to improve the yield of in vitro transcription reactions catalyzed by T7 RNA polymerase. Biotechniques 9:713-714.

D'Souza, V., Dey, A., Habib, D., Summers, M. F. (2004) NMR structure of the 101-nucleotide core encapsidation signal of the Moloney murine leukemia virus. J Mol Biol 337:427-442. 10.1016/j.jmb.2004.01.037

Dale, T., Smith, R., Serra, M. J. (2000) A test of the model to predict unusually stable RNA hairpin loop stability. RNA 6:608-615. 10.1017/s1355838200992495

Damase, T. R., Sukhovershin, R., Boada, C., Taraballi, F., Pettigrew, R. I., Cooke, J. P. (2021) The Limitless Future of RNA Therapeutics. Front Bioeng Biotechnol 9:628137. 10.3389/fbioe.2021.628137

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277-293. 10.1007/BF00197809

Ebrahimi, M., Rossi, P., Rogers, C., Harbison, G. S. (2001) Dependence of 13C NMR chemical shifts on conformations of rna nucleosides and nucleotides. J Magn Reson 150:1-9. 10.1006/jmre.2001.2314

Fares, C., Amata, I., Carlomagno, T. (2007) 13C-detection in RNA bases: revealing structure-chemical shift relationships. J Am Chem Soc 129:15814-15823. 10.1021/ja0727417

Frank, A. T., Bae, S. H., Stelzer, A. C. (2013) Prediction of RNA H-1 and C-13 Chemical Shifts: A Structure Based Approach. Journal of Physical Chemistry B 117:13497-13506. 10.1021/jp407254m

Frank, A. T., Law, S. M., Brooks, C. L., 3rd (2014) A simple and fast approach for predicting (1)H and (13)C chemical shifts: toward chemical shift-guided simulations of RNA. J Phys Chem B 118:12168-12175. 10.1021/jp508342x

Ghose, R., Marino, J. P., Wiberg, K. B., Prestegard, J. H. (1994) Dependence of 13C Chemical Shifts on Glycosidic Torsional Angles in Ribonucleic Acids. J. Am. Chem. Soc. 116:8827-8828. 10.1021/ja00098a060

Giessner-Prettre, C., Pullman, B. (1987) Quantum mechanical calculations of NMR chemical shifts in nucleic acids. Q Rev Biophys 20:113-172. 10.1017/s0033583500004169

Han, B., Liu, Y., Ginzinger, S. W., Wishart, D. S. (2011) SHIFTX2: significantly improved protein chemical shift prediction. J Biomol NMR 50:43-57. 10.1007/s10858-011-9478-4

Helmling, C., Keyhani, S., Sochor, F., Furtig, B., Hengesbach, M., Schwalbe, H. (2015) Rapid NMR screening of RNA secondary structure and binding. J Biomol NMR 63:67-76. 10.1007/s10858-015-9967-y

Howe, J. A., Wang, H., Fischmann, T. O., Balibar, C. J., Xiao, L., Galgoci, A. M., Malinverni, J. C., Mayhood, T., Villafania, A., Nahvi, A., Murgolo, N., Barbieri, C. M., Mann, P. A., Carr, D., Xia, E., Zuck, P., Riley, D., Painter, R. E., Walker, S. S., Sherborne, B., de Jesus, R., Pan, W., Plotkin, M. A., Wu, J., Rindgen, D., Cummings, J., Garlisi, C. G., Zhang, R., Sheth, P. R., Gill, C. J., Tang, H., Roemer, T. (2015) Selective small-molecule inhibition of an RNA structural element. Nature 526:672-677. 10.1038/nature15542

Hwang, T. L., Shaka, A. J. (1995) Water Suppression That Works - Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. J Magn Reson Ser A 112:275-279. DOI 10.1006/jmra.1995.1047

Imai, S., Kumar, P., Hellen, C. U., D'Souza, V. M., Wagner, G. (2016) An accurately preorganized IRES RNA structure enables eIF4G capture for initiation of viral translation. Nat Struct Mol Biol 23:859-864. 10.1038/nsmb.3280

Johnson, B. A., Blevins, R. A. (1994) NMR View: A computer program for the visualization and analysis of NMR data. J Biomol NMR 4:603-614. 10.1007/BF00404272

Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H., Pardi, A. (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops. J Mol Biol 264:968-980. 10.1006/jmbi.1996.0690

Kao, C., Zheng, M., Rudisser, S. (1999) A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. RNA 5:1268-1272. 10.1017/s1355838299991033

Krahenbuhl, B., Lukavsky, P., Wider, G. (2014) Strategy for automated NMR resonance assignment of RNA: application to 48-nucleotide K10. J Biomol NMR 59:231-240. 10.1007/s10858-014-9841-3

Kwok, C. K., Lam, S. L. (2013) NMR proton chemical shift prediction of T center dot T mismatches in B-DNA duplexes. Journal of Magnetic Resonance 234:184-189. 10.1016/j.jmr.2013.06.022

Lam, S. L. (2007) DSHIFT: a web server for predicting DNA chemical shifts. Nucleic Acids Research 35:W713-W717. 10.1093/nar/gkm320

Lam, S. L., Lai, K. F., Chi, L. M. (2007) Proton chemical shift prediction of A.A mismatches in B-DNA duplexes. Journal of Magnetic Resonance 187:105-111. 10.1016/j.jmr.2007.04.005

Legault, P., Hoogstraten, C. G., Metlitzky, E., Pardi, A. (1998) Order, dynamics and metal-binding in the lead-dependent ribozyme. J Mol Biol 284:325-335. 10.1006/jmbi.1998.2181

Marchant, J., Summers, M. F., Johnson, B. A. (2019) Assigning NMR spectra of RNA, peptides and small organic molecules using molecular network visualization software. J Biomol NMR 10.1007/s10858-019-00271-3

Meiler, J. (2003) PROSHIFT: protein chemical shift prediction using artificial neural networks. J Biomol NMR 26:25-37. 10.1023/a:1023060720156

Meyer, S. M., Williams, C. C., Akahori, Y., Tanaka, T., Aikawa, H., Tong, Y., Childs-Disney, J. L., Disney, M. D. (2020) Small molecule recognition of disease-relevant RNA structures. Chem Soc Rev 49:7167-7199. 10.1039/d0cs00560f

Ng, K. S., Lam, S. L. (2015) NMR proton chemical shift prediction of C.C mismatches in B-DNA. Journal of Magnetic Resonance 252:87-93. 10.1016/j.jmr.2015.01.005

Norris, M., Fetler, B., Marchant, J., Johnson, B. A. (2016) NMRFx Processor: a cross-platform NMR data processing program. J Biomol NMR 65:205-216. 10.1007/s10858-016-0049-6

Papasaikas, P., Valcarcel, J. (2016) The Spliceosome: The Ultimate RNA Chaperone and Sculptor. Trends Biochem Sci 41:33-45. 10.1016/j.tibs.2015.11.003

Pardi, N., Hogan, M. J., Porter, F. W., Weissman, D. (2018) mRNA vaccines - a new era in vaccinology. Nat Rev Drug Discov 17:261-279. 10.1038/nrd.2017.243

Pham, V. V., Salguero, C., Khan, S. N., Meagher, J. L., Brown, W. C., Humbert, N., de Rocquigny, H., Smith, J. L., D'Souza, V. M. (2018) HIV-1 Tat interactions with cellular 7SK and viral TAR RNAs identifies dual structural mimicry. Nat Commun 9:4266. 10.1038/s41467-018-06591-6

Popenda, M., Blazewicz, M., Szachniuk, M., Adamiak, R. W. (2008) RNA FRABASE version 1.0: an engine with a database to search for the three-dimensional fragments within RNA structures. Nucleic Acids Res 36:D386-391. 10.1093/nar/gkm786

Popenda, M., Szachniuk, M., Blazewicz, M., Wasik, S., Burke, E. K., Blazewicz, J., Adamiak, R. W. (2010) RNA FRABASE 2.0: an advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures. BMC Bioinformatics 11:231. 10.1186/1471-2105-11-231

Rodnina, M. V. (2016) The ribosome in action: Tuning of translational efficiency and protein folding. Protein Sci 25:1390-1406. 10.1002/pro.2950

Rossi, P., Harbison, G. S. (2001) Calculation of 13C chemical shifts in rna nucleosides: structure-13C chemical shift relationships. J Magn Reson 151:1-8. 10.1006/jmre.2001.2350

Sheehy, J. P., Davis, A. R., Znosko, B. M. (2010) Thermodynamic characterization of naturally occurring RNA tetraloops. RNA 16:417-429. 10.1261/rna.1773110

Sripakdeevong, P., Cevec, M., Chang, A. T., Erat, M. C., Ziegeler, M., Zhao, Q., Fox, G. E., Gao, X., Kennedy, S. D., Kierzek, R., Nikonowicz, E. P., Schwalbe, H., Sigel, R. K., Turner, D. H., Das, R. (2014) Structure determination of noncanonical RNA motifs guided by (1)H NMR chemical shifts. Nat Methods 11:413-416. 10.1038/nmeth.2876

Theimer, C. A., Feigon, J. (2006) Structure and function of telomerase RNA. Curr Opin Struct Biol 16:307-318. 10.1016/j.sbi.2006.05.005

Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C. F., Tolmie, D. E., Kent Wenger, R., Yao, H., Markley, J. L. (2008) BioMagResBank. Nucleic Acids Res 36:D402-408. 10.1093/nar/gkm957

Vallurupalli, P., Moore, P. B. (2003) The solution structure of the loop E region of the 5S rRNA from spinach chloroplasts. J Mol Biol 325:843-856. 10.1016/s0022-2836(02)01270-6

Wang, Y., Han, G., Jiang, X., Yuwen, T., Xue, Y. (2021) Chemical shift prediction of RNA imino groups: application toward characterizing RNA excited states. Nat Commun 12:1595. 10.1038/s41467-021-21840-x

Warner, K. D., Hajdin, C. E., Weeks, K. M. (2018) Principles for targeting RNA with drug-like small molecules. Nat Rev Drug Discov 17:547-558. 10.1038/nrd.2018.93

Winkle, M., El-Daly, S. M., Fabbri, M., Calin, G. A. (2021) Noncoding RNA therapeutics - challenges and potential solutions. Nat Rev Drug Discov 10.1038/s41573-021-00219-z

Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., Sykes, B. D. (1995) 1H, 13C and 15N chemical shift referencing in biomolecular NMR. J Biomol NMR 6:135-140. 10.1007/BF00211777

Zhang, K., Abdallah, K., Ajmera, P., Finos, K., Looka, A., Mekhael, J., Frank, A. T. (2021) CS-Annotate: A Tool for Using NMR Chemical Shifts to Annotate RNA Structure. J Chem Inf Model 61:1545-1549. 10.1021/acs.jcim.1c00006

Zhang, K., Frank, A. T. (2020) Conditional Prediction of Ribonucleic Acid Secondary Structure Using Chemical Shifts. J Phys Chem B 124:470-478. 10.1021/acs.jpcb.9b09814

Figure Legends

Figure 1. Overview of RNA sequences with missing assignments. Four groups of RNAs were identified A) Group 1, RNAs 5-8, B) Group 2, RNAs 21-24, C) Group 3, RNAs 73-76, D) Group 4, RNAs 89-92. Red nucleotide indicates the nucleotide with missing chemical shifts. The common triple of base pairs for each group is highlighted in a colored box. Sequence variations at flanking nucleotides were examined. The occurrence of aromatic proton and carbon chemical shifts in the BMRB for each RNA are indicated. Bold font is used to highlight the RNAs from each group with missing chemical shifts and are the RNAs further examined in this study. Flanking sequences for each RNA are shown. R, purine; Y, pyrimidine.

Figure 2. RNA constructs used in this study. Red nucleotides indicate the variable sequence relative to a single example for each group. The common triple of base pairs for each group is highlighted in a colored box and the central nucleotide is in bold font.

Figure 3. RNA7 chemical shift assignments. A) ^{1}H-^{1}H NOESY, B) ^{1}H-^{13}C HMQC, C) Summary of the sequence, secondary structure, and assignment validation for RNA7. The secondary structure is shown in Vienna format. NMRViewJ assignment summary to validate proton (H6/H8, H5/H2, H1′, H2′, H3′) and carbon (C6/C8, C2) assignments (bottom). Assigned chemical shifts for specific atoms are indicated with open and filled circles. The vertical offset of the circles indicates the deviation from the predicted values for that atom. Filled circles indicate that there are chemical shifts for atoms with the same set of attributes in the BMRB. Open circles indicate atoms that have a prediction, but for which no exact matches of the attributes are available in the BMRB. Grey boxes represent atoms that are not present in a given base.

Figure 4. RNA23 chemical shift assignments. A) ^{1}H-^{1}H NOESY, B) ^{1}H-^{13}C HMQC, C) Summary of the sequence, secondary structure, and assignment validation for RNA23. The secondary structure is shown in Vienna format. NMRViewJ assignment summary to validate proton (H6/H8, H5/H2, H1′, H2′, H3′) and carbon (C6/C8, C2) assignments (bottom). Assigned chemical shifts for specific atoms are indicated with open and filled circles. The vertical offset of the circles indicates the deviation from the predicted values for that atom. Filled circles indicate

that there are chemical shifts for atoms with the same set of attributes in the BMRB. Open circles indicate atoms that have a prediction, but for which no exact matches of the attributes are available in the BMRB. Grey boxes represent atoms that are not present in a given base.

