An RNA dynamic ensemble at atomic resolution

Honglue Shi[1], Atul Rangadurai[2], Hala Abou Assi[2,3], Rohit Roy[4], David A. Case[5]*, Daniel Herschlag[6,7,8]*, Joseph D. Yesselman[9]*, and Hashim M. Al-Hashimi[1,2]*

Affiliations:

[1]Department of Chemistry, Duke University, Durham NC 27710 USA

[2]Department of Biochemistry, Duke University School of Medicine, Durham NC 27710 USA

[3]Department of Medicine, Duke University School of Medicine, Durham NC 27710 USA

[4]Center for Genomics and Computational Biology, Duke University School of Medicine, Durham NC 27710 USA

[5]Department of Chemistry and Chemical Biology, Rutgers University, Piscataway NJ 08854 USA

[6]Department of Biochemistry, Stanford University, Stanford CA 94305 USA

[7]Department of Chemical Engineering, Stanford University, Stanford CA 94305 USA

[8]ChEM-H Institute, Stanford University, Stanford, CA 94305 USA

[9]Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE 68588 USA

*Correspondence to: hashim.al.hashimi@duke.edu, jyesselm@unl.edu, herschla@stanford.edu, david.case@rutgers.edu

Abstract: Solving dynamic ensembles of biomolecules at atomic resolution is a major challenge in structural biology because it requires information which far exceeds that available from experimental measurements. We addressed the data gap in RNA ensemble determination by using structure prediction tools that leverage the growing database of RNA structures to generate a conformational library. Library refinement with NMR residual dipolar couplings enabled determination of an atomic-resolution ensemble for HIV-1 TAR as confirmed by quantum-mechanical calculations of NMR chemical shifts, comparison to a crystal structure of a substate, and atomic mutagenesis. The ensemble reveals how bulge residues cooperatively flip out and undergo sugar repuckering to allow the adjoining helices to stack. The generality of this approach will make determination of atomic-resolution RNA ensembles routine.

Main Text:

Biomolecules do not fold into a single 3D structure but rather form dynamic ensembles of many inter-converting conformations (*1*). Knowledge of dynamic ensembles is key for understanding how biomolecules fold and function, and for rationally manipulating their activities in drug discovery and synthetic biology (*2-4*). However, determining dynamic ensembles at atomic resolution remains an outstanding challenge in structural biology and biophysics, particularly for highly flexible ribonucleic acids (RNAs) (*4-7*).

The information required to specify the position of all atoms in thousands of conformations in an ensemble far exceeds the information content of experimental measurements. NMR spectroscopy is a rich source of ensemble-averaged measurements, and in combination with computational modelling, has been applied with success to determine ensembles of proteins at atomic resolution (*8-10*). In contrast, fewer atomic-level experimental measurements are typically available for characterizing nucleic acids ensembles. Moreover, while computational modeling methods such as molecular dynamics (MD) simulations are needed to address the experimental data gap (*11*), nucleic acid force fields remain underdeveloped relative to proteins, and MD simulations of RNAs often poorly predict experimental data even for simple motifs (*12, 13*). Consequently, relative to proteins, there is a greater danger of over-fitting an RNA ensemble, and assessing ensemble accuracy is more difficult.

To address the data gap, as well as dramatically simplify and accelerate RNA ensemble determination, we took advantage of structure prediction tools that leverage the growing database of RNA structures to directly generate a conformational library from a secondary structure that broadly samples energetically favorable 3D conformations. We used Fragment Assembly of RNA with Full-Atom Refinement (FARFAR) (*14*), given its high performance in extensive tests of blind

prediction of 3D RNA structure (*15*). We then determined RNA ensembles by using NMR residual dipolar coupling (RDC) data (*16-18*) to guide selection of conformers from the FARFAR library (*7, 13, 19*).

We tested our approach on the transactivation response element (TAR) (Fig. 1A) from HIV-1 (*20, 21*), which has served as a model system for bulge motifs. Bulges are one of the most common RNA secondary structural elements that act as dynamic joints connecting helical elements, enabling their relative orientation to change adaptively during folding and function (*13, 20-22*). We used FARFAR to directly generate a conformational library (N = 10,000) from an input TAR secondary structure only constraining Watson-Crick base pairs (bps) inferred by NMR, and assuming an idealized A-form geometry (*23*) for these bps while predicting the structure for all remaining nucleotides (Methods). We then tested and optimized the FARFAR-library using a rich dataset of four independent RDCs (~8 RDCs per nucleotide) previously measured for four TAR molecules variably elongated to modulate alignment relative to the NMR magnetic field (*7, 13*).

Intriguingly, the FARFAR-library showed better agreement (RMSD 8.0 versus 8.6 Hz) with the RDCs (Fig. 1B) as compared to a previously reported (*13*) TAR library (Anton-MD) generated by subjecting an experimentally determined NOE-based NMR structure of TAR (PDB ID 1ANR) (*20*) to MD simulations with the CHARMM36 force field (*24*) (Fig. 1C). An optimized ensemble with N = 20 conformers (fig. S1) was generated (Methods) by using the RDC agreement to guide selection of conformers from the FARFAR-library (Fig. 1D) (*13, 19*). The optimized FARFAR-NMR ensemble also better predicted the RDCs (RMSD 3.1 Hz) relative to the optimized Anton-MD-NMR ensemble (RMSD 3.6 Hz) obtained using a similar procedure and the Anton-MD pool (Fig. 1E and fig. S1). Cross validation (*13, 25*) showed that the improved RDC

agreement was not due to over-fitting (fig. S2). Similar RDC agreement (Fig. 1F) was obtained when using FARFAR-NMR to generate ensembles for TAR in the presence of Mg^{2+}, and three additional TAR mutants (*22*) containing one (U1-TAR), two (U2-TAR), and seven (U7-TAR) bulge nucleotides in the presence and absence of Mg^{2+} (fig. S3 to S4, and Supplementary Text), establishing the generality of the approach.

