

Accurately Modeling RNA Stem-Loops in an Implicit Solvent Environment

Jason T. Linzer, Ethan Aminov, Aalim S. Abdullah, Colleen E. Kirkup, Rebeca I. Diaz Ventura, Vinay R. Bijoor, Jiyun Jung, Sophie Huang, Chi Gee Tse, Emily Álvarez Toucet, Hugo P. Onghai, Arghya P. Ghosh, Alex C. Grodzki, Emilee R. Haines, Aditya S. Iyer, Mark K. Khalil, Alexander P. Leong, Michael A. Neuhaus, Joseph Park, Asir Shahid, Matthew Xie, Jan M. Ziembicki, Carlos Simmerling, and Maria C. Nagan*



Cite This: *J. Chem. Inf. Model.* 2024, 64, 6092–6104



Read Online

ACCESS |



Metrics & More

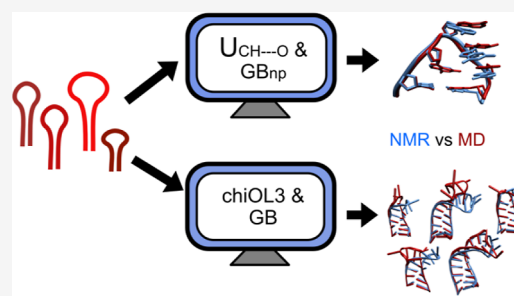


Article Recommendations



Supporting Information

ABSTRACT: Ribonucleic acid (RNA) molecules can adopt a variety of secondary and tertiary structures in solution, with stem-loops being one of the more common motifs. Here, we present a systematic analysis of 15 RNA stem-loop sequences simulated with molecular dynamics simulations in an implicit solvent environment. Analysis of RNA cluster ensembles showed that the stem-loop structures can generally adopt the A-form RNA in the stem region. Loop structures are more sensitive, and experimental structures could only be reproduced with modification of CH $\cdots$ O interactions in the force field, combined with an implicit solvent nonpolar correction to better model base stacking interactions. Accurately modeling RNA with current atomistic physics-based models remains challenging, but the RNA systems studied herein may provide a useful benchmark set for testing other RNA modeling methods in the future.



INTRODUCTION

Ribonucleic acid (RNA) molecules, expressed as single-stranded biopolymers, can adopt a wide variety of secondary structures, including canonical base-paired regions, non-standard base pairs, bulges, internal loops, pseudoknots, and multihelix junctions.^{1,2} These unique RNA secondary structures can give rise to even more complex but interesting tertiary structures including cross-strand loop–loop interactions, loop–helix interactions, and coaxial helical stacking.^{3–5} A number of computational methods are currently under development to predict tertiary RNA structure, employing a wide range of approaches including physics-based models, homology models, those that incorporate experimental information, and machine learning.^{6–8} The focus of these RNA structure prediction calculations is on overall folds with medium and larger-sized RNA molecules, which are often assembled from smaller fragments.

One of the more basic but prolific secondary structural motifs is the stem-loop.^{9,10} Stem-loops or hairpins are involved in many cellular functions including cancer regulation,¹¹ ribosomal tertiary structure,¹² and protein–RNA recognition.¹³ However, the tertiary structures of loops are complex, often containing noncanonical base pairing, base–base stacking interactions, and cross-loop interactions.¹⁴ Accurate modeling of smaller stem-loops and more generally under-

standing the contributions of chemical structure to RNA biochemical function are needed.

Despite a number of both small and large MD simulations of RNA molecules with atomistic detail,^{15,16} modeling RNA remains challenging. For instance, some larger loop structures, which were derived from crystal structures, have been successfully modeled in which strong electrostatic interactions across a loop^{17,18} and base stacking interactions agree with biochemical data.^{19,20} However, cross-loop base–base stacking²¹ and hydrogen bonding between 2'-hydroxyls and bases across the loop are difficult to model.²² A number of atomistic force fields for RNA are available in popular biomolecular simulation packages such as Amber,^{23–26} GROMOS,^{27–29} and CHARMM.^{30–32} The RNA structures that are preferred by a particular physics model are a balance of hydrogen bonding, base stacking, ion binding, and solvation. In development of the Amber protein force fields,^{33–35} they were validated against a number of biologically relevant systems and NMR

Received: May 2, 2024

Revised: June 28, 2024

Accepted: July 3, 2024

Published: July 13, 2024



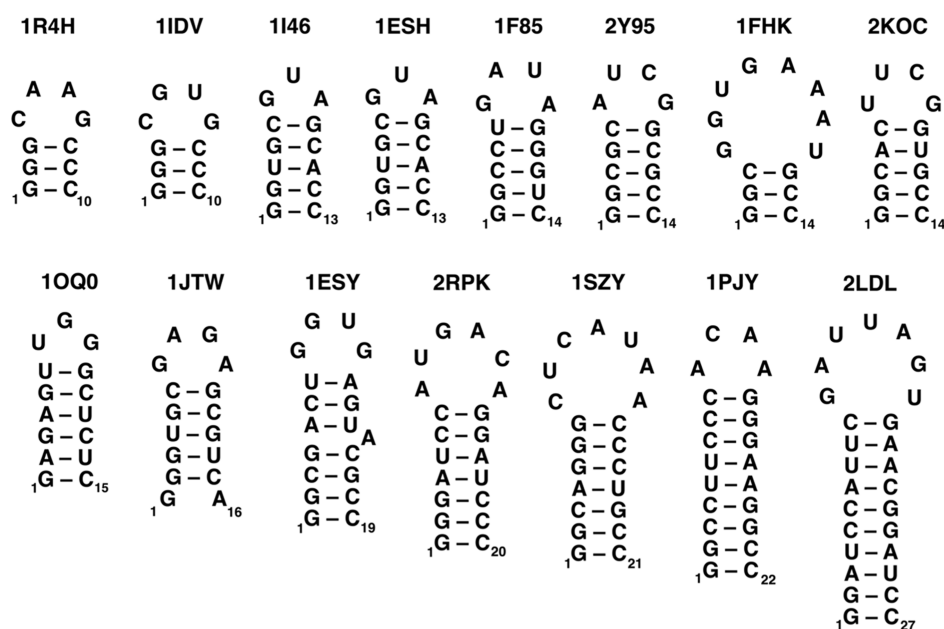


Figure 1. Secondary structures of stem-loops simulated. PDBID codes of the corresponding experimental structure are also given. See [Methods: RNA Systems](#) for references.

spectroscopic parameters to ensure that the physics model is balanced.

In addition to an accurate force field, the reliable sampling of RNA free energy landscapes continues to be difficult. The issues are coupled, and precise simulation data for multiple systems are required to improve the physics models in a more systematic manner. Replica-exchange molecular dynamics (REMD) on one or two systems is difficult,^{36,37} and RNA structures are still not sampling the correct structure without a reservoir.³⁸ State-of-the-art RNA folding studies with REMD have been performed, but the slow time scale of folding in simulations is difficult to test on a variety of systems with these methods. Another issue is that when biologically relevant systems are employed, the systems may be stable but then later new force field parameters do not work more broadly.³⁹ Simulating RNA in an implicit solvent environment is computationally more efficient than in an explicit solvent while still allowing adequate sampling, but accurate implicit solvent models for RNA have lagged behind those available for proteins.

As a starting point for gaining insights into better modeling of the RNA tertiary structure, implicit solvent simulations across a large set of diverse RNA stem-loop structures (Figure 1) are presented herein. Simulations of stem-loop RNA molecules began in the folded, experimentally determined state, and then were unfolded and refolded. Systems were heated to a high enough temperature to avoid kinetically trapped structures, providing adequate thermal energy to sample other basins. At the same time, the temperature is low enough for the MD simulation to sample low-energy favorable structures.

A variety of stem-loop RNA structures, with overall RNA lengths spanning 10 to 27 nucleotides, were chosen to avoid anecdotal conclusions. Loop length varied from 3 to 8 nucleotides and included typical RNA structural motifs such as GNRA,^{9,40} UNCG,⁴¹ and AUCG.⁴² In addition, RNA structures exhibiting base stacking interactions, novel *syn* base orientations, and C2'-endo sugar puckers were also included.

This comprehensive set of implicit solvent simulations on RNA stem-loop structures allows for a systematic identification of current issues of the simulation methods and viable improvements. Enacting such improvements could not only improve the implicit solvent RNA model but possibly RNA atomistic force fields overall and eventually may lead to more accurate MD simulations on a faster time scale.

METHODS

RNA Systems. A total of 15 stem-loop RNA structures (Figure 1) were simulated. Stem-loop sequences ranged from 10 to 27 total nucleotides, and loop length also varied from 3 to 8 nucleotides. When sequences were selected, base pairs known to require an explicit water molecule such as the G-U wobble base pair were excluded from consideration. In addition, those that required specific ion binding were also excluded, although future implicit solvent simulations could include K⁺ or Na⁺.⁴³ Simulations were begun from the lowest energy experimental structures. These included the following PDBID entries: 1R4H,⁴⁴ 1IDV,⁴⁴ 1I46,⁴⁵ 1ESH,⁴⁶ 1F85,⁴⁷ 2Y95,⁴² 1FHK,⁴⁸ 2KOC,⁴⁹ 1OQ0,⁵⁰ 1JTW,⁵¹ 1ESY,⁵² 2PRK,⁵³ 1SZY,⁵⁴ 1PJY,⁵⁵ and 2LDL,⁵⁶ except in the case of 1ESH, which was begun from the average structure as it was the only structure deposited.

