

Adjusting the Energy Profile for CH–O Interactions Leads to Improved Stability of RNA Stem-Loop Structures in MD Simulations

Published as part of *The Journal of Physical Chemistry B* virtual special issue “Recent Advances in Simulation Software and Force Fields”.

Lauren E. Raguetto,[#] Sarah S. Gunasekera,[#] Rebeca I. Diaz Ventura, Ethan Aminov, Jason T. Linzer, Diksha Parwana, Qin Wu, Carlos Simmerling, and Maria C. Nagan^{*}



Cite This: <https://doi.org/10.1021/acs.jpcb.4c01910>



Read Online

ACCESS |



Metrics & More

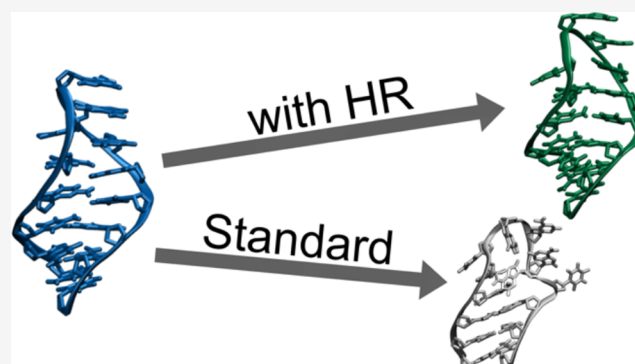


Article Recommendations



Supporting Information

ABSTRACT: The role of ribonucleic acid (RNA) in biology continues to grow, but insight into important aspects of RNA behavior is lacking, such as dynamic structural ensembles in different environments, how flexibility is coupled to function, and how function might be modulated by small molecule binding. In the case of proteins, much progress in these areas has been made by complementing experiments with atomistic simulations, but RNA simulation methods and force fields are less mature. It remains challenging to generate stable RNA simulations, even for small systems where well-defined, thermostable structures have been established by experiments. Many different aspects of RNA energetics have been adjusted in force fields, seeking improvements that are transferable across a variety of RNA structural motifs. In this work, the role of weak CH...O interactions is explored, which are ubiquitous in RNA structure but have received less attention in RNA force field development. By comparing data extracted from high-resolution RNA crystal structures to energy profiles from quantum mechanics and force field calculations, it is shown that CH...O interactions are overly repulsive in the widely used Amber RNA force fields. A simple, targeted adjustment of CH...O repulsion that leaves the remainder of the force field unchanged was developed. Then, the standard and modified force fields were tested using molecular dynamics (MD) simulations with explicit water and salt, amassing over 300 μ s of data for multiple RNA systems containing important features such as the presence of loops, base stacking interactions as well as canonical and noncanonical base pairing. In this work and others, standard force fields lead to reproducible unfolding of the NMR-based structures. Including a targeted CH...O adjustment in an otherwise identical protocol dramatically improves the outcome, leading to stable simulations for all RNA systems tested.



INTRODUCTION

The biological function of ribonucleic acid (RNA), first introduced as a key intermediate in gene expression,^{1,2} is more complicated than its traditionally attributed role. Many viruses have RNA-based genomes³ but RNA also serves as a regulatory molecule⁴ in cancer,⁵ the RNA interference pathway,⁶ long noncoding RNA molecules,⁷ and riboswitches.^{8,9} RNA molecules are expressed as single-stranded polynucleotides and often fold back on themselves to form complicated secondary and tertiary structures,^{10,11} with some of the most simplistic entities being stem-loop structures in which the loop is known to form distinct motifs.^{12,13} In addition, there are a plethora of noncanonical base pairing interactions^{14,15} some of which are pH dependent¹⁶ and some require water binding.¹⁷ RNA molecular dynamics studies have been carried out for more than 20 years¹⁸ but the field is less mature than that for proteins or DNA.¹⁹ This is due in part to the diversity of RNA

structures but also because there is no consensus on the force field for RNA.

At the heart of the Amber RNA force fields is the Cornell et al. additive force field²⁰ with restrained electrostatic potential (RESP) charges.²¹ Over the years, the force field has undergone a number of subtle but important adjustments that stabilize sugar pucker orientations,²² and backbone torsions^{23,24} resulting in the current ff99bsc0 χ OL3 force field, which will be referred to as OL3 hereafter. The phosphate oxygen van der Waals radii in OL3 have also

Received: March 22, 2024

Revised: July 22, 2024

Accepted: July 29, 2024

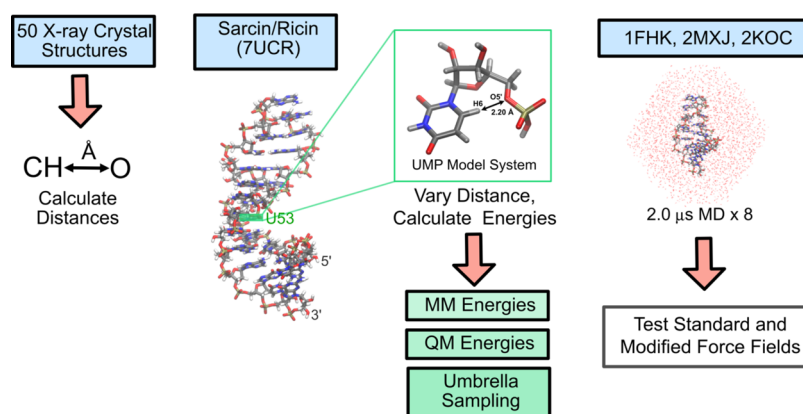


Figure 1. Schematic of Force Field Modification Approach. CH $\cdots$ O distances in naturally occurring X-ray crystal structures were examined. Then, the very stable *Sarcin/Ricin* was examined by MD. In the uridine monophosphate (UMP) model system, coordinates were extracted from residue U53 in PDBID: 7UCR. The distance between atoms H6 and O5' is indicated. From this model system, both molecular mechanical (MM) and quantum mechanical (QM) potential energy profiles were examined and umbrella sampling undertaken. From this, modified force fields were tested in explicit solvent on three biologically important systems, and results were compared to standard force fields.

been adjusted leading to the alternative force field parameters known as LJbb.^{25,26} Others have modified the Amber RNA force field torsions,^{27,28} parametrized backbone oxygen van der Waals radii,^{29,30} adjusted hydrogen bonding with an explicit term³¹ or altogether revised charges.³² Other classical force fields modeling atomistic RNA molecules include CHARMM^{33–35} and GROMOS^{36–38} as well as polarizable force fields that are also under development.^{39,40} The challenge of any accurate force field is to balance subtle interaction energies important for complex RNA structure⁴¹ in a relatively rugged energy landscape.^{42,43}

Weak interactions between CH $\cdots$ O are known to be important to RNA structure. Early analyses of X-ray crystal structures^{44,45} and computations^{46,47} indicated C8–H8 and C6–H6 to phosphate backbone oxygen atoms were important to RNA structure. Subsequently, weak CH $\cdots$ O interactions have also been found in small molecule-RNA recognition,^{48,49} biological structures in general,^{50–52} and more generally in molecular recognition.^{53–55} Nagan and co-workers employed solvated quantum mechanical calculations to demonstrate that C8–H8 and C6–H6 to O backbone interactions are critical for maintaining canonical tRNA anticodon structure.⁵⁶ It was concluded that the treatment of these backbone electrostatic interactions in atomistic force fields should be improved. An analysis herein of 50 high-quality RNA crystal structures in the PDB Databank indicates that distances between oxygen atoms and hydrogens bonded to carbon atoms do occur often in nature and that H to O distances are short across the entire backbone, not exclusive to interactions between the base and phosphate. Recently, an explicit base-phosphate nonbonded term was incorporated into the OL3 force field,⁵⁷ but it was also combined with a number of other changes.

Here, a targeted nonbonded adjustment specifically for CH $\cdots$ O interactions trained against quantum mechanical calculations is introduced and the approach is summarized in Figure 1. This modification is named hydrogen repulsion (HR) and indicates the importance of decreasing the repulsion of hydrogens involved in the RNA CH $\cdots$ O parameters. Since there are many different versions to the RNA force field, only LJbb^{25,26} and OL3²⁴ modifications were considered in this study; although the HR modification could in principle be applied to any variation of the underlying force field. Then, the

HR force field modification is examined in a number of experimentally relevant RNA stem-loop systems. Results presented herein show that some naturally occurring RNA systems spontaneously unfold with standard force fields and that the HR nonbonded term alone is sufficient to produce stable simulations in naturally occurring RNA systems.

METHODS

General Analysis and Visualization Tools. Structure visualization and image generation were performed with VMD.⁵⁸ All analyses of structures including distance, hydrogen bonding, vectors, angles, torsions, and root-mean-square deviations were carried out with the *cpptraj*⁵⁹ module of AmberTools.⁶⁰

Analysis of Distances in Crystal Structures. To determine the prevalence of naturally occurring CH $\cdots$ O interactions, high resolution crystal structures, which contain more precise atomic positions relative to NMR-based models, were examined. The set includes both single-stranded and double-stranded RNA, and RNA-protein complexes, with a variety of loops, canonical and noncanonical base pairing, phosphate-base hydrogen bonding, and stacking motifs. RNA chain lengths ranged from 16 to 30 nucleotides. This diversity helps increase the likelihood of observing whether these close contacts are present in biologically relevant conformations.

Collection of 50-High Resolution RNA Structures. A set of 50 high-resolution structures containing RNA coordinates were collected. The search tool at rcsb.org was used, filtering by RNA for “polymer entity” and resolution under 1.5 Å. Results were sorted by resolution. When multiple structures of related systems were included (such as for the *Sarcin/Ricin* loop, see Supporting Information), only a single example was retained. After filtering, the 50 structures with the highest resolution were retained (Table S1). Structures in each set were downloaded and processed through *pdb4amber*. Distances were calculated from these coordinates using the *cpptraj*⁵⁹ program.

UMP Model System. A small model system of uridine monophosphate (UMP) was constructed to quantify the energy response for varying the distance between the H6 and O5' atoms, with an overall goal of identifying the distance value at the energy minimum.

