Kinetic Resolution of the Atomic 3D Structures Formed by Ground and Excited Conformational States in an RNA Dynamic Ensemble

Rohit Roy,

Center for Genomic and Computational Biology, Duke University School of Medicine, Durham, North Carolina 27710, United States

Ainan Geng,

Department of Biochemistry, Duke University School of Medicine, Durham, North Carolina 27710, United States

Honglue Shi,

Department of Chemistry, Duke University, Durham, North Carolina 27708, United States

Dawn K. Merriman,

Department of Chemistry, Duke University, Durham, North Carolina 27708, United States

Elizabeth A. Dethoff,

Department of Chemistry and Biophysics, University of Michigan, Ann Arbor, Michigan 48109, United States

Loïc Salmon,

Department of Chemistry and Biophysics, University of Michigan, Ann Arbor, Michigan 48109, United States

Hashim M. Al-Hashimi

Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, United States

Corresponding Author: Hashim M. Al-Hashimi – Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, United States; ha2639@cumc.columbia.edu.

Present Address: Innovative Genomics Institute, University of California, Berkeley, CA 94720, USA

Present Address: Alcon, 6201 South Freeway, TC-38, Fort Worth, TX 76134-2099, USA.

Present Address: Ribometrix Inc., 8 Davis Dr Suite 210, Durham, NC 27709, USA.

Present Address: Centre de Resonancé Magnetiqué Nucleairé à Très Hauts Champs, (CRMN), UMR 5082, CNRS, ENS Lyon, UCBL, Université de Lyon, 69100 Villeurbanne, France.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.3c04614.

TAR variants, RDC measurements and optimization, cross-validation of TAR ensembles, trends, chemical shift predictions, apical loop minimally impacting the bulge ensemble, direct evidence for protonation, GS-ES1 dynamics, exchange parameters, list of RF powers, NMR measurements, ensemble optimization and validation, structural comparisons, order tensor analysis, and secondary structural constraints (PDF)

Video rendering of motions observed in kinetically resolved dynamic ensemble of HIV-1 TAR (AVI)

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.3c04614

Abstract

Knowing the 3D structures formed by the various conformations populating the RNA free-energy landscape, their relative abundance, and kinetic interconversion rates is required to obtain a quantitative and predictive understanding of how RNAs fold and function at the atomic level. While methods integrating ensemble-averaged experimental data with computational modeling are helping define the most abundant conformations in RNA ensembles, elucidating their kinetic rates of interconversion and determining the 3D structures of sparsely populated short-lived RNA excited conformational states (ESs) remains challenging. Here, we developed an approach integrating Rosetta-FARFAR RNA structure prediction with NMR residual dipolar couplings and relaxation dispersion that simultaneously determines the 3D structures formed by the ground-state (GS) and ES subensembles, their relative abundance, and kinetic rates of interconversion. The approach is demonstrated on HIV-1 TAR, whose six-nucleotide apical loop was previously shown to form a sparsely populated (~13%) short-lived (lifetime ~ 45 μs) ES. In the GS, the apical loop forms a broad distribution of open conformations interconverting on the pico-to-nanosecond time scale. Most residues are unpaired and preorganized to bind the Tat-superelongation protein complex. The apical loop zips up in the ES, forming a narrow distribution of closed conformations, which sequester critical residues required for protein recognition. Our work introduces an approach for determining the 3D ensemble models formed by sparsely populated RNA conformational states, provides a rare atomic view of an RNA ES, and kinetically resolves the atomic 3D structures of RNA conformational substates, interchanging on time scales spanning 6 orders of magnitude, from picoseconds to microseconds.

Graphical Abstract
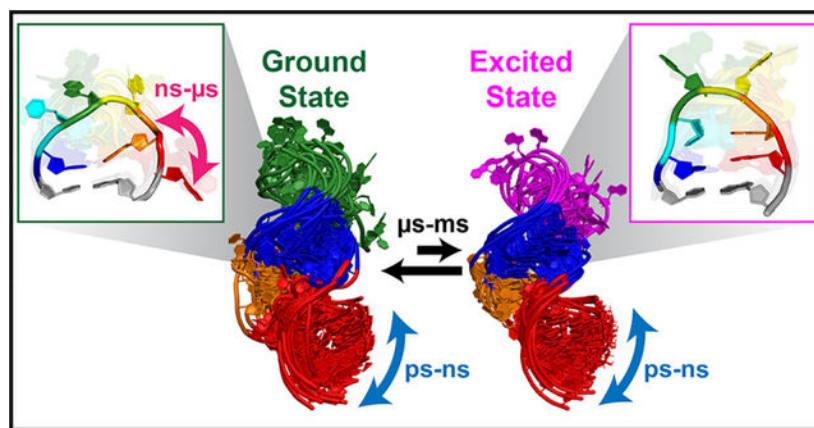
INTRODUCTION

Ribonucleic acids (RNAs) form dynamic ensembles of many conformations that interchange on time scales spanning 12 orders of magnitude in time from picoseconds to seconds.[1] The ensemble includes the energetically most favorable ground-state (GS) and lowly populated and short-lived conformational states commonly referred to as "excited conformational states" (ESs).[2–4] The ESs form on the microto-millisecond time scale typically by reshuffling base-pairs (bps) in and around noncanonical motifs.[2] These transient states

are of great importance, as they perform critical functions in multiple aspects of RNA biology. They serve as RNA-based switches controlling gene expression and regulation[5–7] and as intermediates during processes such as RNA folding,[2,8] processing,[9–11] and viral replication.[12,13] Additionally, these states frequently represent attractive RNA drug-targets.[14,15]

The GS and ESs are themselves subensembles, comprising many conformations sharing a similar overall secondary structure but differing with respect to sugar pucker, base stacking, and the relative orientation of helical domains.[16–19] These conformations typically interconvert on a time scale ranging from picoseconds to microseconds, which is more than an order of magnitude faster than the dynamics between GS and ESs[16,20,21] (Figure 1A). These fast motions lubricate RNA structures, enabling them to adaptively bind to protein and ligand targets.[22–24]

Knowing the atomic three-dimensional (3D) structures populating the GS and ES subensembles, in addition to their kinetic rates of interconversion, is required to obtain a quantitative understanding of how RNAs fold and function at the atomic level and for controlling RNA in RNA-targeted drug discovery and bioengineering.[1] Ensemble-averaged experimental data from NMR,[16,17] SAXS,[25] EPR,[26] and chemical probing[27] have been used in combination with computational modeling[1] to determine RNA conformational ensembles. However, these approaches face challenges when determining the 3D structures of sparsely populated ES, as their contribution to ensemble-averaged measurements can be small, falling below the limits of detection. Furthermore, the conformational landscape of high-energy ESs is more expansive and challenging to define compared to that of the GS. Structures of lowly populated conformations can in some cases be determined using room-temperature X-ray crystallography,[28] femtosecond X-ray free electron laser pulse crystallography,[29] and cryo-electron microscopy.[30–32] However, these approaches require further advancement to efficiently resolve the conformational states within the GS and ES subensembles. Additionally, none of these approaches provide information on the kinetic rates with which the different conformations interconvert over broad range of time scales.

NMR spectroscopy is one of the most powerful techniques for characterizing the conformational ensembles of biomolecules[3,33,34] with recent advances culminating in the determination of atomic structures for several protein ESs.[35–39] Most of these NMR approaches rely on exchange between the dominant NMR-visible GS and the NMR-invisible ES to transfer information related to a particular magnetic resonance property, where this property can be readily detected. In addition, the integration of ensemble averaged measurements such as residual dipolar couplings (RDCs)[40,41] with kinetics information from relaxation dispersion,[42,43] Nuclear overhauser effects (NOEs),[44] and spin relaxation[44–47] has made it possible to simultaneously determine the 3D structures of protein conformational-substates and their kinetic rates of interconversion. Here, we report an approach that integrates Rosetta FARFAR RNA structure prediction[48] with NMR RDCs and relaxation dispersion that simultaneously determines the 3D atomic structures formed by GS and ES subensembles, their relative abundance, and kinetic rate of interconversion.

Many RNA ESs have populations exceeding 10% and exchange with the GS on the submillisecond time scales.[3,49] In theory, such ESs could make a direct and measurable contribution to ensemble-averaged RDCs. However, without additional constraints, resolving the 3D structures of these RNA ESs based on their minor contribution to the measured RDCs would be difficult, if not impossible. Here, we addressed this data gap by using NMR relaxation dispersion experiments to independently determine the secondary structure of the RNA ES based on ^{13}C and ^{15}N chemical shifts.[2,49] Rosetta FARFAR RNA structure prediction[48] is then used to generate a conformational library broadly sampling the conformations of both the GS and ES given their NMR-derived secondary structures. The augmented GS + ES library is then optimized using RDCs, and cross-validated using chemical shifts (ω_{avg}, $\Delta\omega = \omega_{ES} - \omega_{GS}$)[50] and by comparing the ES population (p_{ES}) with values determined independently using relaxation dispersion (Figure 1B). By integrating relaxation dispersion experiments, our work extends the FARFAR-NMR approach[17] recently developed to determine conformational ensembles of the RNA GS to enable 3D structure determination of both the GS and ES subensembles as well as their kinetics rates of interconversion.

We demonstrate this extended FARFAR-NMR approach by determining the conformational ensemble model for HIV-1 TAR RNA.[51,52] TAR activates transcription elongation of the retroviral HIV-1 genome by binding to the viral transactivating protein Tat and superelongation complex (SEC).[53,54] TAR has two crucial motifs: the trinucleotide UCU bulge, which interacts with Tat,[55] and the long hexanucleotide apical loop, which establishes extensive contacts with both Tat and the cyclin-T1 component of SEC.[55,56] Several ensembles[17,57] have been reported for the UCU bulge using a TAR variant with a UUCG loop in place of the wild-type (wt) six-nucleotide apical loop (CUGGGA). The ensembles revealed that the base-triple conformation required for productive Tat binding[54,55] is exceptionally lowly populated[58] in the free RNA, and comes with a substantial energetic penalty (>7 kcal/mol) for binding.[59] In contrast, no conformational ensemble has been reported for the wtTAR apical loop or for other comparably long apical loops, which represent a functionally important class of RNA motifs.[60,61] The degree to which apical loop motifs are preorganized for molecular recognition or must undergo large conformational changes on binding to partner molecules remains largely unexplored. The conformational ensemble of TAR with the wt apical loop is also needed to apply ensemble-based virtual screening in the development of TAR-targeting anti-HIV therapeutics.[14,62]

Prior studies revealed that the TAR apical loop undergoes complex dynamics over a broad range of time scales. ^{13}C NMR relaxation dispersion experiments showed that in wildtype (wt) TAR, under standard experimental conditions (pH 6.4 and temperature 25 °C, see Methods), the apical loop exists in dynamic equilibrium between the GS and a lowly populated (~15%) short-lived (lifetime of ~45 μs) ES termed "ES1"[2,63] (Figure 2A). Based on the ES1 ^{13}C chemical shifts measured using relaxation dispersion,[2] a secondary structure was proposed for ES1 in which the cross-stranded C30–G34 Watson–Crick bp in the GS is replaced by two consecutive C30–A[+]35 and U31–G34$_{syn}$ mismatches, effectively zipping up the apical loop. In addition to these microsecond time scale motions, ^{13}C spin relaxation data[63] indicated that apical loop residues U31, G32, and A35, which are unpaired in the

GS, undergo large amplitude motions ($S^2 < 0.6$; S is the Lipari–Szabo order parameter[64] which is equal to 1 and 0 for minimum and maximum amplitude motion, respectively) on the picosecond-to-nanosecond time scale.

The FARFAR-NMR ensemble of wt TAR reported here resolves the 3D structures formed by the GS and ES1, which interconvert on the microsecond time scale, as well as the conformational substates within each of the GS and ES1 subensembles, which interconvert on the pico-to-nanosecond time scales. In the GS subensemble, the apical loop forms a broad distribution of open conformations in which most residues are unpaired. Conformers in the GS subensemble strongly resemble the structure of TAR in complex with Tat-SEC. Thus, the apical loop is preorganized for protein recognition. The apical loop zips up in the ES subensemble, forming a narrow distribution of closed conformations, which alter the shape and sequester critical residues required for Tat-SEC recognition. Our study resolves the atomic structures of conformational substates in a complex apical loop, which interchange on time scales spanning 6 orders of magnitude in time, provides a rare atomic view of the 3D structure formed by an RNA ES, and extends the utility of FARFAR-NMR to enable resolution of the 3D structures formed by low-populated and short-lived RNA substates.

RESULTS

Measuring Multiple RDC Data Sets in HIV-1 TAR Using Domain-Elongation.