Force Field and Implicit Solvent Parameters. The 15 RNA molecules described above were modeled initially with the ff99bsc0χOL3^{23–26} force field parameters known as chiOL3. In these simulations, GB-Neck2nuc⁵⁷ implicit solvent model and the mbondi2⁵⁸ intrinsic radii were employed with a 0.1 M salt concentration. Simulations of select systems (1F85, 2Y95, 1FHK, 1JTW, 2KOC, and 2LDL) were also performed with the recently developed hydrogen repulsion (HR) modification⁵⁹ to the chiOL3 force field, which will be referred to as chiOL3-HR henceforth. This RNA force field modification improves CH...O interactions and minimizes hydrogenic repulsion to select the oxygen atoms. GB models represent only the polar contribution to the solvation free energy ($\Delta G_{\text{solv,pol}}$), which tends to be much larger in

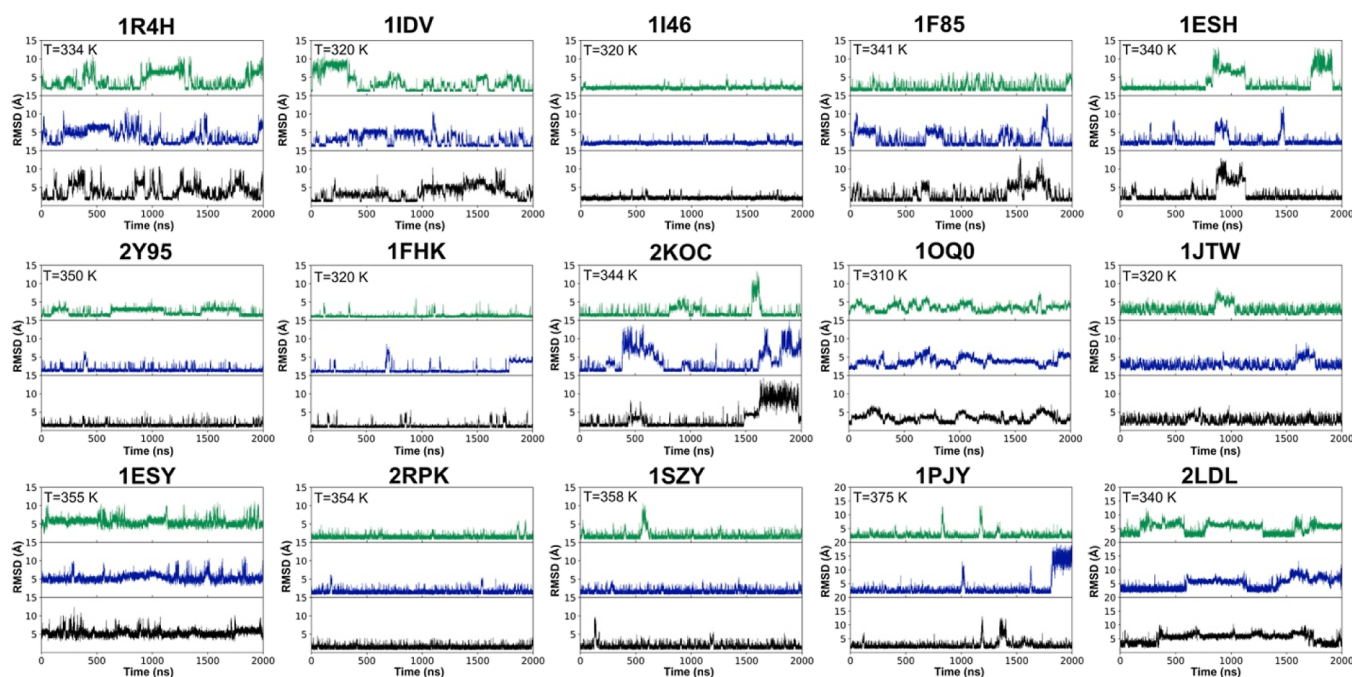


Figure 2. Plots of RMSD (Å) from the NMR structure for each stem-loop. Sequences of each RNA structure, designated by PDBIDs, correspond to those in Figure 1. All RMSDs are shown for non-hydrogenic atoms in the stem region. Each simulation was performed in triplicate (black, blue, or green indicates an independent trajectory), with the chiOL3 force field in implicit solvent at the indicated temperature.

magnitude than nonpolar solvation ($\Delta G_{\text{solv,np}}$) for polar solutes. For proteins, GB is often supplemented with a SASA-based estimate of $\Delta G_{\text{solv,np}}$ (GB/SA), and a similar approach may be required to capture base stacking in an implicit solvation model for nucleic acids. Most SASA algorithms lead to significant slowdown of the MD performance (often 10x or more), leading many to neglect this term in favor of more reliable sampling. A fast, GPU-based SASA estimate was recently developed for proteins (pwSASA⁶⁰), which incurs minimal computational overhead in the calculation of forces. However, the pwSASA model is highly specific for protein atom types and, currently, is not transferable to nucleic acids. For expediency and proof-of-concept, we developed here a specific base-stacking correction, which correlates well ($R^2 = 0.93$) with the SASA energy, that was incorporated into the implicit solvent model (described in depth in Supporting Information), since the hydrophobic effect is more applicable to the RNA bases than the sugar-phosphate backbone. This nonpolar modification of GB-Neck2nuc will be referred to as GB_{np}. Simulations of the six select stem-loops were carried out with both chiOL3-HR and GB_{np}, keeping all other parameters and conditions the same as those of the larger RNA set.

MD Simulation Conditions. Simulations were carried out with the Amber suite of simulation software,^{61–63} including system setup with *tleap*, *pdb4amber*, and *parmed*, as well as *sander* and *pmemd.cuda* for simulations and minimization and *cpttraj* for most post-trajectory analysis. General simulation protocols were similar for all systems, as outlined in previous implicit solvent RNA simulations.³⁸ The SHAKE algorithm⁶⁴ was applied to all bonds with H, and hydrogen mass repartitioning⁶⁵ was performed to enable use of a 4.0 fs time step. Simulations were run for a total of 2.0 μs each in triplicate. Simulations employed a Langevin thermostat with a collision frequency of 0.1 s^{-1} . For most properties, the replicate

trajectories were concatenated into one data set, and then statistics were collected over the whole set.

Simulation Temperature. To optimize sampling of phase space, simulations were carried out as close to the *in silico* melting temperature (50–50 folded and unfolded state) as possible. Initial temperatures were estimated with OligoCalc,⁶⁶ which calculates *in vitro* melting temperatures, and 25 K was added to encourage *in silico* unfolding. The distribution of folded and unfolded states in the simulation, as assessed by the root-mean-square deviation (RMSD) in position from experimental structures (Figures 2, S5 and S7 for chiOL3-HR simulations), was then improved by adjusting the simulation temperatures to make sure that the RMSDs sampled larger and smaller values. Final temperatures of the simulations are listed in all RMSD plots.

Assessing RNA Stem Folding. To determine if the stem region of the RNA was able to reform, the trajectories were clustered employing the *k*-means algorithm⁶⁷ on non-hydrogenic atoms in the stem residues as implemented in *cpttraj*. The most populated or highest ranked cluster is designated cluster 1, and the second most populated cluster is designated cluster 2. The backbone atoms and residues in the stem region of the NMR structure were overlaid with the representative structure of the most populated cluster (Figure 3). RMSD (Å) values between the most populated loop cluster structures and the corresponding NMR structure were also calculated.

RNA Structure Assessment. Hydrogen bonding was assessed with *cpttraj* with heavy atom–heavy atom (X–Y) distances <4.0 Å and an X–H–Y angle cutoff of 135°. A longer distance of 4.0 Å was included to account for the thermal fluctuations found at higher temperatures.

Structural factors indicative of A-form RNA helical structure⁶⁸ were determined for each stem cluster trajectory such as stem base pair formation, α -displacement from the helical axis, and C3'-endo sugar pucker presence (Figure 4). Canonical A-form RNA exhibits a C3'-endo sugar pucker,

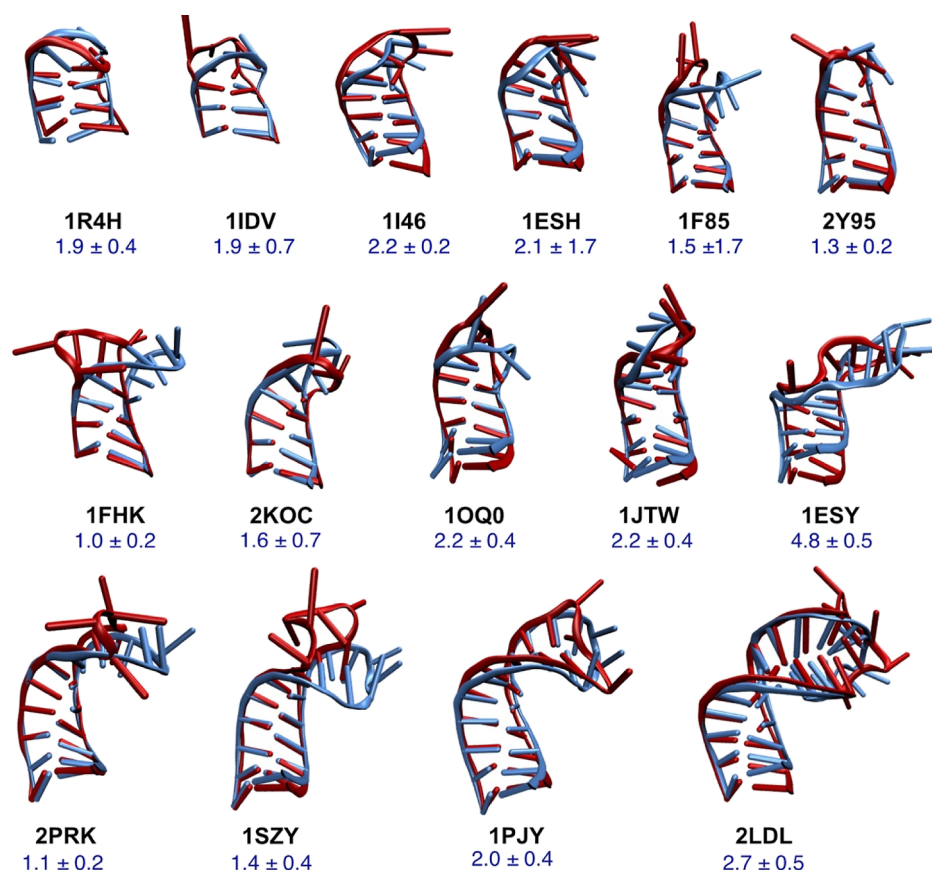


Figure 3. Representative cluster structures with stem residues, chiOL3 and GB, compared to experiment. Overlays of representative structures from the most populated cluster, cluster 1 (red), with the lowest energy NMR structure (blue). Clustering was performed with the *k*-means method on the non-hydrogenic atoms in the stem region. Overlays included all heavy atoms in the stem. The RNA stem-loop structure is labeled with the corresponding PDBID. RMSD (Å) values ± standard deviations between the most populated cluster structures and the experimental structure are shown. Because these are clusters on the stem region of the stem-loop trajectory, the loop structure depicted here is not representative of the trajectory.