Quantum Mechanics Potential Energy Profiles. Orca⁶¹ 3.1 was used with the default optimization convergence to perform a 1-dimensional scan of the distance between H6 and O5' from 1.8 to 3.5 Å with a 0.5 Å increment. In addition to this distance, the dihedrals outside of the two rings were constrained. These included dihedral constraints on atoms HOP3-OP3-P-O5', OP3-P-O5'-C5', P-O5'-C5'-C4', O5'-C5'-C4'-O4', O4'-C1'-N1-C6, HO2'-O2'-C2'-C1', and HO3'-O3'-C3'-C4'. The QM geometries were optimized with B3LYP-D3^{62–64} with the aug-cc-pvDz⁶⁵ basis set. This is an affordable and consistent approach.⁶⁶ Energies were calculated for the optimized geometries using MP2⁶⁷/6-311+G**^{68,69} with TightSCF. Optimized QM structures were reminimized with the MM force fields to obtain potential energy profiles (Supporting Information).

Free Energy Profiles. Free energy profiles for varying the same distance in explicit solvent were obtained (setup detailed in Supporting Information). Following relaxation, a distance restraint was added between H6 and O5' with a force constant of 100 kcal/(mol·Å²). The position of the minimum for this restraint was varied during steered MD and umbrella sampling. Steering involved 5.0 ns MD with a 1.0 fs time step keeping constant moles, volume and temperature (NVT) at 300 K, in which the restraint minimum was varied linearly from 2.0 to 3.5 Å; all other restraints remained unchanged. Sixteen equally spaced snapshots were extracted from the steering trajectory file and used to initiate 16 restrained MD simulations (umbrella sampling windows). Each umbrella sampling window was 1.0 ns in length, propagated with a 1.0 fs time step under constant NVT (300 K). Values of the H6 to O5' distance were saved each time step. The weighted histogram analysis method (WHAM⁷⁰) using code from Grossfield⁷¹ was used to reconstruct the free energy profile from the umbrella sampling data, with 32 bins, a tolerance of 0.001 and 10 Monte Carlo trials. Two independent trials of umbrella sampling and WHAM analysis were generated for each force field setup to estimate precision.

Modification of Lennard–Jones Parameters. In the Amber-associated force fields a Lennard–Jones function is used to model van der Waals interactions. In this work, the form of the Lennard–Jones 12–6 potential is expressed as

$$E = \epsilon \left(\left(\frac{R_{\min}}{r} \right)^{12} - 2 \left(\frac{R_{\min}}{r} \right)^6 \right) \quad (1)$$

Where $R_{\min}$ is the distance at which the function has a minimum. Another common way to represent the same Lennard–Jones potential uses the variable σ instead of $R_{\min}$; these are related through the expression

$$R_{\min} = \sqrt[6]{2} \sigma \quad (2)$$

In order to alter only specific hydrogen–oxygen interactions, the “off-diagonal” $R_{\min}$ values were directly modified. Two different $R_{\min}$ values (2.7 and 2.8 Å) were tested by modifying the Amber parameter-topology files and repeating the entire setup and simulation protocol with the altered parameters. The atom names, atom types and original and modified Lennard–Jones parameters are provided in Tables S3 and S4. A sample *parmred* script for modifying the Amber topology files is also included in Supporting Information.

Example Lennard–Jones energy profiles corresponding to the parameters found in OL3 and LJbb, along with modified $R_{\min}$ values, are shown in Figure 2 for the H6 to O5'

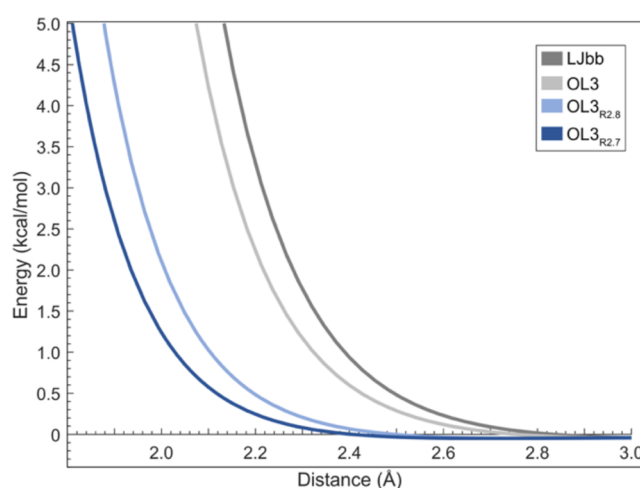


Figure 2. Lennard–Jones function energy profiles corresponding to the H6...O5' parameters for standard force fields LJbb ($R_{\min} = 3.1808$ Å) and OL3 ($R_{\min} = 3.0927$ Å) as well as OL3 modified force fields in which $R_{\min} = 2.7$ Å and $R_{\min} = 2.8$ Å. In all force fields, the default value for ϵ remained unchanged ($\epsilon = 0.050498$ kcal/mol).

interaction. All models show a very shallow minimum due to the small value of ϵ . With the standard parameters, the repulsion rises to ~ 1 kcal/mol at a distance of 2.4 Å, and 2–3 kcal/mol at 2.2 Å. As expected due to the increase in size of the oxygen atom,²⁶ the LJbb energy profile is shifted to the right (more repulsive) compared to OL3.

RNA Systems and Simulation Details. All simulations used OL3 or LJbb. Following recent comparisons of water models used in RNA MD,²⁵ the OPC 4-point water model⁷² was used for the solvent; K^+ and Cl^- ions used the Li/Merz parameter set⁷³ trained for OPC.

Simulation models were built from the experimental structure from the PDB (1FHK,⁷⁴ 2KOC,⁷⁵ 2MXJ,⁷⁶ 7UCR⁷⁷), using the AmberTools *tleap* module and loading the first model from the PDB file in cases where multiple models were present. For 2KOC, the 5' terminal phosphate was removed. For 7UCR, crystallographic water was retained, sulfate was removed, and names for deuterium atoms were changed to hydrogen.

System building proceeded as follows. An initial system was constructed in *tleap* by solvating the RNA in a truncated octahedron periodic box, with a minimum distance of 10.0 Å from solute atom to box boundary and a solute–water *closeness* value of 0.75. The volume of the resulting box was calculated and used to estimate the number of ions needed to produce 200 mM excess salt. Then the system was built again with solvent and ions (using the *addionsrand* command), adding sufficient K^+ ions to neutralize the RNA net charge, followed by 200 mM KCl.

Unless otherwise specified, all simulations used default settings in Amber v22,⁷⁸ a 10.0 Å direct space nonbonded interaction cutoff with particle mesh Ewald⁷⁹ for long-range electrostatics, a Langevin thermostat⁸⁰ with collision frequency of 1.0 ps^{−1}, a Monte Carlo barostat,⁸¹ and SHAKE⁸² on all bonds involving hydrogen atoms with 0.00001 Å tolerance. Simulations were carried out using the *pmemd.cuda* module of Amber v22 using a variety of NVIDIA GPUs.

Each system was relaxed using a 9-step protocol designed to prepare the system for simulation conditions, using the experimental structure as a reference for restraints, and a 1.0

fs time step (summarized Figure S1). All simulation parameters were unchanged in the next step unless noted. The first step included 100 steps of steepest descent minimization with 100.0 kcal/(mol·Å²) Cartesian positional restraints on the RNA heavy atoms. The system was then heated from 100 to 298 K over 100 ps at constant NVT, maintaining the same restraints. The third step was 100 ps MD at constant NPT (1 bar, 298 K) with the same restraints. The fourth step was 100 ps MD at constant NPT with the same restrained atoms and force constant reduced to 10.0 kcal/(mol·Å²). The fifth step was 100 steps of minimization at NVT with only RNA backbone atoms O3', C3', C4', C5', O5' and P restrained with a 10 kcal/(mol·Å²) force constant. Then MD at constant NPT with 10.0 kcal/(mol·Å²) restraints on the RNA backbone was carried out for 100 ps, followed by 100 ps MD easing restraints to 1.0 kcal/(mol·Å²) restraints on RNA backbone, 0.1 kcal/(mol·Å²) restraints and finally no restraints. The final coordinates and velocities from the last relaxation step were used to initiate NVT MD simulations.

Production simulations used different random seeds to avoid synchronization of dynamics,⁸³ and a 4.0 fs time step via hydrogen mass repartitioning.⁸⁴ For 7UCR, 3 independent production simulations of 1.0 μ s each were generated for each of the 6 force field variations (18 simulations total). For the stem-loop systems (1FHK, 2KOC, 2MXJ), 8 independent production simulations of 2.0 μ s each were generated for each of the 6 force field variations (48 simulations for each stem-loop, 288.0 μ s total).

Measurement of RNA Structural Features. All structural factors were calculated with *cptraj*. Presence of a hydrogen bond was defined as a heavy atom-heavy atom (X–Y) distance of <4.0 Å and an X–H–Y angle cutoff of 135°. The glycosidic angle (χ), defined as O4'–C1'–N1–C2 for pyrimidines and O4'–C1'–N9–C4 for purines, was assessed on a 360° scale with *syn* taken as between 0 and 90° or 270–360° while *anti* included angles between 180 $\pm$ 90°. For base stacking interactions, structures were first assessed for spatially overlapping aromatic rings of the bases (6m, 6-membered ring; 5m-5-membered ring) employing only nonhydrogenic cyclic atoms (e.g., purines 6m: N1,C2,N3,C4,C5 and C6). The angle between vectors normal to the planar base rings was calculated. Bases were considered stacked if the distance between the center-of-mass of the 6m or 5m portions of the base was $\leq$ 4.0 Å and the angle between the two vectors was 0 $\pm$ 40° or 180 $\pm$ 40°.

RESULTS AND DISCUSSION

Analysis of CH...O distances in naturally occurring RNA X-ray crystal structures was undertaken. Histograms of distances for each hydrogen–oxygen pair (Figure 3) indicate the pairs most frequently adopting distances closer than $\sim$ 2.4 Å including H6–O5' (typically pyrimidines), H8–O5' (purines), H2'–O4', H6–O4' (purines), H3'–OP2, and to a lesser extent, H8–O4' (purines) and H5'–O2'. It is notable that these H to O distances encompass base-backbone interactions but also backbone–backbone CH...O interactions. Other pairs also sample short distances at reduced frequency, perhaps in less-common structural motifs. Overall, the analysis indicates that these close interactions occur frequently in a variety of RNA structures.

