




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

Yaping Liu¹ · Anita Kotar^{1,5} · Tracy L. Hodges¹ · Kyrillos Abdallah¹ · Mallak H. Taleb¹ · Brayden A. Bitterman¹ · Sara Jaime¹ · Kyle J. Schaubroek¹ · 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 · Sarah C. Keane^{1,4} 

Received: 24 May 2021 / Accepted: 22 August 2021 / Published online: 27 August 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract

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; 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, b) 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;

Yaping Liu and Anita Kotar have contributed equally to this work.

✉ Sarah C. Keane
sckeane@umich.edu

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, 1000 Ljubljana, Slovenia

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, 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 (Fig. 2). 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, 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.

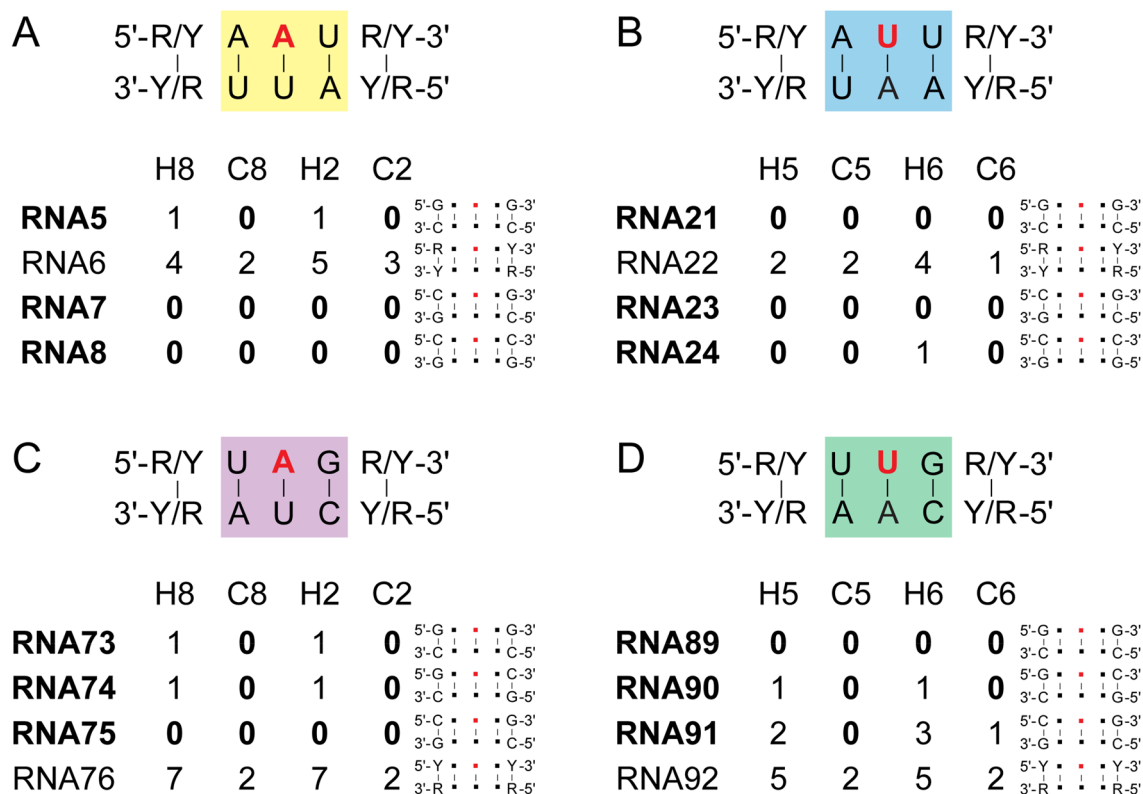


Fig. 1 Overview of RNA sequences with missing assignments. Four groups of RNAs were identified **A** Group 1, RNAs 5–8, **B** Group 2, RNAs 21–24, **C** Group 3, RNAs 73–76, **D** Group 4, RNAs 89–92. Red nucleotide indicates the nucleotide with missing chemical shifts. The common triple of base pairs for each group is highlighted in a colored box. Sequence variations at flanking nucleotides were exam-

ined. 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

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'-TAATAC GACTCACTATA-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 min and left in the water bath to slowly cool to room temperature over at least 2 h. The annealed template was then diluted with 940 μ L H₂O to produce the partially double-stranded DNA templates for transcription.