which is determined by the pseudorotation phase angle (P) of 0 to 36°. ⁶⁹ X -displacement from the helical axis and other helical parameters were determined in *cpptraj*, which has implemented the general 3DNA method of nucleic acid helical parameter analysis. ⁷⁰

A base pair was determined to be present if all expected hydrogen bonds occurred simultaneously for a given frame. Specifically, canonical G/C pairing required all three hydrogen bonds (G O6 to C N4–H1, G N1–H to C N3, and G N2–H1 to C O2) to be present in a given frame. Likewise, the canonical A/U required both hydrogen bonds (A N6–H1–U O4 and A N1–U N3–H3) to be present in a given frame to be considered base paired. Lastly, G/U wobble pairing was defined as hydrogen bonding of G O6 to U N3–H3 and G N1–H1 to U O2. Other noncanonical base pairs were analyzed relative to the experimentally observed hydrogen bonding pattern. In 1ESY, the A5–A15 base pair was present if A5 N1–H1 was hydrogen bonded to A15 N6.

Assessing RNA Loop Structure. RNA trajectories were also clustered employing the *k*-means algorithm ⁶⁷ on non-hydrogenic atoms of the loop and the closing base pair residues. The most populated or highest ranked cluster is designated cluster 1 and the second most populated cluster is designated cluster 2. Overlays of representative structures from the most populated loop cluster with the lowest energy NMR structure included all heavy atoms for ease of visualization in

context (Figure 5). RMSD (Å) values between the most populated loop cluster structures and the corresponding NMR structures were also calculated.

RESULTS AND DISCUSSION

All 15 stem-loop RNA simulations were begun from NMR structures, and the computational results are largely assessed based upon similarities to these structures. These NMR structures are advantageous because they are determined in aqueous solution and capture the dynamics of the RNA. Modern NMR is laborious as it requires resonances to be assigned despite signal overlap ⁷¹ and may need to be combined with complementary methods such as small-angle X-ray scattering. ⁷² As a last step, NMR structure refinement methods may also incorporate molecular dynamics methods to find low-energy structures that satisfy the experimental restraints. ⁷³ The overall structures examined herein are expected to be reliable. Including a larger data set of RNA structures in this study also ensures that the computational results are less dependent on the quality of any one structure.

Each simulation trajectory was performed for 2.0 μ s in triplicate at the temperature indicated (Figure 2). The temperature was chosen to be as close to the *in silico* melting temperature as possible, as assessed by the RMSD from the NMR structure. This approach is common in early peptide folding studies. ⁷⁴ In some of the simulations, such as 1R4H,

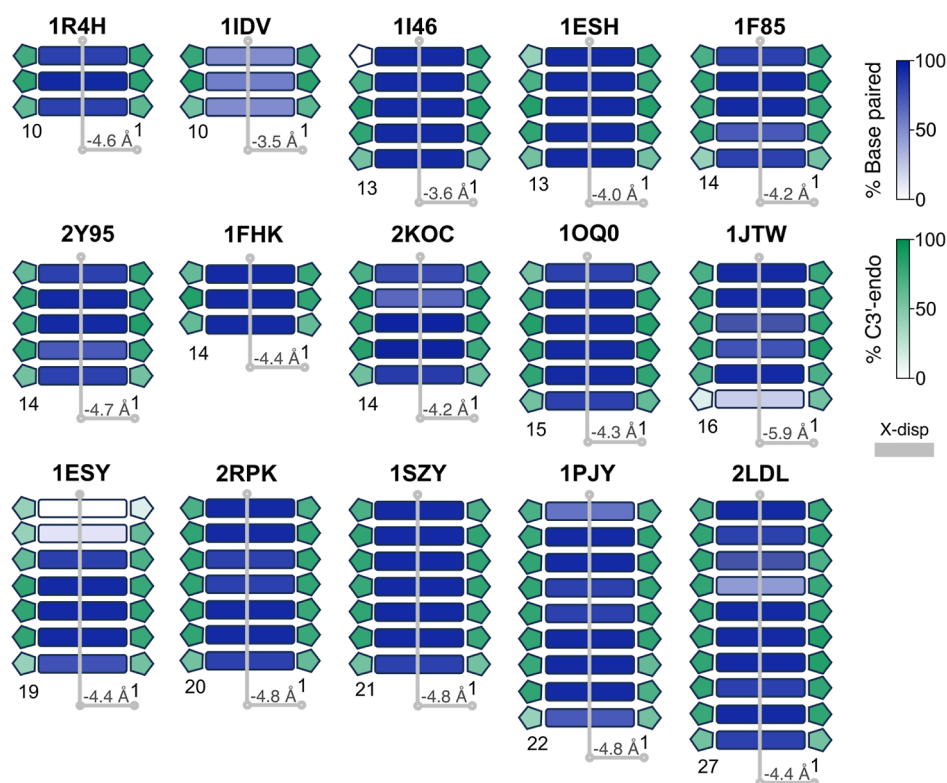


Figure 4. Structural parameters of base-paired regions of highest-ranked stem cluster trajectory with chiOL3 and GB. Pentagons represent sugars and indicate the percentage of the cluster trajectory exhibiting a C3'-endo sugar pucker. Rectangles between stem residues indicate the percentage of the cluster trajectory that a base pair was present. X-displacement from the helical axis is shown as a gray line through the stem region, and the average displacement value in the *x*-direction is also shown in Å. Residue numbers shown match Figure 3 orientations.

1ESH, and 1F85, there is a clear oscillation between larger RMSD values (>5 Å) and lower values (<3 Å). In some of the larger RNA molecules, 1SZY and 1PJY, sampling of any RMSD value that was larger, followed by sampling of a smaller value in the shape of a peak-like line, was considered reasonable. As the RNA molecules get larger, it is difficult to find a temperature at which they will completely unfold but then still be able to refold on the time scale of the simulation. The melting temperature of RNA *in vitro* is dependent upon a number of variables including base-pair content, stability of the loop, and overhang base–base stacking interactions.^{75–78}

Under ideal sampling conditions, the RNA stem-loop would unfold and then refold at the same temperature. While it was not always possible to have an even balance between the two states or to even capture a two-state model, most simulations at least sampled larger and smaller RMSD values. The scale of the RMSD *y*-axis was broadened on the same scale (0–15 Å) for comparison purposes between simulations but enlarged even further for the larger RNA molecules, 1PJY and 2LDL, both of which are greater than 20 nucleotides. While some simulations such as 2RPK did not exhibit large RMSD deviations from the NMR structure, there were some in the 5–6 Å range that subsequently decreased to less than 2 Å. 1I46 showed less tendency to sample high RMSD values, so the thermostat temperature was raised to 335 K; in that simulation, the RMSD rose to >10 Å, returning to a well-folded stem after ~ 200 ns (Figure S6A). Raising the temperature to 360 K resulted in irreversible unfolding, as expected (Figure S6B). Simulating a single temperature is not an ideal choice, as the true melting temperature is not known. Future replica exchange studies could more thoroughly

determine full melting curves for all 15 stem-loop structures. While not all of the stem-loop structures completely unfold to the fully extended structures, they do deviate from the canonical structure as indicated by a higher RMSD and generally adopt stable helical conformations, as discussed below.

We show that in an implicit solvent environment, some RNA stem-loops are reasonably robust to temperature-induced distortions, transiently sampling high RMSD values before returning to a native stem structure in excellent agreement with the NMR-based models. This is achieved by a combination of GB-Neck2nuc⁵⁷ and the chiOL3 force field.²⁶ When the simulations are clustered on the non-hydrogenic atoms of the stem residues (Figure 3), the RMSD between the most populated cluster structures and the NMR structure is usually less than 2.2 Å with small variations (± 1 Å). Of the 15 systems shown here, 13 fit this criterion.

For systems consisting of a small number of exclusively G/C base pairs such as 1R4H, 1IDV, and 1FHK, the low RMSD on the stem region heavy atoms is not surprising. Refolding of stem-loops with as many as five base pairs such as 2Y95 and 2KOC may be unanticipated but is observed in these implicit solvent simulations. Impressively, 1F85 contains two non-canonical G/U base pairs, and the RMSD of the most populated cluster on the stem residues is one of the lowest values (1.5 Å). Even 2RPK and 1SZY, both containing 7 base pairs, refold the stem region in simulations. Most remarkable, however, the large 9 canonical base paired stem of 1PJY refolds back to a structure that closely resembles the stem region of the original hairpin NMR structure.