Figure 5. RNA73 chemical shift assignments. A) ^{1}H-^{1}H NOESY, B) ^{1}H-^{13}C HMQC, C) Summary of the sequence, secondary structure, and assignment validation for RNA73. The secondary structure is shown in Vienna format. NMRViewJ assignment summary to validate proton (H6/H8, H5/H2, H1′, H2′, H3′) and carbon (C6/C8, C2) assignments (bottom). Assigned chemical shifts for specific atoms are indicated with open and filled circles. The vertical offset of the circles indicates the deviation from the predicted values for that atom. Filled circles indicate that there are chemical shifts for atoms with the same set of attributes in the BMRB. Open circles indicate atoms that have a prediction, but for which no exact matches of the attributes are available in the BMRB. Grey boxes represent atoms that are not present in a given base.

Figure 6. RNA89 chemical shift assignments. A) ^{1}H-^{1}H NOESY, B) ^{1}H-^{13}C HMQC, C) Summary of the sequence, secondary structure, and assignment validation for RNA89. The secondary structure is shown in Vienna format. NMRViewJ assignment summary to validate proton (H6/H8, H5/H2, H1′, H2′, H3′) and carbon (C6/C8, C2) assignments (bottom). Assigned chemical shifts for specific atoms are indicated with open and filled circles. The vertical offset of the circles indicates the deviation from the predicted values for that atom. Filled circles indicate that there are chemical shifts for atoms with the same set of attributes in the BMRB. Open circles indicate atoms that have a prediction, but for which no exact matches of the attributes are available in the BMRB. Grey boxes represent atoms that are not present in a given base.

Figure 7. Assigned imino proton NOESY spectra for A) RNA7, B) RNA23, C) RNA73, and D) RNA89. Secondary structures for each RNA are inset.

Figure 8. ^{1}H-^{1}H NOESY spectrum of RNA24, highlighting the chemical shift assignments for the GAGA tetraloop. Tetraloop resonances are connected with green lines, resonances from flanking sequences are connected with purple lines. Dashed lines are used in crowded regions to indicate resonance assignment.

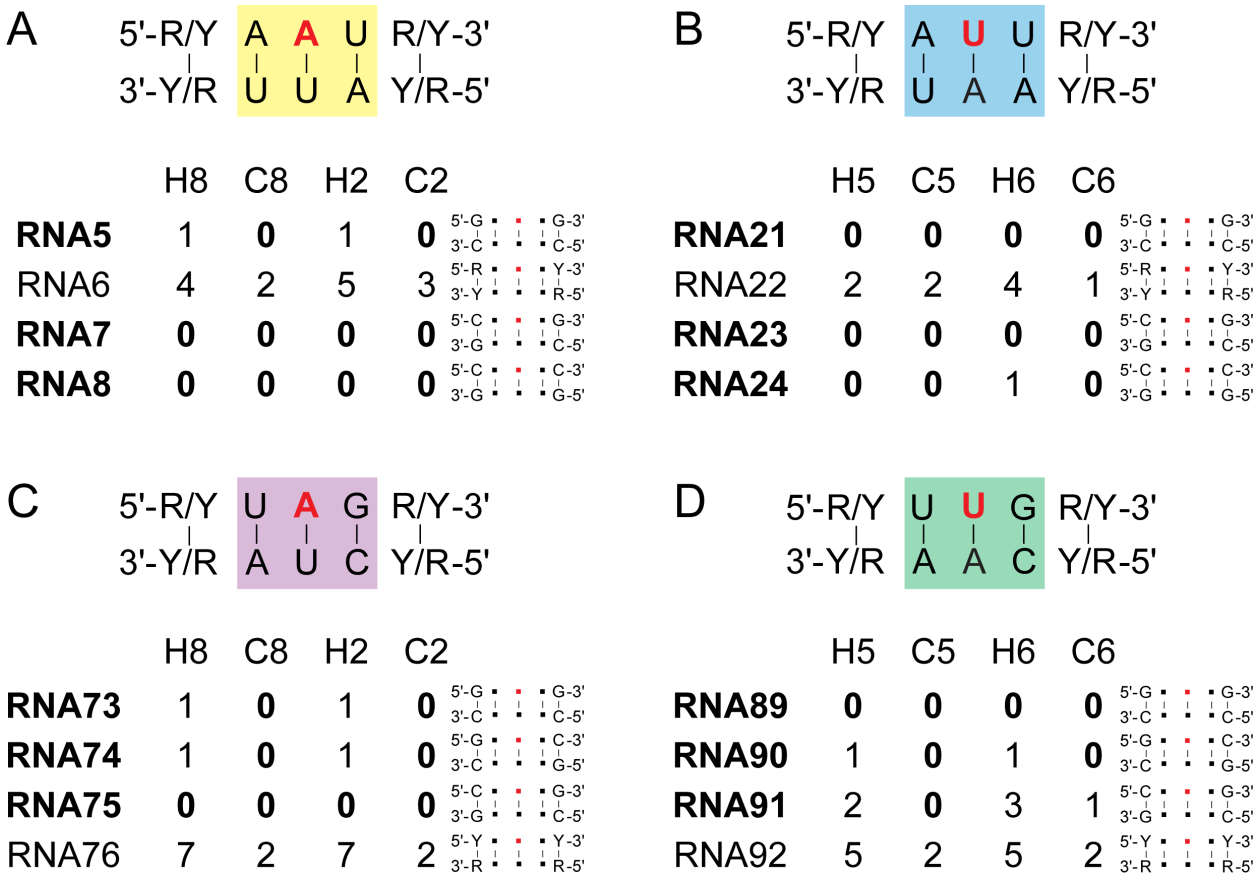

A
5'-R/Y A A U R/Y-3'
3'-Y/R U U A Y/R-5'
H8 C8 H2 C2
RNA5 1 0 1 0
RNA6 4 2 5 3
RNA7 0 0 0 0
RNA8 0 0 0 0
B
5'-R/Y A U U R/Y-3'
3'-Y/R U A A Y/R-5'
H5 C5 H6 C6
RNA21 0 0 0 0
RNA22 2 2 4 1
RNA23 0 0 0 0
RNA24 0 0 1 0
C
5'-R/Y U A G R/Y-3'
3'-Y/R A U C Y/R-5'
H8 C8 H2 C2
RNA73 1 0 1 0
RNA74 1 0 1 0
RNA75 0 0 0 0
RNA76 7 2 7 2
D
5'-R/Y U U G R/Y-3'
3'-Y/R A A C Y/R-5'
H5 C5 H6 C6
RNA89 0 0 0 0
RNA90 1 0 1 0
RNA91 2 0 3 1
RNA92 5 2 5 2

Figure 1

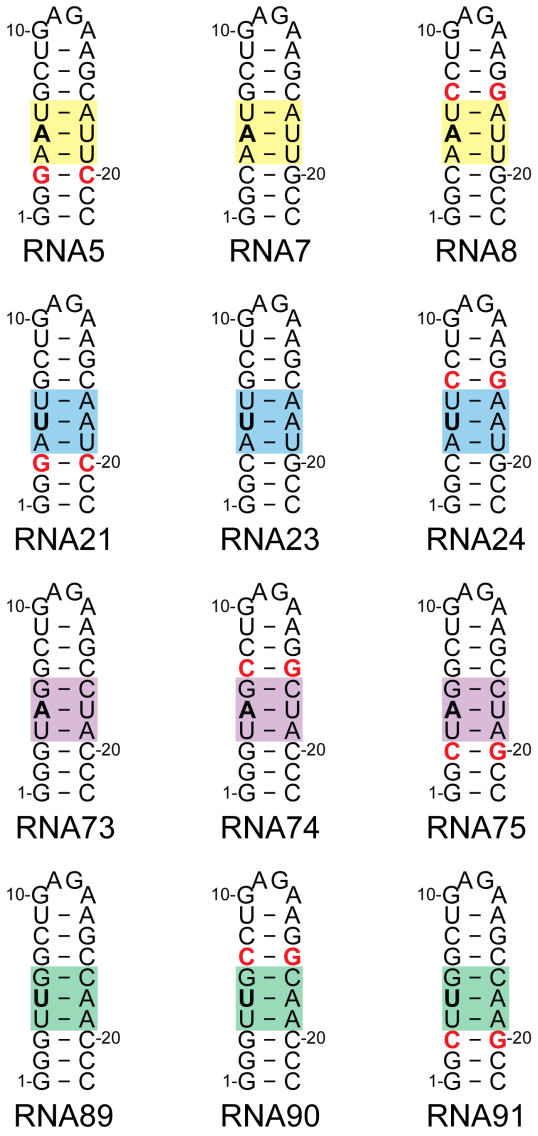

RNA5
RNA7
RNA8
RNA21
RNA23
RNA24
RNA73
RNA74
RNA75
RNA89
RNA90
RNA91

Figure 2

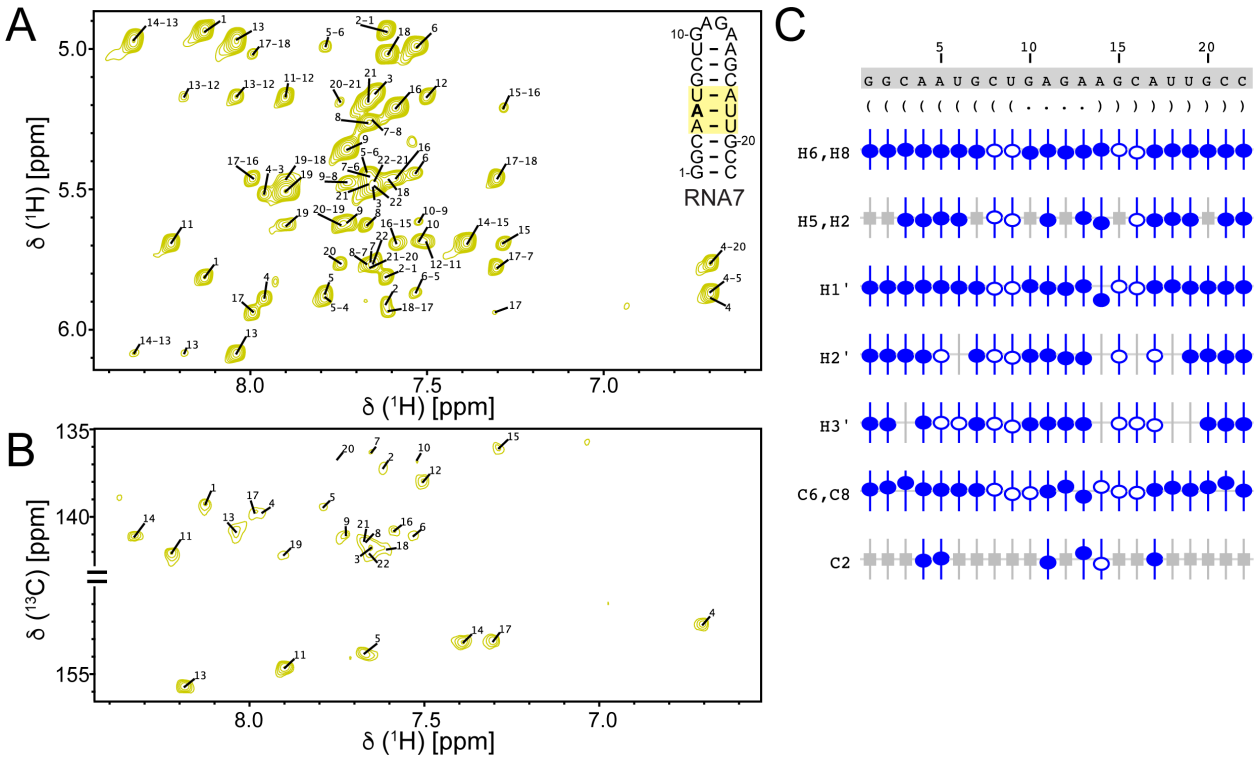

A
B
C
δ (¹H) [ppm]
δ (¹³C) [ppm]
8.0
7.5
7.0
5.0
5.5
6.0
135
140
155
RNA7
H6,H8
H5,H2
H1'
H2'
H3'
C6,C8
C2
G G C A A U G C U G A G A A G C A U U G C C
((((((((. . . .)))))))))