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 MgCl₂, 8–13 pmol annealed DNA template, 0.2 U/mL yeast inorganic pyrophosphatase (New England Biolabs) (Cunningham and Ofengand 1990), ~15 μ M T7 RNA polymerase, and 10–20% (vol/vol) DMSO (Helmling et al. 2015). Reactions were incubated at 37 °C for 3–4 h 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,

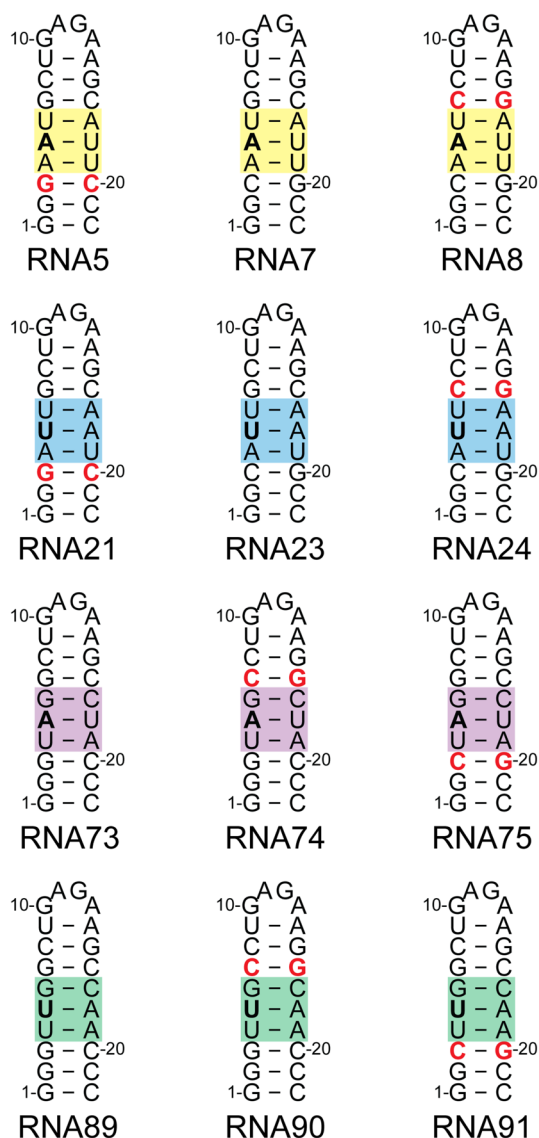


Fig. 2 RNA constructs used in this study. Red nucleotides indicate the variable sequence relative to a single example for each group. The common triple of base pairs for each group is highlighted in a colored box and the central nucleotide is in bold font

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% $\text{D}_2\text{O}/90\%$ H_2O (99.8%, Cambridge Isotope Laboratories, Inc.). 2D ^1H - ^1H NOESY (in both D_2O and H_2O), ^1H - ^1H TOCSY, and ^1H - ^{13}C HMQC spectra were recorded for each RNA at 25 $^\circ\text{C}$ (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 ^{13}C chemical shifts were indirectly referenced from the ^1H chemical shift (Wishart et al. 1995).

Assignments of helical regions

For all 12 RNA oligonucleotides, the nonexchangeable proton assignments were unambiguously assigned using 2D ^1H - ^1H NOESY (Hwang and Shaka 1995), 2D ^1H - ^1H TOCSY (Bax and Davis 1985), and ^1H - ^{13}C 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 - ^{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 ^1H - ^1H NOESY (aromatic) and ^1H - ^{13}C HMQC spectra for the 12 RNAs are presented in Figs. 3, 4, 5, 6, 7 and Supplemental Figs. S1–S9.

Assignments of imino protons

The exchangeable imino protons were unambiguously assigned using the 2D ^1H - ^1H NOESY recorded in 10% $\text{D}_2\text{O}/90\%$ H_2O and confirm the secondary structure of the

Table 1 RNA constructs

RNA	RNA sequence (5'–3') ^a	DNA template (5'–3') ^{b,c}
RNA5	GGG <u>AAUGCUGAGAAGCAU</u> UCCC	mGmGGAATGCTTCTCAGCATTCCCTATAGTGAGTCGTATTA
RNA7	GGCA <u>AUUGCUGAGAAGCAU</u> UGCC	mGmGCAATGCTTCTCAGCATTGCCTATAGTGAGTCGTATTA
RNA8	GGCA <u>AUCCUGAGAAGGAU</u> UGCC	mGmGCAATCCTTCTCAGGATTGCCTATAGTGAGTCGTATTA
RNA21	GGG <u>AAUUGCUGAGAAGCAA</u> UCCC	mGmGGATTGCTTCTCAGCAATCCCTATAGTGAGTCGTATTA
RNA23	GG <u>CAUUGCUGAGAAGCAA</u> UGCC	mGmGCATTGCTTCTCAGCAATGCCTATAGTGAGTCGTATTA
RNA24	GGCA <u>UUCUGAGAAGGAAU</u> GCC	mGmGCATTCCCTTCTCAGGAATGCCTATAGTGAGTCGTATTA
RNA73	GGG <u>UAGGCUGAGAAGCCU</u> ACCC	mGmGGTAGGCTTCTCAGCCTACCCTATAGTGAGTCGTATTA
RNA74	GGG <u>UAGCCUGAGAAGCCU</u> ACCC	mGmGGTAGCCTTCTCAGGCTACCCTATAGTGAGTCGTATTA
RNA75	GG <u>CUAGGCUGAGAAGCCU</u> AGCC	mGmGCTAGGCTTCTCAGCCTAGCCTATAGTGAGTCGTATTA
RNA89	GGG <u>UUGGCUGAGAAGCCA</u> ACCC	mGmGGTTGGCTTCTCAGCCAACCCTATAGTGAGTCGTATTA
RNA90	GGG <u>UUGCCUGAGAAGGCA</u> ACCC	mGmGGTTGCCTTCTCAGGCAACCCTATAGTGAGTCGTATTA
RNA91	GG <u>CUUGGCUGAGAAGCCA</u> AGCC	mGmGCTTGGCTTCTCAGCCAAGCCTATAGTGAGTCGTATTA