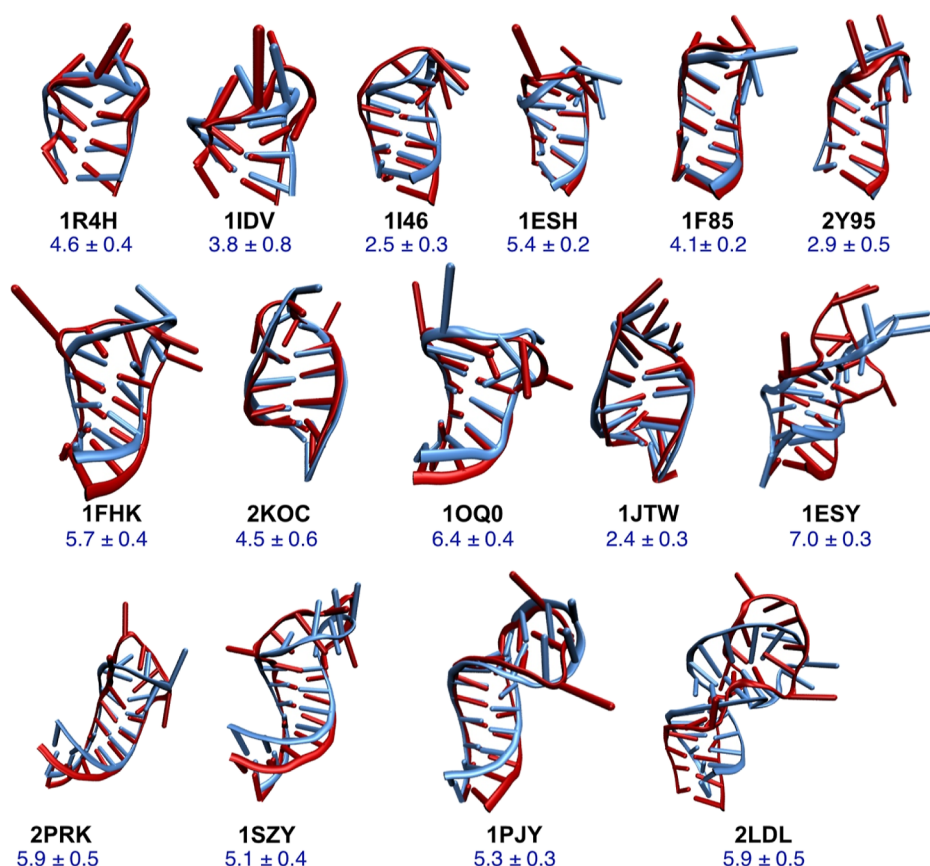


Figure 5. Comparison of structures from clusters on the loop and closing first base pair residues with chiOL3 and GB simulations relative to the experimental structure. Overlays of representative structures from the most populated cluster (cluster 1, red) with the lowest energy NMR structure or X-ray crystal structure (blue). Clustering was performed with the *k*-means method on the non-hydrogenic atoms in the loop and the closing first base pair. Overlays included all heavy atoms. The RNA stem-loop structure is labeled with the corresponding PDBID. RMSD (Å) values $\pm$ standard deviations between the cluster 1 ensemble of structures and the experimental structure are shown.

Structural parameters for the ensemble of structures in cluster 1 were assessed, and average values were calculated (Figure 4 and Table S3). Most base pairs are present for the majority of the cluster simulation time, with % base pairing occurring in greater than 70–80% of the structures. In 2Y95, G/C base pairs were formed in 72–97% of the cluster structures, and every single base pair was reformed. In systems such as 2KOC or 2RPK with one or two A/U base pairs, A/U base pairs were formed in greater than 85% of the structures. Even in 1PJY, a relatively large RNA with 9 base pairs, it was only the terminal base pair (72%) and the closing base pair of the loop (60%) that exhibited a reduction in the frequency of base pairing, with all others present in 85–94% of structures. Notably, IDV exhibited reduced base pairing with G/C base pairs formed in only ~50% of the cluster structures. This is because the G/C base pairs get misaligned when refolding, with G1 base pairing with C9 in 38% of the structures. Interestingly, 1R4H has the identical sequence as 1IDV, containing three G/C base pairs in the stem but with a different loop, and yet it base pairs, reading up from the first to the third base pairs of the stem in 83, 93, and 87% of the cluster stem structures, respectively. Since these two tetraloops only differ between positions 5 and 6 of the loop, the loop structure and helix formation may be correlated. This phenomenon has been observed in other tetraloop structures with the same helix sequence.^{79–81}

Two of the larger systems only partially reform the stem structure, perhaps due to the presence of nonstandard base pairing. For instance, HIV-1 stem-loop 2 (1ESY) contains a bulge, with an extrahelical A15. In the NMR structure, A15 is part of an A5–U14–A15 base triple in which the U14 folds back to base pair through the O2 position to the A5/A15 base pair.⁵² In order to form this structure, a kink in the backbone is required. In the most populated cluster of structures on the RNA stem residues (cluster 1), the first four canonical base pairs from the end (1–19 through 4–16) are able to form (Figure 3 and 4). The A5–A15 base pair is still present (83%), as was in the NMR structure with one hydrogen bond between A5 N1 and A15 N6. However, in the cluster 1 structures, U14 rotates into solution and the U14 O2 does not form a hydrogen bond with A15 N6 to complete the base triple. This resulted in an RMSD value of 4.8 Å between the highest ranked cluster on the stem region and the NMR structure.

A similar but relatively moderate difference from the experimental structure (RMSD of 2.7 Å) was observed with the 28-residue HIV-1 exon splicing silencer 3 (2LDL), which is a larger stem-loop structure with 10 total base pairs capped by a heptaloop sequence. This structure is folded relatively well but contains a nonstandard base pair A7–C21 in the middle of the stem region. In A-form RNA, a standard base pair exhibits a buckle of -0.1° and a propeller twist of -11° ;⁷⁰ however, the A7–C21 base pair in the NMR structure has a more pronounced buckle (10°), and the bases are twisted relative

to each other (propeller -30°). In simulations with chiOL3, these values are moved away from those in the experiments toward the standard A-form helical parameters, resulting in values of buckle -0.6° and propeller -11° . As a result, A7–C21 acts like a hinge, such that when either the bottom six base pairs or the three base pairs closest to the loop align with the NMR structure (RMSD 1.6 or 1.4 Å), but all together the RMSD appears higher (2.7 Å).

Helical formation is dependent on base pairs forming through hydrogen bonding. Structures in stem cluster 1 were examined for the presence of base pairing (Figure 4 and Table S3). A base pair was considered present if all the hydrogen bonds required for the base pair were present simultaneously in a given structure. Due to previously mentioned structural issues in 1ESY and 1IDV, the correct base pairs did not always form. Overall, however, most base pairs were formed in the most populated cluster on the stem residues, and base pairing was present in the majority of the cluster structures with generally 70–99% presence, as indicated by the dark blue color in Figure 4.

The ensemble of structures clustered on the stem simulated with chiOL3 was assessed for RNA classic structural parameters. Sugar pucker orientation was determined based upon the pseudorotation phase angle, with C3'-endo defined as $18 \pm 18^\circ$ and C2'-endo as $162 \pm 18^\circ$.⁶⁹ When clustering on the stem residues, C3'-endo sugar puckers (Figure 4 and Table S4, green pentagons) were observed in most stem residues (70–90%). Residues on the ends of the helix, either the base of the stem or near the loop, exhibited a C3'-endo sugar pucker in only 50–60% of the cluster 1 structures. This seems typical for terminal residues, which are known to have more structural freedom and become frayed in experiments. Even though in the stem region, some residues might exhibit C3'-endo sugar puckers in the majority of the cluster 1 structures, it also seems reasonable that due to the dynamic nature of the sugar pucker sampling, neighboring conformations such as C4'-exo or C2'-exo may occur, and this is indeed the case. In a few exceptional cases, the major sugar pucker present in the cluster 1 helical structures was C2'-endo. In the case of 1ESY, this is probably due to the disruption of the helix already noted. While in 1JTW, the terminal base pair G1/A16 is not paired in the NMR structure resulting in a flexible A16 sugar pucker.

C3'-endo sugar puckers and a significant α -displacement from the helical axis (-4.4 Å) are hallmark features of A-form helical RNA structure.^{68,82} The average α -displacement from the helical axis is shown for each cluster 1 helix (Figure 4, gray line and value). Values generally range from -3.5 to -4.8 Å. Analysis of the original NMR structures indicates that the α -displacement in the helical region can vary from -1.8 to -5.4 Å. In the case of 1JTW, the value is -5.9 Å. While this value is larger than the NMR structure (-4.1 Å), it is not ~ 0 Å as is found in B-DNA,⁸³ and this is qualitatively in line with A-form RNA.

The ability of our simulations in implicit solvent to form the A-form RNA is consistent with results found in explicit solvent simulations. In explicit solvent simulations with TIP3P, Spomer and co-workers assessed duplex RNA structural parameters²⁶ and found that both ff99bsc0^{23–25} and chiOL3 generally retain A-form RNA structure, with α -displacement values -4.85 ± 1.60 and -4.95 Å, respectively. In these simulations, the sugar puckers are generally C3'-endo (P is 19.3 parmbc0 and 17.4 chiOL3). The ability of implicit solvent simulations to

correctly model RNA across such a range of RNA stems is promising for RNA simulations in the long term.

While the RNA stems of the structures studied herein can transiently sample distorted structures and then reform A-form RNA with the chiOL3 force field in an implicit solvent environment, modeling the loop structures accurately remains especially challenging. When the simulations are clustered employing non-hydrogenic atoms of the loop and closing base pair residues (Figure 5), the RMSD values from the NMR structure were relatively large, with most in the 4–7 Å range. This is not a surprising result, as in explicit solvent with a buffer of ions, it is quite difficult to accurately model RNA loop structures.^{84–86} Notably, two loop structures (1I46 and 1JTW) generally exhibited smaller RMSD values of 2.5 and 2.4 Å, respectively, indicating that the loop and first base pair residues are qualitatively close to the experimental structure. In both structures, the bases are on the correct side of the loop, and there are no appreciable distortions in the backbones.