The Lennard–Jones potential function using typical RNA parameters for H and O to determine the distances at which steric clash begins were then analyzed. The CH...O distance

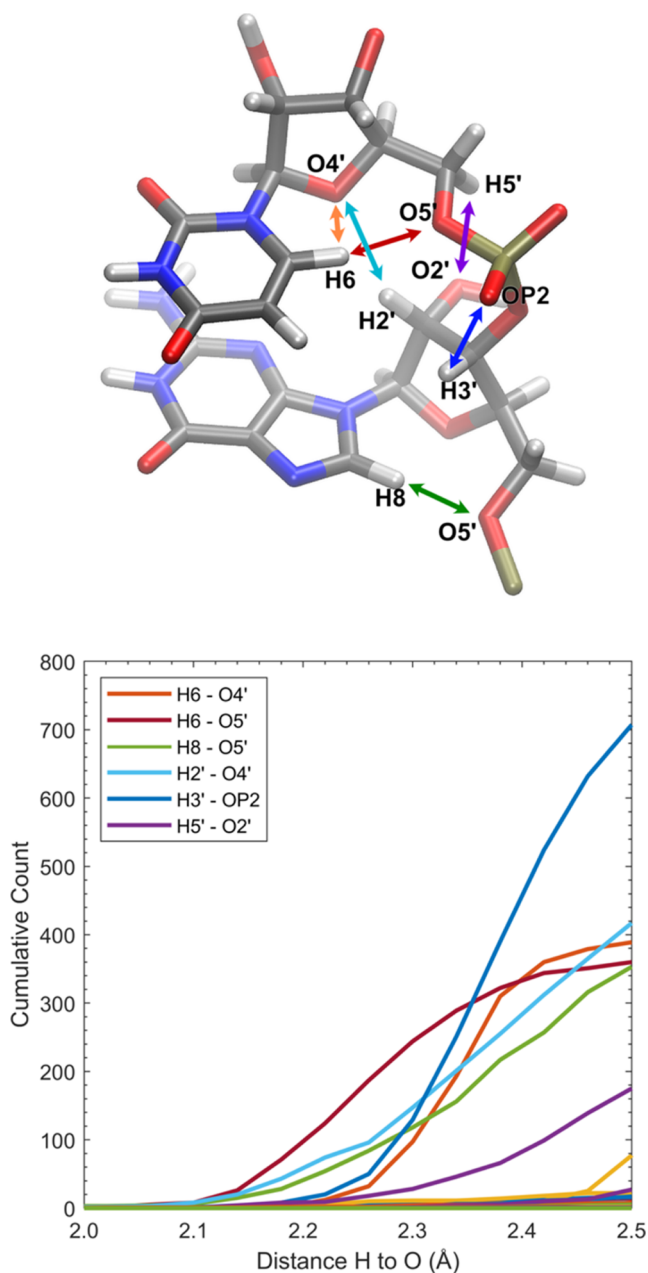


Figure 3. Cumulative histograms of distances between CH...O atom pairs in 50 RNA high-resolution crystal structures (Table S1). Each histogram represents data for a specific pair of H and O atoms, collected over all RNA residues in all structures. Data are shown for distances less than 2.4 Å, the region where the standard Lennard–Jones interaction becomes repulsive for these pairs. The analysis suggests a high occurrence of short CH...O distances, particularly for H6 and H8 to O5', H2', H6 and H8 to O4', H3'–OP2, and H5'–O2'.

was varied in a uridine monophosphate (UMP) model system, comparing quantum mechanics calculations to those obtained from gas-phase energy profiles from the Amber force fields, as well as free energy profiles in explicit water. Finally, these energy profiles were used to suggest simple modifications to the Lennard–Jones function for van der Waals interactions, and the changes were tested on multiple RNA systems via repeated runs of microsecond-time scale MD simulations with explicit water and salt.

Characterizing the Energy Profile for Varying the H6 to O5' Distance in Uridine Monophosphate from U53 in 7UCR. Analysis of crystal structures suggests that RNA experimental structures adopt distances (Figure 3) where the H $\cdots$ O Lennard–Jones repulsion rises (Figure 2). However, this reflects only one component of the overall energy and may not reflect the total energy for this interaction (including other atoms) in the context of geometries observed in RNA. An important component of our strategy for developing successful protein force fields^{87–89} relies on creating model systems that isolate and reproduce a specific weakness of the prior force field, and are tractable for calculating reference data using quantum mechanics (QM) calculations. Following that approach here, a uridine monophosphate model system (UMP, Figure 1) was built by extracting residue U53 from the PDB coordinates 7UCR⁷⁷ of the joint X-ray/neutron structure of the Sarcin-Ricin loop RNA. 7UCR was selected due to the high resolution and resolved positions of the hydrogen atoms. In this structure, U53 exhibits a close approach between H6 and O5' (2.20 Å, Figure 1), as frequently observed in RNA crystal structures (Figure 3). By varying this distance and calculating the MM energy (here using the OL3 force field only), the force field energy can be evaluated to determine whether the energy is unfavorable at the distance seen in the crystal structure. Since the environment of the model system differs from the full crystal structure, the QM energy profile for the same geometries as the MM energy scan were also calculated (see Methods Section for details). This allows direct evaluation of the MM function accuracy.

The energy profiles for varying the H6 – O5' distance in UMP are shown in Figure 4. In the OL3 force field, the energy minimum occurs near 2.5 Å, slightly shorter than seen for the isolated H to O Lennard–Jones function (Figure 2) but still much larger than the 2.2 Å in the full RNA crystal structure. The model lacks the complex RNA environment in 7UCR, which could be responsible for the longer distance preferred in the model system. However, the QM energy profile for UMP shows a minimum near 2.1 Å (Figure 4A), much more consistent with the experimental RNA structure. This consistency between QM and experiment provides additional evidence that the force field is overly repulsive for this important RNA feature.

These potential energy scans were carried out in the gas phase to facilitate comparison of QM and MM energy profiles. However, the partial charge model in the force fields used here includes implicit polarization more suitable for aqueous simulations.^{21,90} To evaluate whether an aqueous environment would change the conclusions drawn from the gas-phase potential energy profiles, the free energy versus distance profiles (potentials of mean force) for the same UMP model system in explicit water using both the OL3 and LJbb force fields were calculated (Figure 4B, see Methods Section for details). The free energy profiles become flatter at longer distances due to the favorable interaction with water that is missing in the gas-phase energy. However, the positions of the minima in water closely match those on the corresponding gas-phase potential energy profiles. Overall, in gas phase and water, both OL3 and LJbb force fields are quite repulsive for H6–O5' distances observed in the crystal structure.

Empirical Modification of the CH $\cdots$ O Lennard–Jones Parameters. The CH $\cdots$ O distances favored by the standard OL3/LJbb RNA force fields are consistently shifted to larger

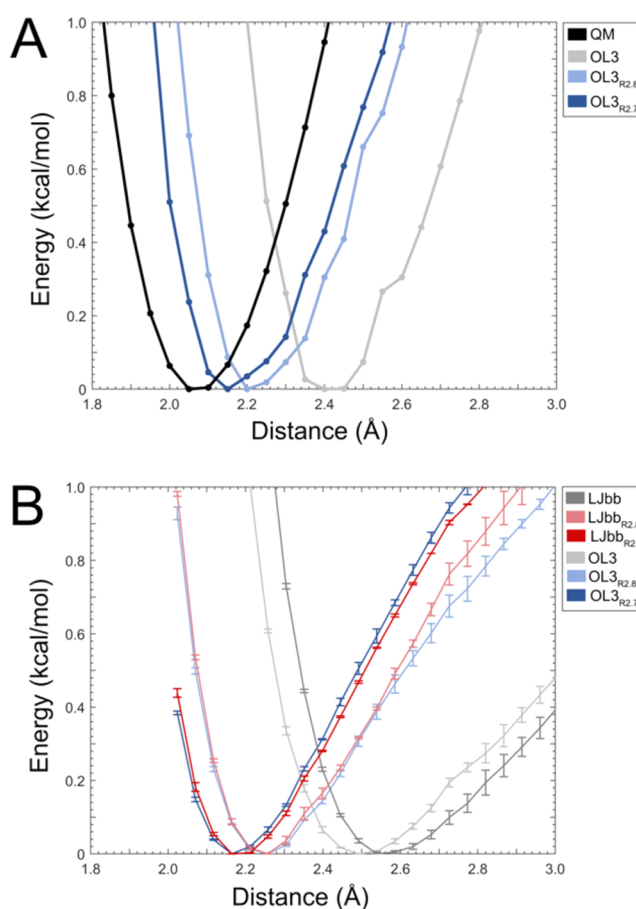


Figure 4. Total potential energy plotted versus distance between H6 and O5' in the UMP model system. UMP was used to compare force field and quantum mechanics energies, and to train modified van der Waals parameters for interaction of certain pairs of hydrogen and oxygen atoms. (A) Gas-phase potential energies at single points along the distance range (circles) for quantum mechanics (QM) and the OL3 RNA force field with standard and modified $R_{\min}$ for the Lennard–Jones interaction between H6 and O5'. (B) Free energy profiles from umbrella sampling simulations in explicit water at 300 K, using standard and modified OL3 and LJbb force fields.

values than are optimal in the QM energy profile and measured in experimental RNA structures. This observation supports the hypothesis that the CH $\cdots$ O interaction is too repulsive at the experimentally sampled distances. It is well established that the r^{12} function commonly used in force fields does not accurately reproduce Pauli repulsion compared to other functional forms such as an exponential.^{91,92} For hydrogen bonded interactions, the position of the overall energy minimum is a balance between strong electrostatic attraction and van der Waals repulsion and is therefore especially sensitive to inaccuracy in the repulsive Lennard–Jones region. We hypothesized that this may also be a source of the inaccurate CH $\cdots$ O energy profiles in Figure 4.