The improved agreement observed with the FARFAR-NMR ensemble was surprising given that the global inter-helical distribution (*26*) of the Anton-MD-NMR TAR ensemble has been independently validated via X-ray scattering interferometry (XSI) (*27*). Indeed, the FARFAR-NMR and Anton-MD-NMR ensembles show comparable agreement with the RDCs measured for helical bps (Fig. 1, D and E) and the two ensembles sample similar inter-helical orientational distributions (Fig. 1G and fig. S5).

Rather, the improved agreement was primarily driven by the RDCs measured in the locally more flexible bulge residues. For the FARFAR-NMR ensemble, the RDC RMSD (1.8 Hz) for bulge residues (Fig. 1D) was within experimental error (~2 Hz), but was substantially higher (3.7 Hz) for the Anton-MD-NMR ensemble (Fig. 1E). The bulge RDC RMSD could not be improved by running MD simulations using different force fields (Fig. 1H), indicating that the improved conformational sampling in the FARFAR-library makes it possible to surpass the accuracy with which the TAR bulge could be described using conventional MD simulations. This improved performance is particularly noteworthy when considering that determining a high-resolution structure and running MD can take several months and often years whereas the FARFAR-library is generated within 24 hours running on 100 cores in parallel.

To further evaluate the accuracy of the FARFAR-NMR TAR ensemble, we substantially expanded the breadth and depth of atomic-level experimental data that can be brought to bear when

evaluating the accuracy of RNA ensembles by predicting ensemble-averaged ^{1}H, ^{13}C, and ^{15}N chemical shifts (~15 chemical shifts per nucleotide) using quantum-mechanical AF-QM/MM calculations (*28, 29*). Although rich in structural information that is complimentary to that obtained from RDCs, the agreement between measured chemical shifts and values predicted from crystal structures of nucleic acids has traditionally been poor (*28, 30*). We recently showed in studies of DNA duplexes that at least some of this disagreement originates from neglecting ensemble averaging when predicting ^{13}C chemical shifts (*29*). This revelation and the improved FARFAR-NMR TAR ensemble led us to test this approach on flexible RNAs, and to extend its reach by incorporating ^{1}H and ^{15}N shifts in addition to ^{13}C chemical shifts (Fig. 2A).

Remarkably, good agreement was observed between the measured ^{1}H, ^{13}C, and ^{15}N base and sugar chemical shifts and values back-calculated for the FARFAR-library (fig. S6 to S7 and Supplementary Text). The agreement improved substantially for the FARFAR-NMR ensemble following RDC optimization for 48% of the atom-types analyzed, while the agreement was unaffected for the remaining atom-types (Fig. 2B and fig. S6 to S7). The agreement deteriorated substantially for any one conformer member of the ensemble, underscoring the critical importance of ensemble-averaging (fig. S8). The agreement with an independent set of measurements not used in ensemble determination that is highly sensitive to various structural features of the base, sugar, and backbone strongly suggests that FARFAR-NMR describes the TAR ensemble with atomic accuracy (< 2 Å, Methods). In sharp contrast, the agreement was much weaker for the Anton-MD-NMR ensemble relative to FARFAR-NMR for 90% of the atom-types, particularly for sugar chemical shifts, some of which show no apparent correlation even following RDC optimization (Fig. 2, C and D, and fig. S6 to S7). These results establish the utility of chemical

shifts in RNA ensemble determination and also show that the improved accuracy of the FARFAR-NMR ensemble relative to Anton-MD-NMR is even greater than alluded to by the RDC data.

We compared the FARFAR-NMR and Anton-MD-NMR ensembles (Fig. 3A) to better understand the features responsible for the more accurate ensemble description of the TAR bulge. Approximately 75% of the conformers in the FARFAR-NMR ensemble have the canonical UCU bulge, while the remaining 25% have a non-canonical AUC bulge that forms through a single nucleotide register shift (fig. S9). In contrast, conformers in the Anton-MD-NMR ensemble sample a broader set of junction topologies some of which diverge from the NMR-derived Watson-Crick pairing (fig. S9 to S10 and Supplementary Text), and this could contribute to the poor agreement observed with the imino $^{15}N/^{1}H$ chemical shifts (Fig. 2C and fig. S6 to S7).

In addition, despite having a narrower range of junction topologies, the sugar pucker distribution (defined by angle δ) for bulge residues in the FARFAR-NMR sample a broader range (Fig. 3, B and C), are substantially enriched in the non-canonical C2′-*endo* conformation (Fig. 3D) relative to the Anton-MD-NMR ensemble, and are in better agreement with the NMR-derived (*31*) sugar puckers (Fig. 3E and fig. S11). Moreover, ~15% of the conformers in the FARFAR-NMR ensemble had all three bulge residues simultaneously in the C2′-*endo* conformation but none did in the Anton-MD-NMR ensemble (Fig. 3F). Excluding all conformers containing C2′-*endo* sugar puckers in U23, C24, U25, A22 or U40 from the FARFAR-library diminished the RDC agreement to a level similar to the Anton-MD derived ensembles again confirming that the improved agreement is not due to over-fitting of the data (fig. S12). A similar behavior was observed for the backbone torsion angle γ (Fig. 3B), which shows greater sampling of non-*gauche*+ angles in the FARFAR relative to the Anton-derived ensembles (Fig. 3C and fig. S11 to S12). This can account

for the better agreement observed for sugar chemical shifts (*31*) in the FARFAR-NMR versus Anton-MD-NMR ensembles (Fig. 2, B to D).

