Accurately Modeling RNA Stem-Loops in an Implicit Solvent Environment

Jason T. Linzer[1], Ethan Aminov[1], Aalim S. Abdullah[1], Colleen E. Kirkup[1], Rebeca I. Ventura Diaz[1], Vinay R. Bijoor[1], Jiyun Jung[1], Sophie Huang[1], Chi Gee Tse[1], Emily Álvarez Toucet[1], Hugo P. Onghai[1], Arghya P. Ghosh[1], Alex C. Grodzki[1], Emilee R. Haines[1], Aditya S. Iyer[1], Mark K. Khalil[1], Alexander P. Leong[1], Michael A. Neuhaus[1], Joseph Park[1], Asir Shahid[1], Matthew Xie[1], Jan M. Ziembicki[1], Carlos Simmerling[1,2] and Maria C. Nagan[1,*]

[1]*Department of Chemistry, Stony Brook University, Stony Brook, NY 11794, United States*

[2]*Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, New York 11794, United States*

*To whom correspondence should be addressed

Author contact information: maria.nagan@stonybrook.edu; Tel. 631-632-5793

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 fifteen 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 A-form RNA in the stem region. Loop structures are more sensitive and experimental structures could only be reproduced with modification of CH---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, nonstandard base pairs, bulges, internal loops, pseudoknots and multi-helix 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 co-axial helical stacking.[3-5] A number of computational methods are currently under development to predict tertiary RNA structure, employing a variety 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 are 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 modelling of smaller stem-loops and more generally understanding 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 stacking[21] 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 spectroscopic parameters to ensure the physics model is balanced.

In addition to an accurate force field, 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 timescale of folding in simulations is difficult to test on a variety of systems with these methods. Another issues 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 explicit solvent while still allowing adequate sampling, but accurate implicit solvent models for RNA have lagged those available for proteins.

As a starting point for gaining insight into better modeling of 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] As well, 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, possibly RNA atomistic force fields overall, and eventually may lead to more accurate MD simulations on a faster timescale.

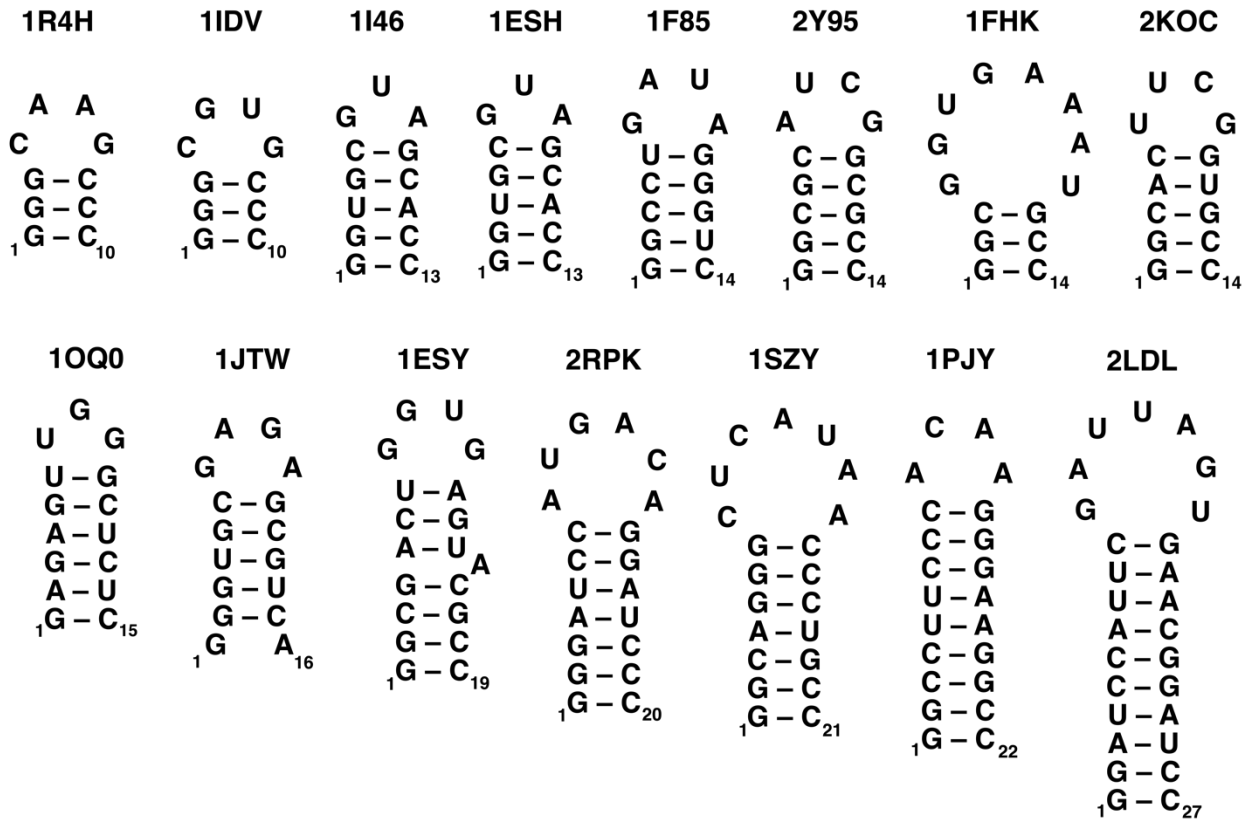

Figure 1. Secondary structures of stem-loops simulated. PDBID codes of the corresponding experimental structure are also given. See *Methods: RNA Systems* for references.

