Adjusting the energy profile for CH–O interactions leads to improved stability of RNA stem-loop structures in MD simulations

Lauren E. Raguette[‡,1,2], Sarah S. Gunasekera[‡,3], Rebeca I. Ventura Diaz[2], Ethan Aminov[2], Jason T. Linzer[2], Diksha Parwana[4], Qin Wu[5], Carlos Simmerling[1,2,3] and Maria C. Nagan[2,*]

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

[2]Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States

[3]Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, New York 11794, United States

[4]Biochemistry & Structural Biology Program, Stony Brook University, Stony Brook, New York 11794, United States

[5]Center for Functional Nanomaterials, Brookhaven National Laboratories, Upton, NY 11973

[‡]These authors contributed equally.

*maria.nagan@stonybrook.edu

Abstract

The role of 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, we explore the role of weak CH···O interactions, 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, we demonstrate that CH···O interactions are overly repulsive in the widely-used Amber RNA force fields. We developed a simple, targeted adjustment of CH···O repulsion that leaves the remainder of the force field unchanged. We then tested the standard and modified force fields using MD simulations with explicit water and salt, amassing over 300 μsec of data for multiple RNA systems containing important features such as presence of loops, base stacking interactions as well as canonical and non-canonical base pairing. Our results demonstrate that the standard force fields lead to reproducible unfolding of the NMR-based structures, as has been reported by others. Including our 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 genomes[3] but RNA also serves as a regulatory molecule[4] 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 dependent[16] and some require water binding.[17] There are some RNA molecular dynamics studies[18] 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 field[20] 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 *ff99bsc0χOL3* force field, which will be referred to as OL3 hereafter. The phosphate oxygen van der Waals radii in OL3 have also been adjusted leading to LJbb.[25, 26] Others have modified the Amber RNA force field torsions,[27, 28] parameterized backbone oxygen van der Waals radii[29, 30], adjusted hydrogen bonding with an explicit term[31] 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 structure in a relatively flat energy landscape.[41]

Weak interactions between CH···O are known to be important to RNA structure. Early analyses of X-ray crystal structures[42, 43] and computations[44, 45] indicated C8-H8 and C6-H6 to phosphate backbone oxygen atoms were important to RNA structure. Subsequently, weak CH···O

interactions have also been found in small molecule-RNA recognition,[46, 47] biological structures in general,[48-50] and more generally in molecular recognition.[51-53] Nagan and coworkers employed solvated quantum mechanical calculations to demonstrate that C8-H8 and C6-H6 to O backbone interactions are critical for maintaining canonical transfer RNA anticodon structure.[54] It was concluded that the treatment of these backbone electrostatic interactions in atomistic force fields should be improved. Recently, an explicit base-phosphate nonbonded term was incorporated into the OL3 force field,[55] but it was also combined with a number of other changes.

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 (see Methods for details). Histograms of distances for each hydrogen-oxygen pair (**Figure 1)** indicate the pairs most frequently adopting distances closer than ~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.