To overcome this short-range repulsion, the Lennard–Jones parameters for specific CH $\cdots$ O pairs in both the OL3 and LJbb force fields were empirically adjusted. Molecular mechanics force fields typically specify the Lennard–Jones ϵ and $R_{\min}$ values for each atom type, and the parameters for interactions between pairs of atom types are obtained using combining rules (Lorentz–Berthelot in the force fields used here). However, this simplification is done solely to avoid the need

to train separate parameters for all atom type pairs. New “off-diagonal” Lennard–Jones parameters were directly specified only for CH $\cdots$ O interactions, leaving unchanged their interaction with other atoms (NBFIX in CHARMM, LJEDIT in Amber). A different approach to reducing the repulsion would be to modify the Lennard–Jones parameters for the atom O, or the atom H, but doing so would modify all Lennard–Jones interactions involving that atom type, potentially altering the force field in unexpected ways (such as changing dihedral energy profiles). By making targeted off-diagonal adjustments applicable only to specific atom pairs, the CH $\cdots$ O repulsion can be reduced without any direct impact on the remainder of the RNA force field.

The impact of adjusting Lennard–Jones $R_{\min}$ parameters for the interaction between defined pairs of hydrogen and oxygen atoms was tested (see Methods Section for full details). While new $R_{\min}$ (and ϵ) values could be trained directly against the QM energy profile, inconsistencies between the gas-phase QM data and MM that uses prepolarized charges intended for solution simulations make the comparison qualitative (as discussed in⁸⁸). Instead, small empirical adjustments were made to the standard $R_{\min}$ values of ~ 3.1 – 3.2 Å (see Table S3), testing two smaller $R_{\min}$ values of 2.8 and 2.7 Å throughout the remainder of this work.

First the gas-phase MM potential energy profiles for UMP were recalculated, employing OL3 with reduced $R_{\min}$ values for CH $\cdots$ O interactions (Figure 4A). The adjustment resulted in the expected shift of the energy minimum to shorter values, approaching the distance of the QM energy minimum.

Next, the umbrella sampling in water for UMP was repeated, calculating free energy profiles for the H6–O5' distance using the adjusted $R_{\min}$ values with both OL3 and LJbb. While these two force fields had differed somewhat in the positions of the free energy minima using their original parameters, with LJbb shifted to even longer distances than OL3, the profiles for modified OL3 and LJbb become nearly identical when the Lennard–Jones function for the H6 to O5' is less repulsive (Figure 4B). This indicates that the differences between OL3 and LJbb behavior in the UMP test arise from the H6–O5' interaction, rather than the many other interactions that change with the oxygen parameter adjustments that were made in LJbb. The new locations of the free energy minima are in good agreement with the shifts observed in the gas-phase potential energy minima (Figure 4A).

MD Simulations of the Complete 7UCR RNA Structure. MD simulations on the full 7UCR system in explicit water with 200 mM KCl for the OL3 and LJbb force fields, as well as the two variations of each that include the empirical correction described above, were performed (1.0- μ s simulations repeated in triplicate for each of the 6 force field variations). 7UCR is an ideal model system due to the high precision of the experimental atomic positions. Simulations with the unmodified force fields can be analyzed to determine the extent to which the erroneously long distances preferred in the UMP model are reproduced in the full context of RNA structure, and simulations with the modified force fields test whether the improved agreement with the QM data for UMP is transferable to RNA simulations in water.

Overall, all six force field variations result in stable simulations on the microsecond time scale, with histograms of RMSD values showing a single strong peak near 1.0–1.2 Å (Figure S2). Fluctuations are somewhat higher when using standard OL3, or OL3 with $R_{\min}$ of 2.7 Å. The GAGA loop is

also stable, with RMSD histograms peaking at 0.7–0.8 Å for all force fields (Figure S3).

For each nucleotide the average RMSD value during the entire simulation, following best-fit of the entire RNA structure (Figure S4), was calculated to evaluate whether some regions of the RNA exhibit greater deviations than others. The trends in RMSD across the structure are similar for the different force fields, with most nucleotides showing small RMSD values of 0.5–1.5 Å. Unmodified OL3 performed more poorly, with one simulation showing RMSD values 1 Å larger compared to the other force fields. In all force fields, RMSD values for the loop region were ~ 2 Å, again consistent across force fields. These RMSDs were measured after fitting the entire structure; the loop RMSD after best-fit of the loop region peaks near 0.8 Å (Figure S3), suggesting that the loop position relative to the stem is moderately shifted in all simulations relative to the orientation in the crystal structure.

After concluding that the Sarcin/Ricin system is highly stable in these microsecond-length MD simulations, the ability of the force fields to reproduce the short distances between HC and O atoms seen in the crystal structure was evaluated in the model system. Distances between H6 (U, C) atoms that are near O5' atoms were measured in 26 high-resolution (<1.5 Å) structures of the Sarcin/Ricin loop in the PDB (9 pairs less than 2.8 Å). As shown in Figure 5, the experimental distance

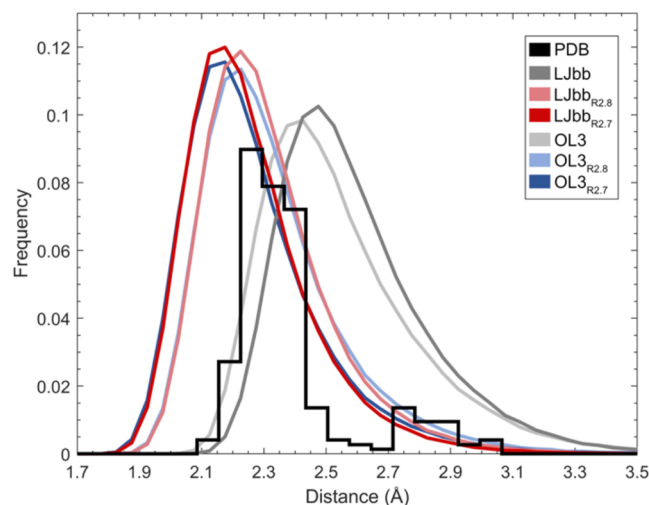


Figure 5. Histograms of distances between H6 and O5' atoms in the Sarcin/Ricin loop, including the PDB (black staircase) and MD simulations (smooth curves). PDB data includes 26 high-resolution crystal structures of this system, and MD data for each parameter set aggregates the 3 independent runs of 1.0 μ s.

distribution is strongly peaked near 2.2 Å, consistent with the analysis of diverse RNA crystal structures in Figure 2. Despite the more complex RNA structure environment in 7UCR, and inclusion of water and KCl, the MD results (Figure 5) are remarkably consistent with those obtained from the UMP model system (Figure 4). The unmodified OL3 and LJbb force fields lead to H6–O5' distances that are shifted to values that are ~ 0.2 Å longer than in the PDB structure, with LJbb even longer due to the increased oxygen size. In contrast, simulations using the QM-based $R_{\min}$ for CH $\cdots$ O pairs show near-quantitative agreement with the crystal structure, with the expected additional broadening due to thermal fluctuations. Unlike their standard versions, modified OL3 and LJbb are in

close agreement when the same $R_{\min}$ value is used, suggesting that the deviation between standard OL3 and LJbb in Figure 4 arises solely from the overly repulsive interaction between the H6/H8 and O5' atoms. As expected, the histograms obtained using $R_{\min}$ of 2.7 are shifted to slightly lower values than $R_{\min}$ of 2.8. While both are in good agreement with the crystal structure histogram, this test on 7UCR lacks the sensitivity to determine which value provides better accuracy in simulations.

Testing the Impact of Parameter Modification on the Stability of RNA Stem-Loops in MD. Simulations for three additional RNA stem-loop systems (Figure 6) were carried out,

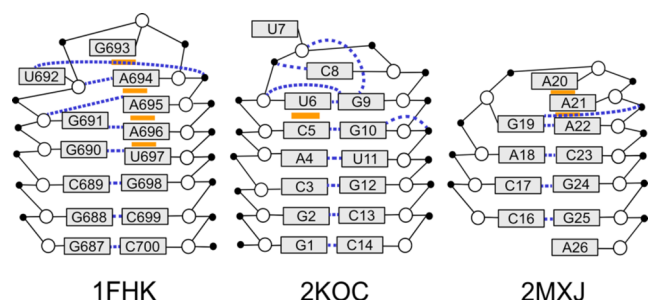


Figure 6. Sequence and topology of the RNA stem-loops studied herein. Numbering reflects biological numbering in the original structures (blue dotted line = hydrogen bonds; orange dash = stacking interactions; circle open = ribose; ● = phosphate; rectangle open = base).

including the solution structure of the *Escherichia coli* 16S rRNA 690 hairpin (PDBID: 1FHK⁷⁴), a model thermostable 14-mer cUUCGg stem-loop (PDBID: 2KOC⁷⁵) and the 3' splice site in influenza A (PDBID: 2MXJ⁷⁶). These systems were selected to explore the impact of the force field modifications on multiple common RNA loop motifs, providing a more reliable evaluation of the transferability of the changes across different RNA structure types. Simulations of each hairpin system were carried out in each of the six force field variations used above. For each, eight independent simulations of 2.0 μ s in length were generated. It remains intractable to generate fully converged equilibrium ensembles for RNA systems of this size; therefore the ability of each force field to reproducibly maintain stable structures during MD simulations initiated from the NMR-based coordinates (probing kinetic, rather than thermodynamic stability) was the focus of this study. For example, it has been reported that it is exceptionally challenging to generate stable MD simulations for 2KOC,⁹³ despite the high stability of the UUCG tetraloop in experiments.⁹⁴ Here, the adjustment of the $R_{\min}$ for CH $\cdots$ O pairs to better match QM reference data leads to significant improvement in the stability of the RNA hairpins during MD simulation.

First the overall stability of each system was analyzed with RMSD calculations. For each stem-loop and force field modification, RMSDs from the experimental structure over time for the eight independent simulations (Figure 7, left) were monitored. These indicate how many independent simulations maintained the experimental structure for each force field. For each stem-loop, Figure 7, right overlays RMSD histograms (merging all eight replicas) for each setup, allowing a more direct comparison of the accuracy of structures sampled using different force field variations.

The unmodified force fields (OL3 and LJbb) showed difficulty maintaining the initial NMR models; for all three stem-loops, neither of these force fields was able to maintain a stable structure in all eight replicates. As discussed below, this stability was markedly improved when the CH $\cdots$ O repulsion was adjusted, achieving reproducible, stable simulations with no further adjustments to the force field. Data to support the discussion below can be found in Figure 7.

