Solution NMR readily reveals distinct structural folds and interactions in doubly $^{13}$C- and $^{19}$F-labeled RNAs

Owen B. Becette1, Guanghui Zong1, Bin Chen1, Kehinde M. Taiwo1, David A. Case2, T. Kwaku Dayie1*

RNAs form critical components of biological processes implicated in human diseases, making them attractive for small-molecule therapeutics. Expanding the sites accessible to nuclear magnetic resonance (NMR) spectroscopy will provide atomic-level insights into RNA interactions. Here, we present an efficient strategy to introduce $^{19}$F-$^{13}$C spin pairs into RNA by using a 5-fluorouridine-5′-triphosphate and T7 RNA polymerase–based in vitro transcription. Incorporating the $^{19}$F-$^{13}$C label in two model RNAs produces linewidths that are twice as sharp as the commonly used $^1$H-$^{13}$C spin pair. Furthermore, the high sensitivity of the $^{19}$F nucleus allows for clear delineation of helical and nonhelical regions as well as GU wobble and Watson-Crick base pairs. Last, the $^{19}$F-$^{13}$C label enables rapid identification of a small-molecule binding pocket within human hepatitis B virus encapsidation signal epsilon (hHBV ε) RNA. We anticipate that the methods described herein will expand the size limitations of RNA NMR and aid with RNA-drug discovery efforts.

INTRODUCTION

RNAs form essential regulators of biological processes and are implicated in human diseases, making them attractive therapeutic targets (1, 2). This extensive functional diversity of RNA derives from its ability to fold into complex three-dimensional (3D) structures. Yet, the number of noncoding RNA sequences far outstrips the number of solved RNA structures deposited in the Protein Data Bank (PDB) necessary for understanding RNA function (3, 4). In comparison to x-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy provides high-resolution structural and dynamic information in solution, making it an ideal biophysical technique to characterize the interactions between target RNAs and small drug-like molecules. Nonetheless, NMR studies of RNA suffer from poor spectral resolution and sensitivity, both of which worsen with increasing molecular weight. In contrast with proteins, which are made up of 20 unique amino acid building blocks, RNAs are composed of only four aromatic residues. These four resonate over a very narrow chemical shift region. At high magnetic field strengths, sizable transverse relaxation rates ($R_2$) cause line broadening and thereby decrease both sensitivity and resolution. These problems are further exacerbated with increasing molecular weight. To overcome these limitations of RNA, novel labeling strategies that expand the number of NMR probes beyond the traditional nonradioactive and stable isotope labels such as hydrogen-1 (1H), phosphorus-31 (31P), carbon-13 (13C), hydrogen-2 (2H), and nitrogen-15 (15N) are needed.

Solution NMR of the magnetically active fluorine-19 ($^{19}$F) isotope offers clear advantages in the study of RNA structure and conformational changes, which occur upon ligand binding. $^{19}$F has high NMR sensitivity (0.83 of 1H) due to a large gyromagnetic ratio that is comparable to 1H (0.94 of 1H), a 100% natural abundance, and ~6 wider chemical shift dispersion than 1H (5, 6). In addition, $^{19}$F is also sensitive to changes in its local chemical environment (5, 6). In contrast with other commonly used NMR nuclei (1H/31P/13C/15N), $^{19}$F is virtually absent in biological systems, thereby rendering $^{19}$F NMR background free. Together, $^{19}$F is an attractive probe for incorporation into nucleic acids to study their structure, interactions, and dynamics in solution.

Given its attractive spectroscopic properties, $^{19}$F was incorporated into RNA for NMR studies in the 1970s (7–9). Since then, $^{19}$F has been successfully incorporated into DNA and RNA oligonucleotides for NMR analysis and used to probe RNA and DNA structure, conformational exchange, and macromolecular interactions (10, 11). Most of these studies were conducted on short oligonucleotides (~30 nucleotides (nt)) prepared by solid-phase synthesis with only a few residues $^{19}$F labeled. Even when 2-fluoroadenine (2FA) was incorporated into a 73-nt (~22 kDa) guanine-sensing riboswitch, only 4 of the 16 signals could be assigned. This 2FA study hinted at the limitations of $^{19}$F NMR for large RNAs (12). Despite its attractiveness, the application of $^{19}$F NMR to study RNA has remained limited because the large $^{19}$F chemical shift anisotropy (CSA) contributes substantially to line broadening as a function of increasing molecular weight and polarizing magnetic fields.