Figure 3

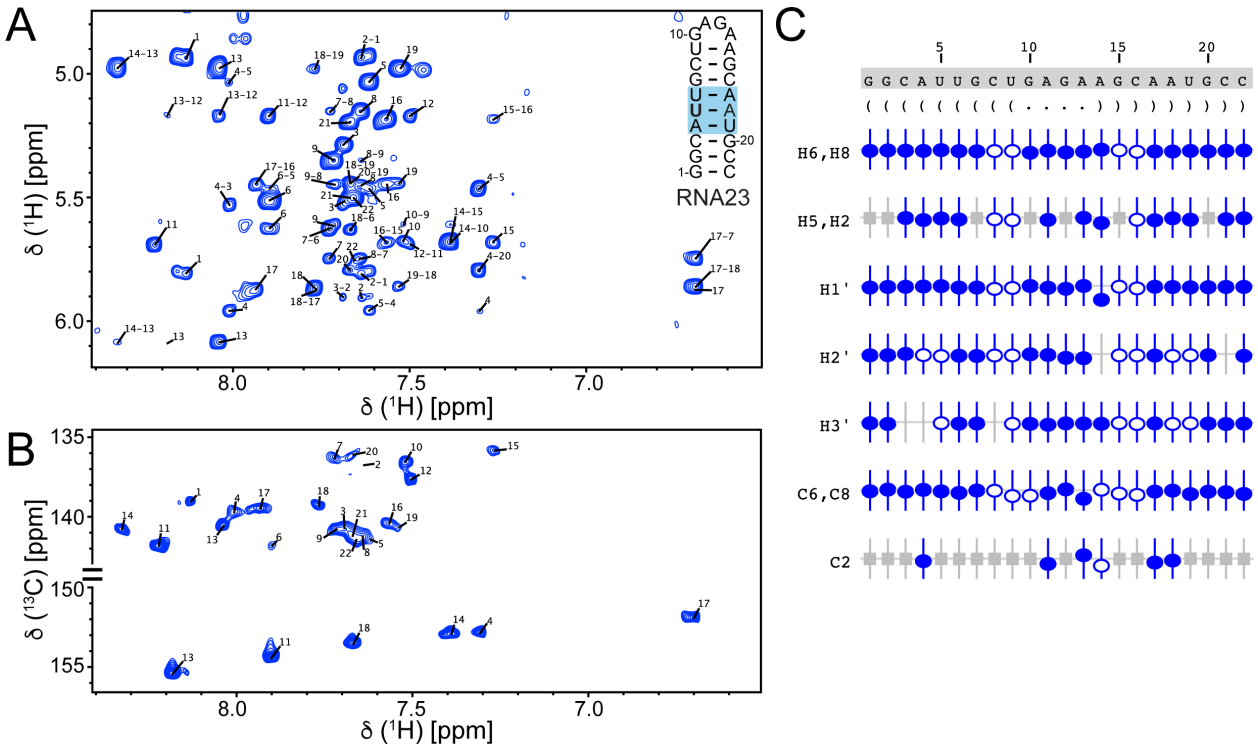

A
B
C
δ (¹H) [ppm]
δ (¹³C) [ppm]
RNA23
G G C A U U G C U G A G A A G C A A U G C C
(((((((((. . . .)))))))))
H6,H8
H5,H2
H1'
H2'
H3'
C6,C8
C2

Figure 4

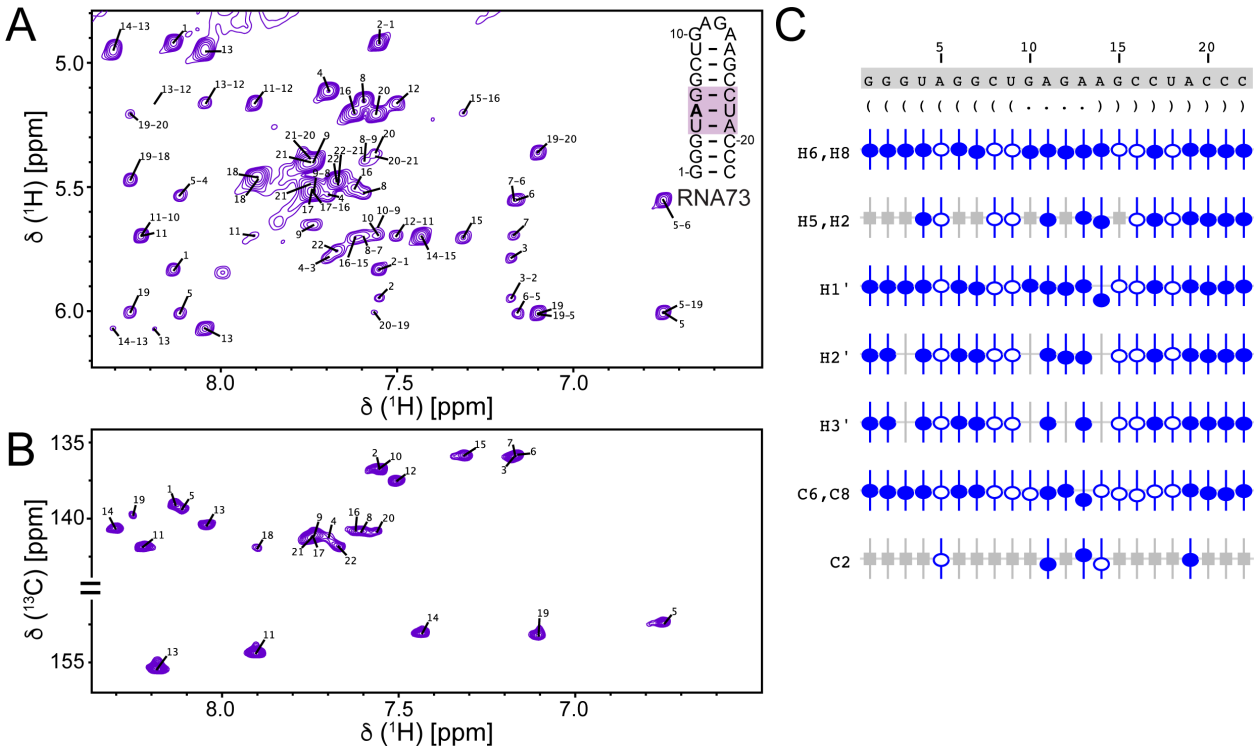

A
B
C
δ (¹H) [ppm]
δ (¹³C) [ppm]
RNA73
G G G U A G G C U G A G A A G C C U A C C C
(((((((((. . . .)))))))))
H6,H8
H5,H2
H1'
H2'
H3'
C6,C8
C2

Figure 5

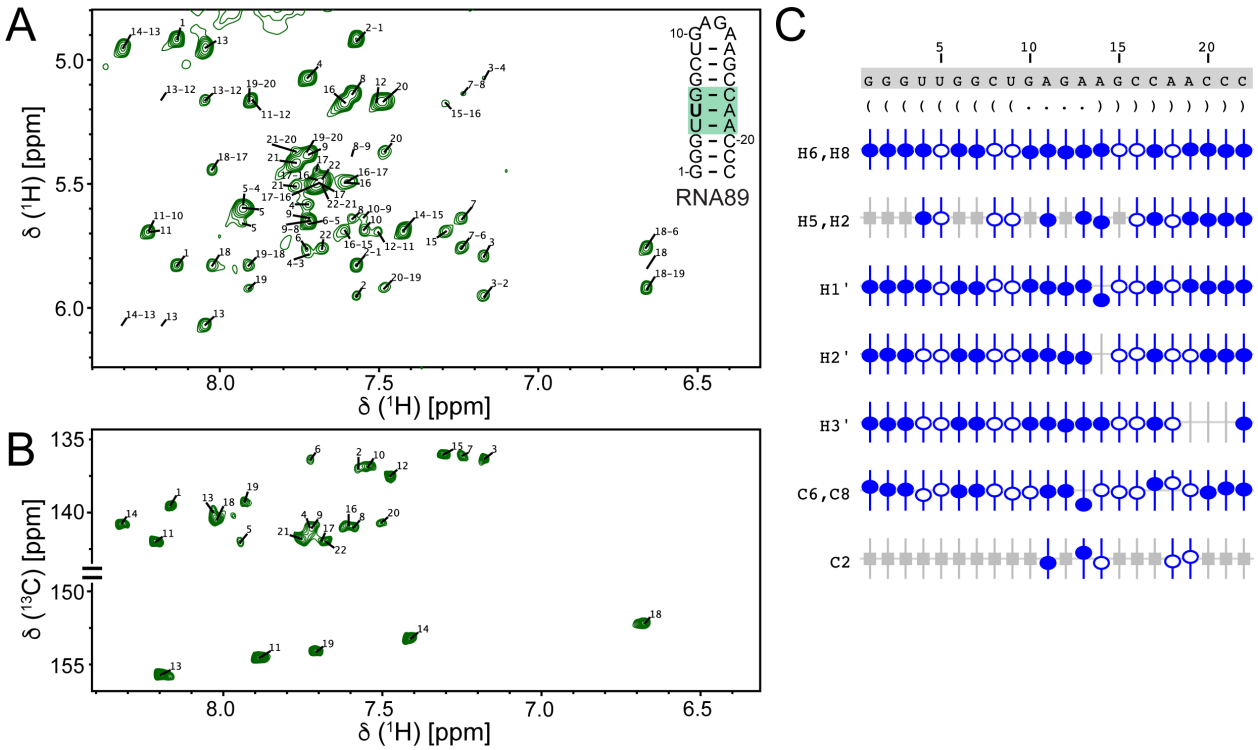

A
B
C
δ (¹H) [ppm]
δ (¹³C) [ppm]
RNA89
H6,H8
H5,H2
H1'
H2'
H3'
C6,C8
C2
G G G U U G G C U G A G A A G C C A A C C C
(((((((((. . . .)))))))))

Figure 6

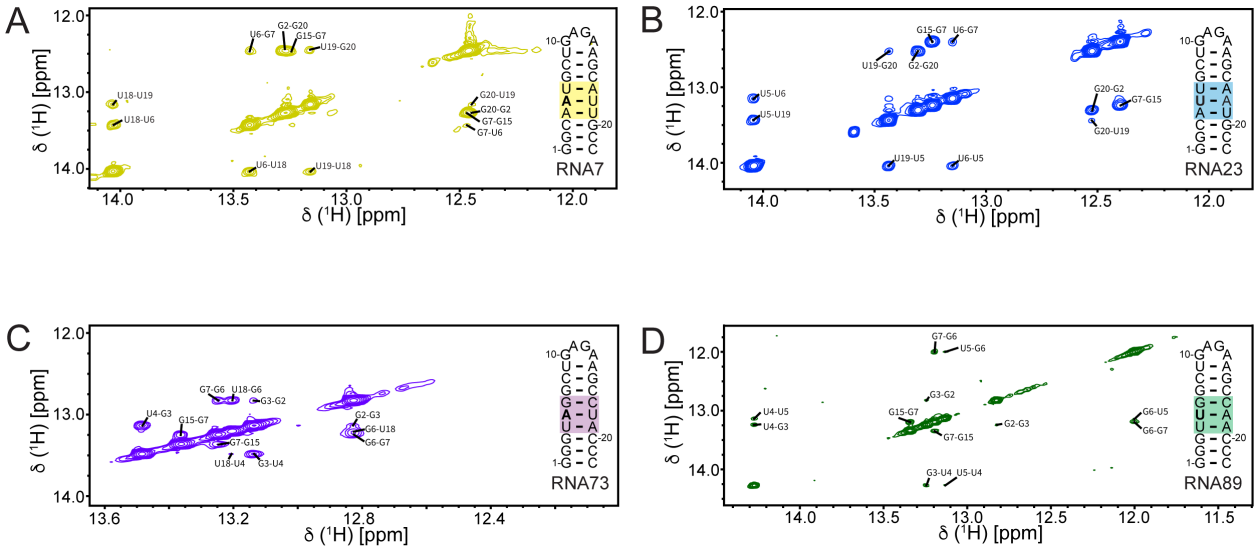
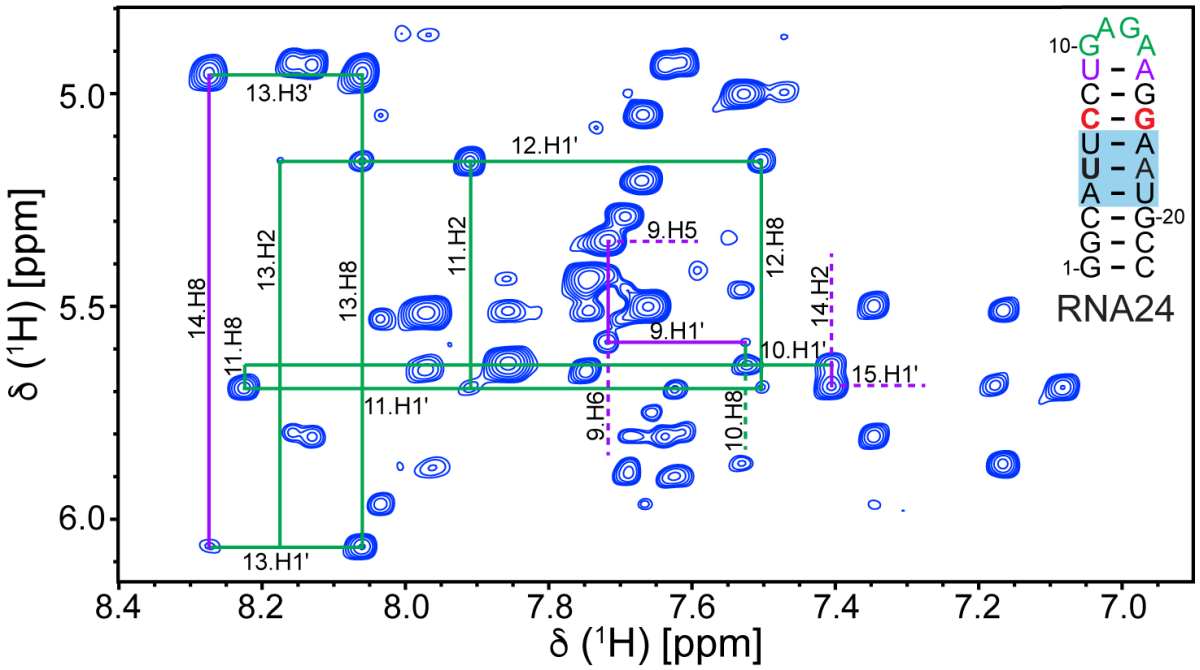