^aUnderlined regions correspond to variable sequences in the RNAs. The GAGA tetraloop sequence is bold

^bm denotes 2'-O-Me modification of the primer

^cItalicized nucleotides correspond to the sequence complementary to the T7 promoter

Table 2 Experimental parameters for NMR data acquisition

Experimental parameter	¹ H– ¹ H NOESY (100% D ₂ O)	¹ H– ¹ H NOESY (90% H ₂ O/10% D ₂ O)	¹ H– ¹ H TOCSY	¹ H– ¹³ C HMQC
pulse sequence	noesyegpph (Hwang and Shaka 1995)	noesyegpph (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)	8192	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

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 ¹H–¹H 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;

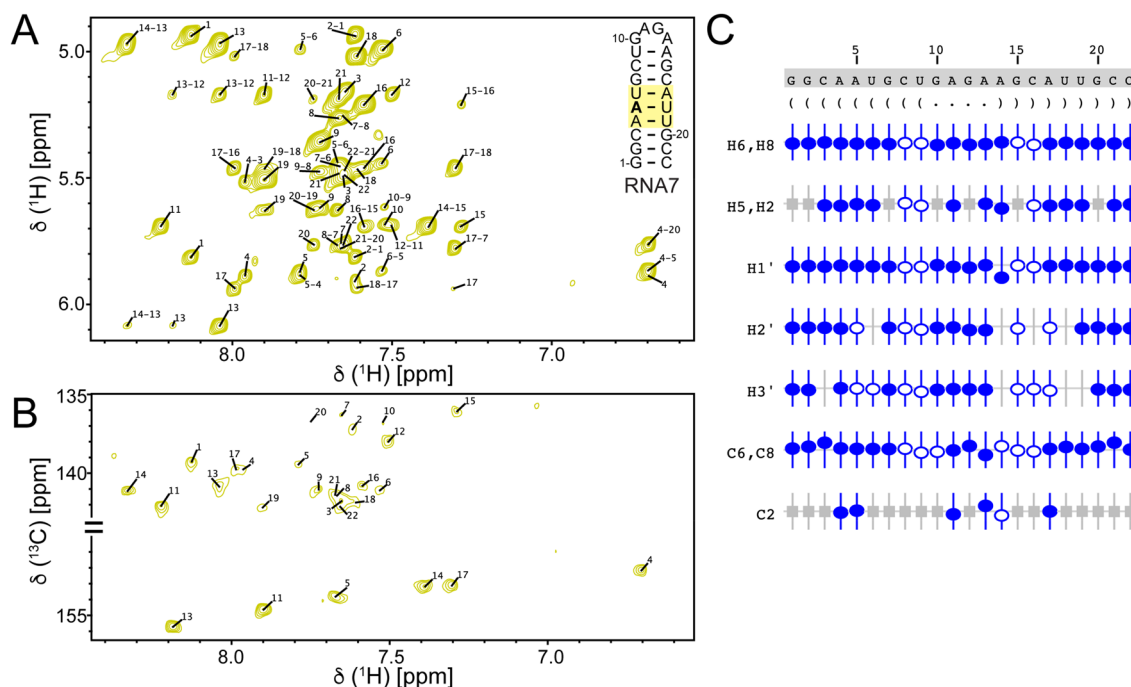


Fig. 3 RNA7 chemical shift assignments. **A** ^1H - ^1H NOESY, **B** ^1H - ^{13}C HMQC, **C** Summary of the sequence, secondary structure, and assignment validation for RNA7. The secondary structure is shown in Vienna format. NMRViewJ assignment summary to validate proton (H6/H8, H5/H2, H1', H2', H3') 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