To circumvent this limitation, Boeszormenyi et al. (13) recently showed that direct coupling of $^{19}$F to $^{13}$C allowed for cancelation of CSA and dipole-dipole (DD) interactions. By incorporating this $^{19}$F-$^{13}$C spin pair into aromatic moieties of proteins and a 16-nt DNA, they showed that a transverse relaxation optimized spectroscopy (TROSY) version of a $^{19}$F-$^{13}$C heteronuclear single-quantum coherence (HSQC) (13) provided improved spectroscopic properties. These exciting results hinted that installing $^{13}$C-$^{19}$F pairs in RNA nucleobases should also lead to improved spectroscopic features.

However, there were no facile methods to readily incorporate $^{19}$F-$^{13}$C spin pairs into RNA. To overcome this technical obstacle of incorporating fluorinated aromatic moieties into RNA, we provide here a straightforward chemoenzymatic synthesis of [5-$^{19}$F, 5-$^{13}$C]-uridine 5′-triphosphate (5FUTP) for incorporation into RNA (Fig. 1) using phage T7 RNA polymerase–based in vitro transcription. To showcase its versatility, we transcribed two model RNAs using these labels: the 30-nt (~10-kDa) human immunodeficiency type 2 transactivation response (HIV-2 TAR) element (6, 14) and the 61-nt (~20-kDa) human hepatitis B virus encapsidation signal epsilon (hHBV ε) element (Fig. 1) (15, 16).
can reduce the 13 C linewidth by a factor >2, compared to a 13 C-1 H pair, using unlabeled potassium cyanide, 13 C-labeled bromoacetic acid, and the molecular weight of the RNA, the TROSY effect in the 19 F-13 C pair regions. With protons substituted with deuterium and depending on widths of all four components. 

...with previous studies (13). Unlike C-H spectra, the resolving power...
The chemical shift dispersion of 19F compared with 1H and similar dispersion in 13C (Fig. 3). All six correlations of HIV-2 TAR are well resolved for both 1H-13C and 19F-13C correlations and are in agreement with previously published 1H-19F and 1H-13C RNA spectra (6, 24, 33). Nonetheless, even for this small RNA, the 19F-13C spin pair markedly improves the spectral resolution. 5FU HIV-2 TAR shows a chemical shift dispersion of 2.6 parts per million (ppm) in the 19F dimension and only 0.5 ppm in the 1H dimension for WT (Fig. 3, A and B). Replacing 1H with 19F at C5 results in a slight reduction in chemical shift dispersion along the 13C dimension from 2.1 to 1.5 ppm, although this effect is much smaller than the gain in resolution for 19F over 1H (Fig. 3, A and B). Similarly, the 19F resonances of 5FU hHBV ε are spread over 4.5 ppm, whereas the WT 1H signals resonate over a narrow 0.8-ppm window. This represents 5.7 times better dispersion (Fig. 3, C and D). Again, substitution of 1H with 19F at C5 results in a reduction in chemical shift dispersion of 2.3 to 1.7 ppm along the 13C dimension for hHBV ε (Fig. 3, C and D). Of the anticipated 18 signals for hHBV ε, 16 are resolved for WT and 17 for 5FU. Together, these results demonstrate the marked gain afforded by the 19F-13C spin pair in 5FU RNAs compared with the 1H-13C spin pair in WT.