What causes conformers in the FARFAR-NMR ensemble to more greatly sample non-canonical sugar-backbone conformations at the bulge relative to conformers in the Anton-MD-NMR ensemble? Unpaired pyrimidine RNA nucleotides unconstrained by other interactions are enriched in the C2′-*endo* conformation (C2′-*endo*:C3′-*endo* is 40:60) (*32*) and C3′-*endo* becomes the predominant sugar-pucker when the nucleotides form bps or stack intra-helically (*31*). Indeed, ~80% of the residues with C2′-*endo* sugar puckers were extra-helical in the FARFAR-NMR ensemble (Fig. 4A and fig. S9). Moreover, for linear inter-helical conformations, there was a strong preference to have all three bulge residues simultaneously flip out (Fig. 4B) and adopt conformations enriched in the C2′-*endo* sugar pucker (Fig. 4C); the bulge residues flip out to allow the two helices to coaxially stack (Movie S1 and fig. S9). Such coaxial conformations have previously been hypothesized to exist within the TAR ensemble based on the Mg^{2+} dependence of the inter-helical ensemble (*22*) and a crystal structure of Ca^{2+} bound TAR (*33*).

Strikingly, two of the coaxial conformations in the FARFAR-NMR ensemble superimpose with the TAR crystal structure with a heavy-atom RMSD of 1.3-1.4 Å (Fig. 4D). Thus, the TAR crystal structure captures substates of the ensemble in solution. In sharp contrast, none of the conformations in the Anton-MD-NMR ensemble had the two helices coaxially stacked, and not a single conformer had all three bulge residues flipped out and having the C2′-*endo* sugar pucker (Fig. 4, A to C and fig. S9). The FARFAR-NMR ensemble (Fig. 4A and Movie S1) further shows that all three bulge residues simultaneously flip out with high cooperativity (~2.3 kcal/mol, see methods) most likely because this then permits favorable coaxial stacking of the helices (Fig. 4A, fig. S9, and Movie S1). This can explain prior results (*34*) showing that a TAR mutation that

promotes inter-helical stacking results in a predominantly coaxially stacked conformation in which all three bulge residues are flipped out.

Finally, we put key atomic features of the FARFAR-NMR ensemble to a test by rationally redistributing the conformer populations using atomic mutagenesis. The ensemble shows that the kinked unstacked conformations are enriched in bulge residues that have the C3′-*endo* sugar pucker relative to the coaxial stacked conformations (Fig. 4C). To test this feature of the ensemble, we incorporated 2′-O-Methyl (Nm) modifications (Fig. 4E) at U23 or C24 to bias the sugar pucker at these positions toward C3′-*endo* by ~0.2 kcal/mol (*35*). Indeed, methylating either U23 or C24 resulted in chemical shift perturbations in 2D NMR spectra of TAR (Fig. 4F and fig. S13) throughout the bulge and neighboring residues that are specifically directed towards the chemical shifts of the unstacked conformation (*22*), as expected for a cooperative redistribution in favor of the kinked conformation.

In conclusion, FARFAR-NMR lays the foundation for a new paradigm for RNA ensemble determination by combining reliably measurable NMR data with 3D structure prediction. The approach can immediately be applied to render many existing NMR structures of RNAs that were determined using conventional approaches into dynamic ensembles, and their accuracies can be tested using the available chemical shift data. The approach is general, rapid, and can also incorporate other sources of experimental data. Given its ease of implementation and higher throughput, FARFAR-NMR has the potential to unleash the ensemble description of RNAs to all corners of biology, from which a deeper and broader understanding of folding and function will undoubtedly emerge.

References and Notes:

1. H. Frauenfelder, S. G. Sligar, P. G. Wolynes, The energy landscapes and motions of proteins. *Science* **254**, 1598-1603. (1991).
2. M. R. Jensen *et al.*, Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. *Structure* **17**, 1169-1185 (2009).
3. H. van den Bedem, J. S. Fraser, Integrative, dynamic structural biology at atomic resolution--it's about time. *Nat. Methods* **12**, 307-318 (2015).
4. L. R. Ganser, M. L. Kelly, D. Herschlag, H. M. Al-Hashimi, The roles of structural dynamics in the cellular functions of RNAs. *Nat. Rev. Mol. Cell Biol.* **20**, 474-489 (2019).
5. S. K. Denny *et al.*, High-Throughput Investigation of Diverse Junction Elements in RNA Tertiary Folding. *Cell* **174**, 377-390.e320 (2018).
6. A. Reining *et al.*, Three-state mechanism couples ligand and temperature sensing in riboswitches. *Nature* **499**, 355-359 (2013).
7. Q. Zhang, A. C. Stelzer, C. K. Fisher, H. M. Al-Hashimi, Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* **450**, 1263-1267 (2007).
8. W. Y. Choy, J. D. Forman-Kay, Calculation of ensembles of structures representing the unfolded state of an SH3 domain. *J. Mol. Biol.* **308**, 1011-1032 (2001).
9. V. Ozenne *et al.*, Mapping the Potential Energy Landscape of Intrinsically Disordered Proteins at Amino Acid Resolution. *J. Am. Chem. Soc.* **134**, 15138-15148 (2012).
10. O. F. Lange *et al.*, Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* **320**, 1471-1475 (2008).
11. S. A. Showalter, R. Brüschweiler, Quantitative molecular ensemble interpretation of NMR dipolar couplings without restraints. *J. Am. Chem. Soc.* **129**, 4158-4159 (2007).
12. J. Sponer *et al.*, RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview. *Chem. Rev.* **118**, 4177-4338 (2018).
13. L. Salmon, G. Bascom, I. Andricioaei, H. M. Al-Hashimi, A General Method for Constructing Atomic-Resolution RNA Ensembles using NMR Residual Dipolar Couplings: The Basis for Interhelical Motions Revealed. *J. Am. Chem. Soc.* **135**, 5457-5466 (2013).
14. A. M. Watkins, R. Das, FARFAR2: Improved de novo Rosetta prediction of complex global RNA folds. *bioRxiv*, 764449 (2019).
15. Z. Miao *et al.*, RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme. *RNA* **23**, 655-672 (2017).
16. N. Tjandra, A. Bax, Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**, 1111-1114 (1997).
17. J. R. Tolman, J. M. Flanagan, M. A. Kennedy, J. H. Prestegard, Nuclear Magnetic Dipole Interactions in Field-Oriented Proteins - Information For Structure Determination in Solution. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 9279-9283 (1995).
18. C. D. Schwieters, G. M. Clore, A physical picture of atomic motions within the Dickerson DNA dodecamer in solution derived from joint ensemble refinement against NMR and large-angle X-ray scattering data. *Biochemistry* **46**, 1152-1166 (2007).
19. Y. Chen, S. L. Campbell, N. V. Dokholyan, Deciphering protein dynamics from NMR data using explicit structure sampling and selection. *Biophys. J.* **93**, 2300-2306 (2007).

20. F. Aboul-ela, J. Karn, G. Varani, Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acids Res.* **24**, 3974-3981 (1996).
21. J. D. Puglisi, R. Tan, B. J. Calnan, A. D. Frankel, J. R. Williamson, Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science* **257**, 76-80 (1992).
22. D. K. Merriman *et al.*, Increasing the length of poly-pyrimidine bulges broadens RNA conformational ensembles with minimal impact on stacking energetics. *RNA* **24**, 1363-1376 (2018).
23. C. Musselman *et al.*, Impact of static and dynamic A-form heterogeneity on the determination of RNA global structural dynamics using NMR residual dipolar couplings. *J. Biomol. NMR* **36**, 235-249 (2006).
24. E. J. Denning, U. D. Priyakumar, L. Nilsson, A. D. Mackerell, Jr., Impact of 2'-hydroxyl sampling on the conformational properties of RNA: update of the CHARMM all-atom additive force field for RNA. *J. Comput. Chem.* **32**, 1929-1943 (2011).
25. L. Salmon, S. Yang, H. M. Al-Hashimi, Advances in the determination of nucleic acid conformational ensembles. *Annu. Rev. Phys. Chem.* **65**, 293-316 (2014).
26. M. H. Bailor, A. M. Mustoe, C. L. Brooks, 3rd, H. M. Al-Hashimi, 3D maps of RNA interhelical junctions. *Nat. Protoc.* **6**, 1536-1545 (2011).
27. X. Shi, P. Walker, P. B. Harbury, D. Herschlag, Determination of the conformational ensemble of the TAR RNA by X-ray scattering interferometry. *Nucleic Acids Res.* **45**, e64 (2017).
28. J. Swails, T. Zhu, X. He, D. A. Case, AFNMR: automated fragmentation quantum mechanical calculation of NMR chemical shifts for biomolecules. *J. Biomol. NMR* **63**, 125-139 (2015).
29. H. Shi *et al.*, Atomic structures of excited state A-T Hoogsteen base pairs in duplex DNA by combining NMR relaxation dispersion, mutagenesis, and chemical shift calculations. *J. Biomol. NMR* **70**, 229-244 (2018).
30. D. A. Case, Chemical shifts in biomolecules. *Curr. Opin. Struct. Biol.* **23**, 172-176 (2013).
31. M. C. Clay, L. R. Ganser, D. K. Merriman, H. M. Al-Hashimi, Resolving sugar puckers in RNA excited states exposes slow modes of repuckering dynamics. *Nucleic Acids Res.* **45**, e134 (2017).
32. D. B. Davies, Conformations of nucleosides and nucleotides. *Prog. Nucl. Magn. Reson. Spectrosc.* **12**, 135-225 (1978).
33. J. A. Ippolito, T. A. Steitz, A 1.3-angstrom resolution crystal structure of the HIV-1 trans-activation response region RNA stem reveals a metal ion- dependent bulge conformation. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9819-9824 (1998).
34. A. C. Stelzer, J. D. Kratz, Q. Zhang, H. M. Al-Hashimi, RNA Dynamics by Design: Biasing Ensembles Towards the Ligand-Bound State. *Angew. Chem. Int. Ed. Engl.* **49**, 5731-5733 (2010).
35. G. Kawai *et al.*, Conformational rigidity of specific pyrimidine residues in tRNA arises from posttranscriptional modifications that enhance steric interaction between the base and the 2'-hydroxyl group. *Biochemistry* **31**, 1040-1046 (1992).
36. T. E. Cheatham, 3rd, P. Cieplak, P. A. Kollman, A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.* **16**, 845-862 (1999).