We next explored several adjustments of the energy function. In explicit solvent, we had previously found a force field modification that minimizes repulsion in the CH $\cdots$ O interactions to be necessary for proper loop structure formation.⁵⁹ Additionally, it is well established that burial of nonpolar groups is a major driving force for protein stability, and as such, it was expected that some of the base stacking interactions may not be strong enough in a GB environment since it models only the polar aspect of solvation. To better model the base stacking interaction with GB-Neck2nuc,⁵⁷ umbrella sampling of a model system was performed (see Supporting Information), in which an adenine is stacked over a G/C base pair (Figure S2). Analysis of the energy profile indicated that the stacking energy measured in GB-Neck2nuc was ~ 2.0 kcal/mol, significantly weaker than the same system modeled in explicit solvent, in which the stacking energy was measured to be ~ 6.5 kcal/mol (Figure S2). We found that the agreement with explicit solvent could be greatly improved in GB simulations by a 60% increase in the Lennard-Jones well depth between pairs of heavy atoms in bases (Figure S2 and Table S1). To capture this interaction in GB-Neck2nuc simulations, while still being a bit conservative, a 30% increase in pairwise Lennard-Jones interaction energy for RNA base–base heavy atoms was implemented for the RNA stem-loops (Supporting Information). This nonpolar modification to the GB-Neck2nuc implicit solvent model will be referred to hereafter as GB_{np}.

Since many of the loops did not form, a subset of six of the 15 RNA structures was chosen for further study with the chiOL3-HR force field and GB_{np} implicit solvent model. These RNA molecules were chosen based upon their distinct structural features and included four tetraloops (1F85, 2Y95, 1JTW, and 2KOC) as well as two larger stem-loops (2LDL and 1FHK). Simulations were carried out at the same temperatures as their original chiOL3 simulations and otherwise comparable MD simulation conditions. Plots of the RMSD values from the NMR structure over the 2.0 μ s simulation time, performed in triplicate, indicated stable simulations with variable unfolding and folding events (Supporting Information). Structures were clustered on the loop and first base pair and compared with the NMR loop structures.

Of the six RNA stem-loop structures simulated with chiOL3-HR and GB_{np}, the model remarkably improved both 1JTW and 1FHK structures (Figure 6). With chiOL3, the RMSD between

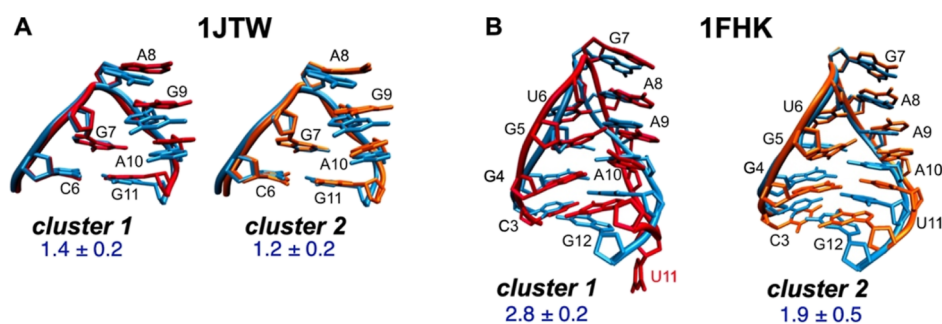


Figure 6. Loop structures with chiOL3-HR and GB_{np}. Overlays of representative structures from the most populated cluster on the loop, cluster 1 (red) and cluster 2 (orange) with the lowest energy NMR structure (blue) for two different systems. Structures are shown for only the loop and closing base pair residues. Clustering was performed with the *k*-means method on the non-hydrogenic atoms in the loop and the closing first base pair. Overlays included all heavy atoms. The RNA stem-loop structure is labeled with the corresponding PDBID. RMSD (Å) values $\pm$ standard deviations between the cluster 1 or cluster 2 trajectories and the experimental structure are shown. (A) 1JTW cluster 1 and cluster 2 overlays with the NMR structure. (B) 1FHK cluster 1 and cluster 2 overlays with the NMR structure.

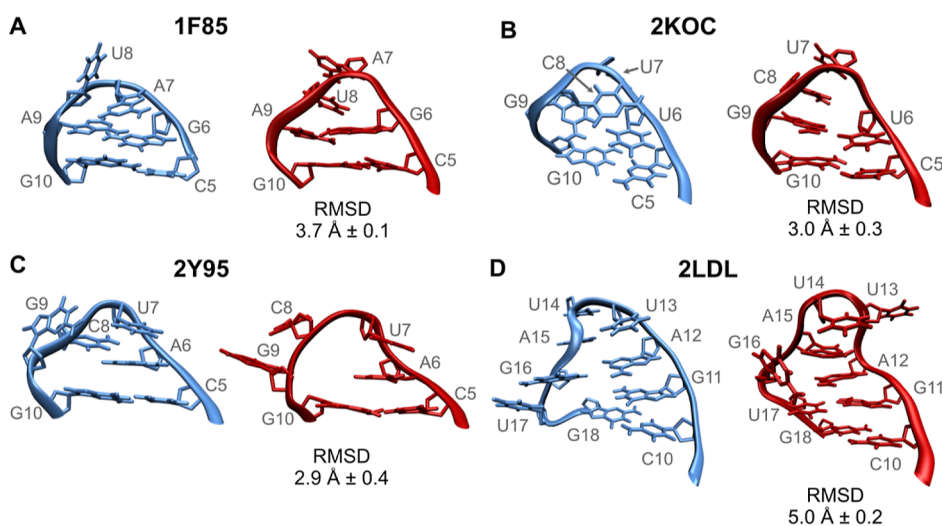


Figure 7. Comparison of loop structures from NMR and those simulated with the chiOL3-HR that still require some improvement. Structures show highest populated cluster (cluster 1) of the loop and closing base pair (red) and the NMR structure (blue) for the corresponding region.

the 1JTW loop structure in the most populated cluster (cluster 1) and the NMR structure was 2.4 ± 0.3 Å. However, with chiOL3-HR, it improved to 1.4 ± 0.2 Å (Figure 6A). The 1FHK structure showed an even greater improvement in loop structure with the force field and GB modifications. With the standard RNA force field chiOL3, the RMSD between the NMR structure and the cluster 1 loop structure is 5.7 ± 0.4 Å, which by all accounts is not accurate (Figure 5). However, with chiOL3-HR, the 1FHK cluster 1 loop (29%) has an RMSD of 2.8 ± 0.2 Å from the NMR structure (Figure 6B), and cluster 2, populated 26%, has an even greater improvement with an RMSD of 1.9 ± 0.5 Å.

Characteristic structural features of 1JTW are preserved when compared with the NMR structure (Figure 6A and Table S5). In 1JTW, the second most populated structure, present for 45% of the simulation (cluster 2), also very closely resembled the NMR structure with an RMSD of 1.2 ± 0.2 Å, and so both were analyzed further. Since both clusters represent over 90% of the total simulation time, it is possible that the loop was not appreciably disrupted at the chiOL3 simulated temperature, but it does indicate that in implicit solvent, the RNA hairpin loop is stable. The 1JTW structure is capped by a GAGA tetraloop closed by a C/G base pair, in which the loop turns between G7 and A8, followed by a 3'-

stack of A8/G9/A10 (100% of all NMR structures). Base stacking in the loop is observed in a similar proportion in both cluster 1 and cluster 2 between the A8 6-membered ring and the G9 5-membered ring (42, 48%, respectively) and more so between the G9 6-membered ring and A10 five-membered ring (70%, 80%, respectively). This 3'-stack is facilitated by a turn of the backbone quantified by an A8 α -torsion ($O3'-P-O5'-C5'$) that is primarily *trans* (*t*) (84.6%). Cluster 1 and cluster 2 retain this distribution (64% *t*, 85% *t*, respectively). The tight turn in the GAGA loop backbone is pinned by a G7-A10 sheared pair in the NMR structure, in which the 2-amino group of G7 hydrogen bonds both to A10 N7 (62%) and to the A10 O2P (46%). Interestingly, both clusters favor these hydrogen bonds a bit more than the experimental structure, with occupancies increased by as much as 20–30 percentage points.

Simulations with chiOL3-HR of a second stem-loop, 1FHK, also closely replicated the experimental structure. This stem-loop contains a distinct octal loop exhibiting a U-turn motif and a stair-like structure with consecutive bases stacking one on another (Figure 6B). In the NMR structure, there is a U-turn motif between the U6 and G7. This is held together by hydrogen bonds between U6 N3–H3 and the A9 backbone O2P (93% of NMR structures) and U6 2'-OH and A8 N7 (also

93% of NMR structures). In the most populated cluster on the loop residues, cluster 1, these hydrogen bonds are present in the majority of the structures (Table S5). They are still present in cluster 2, but to a lesser extent. Base stacking interactions are also preserved between the NMR structure and the cluster structures, with cluster 2 structures showing better base stacking.

In the remaining four stem-loop structures, the RMSD between the loop and first base pair region in the most populated cluster and the NMR structure (Figure 7) is improved with chiOL3-HR from those simulated with chiOL3. Some improvements are incremental such as with 2Y95, where the RMSD values from the NMR structure are essentially the same with both conditions. However, in the other three systems, RMSD values decreased by 0.4–3.0 Å. The variation of the cluster 1 structures or standard deviation in the RMSD values was also smaller, indicating more uniformity in the ensemble of structures sampled.