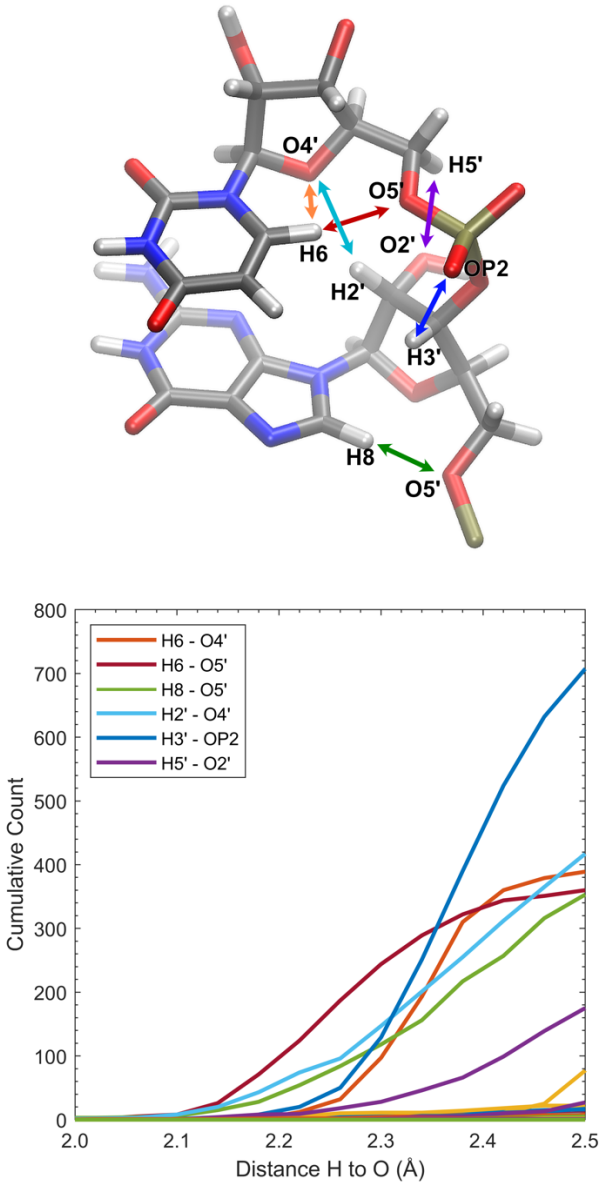

Figure 1. Cumulative histograms of distances between CH···O atom pairs in 50 RNA high-resolution crystal structures (**Table S1**). Each histogram represents data for a specific pair of H and O atoms, collected over all RNA residues in all structures. Data are shown for distances less than 2.4 Å, the 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'.

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

$$E = \varepsilon\left(\left(\frac{R_{min}}{r}\right)^{12} - 2\left(\frac{R_{min}}{r}\right)^{6}\right) \qquad \text{Eq. 1}$$

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

$$R_{min} = \sqrt[6]{2}\ \sigma \qquad \text{Eq. 2}$$

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

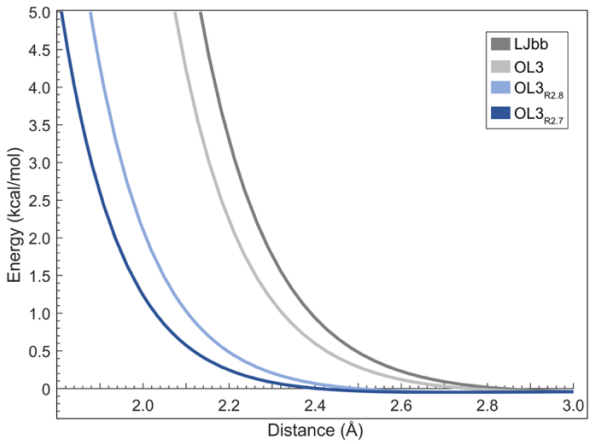

Figure 2: Lennard-Jones function energy profiles corresponding to the H6···O5' parameters for in several force fields (R_{min} = 3.0927 Å (OL3), 3.1808 Å (LJbb); ε = 0.050498 kcal/mol (all)).

Here, we introduce a targeted nonbonded adjustment specifically for CH···O interactions trained against quantum mechanical calculations. This modification is named hydrogen repulsion (HR) and indicates the importance of decreasing repulsion of hydrogens involved in the RNA CH···O parameters. Since there are many different versions to the RNA force field, only LJbb[25, 26] and OL3[24] modifications were considered in this study; although the HR modification could in principle be applied to any variation of the underlying force field. Then we examine the HR force field modification in a number of experimentally relevant RNA stem-loop systems. We show that the non-bonded 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[56]. All analyses of structures including distance, hydrogen bonding, vectors, angles, torsions, and root mean square deviations were carried out with the *cpptraj*[57] module of AmberTools[58].