1FHK. Five of the eight simulations of 1FHK using OL3 unfolded within 2.0 μ s ($\text{RMSD} > 5$ Å), and the others showed high flexibility, sampling RMSD values near 1 Å, but also reversibly sampling peaks near 1.5 and 3 Å. 1FHK with LJbb is somewhat improved, but still was unable to maintain the experimental structure and multiple simulations led to unfolding. The stability of the simulations further improved with both OL3_{R2.8} and LJbb_{R2.8}, sampling a main RMSD peak near 1 Å with smaller peaks up to 3 Å. Only one of the 16 simulations led to irreversible unfolding (with OL3_{R2.8}). LJbb_{R2.7} was similar to LJbb_{R2.8}, showing no unfolding, but frequent reversible excursions to 3 Å. The greatest stability was observed with OL3_{R2.7}; no unfolding was observed, and the transient sampling of RMSD values over 2 Å was infrequent in all eight runs.

2KOC. Once again, the results for the UUCG tetraloop with unmodified OL3 and LJbb were unsatisfactory. With OL3, only one of the eight simulations maintained the experimental structure, with the remaining seven unfolding. LJbb again was somewhat better, with three of eight replicates unfolding. When they stayed folded, the structure using OL3 or LJbb was a reasonable match to the NMR model, with a folded RMSD peak near 1.2 Å. OL3_{R2.8} and LJbb_{R2.8} were markedly improved over their unmodified versions, with only 2 and 1 simulations unfolding, respectively. Additional improvement was observed in simulations using OL3_{R2.7}, with all eight simulations remaining folded; six of eight remained folded with LJbb_{R2.7}. These differences between $R_{\min}$ of 2.7 or 2.8 Å may not be significant due to the overall good stability and few unfolding events. In addition to reduced unfolding, all simulations with $R_{\min}$ of 2.7 or 2.8 Å resulted in a highly accurate folded state with an RMSD peak near 0.8 Å; these are lower than the peak near 1.2 Å sampled by both OL3 and LJbb.

2MXJ. The third stem-loop (with GNRA tetraloop) also unfolded at least once for both standard force fields, with four of eight simulations using OL3 leading to unfolding. As with the other two stem-loops, LJbb showed respectable improvement over OL3, with only one of eight replicates unfolding. With both unmodified force fields, the folded structure showed significant flexibility, with rapid transitions between native-like structures with RMSD peaks near 1 and 2 Å. OL3_{R2.8} again dramatically improved the stability over OL3, with one simulation unfolding, and another unfolding prior to refolding. The native-like peak near 2 Å largely disappeared, and the simulations maintained an accurate native structure with an RMSD peak near 0.9 Å. OL3_{R2.7} performs even better, with no unfolding events and one replicate sampling a transient increase in RMSD to ~ 2.0 Å before returning below 1.0 Å. LJbb_{R2.8} and LJbb_{R2.7} both provided highly stable simulations in all replicates, with a single peak near 0.9 Å in the RMSD histogram.

Next, all three RNA stem-loop structures were assessed for characteristic structural features (Figure 8 and Tables S5–S7). While the RMSD from the experimental structures can be small (e.g., 0.8 Å on average for 2KOC with $R_{\min} = 2.7$ or 2.8

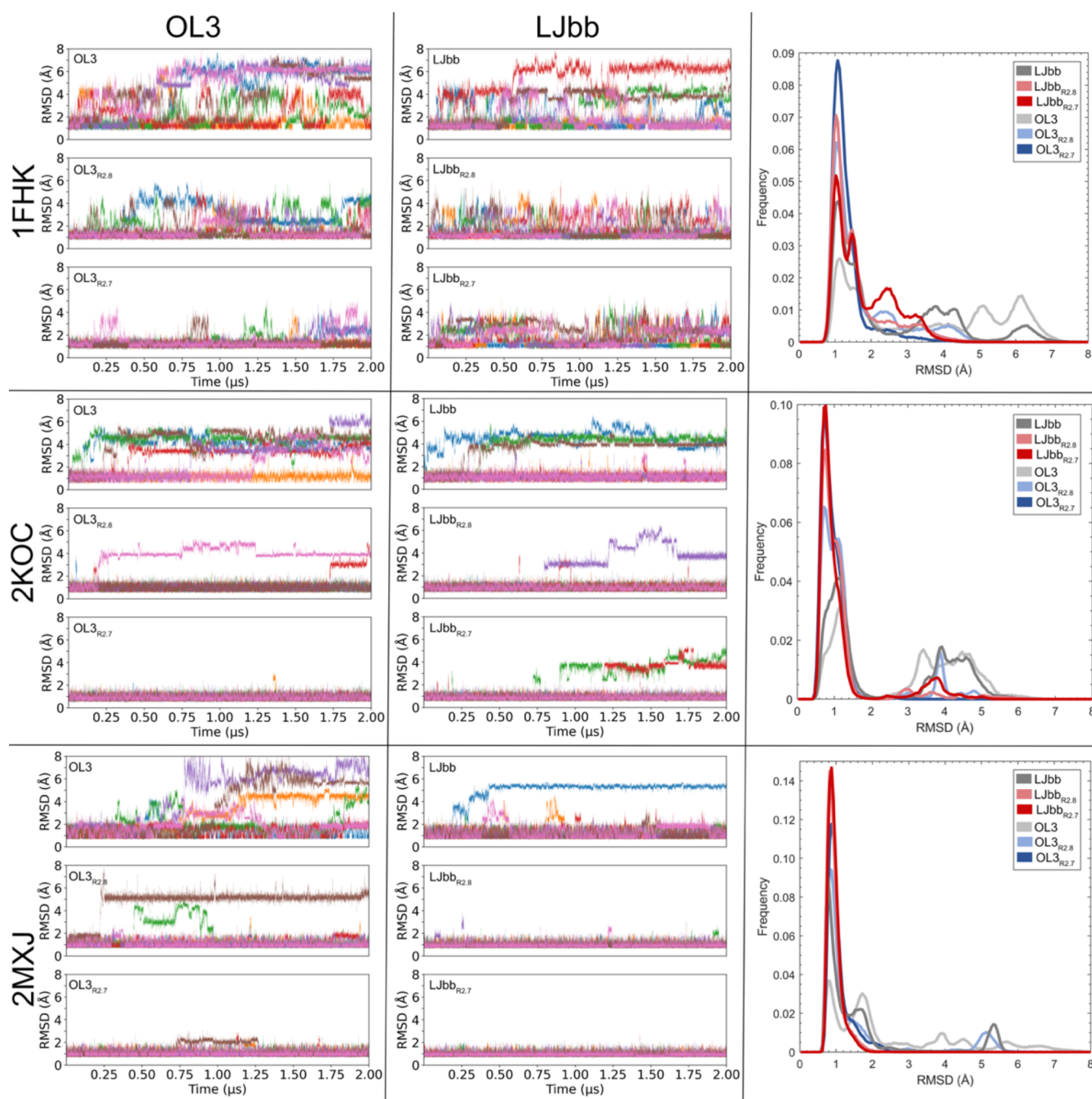


Figure 7. Results of RNA stem-loop simulations. Columns represent the parent RNA force field (OL3 and LJbb), and rows represent the 3 RNA stem-loops (PDBID: 1FHK, 2KOC, and 2MXJ). In each, 3 panels provide results from the force field using different $R_{\min}$ for CH $\cdots$ O pairs: default, 2.8 and 2.7 Å (top to bottom). Inside each of these, different color lines provide the RMSD plotted versus time for eight independent runs. (far right) Histograms of RMSD values sampled during the same MD simulations of the RNA hairpin systems. For each hairpin, multiple histograms are shown; each histogram represents combined data from 8 replicates using that force field variation.

modifications, Figure 7), there could still be subtle differences that are not captured such as loss of a base stacking interaction or hydrogen bond due to a slight shift in the backbone.

The solution structure of *E. coli* 690 loop of 16S rRNA (PDBID: 1FHK⁷⁴) is striking due to the presence of a five-base single-stranded stack (G7 to U11). It also notably contains two unusual sheared base pairs G4:U11 and G5:A10. A complex set of backbone-base hydrogen bonds (4–6, Figure 8A) rounds out the U-turn motif. With the exception of the G5 O2' hydrogen bonding with A9 amino group at the 6'-position (4, Figure 8A), simulations with OL3 did sample structures with

the characteristic hydrogen bonds present between 12 and 30% of the simulation time, while those with LJbb were a bit better with hydrogen bond occupancy between 13 and 52%. Relatively low values are not surprising; as seen in the RMSD plots (Figure 7), the stem-loop unfolded in about a third of the simulations. With the modifications, presence of hydrogen bonds (1–3 and 5–6, Figure 8A) increased by 12–45 percentage points relative to the original force field, with the $R_{\min} = 2.7$ Å modification generally better than that with $R_{\min} = 2.8$ Å. With all force fields and modifications, stacking interactions (S1–S4) were present generally >50% of the