In addition to this considerable gain in resolution, 19F-13C labeling confers favorable 13C TROSY linewidths. We compared the relative linewidths for both RNAs, which we assume to be Lorentzian (Figs. 4 and 5). For 5FU HIV-2 TAR, the 13C TROSY linewidths were 1.5 times sharper on average than the anti-TROSY components, with a range of 1.3 to 1.7 (Fig. 4A). For WT HIV-2 TAR, the 13CH TROSY component was 3.7-fold narrower than the anti-TROSY component (range, 1.6 to 8.7) (Fig. 4B). Similarly, for 5FU HBV ε, the 13CF TROSY linewidths were 2.2-fold narrower than the anti-TROSY ones over a range of 1.5 to 3.3 (Fig. 4C). For WT HBV ε, only 5 of the 16 13CH anti-TROSY signals were observed and were 2.6 times broader than the TROSY resonances (range, 2.0 to 3.3) (Fig. 4D). As predicted from our simulations (Fig. 2), the 13CF TROSY component relaxes ~2 times slower than the 13CH TROSY component in both HIV-2 TAR and hHBV ε. The 13FC TROSY linewidths for 5FU HIV-2 TAR and 5FU HBV ε were 1.4 (range, 1.3 to 1.6) and 1.6 (range, 1.1 to 2.5) times narrower than the anti-TROSY components, respectively (Fig. 5, A and C). For both WT HIV-2 and WT HBV ε, the 13CH TROSY and anti-TROSY linewidths were comparable (Fig. 5, B and D). Consistent with our simulations, the 13FC TROSY linewidth is ~2-fold larger than that of the 13CH component for both RNAs (Fig. S3). Again, this is in line with the poor performance of 19F NMR experiments due to the large CSA-induced relaxation. Thus, the incorporation of the 13C label mitigates the deleterious relaxation of the 19F nuclei within a 19F-13C spin pair. However, even for medium-sized RNAs ~20 kDa, 19F TROSY detection of the 19F-13C spin pair still outperforms that for a 1H-13C
In addition to these gains in resolution and favorable linewidths, the helical residues U38, U40, and U42 of 5FU HIV-2 TAR (Fig. 3C) are centered around ~−167.5 ppm in 19 F and ~141.5 ppm in 13 C in the 13 C dimension for a 1 H-13 C spin pair. Therefore, to reap the maximum benefits of this label, it is advantageous to monitor the 13 C nuclei rather than the 19 F nuclei. We anticipate that the 19 F-13 C TROSY effect will continue to scale with molecular weight for RNAs as was seen recently with proteins (13) and our simulations.

19F chemical shifts reveal RNA structural equilibria

In addition to these gains in resolution and favorable linewidths, previous work suggested the 19 F chemical shifts serve as sensitive markers of RNA secondary structure (10, 11). For example, GU wobble base pairs are deshielded and shifted by ~4.5 ppm to lower fields compared with AUs within Watson-Crick geometries (34). On the basis of these earlier observations, we hypothesized that 19 F-13 C correlations of HIV-2 TAR and hHBV ε can be grouped on the basis of whether or not they are in helical, nonhelical, or GU base-paired regions of the RNA. As a positive control, we note that nonhelical U23, U25, and U31 in 5FU HIV-2 TAR resonate around ~−165.5 ppm in 19 F and ~142.5 ppm in 13 C (Fig. 3A). On the other hand, the helical residues U38, U40, and U42 of 5FU HIV-2 TAR are centered around ~−167.5 ppm in 19 F and ~141.5 ppm in 13 C in line with previous observations for 19 F-1 H samples of HIV-2 TAR (6) and tRNA (34). Comparison of the equivalent 1 H-13 C spectra shown in Fig. 3B indicates that even though helical residues cannot be distinguished from nonhelical residues, nonhelical residues cannot be differentiated from GU base pairs for a 1 H-13 C spin pair. Thus, the spectroscopic discrimination of helical and nonhelical regions as well as GU wobble and Watson-Crick base pairs in RNA structures becomes possible with the high sensitivity of 19 F to the local chemical environment of a 19 F-13 C spin pair. This distinguishing feature is not readily available for a 1 H-13 C spin pair.