Figure 7



508

509 **Figure 8**

510

Supporting information

NMR chemical shift assignments of RNA oligonucleotides to expand the RNA chemical shift database

Yaping Liu[1#], Anita Kotar[1,5#], Tracy L. Hodges[1], Kyrillos Abdallah[1], Mallak H. Taleb[1], Brayden A. Bitterman[1], Sara Jaime[1], Kyle J. Schaubroeck[1], Ethan Mathew[1], Nicholas W. Morgenstern[1], Anthony Lohmeier[1], Jordan L. Page[1], Matt Ratanapanichkich[1], Grace Arhin[1], Breanna L. Johnson[1], Stanislav Cherepanov[1], Stephen C. Moss[1], Gisselle Zuniga[1], Nicholas J. Tilson[1], Zoe C. Yeoh[2], Bruce A. Johnson[3], and Sarah C. Keane*[1,4]

[1] Biophysics Program, University of Michigan, 930 N. University Avenue, Ann Arbor, MI 48109, USA

[2] Department of Biological Chemistry, University of Michigan, 1150 W. Medical Center Drive, Ann Arbor, MI 48109, USA

[3] Structural Biology Initiative, CUNY Advanced Science Research Center, 85 St. Nicholas Terrace, New York, NY, 10031, USA

[4] Department of Chemistry, University of Michigan, 930 N. University Avenue, Ann Arbor, MI 48109, USA

[5] Current Address: Slovenian NMR Centre, National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana, Slovenia

*Corresponding author: sckeane@umich.edu

Table S1. Summary of 5-base pair RNA sequences and their reported chemical shifts.

	Sequence[a]	Number of measured chemical shifts								Complete chemical shift data?
		H8	C8	H2	C2	H5	C5	H6	C6	
RNA1	RY.AU.AU.AU.RY	4	4	4	4	0	0	0	0	TRUE
RNA2	RY.AU.AU.AU.YR	5	3	5	3	0	0	0	0	TRUE
RNA3	YR.AU.AU.AU.RY	0	0	0	0	0	0	0	0	FALSE
RNA4	YR.AU.AU.AU.YR	3	3	3	3	0	0	0	0	TRUE
RNA5	RY.AU.AU.UA.RY	1	0	1	0	0	0	0	0	FALSE
RNA6	RY.AU.AU.UA.YR	4	2	5	3	0	0	0	0	TRUE
RNA7	YR.AU.AU.UA.RY	0	0	0	0	0	0	0	0	FALSE
RNA8	YR.AU.AU.UA.YR	0	0	0	0	0	0	0	0	FALSE
RNA9	RY.AU.AU.GC.RY	6	3	7	3	0	0	0	0	TRUE
RNA10	RY.AU.AU.GC.YR	4	3	4	3	0	0	0	0	TRUE
RNA11	YR.AU.AU.GC.RY	1	0	2	1	0	0	0	0	FALSE
RNA12	YR.AU.AU.GC.YR	1	1	1	1	0	0	0	0	TRUE
RNA13	RY.AU.AU.CG.RY	5	4	5	4	0	0	0	0	TRUE
RNA14	RY.AU.AU.CG.YR	5	4	5	4	0	0	0	0	TRUE
RNA15	YR.AU.AU.CG.RY	2	1	2	1	0	0	0	0	TRUE
RNA16	YR.AU.AU.CG.YR	1	0	1	0	0	0	0	0	FALSE
RNA17	RY.UA.AU.AU.RY	1	0	1	0	0	0	0	0	FALSE
RNA18	RY.UA.AU.AU.YR	0	0	0	0	0	0	0	0	FALSE
RNA19	YR.UA.AU.AU.RY	1	1	1	1	0	0	0	0	TRUE
RNA20	YR.UA.AU.AU.YR	2	1	2	1	0	0	0	0	TRUE
RNA21	RY.UA.AU.UA.RY	9	5	9	8	0	0	0	0	TRUE
RNA22	RY.UA.AU.UA.YR	6	4	6	4	0	0	0	0	TRUE
RNA23	YR.UA.AU.UA.RY	3	2	3	3	0	0	0	0	TRUE
RNA24	YR.UA.AU.UA.YR	1	0	1	0	0	0	0	0	FALSE
RNA25	RY.UA.AU.GC.RY	1	0	1	0	0	0	0	0	FALSE
RNA26	RY.UA.AU.GC.YR	1	0	1	0	0	0	0	0	FALSE
RNA27	YR.UA.AU.GC.RY	0	0	0	0	0	0	0	0	FALSE
RNA28	YR.UA.AU.GC.YR	7	2	7	2	0	0	0	0	TRUE
RNA29	RY.UA.AU.CG.RY	0	0	0	0	0	0	0	0	FALSE
RNA30	RY.UA.AU.CG.YR	6	5	6	6	0	0	0	0	TRUE
RNA31	YR.UA.AU.CG.RY	1	1	1	1	0	0	0	0	TRUE
RNA32	YR.UA.AU.CG.YR	1	1	1	1	0	0	0	0	TRUE
RNA33	RY.GC.AU.AU.RY	9	6	10	6	0	0	0	0	TRUE
RNA34	RY.GC.AU.AU.YR	7	3	7	4	0	0	0	0	TRUE
RNA35	YR.GC.AU.AU.RY	2	0	2	0	0	0	0	0	FALSE
RNA36	YR.GC.AU.AU.YR	0	0	0	0	0	0	0	0	FALSE
RNA37	RY.GC.AU.UA.RY	2	2	2	2	0	0	0	0	TRUE
RNA38	RY.GC.AU.UA.YR	8	4	9	4	0	0	0	0	TRUE
RNA39	YR.GC.AU.UA.RY	3	2	3	2	0	0	0	0	TRUE
RNA40	YR.GC.AU.UA.YR	5	3	6	3	0	0	0	0	TRUE
RNA41	RY.GC.AU.GC.RY	8	4	9	4	0	0	0	0	TRUE

RNA42	RY.GC.AU.GC.YR	13	10	13	12	0	0	0	0	TRUE
RNA43	YR.GC.AU.GC.RY	5	1	5	1	0	0	0	0	TRUE
RNA44	YR.GC.AU.GC.YR	2	1	2	1	0	0	0	0	TRUE
RNA45	RY.GC.AU.CG.RY	8	7	8	6	0	0	0	0	TRUE
RNA46	RY.GC.AU.CG.YR	13	6	12	5	0	0	0	0	TRUE
RNA47	YR.GC.AU.CG.RY	0	0	0	0	0	0	0	0	FALSE
RNA48	YR.GC.AU.CG.YR	3	2	3	2	0	0	0	0	TRUE
RNA49	RY.CG.AU.AU.RY	5	2	5	2	0	0	0	0	TRUE
RNA50	RY.CG.AU.AU.YR	2	2	3	2	0	0	0	0	TRUE
RNA51	YR.CG.AU.AU.RY	2	2	1	1	0	0	0	0	TRUE
RNA52	YR.CG.AU.AU.YR	0	0	0	0	0	0	0	0	FALSE
RNA53	RY.CG.AU.UA.RY	7	5	7	6	0	0	0	0	TRUE
RNA54	RY.CG.AU.UA.YR	0	0	0	0	0	0	0	0	FALSE
RNA55	YR.CG.AU.UA.RY	0	0	0	0	0	0	0	0	FALSE
RNA56	YR.CG.AU.UA.YR	4	4	4	4	0	0	0	0	TRUE
RNA57	RY.CG.AU.GC.RY	9	3	9	3	0	0	0	0	TRUE
RNA58	RY.CG.AU.GC.YR	8	8	7	7	0	0	0	0	TRUE
RNA59	YR.CG.AU.GC.RY	7	3	7	1	0	0	0	0	TRUE
RNA60	YR.CG.AU.GC.YR	5	3	5	2	0	0	0	0	TRUE
RNA61	RY.CG.AU.CG.RY	6	5	6	5	0	0	0	0	TRUE
RNA62	RY.CG.AU.CG.YR	13	10	13	10	0	0	0	0	TRUE
RNA63	YR.CG.AU.CG.RY	0	0	0	0	0	0	0	0	FALSE
RNA64	YR.CG.AU.CG.YR	4	1	4	1	0	0	0	0	TRUE
RNA65	RY.AU.UA.AU.RY	0	0	0	0	1	0	1	0	FALSE
RNA66	RY.AU.UA.AU.YR	0	0	0	0	3	2	5	2	TRUE
RNA67	YR.AU.UA.AU.RY	0	0	0	0	1	1	2	0	FALSE
RNA68	YR.AU.UA.AU.YR	0	0	0	0	7	4	8	3	TRUE
RNA69	RY.AU.UA.UA.RY	0	0	0	0	0	0	0	0	FALSE
RNA70	RY.AU.UA.UA.YR	0	0	0	0	2	2	4	1	TRUE
RNA71	YR.AU.UA.UA.RY	0	0	0	0	0	0	0	0	FALSE
RNA72	YR.AU.UA.UA.YR	0	0	0	0	0	0	1	0	FALSE
RNA73	RY.AU.UA.GC.RY	0	0	0	0	3	3	3	3	TRUE
RNA74	RY.AU.UA.GC.YR	0	0	0	0	0	0	0	0	FALSE
RNA75	YR.AU.UA.GC.RY	0	0	0	0	0	0	0	0	FALSE
RNA76	YR.AU.UA.GC.YR	0	0	0	0	7	7	7	6	TRUE
RNA77	RY.AU.UA.CG.RY	0	0	0	0	8	4	8	7	TRUE
RNA78	RY.AU.UA.CG.YR	0	0	0	0	11	9	12	8	TRUE
RNA79	YR.AU.UA.CG.RY	0	0	0	0	4	3	4	3	TRUE
RNA80	YR.AU.UA.CG.YR	0	0	0	0	2	2	2	2	TRUE
RNA81	RY.UA.UA.AU.RY	0	0	0	0	2	1	2	1	TRUE
RNA82	RY.UA.UA.AU.YR	0	0	0	0	0	0	0	0	FALSE
RNA83	YR.UA.UA.AU.RY	0	0	0	0	1	1	1	1	TRUE
RNA84	YR.UA.UA.AU.YR	0	0	0	0	1	0	1	0	FALSE
RNA85	RY.UA.UA.UA.RY	0	0	0	0	2	2	2	2	TRUE

RNA86	RY.UA.UA.UA.YR	0	0	0	0	5	3	5	3	TRUE
RNA87	YR.UA.UA.UA.RY	0	0	0	0	0	0	0	0	FALSE
RNA88	YR.UA.UA.UA.YR	0	0	0	0	5	3	5	4	TRUE
RNA89	RY.UA.UA.GC.RY	0	0	0	0	0	0	0	0	FALSE
RNA90	RY.UA.UA.GC.YR	0	0	0	0	1	0	1	0	FALSE
RNA91	YR.UA.UA.GC.RY	0	0	0	0	2	0	3	1	FALSE
RNA92	YR.UA.UA.GC.YR	0	0	0	0	5	2	5	2	TRUE
RNA93	RY.UA.UA.CG.RY	0	0	0	0	0	0	0	0	FALSE
RNA94	RY.UA.UA.CG.YR	0	0	0	0	5	4	6	3	TRUE
RNA95	YR.UA.UA.CG.RY	0	0	0	0	4	0	4	1	FALSE
RNA96	YR.UA.UA.CG.YR	0	0	0	0	10	7	10	6	TRUE
RNA97	RY.GC.UA.AU.RY	0	0	0	0	1	1	1	1	TRUE
RNA98	RY.GC.UA.AU.YR	0	0	0	0	6	5	6	4	TRUE
RNA99	YR.GC.UA.AU.RY	0	0	0	0	1	0	1	0	FALSE
RNA100	YR.GC.UA.AU.YR	0	0	0	0	1	1	1	1	TRUE
RNA101	RY.GC.UA.UA.RY	0	0	0	0	1	0	1	0	FALSE
RNA102	RY.GC.UA.UA.YR	0	0	0	0	4	3	4	3	TRUE
RNA103	YR.GC.UA.UA.RY	0	0	0	0	2	1	2	1	TRUE
RNA104	YR.GC.UA.UA.YR	0	0	0	0	4	4	5	4	TRUE
RNA105	RY.GC.UA.GC.RY	0	0	0	0	5	1	5	1	TRUE
RNA106	RY.GC.UA.GC.YR	0	0	0	0	12	9	12	9	TRUE
RNA107	YR.GC.UA.GC.RY	0	0	0	0	2	2	2	2	TRUE
RNA108	YR.GC.UA.GC.YR	0	0	0	0	3	3	3	3	TRUE
RNA109	RY.GC.UA.CG.RY	0	0	0	0	2	1	2	1	TRUE
RNA110	RY.GC.UA.CG.YR	0	0	0	0	15	8	16	9	TRUE
RNA111	YR.GC.UA.CG.RY	0	0	0	0	0	0	0	0	FALSE
RNA112	YR.GC.UA.CG.YR	0	0	0	0	6	6	6	6	TRUE
RNA113	RY.CG.UA.AU.RY	0	0	0	0	7	2	7	2	TRUE
RNA114	RY.CG.UA.AU.YR	0	0	0	0	1	1	1	1	TRUE
RNA115	YR.CG.UA.AU.RY	0	0	0	0	0	0	0	0	FALSE
RNA116	YR.CG.UA.AU.YR	0	0	0	0	1	1	1	1	TRUE
RNA117	RY.CG.UA.UA.RY	0	0	0	0	1	0	1	1	FALSE
RNA118	RY.CG.UA.UA.YR	0	0	0	0	4	3	4	2	TRUE
RNA119	YR.CG.UA.UA.RY	0	0	0	0	1	0	3	0	FALSE
RNA120	YR.CG.UA.UA.YR	0	0	0	0	6	4	6	2	TRUE
RNA121	RY.CG.UA.GC.RY	0	0	0	0	4	2	4	3	TRUE
RNA122	RY.CG.UA.GC.YR	0	0	0	0	5	5	5	5	TRUE
RNA123	YR.CG.UA.GC.RY	0	0	0	0	9	6	9	5	TRUE
RNA124	YR.CG.UA.GC.YR	0	0	0	0	11	5	11	5	TRUE
RNA125	RY.CG.UA.CG.RY	0	0	0	0	2	1	2	1	TRUE
RNA126	RY.CG.UA.CG.YR	0	0	0	0	16	12	16	12	TRUE
RNA127	YR.CG.UA.CG.RY	0	0	0	0	2	0	2	0	FALSE
RNA128	YR.CG.UA.CG.YR	0	0	0	0	15	7	15	8	TRUE
RNA129	RY.AU.GC.AU.RY	2	1	0	0	0	0	0	0	TRUE