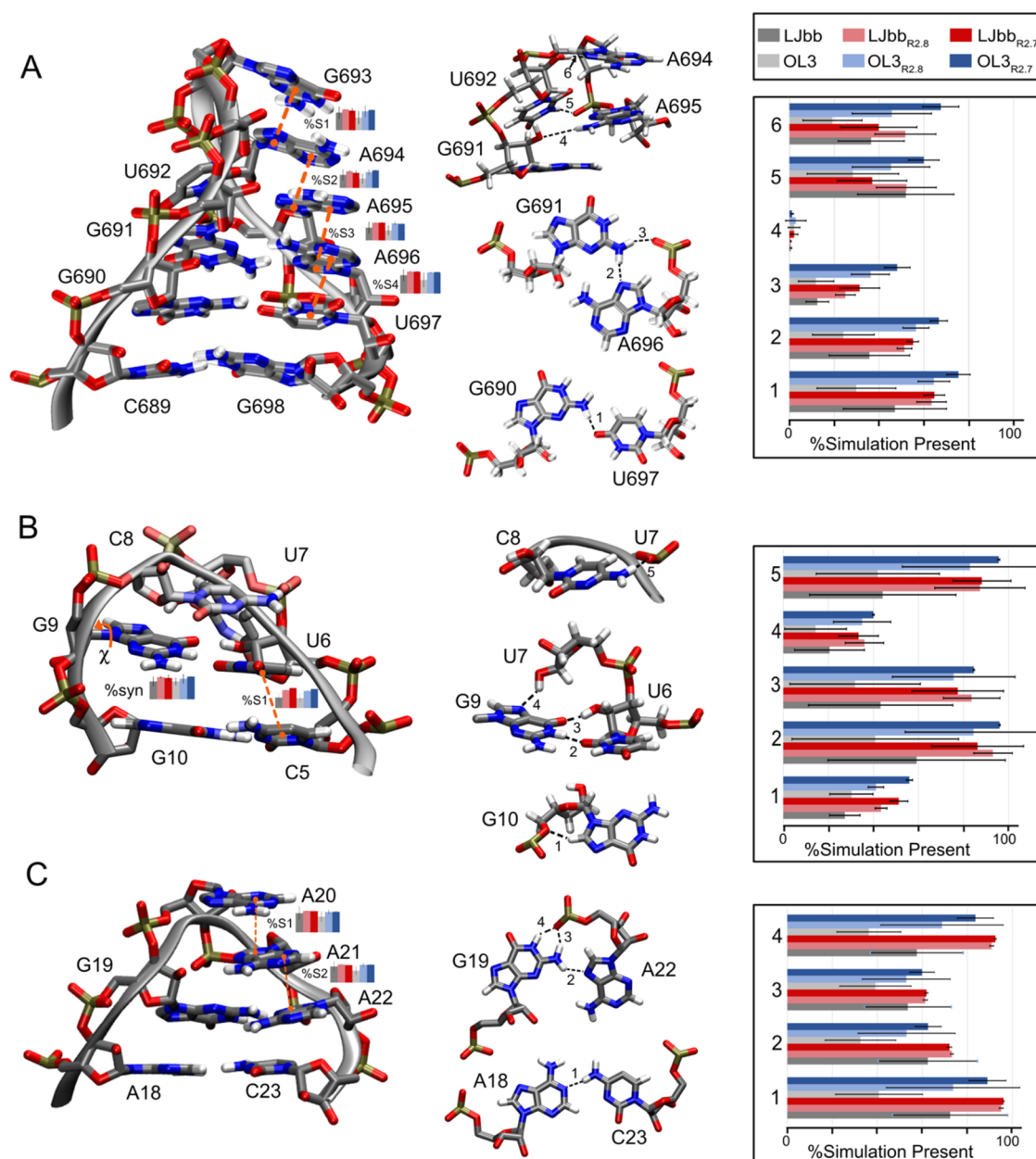


Figure 8. Presence of key structural features for the three RNA stem-loop systems during MD simulations. (A) Structure of the *E. coli* 690 loop of 16S rRNA (PDBID: 1FHK⁷⁴) including S1–S4 stacking interactions, and hydrogen bonds in the loop region (1–6). (B) Structure of the UUCG tetraloop (PDBID: 2KOC⁷⁵), including a S1 stacking interaction, the G9%*syn*, and hydrogen bonds in the loop region (1–5). (C) Structure of the 3' splice site of influenza A containing a GAAA tetraloop (PDBID: 2MXJ⁷⁶) including S1 and S2 stacking interactions, and hydrogen bonds in the loop region (1–4). Bar graphs indicating presence of the corresponding interaction are shown. Hydrogen bonds defined in the middle are shown to the far right. All values are in % simulation present and are reported as averages over all 8 trajectories. Error bars correspond to ± 1 standard deviation. Stacking interactions were calculated between either 6-membered (6m) rings or 5-membered (5m) rings of the bases indicated. Numerical values are provided in Tables S5–S7, respectively.

simulation time. However, they also universally markedly improved with modifications, with MD using OL3_{R2.7} being the most stable ($70\text{--}90\% \pm 3$). These stacking data, taken together with the very small standard deviation in the values, general stable hydrogen bonding pattern, the smaller RMSD values (Figure 7), indicate that OL3_{R2.7} results in the most stable simulation of 1FHK.

The thermostable UUCG⁹⁴ tetraloop, represented here as 2KOC,⁷⁵ contains a series of mostly backbone-base hydrogen bonds and one U-G *trans*-wobble base pair (1–5 in Figure 8B). In addition, U6 stacks on the C5 of the closing base pair

(S1) and G9 is characterized by a *syn* glycosidic conformation. Simulations with OL3 did sample the correct hydrogen bonding pattern but these were only present 30–40% of the simulation time. Similar results were observed with LJbb, but present slightly more often (20–60%). However, when LJbb or OL3 have the $R_{\min}$ modifications, loop hydrogen bonds are present between 10 and 40 percentage points more often. The G9 *syn* conformation is observed for a significant proportion of simulations conducted with both OL3 ($76 \pm 15\%$) and LJbb ($76 \pm 19\%$). However, when $R_{\min} = 2.7$ or 2.8 Å modifications are added, G9 *syn* preference is more pronounced, present for

87–96% of the simulation time on average. Base stacking interactions are also markedly increased from ~40% presence with OL3 and LJbb to 69–81% with the $R_{\min}$ modifications, with OL3_{R2.7} performing best.

Another stable⁹⁵ and prolific tetraloop motif is the GNRA class of tetraloops⁹⁶ in which N is any base and R is A or G. The GAAA tetraloop (PDBID: 2MXJ⁷⁶) is representative, and this particular loop is closed by a C:A base pair. The GAAA loop itself contains a sheared G19:A22 base pair stabilized by two backbone to G19 hydrogen bonds (2–4, Figure 8C). In simulations with OL3, characteristic hydrogen bonds are present but not often (33–41%). With OL3_{R2.7} and OL3_{R2.8}, the hydrogen bond presence is increased to 53–89%. In this system, LJbb starts out better with the baseline hydrogen bonds present 54–72% of the simulation time but with large standard deviations (19–26%) due to unfolding of multiple simulations. With LJbb_{R2.7} and LJbb_{R2.8}, these hydrogen bonds increase to as much as 96% of the simulation time and the standard deviations become $\leq 1\%$. Stacking interactions are also well populated in general (50–78%) and only improve by ~10–20 percentage points. With this system, LJbb_{R2.7} and LJbb_{R2.8} result in generally higher hydrogen bonding, the OL3_{R2.7} and OL3_{R2.8} are not appreciably different.

CONCLUSIONS

Despite many incremental yet important adjustments over the past decade, Amber-related RNA force fields have continued to be challenged by poor stability of RNA, which can include complex structural motifs. The native structures of biomolecules as observed or inferred from experiments involve a delicate balance between many terms in the force field; often, the small free energy difference between alternate conformations arises from near-cancellation of contributions that have large but opposing magnitudes. For example, the small free energy for the formation of a hydrogen bond in aqueous solution arises from opposition between the large and favorable intramolecular electrostatics when the hydrogen bond is present, and the favorable solvation of the exposed polar groups. This presents a significant challenge to force field developers, since small errors in these large individual energies can alter the free energy landscape such that native structures become unstable. This also hampers the transferability of empirical adjustments intended to improve agreement between simulation and experiment for a given system; an adjustment of the wrong force field term may work in the training system due to cancellation of error, but fail in a different context that involves a balance between a different set of interactions. A key aspect to the modern relevance of Peter Kollman's decades-old Amber force field philosophy is isolating the competing terms as much as possible by constructing a small model system that is simple yet maintains relevance to the biological context of the interaction. Subsequent physics-based parameter adjustment leads to reduced likelihood of cancellation of error, which of course must then be tested across a variety of systems to ensure good transferability.

In this work, the role of weak CH $\cdots$ O interactions in RNA structure⁵⁶ was revisited and improvements to the model were sought. Built on the insights from that work, by analyzing a set of high-resolution RNA crystal structures, it was shown that RNA CH $\cdots$ O distances frequently adopt values that lie in the repulsive region of the Lennard–Jones function for these atoms. Short CH $\cdots$ O pairs are observed for common base-phosphate interactions, but short CH $\cdots$ O pairs involving the

ribose ring are also abundant. A simple model, consisting of a single nucleotide (UMP) extracted from a high-quality experimental RNA structure, was employed to compare locations of energy minima for the RNA force fields to those obtained from QM calculations. These results supported the hypothesis that the Lennard–Jones function for these interactions is overly repulsive at distances seen in experimental RNA structures. In this work, only the energy profile for these atom pairs were altered, leaving all other interactions unchanged. This is especially important in the case of modifying parameters for atoms capable of forming hydrogen bonds, since the strength of the hydrogen bond in the force field involves a delicate balance of unfavorable van der Waals repulsion and strong electrostatic attraction. Importantly, altering only the CH $\cdots$ O pairwise term allows unambiguous adjustment of this interaction, with no possibility of degrading performance for hydrogen bonds. This targeted adjustment also facilitates the interpretation of simulation results using the modified force field, since the CH $\cdots$ O alteration is not convoluted with changes to any other interactions.

The effects of decreasing the repulsion of hydrogens involved in the RNA CH $\cdots$ O interactions in the context of two different Amber RNA force fields (OL3 and LJbb), and on four different RNA systems exhibiting a variety of biologically relevant structural motifs were examined. Results with the unmodified force fields recapitulate other reports in the literature,^{97–99} such as the rapid and reproducible unfolding of the thermostable UUCG tetraloop. In all cases, the modification improved the match between simulations and experiments; excellent transferability of the CH $\cdots$ O improvement was seen for both RNA force fields, and across the different RNA structures. MD simulations of RNA stem-loops better matched experiment with an HR modification ($R_{\min} = 2.7$) combined with OL3, and should be used in RNA simulations going forward. Similar adjustments may be important for modeling other biopolymers. Additionally, the water model OPC was chosen due to established improvement with RNA,²⁵ but these HR modifications could be transferable to other water models and should not affect ion parameters, as it is a focused change between two specific atoms. No side-effects (i.e., degradation in performance) could be identified in these biologically relevant systems. For this reason, combined with the better match to QM data for the model system, we believe that the overly long CH $\cdots$ O distances probe a fundamental weakness in the RNA force fields, rather than being a symptom of a different problem. The observation that a relatively small inaccuracy in the CH $\cdots$ O minimum energy distance can lead to notable failure of RNA simulations attests to the complexity of the balance of interactions in RNA. While this corrects a significant weakness in the Amber RNA models, further improvements are still needed and inevitably will be unearthed in future studies.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcb.4c01910>.