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

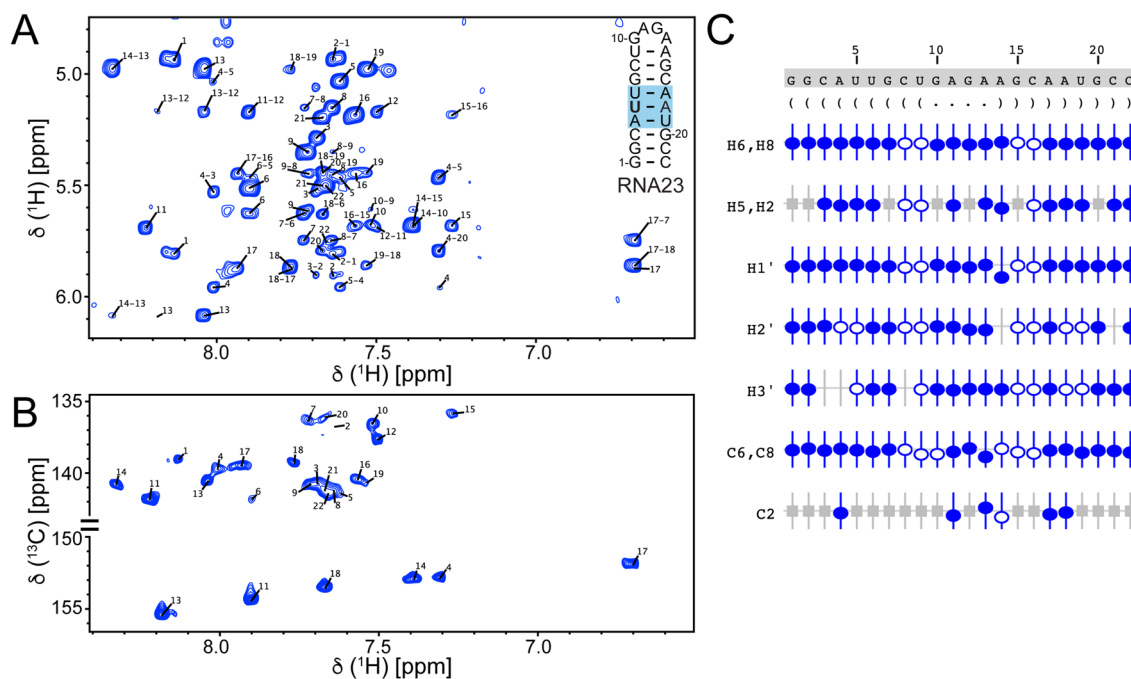


Fig. 4 RNA23 chemical shift assignments. **A** ^1H - ^1H NOESY, **B** ^1H - ^{13}C HMQC, **C** summary of the sequence, secondary structure, and assignment validation for RNA23. The secondary structure is shown in Vienna format. NMRViewJ assignment summary to validate proton (H6/H8, H5/H2, H1', H2', 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

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: 50,933, RNA7: 50,932, RNA8: 50,931, RNA21: 50,930, RNA23: 50,929, RNA24: 50,928, RNA73: 50,927, RNA74: 50,926,

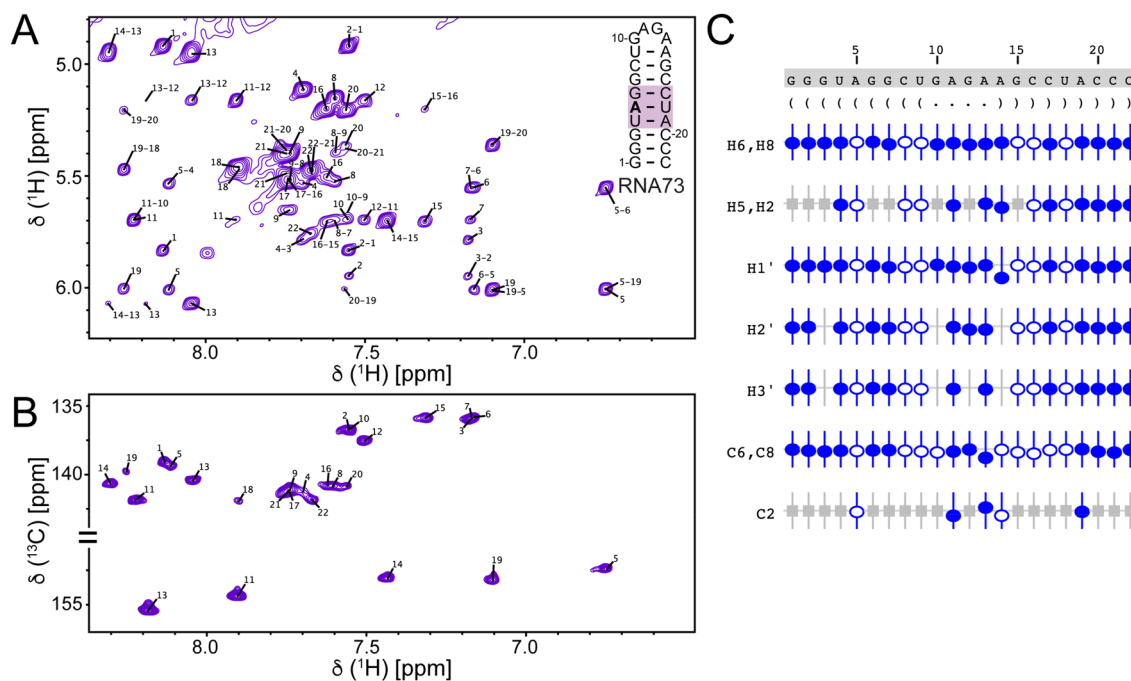


Fig. 5 RNA73 chemical shift assignments. **A** ^1H - ^1H NOESY, **B** ^1H - ^{13}C HMQC, **C** summary of the sequence, secondary structure, and assignment validation for RNA73. The secondary structure is shown in Vienna format. NMRViewJ assignment summary to validate proton (H6/H8, H5/H2, H1', H2', 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

RNA75: 50,925, RNA89: 50,924, RNA90: 50,923, RNA91: 50,922. 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).