RNA130	RY.AU.GC.AU.YR	3	2	0	0	0	0	0	0	TRUE
RNA131	YR.AU.GC.AU.RY	3	1	0	0	0	0	0	0	TRUE
RNA132	YR.AU.GC.AU.YR	2	1	0	0	0	0	0	0	TRUE
RNA133	RY.AU.GC.UA.RY	9	9	0	0	0	0	0	0	TRUE
RNA134	RY.AU.GC.UA.YR	3	0	0	0	0	0	0	0	FALSE
RNA135	YR.AU.GC.UA.RY	3	1	0	0	0	0	0	0	TRUE
RNA136	YR.AU.GC.UA.YR	4	2	0	0	0	0	0	0	TRUE
RNA137	RY.AU.GC.GC.RY	7	2	0	0	0	0	0	0	TRUE
RNA138	RY.AU.GC.GC.YR	9	6	0	0	0	0	0	0	TRUE
RNA139	YR.AU.GC.GC.RY	1	0	0	0	0	0	0	0	FALSE
RNA140	YR.AU.GC.GC.YR	2	2	0	0	0	0	0	0	TRUE
RNA141	RY.AU.GC.CG.RY	6	4	0	0	0	0	0	0	TRUE
RNA142	RY.AU.GC.CG.YR	8	7	0	0	0	0	0	0	TRUE
RNA143	YR.AU.GC.CG.RY	0	0	0	0	0	0	0	0	FALSE
RNA144	YR.AU.GC.CG.YR	9	6	0	0	0	0	0	0	TRUE
RNA145	RY.UA.GC.AU.RY	2	0	0	0	0	0	0	0	FALSE
RNA146	RY.UA.GC.AU.YR	1	1	0	0	0	0	0	0	TRUE
RNA147	YR.UA.GC.AU.RY	2	2	0	0	0	0	0	0	TRUE
RNA148	YR.UA.GC.AU.YR	3	0	0	0	0	0	0	0	FALSE
RNA149	RY.UA.GC.UA.RY	2	1	0	0	0	0	0	0	TRUE
RNA150	RY.UA.GC.UA.YR	6	4	0	0	0	0	0	0	TRUE
RNA151	YR.UA.GC.UA.RY	0	0	0	0	0	0	0	0	FALSE
RNA152	YR.UA.GC.UA.YR	2	2	0	0	0	0	0	0	TRUE
RNA153	RY.UA.GC.GC.RY	3	2	0	0	0	0	0	0	TRUE
RNA154	RY.UA.GC.GC.YR	6	5	0	0	0	0	0	0	TRUE
RNA155	YR.UA.GC.GC.RY	3	1	0	0	0	0	0	0	TRUE
RNA156	YR.UA.GC.GC.YR	7	6	0	0	0	0	0	0	TRUE
RNA157	RY.UA.GC.CG.RY	2	0	0	0	0	0	0	0	FALSE
RNA158	RY.UA.GC.CG.YR	6	5	0	0	0	0	0	0	TRUE
RNA159	YR.UA.GC.CG.RY	1	0	0	0	0	0	0	0	FALSE
RNA160	YR.UA.GC.CG.YR	18	8	0	0	0	0	0	0	TRUE
RNA161	RY.GC.GC.AU.RY	11	7	0	0	0	0	0	0	TRUE
RNA162	RY.GC.GC.AU.YR	10	6	0	0	0	0	0	0	TRUE
RNA163	YR.GC.GC.AU.RY	4	2	0	0	0	0	0	0	TRUE
RNA164	YR.GC.GC.AU.YR	1	0	0	0	0	0	0	0	FALSE
RNA165	RY.GC.GC.UA.RY	7	3	0	0	0	0	0	0	TRUE
RNA166	RY.GC.GC.UA.YR	6	2	0	0	0	0	0	0	TRUE
RNA167	YR.GC.GC.UA.RY	3	3	0	0	0	0	0	0	TRUE
RNA168	YR.GC.GC.UA.YR	2	3	0	0	0	0	0	0	TRUE
RNA169	RY.GC.GC.GC.RY	8	7	0	0	0	0	0	0	TRUE
RNA170	RY.GC.GC.GC.YR	3	0	0	0	0	0	0	0	FALSE
RNA171	YR.GC.GC.GC.RY	6	2	0	0	0	0	0	0	TRUE
RNA172	YR.GC.GC.GC.YR	1	0	0	0	0	0	0	0	FALSE
RNA173	RY.GC.GC.CG.RY	5	3	0	0	0	0	0	0	TRUE

RNA174	RY.GC.GC.CG.YR	12	8	0	0	0	0	0	0	TRUE
RNA175	YR.GC.GC.CG.RY	1	0	0	0	0	0	0	0	FALSE
RNA176	YR.GC.GC.CG.YR	4	3	0	0	0	0	0	0	TRUE
RNA177	RY.CG.GC.AU.RY	1	0	0	0	0	0	0	0	FALSE
RNA178	RY.CG.GC.AU.YR	6	6	0	0	0	0	0	0	TRUE
RNA179	YR.CG.GC.AU.RY	3	1	0	0	0	0	0	0	TRUE
RNA180	YR.CG.GC.AU.YR	3	2	0	0	0	0	0	0	TRUE
RNA181	RY.CG.GC.UA.RY	1	1	0	0	0	0	0	0	TRUE
RNA182	RY.CG.GC.UA.YR	3	2	0	0	0	0	0	0	TRUE
RNA183	YR.CG.GC.UA.RY	4	4	0	0	0	0	0	0	TRUE
RNA184	YR.CG.GC.UA.YR	3	3	0	0	0	0	0	0	TRUE
RNA185	RY.CG.GC.GC.RY	2	1	0	0	0	0	0	0	TRUE
RNA186	RY.CG.GC.GC.YR	3	2	0	0	0	0	0	0	TRUE
RNA187	YR.CG.GC.GC.RY	4	0	0	0	0	0	0	0	FALSE
RNA188	YR.CG.GC.GC.YR	3	0	0	0	0	0	0	0	FALSE
RNA189	RY.CG.GC.CG.RY	3	1	0	0	0	0	0	0	TRUE
RNA190	RY.CG.GC.CG.YR	2	2	0	0	0	0	0	0	TRUE
RNA191	YR.CG.GC.CG.RY	0	0	0	0	0	0	0	0	FALSE
RNA192	YR.CG.GC.CG.YR	7	6	0	0	0	0	0	0	TRUE
RNA193	RY.AU.CG.AU.RY	0	0	0	0	2	2	2	3	TRUE
RNA194	RY.AU.CG.AU.YR	0	0	0	0	4	4	5	4	TRUE
RNA195	YR.AU.CG.AU.RY	0	0	0	0	2	2	2	2	TRUE
RNA196	YR.AU.CG.AU.YR	0	0	0	0	2	2	2	2	TRUE
RNA197	RY.AU.CG.UA.RY	0	0	0	0	5	3	5	2	TRUE
RNA198	RY.AU.CG.UA.YR	0	0	0	0	2	0	2	0	FALSE
RNA199	YR.AU.CG.UA.RY	0	0	0	0	2	2	2	2	TRUE
RNA200	YR.AU.CG.UA.YR	0	0	0	0	9	7	9	7	TRUE
RNA201	RY.AU.CG.GC.RY	0	0	0	0	3	3	3	3	TRUE
RNA202	RY.AU.CG.GC.YR	0	0	0	0	3	2	3	2	TRUE
RNA203	YR.AU.CG.GC.RY	0	0	0	0	4	4	4	3	TRUE
RNA204	YR.AU.CG.GC.YR	0	0	0	0	1	0	1	1	FALSE
RNA205	RY.AU.CG.CG.RY	0	0	0	0	2	1	2	1	TRUE
RNA206	RY.AU.CG.CG.YR	0	0	0	0	6	2	6	1	TRUE
RNA207	YR.AU.CG.CG.RY	0	0	0	0	3	2	3	3	TRUE
RNA208	YR.AU.CG.CG.YR	0	0	0	0	8	3	9	3	TRUE
RNA209	RY.UA.CG.AU.RY	0	0	0	0	2	0	2	1	FALSE
RNA210	RY.UA.CG.AU.YR	0	0	0	0	1	1	1	1	TRUE
RNA211	YR.UA.CG.AU.RY	0	0	0	0	1	0	1	0	FALSE
RNA212	YR.UA.CG.AU.YR	0	0	0	0	2	0	2	0	FALSE
RNA213	RY.UA.CG.UA.RY	0	0	0	0	6	3	6	2	TRUE
RNA214	RY.UA.CG.UA.YR	0	0	0	0	4	3	5	2	TRUE
RNA215	YR.UA.CG.UA.RY	0	0	0	0	9	4	9	4	TRUE
RNA216	YR.UA.CG.UA.YR	0	0	0	0	9	4	8	4	TRUE
RNA217	RY.UA.CG.GC.RY	0	0	0	0	1	0	1	0	FALSE

RNA218	RY.UA.CG.GC.YR	0	0	0	0	8	5	8	5	TRUE
RNA219	YR.UA.CG.GC.RY	0	0	0	0	2	1	2	1	TRUE
RNA220	YR.UA.CG.GC.YR	0	0	0	0	1	0	1	0	FALSE
RNA221	RY.UA.CG.CG.RY	0	0	0	0	0	0	0	0	FALSE
RNA222	RY.UA.CG.CG.YR	0	0	0	0	12	6	12	6	TRUE
RNA223	YR.UA.CG.CG.RY	0	0	0	0	3	1	3	1	TRUE
RNA224	YR.UA.CG.CG.YR	0	0	0	0	12	7	12	6	TRUE
RNA225	RY.GC.CG.AU.RY	0	0	0	0	17	6	18	6	TRUE
RNA226	RY.GC.CG.AU.YR	0	0	0	0	6	4	6	5	TRUE
RNA227	YR.GC.CG.AU.RY	0	0	0	0	2	0	2	0	FALSE
RNA228	YR.GC.CG.AU.YR	0	0	0	0	1	0	1	0	FALSE
RNA229	RY.GC.CG.UA.RY	0	0	0	0	6	3	6	3	TRUE
RNA230	RY.GC.CG.UA.YR	0	0	0	0	6	6	6	6	TRUE
RNA231	YR.GC.CG.UA.RY	0	0	0	0	2	2	2	2	TRUE
RNA232	YR.GC.CG.UA.YR	0	0	0	0	6	3	6	3	TRUE
RNA233	RY.GC.CG.GC.RY	0	0	0	0	3	1	3	1	TRUE
RNA234	RY.GC.CG.GC.YR	0	0	0	0	2	1	2	1	TRUE
RNA235	YR.GC.CG.GC.RY	0	0	0	0	1	1	1	1	TRUE
RNA236	YR.GC.CG.GC.YR	0	0	0	0	2	0	2	1	FALSE
RNA237	RY.GC.CG.CG.RY	0	0	0	0	6	4	6	4	TRUE
RNA238	RY.GC.CG.CG.YR	0	0	0	0	12	8	11	7	TRUE
RNA239	YR.GC.CG.CG.RY	0	0	0	0	1	0	1	0	FALSE
RNA240	YR.GC.CG.CG.YR	0	0	0	0	6	3	7	3	TRUE
RNA241	RY.CG.CG.AU.RY	0	0	0	0	6	2	6	5	TRUE
RNA242	RY.CG.CG.AU.YR	0	0	0	0	3	4	5	3	TRUE
RNA243	YR.CG.CG.AU.RY	0	0	0	0	3	1	3	2	TRUE
RNA244	YR.CG.CG.AU.YR	0	0	0	0	2	2	2	1	TRUE
RNA245	RY.CG.CG.UA.RY	0	0	0	0	3	2	3	1	TRUE
RNA246	RY.CG.CG.UA.YR	0	0	0	0	8	4	9	3	TRUE
RNA247	YR.CG.CG.UA.RY	0	0	0	0	0	0	0	0	FALSE
RNA248	YR.CG.CG.UA.YR	0	0	0	0	7	3	7	3	TRUE
RNA249	RY.CG.CG.GC.RY	0	0	0	0	3	2	3	2	TRUE
RNA250	RY.CG.CG.GC.YR	0	0	0	0	3	2	3	2	TRUE
RNA251	YR.CG.CG.GC.RY	0	0	0	0	4	1	4	0	FALSE
RNA252	YR.CG.CG.GC.YR	0	0	0	0	2	2	2	2	TRUE
RNA253	RY.CG.CG.CG.RY	0	0	0	0	7	0	7	0	FALSE
RNA254	RY.CG.CG.CG.YR	0	0	0	0	3	1	3	1	TRUE
RNA255	YR.CG.CG.CG.RY	0	0	0	0	5	0	5	2	FALSE
RNA256	YR.CG.CG.CG.YR	0	0	0	0	11	5	11	7	TRUE

534 [a] R, purine; Y, pyrimidine
535
536

Table S2. Biological RNAs with base pairing represented by the 12 RNAs investigated in this study.