Still, key structural attributes were not conserved in simulations of these systems (Figure 7 and Table S6). In all four stem-loop structures, one of the base flips in the loop (1F85-A7, 2KOC-C8, 2Y95-C8, and 2LDL-U14). In 2Y95, G9 also flips, but it was dynamic in the NMR structure. In another example, in the NMR structures of 1F85, 2KOC, and 2Y95, there are one or more C2'-endo sugar puckers in the loop region (U8, U7 and C8, and U7, respectively). In the 2LDL experimental structure, A15 has a C1'-exo sugar pucker, and in 1F85, A7 is a C3'-exo sugar pucker, which are both directly adjacent to C2'-endo. However, in the simulations, even with the chiOL3-HR and GB_{np}, the most populated cluster on the loop region tends toward C3'-endo. A common feature also observed is that the α -torsion (O3'-P-O5'-C5') of the same residue containing a C2'-endo-like sugar pucker is also noncanonical. In A-form RNA, the α -torsion tends to be gauche (g^-) with an average value of 292°. ^{68,87} In the NMR structures, when the sugar pucker is C2'-endo, it does sample the g^- conformation, but more often it is instead displaying c , g^+ , a^- , or t conformations. In the four systems with unusual sugar puckers in the loop, the simulations do not match the experimental α -torsion values. This may be due to the inherent flexibility of the backbone. All of these structures are NMR structures and some backbone positions are difficult to resolve by NMR. ^{88,89} Interestingly, in 1JTW and 1FHK, when the loop turns, the sugar pucker at the turn (A8 and G7, respectively) remains C3'-endo, and the α -torsion generally matches between experiment and simulation preferring a t conformation. This indicates that the chiOL3-HR force field with the GB_{np} can successfully model loop structures with C3'-endo sugar puckers but not C2'-endo. It cannot be determined whether this is due to sugar pucker conformation misalignment in the force field or whether it is instead due to other intermolecular interactions that are not strong enough in the force field. For instance, perhaps the sugar pucker is flexible and the base is not held in tightly by hydrogen bonds, or perhaps the sugar pucker favors C3'-endo and the bias in the sugar pucker prevents the base from compensating and flips.

CONCLUSIONS

Atomistic simulations of RNA systems have lagged behind the progress for protein simulations, in part due to the slow time scales of RNA motion that make it difficult to obtain precise simulation ensembles to compare to experimental measurements. The slow convergence also hinders improvement in

RNA force fields; RNA models have additional challenges when compared to proteins due to the numerous rotatable bonds in the nucleic acid sugar-phosphate backbone, the high charge density, and the overall complexity of the RNA structure. A key step in RNA model improvement would be large-scale tests of the force field performance on a variety of RNA systems with diverse structural motifs. Here, we present a step in that direction, carrying out simulations on a number of RNA systems and exploring trends in the features that are well reproduced in MD, along with aspects that challenge the current RNA force field and solvent models. The study is enabled by the use of a generalized Born implicit solvent model developed for DNA and RNA, which speeds the sampling of alternate conformations.

Analysis of the simulation data indicates that the chiOL3 RNA model combined with GB-Neck2nuc solvation performs remarkably well on the RNA stem regions, reproducing expected structure features. We had previously reported that subtle force field details such as treatment of CH...O interactions in RNA are important for accurately modeling RNA in explicit solvent, and these changes improve performance in implicit solvent as well. Reproduction of loop structures was less satisfying, and the results suggested that base stacking is too weak in the implicit solvent model. We trained a simple, fast correction to the GB model that led to improved stacking and notably better reproduction of experimental structures for some of the RNA systems. It is important to note that base stacking for all residues, not just the purines, was changed in the GB_{np} correction. Although the GB model likely needs further improvement, we recommend that future implicit solvent simulations of RNA include the base stacking and CH...O modifications that we tested here.

When considered together, this is one of the first systematic analyses of RNA loop structures in implicit solvent over a range of structures, and the results can inform improvements to future GB models, as we showed here for the base stacking adjustment. In addition, the RNA model systems that we curated here may provide a useful benchmark set for testing other RNA force fields. Continued improvements to computer hardware and simulation methods will permit the evaluation of the benchmark test via converged ensembles in explicit water, providing additional guidance for the improvement of RNA simulation models.

ASSOCIATED CONTENT

Data Availability Statement

All starting structures were downloaded from the RCSB protein data bank (<https://www.rcsb.org/>); PDBIDs listed in Methods. Please see <https://ambermd.org> for licensing and distribution of Amber and AmberTools software suites. VMD molecular visualization software ⁹⁰ was employed to visualize and render RNA structures. Figures employed ChemDraw, Matplotlib, Affinity Photo 2, Microsoft Excel, and *cpptraj*. MD trajectories are available upon request to the authors. Topology files, input files, and a parmed script used to generate the results in this work are provided here <https://github.com/naganlab/rnaGBinputs>.

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.4c00756>.

Description of training data for the implicit solvent base stacking calculations, including changes to the Lennard-

Jones parameters; RMSD plots for chiOL3 and chiOL3-HR with GB_{np} simulations; and RNA structural analysis for base pairing and sugar puckers shown in Figures 4–6 (PDF)

AUTHOR INFORMATION

Corresponding Author

Maria C. Nagan – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States; orcid.org/0000-0003-2678-6825; Phone: 631-632-5793; Email: maria.nagan@stonybrook.edu

Authors

Jason T. Linzer – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States; orcid.org/0009-0008-2739-5432

Ethan Aminov – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States; orcid.org/0009-0003-2787-1869

Aalim S. Abdullah – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States

Colleen E. Kirkup – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States

Rebeca I. Diaz Ventura – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States; orcid.org/0009-0006-0207-3053

Vinay R. Bijoor – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States; orcid.org/0009-0009-5837-376X

Jiyun Jung – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States

Sophie Huang – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States; orcid.org/0009-0003-4838-9916

Chi Gee Tse – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States

Emily Alvarez Toucet – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States

Hugo P. Onghai – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States; orcid.org/0009-0006-7779-0380

Arghya P. Ghosh – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States; orcid.org/0000-0002-0780-9780

Alex C. Grodzki – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States; orcid.org/0009-0007-7041-9041

Emilee R. Haines – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States

Aditya S. Iyer – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States

Mark K. Khalil – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States; orcid.org/0009-0003-4380-229X

Alexander P. Leong – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States; orcid.org/0009-0005-2693-7496

Michael A. Neuhaus – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States

Joseph Park – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States

Asir Shahid – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States; orcid.org/0009-0009-6232-387X

Matthew Xie – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States; orcid.org/0009-0002-8874-2959

Jan M. Ziembicki – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States; orcid.org/0009-0009-1569-0343

Carlos Simmerling – Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States; Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, New York 11794, United States; orcid.org/0000-0002-7252-4730

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jcim.4c00756>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Computational resources were provided in part by the MERCURY consortium (<http://mercuryconsortium.org>) under NSF grants CHE-0521063, CHE-0849677, CHE 1229354, and CHE-1662030. E.R.H., E.A.T., A.S. Abdullah, and M.C. Nagan were supported by the NSF REU program, CHE-2050541. H.P.O. was supported by the Simons Summer Research Program. C.S. was supported by NIH GM105107 and funding from the Laufer family.

REFERENCES

- (1) Leontis, N. B.; Westhof, E. Analysis of RNA Motifs. *Curr. Opin. Struct. Biol.* **2003**, *13* (3), 300–308.
- (2) Mathews, D. H. Revolutions in RNA Secondary Structure Prediction. *J. Mol. Biol.* **2006**, *359* (3), 526–532.
- (3) Ferre-D'Amaré, A. R.; Doudna, J. A. RNA Folds: Insights from Recent Crystal Structures. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 57–73.
- (4) Butcher, S. E.; Pyle, A. M. The Molecular Interactions That Stabilize RNA Tertiary Structure: RNA Motifs, Patterns, and Networks. *Acc. Chem. Res.* **2011**, *44* (12), 1302–1311.
- (5) Vicens, Q.; Kieft, J. S. Thoughts on How to Think (and Talk) About RNA Structure. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119* (17), No. e2112677119.
- (6) Das, R.; Baker, D. Automated De Novo Prediction of Native-Like RNA Tertiary Structures. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (37), 14664–14669.
- (7) Miao, Z.; Adamiak, R. W.; Antczak, M.; Boniecki, M. J.; Bujnicki, J.; Chen, S.-J.; Cheng, C. Y.; Cheng, Y.; Chou, F.-C.; Das, R.; Dokholyan, N. V.; Ding, F.; Geniesse, C.; Jiang, Y.; Joshi, A.; Krokhotin, A.; Magnus, M.; Mailhot, O.; Major, F.; Mann, T. H.; Piątkowski, P.; Pluta, R.; Popenda, M.; Sarzynska, J.; Sun, L.; Szachniuk, M.; Tian, S.; Wang, J.; Wang, J.; Watkins, A. M.; Wiedemann, J.; Xiao, Y.; Xu, X.; Yesselman, J. D.; Zhang, D.; Zhang, Y.; Zhang, Z.; Zhao, C.; Zhao, P.; Zhou, Y.; Zok, T.; Żyła, A.; Ren, A.; Batey, R. T.; Golden, B. L.; Huang, L.; Lilley, D. M.; Liu, Y.; Patel, D. J.; Westhof, E. RNA-Puzzles Round IV: 3D Structure Predictions of Four Ribozymes and Two Aptamers. *RNA* **2020**, *26* (8), 982–995.
- (8) Xu, X.; Chen, S. J. Hierarchical Assembly of RNA Three-Dimensional Structures Based on Loop Templates. *J. Phys. Chem. B* **2018**, *122* (21), 5327–5335.
- (9) Woese, C. R.; Winker, S.; Gutell, R. R. Architecture of Ribosomal RNA: Constraints on the Sequence of Tetra-Loops. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87* (21), 8467–8471.