37. M. Zgarbova *et al.*, Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **7**, 2886-2902 (2011).
38. D. Tan, S. Piana, R. M. Dirks, D. E. Shaw, RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E1346-e1355 (2018).

Acknowledgments: We thank members of the Al-Hashimi laboratory and Dr. Dawn Merriman (University of Dayton) for assistance and critical comments on the manuscript. We would like to thank Dr. Andrew Watkins (Stanford University) for advice about FARFAR. **Funding:** This work was supported by US National Institute for General Medical Sciences (1R01GM132899 to H.M.A. and D.H.), and US National Institute of Health (U54 AI150470 to H.M.A. and D.A.C.) and by Tobacco Settlement Fund (21-5734-0010 to J.D.Y.). **Author contributions:** H.S., D.H., J.D.Y. and H.M.A. conceived the project and experimental design. H.S. and J.D.Y. performed FARFAR calculations. A.R. performed MD simulations. H.S. performed sample and select and other ensemble analysis, with assistance from R.R. H.A.A. prepared NMR samples, performed NMR experiments and analyzed NMR data. D.A.C. performed QM/MM chemical shift calculation. H.M.A. and H.S. wrote the manuscript with critical input from A.R., D.H., D.A.C. and J.D.Y. **Competing interests:** H.M.A. is advisor to and holds an ownership interest in Nymirum, an RNA-based drug discovery company. The remaining authors declare no competing interests. **Data and materials availability:** Rosetta FARFAR commands as well as its required input files are included in Supplementary Materials. The raw RDC data can be downloaded from https://sites.duke.edu/alhashimilab/resources/. The AFNMR programs are available at https://github.com/dacase/afnmr. All other in-house Python scripts for SAS, structural analysis as well as other data analysis are available on request to the corresponding authors. The authors declare that the data supporting the findings are available within the article and its Supplementary Materials, or are available upon request to the corresponding authors.