We employed NMR RDCs to determine the conformational ensemble of the HIV-1 TAR. RDCs measured between directly bonded nuclei depend on the orientation of the internuclear vector relative to a principal axis system (PAS) of an alignment tensor describing the average orientation of a molecule relative to the applied magnetic field.[65–67] The RDCs are ensemble-averaged over all conformations interchanging on the submillisecond time scale[68] and are exquisitely sensitive to details of conformational distributions.

Using the domain-elongation strategy,[16] we altered the overall shape of TAR and its alignment in Pf-1 phage and measured multiple RDC data sets reporting on the conformational distribution of bond vectors relative to different molecule-fixed PAS frames (Figure 2B). Having multiple data sets was especially important given the minor ES1 contribution and the need to resolve degeneracies and assess overfitting of the RDCs during ensemble determination.[69,70] In addition to the canonical E-0 TAR construct used in prior studies,[1,63] three additional variants (Figure 2B) were used to modulate TAR alignment in the Pf-1 ordering medium.[71] In EI-22,[16] the lower helix was elongated by twenty-two bps. In EII-3 and EII-13, the bulge and lower stem were omitted, and the upper stem elongated by three (EII-3) or 13 (EII-13) bps, respectively. The 2D heteronuclear single quantum coherence (HSQC) NMR spectra (Figure S1) of the three variants were in excellent agreement with those of E-0, indicating that these changes to the TAR secondary structure do not significantly alter the conformational ensemble of the apical loop.

We measured one-bond $^{13}C – {}^1H$ ($^1D_{CH}$) and $^{15}N – {}^1H$ ($^1D_{NH}$) RDCs in Pf1 phage (15–22 mg/mL) for all four TAR variants using two independent frequency-based experiments in

which the splittings are encoded either along the $^{13}C/^{15}N$ or ^{1}H dimension. The root-mean-square deviation (RMSD) between the two sets of RDC measurements (~2.8 Hz) was used to estimate the RDC uncertainty (Figure S2A,B).

To assess the independence of the four data sets, we subjected RDCs measured in the upper helix to an order tensor analysis assuming an idealized A-form geometry[72] and determined the five elements of the order tensor describing overall alignment. As expected, the average orientation of the magnetic field (S_{zz}) deviated most strongly ($\beta\sim31°$) from the axis of the upper helix for the kinked-elongated EI-22 followed by the nonelongated E-0 ($\beta\sim10°$), whereas it was nearly parallel ($\beta\sim2°$) for the more linear EII-13 (Figure 2C). Interestingly, the shorter EII-3 variant, which amplifies the contribution of the apical loop to the overall shape, yielded a distinct S_{zz} orientation ($\beta\sim20°$). Thus. Thus, the four TAR variants yielded four semi-independent RDC data sets, as also verified by pairwise comparison of the RDCs (Figure S2C). These results establish the utility of domain-elongation to modulate the alignment of RNAs containing long apical loops and demonstrate that the alignment can be modulated by shortening helices in addition to their elongation.

Determining the TAR Conformational Ensemble by Integrating FARFAR-NMR and NMR Relaxation Dispersion.

We used the four semi-independent sets of RDCs to determine a conformational ensemble for wtTAR using FARFAR-NMR (Figure 2B). We used FARFAR and the previously reported NMR-derived secondary structural constraints (Table S1) to generate two conformational libraries, one for the TAR GS ($N = 5000$) and another for ES1 ($N = 5000$). For the open GS apical loop, only C30–G34 was constrained to a Watson–Crick bp. For the closed ES1, U31 and G34 were constrained to form a trans sugar-edge/Watson–Crick bp. This conformation could be directly inferred based on the G34–C8 and G34–C1$'$ chemical shifts measured for ES1 as well as NOEs measured at low pH conditions stabilizing ES1[2]. In addition, C30 and A35 were constrained to form a C30–A^{+}35 *cis*-wobble. While we did not have direct chemical shift evidence for this wobble bp geometry, its existence was indirectly inferred based on the overall pH dependence of TAR apical loop chemical shifts that indicated that formation of ES1 is coupled to protonation, as well as the appearance of a potential protonated A35^{+} at low pH based on A35–C2H2.[2] However, these results cannot rule out the possibility that under the conditions (pH = 6.4) used to measure RDCs, a substantial fraction of the A35 is not protonated in ES1 and possibly forms alternative conformations deviating from the wobble geometry. Assuming all of ES1 is protonated at pH = 6.4; the estimated p$K_{a}\sim5.6$ predicts that the ES1 population would decrease to ~1.6% when increasing the pH to 7.4. This, in turn, would make the ES1 invisible to detection using the previously used ES1 probes, G34–C8, G34–C1$'$, and A35–C1$'$. We tested this prediction by preparing a uniformly $^{13}C/^{15}N$ labeled TAR sample and performing pH-dependent RD measurements on the previously used ES1 probes G34–C8, G34–C1$'$, and A35–C1$'$. Indeed, the population was decreased to ~1.3% in excellent agreement with the predicted value of ~1.6%. In addition, we were also able to obtain direct evidence for A35 protonation based on the A35–C2 RD profile that was measurable under these pH

conditions yielding the large $\Delta\omega$ of ~-10 ppm expected for a protonated A35–N1 in ES1 (Figure S9A,B).

We then augmented the GS and ES1 libraries ($N = 10,000$) in an unbiased 1:1 ratio (FARFAR-library) and used the agreement with the four RDC data sets to guide selection of conformers for inclusion in an RDC-satisfying ensemble (FARFAR-NMR) with the option of selecting either GS and/or ES1 conformations.

We used sample and select (SAS)[17,73] to test increasingly large ensemble sizes (N) to find the smallest size ensemble model satisfying the RDC data to within experimental uncertainty (Figure 3A). None of the conformers ($N = 1$) in the FARFAR library satisfied the RDC data; the lowest RDC RMSD was ~10 Hz, which was significantly larger than the experimental uncertainty of ~2.8 Hz. This poor agreement is to be expected for a highly dynamic RNA molecule in which the RDCs are averaged over many conformations. Indeed, increasing the ensemble size gradually reduced the RDC RMSD reaching a plateau at $N = 20$, at which point the RMSD of 3.0 Hz was near experimental uncertainty.

The FARFAR-NMR ensemble (RMSD = 3.0 Hz) (Movie S1) showed substantially improved agreement with the RDCs compared to the entire FARFAR-library (RMSD = 9.6 Hz) (Figure 3B) and two reference ensembles obtained by selecting either 20 conformers randomly (FARFAR-random, RMSD = 9.7 Hz) or the 20 lowest-energy conformers (FARFAR-lowest-energy; RMSD = 11.6 Hz) from the parent library (Figure S3A,B). The improved RDC agreement was observed robustly across all motifs (Figure 3B), including the apical loop for which the RMSD decreased from 11.4 Hz in the FARFAR-random to 3.1 Hz in the optimized FARFAR-NMR ensemble. Cross validation[17,57] in which a subset of RDCs were left out of SAS and used to evaluate the quality of the ensemble indicated that the decrease in RMSD in the selected FARFAR-NMR ensemble was not purely due to overfitting to the RDC data (Figure S3C). Thus, a population reweighting of the FARFAR-library was necessary to satisfy the RDCs, resulting in the selection of many conformers that have comparatively poor energy scores in the FARFAR-library (Figure S3F).

Interestingly, the unbiased RDC optimization resulted in selection of conformations from both the GS and ES1 libraries, with a stronger preference for the GS. Averaging over 100 SAS repeats with $N = 20$, we found the fractional ES1 population in the FARFAR-NMR ensemble to be $p_{ES1} = 17 \pm 4\%$ (Figure 3C), in excellent agreement with $p_{ES1} = 13 \pm 2\%$ obtained independently using NMR relaxation dispersion.[2] This agreement is noteworthy, since no constraints on p_{ES1} were applied during ensemble determination.

To further verify that selection of ES1 conformers was not merely the result of overfitting the RDCs, we performed additional SAS optimizations replacing ES1 with another library generated using FARFAR for a different TAR conformation termed "ES2".[74] Relative to the GS, the ES2 reshuffles the bps in the upper stem and apical loop. ES2 also forms in TAR[74] but has too low of a population (0.4%) to make measurable contribution to RDC data. If the selection of ES1 conformers is not due to overfitting of the RDCs, we would expect fewer ES2 conformers to be selected during SAS. Indeed, none of the conformers in the ES2 library were selected (Figure 3D).

We also assessed whether GS conformations would be enough to satisfy the RDCs. To test this possibility, we performed SAS optimization using only the GS library after removing ES1-like conformations (see Methods). This resulted in an increase in the overall RDC RMSD from 3.0 to 3.9 Hz, and from 3.1 to 3.8 Hz for apical loop residues (Figure S3C). Even when conformers from the GS library with certain ES1 features were omitted, the SAS optimization consistently retrieved conformers having other ES1-like features (Figure S3D). Taken together, these results indicate that our RDC data sets sense ES1 and can accurately determine its population.

Testing Ensemble Accuracy Using Chemical Shifts and ES1 Mutants.

We used ensemble-averaged ^{1}H, ^{13}C, and ^{15}N chemical shifts measured throughout the base and sugar moieties to independently test the accuracy of FARFAR-NMR TAR ensemble.[17,50] The chemical shifts were not used in the ensemble determination and are exquisitely sensitive to base pairing, stacking, and sugar pucker distributions.[17,50] Importantly, not only did we have access to chemical shifts (ω_{avg}) averaged over the entire (GS + ES1) TAR ensemble ($\omega_{avg} = p_{GS} \times \omega_{GS} + p_{ES1} \times \omega_{ES1}$), but through relaxation dispersion experiments[2], we also have the chemical shifts averaged over the GS (ω_{GS}) and ES1 (ω_{ES1}) subensembles for a few residues with detectable chemical exchange. Therefore, we could use chemical shifts to evaluate both the overall ensemble as well as its kinetic partitioning into GS and ES1 subensembles. Note that while chemical shifts were used to derive secondary structural constraints used in the FARFAR structure prediction, their values strongly depend on details of the 3D structure, including, for example, the specific torsion angles that define the electron density at a nucleus as well as stacking interactions from aromatic rings and unsaturated groups defining ring currents[75] (Figure S5D).

Using the quantum mechanics/molecular mechanics (QM/MM)-based automated fragmentation quantum mechanical calculation of NMR chemical shifts (AFNMR)[50] approach, we computed the chemical shifts for every conformer in the FARFAR-NMR TAR ensemble. We then averaged the predicted chemical shifts over conformers in the GS + ES1 ensemble (ω_{avg}), GS (ω_{GS}), and ES1 (ω_{ES1}) subensembles and compared predicted values to their experimentally measured counterparts.

We observed very good agreement (average RMSD = 0.60 and 0.15 ppm for ^{13}C/^{15}N and ^{1}H nuclei, respectively) between the measured and predicted ω_{avg} values robustly across the different spins including for residues in the apical loop (Figure 4A,B). Furthermore, the agreement was consistently weaker across ^{15}N, ^{13}C, and ^{1}H nuclei for the FARFAR-random and FARFAR-lowest-energy ensembles (all $N = 20$) determined without RDCs (Figures 4B, S4 and S5) as well as for any single conformer within the FARFAR-NMR ensemble (Figure S6).

Equally good agreement was also obtained for ω_{GS} and ω_{ES1} as well as their differences ($\Delta\omega = \omega_{ES} - \omega_{GS}$) (Figure 4C,D), indicating that the FARFAR-NMR ensemble accurately partitions the GS and ES1 subensembles. As a negative control, we took the 20 conformers in the FARFAR-NMR ensemble and randomly labeled them "GS" and "ES1" making sure to keep the ES1 population ~20%. This ensemble, in which conformers were randomly

partitioned into GS and ES1 resulted in much weaker agreement with the measured $\Delta\omega$ values (Figure 4D).

We previously showed that ES1 could be stabilized to become the predominant state using two TAR mutants (C30U and A35G) in which the C30–A$^+$35 wobble in ES1 was replaced with either U30–A35 or C30–G35 Watson–Crick bps, respectively.[2] We prepared the two TAR-C30U and TAR-A35G mutants and measured C–H RDCs in natural abundance (Figure S10A and Table S6). We then used the measured RDCs to test how well our ES1 subensemble predicts the ES1 RDCs. Indeed, we found very good agreement between the mutant RDCs and values predicted from the ES1 ensemble (RMSD = 3.9 Hz, R^2 = 0.58 for C 30 U and RMSD = 3.2 Hz, R^2 = 0.75 for A35G) (Figure S10B). In addition, the agreement deteriorated considerably when comparing the mutant ES1 RDCs with those predicted for the GS subensemble (RMSD = 4.7 Hz, R^2 = 0.45 for C 30 U and RMSD = 3.9 Hz, R^2 = 0.69 for A35G) (Figure S10B). While this agreement is not as good as that observed for the overall RDC-optimized ensemble (Figure 3B), this is to be expected given that the RDCs were measured at natural abundance and that the mutants replace noncanonical mismatches with Watson–Crick bps and are not expected to perfectly recapitulate the structure of the ES1 in wt TAR. Thus, these data collectively further validate the ability to resolve GS and ES subensembles using this approach.