19F chemical shift perturbation enables facile identification of site-specific RNA binders

Ligand-based (35) and protein-observed (36) 19 F NMR screening methods are important for identifying small drug-like molecules that act as protein inhibitors. Although most work to date has focused on proteins, recent work suggests that RNAs also contain specific binding pockets that could be easily distinguished and targeted with small molecules (1, 2). hHBV ε is at the center of the viral replication cycle since the first two residues in its internal bulge are used
by the virus to initiate synthesis of the minus-strand DNA. Thus, targeting this RNA structure will notably expand the repertoire of HBV drug targets beyond the current focus on viral proteins (37). Given $^{19}$F chemical shifts serve as sensitive markers of RNA secondary structure, we reasoned that $^{19}$F-$^{13}$C spectroscopy will likely pinpoint loop over helical region binders. Rather satisfyingly, we found a small molecule that specifically binds a subset of nonhelical residues in 5FU hHBV ε (Fig. 6). Overlay of the full spectra of 5FU hHBV ε with and without the small-molecule shows chemical shift perturbations (CSPs) (38) predominantly confined to nonhelical regions (Fig. 6). Within the nonhelical residues, only four of the seven signals shift with the addition of the small molecule, which suggests selectivity for certain nonhelical residues over others (Fig. 6). We propose a model whereby our small molecule binds hHBV ε in the 6-nt bulge formed between C14 and C19, but not anywhere else in the RNA. The minor CSPs seen in the helical portion of the 5FU hHBV spectra are from U residues flanking the 6-nt bulge, specifically U47, U48, and U49. Last, the CSP seen in the GU portion is from U12, which also flanks our proposed binding pocket.

**DISCUSSION**

$^{19}$F is an attractive spectroscopic probe to study biomolecular structure, interactions, and dynamics in solution. Nonetheless, a number of obstacles must be overcome for it to become widely useful. First, we must be able to easily install the label into any biopolymer. While incorporation of fluorinated aromatic amino acids and nucleobases into proteins and nucleic acids is usually not a technical challenge, until now, synthesis of carbon-labeled and fluorinated nucleobase to create a $^{19}$F-$^{13}$C spin pair has been problematic for RNA. Here, we present a facile strategy to incorporate $^{19}$F-$^{13}$C 5-fluorouridine into RNA using in vitro transcription for characterization of small-molecule binding interactions by NMR. Our protocol to prepare $^{19}$F-$^{13}$C 5-fluorouridine-5′-triphosphate (5FUTP) involves chemically synthesizing 5FU and then enzymatically coupling it to $^{13}$C-labeled D-ribose. Our synthetic strategy can be generalized to selectively place labels in the pyrimidine nucleobase at either $^{15}$N1, $^{15}$N3, $^{13}$C2, $^{13}$C4, $^{13}$C5, or $^{13}$C6 or any combinations thereof, and then enzymatically couple ribose labeled at either $^{13}$C1, $^{13}$C2, $^{13}$C3, $^{13}$C4, or $^{13}$C5 or any of the preceding ribose combinations to the base. The resulting isotopically enriched 5FUTP is then readily incorporated into any desired RNA using DNA template–directed T7 RNA polymerase–based in vitro transcription. This enzymatic approach, unlike solid-phase RNA synthesis, is not limited to RNAs less than 70 nt or to nucleotides made of labeled nucleobase coupled to unlabeled ribose. Although fluorine substitution at C5 in pyrimidines strongly affects the shielding of the nearby H6, it has little effect on the anomic H1’ chemical shifts (24). We therefore anticipate that our unique strategy that combines ribose $^{13}$C1’ label with $^{19}$F-$^{13}$C uracil should allow the transfer of assignments from unmodified RNAs to 5-fluoropyrimidine–substituted RNAs made with our labels.

Second, because of van der Waals radii comparable to that of $^{1}$H, $^{19}$F is considered minimally perturbing when incorporated into bio-polymers (24). Although fluorine substitution in 5FU RNAs leads to sizeable line broadening of the imino protons, thermal melting analysis indicates that the 5FU RNAs are thermodynamically equivalent to the nonfluorinated RNAs (6, 7, 24). In future work, it will be important to systematically investigate the effect of fluorine substitution not only on thermodynamic stability but also on folding kinetics of RNAs. Insights derived from solving, at high-resolution, the 3D structures of fluorinated and nonfluorinated RNA could potentially guide the use of these spin pairs to spy on the biological processes within the cell.