Supplementary Materials

Materials and Methods

Supplementary Text

Figures S1-S16

Tables S1-S5

Movie S1

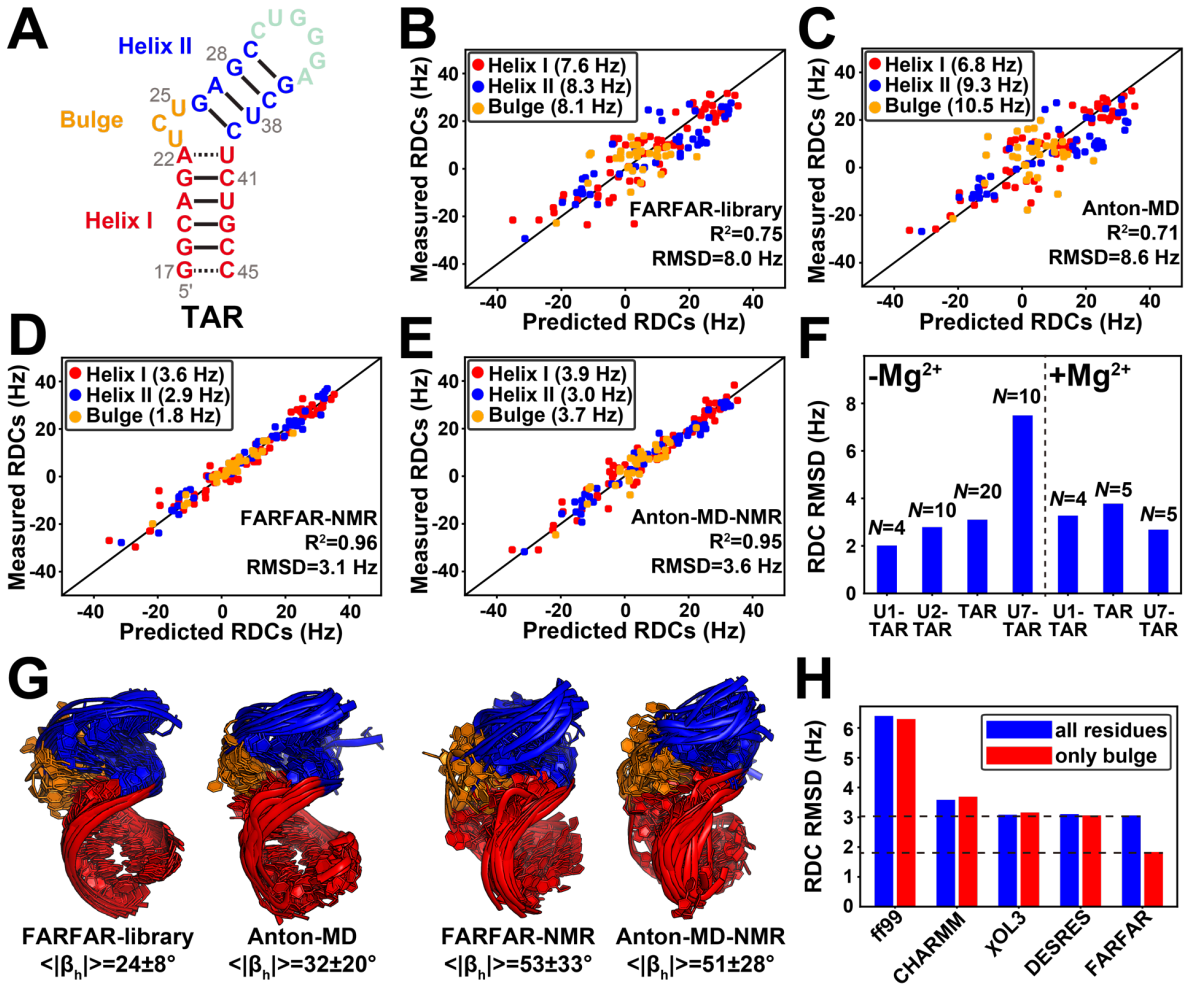

Fig. 1. Using FARFAR-NMR to determine a TAR ensemble. (A) Secondary structure of TAR. (B-E) Comparison between measured and predicted TAR RDCs for the (B) FARFAR-library (N = 10,000), (C) Anton-MD library (N = 10,000), (D) FARFAR-NMR (N = 20) and (E) Anton-MD-NMR (N = 20). Values in parentheses denote the RDC RMSD. RDCs are color-coded according to the structural elements in Fig. 1A. (F) RDC RMSD for ensembles of TAR variants with ensemble size N obtained using FARFAR-NMR. (G) Structural overlay of the TAR ensembles (N = 20) (Methods). $|\beta_h|$ is the absolute magnitude of the bend angle. (H) Comparison of RDC RMSD over all (blue) or bulge (red) residues for ensembles generated using different MD force fields (ff99 (*36*), CHARMM36 (*24*), ff99bsc0χOL3 (*37*) and modified ff14 DESRES (*38*)) and FARFAR.

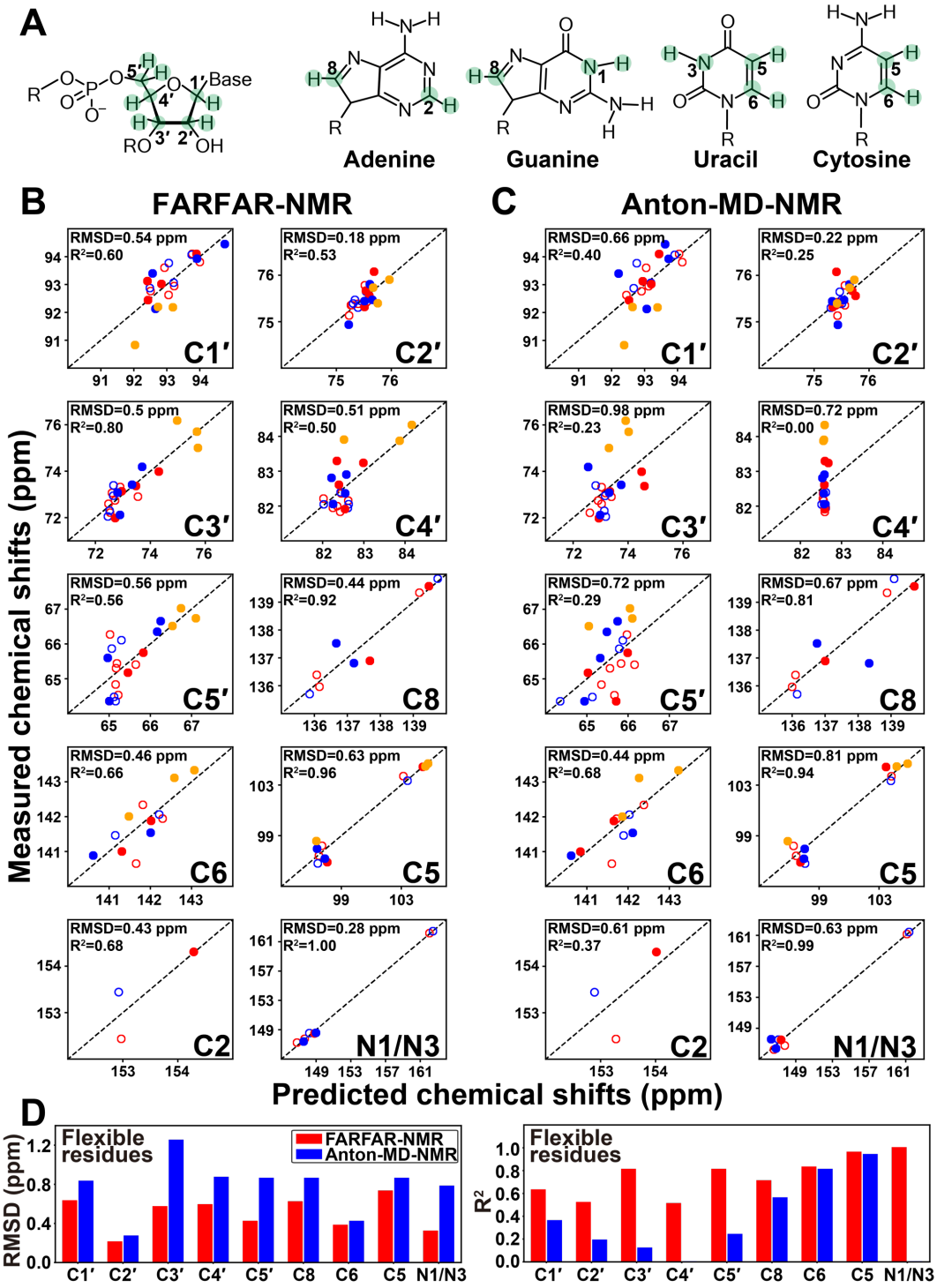

Fig. 2. Evaluating TAR ensembles using chemical shifts. (A) Chemical structures of the sugar and base moieties with chemical shift probes used to test ensemble accuracy highlighted in green. (B-C) Comparison of measured and predicted $^{13}C/^{15}N$ chemical shifts for (B) FARFAR-NMR (*N*

= 20) (C) Anton-MD-NMR (N = 20) by AF-QM/MM (Methods). Values are color-coded according to the structural elements in Fig. 1A. Chemical shifts for central Watson-Crick bps within A-form helices (C19-G43, A20-U42, G21-C41, A27-U38, G28-C37) are denoted using open circles. A correction was applied to the predicted chemical shifts (Methods) as described previously (*29*). For ^{1}H chemical shifts see fig. S7. (D) Comparison of RMSD (left) and R^2 (right) between measured and predicted ^{13}C/^{15}N chemical shifts for flexible residues (U23, C24, U25, A22-U40, G26-C39, C29-G36, G18-C44) for FARFAR-NMR (red) and Anton-MD-NMR (blue).