Comparison of the GS and ES Apical Loop Sub-ensembles.

In the GS subensemble, the apical loop forms a broad distribution of open conformations that differ with respect to stacking interactions between bases (Figures 3C, 5A). Except for C30 and G34, which robustly form a cross-stranded C30–G34 Watson–Crick bp as enforced by the NMR-derived constraints, all other residues (U31, G32, G33, and A35) were unpaired, primarily flipped out (~75%), and enriched (>55%) in the noncanonical C2$'$-endo sugar pucker relative to a canonical Watson–Crick bp.

Conversely, in the ES1 subensemble, the apical loop zips up (Figure 5A) forming a narrow distribution of conformations stabilized by the two adjacent mismatches that were enforced via NMR-derived constraints: C30–A$^+$35 wobble and U31–G34$_{syn}$ trans sugar-edge/Watson–Crick bp. C30, U31, and A35 are primarily intrahelical and enriched in the C3$'$-endo pucker (>90%) whereas G32 and G33 remain unpaired and sample a broad distribution of conformations. We previously showed[17] that when bulge residues are intrahelical, they preferentially adopt (>91%) the canonical C3$'$-endo sugar pucker whereas when extra-helical they tend to be enriched (>50%) in the noncanonical C2$'$-endo sugar pucker. We observe a similar behavior here for C30, U31, and A35 in the context of an apical loop (Figure 5A). Interestingly, whereas G34 was primarily intrahelical in both the GS and ES, it was enriched in the C2$'$-endo sugar pucker most likely because it engages in noncanonical base-pairing interactions. The GS and ES sugar pucker distributions in the FARFAR-NMR ensemble (Figure S11) were in excellent agreement with the sugar pucker distributions deduced independently based on analysis of C1$'$ and C4$'$ chemical shifts.[76]

Interestingly, approximately 10% of the conformers in the GS subensemble had ES1-like features (Figure 5B). These hybrid conformers had the cross-stranded C30–G34 bp

characteristic of the GS, but U31 and A35 were also intrahelical like in ES1, and some conformers had the A35–C30 wobble characteristic of ES1 (Figure 5C). These hybrid conformations imply a relatively shallow free-energy landscape without a sharp demarcation between the GS and ES1. This is not too surprising considering that the GS–ES1 exchange rate is fast $\left(k_{ex} = 25,400 \; s^{-1}\right)$. A shallow free energy landscape was also previously reported in the ensemble of the UUCG RNA tetraloop determined using enhanced MD simulations and exact NOEs, scalar couplings, RDCs, and solvent paramagnetic resonance enhancements.[18]

As expected, based on their near identical chemical shifts, the bulge ensemble was very similar in the GS and ES, and in excellent agreement with the FARFAR-NMR ensemble reported previously for a TAR variant containing the UUCG apical loop (Figure S7). Thus, the bulge ensemble is largely independent of the apical loop, consistent with prior NMR studies.[63]

GS Subensemble is Preorganized for Protein Recognition.

Interestingly, when comparing the GS subensemble of the TAR apical loop with previously determined structures of TAR, we found the best agreement with the X-ray structure of TAR[56] (PDBID 6CYT, Figure 5D) in complex with viral protein Tat and the host SEC. For the GS subensemble, the best conformer overlaid with the TAR-Tat:SEC complex with heavy all-atom RMSD of 3.3 Å (Figure 5D), where the apical loop residues were well-positioned to form the interactions observed in the complex. Thus, unlike the bulge motif, the TAR apical loop is largely preorganized and the conformational penalty[58] accompanying protein recognition is estimated to be <0.2 kcal/mol.

In contrast, conformers in the ES1 subensemble form a secondary structure unlike that of protein-bound TAR with the best conformer overlaying with RMSD of 4.6 Å (Figure 5D), higher than the best GS conformer. In addition, many residues critical for SEC binding, namely, the C30–G34 base pair, U31 and A35, are sequestered into noncanonical mismatches (Figure 5D), potentially disrupting specific interactions required to form the complex. This can help explain why mutations stabilizing ES1 relative to the GS inhibit cellular transactivation.[14]

Resolving Supra-τ_c Motions in the Apical Loop.

Approaches based on NMR spin relaxation data[20,77,78] can be used to characterize fast pico-to-nanosecond time scale motion, however, sensitivity to nanosecond and slower time scale motion is truncated around the overall correlation time for molecular rotational diffusion (τ_c), which is ~7 ns for E-0 TAR.[78] On the other hand, the sensitivity of relaxation dispersion to microto-millisecond time scale motion is typically truncated at around tens of microseconds. This leaves a window spanning 3 orders of magnitude in time from ~10 ns to ~10 μs (Figure 6A), in which motions are difficult to detect with either method. As extensively demonstrated in application to proteins,[45–47,79] RDCs can bridge the gap and help resolve these supra-τ_c motions.

We, therefore, compared the order parameters deduced from the overall FARFAR-NMR ensemble and for the individual GS and ES subensembles with those measured previously[63] by spin relaxation. The order parameters deduced by both methods $\left(S_{\text{rel}}^2 > 0.8\right)$ indicated low-amplitude motion for C30 and G34, which form the hydrogen-bonded C30–G34 bp in the GS (Figure 6A). The lower overall ensemble-derived order parameter for G34 is most likely due to transitions toward ES1 (Figure 6A), which are detected by relaxation dispersion. Thus, for C30 and G34, we have no evidence for supra-τ_c type motions.

However, for G32, G33, and A35, which are unpaired in the GS, the ensemble-derived order parameters were robustly smaller than counterparts measured using spin relaxation, even for the GS subensemble. These results provide evidence that some of the conformers in the GS subensemble interconvert on the supra-τ_c time scale, slower than ~7 ns (Figure 6B) but faster than 10 μs. These motions could represent changes in stacking interactions, including intra- versus extra-helical base flipping, which do not require the breaking of h-bonds, and which are likely too fast (<10 μs) to be detectable by conventional relaxation dispersion experiments. These supra-τ_c motions can be verified in the future with the use of alternative methods such as domain-elongation spin relaxation,[78] high-power RD,[3] and nanoparticle-assisted spin relaxation.[80]

DISCUSSION

Our approach for determining the 3D structures formed by the GS and ES subensembles utilizes chemical shift data from chemical exchange experiments to constrain the secondary structure of the ES and employs FARFAR structure prediction to generate a finite library of plausible 3D conformations for both the GS and ES, given their secondary structures. Without these constraints on the conformational space, it would be difficult if not impossible to disambiguate multiple ensemble solutions given the minor ES1 contribution to the measured RDCs. While we were able to determine a FARFAR-NMR ensemble model for HIV-1 TAR, which satisfies the NMR relaxation dispersion, RDC, and chemical shift data, additional data will ultimately be required to further refine and test the ensemble. In this regard, the versatile FARFAR-NMR pipeline can readily incorporate other types of data used previously to determine RNA ensembles.[18,25,26] For example, enhanced accuracy and computational efficiency in QM/MM chemical shift predictions could enable future optimization of FARFAR-NMR ensembles using easily measurable chemical shifts, and this approach could be assessed by cross-validating the ensembles against RDCs.

Significant effort has been directed recently toward predicting 3D structures of proteins[81] and RNA molecules[48,82] based on their sequences. However, the TAR conformational ensemble presented here, including the GS and ES subensembles, highlights the inadequacy of using a single static structure to represent a dynamic molecule with multiple conformational states. In fact, the lowest energy FARFAR structures were unable to accurately predict the ensemble-averaged RDCs and chemical shifts (Figures S3B and S4C). Consequently, there remains a pressing need to integrate experimental data with computational modeling to improve ensemble determination methodologies, which may eventually pave the way for predicting ensembles from sequences alone.

Several NOE-based NMR structures[54,83–85] have been reported for TAR in free and ligand or peptide bound states. Because these structures were determined by finding the single conformation that best satisfies the NMR data, it is unsurprising that none of them included the ES1 conformation with the C30–A$^+$35 and U31–G34$_{syn}$ mismatches. These NOE-based structures more closely resemble the GS subensemble reported here, but they were generally more disordered often lacking the C30–G34 bp (Figure S8). Some of these discrepancies could arise from differences in conditions, but it could also be that the assumption of a single static structure combined with the semiquantitative nature of NOE-based distance constraints diminishes the accuracy of the resulting structures. Indeed, an RNA ensemble for the UUCG apical loop determined using quantitative eNOEs, which explicitly accounted for ensemble averaging,[19] was shown to be in good agreement with the ensembles obtained using RDCs.[86] These and other studies[17,87] highlight the importance of properly accounting for ensemble averaging during structure refinement.

The FARFAR-NMR ensemble reported here provides insights into the 3D structure of an RNA ES and how it differs from the GS. Counterintuitively and as anticipated by the secondary structure, the high energy ES1 has one extra bp relative to the GS, enjoys greater stacking interactions, and forms a narrower more ordered ensemble. Why then is ES1 less energetically favorable than the GS? Prior studies[2] showed that simply lowering the pH is sufficient to render ES1 the dominant GS. Future studies should further dissect the energetic contribution associated with protonating A35 to form the C30–A$^+$35 wobble bp. Based on their secondary structures, we can anticipate ES ensembles that sample narrower conformational spaces relative to the GS for several other RNAs, including the TAR ES2, RRE stem-IIb,[88] pre-Q class-I,[89] and SAM-II[6] riboswitches.

The TAR ensemble also provided insights into the GS–ES1 transition pathway. A recent phi-value analysis[90] of the GS–ES1 transition showed that pairing of A35 with C30 was not rate limiting. These findings are consistent with our observation of hybrid apical loop conformations in which the ES1 C30–A$^+$35 wobble pair is formed without disrupting the cross stranded C30–G34. Rather, the melting of the cross stranded C30–G34 bp, which formed in all GS conformers, is most likely the rate limiting step during the GS–ES1 transition (Figure 7A). The C30–G34 bp could melt after the hybrid structure formed with the C30–A$^+$35 mismatch, and this could be followed by rapid *anti* to *syn* isomerization of G34 to form the U31–G34$_{syn}$ mismatch (Figure 7A). This proposed kinetic pathway can be tested in the future using phi-value analysis and modifications specifically targeting C30 and G34.

The FARFAR-NMR ensemble allowed us to resolve supra-τ_c motions in the TAR apical loop that are invisible to conventional spin relaxation or chemical exchange measurements, which likely involve base stacking/unstacking dynamics. Based on an RDC-derived conformational ensemble, we previously reported similar lines of evidence for supra-τ_c interhelical motions in TAR linked to base stacking/unstacking dynamics at the bulge.[20,57] FARFAR-NMR, in conjunction with spin relaxation and relaxation dispersion measurements, should help illuminate nanosecond-to-microsecond time scale motions in other RNA molecules.

The TAR ensemble also uncovered connections between motional modes occurring on different time scales. In particular, we find that microsecond time scale transitions redistribute the populations associated with faster motional modes occurring on the pico-to-nanosecond time scale (Figure 7B). For example, in the GS, U31 exists in a rapid equilibrium between C3′-endo (62.5%) ⇌ C2′-endo (37.5%). This equilibrium is slowly redistributed on the microsecond time scale to C3′-endo (>99%) ⇌ C2′-endo (<1%) in ES1. Likewise, in the GS, A35 exists in a rapid equilibrium between extra-helical (~78%) ⇌ intrahelical (32%) conformations. This equilibrium is slowly redistributed to extra-helical (<1%) ⇌ intrahelical (>99%) in ES1.

In summary, by integrating NMR chemical exchange data, we enhanced the utility of FARFAR-NMR, enabling the determination of conformational ensembles for both the RNA GS and ES. This is achievable if the ES population exceeds 10% and interconverts with the GS on submillisecond time scale. Thus, FARFAR-NMR holds great promise in going beyond 3D structure determination to provide deeper and broader descriptions of the RNA free energy landscape.

METHODS

Sample Preparation.