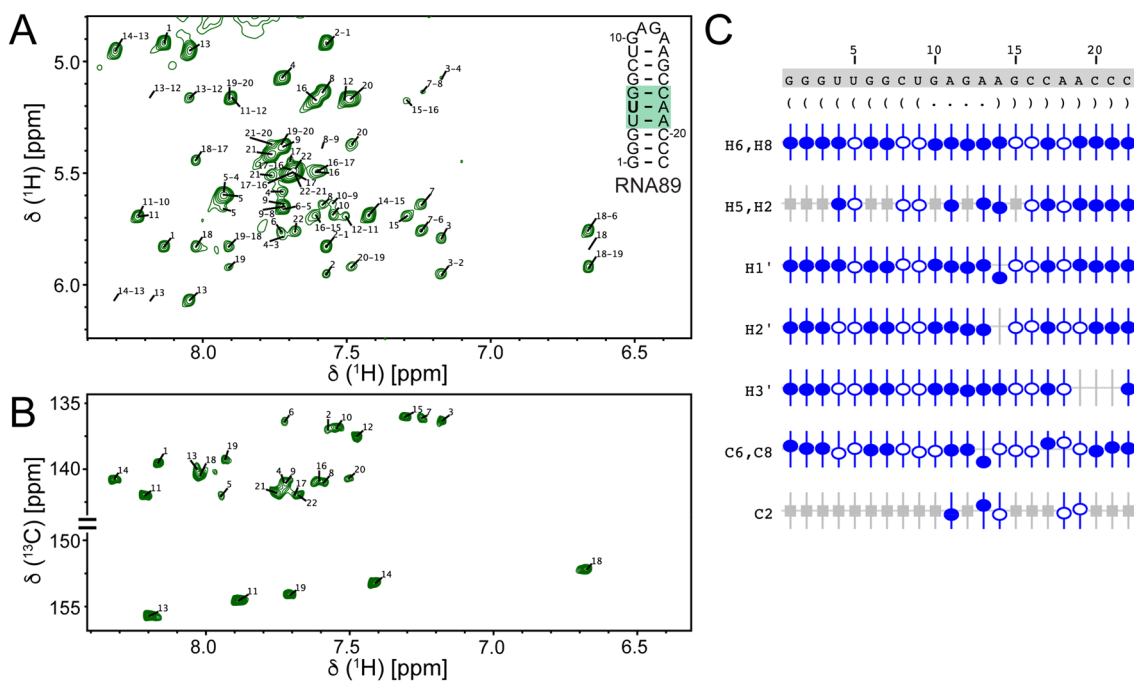


Fig. 6 RNA89 chemical shift assignments. **A** ^1H - ^1H NOESY, **B** ^1H - ^{13}C HMQC, **C** summary of the sequence, secondary structure, and assignment validation for RNA89. The secondary structure is shown in Vienna format. NMRViewJ assignment summary to validate proton (H6/H8, H5/H2, H1', H2', 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