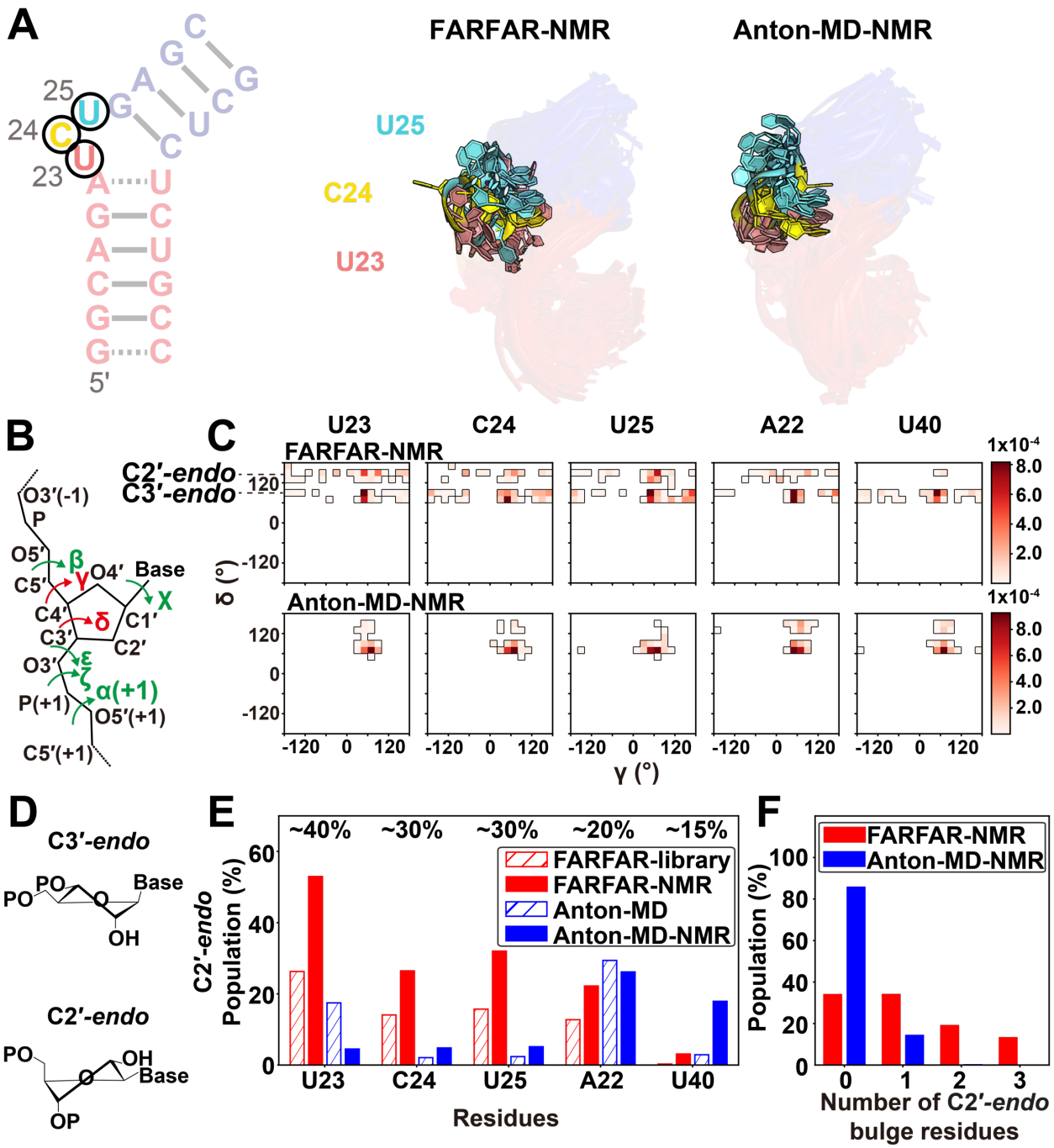

Fig. 3. FARFAR-NMR ensemble more broadly samples non-canonical sugar-backbone torsion angles relative to Anton-MD-NMR. (A) Overlay of the dynamic ensembles of the TAR bulge. (B) RNA backbone torsion angles exhibiting different and similar distributions between Anton-MD-NMR and FARFAR-NMR are colored red and green, respectively (Methods). (C) 2D density maps of δ versus γ comparing Anton-MD-NMR and FARFAR-NMR ensembles (N = 2,000) for

bulge residues as well as A22 and U40. The bin width is 20°. (D) Structure of the ribose moiety in C3′-*endo* and C2′-*endo* conformations. (E) Population of C2′-*endo* pucker at bulge residues as well as A22 and U40 in the FARFAR-library (N = 10,000, red open), FARFAR-NMR (N = 2,000, red fill), Anton-MD library (N = 10,000, blue open) and Anton-MD-NMR (N = 2,000, blue fill). Experimental estimates of the C2′-*endo* population based on ^{13}C chemical shifts are indicated above the bars (Methods). (F) The population of conformers in the ensemble as a function of the number of C2′-*endo* bulge residues for FARFAR-NMR (red, N = 2,000) and Anton-MD-NMR (blue, N = 2,000).