In Vitro Transcription.—The EI-22 and EII-13 variants were prepared by in vitro transcription using T7 RNA polymerase (New England Biolabs Inc.), uniformly $^{13}C/^{15}N$ labeled nucleotide triphosphates (Cambridge Isotopes, Inc.), and synthetic DNA templates (Integrated DNA Technologies, Inc.) containing the T7 promoter and sequence of interest. The transcription product was purified by 20% (w/v) denaturing polyacrylamide gel electrophoresis (PAGE), using 8 M urea and 1× TBE [89 mM Tris-borate, 89 mM boric acid, 2 mM ethylenediaminetetraacetic acid (EDTA)]. The RNA was electro-eluted from the gel in 20 mM Tris pH 8 buffer followed by ethanol precipitation. The RNA pellet was dissolved in water, annealed by heating to 95 °C for 5 min and rapid cooling on ice, and then exchanged into NMR buffer (15 mM sodium phosphate, 0.1 mM EDTA and 25 mM NaCl at pH 6.4) multiple times using an Amicon Ultra-4 Centrifugal Filter Unit (Millipore Corp. MWCO 3 kDa). Two samples were prepared for each of the EI-22 and EII-13 variants to minimize spectral overlap. For EI-22, samples were either $^{13}C/^{15}N$ A/U or $^{13}C/^{15}N$ G/C labeled while for EII-13 the samples were either $^{13}C/^{15}N$ uniformly or $^{13}C/^{15}N$ G/C labeled. Unlabeled EII-3 was purchased from Dharmacon (Thermo Fischer Scientific), dissolved in water, refolded, and exchanged into a NMR buffer. All samples were buffer exchanged to a final RNA concentration of ~1–2 mM. Aligned samples for RDC measurements were prepared by concentrating the RNA sample by a factor of 2 then adding Pf1 phage (Asla Biotech) solution (50 mg/mL) to the desired final concentration of 19, 25, and 17 mg/mL for EI-22, EII-3, and EII-13, respectively.

Solid-Phase Synthesis.—Unlabeled TAR mutants A35G and C30U were synthesized using MerMade 12 Oligo Synthesizer (BioAutomation) for solid-phase synthesis using standard protocols for phosphoramidite chemistry and 2′-hydroxyl deprotection in prior studies.[90] Unlabeled RNA phosphoramidites were purchased from ChemGenes and

both samples used 2′-TBDMS protected phosphoramidites with 1 μmol standard (1000 Å) synthesis columns. The 5′-DMT (4,4′-dimethoxytrityl) was removed for DMT-off deprotection, and standard PAGE purification was performed as previously described. Nucleobase protecting groups were removed, and oligonucleotides were cleaved from the synthesis columns using 1 mL of 30% ammonium hydroxide and 30% methylamine (1:1) followed by incubation at room temperature for 2 h. Deprotected samples were then air-dried followed by addition of 100 μL of DMSO and incubation at 65 °C for 5 min to ensure the samples were fully dissolved. 125 μL of TEA-3HF was added to the dissolved samples and the mixture incubated at 65 °C for 2.5 h. Finally, samples were precipitated overnight using 3 M sodium acetate and 100% ethanol, air-dried, then dissolved in water for gel purification, elution, ethanol washing, and buffer exchange as described for in vitro transcription above. The final concentration of the two samples were ~2 mM, and each sample was concentrated 2-fold before addition of Pf1 phage to achieve a final phage concentration of ~20–25 mg/mL.

NMR Experiments.

Unless indicated otherwise, all NMR experiments were performed at 298 K on a Bruker 600 MHz spectrometer equipped with a 5 mm triple-resonance HCN cryogenic probe. All experiments were processed using NMRPipe[91] and visualized in SPARKY.[92]

Resonance Assignments.—^{1}H, ^{13}C, and ^{15}N Resonance assignments for E-0 and EI-22 TAR as well as the mutants C30U and A35G were obtained from prior studies.[2,63] Resonances in EII-3 and EII-13 were assigned using standard homonuclear and heteronuclear 2D experiments and by comparison with wtTAR resonance assignments. The ^{1}H, ^{13}C, and ^{15}N chemical shifts for E-0[76] were used to cross-validate the ensemble. Chemical shifts for the GS (ω_{GS}) and ES1 (ω_{ES1}) were computed from the measured average chemical shift (ω_{avg}), exchange parameters ($\Delta\omega = \omega_{ES} - \omega_{GS}$ and p_B) deduced from relaxation dispersion experiments[2,76] using the following equation applicable under fast-exchange

$$\omega_{GS} = \omega_{avg} - p_{ES} \times \Delta\omega$$

$$\omega_{ES} = \omega_{avg} + p_{GS} \times \Delta\omega$$

RDC Measurements.—2D ^{13}C – ^{1}H or ^{15}N – ^{1}H S^{3}E HSQC and transverse relaxation-optimized spectroscopy (TROSY) experiments were used to measure one-bond C–H and N–H RDCs ($^1D_{CH}$) in aromatic and sugar moieties (C6H6, C8H8, C2H2, C5H5, C1′H1', and N1H1/N3H3) in EII-3 and EII-13. The 2D TROSY experiment[16] was used to measure RDCs (C8H8, C6H6, C2H2, C5H5, C1′H1') in the elongated EI-22, at 298 K using a Varian 800 MHz spectrometer equipped with a 5 mm triple-resonance HCN cryogenic probe. RDCs were calculated as the difference in splittings measured in the absence (J) and presence ($J + D$) of the Pf1 phage ordering medium. Comparison of 2D HSQC spectra showed little to no perturbations in the chemical shifts in the absence and presence of Pf1 phage, consistent with prior studies.[16] For EII-3, RDCs were measured using two

experiments in which the splittings were encoded along either the $^{13}C/^{15}N$ (indirect) or ^1H (direct) dimension. The RDC uncertainty was estimated based on the RMSD between values measured using two different experiments as described previously.[93] RDCs for EII-3 and E-0[63] were obtained by taking the average of the RDCs measured along the $^{13}C/^{15}N$ and ^1H dimensions, while for the elongated EII-13 and EI-22, only the RDCs measured along the ^1H dimension were used.[16] One bond aromatic and sugar C–H RDCs for TAR mutants C30U and A35G were measured using experiments identical to those used for EII-3, and these measurements were performed on a Bruker 800 MHz spectrometer equipped with a 5 mm triple-resonance HCN cryogenic probe. All measured RDCs are summarized in Tables S1, S2, and S6 have been deposited in BMRB and are additionally also available on GitHub: https://github.com/alhashimilab/Kinetically_resolved_ensemble_TAR.

Order Tensor Analysis of RDCs.—RDCs measured in EI-22, EII-3 and EII-13 for nonterminal Watson–Crick bps in the upper stem were subjected to an order tensor analysis using idealized A-form geometry with the helix axis oriented along the z-axis of the molecular frame, as previously described.[94] Best fit order tensor parameters are summarized in Table S2.

Off-Resonance ^{13}C $R_{1\rho}$ Relaxation Dispersion Experiments.—Off-resonance ^{13}C $R_{1\rho}$ measurements were performed using a 1D scheme that uses selective Hartman-Hahn magnetization transfers as described previously.[49] Briefly, weakly matched ^1H and ^{13}C RF fields were used to selectively transfer magnetization from protons to the ^{13}C nucleus of interest. The longitudinal ^{13}C magnetization was allowed to relax for 5 ms to allow equilibration of the substates and then tilted along the effective field direction. Then, a ^{13}C spin-lock was applied for a maximal duration (<60 ms for ^{13}C) to obtain ~70% loss in signal intensity at the end of the relaxation period. The signal intensity was recorded for 6 delays spaced over the total relaxation period. Spin-lock powers and offsets used are given in Table S5. The RD measurements were performed on samples in D_2O at pH 7.4 and $T = 25°C$ in NMR buffer unless stated otherwise. The spin-lock powers used, along with associated offsets and delay times, are summarized for each resonance in Table S5.

Analysis of Off-Resonance ^{13}C $R_{1\rho}$ Data.—Fit values for the relaxation rate in rotation frame $R_{1\rho}$ for a given spin-lock power and offset combination were obtained by fitting the measured peak intensities (extracted from NMRPipe) to a monoexponential function. Numerical integration of the Bloch–McConnell equations was then used to globally fit $R_{1\rho}$ values as a function of spin-lock power and offset to a two-state exchange model. A Monte Carlo procedure was used to estimate errors in the fitted parameters as described in prior studies.[49] Exchange parameters of interest, such as the population of ES (p_{ES}), the exchange rate ($k_{ex} = k_{forward} + k_{reverse}$), and the chemical shift difference between ES and GS ($\Delta\omega = \omega_{ES} - \omega_{GS}$) were extracted from the two-state exchange model. The fitted parameters are summarized in Table S4.

The final off-resonance profile plots were generated by plotting $R_2 + R_{ex} = (R_{1\rho} - R_{1\rho} \cos 2\theta)/\sin 2\theta$, where θ is the angle between the effective and the z axis in radians as a function of $\Omega_{OBS} = \omega_{OBS} - \omega_{RF}$, in which ω_{OBS} is the observed Larmor frequency, and ω_{RF} is the angular frequency of the applied spin-lock power.

FARFAR-NMR.

Generation of Conformation Library Using FARFAR.—FARFAR,[48] available in the Rosetta Software Suite as *rna_denovo*, was used to generate conformational libraries for the GS and ES1 as previously described.[17] Constraints on the secondary structure and base pairing were implemented for the different GS + ES libraries, as summarized in Table S3. An initial library of $N = 20,000$ conformers was generated for each state followed by Rosetta energy-based filtering (Rosetta energy 0) to remove models with potential steric clashes to generate a final library of $N = 10,000$ conformations. For ES1, an additional filter was applied to constrain G34 to a *syn* conformation with χ values between $0°$ and $100°$. (Note that the base pair mode constraint was insufficient to fix the orientation of the base in the *syn* conformation observed by using NMR relaxation dispersion). The GS and ES libraries were then combined in 1:1 ratio to form the final unbiased conformation library ($N_{Total} = 10,000$, $N_{GS} = 5000$ and $N_{ES} = 5000$).

RDC Calculations.—Ensemble-averaged RDCs were calculated by computing the RDCs for each conformer in an ensemble using the program PALES.[95] Ensemble-averaged RDCs were calculated by averaging over all conformers in the ensemble, assuming each conformer to be equally probable. Predictions for each of the four RDC data sets were obtained as in prior studies by in silico helix elongation using an idealized A-form geometry prior to the PALES calculations. Predictions for each data set were scaled independently to account for variations in phage concentrations across experiments using the scaling factor defined as follows:

$$L_i = \frac{\sum_j D_{i,j}^{meas} \times D_{i,j}^{pred}}{\sum_k D_{i,k}^{pred} \times D_{i,k}^{pred}}$$

where $D_{i,j}^{pred}$ and $D_{i,j}^{meas}$ are the predicted and measured RDC for the j th bond vector in the i th data set.

Sample and Select.—To select a subset of conformations from the library that best satisfied the measured RDCs, a simulated annealing Monte Carlo sampling algorithm was used to minimize the cost function $\left(\chi^2\right)$ representing the squared error between predicted and measured values

$$\chi^2 = \frac{\sum_i \sum_j \left(L_i \times D_{i,j}^{pred} - D_{i,j}^{meas}\right)}{n}$$

in which $D_{i,j}^{pred}$ and $D_{i,j}^{meas}$ are the predicted and measured RDC for the j th bond vector in the i th data set, respectively; L_i is the scaling factor; and n is the number of bond

vectors. Simulated annealing was performed from a starting effective temperature of 100 and decreased by *a* factor of 0.9 every 500,000 MC steps for 100 decrements. SAS was performed initially by varying the ensemble size from 1 to 50 and the optimal ensemble size was selected to be the smallest size satisfying the RDCs to avoid overfitting. The optimal ensemble size was used to run SAS 100 times, and the ensemble with the lowest RMSD from these repeats was selected as the FARFAR-NMR ensemble. RDCs from terminal ends or bps adjacent to inserted elongations in the various RDC constructs were not used in ensemble optimization.

RDC Cross-Validation.—The FARFAR-NMR ensemble was cross-validated using a 10× cross-validation in which 10% of the RDCs were systematically removed in 10 repeats to test the ability of the ensemble generated with the remaining data to predict the RDCs that were left out, ensuring that each residue had at least one RDC within the training data. The RDCs were predicted with a slightly higher RDC RMSD = 5.6 Hz than the optimized FARFAR-NMR ensemble (RMSD = 3.0 Hz), but significantly lower than the unoptimized FARFAR-library (RMSD = 9.6 Hz) indicating that the improvement in RDC agreement in the optimized ensemble was not due to overfitting the data (Figure S3B).

Subensemble Validation Using Mutants.—The ES1 and GS subensembles obtained from FARFAR-NMR were used to predict ES1 and GS RDCs that were compared against RDCs measured on the ES1-trapping TAR mutants (C30U and A35G) using PALES prediction of alignment along with appropriate scaling as previously described. RDCs measured on the residues being mutated were removed from the analysis.