Analysis of Distances in Crystal Structures. To determine the prevalence of naturally occurring CH···O interactions, high resolution crystal structures were examined. The set includes both single-stranded and double-stranded RNA, and RNA-protein complexes, with a variety of loops, canonical and non-canonical base pairing, phosphate-base hydrogen bonding, and a variety of 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), only a

single example was retained. Structures were visually examined to find instances of overlapping coordinates (such as present in 1QCU); these were removed from the list. After filtering, the 50 structures with highest resolution were retained (**Table S1**).

Collection of Sarcin/Ricin Loop Structures. The set of 26 PDB structures for the Sarcin/Ricin loop was identified using the RCSB search with polymer entity of "RNA", text matching "sarcin", and resolution under 1.5 Å (**Table S2**).

Structures in each set were downloaded and processed through *pdb4amber*, retaining only alternate location A, and `–reduce` to add any missing hydrogen atoms. Distances were calculated from these coordinates using the *cpptraj*[57] program. For each structure, distances were calculated between all pairs of oxygen atoms (OP1, OP2, O2', O3', O4', O5') and carbon-attached hydrogen atoms (H5, H6, H8, H1', H2', H3', H4', H5', H5''). Positions for all hydrogen atoms used in the analysis can be determined to reasonable accuracy from the relevant heavy atom geometries (e.g. no hydroxyl hydrogen atoms were included). Many intra-residue pairs exhibit close distances that are conformationally constrained, such as across the ribose (e.g. H3' to O4'), or are excluded from nonbonded interaction, such as those involved in an angle term (e.g. H3' to O3'). Including these distances would complicate the use of the histograms to provide insight into the possible impact of steric clash on RNA dynamics. Therefore, the crystal structure analysis removed intra-residue oxygen-hydrogen pairs except those involving base hydrogens.

UMP Model System for evaluating H6 to O5' Interaction Energies. 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. The UMP coordinates were extracted from the 7UCR PDB file due to the high-resolution of this neutron crystal experimental structure.[59] In the experimental structure, the H6 and O5' atoms of U53 are positioned 2.2 Å apart, consistent with commonly sampled distances for this atom pair in our set of crystal structures. We retained only U53, along with the O3' from C52 (**Figure 3**). Atom P of C52 was renamed OP3 with residue name OHE

(creating a 5'-phosphate). Capping hydrogen atoms were added to the OP3 and O3' using the *tleap*. For UMP only, Lennard-Jones parameters were added to the ribose and phosphate hydroxyl hydrogens (R_{min} = 0.6 Å, ε = 0.01 kcal/mol) to avoid instability from lack of Lennard-Jones interactions in the standard force field. To focus sampling on the H6 to O5' distance, additional restraints (MM) or constraints (QM) were added to maintain the orientation of the base toward the O5', and to reduce sampling of alternate rotamers that could change during distance scanning, or be slow to converge during umbrella sampling. These included dihedral restraints 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' (details provided below).

Quantum Mechanics Potential Energy Profiles for the Model System. Orca[60] 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 Å to 3.5 Å with a 0.5 Å increment. In addition to this distance, the dihedrals outside of the two rings were constrained as discussed above. These included dihedral restraints on atoms HOP3-OP3-P-O5' (-79.6166°), OP3-P-O5'-C5' (-63.3380°), P-O5'-C5'-C4' (-170.3658°), O5'-C5'-C4'-O4' (-63.1787°), O4'-C1'-N1-C6 (48.6414°), HO2'-O2'-C2'-C1'(126.2477°), and HO3'-O3'-C3'-C4' (151.6499°). The QM geometries were optimized with B3LYP-D3[61-63] with the aug-cc-pvDz[64] basis set. This is an affordable and consistent approach.[65] Energies were calculated for the optimized geometries using MP2[66]/6-311+G**[67, 68] with TightSCF.