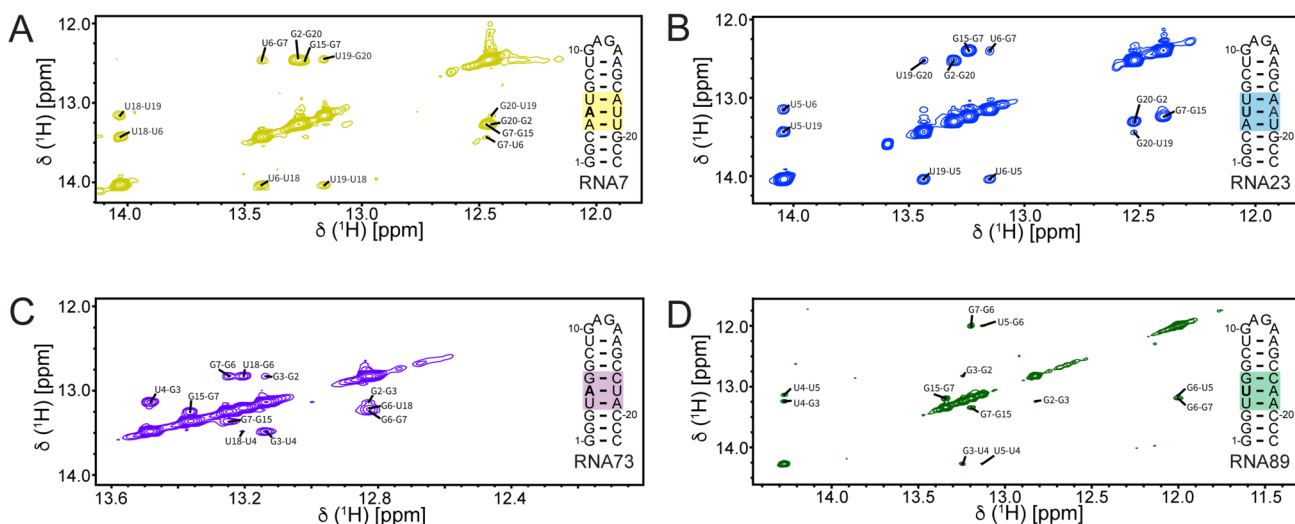
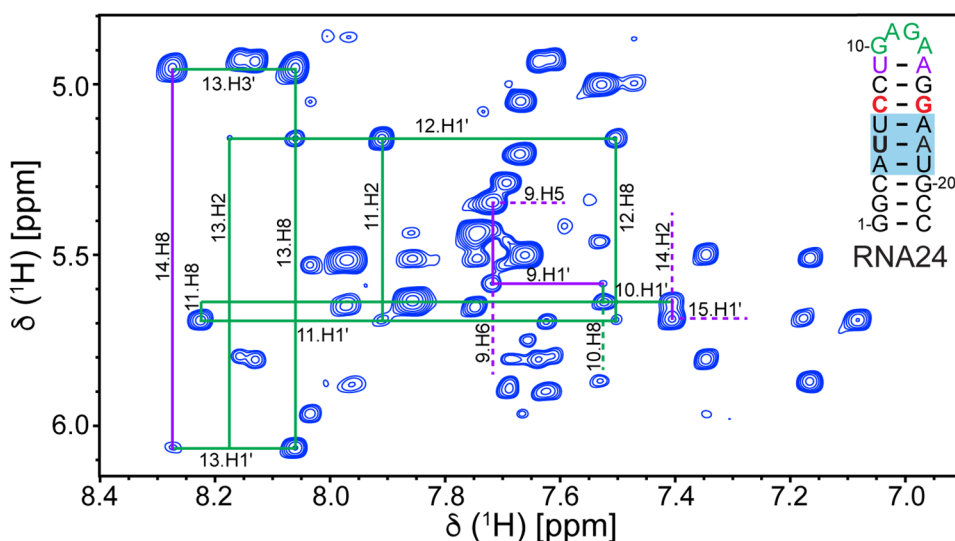


Fig. 7 Assigned imino proton NOESY spectra for **A** RNA7, **B** RNA23, **C** RNA73, and **D** RNA89. Secondary structures for each RNA are inset

Fig. 8 ^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



Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12104-021-10049-0>.

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.

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 103,297 (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.

Declarations

Conflict of interest 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.