Third, despite its huge potential, nucleic acid observed $^{19}$F (NOF) NMR has remained underused because the large $^{19}$F CSA induces severe line broadening at high molecular weights and magnetic fields. Using DFT calculations of CST parameters, we show that an optimal $^{19}$F-$^{13}$C TROSY enhancement occurs at 600-MHz $^{1}$H frequency to enable slow relaxation of $^{13}$C bonded to $^{19}$F. Our RNAs show an enhanced $^{19}$F-$^{13}$C TROSY effect with increasing molecular weight and $^{13}$C linewidths that are twice as sharp as seen with traditional $^{1}$H-$^{13}$C spin pairs. Thus, nucleobase $^{19}$F-$^{13}$C TROSY will expand the applicability of RNA NMR beyond the ~30-nt (~10-kDa) average.

Fourth, the RNA secondary structure is made up of segments of nucleotides that are either base paired or not. The arrangements of base-paired with unpaired regions can leave distinct NMR chemical shift signatures that can provide low-resolution structural information with minimum expenditure of time and cost. For example, the H5 of a pyrimidine is sensitive to the nature of the residue that comes before it within a triplet of canonical Watson-Crick AU and GC base pairs. When the A in a central UA base pair is substituted by a G, the H5 resonance shifts downfield because of the formation of the GU base pair. Yet, an analysis of the commonly used $^{1}$H-$^{13}$C probes fails to unambiguously separate nonhelical residues from helical ones (39). In contrast, the $^{19}$F-$^{13}$C labels resonate in distinct chemical shift regions based on their secondary structure. For instance, nonhelical residues resonate in spectral regions distinct from

---

**Fig. 6. Small-molecule binding to 5FU hHBV ε.** (A) Overlay of $^{19}$F-$^{13}$C TROSY spectra for hHBV ε without (black) and with small molecule (SM, magenta). (B) Zoom-in of nonhelical residues showing chemical shift perturbations (CSPs) upon addition of SM. (C) Quantification of the CSPs upon addition of SM. The average (Ave) CSP is shown as a dashed line.
helical ones, which are further separated into GU wobble and AU Watson-Crick base-paired regions. The ability to differentiate between different structural features in an RNA simply based on chemical shifts removes the need for the time-consuming and laborious process of resonance assignment.

Given the ubiquity and functional importance of GU wobble base pairs (40) in all kingdoms of life (41), the ability to easily distinguish GU from canonical GC and AU base pairs has several important implications. For instance, in the minor groove, a GU base pair presents a distinctive exocyclic amino group that is unpaired and the U’s C1’ atom rotates counterclockwise compared with the C’s C1’ atom in a canonical GC base pair. This region serves as an important site for protein–RNA interactions. Similarly, in the major groove, G N7 and O6 together with U O4 create an area of intense negative electrostatic potential conducive for binding divalent metal ions. Furthermore, all canonical Watson-Crick base pairs are circumscribed by ~10.6-Å diameters formed by a line connecting their C1’-C1’ centers. These ribose-connected centers are superimposable with almost perfect alignment. In contrast, a GU base pair is misaligned counterclockwise by a residual twist of +14°, and an UG base pair is misaligned clockwise by a residual twist of −11° (42). That is, the GU base pair is not isosteric with canonical Watson-Crick pairs. Rather, these wobble base pairs either overtwist or undertwist the RNA double helix. 19F-13C labeling might aid in elucidating the structural and dynamic basis of these twists depending on the identity of the base pairs neighboring the wobble pair. We, therefore, anticipate that our new label could potentially open up avenues for probing GU wobble pairs in various structural contexts outlined above, such as 19F-13C-labeled RNA-protein interactions and metalloribosome interactions.

In summary, the labeling technologies presented here open the door for characterizing the structure, dynamics, and interactions of RNA, RNA-RNA, RNA-DNA, RNA-protein, and RNA-drug complexes in vitro and in vivo for complexes as large as 100 kDa or higher with the appropriate fluorine NMR hardware. This 19F-13C labeling approach will also enable correlating chemical shift–structure relationships to aid chemical shift–centered probing of RNA structure, dynamics, and interactions. We envision that the 19F-13C spin pair, by providing a clear demarcation of RNA structural elements, may facilitate the discovery and identification of small drug-like molecules that target RNA binding pockets in vitro and in vivo.