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 selecting sequences, base pairs known to require an explicit water molecule such as the G•U wobble base pair were excluded from consideration. Simulations were begun from the lowest energy experimental structures. These included the following PDBID entries: 1R4H,[43] 1IDV,[43] 1I46,[44] 1ESH,[45] 1F85,[46] 2Y95,[42] 1FHK,[47] 2KOC,[48] 1OQ0,[49] 1JTW,[50] 1ESY,[51] 2PRK,[52] 1SZY,[53] 1PJY,[54] and 2LDL,[55] 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-Neck2nuc[56] implicit solvent model and the mbondi2[57] 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 the with the recently developed hydrogen repulsion (HR) modification[58] 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 oxygen atoms. GB models represent only the polar contribution to the solvation free energy ($\Delta G_{solv,pol}$), which tends to be much larger in magnitude than nonpolar solvation ($\Delta G_{solv,np}$) for polar solutes. For proteins, GB is often supplemented with a SASA-based estimate of $\Delta G_{solv,np}$ (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 (pwSASA[59]), 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 that was incorporated into the implicit solvent model (described in depth in Supplementary 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 the larger RNA set.

MD Simulation Conditions. Simulations were carried out with the Amber suite of simulation software,[60-62] including system setup with *tleap*, *pdb4amber*, and *parmed* as well as *sander* and *pmemd.cuda* for simulations and minimization, and *cpptraj* for most post-trajectory analysis. General simulation protocols were similar for all systems, as outlined in previous implicit solvent RNA simulations.[38] The SHAKE algorithm[63] was applied to all bonds with H and hydrogen mass repartitioning[64] 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 the *in silico* melting temperature (50-50 folded and unfolded state) as possible. Initial temperatures were estimated with OligoCalc,[65] 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 (Figure 2 and Figure S4 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 simulations are listed in both Figure 2 and S4.

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 algorithm[66] on non-hydrogenic atoms in the stem residues as implemented in *cpptraj*. 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 overlayed 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 *cpptraj* with heavy atom-heavy atom (X-Y) distances < 4.0 Å and an X-H-Y angle cutoff of 135°.

Structural factors indicative of A-form RNA helical structure[67] were determined for each stem cluster trajectory such as stem base pair formation, x-displacement from the helical axis and C3'-endo sugar pucker presence (Figure 4). Canonical A-form RNA exhibits a C3'-endo sugar pucker, which is determined by the pseudorotation phase angle (P) of 0 to 36°.[68] 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.[69]

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 hydrogen bonding of G O6 to U N3-H3 and G N1-H1 to U O2. Other non-canonical 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[66] 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 and 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 structure were also calculated.

Results and Discussion

Each simulation trajectory was performed for 2.0 μs 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.[70] In some of the simulations, such as 1R4H, 1ESH and 1F85, there is a clear oscillation between a 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 timescale 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.[71-74]

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 be 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 lowered to less than 2 Å.

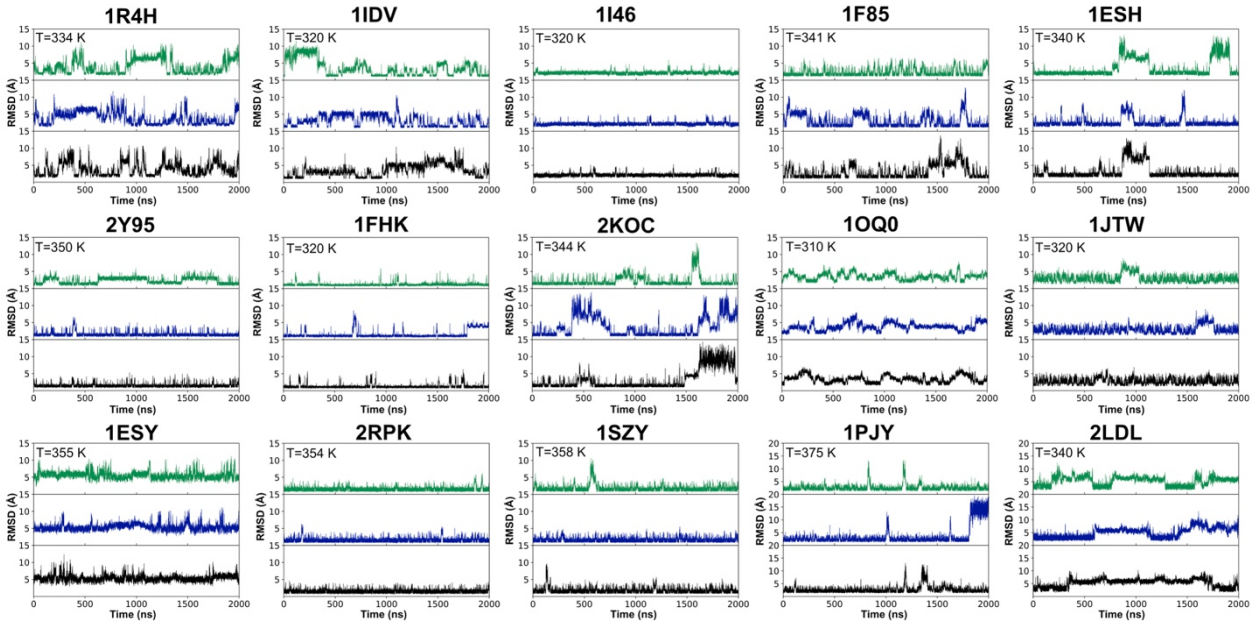

Figure 2. Plots of RMSD (Å) from the NMR structure for each stem-loop. Sequences of each RNA structure, designated by PDBIDs correspond to those in Figure 1. All RMSDs are shown for non-hydrogenic atoms in the stem region. Each trajectory was performed in triplicate (black, blue or green indicates an independent trajectory), with the ChiOL3 force field in implicit solvent at the indicated temperature.