Automated Fragmentation QM/MM Calculations.

Fragment Generation.—The FARFAR-NMR ensemble was subjected to ab initio chemical shift calculations using the previously described AFNMR software.[50] Each RNA conformer in the ensemble was subjected to five conjugate gradient energy minimization steps with 2 kcal/mol $Å^2$ harmonic restraints on heavy atoms to regularize bond lengths and minimize noise in predictions. Each residue was modeled as a quantum mechanical fragment with a full quantum mechanical representation of all atoms within a 3.4 Å distance cutoff. Atoms outside this quantum core, including water and ions were modeled as an equivalent coarse grained uniform distribution of point charges on the surface of the quantum core, obtained using the Poisson–Boltzmann equation (solinprot from MEAD[96]). The quantum core, regions occupied by the conformer outside the core, and the solvent were assigned local dielectric constants (ϵ) of 1, 4, and 80, respectively.

Quantum Mechanical Simulations.—The generated fragments were used to perform GIAO–DFT calculations in Orca5[97] (version 5.0.4) using the OLYP[98] functional and pcSseg-1 (triple-*z* plus polarization) basis set optimized for nuclear magnetics shielding.[99] Reference shielding computations on tetramethylsilane were used to calculate the predicted chemical shifts obtained from the isotropic components of the computed shielding tensor. Predicted shifts were averaged for all conformers in the ensemble. A linear correction obtained from least-squares fitting was applied to the predicted ensemble-averaged chemical shifts, separately for each resonance type, for comparison against measurements as

previously described.[17,87] Predicted chemical shifts for the GS and ES1 subensembles were obtained by averaging over conformers in the GS and ES1 subensembles, respectively. For a few conformers, the predicted chemical shifts for residues C30–C1′, G34–C1′, A35–C1′, C4′ lied outside the range of typically observed values in the Biological Magnetics Resonance Data Bank (BMRB).[100] In such cases, when possible, the calculations were repeated when replacing the conformer by a nearly identical conformer in the conformational library (apical loop structural RMSD < 0.1 Å), ensuring that such a replacement does not impact the predicted RDCs (RMSD < 0.05 Hz). The conformer would be selected if the chemical shifts fell within the BMRB range. This approach was applied to the FARFAR-NMR ensemble in which one conformer with an unusually upfield shifted A35–C1′ ES1 chemical shift (~83 ppm) was replaced with a structurally similar (RMSD ~ 0.1 Å) conformer that did not impact the RDC agreement (difference in predicted RDC RMSD ~ 0.02 Hz) but showed a significantly different chemical shift prediction (~90 ppm) despite minimal structural difference in the orientation of A35 compared to the original conformer. Such outliers in QM/MM predictions need to be investigated further to enhance the robustness of chemical shift predictions.

Ensemble Analysis.

Structural Characterization.—All structural visualization for analyzing structures and generating 3D models in figures was performed using PyMOL (version 2.5.4). Structural features such as base pair and base pair step parameters, sugar backbone torsional angles (alpha, beta, gamma, delta, epsilon), sugar puckering, coaxial stacking between helices, and hydrogen bonding interactions were parsed from PDB structural coordinates using 3DNA[101] (version). Interhelical Euler angles (α_h, β_h, γ_h) about the TAR bulge were computed as previously described[102] by superimposing idealized A-form geometry on three consecutive bps (lower helix: C19–G43, A20–U42, and G21–C41; upper helix: G26–C39, A27–U38, and G28–C37) flanking both sides of the bulge and computing the relative orientation of the helical axis of the two fit idealized A-form helices. The values of α_h and γ_h were inverted relative to the previously reported values[102] ensure that positive and negative values of ($\alpha_h + \gamma_h$) correspond to over and under-twisting, respectively. All heavy atom structural RMSDs were computed using the *rmsd* command available in CPPTRAJ.[103] Sugar pucker populations for apical loop residues were deduced from C1′ chemical shifts as previously described.[17]

Filtering of ES1-Like Conformers from Conformational Library.—To test the quality of fit for ensemble optimized on conformational library without any ES1 or ES1-like states, a new conformational library ($N = 10,000$) using GS secondary structural constraints was generated as mentioned previously for the unbiased library. In addition to energy score-based filtering to obtain the final library, conformations were additionally filtered to remove ES1-like features, which was characterized using the angle between sugar-base vectors (vector from geometric center of sugar atoms to that of nucleobase atoms) for residues G34 and A35, defined as cos φ. In GS, A35 is flipped out relative to G34, and hence the dot product is lower (cos$\varphi \leq 0.5$). In ES1, A35 is stacked in with G34, hence the dot product is

higher (cos φ > 0.5). All conformers with dot product >0.5 were filtered out to obtain the final GS-only library.

Comparison with X-ray Structure.—The contacts formed between the TAR apical loop and proteins Tat and CyclinT1 were identified based on the previously solved X-ray structure of the complex[56] (PDBID 6CYT). Pairwise heavy-atom RMSDs (residues 29–36) were computed for all 20 conformers in the FARFAR-NMR ensemble against the solved structure of the apical loop, and the lowest RMSD conformers within each subensemble was overlaid onto the X-ray structure using *align* command in PyMOL (residues 29–36 used for alignment).

Calculation of Spin Relaxation Order Parameters.—The Lipari–Szabo order parameter $\left(S^2\right)$ was calculated for the FARFAR-NMR ensemble from 3D coordinates of the conformers as previously described,[46] by first aligning all backbone atoms in E-0 and then computing the normalized C–H internuclear bond vector (μ_i) for each resonance of interest to obtain the predicted S^2 values using the following equation

$$S^2 = \frac{1}{2}\left(3\sum_{i=1}^{3}\ \sum_{j=1}^{3}\ \langle \mu_i\mu_j \rangle^2 - 1\right)$$

where μ_i represent the Cartesian coordinates of the normalized internuclear vector and $\langle\ \rangle$ denotes averaging over all conformations in the ensemble. The relative order parameter $\left(S_{\text{relative}}^2\right)$ was then computed from the predicted S^2 value by normalizing relative to the largest corresponding value within the helical residues for each resonance type (C6, C8, C2) independently, in line with the normalization procedure used in prior spin relaxation measurements.[63]

PDB Survey of NMR Structures of TAR Apical Loop.—All solution NMR structures containing the HIV TAR apical loop (both apo and ligand-bound states) were downloaded from the RCSB PDB and filtered to select the subset of structures without any ligand interactions with the apical loop. The filtered set of structures were processed by X3DNA-DSSR[101] (version 1.9) to characterize structural features in the apical loop as described above.

Figure Design.—The results from all analyses were converted into the csv format, and final plots were made within either Jupyter notebooks running Python (version 3.7) or Mathematica (version 13.1) or GraphPad Prism (version 8.2.1). Chemical structures were generated by using ChemDraw (version 20.1). Images originally generated from other sources were exported into Adobe Illustrator 2020 (version 24.5) to be incorporated into the final figures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

Data Availability Statement

The ensemble model of TAR with the GS and ES substates along with all relevant NMR data used for RDC optimization and chemical shift cross-validation has been deposited in the PDB, available with PDB ID 8THV. RDCs measured on ES1-stabilizing mutants have been deposited in BMRB with IDs 52052 for C30U and 52053 for A35G. All raw data, structural models, PyMOL sessions, and scripts or Jupyter notebooks used are available on GitHub at https://github.com/alhashimilab/Kinetically_resolved_ensemble_TAR. Any further information is available from the corresponding authors upon request.

REFERENCES

(1). Ganser LR; Kelly ML; Herschlag D; Al-Hashimi HM The roles of structural dynamics in the cellular functions of RNAs. Nat. Rev. Mol. Cell Biol. 2019, 20, 474–489. [PubMed: 31182864]

(2). Dethoff EA; Petzold K; Chugh J; Casiano-Negroni A; Al-Hashimi HM Visualizing transient low-populated structures of RNA. Nature 2012, 491, 724–728. [PubMed: 23041928]

(3). Liu B; Shi H; Al-Hashimi HM Developments in solution-state NMR yield broader and deeper views of the dynamic ensembles of nucleic acids. Curr. Opin. Struct. Biol. 2021, 70, 16–25. [PubMed: 33836446]

(4). Mulder FAA; Mittermaier A; Hon B; Dahlquist FW; Kay LE Studying excited states of proteins by NMR spectroscopy. Nat. Struct. Biol. 2001, 8, 932–935. [PubMed: 11685237]

(5). Helmling C; Klötzner DP; Sochor F; Mooney RA; Wacker A; Landick R; Fürtig, B.; Heckel, A.; Schwalbe, H. Life times of metastable states guide regulatory signaling in transcriptional riboswitches. Nat. Commun. 2018, 9, 944. [PubMed: 29507289]

(6). Chen B; LeBlanc R; Dayie TK SAM-II Riboswitch Samples at least Two Conformations in Solution in the Absence of Ligand: Implications for Recognition. Angew. Chem., Int. Ed. 2016, 55, 2724–2727.

(7). Zhao B; Guffy SL; Williams B; Zhang Q An excited state underlies gene regulation of a transcriptional riboswitch. Nat. Chem. Biol. 2017, 13, 968–974. [PubMed: 28719589]

(8). Xue Y; Gracia B; Herschlag D; Russell R; Al-Hashimi HM Visualizing the formation of an RNA folding intermediate through a fast highly modular secondary structure switch. Nat. Commun. 2016, 7, ncomms11768. [PubMed: 27292179]

(9). Plangger R; Juen MA; Hoernes TP; Nußbaumer F; Kremser J; Strebitzer E; Klingler D; Erharter K; Tollinger M; Erlacher MD; et al. Branch site bulge conformations in domain 6 determine functional sugar puckers in group II intron splicing. Nucleic Acids Res. 2019, 47, 11430–11440. [PubMed: 31665419]

(10). Baronti L; Guzzetti I; Ebrahimi P; Friebe Sandoz S; Steiner E; Schlagnitweit J; Fromm B; Silva L; Fontana C; Chen AA; et al. Base-pair conformational switch modulates miR-34a targeting of Sirt1 mRNA. Nature 2020, 583, 139–144. [PubMed: 32461691]

(11). Baisden JT; Boyer JA; Zhao B; Hammond SM; Zhang Q Visualizing a protonated RNA state that modulates microRNA-21 maturation. Nat. Chem. Biol. 2021, 17, 80–88. [PubMed: 33106660]

(12). Brown JD; Kharytonchyk S; Chaudry I; Iyer AS; Carter H; Becker G; Desai Y; Glang L; Choi SH; Singh K; et al. Structural basis for transcriptional start site control of HIV-1 RNA fate. Science 2020, 368, 413–417. [PubMed: 32327595]

(13). Houck-Loomis B; Durney MA; Salguero C; Shankar N; Nagle JM; Goff SP; D'Souza VM An equilibrium-dependent retroviral mRNA switch regulates translational recoding. Nature 2011, 480, 561–564. [PubMed: 22121021]

(14). Ganser LR; Chu CC; Bogerd HP; Kelly ML; Cullen BR; Al-Hashimi HM Probing RNA Conformational Equilibria within the Functional Cellular Context. Cell Rep. 2020, 30, 2472–2480.e4. [PubMed: 32101729]

(15). Ganser LR; Kelly ML; Patwardhan NN; Hargrove AE; Al-Hashimi HM Demonstration that Small Molecules can Bind and Stabilize Low-abundance Short-lived RNA Excited Conformational States. J. Mol. Biol. 2020, 432, 1297–1304. [PubMed: 31863746]

(16). Zhang Q; Stelzer AC; Fisher CK; Al-Hashimi HM Visualizing spatially correlated dynamics that directs RNA conformational transitions. Nature 2007, 450, 1263–1267. [PubMed: 18097416]

(17). Shi H; Rangadurai A; Abou Assi H; Roy R; Case DA; Herschlag D; Yesselman JD; Al-Hashimi HM Rapid and accurate determination of atomistic RNA dynamic ensemble models using NMR and structure prediction. Nat. Commun. 2020, 11, 5531. [PubMed: 33139729]

(18). Bottaro S; Nichols PJ; Vögeli B; Parrinello M; Lindorff-Larsen K Integrating NMR and simulations reveals motions in the UUCG tetraloop. Nucleic Acids Res. 2020, 48, 5839–5848. [PubMed: 32427326]

(19). Nichols PJ; Henen MA; Born A; Strotz D; Güntert, P.; Vögeli, B. High-resolution small RNA structures from exact nuclear Overhauser enhancement measurements without additional restraints. Commun. Biol. 2018, 1, 61. [PubMed: 30271943]

(20). Hansen AL; Al-Hashimi HM Dynamics of Large Elongated RNA by NMR Carbon Relaxation. J. Am. Chem. Soc. 2007, 129, 16072–16082. [PubMed: 18047338]