MATERIALS AND METHODS
The full description of Materials and Methods can be found in the Supplementary Materials. A brief summary is provided here.

Chemoenzymatic synthesis of SFUTPs
[5-19F, 5-13C, 6-2H]− and [5-19F, 5-13C, 6-2H, 1,3-15N2]−SFU were synthesized from unlabeled potassium cyanide, 13C-labeled bromoacetic acid, and 15N-labeled urea as described elsewhere (3, 17, 18, 24). The resulting uracil was converted to SFU by direct fluorination with Selectfluor and deuteration (19–21). [1’’5-13C2, 5-19F, 6-2H]−SFUTP and [1’’5-13C2, 5-19F, 6-2H, 1,3-15N2]−SFUTP were synthesized using PPP enzymes (3, 4, 6, 22, 24, 43).

RNA in vitro transcription
All RNAs were prepared by in vitro transcription and purified as previously described (3, 4). RNA concentrations were approximated by UV absorbance using extinction coefficients of 387.5 mM⁻¹ cm⁻¹ for HIV-2 TAR and 768.3 mM⁻¹ cm⁻¹ for hHBV e. All RNA concentrations were >0.5 mM (~0.3 ml) in Shigemi NMR tubes.

Thermal melt analysis
We collected thermal melting profiles for both WT and 5FU-substituted HIV-2 TAR and hHBV e as previously described (24, 25).

Electronic structure calculations
Calculations were carried out on 1-methyl-uracil and 5-fluorouracil, 1-methyl uracil using optimized geometries (44). All calculations used the Gaussian-16 program (29). Details are provided in the Supplementary Materials.

Solution NMR spectroscopy
All 19F-13C TROSY spectra were collected at 298 K using a Bruker 600 MHz Avance III spectrometer equipped with TXI (triple resonance inverse) and BBI (broad band inverse) probes. All data were processed with Bruker’s Topspin 4.0.7 software. 1H chemical shifts were internally referenced to DSS (0.00 ppm), with the 13C chemical shifts referenced indirectly using the gyromagnetic ratios of 13C/1H (44). The 19F chemical shifts were internally referenced to trifluoroacetic acid (~75.51 ppm) (45). Experiments showing each component of the 1H/19F/13C correlations were adapted from a sensitivity- and gradient-enhanced 1H-15N TROSY used for proteins (31).

REFERENCES AND NOTES


Acknowledgments: We thank P. Deshong, J. Kahn, L.-X. Wang, and P. Y. Zavalij (University of Maryland) and H. Arthanari (Harvard University) for the helpful comments. We thank S. Bentz and D. Oh for help in preparing samples for thermal melt analysis, and M. Swirydova for help in analyzing samples by mass spectrometry. Funding: We thank the National Science Foundation (DBI1040158 to T.K.D. for NMR instrumentation) and the NIH (U54AI05470 to T.K.D. and D.A.C.) for support. Author contributions: T.K.D.: conceptualization. T.K.D. and O.B.B.: implementation of the project and manuscript preparation. G.Z., B.C., K.M.T., and T.K.D.: database curation. O.B.B., H. Wagner-Rundell, P. Chhabra, A. Marenich, A. Petersson, and H. Johnson, Database proton NMR chemical shifts for RNA datasets. V. Marenich, A. Petersson, and H. Johnson, Database proton NMR chemical shifts for RNA datasets. J. Biomol. NMR 55, 33–46 (2013).

Submitted 6 May 2020 Accepted 18 August 2020 Published 7 October 2020 10.1126/sciadv.abc6572

Solution NMR readily reveals distinct structural folds and interactions in doubly $^{13}$C- and $^{19}$F-labeled RNAs

Owen B. Becette, Guanghui Zong, Bin Chen, Kehinde M. Taiwo, David A. Case and T. Kwaku Dayie