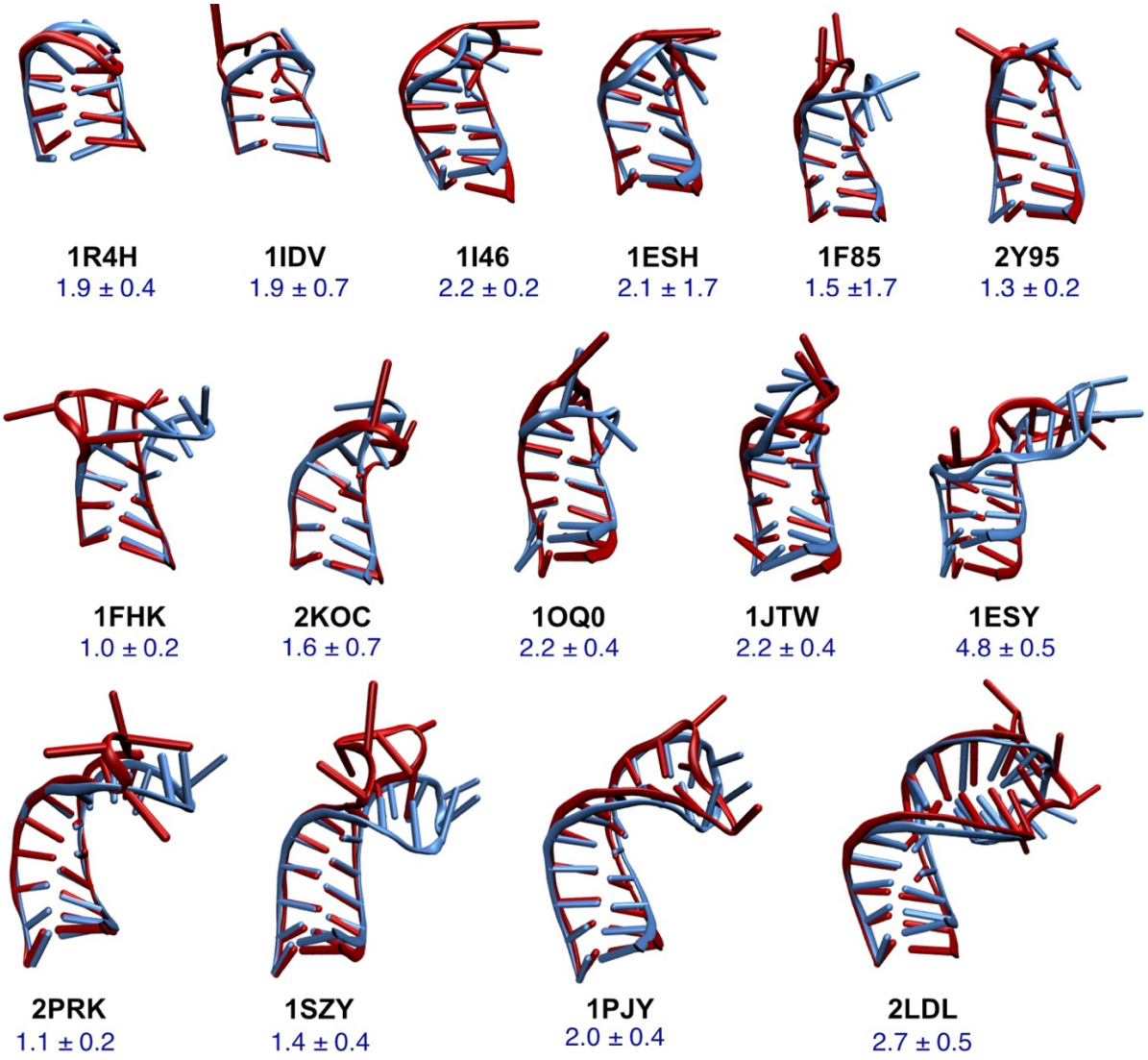

Figure 3. Representative cluster structures with stem residues, chiOL3 and GB compared to experiment. Overlays of representative structures from the most populated cluster *cluster 1* (red) with the lowest energy NMR structure (blue). Clustering was performed with the k-means method on the non-hydrogenic 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.

We show that in an implicit solvent environment RNA stem-loops can unfold and refold; meaning after melting, the base pairs in the stem-region can reform. This is achieved by a combination of GB-Neck2nuc[56]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, 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 are observed in these implicit solvent simulations. Impressively 1F85 contains two noncanonical 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.

Structural parameters for the ensemble of structures in *cluster 1* were assessed and average values were calculated (Figure 4, Table S3). Most base pairs are present 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 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.[75-77]

Two of the larger systems only partially reform the stem structure, perhaps due to the presence of nonstandard base pairing. For instance, the 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.[51] 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 it 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 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 it 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°,[69] 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 towards 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 Å 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 the stem *cluster 1* were examined for presence of base pairing (Figure 4, 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 were assessed for RNA classic structural parameters. Sugar pucker orientation was determined based upon the pseudorotation phase angle, with C3'-endo defined as 18 ± 18° and C2'-endo as 162 ± 18°.[68] When clustering on the stem residues, C3'-endo sugar puckers (Figure 4, 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 that sampling of 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 1ESY and also in 1JTW terminal A16. In the case of 1ESY, this is probably due to the disruption of the helix already noted and in 1JTW, because the terminal base pair G1:A16 is not paired in the NMR structure.

C3'-endo sugar puckers and a significant x-displacement from the helical axis (–4.4 Å) are hallmark features of A-form helical RNA structure.[67, 78] The average x-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 x-displacement in the helical region can vary from –1.8 Å 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[79] and this is qualitatively in line with A-form RNA.