Molecular Mechanics Energy Profiles for the Model System. The optimized QM structures were re-minimized with the MM force fields, using the AmberTools *sander* program *in vacuo* with a maximum of 1000 cycles, switching from steepest descent to conjugate gradient after 500 cycles. No periodicity was applied and the nonbonded cutoff was 12 Å. Restraints matched the constraints for the QM calculations listed above. A force constant of 100 kcal/mol•Å^2 was used for the distance and 100 kcal/mol•rad^2 for the dihedrals.

Calculation of Free Energy Profiles in Water. The UMP coordinates were loaded into *tleap* and solvated with ~3400 OPC water molecules in a truncated octahedral box with a minimum of 20 Å distance between RNA atoms and the box boundary. Other setup and the initial 4 steps of relaxation followed the protocol described below for the longer RNA systems. Beginning at step 5 of relaxation, positional restraints were replaced by harmonic restraints on dihedrals as described below in RNA simulation details. A restraint was added to the distance between the H4' and O2' atoms to reduce sugar pucker changes that could vary between windows. Positions of minima for the restraints matched values in the 7UCR structure. These restraints were included for 100 steps of minimization followed by 300 ps MD at constant NPT (1 bar, 300 K), completing the relaxation for UMP in water.

Following relaxation, a distance restraint was added between atoms H6 and O5' with a force constant of 100 kcal/(mol•Å^2). 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 at NVT and 300 K, in which the restraint minimum was varied linearly from 2.0 Å 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 (windows). Each was 1.0 ns long and carried out with a 1.0 fs time step and constant NVT (300 K). Values of the H6 to O5' distance were saved each time step. The weighted histogram analysis method (WHAM[69]) using code from Grossfield et al.[70] 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 order to alter only specific hydrogen-oxygen interactions, we directly modified the "off-diagonal" R_{min} values for the desired atom type pairs using the "changeLJpair" command in the AmberTools *parmed* program. Using this approach, parameters for interactions between these atoms and any other atoms were

unchanged. When needed, the *parmed* "addljtype" command was used to isolate the desired atom from other atoms that previously shared the same atom type. The same R_{min} was used for all modified atom type pairs. Two different R_{min} values (2.7 Å 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 **Table S1 and Table S2**. A sample *parmed* script for modifying the Amber topology files is included in **Supporting Information**.

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 model[71] was used for the solvent; K^+ and Cl^+ ions used the Li/Merz parameter set[72] trained for OPC.

Simulation models were built from the experimental structure from the PDB (1FHK [73], 2KOC [74], 2MXJ [75], 7UCR[59]), 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[76], a 10.0 Å direct space nonbonded interaction cutoff with particle mesh Ewald[77] for long-range electrostatics, a Langevin thermostat[78] with collision frequency of 1.0 ps^{-1}, a Monte Carlo barostat[79], and SHAKE[80] 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 timestep. 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•Å^2) Cartesian positional restraints on the RNA heavy atoms. The second step involved heating from 100 K 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•Å^2). The fifth step was 100 steps of minimization at NVT with only RNA backbone atoms (O3', C3', C4', C5', O5', P) restrained with a 10 kcal/(mol•Å^2) force constant. The sixth step was 100 ps MD at constant NPT with 10.0 kcal/(mol•Å^2) restraints on the RNA backbone, followed by another 100 ps MD at constant NPT with 1.0 kcal/(mol•Å^2) restraints on RNA backbone, then another 100 ps with 0.1 kcal/(mol•Å^2) restraints, followed by a final 100 ps of unrestrained MD at constant NPT. 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[81], and a 4.0 fs timestep via hydrogen mass repartitioning[82]. For 7UCR, 3 independent production simulations of 1.0 μsec 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 μsec each were generated for each of the 6 force field variations (48 simulations for each stem-loop, 288.0 μsec total).

Measurement of RNA Structural Features. All structural factors were calculated with *cpptraj*. 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 0-90° or 270-360° while *anti* included angles between 180°±90.[83, 84] 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 were <4.0 Å and the angle between the two vectors was 0±40° or 180±40°.