(21). Mustoe AM; Brooks CL; Al-Hashimi HM Hierarchy of RNA Functional Dynamics. Annu. Rev. Biochem. 2014, 83, 441–466. [PubMed: 24606137]

(22). Hermann T; Patel DJ Adaptive Recognition by Nucleic Acid Aptamers. Science 2000, 287, 820–825. [PubMed: 10657289]

(23). Williamson JR Induced fit in RNA-protein recognition. Nat. Struct. Biol. 2000, 7, 834–837. [PubMed: 11017187]

(24). Dethoff EA; Chugh J; Mustoe AM; Al-Hashimi HM Functional complexity and regulation through RNA dynamics. Nature 2012, 482, 322–330. [PubMed: 22337051]

(25). Chen YL; Lee T; Elber R; Pollack L Conformations of an RNA Helix-Junction-Helix Construct Revealed by SAXS Refinement of MD Simulations. Biophys. J. 2019, 116, 19–30. [PubMed: 30558889]

(26). Vazquez Reyes C; Tangprasertchai NS; Yogesha SD; Nguyen RH; Zhang X; Rajan R; Qin PZ Nucleic Acid-Dependent Conformational Changes in CRISPR–Cas9 Revealed by Site-Directed Spin Labeling. Cell Biochem. Biophys. 2017, 75, 203–210. [PubMed: 27342128]

(27). Tomezsko PJ; Corbin VDA; Gupta P; Swaminathan H; Glasgow M; Persad S; Edwards MD; Mcintosh L; Papenfuss AT; Emery A; et al. Determination of RNA structural diversity and its role in HIV-1 RNA splicing. Nature 2020, 582, 438–442. [PubMed: 32555469]

(28). Fraser JS; van den Bedem H; Samelson AJ; Lang PT; Holton JM; Echols N; Alber T Accessing protein conformational ensembles using room-temperature X-ray crystallography. Proc. Natl. Acad. Sci. U.S.A. 2011, 108, 16247–16252. [PubMed: 21918110]

(29). Stagno JR; Liu Y; Bhandari YR; Conrad CE; Panja S; Swain M; Fan L; Nelson G; Li C; Wendel DR; et al. Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography. Nature 2017, 541, 242–246. [PubMed: 27841871]

(30). Chen B; Frank J Two promising future developments of cryo-EM: capturing short-lived states and mapping a continuum of states of a macromolecule. Microscopy 2016, 65, 69–79. [PubMed: 26520784]

(31). Bonilla SL; Sherlock ME; MacFadden A; Kieft JS A viral RNA hijacks host machinery using dynamic conformational changes of a tRNA-like structure. Science 2021, 374, 955–960. [PubMed: 34793227]

(32). Bonilla SL; Vicens Q; Kieft JS Cryo-EM reveals an entangled kinetic trap in the folding of a catalytic RNA. Sci. Adv. 2022, 8, No. eabq4144. [PubMed: 36026457]

(33). Alderson TR; Kay LE Unveiling invisible protein states with NMR spectroscopy. Curr. Opin. Struct. Biol. 2020, 60, 39–49. [PubMed: 31835059]

(34). Jensen MR; Ruigrok RWH; Blackledge M Describing intrinsically disordered proteins at atomic resolution by NMR. Curr. Opin. Struct. Biol. 2013, 23, 426–435. [PubMed: 23545493]

(35). Vallurupalli P; Hansen DF; Kay LE Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. Proc. Natl. Acad. Sci. U.S.A. 2008, 105, 11766–11771. [PubMed: 18701719]

(36). Korzhnev DM; Religa TL; Banachewicz W; Fersht AR; Kay LE A transient and low-populated protein-folding intermediate at atomic resolution. Science 2010, 329, 1312–1316. [PubMed: 20829478]

(37). Bouvignies G; Vallurupalli P; Hansen DF; Correia BE; Lange O; Bah A; Vernon RM; Dahlquist FW; Baker D; Kay LE Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. Nature 2011, 477, 111–114. [PubMed: 21857680]

(38). Stiller JB; Otten R; Häussinger D; Rieder PS; Theobald DL; Kern D Structure determination of high-energy states in a dynamic protein ensemble. Nature 2022, 603, 528–535. [PubMed: 35236984]

(39). Tang C; Schwieters CD; Clore GM Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. Nature 2007, 449, 1078–1082. [PubMed: 17960247]

(40). Prestegard JH New techniques in structural NMR — anisotropic interactions. Nat. Struct. Biol. 1998, 5, 517–522. [PubMed: 9665182]

(41). Bax A; Kontaxis G; Tjandra N Dipolar couplings in macromolecular structure determination. Methods Enzymol. 2001, 339, 127–174. [PubMed: 11462810]

(42). Smith CA; Ban D; Pratihar S; Giller K; Schwiegk C; de Groot BL; Becker S; Griesinger C; Lee D Population Shuffling of Protein Conformations. Angew. Chem., Int. Ed. 2015, 54, 207–210.

(43). Pratihar S; Sabo TM; Ban D; Fenwick RB; Becker S; Salvatella X; Griesinger C; Lee D Kinetics of the Antibody Recognition Site in the Third IgG-Binding Domain of Protein G. Angew. Chem., Int. Ed. 2016, 55, 9567–9570.

(44). Smith CA; Mazur A; Rout AK; Becker S; Lee D; de Groot BL; Griesinger C Enhancing NMR derived ensembles with kinetics on multiple timescales. J. Biomol. NMR 2020, 74, 27–43. [PubMed: 31838619]

(45). Ban D; Funk M; Gulich R; Egger D; Sabo TM; Walter KFA; Fenwick RB; Giller K; Pichierri F; de Groot BL; et al. Kinetics of Conformational Sampling in Ubiquitin. Angew. Chem., Int. Ed. 2011, 50, 11437–11440.

(46). Markwick PRL; Bouvignies G; Blackledge M Exploring Multiple Timescale Motions in Protein GB3 Using Accelerated Molecular Dynamics and NMR Spectroscopy. J. Am. Chem. Soc. 2007, 129, 4724–4730. [PubMed: 17375925]

(47). Bernadó P; Blackledge M Local Dynamic Amplitudes on the Protein Backbone from Dipolar Couplings: Toward the Elucidation of Slower Motions in Biomolecules. J. Am. Chem. Soc. 2004, 126, 7760–7761. [PubMed: 15212507]

(48). Watkins AM; Rangan R; Das R FARFAR2: Improved De Novo Rosetta Prediction of Complex Global RNA Folds. Structure 2020, 28, 963–976.e6. [PubMed: 32531203]

(49). Rangadurai A; Szymaski ES; Kimsey IJ; Shi H; Al-Hashimi HM Characterizing micro-to-millisecond chemical exchange in nucleic acids using off-resonance R1ρ relaxation dispersion. Prog. Nucl. Magn. Reson. Spectrosc. 2019, 112–113, 55–102.

(50). Swails J; Zhu T; He X; Case DA AFNMR: automated fragmentation quantum mechanical calculation of NMR chemical shifts for biomolecules. J. Biomol. NMR 2015, 63, 125–139. [PubMed: 26232926]

(51). Muesing MA; Smith DH; Capon DJ Regulation of mRNA accumulation by a human immunodeficiency virus trans-activator protein. Cell 1987, 48, 691–701. [PubMed: 3643816]

(52). Cullen BR Trans-activation of human immunodeficiency virus occurs via a bimodal mechanism. Cell 1986, 46, 973–982. [PubMed: 3530501]

(53). Frankel AD Activation of HIV transcription by Tat. Curr. Opin. Genet. Dev. 1992, 2, 293–298. [PubMed: 1638124]

(54). Puglisi J; Tan R; Calnan B; Frankel A; Williamson JR Conformation of the TAR RNA-arginine complex by NMR spectroscopy. Science 1992, 257, 76–80. [PubMed: 1621097]

(55). Pham VV; Salguero C; Khan SN; Meagher JL; Brown WC; Humbert N; de Rocquigny H; Smith JL; D'Souza VM HIV-1 Tat interactions with cellular 7SK and viral TAR RNAs identifies dual structural mimicry. Nat. Commun. 2018, 9, 4266. [PubMed: 30323330]

(56). Schulze-Gahmen U; Hurley JH Structural mechanism for HIV-1 TAR loop recognition by Tat and the super elongation complex. Proc. Natl. Acad. Sci. U.S.A. 2018, 115, 12973–12978. [PubMed: 30514815]

(57). Salmon L; Bascom G; Andricioaei I; Al-Hashimi HM A General Method for Constructing Atomic-Resolution RNA Ensembles using NMR Residual Dipolar Couplings: The Basis for Interhelical Motions Revealed. J. Am. Chem. Soc. 2013, 135, 5457–5466. [PubMed: 23473378]

(58). Ken ML; Roy R; Geng A; Ganser LR; Manghrani A; Cullen BR; Schulze-Gahmen U; Herschlag D; Al-Hashimi HM RNA conformational propensities determine cellular activity. Nature 2023, 617, 835–841. [PubMed: 37198487]

(59). Orlovsky NI; Al-Hashimi HM; Oas TG Exposing Hidden High-Affinity RNA Conformational States. J. Am. Chem. Soc. 2020, 142, 907–921. [PubMed: 31815464]

(60). Yates LA; Norbury CJ; Gilbert RJC The Long and Short of MicroRNA. Cell 2013, 153, 516–519. [PubMed: 23622238]

(61). Gong Z; Schwieters CD; Tang C Theory and practice of using solvent paramagnetic relaxation enhancement to characterize protein conformational dynamics. Methods 2018, 148, 48–56. [PubMed: 29656079]

(62). Kelly ML; Chu CC; Shi H; Ganser LR; Bogerd HP; Huynh K; Hou Y; Cullen BR; Al-Hashimi HM Understanding the characteristics of nonspecific binding of drug-like compounds to canonical stem–loop RNAs and their implications for functional cellular assays. RNA 2021, 27, 12–26. [PubMed: 33028652]

(63). Dethoff EA; Hansen AL; Musselman C; Watt ED; Andricioaei I; Al-Hashimi HM Characterizing complex dynamics in the transactivation response element apical loop and motional correlations with the bulge by NMR, molecular dynamics, and mutagenesis. Biophys. J. 2008, 95, 3906–3915. [PubMed: 18621815]

(64). Lipari G; Szabo A Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. J. Am. Chem. Soc. 1982, 104, 4546–4559.

(65). Tjandra N; Bax A Direct Measurement of Distances and Angles in Biomolecules by NMR in a Dilute Liquid Crystalline Medium. Science 1997, 278, 1111–1114. [PubMed: 9353189]

(66). Tolman JR; Flanagan JM; Kennedy MA; Prestegard JH Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. Proc. Natl. Acad. Sci. U.S.A. 1995, 92, 9279–9283. [PubMed: 7568117]

(67). Hansen MR; Mueller L; Pardi A Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. Nat. Struct. Biol. 1998, 5, 1065–1074. [PubMed: 9846877]

(68). Tolman JR; Flanagan JM; Kennedy MA; Prestegard JH NMR evidence for slow collective motions in cyanometmyoglobin. Nat. Struct. Biol. 1997, 4, 292–297. [PubMed: 9095197]

(69). Ulmer TS; Ramirez BE; Delaglio F; Bax A Evaluation of Backbone Proton Positions and Dynamics in a Small Protein by Liquid Crystal NMR Spectroscopy. J. Am. Chem. Soc. 2003, 125, 9179–9191. [PubMed: 15369375]

(70). Yang S; Al-Hashimi HM Unveiling inherent degeneracies in determining population-weighted ensembles of interdomain orientational distributions using NMR residual dipolar couplings: application to RNA helix junction helix motifs. J. Phys. Chem. B 2015, 119, 9614–9626. [PubMed: 26131693]

(71). Hansen MR; Hanson P; Pardi A Pf1 Filamentous Phage as an Alignment Tool for Generating Local and Global Structural Information in Nucleic Acids. J. Biomol. Struct. Dyn. 2000, 17, 365–369. [PubMed: 22607445]

(72). Bailor MH; Musselman C; Hansen AL; Gulati K; Patel DJ; Al-Hashimi HM Characterizing the relative orientation and dynamics of RNA A-form helices using NMR residual dipolar couplings. Nat. Protoc. 2007, 2, 1536–1546. [PubMed: 17571061]

(73). Chen Y; Campbell SL; Dokholyan NV Deciphering Protein Dynamics from NMR Data Using Explicit Structure Sampling and Selection. Biophys. J. 2007, 93, 2300–2306. [PubMed: 17557784]

(74). Lee J; Dethoff EA; Al-Hashimi HM Invisible RNA state dynamically couples distant motifs. Proc. Natl. Acad. Sci. U.S.A. 2014, 111, 9485–9490. [PubMed: 24979799]