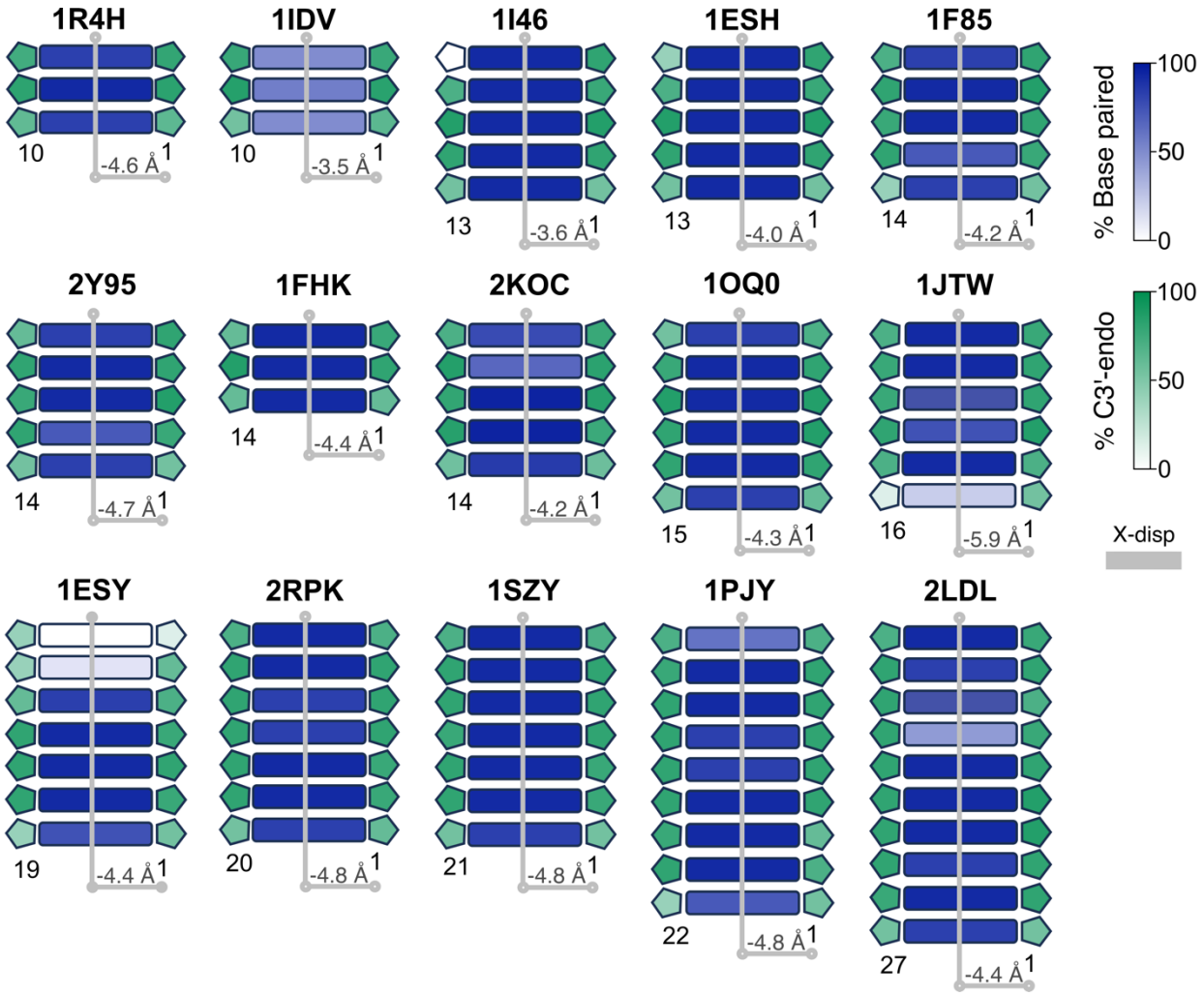

Figure 4. Structural parameters of base-paired regions of highest-ranked stem cluster trajectory with chiOL3 and GB. Pentagons represent sugars and indicate the percentage of the cluster trajectory exhibiting a C3′-endo sugar pucker. Rectangles between stem residues indicate the percentage of the cluster 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.

The ability of our simulations in implicit solvent to form A-form RNA is consistent with results found in explicit solvent simulations. In explicit solvent simulations with TIP3P, Sponer and coworkers assessed duplex RNA structural parameters[26] and found that both ff99bsc0[23-25] and chiOL3 generally retain A-form RNA structure, with x-displacement values –4.85 ± 1.60 Å and –4.95 Å, respectively. In these simulations, the sugar puckers are generally C3'-endo (P is 19.3 parmbsc0 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 stem-loops can fold and unfold 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.[80-82] 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 to the energy function. In explicit solvent, we had previously found a force field modification that minimizes repulsion in the CH•••O interactions to be necessary for proper loop structure formation.[58] Additionally, it is well established that burial of nonpolar groups is a major driving force for protein stability and as such it was expected 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 in with GB-Neck2nuc,[56] umbrella sampling of a model system was performed (see Supplementary Information) in which an adenine is stacked over a G:C base pair (Figure S2). Analysis of the energy profile indicated that 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