Methods details for X-ray structure processing; UMP energy profiles; LJ parameter modifications; explicit water system relaxation; parmed input file for the HR modification of OL3; tables of PDBIDs for X-ray crystal

structural analysis; force field parameters for Lennard–Jones for LJbb and OL3 force fields as well as new HR parameters; RNA structural analysis values for Figure 8; histograms of RMSD values for 7UCR simulations (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; Email: maria.nagan@stonybrook.edu

Authors

Lauren E. Raguetta – Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, New York 11794, United States; Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States; orcid.org/0000-0003-0183-6275

Sarah S. Gunasekera – Department of Applied Mathematics and Statistics, 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

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

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

Diksha Parwana – Biochemistry & Structural Biology Program, Stony Brook University, Stony Brook, New York 11794, United States

Qin Wu – Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton, New York 11973, United States; orcid.org/0000-0001-6350-6672

Carlos Simmerling – Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, New York 11794, United States; Department of Chemistry and Department of Applied Mathematics and Statistics, 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.jpcb.4c01910>

Author Contributions

[#]L.E.R. and S.S.G. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by funding from the National Science Foundation (award CHE- 2018427 and CHE-2050541 to M.C.N. and CTMC-1665159 to C.S.), National Institutes of Health (GM107104 to C.S.), the Panaphil Foundation, and URECA fellowships (S.G. and J.T.L.). Q.W. is supported by the Center for Functional Nanomaterials, which is a U.S. Department of Energy Office of Science User Facility, at Brookhaven National Laboratory under Contract No. DE-SC0012704. We gratefully acknowledge support from Henry and Marsha Laufer.

REFERENCES

- (1) Crick, F. H. On Protein Synthesis. *Symp. Soc. Exp. Biol.* **1958**, *12*, 138–163.
- (2) Crick, F. Central Dogma of Molecular Biology. *Nature* **1970**, *227* (5258), 561–563.
- (3) Woolhouse, M. E. J.; Brierley, L. Epidemiological Characteristics of Human-Infective RNA Viruses. *Sci. Data* **2018**, *5*, No. 180017.
- (4) Morris, K. V.; Mattick, J. S. The Rise of Regulatory RNA. *Nat. Rev. Genet.* **2014**, *15* (6), 423–437.
- (5) Goodall, G. J.; Wickramasinghe, V. O. RNA in Cancer. *Nat. Rev. Cancer* **2021**, *21* (1), 22–36.
- (6) Fire, A.; Xu, S.; Montgomery, M. K.; Kostas, S. A.; Driver, S. E.; Mello, C. C. Potent and Specific Genetic Interference by Double-Stranded RNA in *Caenorhabditis Elegans*. *Nature* **1998**, *391* (6669), 806–811.
- (7) Mattick, J. S.; Amaral, P. P.; Carninci, P.; Carpenter, S.; Chang, H. Y.; Chen, L.-L.; Chen, R.; Dean, C.; Dinger, M. E.; Fitzgerald, K. A.; et al. Long Non-Coding RNAs: Definitions, Functions, Challenges and Recommendations. *Nat. Rev. Mol. Cell Biol.* **2023**, *24* (6), 430–447.
- (8) Kavita, K.; Breaker, R. R. Discovering Riboswitches: The Past and the Future. *Trends Biochem. Sci.* **2023**, *48* (2), 119–141.
- (9) Winkler, W.; Nahvi, A.; Breaker, R. R. Thiamine Derivatives Bind Messenger RNAs Directly to Regulate Bacterial Gene Expression. *Nature* **2002**, *419* (6910), 952–956.
- (10) Cruz, J. A.; Westhof, E. The Dynamic Landscapes of RNA Architecture. *Cell* **2009**, *136* (4), 604–609.
- (11) 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.
- (12) Bevilacqua, P. C.; Blose, J. M. Structures, Kinetics, Thermodynamics, and Biological Functions of RNA Hairpins. *Annu. Rev. Phys. Chem.* **2008**, *59*, 79–103.
- (13) Thapar, R.; Denmon, A. P.; Nikonowicz, E. P. Recognition Modes of RNA Tetraloops and Tetraloop-Like Motifs by RNA-Binding Proteins. *WIREs RNA* **2014**, *5* (1), 49–67.
- (14) Nagaswamy, U.; Voss, N.; Zhang, Z.; Fox, G. E. Database of Non-Canonical Base Pairs Found in Known RNA Structures. *Nucleic Acids Res.* **2000**, *28* (1), 375–376.
- (15) Westhof, E.; Fritsch, V. RNA Folding: Beyond Watson-Crick Pairs. *Structure* **2000**, *8* (3), R55–R65.
- (16) Durant, P. C.; Davis, D. R. Stabilization of the Anticodon Stem-Loop of tRNA^{Lys} by an A+-C Base-Pair and by Pseudouridine. *J. Mol. Biol.* **1999**, *285* (1), 115–131.
- (17) Holbrook, S. R.; Cheong, C.; Tinoco, I.; Kim, S.-H. Crystal Structure of an RNA Double Helix Incorporating a Track of Non-Watson–Crick Base Pairs. *Nature* **1991**, *353* (6344), 579–581.
- (18) Auffinger, P. Molecular Dynamics Simulations of RNA Systems. In *Handbook of RNA Biochemistry*; Wiley, 2014; Chapter 32, pp 687–718.
- (19) Schlick, T.; Portillo-Ledesma, S.; Myers, C. G.; Beljak, L.; Chen, J.; Dakhel, S.; Darling, D.; Ghosh, S.; Hall, J.; Jan, M.; et al. Biomolecular Modeling and Simulation: A Prospering Multidisciplinary Field. *Annu. Rev. Biophys.* **2021**, *50* (1), 267–301.
- (20) 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.
- (21) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the RESP Model. *J. Phys. Chem. A* **1993**, *97* (40), 10269–10280.
- (22) 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.
- (23) Pérez, A.; Marchán, I.; Svozil, D.; Spöner, J.; Cheatham, T. E., III; 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.
- (24) Zgarbová, 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.
- (25) Bergonzo, C.; Cheatham, T. E., III Improved Force Field Parameters Lead to a Better Description of RNA Structure. *J. Chem. Theory Comput.* **2015**, *11* (9), 3969–3972.
- (26) Steinbrecher, T.; Latzer, J.; Case, D. A. Revised Amber Parameters for Bioorganic Phosphates. *J. Chem. Theory Comput.* **2012**, *8* (11), 4405–4412.
- (27) Yildirim, I.; Stern, H. A.; Kennedy, S. D.; Tubbs, J. D.; Turner, D. H. Reparameterization of RNA χ Torsion Parameters for the Amber Force Field and Comparison to NMR Spectra for Cytidine and Uridine. *J. Chem. Theory Comput.* **2010**, *6* (5), 1520–1531.
- (28) Aytenfis, A. H.; Spasic, A.; Grossfield, A.; Stern, H. A.; Mathews, D. H. Revised RNA Dihedral Parameters for the Amber Force Field Improve RNA Molecular Dynamics. *J. Chem. Theory Comput.* **2017**, *13* (2), 900–915.
- (29) Yang, C.; Lim, M.; Kim, E.; Pak, Y. Predicting RNA Structures via a Simple Van Der Waals Correction to an All-Atom Force Field. *J. Chem. Theory Comput.* **2017**, *13* (2), 395–399.
- (30) 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.
- (31) Kührová, P.; Mlynsky, V.; Zgarbova, M.; Krepl, M.; Bussi, G.; Best, R. B.; Otyepka, M.; Sponer, J.; Banas, P. Improving the Performance of the Amber RNA Force Field by Tuning the Hydrogen-Bonding Interactions. *J. Chem. Theory Comput.* **2019**, *15* (5), 3288–3305.
- (32) Tan, D.; Piana, S.; Dirks, R. M.; Shaw, D. E. RNA Force Field with Accuracy Comparable to State-of-the-Art Protein Force Fields. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115* (7), E1346–E1355.
- (33) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell, A. D., Jr Impact of 2'-Hydroxyl Sampling on the Conformational Properties of RNA: Update of the CHARMM All-Atom Additive Force Field for RNA. *J. Comput. Chem.* **2011**, *32* (9), 1929–1943.
- (34) 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.
- (35) MacKerell, A. D., Jr; 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.
- (36) 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 S4A7 and S4B7. *Eur. Biophys. J.* **2011**, *40* (7), 843–856.
- (37) Oostenbrink, C.; Soares, T. A.; van der Vegt, N. F.; van Gunsteren, W. F. Validation of the S3A6 GROMOS Force Field. *Eur. Biophys. J.* **2005**, *34* (4), 273–284.
- (38) 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.
- (39) Sengul, M. Y.; MacKerell, A. D., Jr. Accurate Modeling of RNA Hairpins through the Explicit Treatment of Electronic Polarizability with the Classical Drude Oscillator Force Field. *J. Comput. Biophys. Chem.* **2022**, *21* (4), 461–471.
- (40) Zhang, C.; Lu, C.; Jing, Z.; Wu, C.; Piquemal, J.-P.; Ponder, J. W.; Ren, P. AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids. *J. Chem. Theory Comput.* **2018**, *14* (4), 2084–2108.
- (41) Vangaveti, S.; Ranganathan, S. V.; Chen, A. A. Advances in RNA Molecular Dynamics: A Simulator's Guide to RNA Force Fields. *WIREs RNA* **2017**, *8* (2), No. e1396.
- (42) Thirumalai, D.; Lee, N.; Woodson, S. A.; Klimov, D. Early Events in RNA Folding. *Annu. Rev. Phys. Chem.* **2001**, *52*, 751–762.
- (43) Thirumalai, D.; Hyeon, C. RNA and Protein Folding: Common Themes and Variations. *Biochemistry* **2005**, *44* (13), 4957–4970.
- (44) Wahl, M. C.; Sundaralingam, M. C-H...O Hydrogen Bonding in Biology. *Trends Biochem. Sci.* **1997**, *22* (3), 97–102.
- (45) Brandl, M.; Lindauer, K.; Meyer, M.; Sühnel, J. C-H...O and C-H...N Interactions in RNA Structures. *Theor. Chem. Acc.* **1999**, *101* (1), 103–113.
- (46) Shen, J.; Wang, H.; Xia, Y. A Dft Study of Hydrogen Bond Interactions between Oxidative 2'-Deoxyadenosine Nucleotides and RNA Nucleotides. *Struct. Chem.* **2013**, *24* (2), 559–571.
- (47) Louise-May, S.; Auffinger, P.; Westhof, E. Calculations of Nucleic Acid Conformations. *Curr. Opin. Struct. Biol.* **1996**, *6* (3), 289–298.
- (48) Padroni, G.; Patwardhan, N. N.; Schapira, M.; Hargrove, A. E. Systematic Analysis of the Interactions Driving Small Molecule-RNA Recognition. *RSC Med. Chem.* **2020**, *11* (7), 802–813.
- (49) Hermann, T. Small Molecules Targeting Viral RNA. *WIREs RNA* **2016**, *7* (6), 726–743.
- (50) Horowitz, S.; Trievel, R. C. Carbon-Oxygen Hydrogen Bonding in Biological Structure and Function. *J. Biol. Chem.* **2012**, *287* (50), 41576–41582.
- (51) Newberry, R. W.; Raines, R. T. Secondary Forces in Protein Folding. *ACS Chem. Biol.* **2019**, *14* (8), 1677–1686.
- (52) Driver, R. W.; Claridge, T. D.; Scheiner, S.; Smith, M. D. Torsional and Electronic Factors Control the C-H...O Interaction. *Chem. - Eur. J.* **2016**, *22* (46), 16513–16521.
- (53) Nikolova, E. N.; Stanfield, R. L.; Dyson, H. J.; Wright, P. E. CH...O Hydrogen Bonds Mediate Highly Specific Recognition of Methylated CpG Sites by the Zinc Finger Protein Kaiso. *Biochemistry* **2018**, *57* (14), 2109–2120.
- (54) Derewenda, Z. S.; Hawro, I.; Derewenda, U. C-H...O Hydrogen Bonds in Kinase-Inhibitor Interfaces. *IUBMB Life* **2020**, *72* (6), 1233–1242.
- (55) Zhou, J.; Horton, J. R.; Blumenthal, R. M.; Zhang, X.; Cheng, X. *Clostridioides difficile* Specific DNA Adenine Methyltransferase Cama Squeezes and Flips Adenine out of DNA Helix. *Nat. Commun.* **2021**, *12* (1), No. 3436.
- (56) Witts, R. N.; Hopson, E. C.; Koballa, D. E.; Van Boening, T. A.; Hopkins, N. H.; Patterson, E. V.; Nagan, M. C. Backbone-Base Interactions Critical to Quantum Stabilization of Transfer RNA Anticodon Structure. *J. Phys. Chem. B* **2013**, *117* (25), 7489–7497.
- (57) Mlýnský, V.; Kührová, P.; Stadlbauer, P.; Krepl, M.; Otyepka, M.; Banáš, P.; Sponer, J. Simple Adjustment of Intracucleotide Base-Phosphate Interaction in the OL3 Amber Force Field Improves RNA Simulations. *J. Chem. Theory Comput.* **2023**, *19* (22), 8423–8433.
- (58) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14* (1), 33–38.
- (59) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084–3095.
- (60) Case, D. A.; Aktulga, H. M.; Belfon, K.; Cerutti, D. S.; Cisneros, G. A.; Cruzeiro, V. W. D.; Forouzes, N.; Giese, T. J.; Götz, A. W.; Gohlke, H.; et al. AmberTools. *J. Chem. Inf. Model.* **2023**, *63* (20), 6183–6191.
- (61) Neese, F. The ORCA Program System. *WIREs Comput. Mol. Sci.* **2012**, *2* (1), 73–78.
- (62) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate *Ab Initio* Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132* (15), No. 154104.
- (63) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32* (7), 1456–1465.
- (64) Smith, D. G. A.; Burns, L. A.; Patkowski, K.; Sherrill, C. D. Revised Damping Parameters for the D3 Dispersion Correction to Density Functional Theory. *J. Phys. Chem. Lett.* **2016**, *7* (12), 2197–2203.