(75). Case DA Chemical shifts in biomolecules. Curr. Opin. Struct. Biol. 2013, 23, 172–176. [PubMed: 23422068]

(76). Clay MC; Ganser LR; Merriman DK; Al-Hashimi HM Resolving sugar puckers in RNA excited states exposes slow modes of repuckering dynamics. Nucleic Acids Res. 2017, 45, No. e134. [PubMed: 28609788]

(77). Palmer AG NMR Characterization of the Dynamics of Biomacromolecules. Chem. Rev. 2004, 104, 3623–3640. [PubMed: 15303831]

(78). Zhang Q; Sun X; Watt ED; Al-Hashimi HM Resolving the Motional Modes That Code for RNA Adaptation. Science 2006, 311, 653–656. [PubMed: 16456078]

(79). Lange OF; Lakomek NA; Farès C; Schröder GF; Walter KFA; Becker S; Meiler J; Grubmüller H; Griesinger C; de Groot BL Recognition Dynamics Up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution. Science 2008, 320, 1471–1475. [PubMed: 18556554]

(80). Xie M; Yu L; Bruschweiler-Li L; Xiang X; Hansen AL; Brüschweiler, R. Functional protein dynamics on uncharted time scales detected by nanoparticle-assisted NMR spin relaxation. Sci. Adv. 2019, 5, No. eaax5560. [PubMed: 31453342]

(81). Jumper J; Evans R; Pritzel A; Green T; Figurnov M; Ronneberger O; Tunyasuvunakool K; Bates R; Žídek A; Potapenko A; et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021, 596, 583–589. [PubMed: 34265844]

(82). Townshend RJL; Eismann S; Watkins AM; Rangan R; Karelina M; Das R; Dror RO Geometric deep learning of RNA structure. Science 2021, 373, 1047–1051. [PubMed: 34446608]

(83). Aboul-ela F; Karn J; Varani G Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. Nucleic Acids Res. 1996, 24, 3974–3981. [PubMed: 8918800]

(84). Du Z; Lind KE; James TL Structure of TAR RNA Complexed with a Tat-TAR Interaction Nanomolar Inhibitor that Was Identified by Computational Screening. Chem. Bio. 2002, 9, 707–712. [PubMed: 12079782]

(85). Faber C; Sticht H; Schweimer K; Rosch P Structural rearrangements of HIV-1 Tat-responsive RNA upon binding of neomycin B. J. Biol. Chem. 2000, 275, 20660–20666. [PubMed: 10747964]

(86). Salmon L; Giambasu GM; Nikolova EN; Petzold K; Bhattacharya A; Case DA; Al-Hashimi HM Modulating RNA alignment using directional dynamic kinks: application in determining an atomic-resolution ensemble for a hairpin using NMR residual dipolar couplings. J. Am. Chem. Soc. 2015, 137, 12954–12965. [PubMed: 26306428]

(87). Shi H; Clay MC; Rangadurai A; Sathyamoorthy B; Case DA; Al-Hashimi HM 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 2018, 70, 229–244. [PubMed: 29675775]

(88). Chu C-C; Plangger R; Kreutz C; Al-Hashimi HM Dynamic ensemble of HIV-1 RRE stem IIB reveals non-native conformations that disrupt the Rev-binding site. Nucleic Acids Res. 2019, 47, 7105–7117. [PubMed: 31199872]

(89). Santner T; Rieder U; Kreutz C; Micura R Pseudoknot Preorganization of the PreQ1 Class I Riboswitch. J. Am. Chem. Soc. 2012, 134, 11928–11931. [PubMed: 22775200]

(90). Abou Assi H; Rangadurai AK; Shi H; Liu B; Clay MC; Erharter K; Kreutz C; Holley CL; Al-Hashimi H 2′-O-Methylation can increase the abundance and lifetime of alternative RNA conformational states. Nucleic Acids Res. 2020, 48, 12365–12379. [PubMed: 33104789]

(91). Delaglio F; Grzesiek S; Vuister G; Zhu G; Pfeifer J; Bax A NMRPipe: A multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR 1995, 6, 277–293. [PubMed: 8520220]

(92). Lee W; Tonelli M; Markley JL NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. Bioinformatics 2015, 31, 1325–1327. [PubMed: 25505092]

(93). Al-Hashimi HM; Gorin A; Majumdar A; Gosser Y; Patel DJ Towards structural genomics of RNA: rapid NMR resonance assignment and simultaneous RNA tertiary structure determination using residual dipolar couplings. J. Mol. Biol. 2002, 318, 637–649. [PubMed: 12054812]

(94). Hansen AL; Al-Hashimi HM Insight into the CSA tensors of nucleobase carbons in RNA polynucleotides from solution measurements of residual CSA: Towards new long-range orientational constraints. J. Magn. Reson. 2006, 179, 299–307. [PubMed: 16431143]

(95). Zweckstetter M NMR: prediction of molecular alignment from structure using the PALES software. Nat. Protoc. 2008, 3, 679–690. [PubMed: 18388951]

(96). Bashford D; Karplus M pKa's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. Biochemistry 1990, 29, 10219–10225. [PubMed: 2271649]

(97). Neese F Software update: The ORCA program system Version 5.0. Wiley Interdiscip. Rev.: Comput. Mol. Sci. 2022, 12, No. e1606.

(98). Handy NC; Cohen AJ Left-right correlation energy. Mol. Phys. 2001, 99, 403–412.

(99). Jensen F Segmented contracted basis sets optimized for nuclear magnetic shielding. J. Chem. Theory Comput. 2015, 11, 132–138. [PubMed: 26574211]

(100). Romero PR; et al. In Structural Bioinformatics: Methods and Protocols; Gáspári Z, Ed.; Springer US, 2020; pp 187–218.

(101). Lu X-J; Bussemaker HJ; Olson WK DSSR: an integrated software tool for dissecting the spatial structure of RNA. Nucleic Acids Res. 2015, 43, gkv716.

(102). Bailor MH; Mustoe AM; Brooks CL; Al-Hashimi HM; Al-Hashimi HM 3D maps of RNA interhelical junctions. Nat. Protoc. 2011, 6, 1536–1545. [PubMed: 21959236]

(103). Roe DR; Cheatham TE III PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. J. Chem. Theory Comput. 2013, 9, 3084–3095. [PubMed: 26583988]

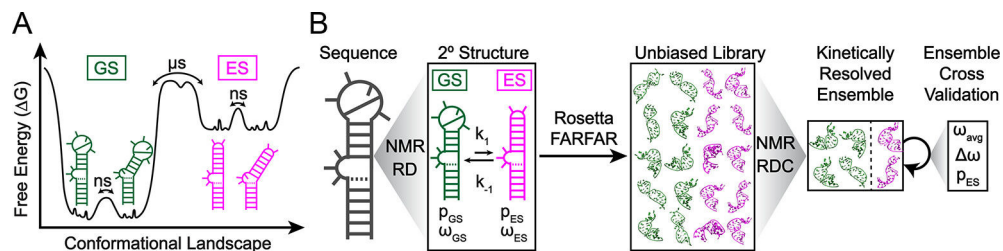**Figure 1.**

Kinetic resolution of RNA ensembles by integrating FARFAR-NMR with NMR chemical exchange measurements. (A) A representative RNA free energy landscape. The energetically most favorable GS (green) interconverts slowly on the microsecond time scale with a lowly populated ES (pink). The GS and ES form ensembles of many conformations interconverting on the faster pico-to-nanosecond time scale. (B) Pipeline for determining kinetically resolved dynamic ensembles of the RNA GS and ES by integrating FARFAR-NMR[17] with NMR chemical exchange measurements. NMR chemical exchange is used to characterize the secondary structures of GS and ES and their kinetic rates of interconversion. A conformational library is generated using Rosetta FARFAR structure prediction[48] for the NMR-derived GS and ES secondary structures. The library is subjected to RDC optimization to obtain a kinetically resolved ensemble containing GS and ES conformations. The ensemble is evaluated using shifts (ω) averaged over the entire ensemble (ω_{avg}), the GS (ω_{GS}), and the ES (ω_{ES}), and by comparing the ES population (p_{ES}) to values determined independently using NMR chemical exchange measurements.
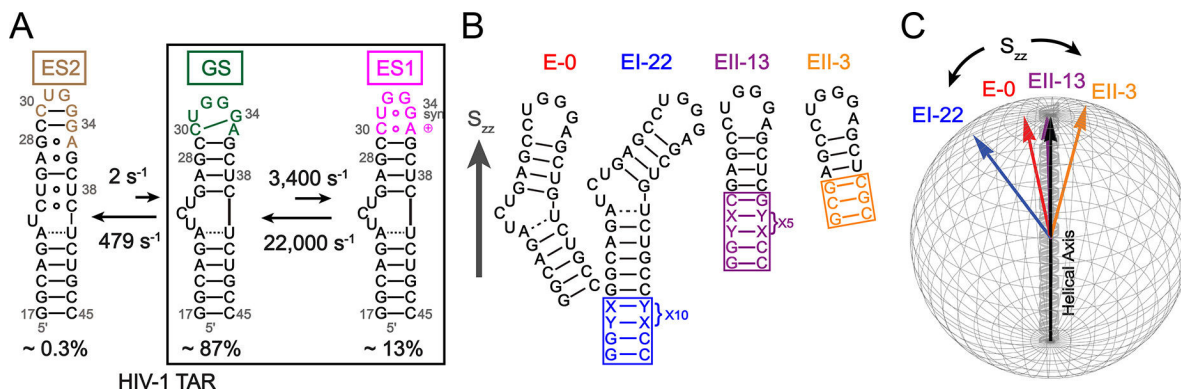
Figure 2.
Modulating partial alignment of the HIV-1 TAR using domain elongation. (A) TAR exists in a dynamic equilibrium between the GS and two ESs (ES1 and ES2). Shown are the secondary structures, populations, and kinetic rates of interconversion deduced in prior NMR relaxation dispersion studies.[2,74] The apical loop is highlighted in each state. (B) Secondary structures of the four HIV-1 TAR variants used to measure RDCs and to modulate partial alignment relative to the magnetic field direction (on average oriented along S_{zz}) when dissolved in the Pf1 phage ordering medium. (C) Globes showing the experimentally determined orientation of the principal axis of the alignment tensor (S_{zz}) for all four variants relative to the axis of the TAR upper helix.
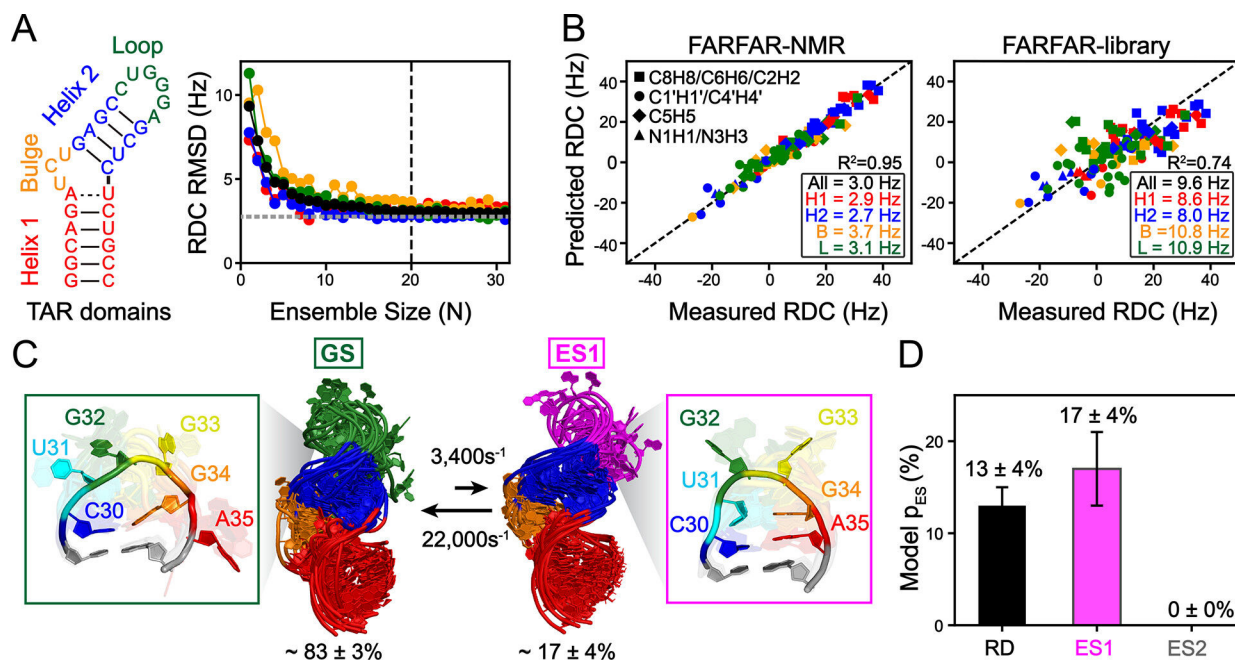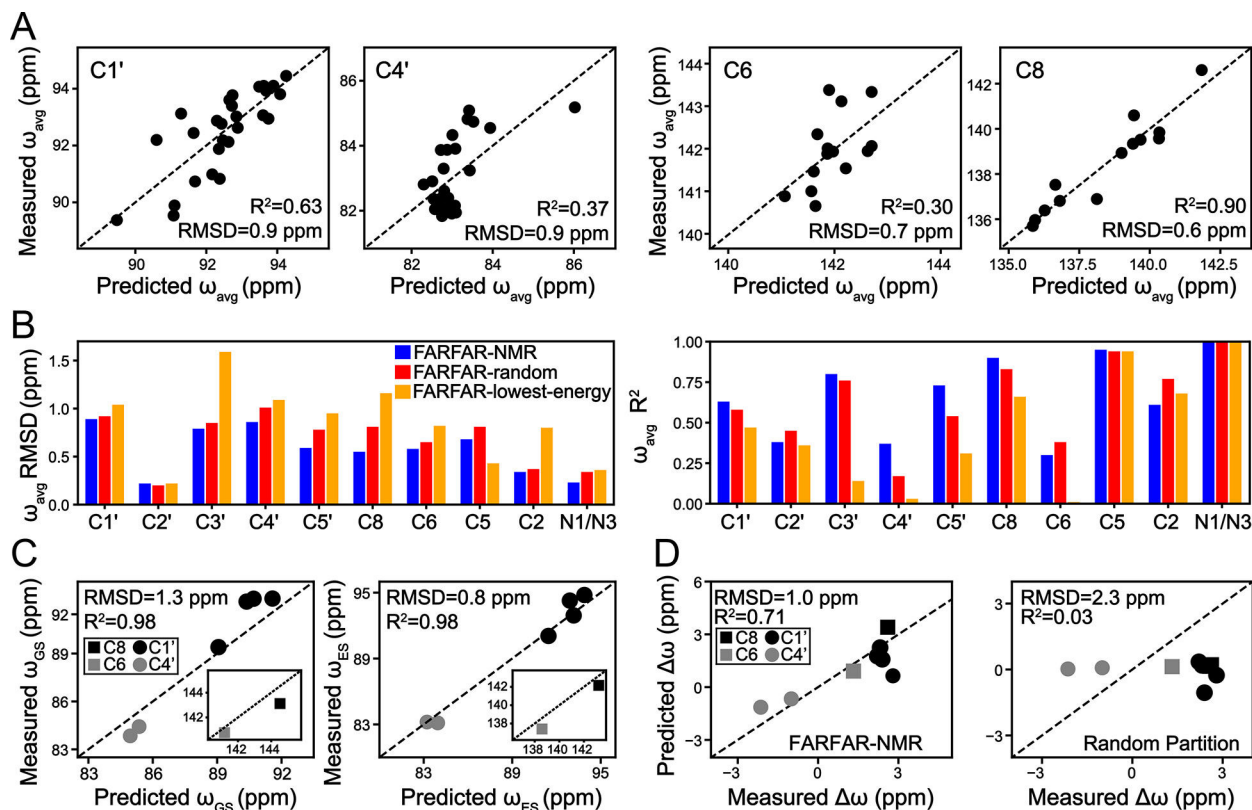
Figure 3.
Determining the HIV-1 TAR ensemble by integrating FARFAR-NMR and NMR relaxation dispersion. (A) (Left) Secondary structure of TAR with color-coded motifs. (Right) The RMSD between measured and predicted RDCs as a function of increasing ensemble size (N) during SAS. Values are shown for all RDCs (in black) as well as for the color-coded motifs. Also shown is the optimal ensemble size (N = 20) used to generate the FARFAR-NMR ensemble. Gray dashed line indicates the estimated RDC uncertainty (2.8 Hz). (B) Comparison between measured and predicted RDCs for the FARFAR-NMR ensemble (N = 20) and the FARFAR-library (N = 10,000). Motif-specific RDC RMSDs are shown in legend. (C) Overlay of ES1 and GS conformations in the FARFAR-NMR ensemble. Motifs are color-coded. The ES1 is represented using the same number of conformations as the GS (N = 16) obtained from repeated SAS runs. (D) Comparison of the ES1 population in the FARFAR-NMR ensemble (in magenta, averaged over 100 SAS runs) with values obtained using NMR relaxation dispersion measurements (black).[2,76] As a negative control, also shown is the ES population obtained using SAS when replacing the ES1 conformational library with a conformational library generated for an alternative ES2 (in gray) with a different secondary structure (Figure 2A).