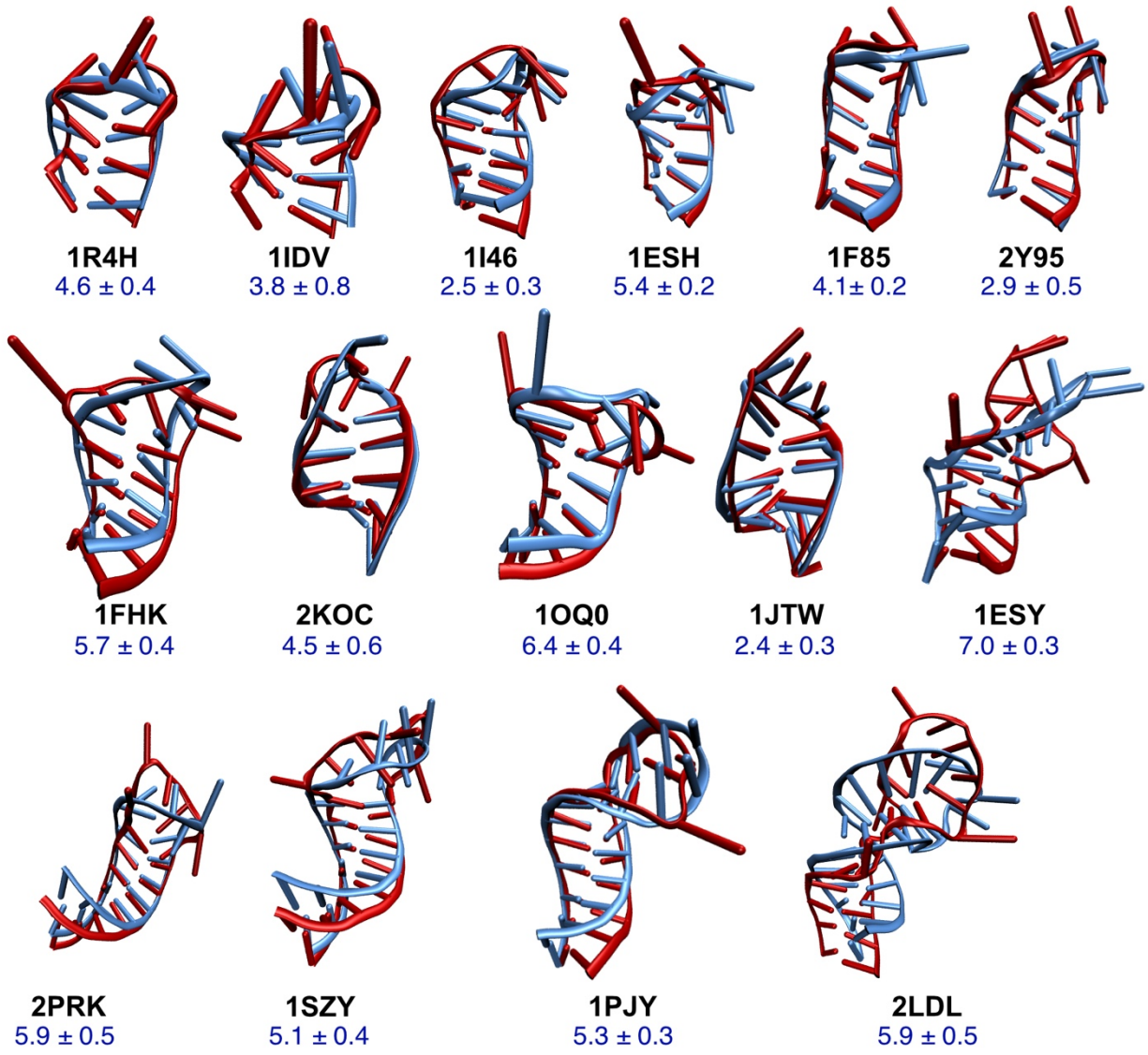

Figure 5. Comparison of structures from clusters on the loop and closing first base pair residues with ChiOL3 and GB simulations relative to the experimental structure. Overlays of representative structures from the most populated cluster (*cluster 1*, red) with the lowest energy NMR structure or x-ray 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 ± standard deviations between the *cluster 1* ensemble of structures and the experimental structure are shown.

pairs of heavy atoms in bases (Figure S2, Table S1). To capture this interaction in GB-Neck2nuc simulations, while still also 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 (see Supplementary Information). This nonpolar modification to GB-Neck2nuc implicit solvent model will be referred to hereafter as GB_{np}.

Six of the 15 RNA structures were unfolded and refolded with 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, 2KOC) as well as two larger stem-loops (2LDL and 1FHK). Simulations were carried out 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 μs in triplicate indicated stable simulations with variable unfolding and folding events (Supplementary Information). Structures were clustered on the loop and first base pair and compared to 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 the 1JTW loop structure in the most populated cluster (*cluster 1*) and the NMR structure was 2.4 ± 0.3 Å. However, with the chiOL3-HR, it improved to 1.4 ± 0.2 Å (Figure 6A). The 1FHK structure showed an even greater improvement in loop structure 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 and by all accounts is not accurate (Figure 5). However, with chiOL3-HR, the 1FHK *cluster 1* loop (29%) has an RMSD is 2.8 ±0.2 Å from the NMR structure (Figure 6B) and the *cluster 2*, populated 26%, has an even greater improvement with an RMSD of 1.9 ± 0.5 Å.