RNA#	Nucleotide numbers 5′→3′:3′→5′	Biological RNA	PDB	Nucleotide numbers 5′→3′:3′→5′
5	1-6:17-22	Ribosome	1VW7	2122-2128:2326-2331
	1-5:18-22	SAM-III riboswitch	3E5C	2-6:48-52
7	1-5:18-22	HIV dimer initiation complex	1BAU	1-5:19-23
	2-7:16-21	Ribosome	5ANC	2418-2423:2446-2452
8	3-7:16-20	Ribosome	1FFK	906-910:1295-1299
	4-8:15-19	HCV IRES	1KH6	24-28:33-37
21	4-8:15-19	Ribosome	1FFK	132-136:142-146
	5-9:14-18	HIV (MAL) dimer initiation complex	1K9W	2-6:18-22
23	3-8:15-20	Ribosome	1VW7	2151-2156:2180-2186
	4-8:15-19	Signal recognition particle RNA	2XXA	3-7:96-100
24	2-6:17-21	tRNA Gly acceptor stem	2VUQ	1-5:68-72
	3-6:17-20	Ribosome	1VW7	2151-2154:2182-2185
73	2-8:15-21	Pre-5S rRNA	2I91	(C)1-7:(D)2-8
	5-9:15-18	Smaug response element RNA	2ES5	5-9:15-19
74	2-7:16-21	*E. coli* copy control RNA duplex	1QC0	(C)109-114:(D)125-130
	3-7:16-20	2'-Deoxyguanosine riboswitch	3SKI	(A)35-39:47-51
	3-7:16-20	Ribosome	4V8P	2649-2653:2695-2699
	3-7:16-20	Ribosome	1NJM	2029-2033:2597-2601
	2-7:16-21	*E. coli* copy control RNA duplex	1QC0	(C)109-114:(D)125-130
75	2-7:16-21	Mitochondrial ribosome	5AJ3	409-414:445-450
	3-6:17-20	Ribonuclease P RNA	1NBS	92-95:108-111
	3-6:17-20	Ribosome	4L47	998-1001:1039-1042
	3-6:17-20	MLV encapsidation signal	16UP	283-286:295-298
89	1-7:16-22	Hairpin ribozyme	1M5K	29-35:46-52
	3-7:16-20	*E. coli* copy control RNA duplex	1QC0	106-110:129-133
	3-7:16-20	Ribosome	1FFK	1734-1738:2041-2045
	3-6:17-20	Ribosome	3J3E	3-6:110-113
90	2-7:16-21	Selenocysteine insertion sequence RNA	2RLU	1-6:14-19
	3-8:15-20	Cole1 inverted loop	1BJ2	1-6:16-21
	3-8:15-20	Selenocysteine insertion sequence RNA	2PLY	16-21:27-32
	3-7:16-20	HIV Lai kissing complex	2F4X	2-6:20-24
	3-7:16-20	tRNA	1ZL3	402-406:418-422
	3-7:16-20	Ribosome	3J3E	1428-1432:1503-1507
91	2-7:16-21	PreQ1 riboswitch	2MIY	-1-5:22-27
	3-7:16-20	Ribosome	3ZEX	729-733:1587-1591
	3-7:16-20	tRNA His	3WC1	48-52:60-64
	3-7:16-20	Hepatitis B virus encapsidation signal	2IXY	6-10:17-21
	3-7:16-20	Ribosome	4V7R	2365-2369:2378-2382

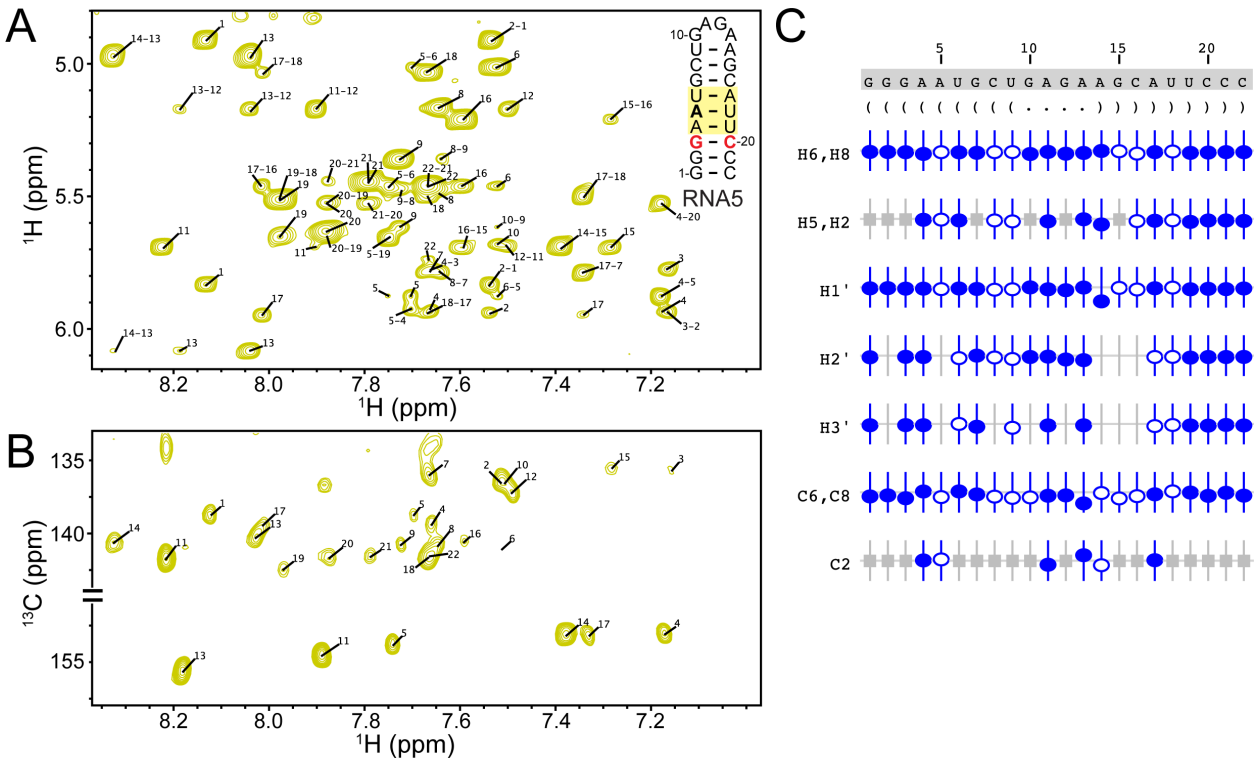
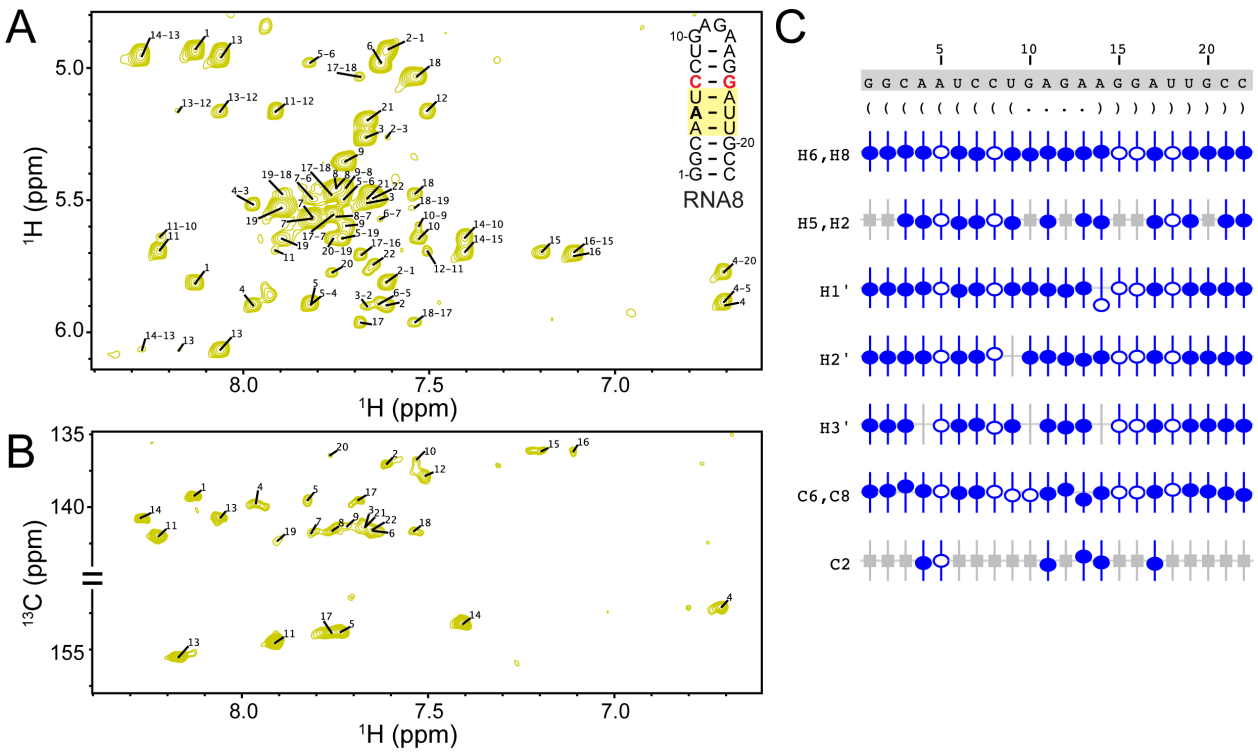


Figure S1. RNA5 chemical shift assignments. A) ^{1}H-^{1}H NOESY, B) ^{1}H-^{13}C HMQC, C) Summary of the sequence, secondary structure, and assignment validation for RNA5. The secondary structure is shown in Vienna format. NMRViewJ assignment summary to validate proton (H6/H8, H5/H2, H1′, H2′, H3′) and carbon (C6/C8, C2) assignments (bottom). Assigned chemical shifts for specific atoms are indicated with open and filled circles. The vertical offset of the circles indicates the deviation from the predicted values for that atom. Filled circles indicate that there are chemical shifts for atoms with the same set of attributes in the BMRB. Open circles indicate atoms that have a prediction, but for which no exact matches of the attributes are available in the BMRB. Grey boxes represent atoms that are not present in a given base.



552
553

554 **Figure S2.** RNA8 chemical shift assignments. A) ^{1}H-^{1}H NOESY, B) ^{1}H-^{13}C HMQC, C) Summary of the sequence,
555 secondary structure, and assignment validation for RNA8. The secondary structure is shown in Vienna format.
556 NMRViewJ assignment summary to validate proton (H6/H8, H5/H2, H1′, H2′, H3′) and carbon (C6/C8, C2)
557 assignments (bottom). Assigned chemical shifts for specific atoms are indicated with open and filled circles. The
558 vertical offset of the circles indicates the deviation from the predicted values for that atom. Filled circles indicate
559 that there are chemical shifts for atoms with the same set of attributes in the BMRB. Open circles indicate atoms that
560 have a prediction, but for which no exact matches of the attributes are available in the BMRB. Grey boxes represent
561 atoms that are not present in a given base.

