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

Yaping Liu^{1,5}, Anita Kotar^{1,5}, Tracy L. Hodges¹, Kyrillos Abdallah¹, Mallak H. Taleb¹, Brayden A. Bitterman¹, Sara Jaime¹, Kyle J. Schaubroeck¹, Ethan Mathew¹, Nicholas W. Morgenstern¹, Anthony Lohmeier¹, Jordan L. Page¹, Matt Ratanapanichkich¹, Grace Arhin¹, Breanna L. Johnson¹, Stanislav Cherepanov¹, Stephen C. Moss¹, Gisselle Zuniga¹, Nicholas J. Tilson¹, Zoe C. Yeoh², Bruce A. Johnson³, 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

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

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), 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). In our experience, a good prediction 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. 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). There are 64 possible canonical base pair triplets, which we have defined as a group. Because the specific identity of the base pairs flanking this core sequence has a lesser effect on the chemical shift of the central nucleotide, we considered nucleotide type (purine or pyrimidine) rather than nucleotide identity. This is consistent with the attributes used in the current prediction algorithm used in NMRFX Analyst (Marchant et al., 2019). If each of these “groups” 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. With these 256 sequences in hand, we examined the BMRB for aromatic protons and carbon chemical shifts and identified four “groups” of RNAs whose chemical shifts were underrepresented (**Fig. 1**). We prepared 12 unique RNAs, three from each group, for chemical shift assignment.

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 S1**) 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. 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 to generate secondary structures from 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 S2**). 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 μ L of each DNA oligonucleotide (200 μ M) was individually mixed with 20 μ L of the T7 promoter sequence oligonucleotide (600 μ M). 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 μ L H₂O 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	GGGAAUGCU <u>GAGA</u> AGCAUUGCC	mGmGGAATGCTTCTCAGCATTCCCTATAGTGAGTCGTATTA
RNA7	GGCAAUGCU <u>GAGA</u> AGCAUUGCC	mGmGCAATGCTTCTCAGCATTGCCTATAGTGAGTCGTATTA
RNA8	GGCAAUCCU <u>GAGA</u> AGGAUUGCC	mGmGCAATCCTTCTCAGGATTGCCTATAGTGAGTCGTATTA
RNA21	GGGAUUGCU <u>GAGA</u> AGCAAUCC	mGmGGATTGCTTCTCAGCAATCCCTATAGTGAGTCGTATTA
RNA23	GGCAUUGCU <u>GAGA</u> AGCAAUGCC	mGmGCATTGCTTCTCAGCAATGCCTATAGTGAGTCGTATTA
RNA24	GGCAUUGCU <u>GAGA</u> AGGAAUGCC	mGmGCATTCCCTTCTCAGGAATGCCTATAGTGAGTCGTATTA
RNA73	GGGUAGGCU <u>GAGA</u> AGCCUAGCC	mGmGGTAGGCTTCTCAGCCTACCCTATAGTGAGTCGTATTA
RNA74	GGGUAGGCU <u>GAGA</u> AGGCUAGCC	mGmGGTAGCCTTCTCAGGCTACCCTATAGTGAGTCGTATTA
RNA75	GGCUAGGCU <u>GAGA</u> AGCCUAGCC	mGmGCTAGGCTTCTCAGCCTAGCCTATAGTGAGTCGTATTA
RNA89	GGGUUGGCU <u>GAGA</u> AGCCAAGCC	mGmGGTTGGCTTCTCAGCCAACCCTATAGTGAGTCGTATTA
RNA90	GGGUUGGCU <u>GAGA</u> AGGCAAGCC	mGmGGTTGCCTTCTCAGGCAACCCTATAGTGAGTCGTATTA
RNA91	GGCUUGGCU <u>GAGA</u> AGCCAAGCC	mGmGCTTGGCTTCTCAGCCAAGCCTATAGTGAGTCGTATTA

^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 NTPs, 10-20 mM MgCl₂, 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, Whatman). 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 D₂O (99.8%, Cambridge Isotope Laboratories, Inc.). 2D ¹H-¹H NOESY, ¹H-¹H TOCSY, and ¹H-¹³C 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). ¹H chemical shifts were referenced to water and ¹³C chemical shifts were indirectly referenced from the ¹H chemical shift (Wishart et al., 1995).

Table 2. Experimental parameters for NMR data acquisition.

Experimental parameter	¹ H- ¹ H NOESY	¹ H- ¹ H TOCSY	¹ H- ¹³ C HMQC
pulse sequence	noesyegpph (Hwang and Shaka, 1995)	mlevphpr.2 (Bax and Davis, 1985)	hmqcphpr (Bax et al., 1983)
ds	16	16	16
ns	24	24	336
sw(F2)	9.8104	9.8104	9.8104
sw(F1)	9.8104	9.8104	50
TD(F2)	8129	2048	1058
TD(F1)	512	512	88
O1 (ppm)	4.703	4.703	4.703
O2 (ppm)	-	-	148.00
D1 (s)	3	2	1.5
mixing time (ms)	300	80	-
experiment time (h)	~27	~8	~12

Assignments of helical regions

For all the RNA oligos, the nonexchangeable proton assignments were unambiguously assigned through using 2D ¹H-¹H NOESY (Hwang and Shaka, 1995), 2D ¹H-¹H TOCSY (Bax and Davis, 1985), and ¹H-¹³C HMQC (Bax et al., 1983) experiments. The 2D ¹H-¹H 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 ¹H-¹H NOESY spectrum. The ¹H-¹H 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 ¹H-¹³C HMQC was used to help distinguish C2-H2 (¹³C δ ~152 ppm) from C6-H6 and C8-H8 resonances (¹³C δ ~139 ppm).