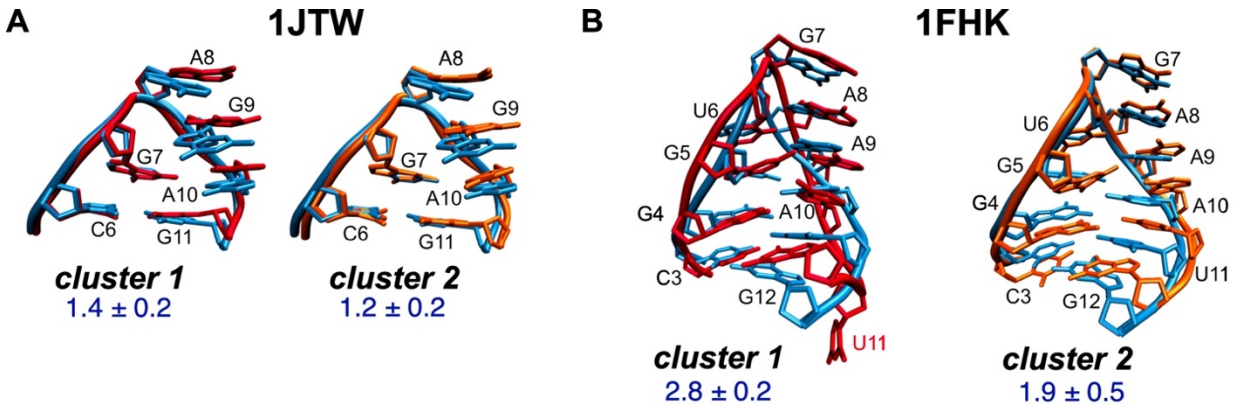

Figure 6. Loop Structures with chiOL3-HR and GB_{np}. Overlays of representative structures from the most populated cluster on the loop, *cluster 1*(red) and *cluster 2*(orange) with the lowest energy NMR structure (blue) for two different systems. Structures are shown for only the loop and closing base 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 ± 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.

Characteristic structural features of 1JTW are preserved when compared to 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 a 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 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 a 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 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 octaloop 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 U6 and G7. This is held together by hydrogen bonds between U6 N3-H3 and the A9 backbone O2P (93% of NMR structures) and a U6 2'OH to 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) are 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 to 3.0 Å. The variation the *cluster 1* structures or standard deviation in the RMSD values was also smaller, indicating more uniformity in the ensemble of structures sampled.

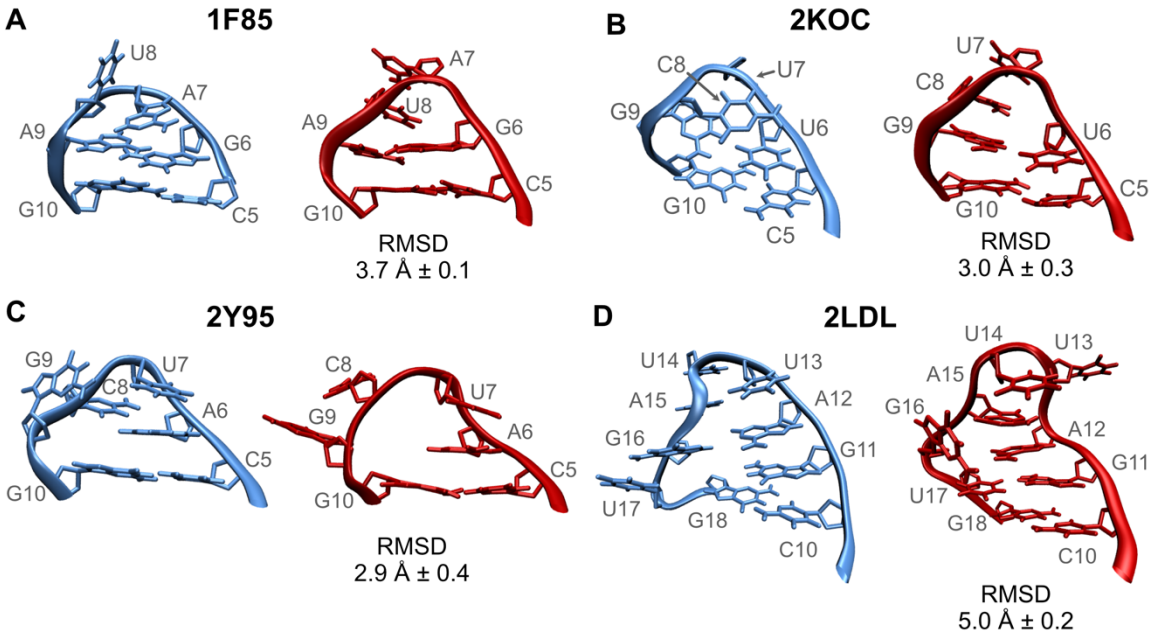

Figure 7. Comparison of loop structures and NMR structures simulated with the chiOL3-HR that still require some improvement. Structures show highest populated cluster (*cluster 1*) of the loop and closing base pair (red) and the NMR structure (blue) for the corresponding region.

Still, key structural attributes were not conserved in simulations of these systems (Figure 7, Table S6). In all four stem-loop structures, one of the base flips in the loop (1F85-A7, 2KOC-C8, 2Y95-C8, 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 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 towards 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 also is noncanonical. In A-form RNA, α tends to be *gauche*$^-$ *(g*$^-$*)* with an average value of 292°.[67, 83] 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, simulations do not match the experimental α-torsion values. This may be due to the inherent flexibility in the backbone. All of these structures are NMR structures and some backbone positions are difficult to resolve by NMR.[84, 85] 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 timescales 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 RNA structure. A key step in RNA model improvement would be large-scale tests of 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 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 sampling of alternate conformations.

Analysis of the simulation data indicates that the chiOL3 RNA model combined with GB-Neck2nuc solvation perform 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 lead 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, were changed in the GB_{np} correction.

When considered all 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 of other RNA force fields. Continued improvements to computer hardware and simulation methods will permit evaluation of the benchmark test via converged ensembles in explicit water, providing additional guidance for the improvement of RNA simulation models.

Data and Software Availability

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 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 and includes:

•A description of training data for the implicit solvent base stacking calculations is described, including changes to the Lennard-Jones parameters. RMSD plots for chiOL3 and chiOL3-HR with GBnp simulations, RNA structural analysis for base pairing and sugar puckers shown in Figures 4, 5 and 6.