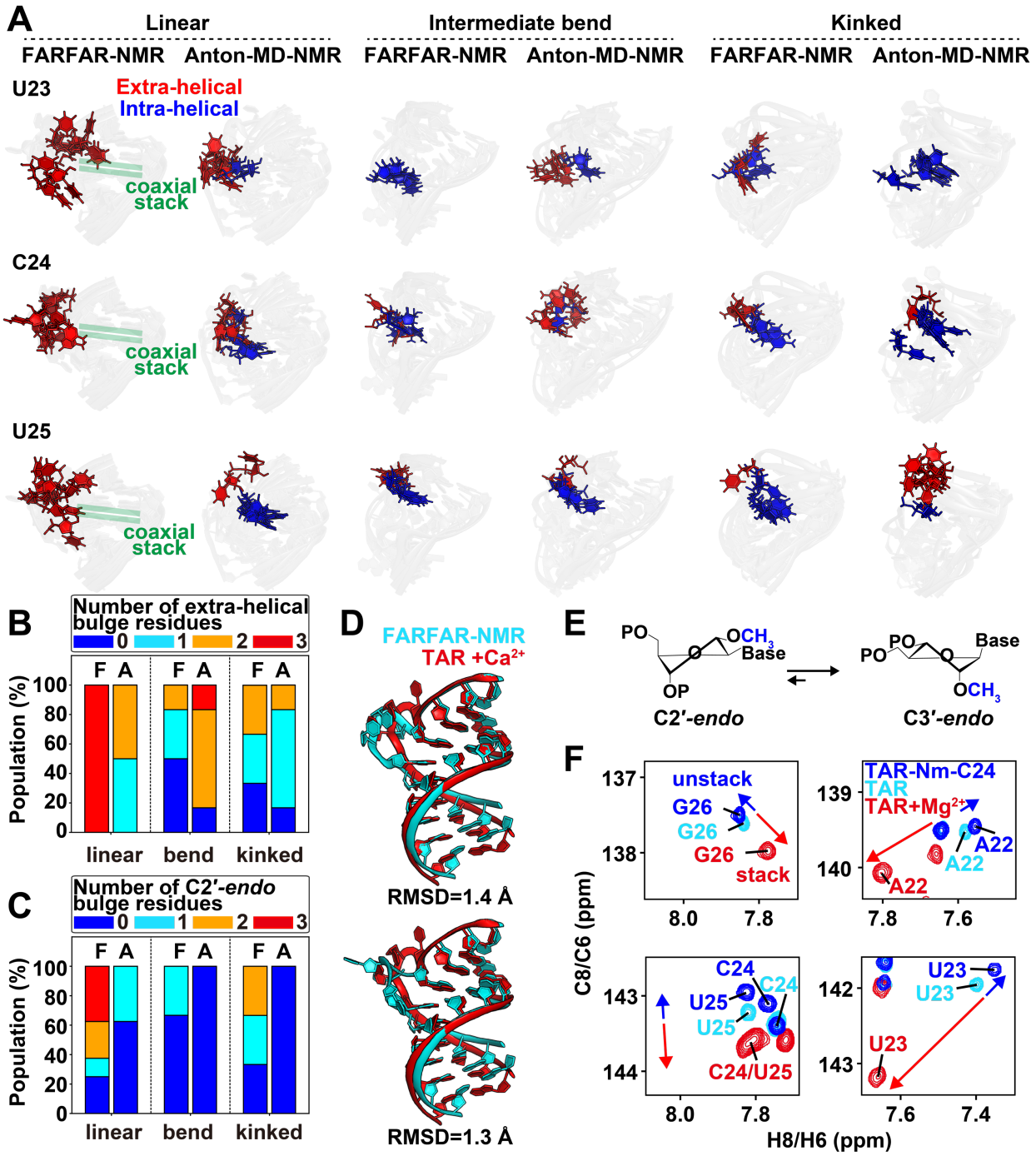

Fig. 4. Cooperative extra-helical flipping and sugar repuckering of bulge residues are coupled to coaxial stacking. (A) Overlay of conformers showing motions in bulge residues for linear ($|\beta_h| < 45°$), intermediate bend ($45° < |\beta_h| < 70°$) and kinked ($|\beta_h| > 70°$) inter-helical conformations in the FARFAR-NMR and Anton-MD-NMR ensembles. (B-C) The fractional populations of

conformers with (B) extra-helical and (C) C2′-*endo* bulge residues (color-coded) as a function of bending angle in the FARFAR-NMR and Anton-MD-NMR ensembles. F and A denote FARFAR-NMR and Anton-MD-NMR, respectively. (D) Comparison of two coaxially stacked FARFAR-NMR conformers with the crystal structure of Ca^{2+} bound TAR (PDB: 397D) (*33*). (E) Nm shifts sugar pucker equilibrium towards C3′-*endo*. (F) Overlay of 2D [^{13}C, ^{1}H] NMR HSQC spectra of the aromatic spins for TAR-Nm-C24 without Mg^{2+} (blue, inducing unstacking), TAR without Mg^{2+} (cyan) and TAR with Mg^{2+} (red, inducing stacking).