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 ¹H-¹H NOESY spectrum (**Fig. 3**). 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

Assigned ¹H-¹H NOESY and ¹H-¹³C HMQC spectra for the 12 RNAs are presented in **Figs. 4-7** and **Supplemental Figs. S1-S8**. 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).

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

NMRx 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. <https://doi.org/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. <https://doi.org/10.1093/nar/20.4.819>
- Bahrami, A., Clos, L. J., 2nd, Markley, J. L., Butcher, S. E., Eghbalian, H. R. (2012) RNA-PAIRS: RNA probabilistic assignment of imino resonance shifts. *J Biomol NMR* 52:289-302. <https://doi.org/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. <https://doi.org/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. [https://doi.org/10.1016/0022-2364\(85\)90018-6](https://doi.org/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. [https://doi.org/10.1016/0022-2364\(83\)90241-X](https://doi.org/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. <https://doi.org/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. <https://doi.org/10.1093/nar/28.1.235>

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. <https://doi.org/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. <https://doi.org/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., Dinno, 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. <https://doi.org/10.1101/2020.06.11.145920>

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.

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. <https://doi.org/10.1017/s1355838200992495>

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. <https://doi.org/10.1007/BF00197809>

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

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

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

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

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

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

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

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

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

265 Hwang, T. L., Shaka, A. J. (1995) Water Suppression That Works - Excitation Sculpting Using Arbitrary Wave-
 266 Forms and Pulsed-Field Gradients. *J Magn Reson Ser A* 112:275-279.
 267 <https://doi.org/10.1006/jmra.1995.1047>

268 Johnson, B. A., Blevins, R. A. (1994) NMR View: A computer program for the visualization and analysis of NMR
 269 data. *J Biomol NMR* 4:603-614. <https://doi.org/10.1007/BF00404272>

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

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

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

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

279 Lam, S. L. (2007) DSHIFT: a web server for predicting DNA chemical shifts. *Nucleic Acids Research* 35:W713-
 280 W717. <https://doi.org/10.1093/nar/gkm320>

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

283 Marchant, J., Summers, M. F., Johnson, B. A. (2019) Assigning NMR spectra of RNA, peptides and small organic
 284 molecules using molecular network visualization software. *J Biomol NMR* [https://doi.org/10.1007/s10858-](https://doi.org/10.1007/s10858-019-00271-3)
 285 [019-00271-3](https://doi.org/10.1007/s10858-019-00271-3)

286 Meiler, J. (2003) PROSHIFT: protein chemical shift prediction using artificial neural networks. *J Biomol NMR*
 287 26:25-37. <https://doi.org/10.1023/a:1023060720156>

288 Meyer, S. M., Williams, C. C., Akahori, Y., Tanaka, T., Aikawa, H., Tong, Y., Childs-Disney, J. L., Disney, M. D.
 289 (2020) Small molecule recognition of disease-relevant RNA structures. *Chem Soc Rev* 49:7167-7199.
 290 <https://doi.org/10.1039/d0cs00560f>

291 Ng, K. S., Lam, S. L. (2015) NMR proton chemical shift prediction of C.C mismatches in B-DNA. *Journal of*
 292 *Magnetic Resonance* 252:87-93. <https://doi.org/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. <https://doi.org/10.1007/s10858-016-0049-6>

Papasaïkas, P., Valcarcel, J. (2016) The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends Biochem Sci* 41:33-45. <https://doi.org/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. <https://doi.org/10.1038/nrd.2017.243>

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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/10.1038/nmeth.2876>

Theimer, C. A., Feigon, J. (2006) Structure and function of telomerase RNA. *Curr Opin Struct Biol* 16:307-318. <https://doi.org/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. <https://doi.org/10.1093/nar/gkm957>

- 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.
<https://doi.org/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. <https://doi.org/10.1038/nrd.2018.93>
- 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.
<https://doi.org/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.
<https://doi.org/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. <https://doi.org/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. 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.

Figure 3. ^1H - ^1H 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.

Figure 4. RNA7 chemical shift assignments. A) ^1H - ^1H NOESY, B) ^1H - ^{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 5. RNA23 chemical shift assignments. A) ^1H - ^1H NOESY, B) ^1H - ^{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 6. RNA73 chemical shift assignments. A) ^1H - ^1H NOESY, B) ^1H - ^{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 7. RNA89 chemical shift assignments. A) ^1H - ^1H NOESY, B) ^1H - ^{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 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6