Figure 4.

Cross-validating the TAR ensemble using chemical shifts. (A) Representative correlation plots comparing measured versus FARFAR-NMR ($N = 20$) predicted base (C6, C8) and sugar (C1′, C4′) ^{13}C chemical shifts (ω_{avg}) averaged over the entire GS + ES1 ensemble. The chemical shifts were computed using AF-QM/MM (see Methods). All other comparisons of ^{13}C, ^{15}N and ^{1}H ω_{avg} values are included in Figure S4. A linear correction was applied to predict the chemical shifts (see Methods) as described previously.[87] (B) Comparison of RMSD (left) and R^2 (right) between measured and predicted ^{13}C/^{15}N chemical shifts (ω_{avg}) for ensembles: FARFAR-NMR (blue), FARFAR-random (red), and FARFAR-lowest-energy (orange), all with $N = 20$. (C) Comparison of measured and FARFAR-NMR predicted ^{13}C chemical shifts for the GS (ω_{GS}, left) and ES1 (ω_{ES1}, right) subensembles and values measured experimentally using relaxation dispersion.[2,76] (D) Correlation plots comparing measured and predicted $\Delta\omega$ values ($\Delta\omega = \omega_{ES1} - \omega_{GS}$) for base (black) and sugar (gray) resonances (as shown in C) for FARFAR-NMR and an ensemble in which the 20 conformers in the FARFAR-NMR were randomly partitioned to the GS and ES1 keeping the correct population of the ES1.
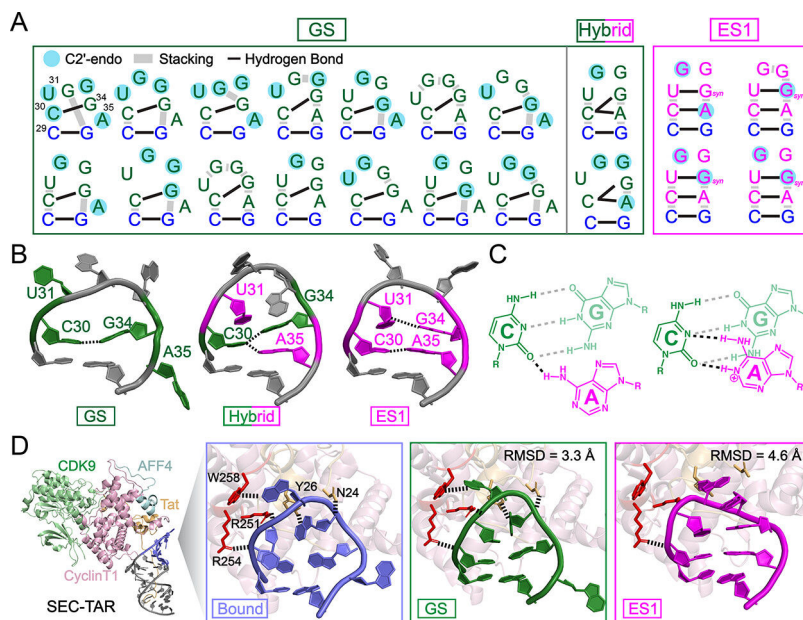
Figure 5.

Comparison of the GS and ES subensembles. (A) Schematic representation of the apical loop conformations in the FARFAR-NMR ensemble ($N = 20$) for the GS (left), hybrid GS–ES1, and ES1 (right). The two hybrid conformations are part of the GS subensemble. Shown are hydrogen bonds (black line), base-stacking (gray lines), and residues enriched in the noncanonical C2′-endo sugar pucker (highlighted in blue). (B) Representative 3D atomic structures of the GS, ES1, and hybrid conformations. Residues C30, U31, G34, and A35 are color-coded to represent whether they are GS-like (green) or ES1-like (magenta). (C) Chemical structures of the C30–G34 Watson–Crick and C30–A$^+$35 wobble bp simultaneously formed in hybrid conformations. (D) Comparison of the 3D structures of the GS and ES with crystal structure of TAR in SEC (PDB ID = 6CMN).[56] Shown is the protein bound structure of TAR using X-ray crystallography (left, blue), as well as the lowest RMSD conformations from the GS (center, green) and ES1 (right, magenta) subensembles projected onto the X-ray structure using heavy atom RMSD-based alignment (see Methods). Also shown are the RNA–protein contacts observed in the crystal structure (PDB ID = 6CMN),[56] and those which can also form with the projected GS or ES1 conformations.
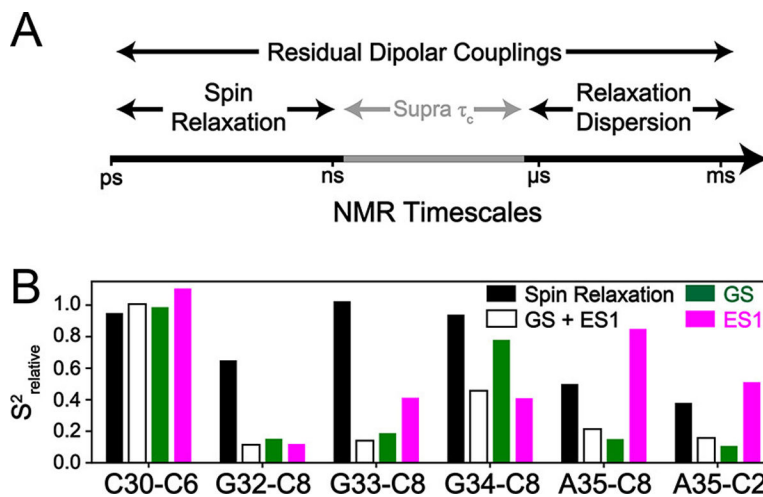
Figure 6.

The FARFAR-NMR ensemble resolves supra-τ_c motions in the TAR apical loop. (A) The range of time scales covered by NMR spin relaxation, relaxation dispersion, and RDC measurements. RDCs provide a unique opportunity to resolve supra-τ_c motions. (B) Comparison of the relative order parameter $\left(S_{relative}^2\right)$ of bond vectors from prior NMR spin relaxation measurements[63] (solid black) with corresponding values computed (see Methods) for the entire GS + ES1 FARFAR-NMR ensemble (hollow black) or separately for the GS (green) and ES1 (pink) subensembles.
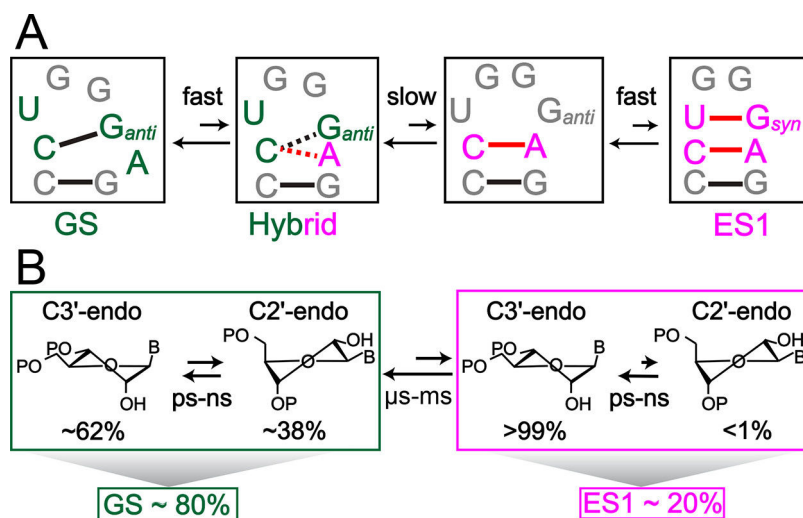
Figure 7.
Ensemble provides insights into the GS to ES1 transition. (A) Proposed mechanism for the conformational transition between the GS and ES1 involving hybrid conformations. (B) Slow (micro-to-millisecond) redistribution of fast (pico-to-nanosecond) sugar repuckering motions between C2′-endo and C3′-endo during transitions between the GS and ES1. The example shown is that for residue U31.