Acknowledgements

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. Haines, E. A. Toucet, A.S. Abdullah and M.C. Nagan were supported by NSF REU program, CHE-2050541. H. Onghai was supported by Simons Summer Research Program. C. Simmerling 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, DOI: 10.1016/s0959-440x(03)00076-9.

(2) Mathews, D. H. Revolutions in RNA Secondary Structure Prediction. *J. Mol. Biol.* **2006**, *359* (3), 526-532, DOI: 10.1016/j.jmb.2006.01.067.

(3) Ferre-D'Amare, A. R.; Doudna, J. A. RNA Folds: Insights from Recent Crystal Structures. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 57-73, DOI: 10.1146/annurev.biophys.28.1.57.

(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, DOI: 10.1021/ar200098t.

(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), e2112677119, DOI: 10.1073/pnas.2112677119.

(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, DOI: 10.1073/pnas.0703836104.

(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, DOI: 10.1261/rna.075341.120.

(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, DOI: 10.1021/acs.jpcb.7b10102.

(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, DOI: 10.1073/pnas.87.21.8467.

(10) Bevilacqua, P. C.; Blose, J. M. Structures, Kinetics, Thermodynamics, and Biological Functions of RNA Hairpins. *Annu. Rev. Phys. Chem.* **2008**, *59*, 79-103, DOI: 10.1146/annurev.physchem.59.032607.093743.

(11) Goodall, G. J.; Wickramasinghe, V. O. RNA in Cancer. *Nat. Rev. Cancer* **2021**, *21* (1), 22-36, DOI: 10.1038/s41568-020-00306-0.

(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, DOI: 10.1126/science.289.5481.905.

(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, DOI: 10.1002/wrna.1196.

(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, DOI: 10.1021/acs.biochem.9b00535.

(15) Auffinger, P. Molecular Dynamics Simulations of RNA Systems. In *Handbook of RNA Biochemistry*, 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, DOI: 10.1021/acs.chemrev.7b00427.

(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, DOI: 10.1093/nar/gkl580.

(18) McCarthy, E.; Ekesan, Ş.; 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, DOI: 10.1093/nar/gkad260.

(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, DOI: 10.1093/nar/28.13.2527.

(20) Spackova, N.; Sponer, J. Molecular Dynamics Simulations of Sarcin-Ricin rRNA Motif. *Nucleic Acids Res.* **2006**, *34* (2), 697-708, DOI: 10.1093/nar/gkj470.

(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, DOI: 10.1021/ct501025q.

(22) Mrazikova, 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, DOI: 10.1021/acs.jctc.0c00801.

(23) Cheatham, T. E., 3rd; 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, DOI: 10.1080/07391102.1999.10508297.

(24) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; 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, DOI: 10.1021/ja00124a002.

(25) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III; Laughton, C. A.; Orozco, M. Refinement of the Amber Force Field for Nucleic Acids: Improving the Description of A/Γ Conformers. *Biophys. J.* **2007**, *92* (11), 3817-3829, DOI: 10.1529/biophysj.106.097782.

(26) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E., 3rd; 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, DOI: 10.1021/ct200162x.

(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, DOI: 10.1007/s00249-004-0448-6.

(28) Soares, T. A.; Hunenberger, P. H.; Kastenholz, 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, DOI: 10.1002/jcc.20193.

(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, DOI: 10.1007/s00249-011-0700-9.

(30) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell Jr, 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, DOI: 10.1002/jcc.21777.

(31) Foloppe, N.; MacKerell, J., Alexander D. 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, DOI: 10.1002/(SICI)1096-987X(20000130)21:2<105::AID-JCC3>3.0.CO;2-P.

(32) MacKerell Jr, 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, DOI: 10.1002/(SICI)1096-987X(20000130)21:2<105::AID-JCC3>3.0.CO;2-P.

(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, DOI: 10.1002/prot.21123.

(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, DOI: 10.1021/acs.jctc.5b00255.

(35) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, 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, DOI: 10.1021/acs.jctc.9b00591.

(36) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Cheatham, T. E., 3rd. Highly Sampled Tetranucleotide and Tetraloop Motifs Enable Evaluation of Common RNA Force Fields. *RNA* **2015**, *21* (9), 1578-1590, DOI: 10.1261/rna.051102.115.

(37) Kührová, 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, DOI: 10.1021/acs.jctc.6b00300.

(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, DOI: 10.1021/acs.jctc.2c00065.

(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, DOI: 10.1021/acs.jcim.1c01148.

(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, DOI: 10.1016/0022-2836(90)90386-Z.

(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., Jr.; et al. Cuucgg Hairpins: Extraordinarily Stable RNA Secondary Structures Associated with Various Biochemical Processes. *Proc.Natl. Acad. Sci. U.S.A.* **1988**, *85* (5), 1364-1368, DOI: 10.1073/pnas.85.5.1364.

(42) Duszczyk, 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, DOI: 10.1261/rna.2747411.

(43) Rijnbrand, R.; Thiviyanathan, 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, DOI: 10.1016/j.jmb.2004.08.095.

(44) 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, DOI: 10.1006/jmbi.2001.4497.

(45) 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, DOI: 10.1038/75202.

(46) 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, DOI: 10.1038/81951.

(47) 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, DOI: 10.1006/jmbi.2000.4431.

(48) Nozinovic, S.; Fürtig, B.; Jonker, H. R. A.; Richter, C.; Schwalbe, H. High-Resolution NMR Structure of an RNA Model System: The 14-Mer cUUCGg Tetraloop Hairpin RNA. *Nucleic Acids Res.* **2009**, *38* (2), 683-694, DOI: 10.1093/nar/gkp956.