562
563
564

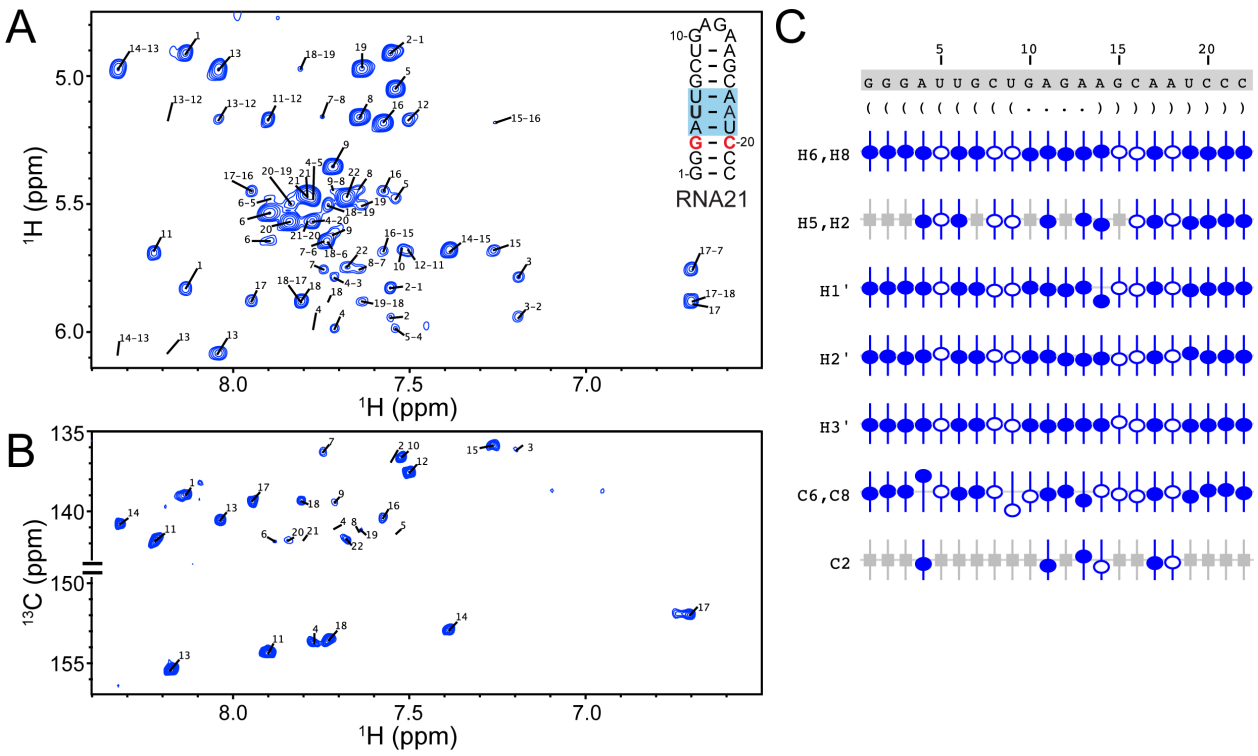


Figure S3. RNA21 chemical shift assignments. A) ^{1}H-^{1}H NOESY, B) ^{1}H-^{13}C HMQC, C) Summary of the sequence, secondary structure, and assignment validation for RNA21. The secondary structure is shown in Vienna format. NMRViewJ assignment summary to validate proton (H6/H8, H5/H2, H1′, H2′, H3′) and carbon (C6/C8, C2) assignments (bottom). Assigned chemical shifts for specific atoms are indicated with open and filled circles. The vertical offset of the circles indicates the deviation from the predicted values for that atom. Filled circles indicate that there are chemical shifts for atoms with the same set of attributes in the BMRB. Open circles indicate atoms that have a prediction, but for which no exact matches of the attributes are available in the BMRB. Grey boxes represent atoms that are not present in a given base.

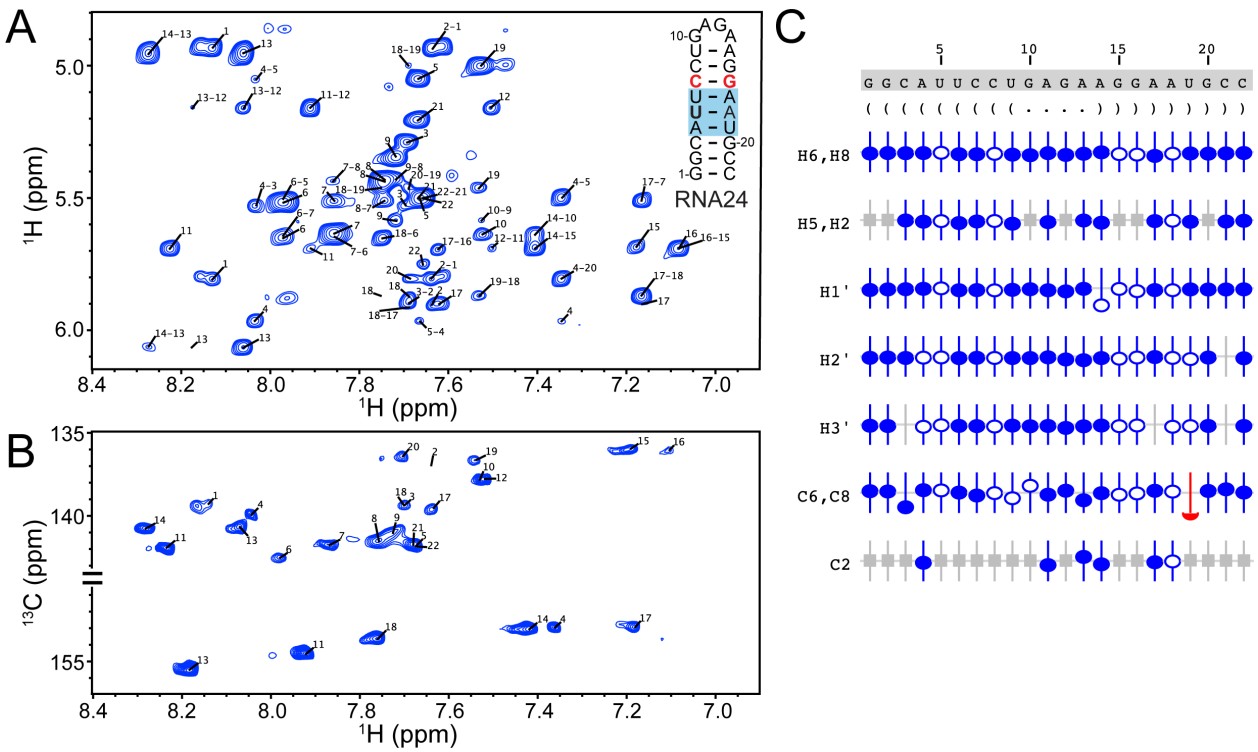


Figure S4. RNA24 chemical shift assignments. A) ^{1}H-^{1}H NOESY, B) ^{1}H-^{13}C HMQC, C) Summary of the sequence, secondary structure, and assignment validation for RNA24. The secondary structure is shown in Vienna format. NMRViewJ assignment summary to validate proton (H6/H8, H5/H2, H1′, H2′, H3′) and carbon (C6/C8, C2) assignments (bottom). Assigned chemical shifts for specific atoms are indicated with open and filled circles. The vertical offset of the circles indicates the deviation from the predicted values for that atom. Filled circles indicate that there are chemical shifts for atoms with the same set of attributes in the BMRB. Open circles indicate atoms that have a prediction, but for which no exact matches of the attributes are available in the BMRB. Red symbols indicate assignments that fall outside of the tolerance range. Grey boxes represent atoms that are not present in a given base.

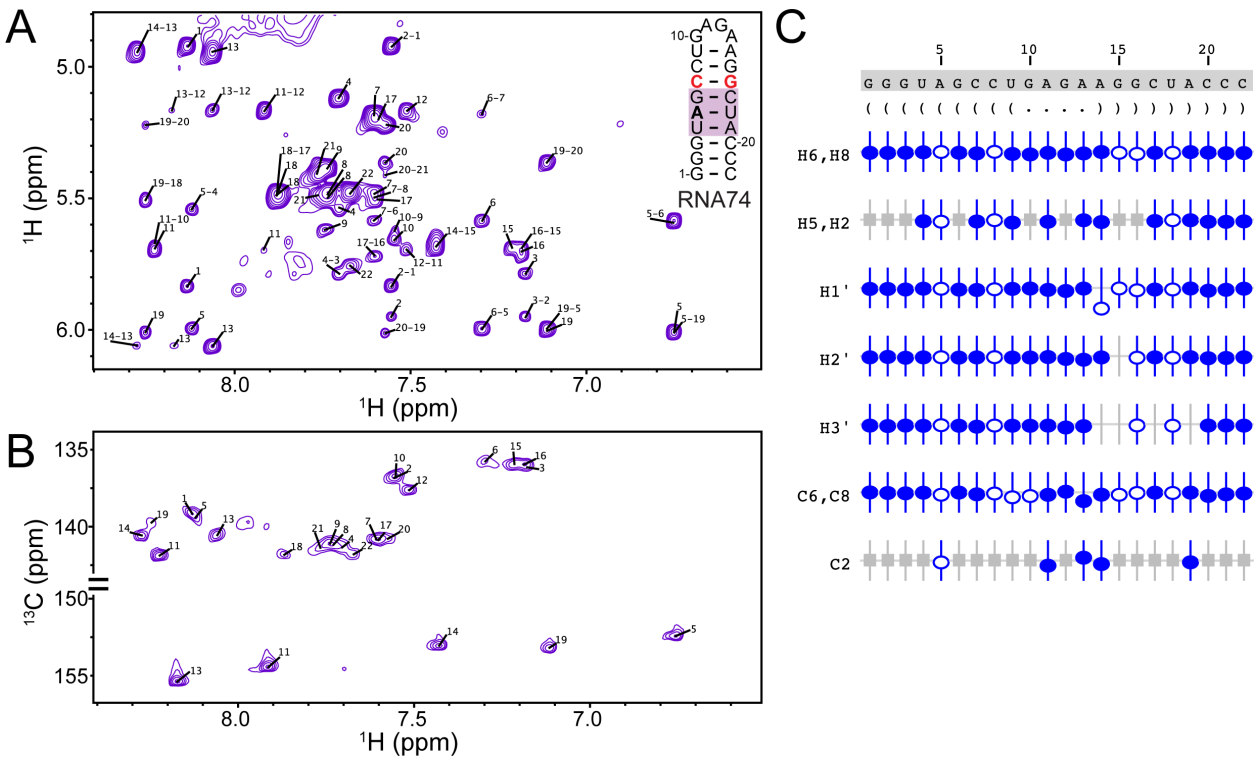
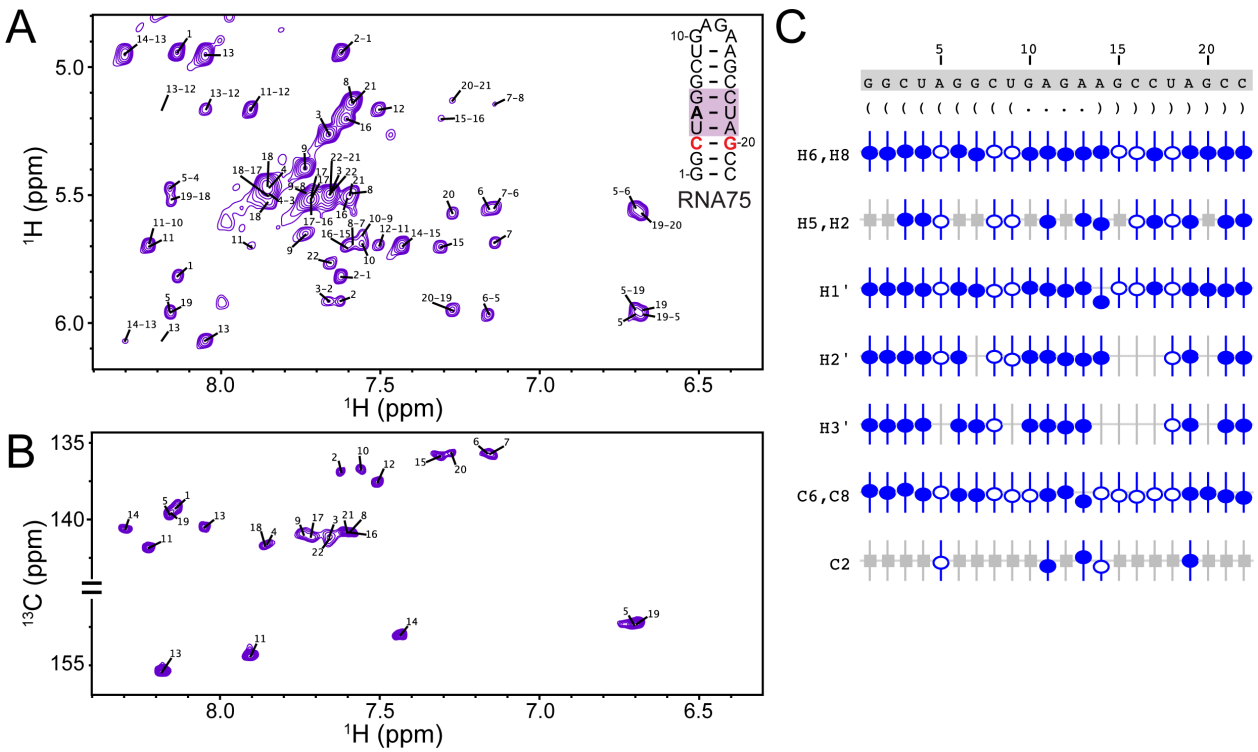


Figure S5. RNA74 chemical shift assignments. A) ^{1}H-^{1}H NOESY, B) ^{1}H-^{13}C HMQC, C) Summary of the sequence, secondary structure, and assignment validation for RNA74. The secondary structure is shown in Vienna format. NMRViewJ assignment summary to validate proton (H6/H8, H5/H2, H1′, H2′, H3′) and carbon (C6/C8, C2) assignments (bottom). Assignments are indicated with open and filled circles. Assigned chemical shifts for specific atoms are indicated with open and filled circles. The vertical offset of the circles indicates the deviation from the predicted values for that atom. Filled circles indicate that there are chemical shifts for atoms with the same set of attributes in the BMRB. Open circles indicate atoms that have a prediction, but for which no exact matches of the attributes are available in the BMRB. Grey boxes represent atoms that are not present in a given base.