References

- Aeschbacher T, Schmidt E, Blatter M, Maris C, Duss O, Allain FH, Güntert 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 VP, 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 LJ 2nd, Markley JL, Butcher SE, Eghbalian HR (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 BA, Summers MF (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 DG (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 RH, Hawkins BL (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 HM, Bhat TN, Bourne PE, 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 HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000b) The Protein Data Bank. *Nucleic Acids Res* 28:235–242. <https://doi.org/10.1093/nar/28.1.235>
- Brown JD, Summers MF, Johnson BA (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>
- Brown JD, Kharytonchik S, Chaudry I, Iyer AS, Carter H, Becker G, Desai Y, Glang L, Choi SH, Singh K, Lopresti MW, Orellana M, Rodriguez T, Oboh U, Hijji J, Ghinger FG, Stewart K, Francis D, Edwards B, Chen P, Case DA, Telesnitsky A, Summers MF (2020) Structural basis for transcriptional start site control of HIV-1 RNA fate. *Science* 368:413–417. <https://doi.org/10.1126/science.aaz7959>
- Burnett JC, Rossi JJ (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 KS, Edwards DK, Leist SR, Abiona OM, Boyoglu-Barnum S, Gillespie RA, Himansu S, Schäfer A, Ziwawo CT, DiPiazza AT, Dinnon KH, Elbashir SM, Shaw CA, Woods A, Fritch EJ, Martinez DR, Bock KW, Minai M, Nagata BM, Hutchinson GB, Wu K, Henry C, Bahl K, Garcia-Dominguez D, Ma L, Renzi I, Kong W-P, Schmidt SD, Wang L, Zhang Y, Phung E, Chang LA, Loomis RJ, Altaras NE, Narayanan E, Metkar M, Presnyak V, Liu C, Louder MK, Shi W, Leung K, Yang ES, West A, Gully KL, Stevens LJ, Wang N, Wrapp D, Doria-Rose NA, Stewart-Jones G, Bennett H, Alvarado GS, Nason MC, Ruckwardt TJ, McLellan JS, Denison MR, Chappell JD, Moore IN, Morabito KM, Mascola JR, Baric RS, Carfi A, Graham BS (2020) SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. *Nature* 586:567–571. <https://doi.org/10.1038/s41586-020-2622-0>
- Cornilescu G, Didychuk AL, Rodgers ML, Michael LA, Burke JE, Montemayor EJ, Hoskins AA, Butcher SE (2016) Structural analysis of multi-helical RNAs by NMR-SAXS/WAXS: application to the U4/U6 di-snRNA. *J Mol Biol* 428:777–789. <https://doi.org/10.1016/j.jmb.2015.11.026>
- Cunningham PR, 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 MJ (2000) A test of the model to predict unusually stable RNA hairpin loop stability. *RNA* 6:608–615. <https://doi.org/10.1017/s1355838200992495>
- Damase TR, Sukhovshin R, Boada C, Taraballi F, Pettigrew RI, Cooke JP (2021) The Limitless Future of RNA Therapeutics. *Front Bioeng Biotechnol* 9:628137. <https://doi.org/10.3389/fbioe.2021.628137>
- Delaglio F, Grzesiek S, Vuister GW, 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>
- D'Souza V, Dey A, Habib D, Summers MF (2004) NMR structure of the 101-nucleotide core encapsidation signal of the Moloney murine leukemia virus. *J Mol Biol* 337:427–442. <https://doi.org/10.1016/j.jmb.2004.01.037>
- Ebrahimi M, Rossi P, Rogers C, Harbison GS (2001) Dependence of ^{13}C NMR chemical shifts on conformations of RNA nucleosides and nucleotides. *J Magn Reson* 150:1–9. <https://doi.org/10.1006/jmre.2001.2314>
- Farés C, Amata I, Carlomagno T (2007) ^{13}C -detection in RNA bases: revealing structure-chemical shift relationships. *J Am Chem Soc* 129:15814–15823. <https://doi.org/10.1021/ja0727417>
- Frank AT, Bae S-H, Stelzer AC (2013) Prediction of RNA ^1H and ^{13}C chemical shifts: a structure based approach. *J Phys Chem B* 117:13497–13506. <https://doi.org/10.1021/jp407254m>
- Frank AT, Law SM, Brooks CL 3rd (2014) A simple and fast approach for predicting ^1H and ^{13}C chemical shifts: toward chemical shift-guided simulations of RNA. *J Phys Chem B* 118:12168–12175. <https://doi.org/10.1021/jp508342x>
- Ghose R, Marino JP, Wiberg KB, Prestegard JH (1994) Dependence of ^{13}C Chemical Shifts on Glycosidic Torsional Angles in Ribonucleic Acids. *J Am Chem Soc* 116:8827–8828. <https://doi.org/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. <https://doi.org/10.1017/s0033583500004169>
- Han B, Liu Y, Ginzinger SW, Wishart DS (2011) SHIFTX2: significantly improved protein chemical shift prediction. *J Biomol NMR* 50:43–57. <https://doi.org/10.1007/s10858-011-9478-4>
- Helmling C, Keyhani S, Sochor F, Fürtig B, Hengesbach M, Schwalbe H (2015) Rapid NMR screening of RNA secondary structure and binding. *J Biomol NMR* 63:67–76. <https://doi.org/10.1007/s10858-015-9967-y>
- Howe JA, Wang H, Fischmann TO, Balibar CJ, Xiao L, Galgoci AM, Malinverni JC, Mayhood T, Villafania A, Nahvi A, Murgolo N, Barbieri CM, Mann PA, Carr D, Xia E, Zuck P, Riley D, Painter RE, Walker SS, Sherborne B, de Jesus R, Pan W, Plotkin MA, Wu J, Rindgen D, Cummings J, Garlisi CG, Zhang R, Sheth PR, Gill CJ, Tang H, Roemer T (2015) Selective small-molecule inhibition of an RNA structural element. *Nature* 526:672–677. <https://doi.org/10.1038/nature15542>
- Hwang TL, Shaka AJ (1995) Water suppression that works. Excitation sculpting using arbitrary wave-forms and pulsed-field gradients. *J Magn Reson Ser A* 112:275–279. <https://doi.org/10.1006/jmra.1995.1047>
- Imai S, Kumar P, Hellen CU, D'Souza VM, Wagner G (2016) An accurately preorganized IRES RNA structure enables eIF4G capture for initiation of viral translation. *Nat Struct Mol Biol* 23:859–864. <https://doi.org/10.1038/nsmb.3280>
- Johnson BA, Blevins RA (1994) NMR view: a computer program for the visualization and analysis of NMR data. *J Biomol NMR* 4:603–614. <https://doi.org/10.1007/BF00404272>
- Jucker FM, Heus HA, Yip PF, Moors EH, Pardi A (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J Mol Biol* 264:968–980. <https://doi.org/10.1006/jmbi.1996.0690>
- Kao C, Zheng M, Rüdiger 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. <https://doi.org/10.1017/s1355838299991033>
- Krähenbühl 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. <https://doi.org/10.1007/s10858-014-9841-3>
- Kwok CK, Lam SL (2013) NMR proton chemical shift prediction of T•T mismatches in B-DNA duplexes. *J Magn Reson* 234:184–189. <https://doi.org/10.1016/j.jmr.2013.06.022>
- Lam SL (2007) DSHIFT: a web server for predicting DNA chemical shifts. *Nucleic Acids Res* 35:W713–W717. <https://doi.org/10.1093/nar/gkm320>
- Lam SL, Lai KF, Chi LM (2007) Proton chemical shift prediction of A•A mismatches in B-DNA duplexes. *J Magn Reson* 187:105–111. <https://doi.org/10.1016/j.jmr.2007.04.005>
- Legault P, Hoogstraten CG, Metlitzky E, Pardi A (1998) Order, dynamics and metal-binding in the lead-dependent ribozyme. *J Mol Biol* 284:325–335. <https://doi.org/10.1006/jmbi.1998.2181>
- Marchant J, Summers MF, Johnson BA (2019) Assigning NMR spectra of RNA, peptides and small organic molecules using molecular network visualization software. *J Biomol NMR*. <https://doi.org/10.1007/s10858-019-00271-3>
- Meiler J (2003) PROSHIFT: protein chemical shift prediction using artificial neural networks. *J Biomol NMR* 26:25–37. <https://doi.org/10.1023/a:1023060720156>
- Meyer SM, Williams CC, Akahori Y, Tanaka T, Aikawa H, Tong Y, Childs-Disney JL, Disney MD (2020) Small molecule recognition of disease-relevant RNA structures. *Chem Soc Rev* 49:7167–7199. <https://doi.org/10.1039/d0cs00560f>
- Ng KS, Lam SL (2015) NMR proton chemical shift prediction of C•C mismatches in B-DNA. *J Magn Reson* 252:87–93. <https://doi.org/10.1016/j.jmr.2015.01.005>
- Norris M, Fetter B, Marchant J, Johnson BA (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>
- Papasaikas P, Valcárcel 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 MJ, Porter FW, 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>