(49) 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, DOI: 10.1093/nar/gkg351.

(50) 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, DOI: 10.1021/bi0111909.

(51) 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. *J. Mol. Biol.* **2000**, *299* (1), 145-156, DOI: 10.1006/jmbi.2000.3710.

(52) 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.* **2008**, *37* (2), 368-381, DOI: 10.1093/nar/gkn918.

(53) 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, DOI: 10.1006/jmbi.1996.0903.

(54) 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, DOI: 10.1093/nar/gkg654.

(55) 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, DOI: 10.1016/j.jmb.2011.11.034.

(56) 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, DOI: 10.1021/acs.jctc.5b00271.

(57) 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, DOI: 10.1002/prot.20033.

(58) Raguette, 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•••O Interactions Leads to Improved Stability of RNA Stem-Loop Structures in MD Simulations. *J. Phys. Chem. B* **2024**, *submitted*, DOI: 10.26434/chemrxiv-2024-9rbjw.

(59) 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, DOI: 10.1021/acs.jctc.8b00413.

(60) D.A. Case, H. M. A., K. Belfon, I.Y. Ben-Shalom, J.T. Berryman, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, N. Forouzesh, G. Giamba¸su, T. Giese, M.K. Gilson, H. Gohlke, A.W. Goetz, J. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails,

R.C. Walker, J. Wang, J. Wang, H. Wei, X. Wu, Y. Wu, Y. Xiong, Y. Xue, D.M. York, S. Zhao, Q. Zhu, and P.A. Kollman *Amber 2023*; University of California, San Francisco, 2023.

(61) D.A. Case, H. M. A., K. Belfon, I.Y. Ben-Shalom, J.T. Berryman, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, J. Wang, H. Wei, R.M. Wolf, X. Wu, Y. Xiong, Y. Xue, D.M. York, S. Zhao, and P.A. Kollman. *Amber22*; University of California, San Francisco, 2022.

(62) 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., III; Roe, D. R.; Roitberg, A.; Simmerling, C.; York, D. M.; Nagan, M. C.; Merz, K. M., Jr. Ambertools. *J. Chem. Inf. Model.* **2023**, *63* (20), 6183-6191, DOI: 10.1021/acs.jcim.3c01153.

(63) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a Systems with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327-341, DOI: 10.1016/0021-9991(77)90098-5.

(64) 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, DOI: 10.1021/acs.jctc.1c00185.

(65) Kibbe, W. A. Oligocalc: An Online Oligonucleotide Properties Calculator. *Nucleic Acids Res.* **2007**, *35* (suppl_2), W43-W46, DOI: 10.1093/nar/gkm234.

(66) Lloyd, S. Least Squares Quantization in Pcm. *IEEE Transactions on Information Theory* **1982**, *28* (2), 129-137, DOI: 10.1109/TIT.1982.1056489.

(67) 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, DOI: 10.1016/0022-2836(73)90183-6.

(68) Altona, C.; Sundaralingam, M. Conformational Analysis of the Sugar Ring in Nucleosides and Nucleotides. A New Description Using the Concept of Pseudorotation. *J. Am. Chem. Soc.* **1972**, *94* (23), 8205-8212, DOI: 10.1021/ja00778a043.

(69) Lu, X.; 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, DOI: 10.1093/nar/gkg680.

(70) 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, DOI: 10.1021/ja0273851.

(71) Tinoco, I.; Uhlenbeck, O. C.; Levine, M. D. Estimation of Secondary Structure in Ribonucleic Acids. *Nature* **1971**, *230* (5293), 362-367, DOI: 10.1038/230362a0.

(72) 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, DOI: 10.1073/pnas.83.24.9373.

(73) Sheehy, J. P.; Davis, A. R.; Znosko, B. M. Thermodynamic Characterization of Naturally Occurring RNA Tetraloops. *RNA* **2010**, *16* (2), 417-429, DOI: 10.1261/rna.1773110.

(74) 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, DOI: 10.1021/acs.jpcb.2c08709.

(75) Antao, V. P.; Tinoco, I., Jr. Thermodynamic Parameters for Loop Formation in RNA and DNA Hairpin Tetraloops. *Nucleic Acids Res.* **1992**, *20* (4), 819-824, DOI: 10.1093/nar/20.4.819.

(76) 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, DOI: 10.1006/jmbi.2000.3623.

(77) 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, DOI: 10.1073/pnas.1901530116.

(78) 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.

(79) Arnott, S.; Hukins, D. W. Optimised Parameters for a-DNA and B-DNA. *Biochem. Biophys. Res. Commun.* **1972**, *47* (6), 1504-1509, DOI: 10.1016/0006-291x(72)90243-4.

(80) Bergonzo, C.; Iii, T. E. C. Improved Force Field Parameters Lead to a Better Description of RNA Structure. *J. Chem. Theory Comput.* **2015**, *11* (9), 3969-3972, DOI: 10.1021/acs.jctc.5b00444.

(81) Love, O.; Winkler, L.; Cheatham, T. E., III. 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, DOI: 10.1021/acs.jctc.3c01164.

(82) Mlýnský, V.; Kührová, P.; Stadlbauer, P.; Krepl, M.; Otyepka, M.; Banáš, P.; Šponer, 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, DOI: 10.1021/acs.jctc.3c00990.

(83) Schneider, B.; Morávek, Z.; Berman, H. M. RNA Conformational Classes. *Nucleic Acids Res.* **2004**, *32* (5), 1666-1677, DOI: 10.1093/nar/gkh333.

(84) 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, DOI: 10.1038/s42003-018-0067-x.

(85) 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. *Nat. Methods* **2011**, *8* (11), 919-931, DOI: 10.1038/nmeth.1735.