606
607

608 **Figure S6.** RNA75 chemical shift assignments. A) ^{1}H-^{1}H NOESY, B) ^{1}H-^{13}C HMQC, C) Summary of the sequence,
609 secondary structure, and assignment validation for RNA75. The secondary structure is shown in Vienna format.
610 NMRViewJ assignment summary to validate proton (H6/H8, H5/H2, H1′, H2′, H3′) and carbon (C6/C8, C2)
611 assignments (bottom). Assignments are indicated with open and filled circles. Assigned chemical shifts for specific
612 atoms are indicated with open and filled circles. The vertical offset of the circles indicates the deviation from the
613 predicted values for that atom. Filled circles indicate that there are chemical shifts for atoms with the same set of
614 attributes in the BMRB. Open circles indicate atoms that have a prediction, but for which no exact matches of the
615 attributes are available in the BMRB. Grey boxes represent atoms that are not present in a given base.

616

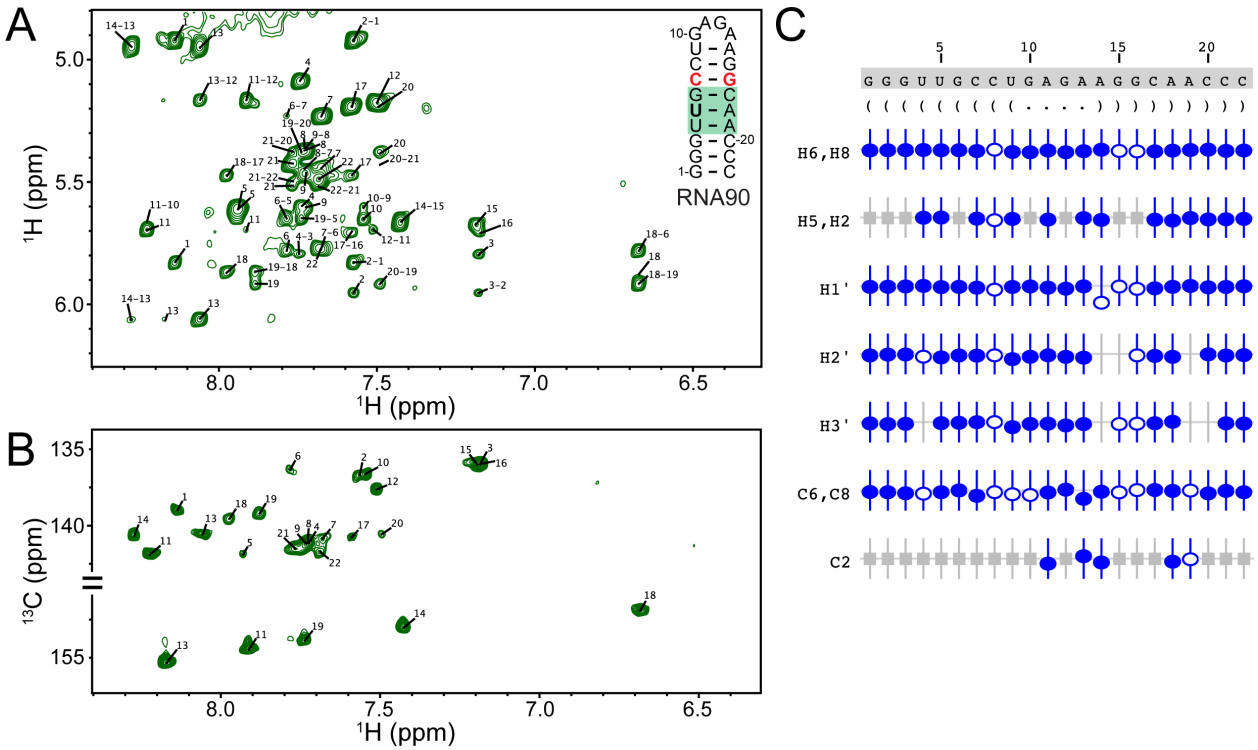
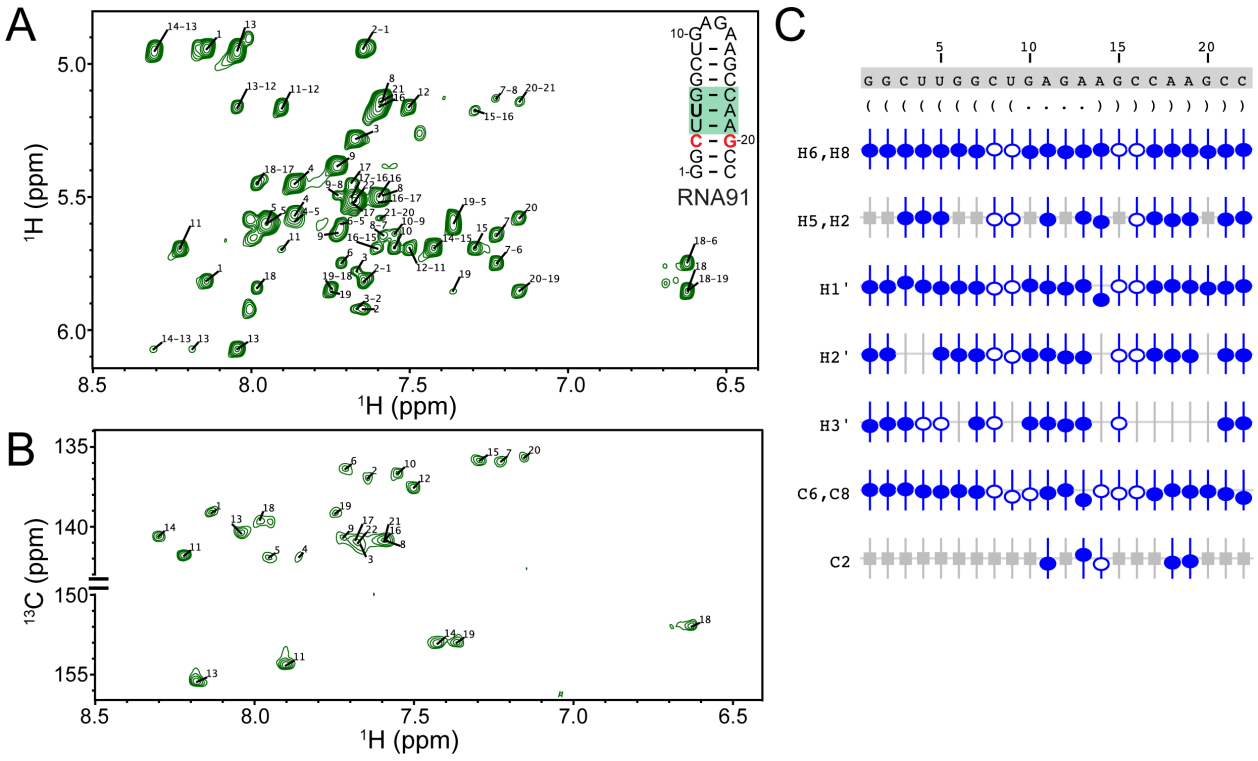


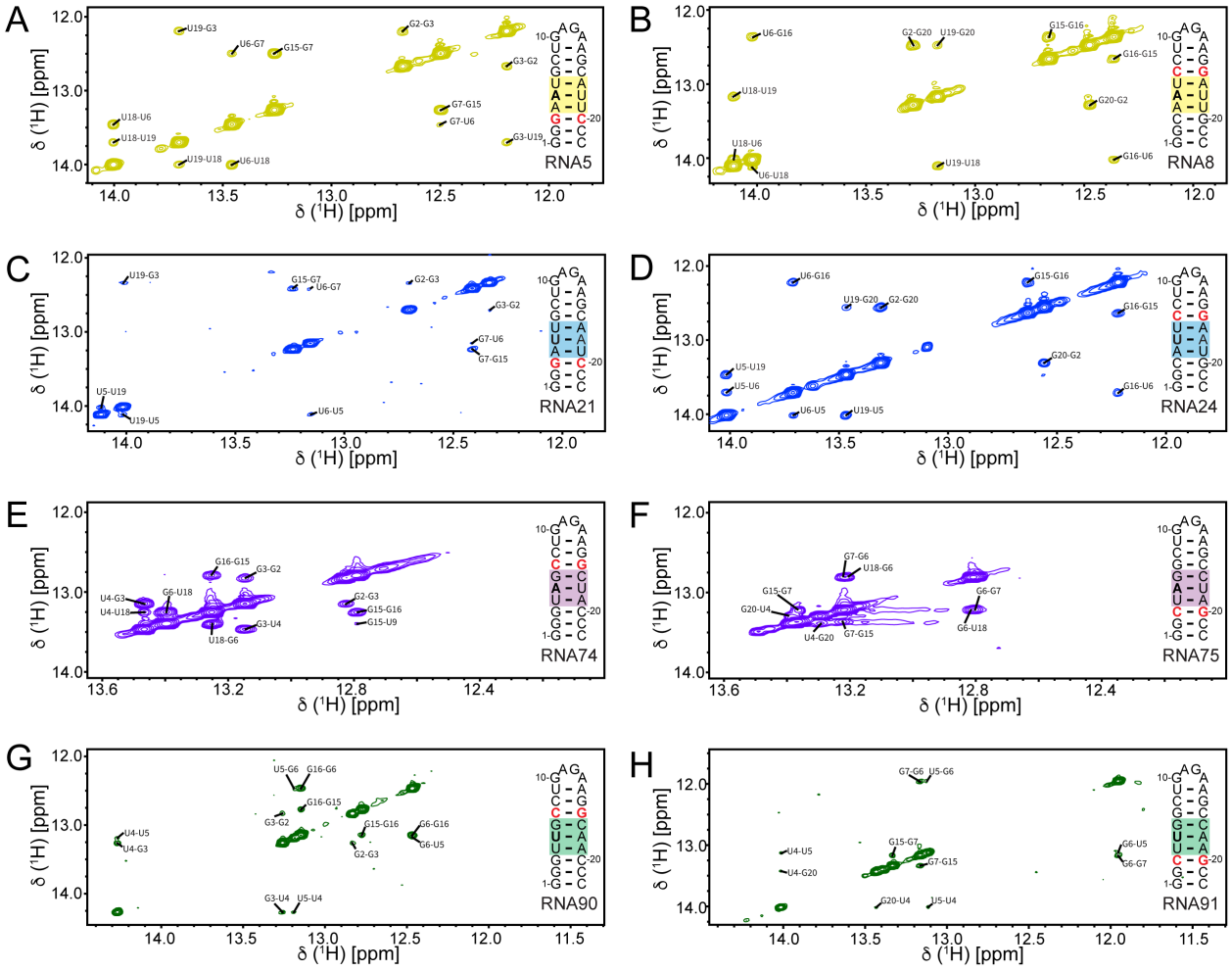
Figure S7. RNA90 chemical shift assignments. A) ^{1}H-^{1}H NOESY, B) ^{1}H-^{13}C HMQC, C) Summary of the sequence, secondary structure, and assignment validation for RNA90. The secondary structure is shown in Vienna format. NMRViewJ assignment summary to validate proton (H6/H8, H5/H2, H1′, H2′, H3′) and carbon (C6/C8, C2) assignments (bottom). Assigned chemical shifts for specific atoms are indicated with open and filled circles. The vertical offset of the circles indicates the deviation from the predicted values for that atom. Filled circles indicate that there are chemical shifts for atoms with the same set of attributes in the BMRB. Open circles indicate atoms that have a prediction, but for which no exact matches of the attributes are available in the BMRB. Grey boxes represent atoms that are not present in a given base.



628
629

630 **Figure S8.** RNA91 chemical shift assignments. A) ^{1}H-^{1}H NOESY, B) ^{1}H-^{13}C HMQC, C) Summary of the sequence,
631 secondary structure, and assignment validation for RNA91. The secondary structure is shown in Vienna format.
632 NMRViewJ assignment summary to validate proton (H6/H8, H5/H2, H1′, H2′, H3′) and carbon (C6/C8, C2)
633 assignments (bottom). Assigned chemical shifts for specific atoms are indicated with open and filled circles. The
634 vertical offset of the circles indicates the deviation from the predicted values for that atom. Filled circles indicate
635 that there are chemical shifts for atoms with the same set of attributes in the BMRB. Open circles indicate atoms that
636 have a prediction, but for which no exact matches of the attributes are available in the BMRB. Grey boxes represent
637 atoms that are not present in a given base.

638
639
640
641
642
643
644



645

646 **Figure S9.** Assigned imino proton NOESY spectra for A) RNA5, B) RNA8, C) RNA21, D) RNA24, E) RNA74, F)

647 RNA75, G) RNA90, H) RNA91. Secondary structures for each RNA are inset.

648
649