- (65) Dunning, T. H., Jr. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90* (2), 1007–1023.
- (66) Burns, L. A.; Mayagotia, A. V.; Sumpter, B. G.; Sherrill, C. D. Density-Functional Approaches to Noncovalent Interactions: A Comparison of Dispersion Corrections (DFT-D), Exchange-Hole Dipole Moment (XDM) Theory, and Specialized Functionals. *J. Chem. Phys.* **2011**, *134* (8), No. 084107.
- (67) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. MP2 Energy Evaluation by Direct Methods. *Chem. Phys. Lett.* **1988**, *153* (6), 503–506.
- (68) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. Efficient Diffuse Function-Augmented Basis Sets for Anion Calculations. III. The 3-21+G Basis Set for First-Row Elements, Li–F. *J. Comput. Chem.* **1983**, *4* (3), 294–301.
- (69) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XX. A Basis Set for Correlated Wave Functions. *J. Chem. Phys.* **1980**, *72* (1), 650–654.
- (70) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13* (8), 1011–1021.
- (71) Grossfield, A. WHAM: An Implementation of the Weighted Histogram Analysis Method, Version 2.0.10 <http://membrane.urmc.rochester.edu/content/wham/>. (accessed July 22, 2024).
- (72) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, *5* (21), 3863–3871.
- (73) Sengupta, A.; Li, Z.; Song, L. F.; Li, P.; Merz, K. M., Jr. Parameterization of Monovalent Ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB Water Models. *J. Chem. Inf. Model.* **2021**, *61* (2), 869–880.
- (74) 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.
- (75) 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.
- (76) Chen, J. L.; Kennedy, S. D.; Turner, D. H. Structural Features of a 3' Splice Site in Influenza A. *Biochemistry* **2015**, *54* (21), 3269–3285.
- (77) Harp, J. M.; Lybrand, T. P.; Pallan, P. S.; Coates, L.; Sullivan, B.; Egli, M. Cryo Neutron Crystallography Demonstrates Influence of RNA 2'-OH Orientation on Conformation, Sugar Pucker and Water Structure. *Nucleic Acids Res.* **2022**, *50* (13), 7721–7738.
- (78) Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I. Y.; Berryman, J. T.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cisneros, G. A.; Cruzeiro, V. W. D. et al. *Amber22*; University of California: San Francisco, 2022.
- (79) York, D. M.; Darden, T. A.; Pedersen, L. G. The Effect of Long-Range Electrostatic Interactions in Simulations of Macromolecular Crystals: A Comparison of the Ewald and Truncated List Methods. *J. Chem. Phys.* **1993**, *99* (10), 8345–8348.
- (80) Langevin, P. Sur La the Aerie Du Mouvement Brownien. *C. R. Acad. Sci.* **1908**, *146*, 530–533.
- (81) Åqvist, J.; Wennerström, P.; Nervall, M.; Bjelic, S.; Brandsdal, B. O. Molecular Dynamics Simulations of Water and Biomolecules with a Monte Carlo Constant Pressure Algorithm. *Chem. Phys. Lett.* **2004**, *384* (4), 288–294.
- (82) 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*, 327–341.
- (83) Sindhikara, D. J.; Kim, S.; Voter, A. F.; Roitberg, A. E. Bad Seeds Sprout Perilous Dynamics: Stochastic Thermostat Induced Trajectory Synchronization in Biomolecules. *J. Chem. Theory Comput.* **2009**, *5* (6), 1624–1631.
- (84) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2015**, *11* (4), 1864–1874.
- (85) Donohue, J.; Trueblood, K. N. Base Pairing in DNA. *J. Mol. Biol.* **1960**, *2*, 363–371.
- (86) Haschemeyer, A. E.; Rich, A. Nucleoside Conformations: An Analysis of Steric Barriers to Rotation About the Glycosidic Bond. *J. Mol. Biol.* **1967**, *27* (2), 369–384.
- (87) 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.
- (88) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Miguës, A. N.; Bickel, J.; Wang, Y. Z.; 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.
- (89) 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* **2006**, *65* (3), 712–725.
- (90) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the Multimolecule and Multiconformational RESP Methodology to Biopolymers: Charge Derivation for DNA, RNA, and Proteins. *J. Comput. Chem.* **1995**, *16* (11), 1357–1377.
- (91) Oliveira, M. P.; Hünenberger, P. H. Influence of the Lennard-Jones Combination Rules on the Simulated Properties of Organic Liquids at Optimal Force-Field Parametrization. *J. Chem. Theory Comput.* **2023**, *19* (7), 2048–2063.
- (92) Bernhardt, M. P.; Nagata, Y.; van der Vegt, N. F. A. Where Lennard-Jones Potentials Fail: Iterative Optimization of Ion–Water Pair Potentials Based on *Ab Initio* Molecular Dynamics Data. *J. Phys. Chem. Lett.* **2022**, *13* (16), 3712–3717.
- (93) 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.
- (94) Sakata, T.; Hiroaki, H.; Oda, Y.; Tanaka, T.; Ikehara, M.; Uesugi, S. Studies on the Structure and Stabilizing Factor of the Cuucg Hairpin RNA Using Chemically Synthesized Oligonucleotides. *Nucleic Acids Res.* **1990**, *18* (13), 3831–3839.
- (95) Heus, H. A.; Pardi, A. Structural Features That Give Rise to the Unusual Stability of RNA Hairpins Containing GNRA Loops. *Science* **1991**, *253* (5016), 191–194.
- (96) 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.
- (97) Banáš, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham, T. E., 3rd; Sponer, J.; Otyepka, M. Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theory Comput.* **2010**, *6* (12), 3836–3849.
- (98) 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.
- (99) Kührová, P.; Best, R. B.; Bottaro, S.; Bussi, G.; Sponer, J.; Otyepka, M.; Banas, P. Computer Folding of RNA Tetraloops: Identification of Key Force Field Deficiencies. *J. Chem. Theory Comput.* **2016**, *12* (9), 4534–4548.