- Pham VV, Salguero C, Khan SN, Meagher JL, Brown WC, Humbert N, de Rocquigny H, Smith JL, D'Souza VM (2018) HIV-1 Tat interactions with cellular 7SK and viral TAR RNAs identifies dual structural mimicry. *Nat Commun* 9:4266. <https://doi.org/10.1038/s41467-018-06591-6>
- Popenda M, Błażewicz M, Szachniuk M, Adamiak RW (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, Błażewicz M, Wasik S, Burke EK, Błażewicz J, Adamiak RW (2010) RNA FRABASE 2.0: an advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures. *BMC Bioinform* 11:231. <https://doi.org/10.1186/1471-2105-11-231>
- Rodnina MV (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 GS (2001) Calculation of ^{13}C chemical shifts in rna nucleosides: structure- ^{13}C chemical shift relationships. *J Magn Reson* 151:1–8. <https://doi.org/10.1006/jmre.2001.2350>
- Sheehy JP, Davis AR, Znosko BM (2010) Thermodynamic characterization of naturally occurring RNA tetraloops. *RNA* 16:417–429. <https://doi.org/10.1261/rna.1773110>
- Sripakdeevong P, Cevce M, Chang AT, Erat MC, Ziegeler M, Zhao Q, Fox GE, Gao X, Kennedy SD, Kierzek R, Nikonowicz EP, Schwalbe H, Sigel RK, Turner DH, Das R (2014) Structure determination of noncanonical RNA motifs guided by ^1H NMR chemical shifts. *Nat Methods* 11:413–416. <https://doi.org/10.1038/nmeth.2876>
- Theimer CA, 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 EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Kent Wenger R, Yao H, Markley JL (2008) BioMagResBank. *Nucleic Acids Res* 36:D402–408. <https://doi.org/10.1093/nar/gkm957>
- Vallurupalli P, Moore PB (2003) The solution structure of the loop E region of the 5S rRNA from spinach chloroplasts. *J Mol Biol* 325:843–856. [https://doi.org/10.1016/s0022-2836\(02\)01270-6](https://doi.org/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. <https://doi.org/10.1038/s41467-021-21840-x>
- Warner KD, Hajdin CE, Weeks KM (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>
- Winkle M, El-Daly SM, Fabbri M, Calin GA (2021) Noncoding RNA therapeutics - challenges and potential solutions. *Nat Rev Drug Discov*. <https://doi.org/10.1038/s41573-021-00219-z>
- Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD (1995) ^1H , ^{13}C , and ^{15}N chemical shift referencing in biomolecular NMR. *J Biomol NMR* 6:135–140. <https://doi.org/10.1007/BF00211777>
- Zhang K, Frank AT (2020) Conditional prediction of ribonucleic acid secondary structure using chemical shifts. *J Phys Chem B* 124:470–478. <https://doi.org/10.1021/acs.jpcc.9b09814>
- Zhang K, Abdallah K, Ajmera P, Finos K, Looka A, Mekhael J, Frank AT (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>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.