- (10) Bevilacqua, P. C.; Blose, J. M. Structures, Kinetics, Thermodynamics, and Biological Functions of RNA Hairpins. *Annu. Rev. Phys. Chem.* **2008**, *59*, 79–103.
- (11) Goodall, G. J.; Wickramasinghe, V. O. RNA in Cancer. *Nat. Rev. Cancer* **2021**, *21* (1), 22–36.
- (12) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. *Science* **2000**, *289* (5481), 905–920.
- (13) Thapar, R.; Denmon, A. P.; Nikonowicz, E. P. Recognition Modes of RNA Tetraloops and Tetraloop-Like Motifs by RNA-Binding Proteins. *Wiley Interdiscip. Rev.: RNA* **2014**, *5* (1), 49–67.
- (14) Richardson, K. E.; Adams, M. S.; Kirkpatrick, C. C.; Gohara, D. W.; Znosko, B. M. Identification and Characterization of New RNA Tetraloop Sequence Families. *Biochemistry* **2019**, *58* (48), 4809–4820.
- (15) Auffinger, P. Molecular Dynamics Simulations of RNA Systems. *Handbook of RNA Biochemistry*; John Wiley & Sons, 2014; pp 687–718.
- (16) Sponer, J.; Bussi, G.; Krepl, M.; Banas, P.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurecka, P.; Walter, N. G.; Otyepka, M. RNA Structural Dynamics as Captured by Molecular Simulations: A Comprehensive Overview. *Chem. Rev.* **2018**, *118* (8), 4177–4338.
- (17) McCrate, N. E.; Varner, M. E.; Kim, K. I.; Nagan, M. C. Molecular Dynamics Simulations of Human tRNA^{Lys}₃^{UUU}: The Role of Modified Bases in mRNA Recognition. *Nucleic Acids Res.* **2006**, *34* (19), 5361–5368.
- (18) McCarthy, E.; Ekesan, S.; Giese, T. J.; Wilson, T. J.; Deng, J.; Huang, L.; Lilley, D. M. J.; York, D. M. Catalytic Mechanism and Ph Dependence of a Methyltransferase Ribozyme (MTR1) from Computational Enzymology. *Nucleic Acids Res.* **2023**, *51* (9), 4508–4518.
- (19) Nagan, M. C.; Beuning, P.; Musier-Forsyth, K.; Cramer, C. J. Importance of Discriminator Base Stacking Interactions: Molecular Dynamics Analysis of A73 Microhelix(Ala) Variants. *Nucleic Acids Res.* **2000**, *28* (13), 2527–2534.
- (20) Špačková, N.; Sponer, J. Molecular Dynamics Simulations of Sarcin-Ricin rRNA Motif. *Nucleic Acids Res.* **2006**, *34* (2), 697–708.
- (21) Condon, D. E.; Kennedy, S. D.; Mort, B. C.; Kierzek, R.; Yildirim, I.; Turner, D. H. Stacking in RNA: NMR of Four Tetramers Benchmark Molecular Dynamics. *J. Chem. Theory Comput.* **2015**, *11* (6), 2729–2742.
- (22) Mráziková, K.; Mlynsky, V.; Kuhrova, P.; Pokorna, P.; Kruse, H.; Krepl, M.; Otyepka, M.; Banas, P.; Sponer, J. UUCG RNA Tetraloop as a Formidable Force-Field Challenge for MD Simulations. *J. Chem. Theory Comput.* **2020**, *16* (12), 7601–7617.
- (23) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. A Modified Version of the Cornell et al. Force Field with Improved Sugar Pucker Phases and Helical Repeat. *J. Biomol. Struct. Dyn.* **1999**, *16* (4), 845–862.
- (24) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179–5197.
- (25) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Loughton, C. A.; Orozco, M. Refinement of the Amber Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.* **2007**, *92* (11), 3817–3829.
- (26) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E.; Jurecka, P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7* (9), 2886–2902.
- (27) Oostenbrink, C.; Soares, T. A.; van der Vegt, N. F.; van Gunsteren, W. F. Validation of the 53A6 GROMOS Force Field. *Eur. Biophys. J.* **2005**, *34* (4), 273–284.
- (28) Soares, T. A.; Hunenberger, P. H.; Kastenhof, M. A.; Krautler, V.; Lenz, T.; Lins, R. D.; Oostenbrink, C.; van Gunsteren, W. F. An Improved Nucleic Acid Parameter Set for the GROMOS Force Field. *J. Comput. Chem.* **2005**, *26* (7), 725–737.
- (29) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. Definition and Testing of the GROMOS Force-Field Versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40* (7), 843–856.
- (30) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell, A. D. 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.* **2011**, *32* (9), 1929–1943.
- (31) Foloppe, N.; MacKerell, J. All-atom empirical force field for nucleic acids: I. Parameter Optimization Based on small Molecule and Condensed Phase Macromolecular Target Data. *J. Comput. Chem.* **2000**, *21* (2), 86–104.
- (32) MacKerell, A. D.; Banavali, N. K. All-Atom Empirical Force Field for Nucleic Acids: II. Application to Molecular Dynamics Simulations of DNA and RNA in Solution. *J. Comput. Chem.* **2000**, *21* (2), 105–120.
- (33) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65* (3), 712–725.
- (34) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713.
- (35) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Miguës, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16* (1), 528–552.
- (36) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Cheatham, T. E. Highly Sampled Tetranucleotide and Tetraloop Motifs Enable Evaluation of Common RNA Force Fields. *RNA* **2015**, *21* (9), 1578–1590.
- (37) Kůhrová, P.; Best, R. B.; Bottaro, S.; Bussi, G.; Šponer, J.; Otyepka, M.; Banáš, P. Computer Folding of RNA Tetraloops: Identification of Key Force Field Deficiencies. *J. Chem. Theory Comput.* **2016**, *12* (9), 4534–4548.
- (38) Lam, K.; Kasavajhala, K.; Gunasekera, S.; Simmerling, C. Accelerating the Ensemble Convergence of RNA Hairpin Simulations with a Replica Exchange Structure Reservoir. *J. Chem. Theory Comput.* **2022**, *18* (6), 3930–3947.
- (39) Chen, J.; Liu, H.; Cui, X.; Li, Z.; Chen, H.-F. RNA-Specific Force Field Optimization with CMAP and Reweighting. *J. Chem. Inf. Model.* **2022**, *62* (2), 372–385.
- (40) Michel, F.; Westhof, E. Modelling of the Three-Dimensional Architecture of Group I Catalytic Introns Based on Comparative Sequence Analysis. *J. Mol. Biol.* **1990**, *216* (3), 585–610.
- (41) Tuerk, C.; Gauss, P.; Thermes, C.; Groebe, D. R.; Gayle, M.; Guild, N.; Stormo, G.; d'Aubenton-Carafa, Y.; Uhlenbeck, O. C.; Tinoco, I.; et al. CUUCG Hairpins: Extraordinarily Stable RNA Secondary Structures Associated with Various Biochemical Processes. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85* (5), 1364–1368.
- (42) Duszczek, M. M.; Wutz, A.; Rybin, V.; Sattler, M. The Xist RNA A-Repeat Comprises a Novel AUCG Tetraloop Fold and a Platform for Multimerization. *RNA* **2011**, *17* (11), 1973–1982.
- (43) Tolokh, I. S.; Thomas, D. G.; Onufriev, A. V. Explicit Ions/Implicit Water Generalized Born Model for Nucleic Acids. *J. Chem. Phys.* **2018**, *148* (19), 195101.
- (44) Rijnbrand, R.; Thivyanathan, V.; Kaluarachchi, K.; Lemon, S. M.; Gorenstein, D. G. Mutational and Structural Analysis of Stem-Loop IIIc of the Hepatitis C Virus and GB Virus B Internal Ribosome Entry Sites. *J. Mol. Biol.* **2004**, *343* (4), 805–817.
- (45) Kim, C.-H.; Tinoco, I. Structural and Thermodynamic Studies on Mutant RNA Motifs that Impair the Specificity Between a Viral Replicase and Its Promoter. *J. Mol. Biol.* **2001**, *307* (3), 827–839.

- (46) Kim, C.-H.; Kao, C. C.; Tinoco, I. RNA Motifs That Determine Specificity between a Viral Replicase and Its Promoter. *Nat. Struct. Biol.* **2000**, *7* (5), 415–423.
- (47) Lukavsky, P. J.; Otto, G. A.; Lancaster, A. M.; Sarnow, P.; Puglisi, J. D. Structures of Two RNA Domains Essential for Hepatitis C Virus Internal Ribosome Entry Site Function. *Nat. Struct. Biol.* **2000**, *7* (12), 1105–1110.
- (48) Morosyuk, S. V.; Cunningham, P. R.; SantaLucia, J. Structure and Function of the Conserved 690 Hairpin in *Escherichia Coli* 16 S Ribosomal RNA. II. NMR Solution Structure. *J. Mol. Biol.* **2001**, *307* (1), 197–211.
- (49) Nozinovic, S.; Fürtig, B.; Jonker, H. R. A.; Richter, C.; Schwalbe, H. High-Resolution NMR Structure of an RNA Model System: The 14-Mer cUUCG Tetraloop Hairpin RNA. *Nucleic Acids Res.* **2010**, *38* (2), 683–694.
- (50) Leeper, T.; Leulliot, N.; Varani, G. The Solution Structure of an Essential Stem–Loop of Human Telomerase RNA. *Nucleic Acids Res.* **2003**, *31* (10), 2614–2621.
- (51) Kerwood, D. J.; Cavaluzzi, M. J.; Borer, P. N. Structure of SL4 RNA from the HIV-1 Packaging Signal. *Biochemistry* **2001**, *40* (48), 14518–14529.
- (52) Amarasinghe, G. K.; De Guzman, R. N.; Turner, R. B.; Summers, M. F. NMR Structure of Stem-Loop SL2 of the HIV-1 Ψ RNA Packaging Signal Reveals a Novel A-U-A Base-Triple Platform 1 Edited by I. Tinoco. *J. Mol. Biol.* **2000**, *299* (1), 145–156.
- (53) Dufour, D.; de la Peña, M.; Gago, S.; Flores, R.; Gallego, J. Structure–Function Analysis of the Ribozymes of *Chrysanthemum Chlorotic Mottle Viroid*: A Loop–Loop Interaction Motif Conserved in Most Natural Hammerheads. *Nucleic Acids Res.* **2009**, *37* (2), 368–381.
- (54) Schweisguth, D. C.; Moore, P. B. On the Conformation of the Anticodon Loops of Initiator and Elongator Methionine tRNAs. *J. Mol. Biol.* **1997**, *267* (3), 505–519.
- (55) Staple, D. W.; Butcher, S. E. Solution Structure of the HIV-1 Frameshift Inducing Stem–Loop RNA. *Nucleic Acids Res.* **2003**, *31* (15), 4326–4331.
- (56) Levengood, J. D.; Rollins, C.; Mishler, C. H. J.; Johnson, C. A.; Miner, G.; Rajan, P.; Znosko, B. M.; Tolbert, B. S. Solution Structure of the HIV-1 Exon Splicing Silencer 3. *J. Mol. Biol.* **2012**, *415* (4), 680–698.
- (57) Nguyen, H.; Pérez, A.; Bermeo, S.; Simmerling, C. Refinement of Generalized Born Implicit Solvation Parameters for Nucleic Acids and Their Complexes with Proteins. *J. Chem. Theory Comput.* **2015**, *11* (8), 3714–3728.
- (58) Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins: Struct., Funct., Bioinf.* **2004**, *55* (2), 383–394.
- (59) Raguet, L. E.; Gunasekera, S. S.; Ventura Diaz, R. I.; Aminov, E.; Linzer, J. T.; Parwana, D.; Wu, Q.; Simmerling, C.; Nagan, M. C. Adjusting the Energy Profile for CH $\cdots$ O Interactions Leads to Improved Stability of RNA Stem-Loop Structures in MD Simulations. **2024**, Chemrxiv-2024-9rbjw
- (60) Huang, H.; Simmerling, C. Fast Pairwise Approximation of Solvent Accessible Surface Area for Implicit Solvent Simulations of Proteins on CPUs and GPUs. *J. Chem. Theory Comput.* **2018**, *14* (11), 5797–5814.
- (61) Case, D. A.; Aktulga, K.; Ben-Shalom, I. Y.; Berryman, J. T.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E.; Cisneros, G. A.; Cruzeiro, V. W. D.; Darden, T. A.; Forouzesh, N.; Giambasu, G.; Giese, T.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, J.; Izadi, S.; Izmailov, S. A.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Machado, M.; Man, V.; Manathunga, M.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; O'Hearn, K. A.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shajan, A.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wang, J.; Wei, H.; Wu, X.; Wu, Y.; Xiong, Y.; Xue, Y.; York, D. M.; Zhao, S.; Zhu, Q.; Kollman, P. A. *Amber 2023*; University of California: San Francisco, 2023.
- (62) Case, D. A.; Belfon, K.; Ben-Shalom, I. Y.; Berryman, J. T.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E.; Cisneros, G. A.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Machado, M.; Man, V.; Manathunga, M.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; O'Hearn, K. A.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shajan, A.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiong, Y.; Xue, Y.; York, D. M.; Zhao, S.; Kollman, P. A. *Amber22*; University of California: San Francisco, 2022.
- (63) Case, D. A.; Aktulga, H. M.; Belfon, K.; Cerutti, D. S.; Cisneros, G. A.; Cruzeiro, V. W. D.; Forouzesh, N.; Giese, T. J.; Götz, A. W.; Gohlke, H.; Izadi, S.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kurtzman, T.; Lee, T.-S.; Li, P.; Liu, J.; Luchko, T.; Luo, R.; Manathunga, M.; Machado, M. R.; Nguyen, H. M.; O'Hearn, K. A.; Onufriev, A. V.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Risheh, A.; Schott-Verdugo, S.; Shajan, A.; Swails, J.; Wang, J.; Wei, H.; Wu, X.; Wu, Y.; Zhang, S.; Zhao, S.; Zhu, Q.; Cheatham, T. E.; Roe, D. R.; Roitberg, A.; Simmerling, C.; York, D. M.; Nagan, M. C.; Merz, K. M. *AmberTools*. *J. Chem. Inf. Model.* **2023**, *63* (20), 6183–6191.
- (64) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327–341.
- (65) Jung, J.; Kasahara, K.; Kobayashi, C.; Oshima, H.; Mori, T.; Sugita, Y. Optimized Hydrogen Mass Repartitioning Scheme Combined with Accurate Temperature/Pressure Evaluations for Thermodynamic and Kinetic Properties of Biological Systems. *J. Chem. Theory Comput.* **2021**, *17* (8), 5312–5321.
- (66) Kibbe, W. A. Oligocalc: An Online Oligonucleotide Properties Calculator. *Nucleic Acids Res.* **2007**, *35*, W43–W46.
- (67) Lloyd, S. Least Squares Quantization in PCM. *IEEE Trans. Inf. Theor.* **1982**, *28* (2), 129–137.
- (68) Arnott, S.; Hukins, D. W. L.; Dover, S. D.; Fuller, W.; Hodgson, A. R. Structures of Synthetic Polynucleotides in the A-RNA and A'-RNA Conformations: X-ray Diffraction analyses of the Molecular Conformations of Polyadenylic Acid · Polyuridylic Acid and Polyinosinic Acid · Polycytidylic Acid. *J. Mol. Biol.* **1973**, *81* (2), 107–122.
- (69) Altona, C.; Sundaralingam, M. Conformational Analysis of the Sugar Ring in Nucleosides and Nucleotides. New Description Using the Concept of Pseudorotation. *J. Am. Chem. Soc.* **1972**, *94* (23), 8205–8212.
- (70) Lu, X.-J.; Olson, W. K. 3DNA: A Software Package for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic Acid Structures. *Nucleic Acids Res.* **2003**, *31* (17), 5108–5121.
- (71) Zhang, H.; Keane, S. C. Advances That Facilitate the Study of Large RNA Structure and Dynamics by Nuclear Magnetic Resonance Spectroscopy. *Wiley Interdiscip. Rev.: RNA* **2019**, *10* (5), No. e1541.
- (72) Tants, J.-N.; Schlundt, A. Advances, Applications, and Perspectives in Small-Angle X-Ray Scattering of RNA. *ChemBioChem* **2023**, *24* (17), No. e202300110.
- (73) Allison, J. R.; Hertig, S.; Missimer, J. H.; Smith, L. J.; Steinmetz, M. O.; Dolenc, J. Probing the Structure and Dynamics of Proteins by Combining Molecular Dynamics Simulations and Experimental NMR Data. *J. Chem. Theory Comput.* **2012**, *8* (10), 3430–3444.
- (74) Simmerling, C.; Strockbine, B.; Roitberg, A. E. All-Atom Structure Prediction and Folding Simulations of a Stable Protein. *J. Am. Chem. Soc.* **2002**, *124* (38), 11258–11259.
- (75) Tinoco, I.; Uhlenbeck, O. C.; Levine, M. D. Estimation of Secondary Structure in Ribonucleic Acids. *Nature* **1971**, *230* (5293), 362–367.

- (76) Freier, S. M.; Kierzek, R.; Jaeger, J. A.; Sugimoto, N.; Caruthers, M. H.; Neilson, T.; Turner, D. H. Improved Free-Energy Parameters for Predictions of RNA Duplex Stability. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83* (24), 9373–9377.
- (77) Sheehy, J. P.; Davis, A. R.; Znosko, B. M. Thermodynamic Characterization of Naturally Occurring RNA Tetraloops. *RNA* **2010**, *16* (2), 417–429.
- (78) Howe, C. P.; Greetham, G. M.; Procacci, B.; Parker, A. W.; Hunt, N. T. Sequence-Dependent Melting and Refolding Dynamics of RNA Uncg Tetraloops Using Temperature-Jump/Drop Infrared Spectroscopy. *J. Phys. Chem. B* **2023**, *127* (7), 1586–1597.
- (79) Antao, V. P.; Tinoco, I. Thermodynamic Parameters for Loop Formation in RNA and DNA Hairpin Tetraloops. *Nucleic Acids Res.* **1992**, *20* (4), 819–824.
- (80) Williams, D. J.; Hall, K. B. Experimental and Computational Studies of the G[UUCG]C RNA Tetraloop. *J. Mol. Biol.* **2000**, *297* (5), 1045–1061.
- (81) Yesselman, J. D.; Denny, S. K.; Bisaria, N.; Herschlag, D.; Greenleaf, W. J.; Das, R. Sequence-Dependent RNA Helix Conformational Preferences Predictably Impact Tertiary Structure Formation. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116* (34), 16847–16855.
- (82) Arnott, S. *Polynucleotide Secondary Structures: An Historical Perspective*. In *Oxford Handbook of Nucleic Acid Structure*; Neidle, S., Ed.; Oxford University Press, 1999; pp 1–38.
- (83) Arnott, S.; Hukins, D. W. Optimised Parameters for A-DNA and B-DNA. *Biochem. Biophys. Res. Commun.* **1972**, *47* (6), 1504–1509.
- (84) Bergonzo, C.; Cheatham, T. E. Improved Force Field Parameters Lead to a Better Description of RNA Structure. *J. Chem. Theory Comput.* **2015**, *11* (9), 3969–3972.
- (85) Love, O.; Winkler, L.; Cheatham, T. E. Van Der Waals Parameter Scanning with Amber Nucleic Acid Force Fields: Revisiting Means to Better Capture the RNA/DNA Structure through MD. *J. Chem. Theory Comput.* **2024**, *20* (2), 625–643.
- (86) Mlýnský, V.; Kührová, P.; Stadlbauer, P.; Krepl, M.; Otyepka, M.; Banáš, P.; Sponer, J. Simple Adjustment of Intranucleotide Base-Phosphate Interaction in the OL3 Amber Force Field Improves RNA Simulations. *J. Chem. Theory Comput.* **2023**, *19* (22), 8423–8433.
- (87) Schneider, B.; Morávek, Z.; Berman, H. M. RNA Conformational Classes. *Nucleic Acids Res.* **2004**, *32* (5), 1666–1677.
- (88) Nichols, P. J.; Henen, M. A.; 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* (1), 61.
- (89) Bothe, J. R.; Nikolova, E. N.; Eichhorn, C. D.; Chugh, J.; Hansen, A. L.; Al-Hashimi, H. M. Characterizing RNA Dynamics at Atomic Resolution Using Solution-State NMR Spectroscopy. *Methods* **2011**, *8* (11), 919–931.
- (90) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38.