Results and Discussion

We first analyzed the Lennard-Jones potential function using typical RNA parameters for H and O to determine the distances at which steric clash begins. Then, we varied the CH···O distance 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, we used these energy profiles to suggest simple modifications to the Lennard-Jones function for van der Waals interactions, and tested the changes on multiple RNA systems via repeated runs of microsecond-timescale 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 where the H···O Lennard-Jones repulsion rises (**Figure 1**). 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[85-87] 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, we built a uridine monophosphate model system (UMP, **Figure 3A**) by extracting residue U53 from the PDB coordinates 7UCR[59] of the joint X-ray/neutron structure

of the Sarcin-Ricin loop RNA (see Methods). 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 3A**), as frequently observed in RNA crystal structures (**Figure 1**). By varying this distance and calculating the MM energy (here using the OL3 force field only), we can evaluate whether the force field energy is unfavorable at the distance seen in the crystal structure. Since the environment of the model system differs from the full crystal structure, we also calculated the QM energy profile for the same geometries as the MM energy scan (see Methods for details). This allows us to evaluate directly the accuracy of the MM function.

The energy profiles for varying the H6 – O5' distance in UMP are shown in **Figure 3**. 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 3B**), 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, 88] To evaluate whether an aqueous environment would change the conclusions drawn from the gas-phase potential energy profiles, we calculated 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 (**Figure 3C**, see Methods 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, the OL3 and LJbb force fields both prefer longer H6-O5' distances than seen in the crystal structures.

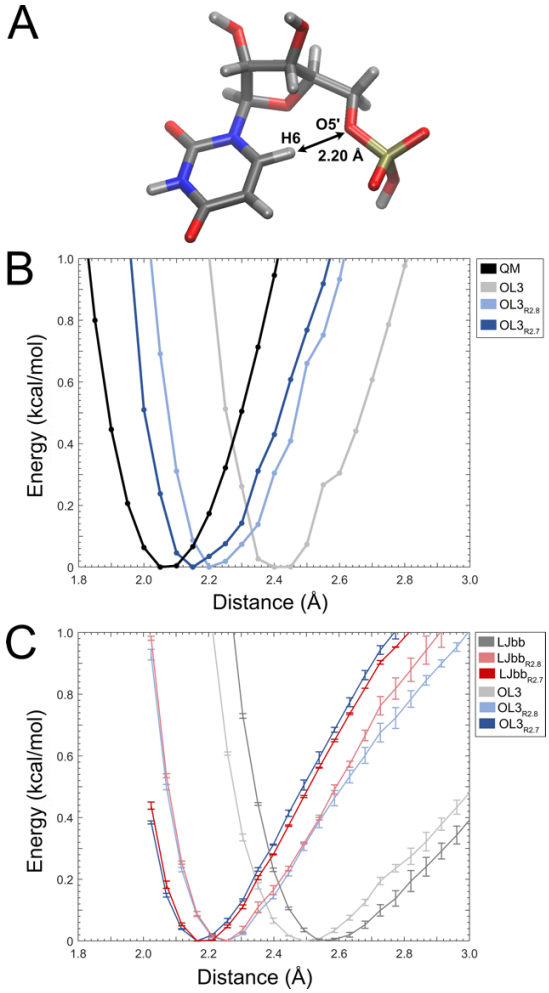

Figure 3. Energy plotted versus distance between H6 and O5' in the UMP model system. (**A**) Uridine monophosphate (UMP) model system with coordinates extracted from residue U53 in PDBID: 7UCR. The distance between atoms H6 and O5' is indicated. UMP was used for comparing force field and quantum mechanics energies, and for training modified van der Waals parameters for interaction of certain pairs of hydrogen and oxygen atoms. (**B**) 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'. (**C**) Free energy profiles from umbrella sampling simulations in explicit water at 300 K, using standard and modified OL3 and LJbb force fields.

Empirical Modification of the CH···O Lennard-Jones Parameters. The CH···O distances favored by the standard OL3/LJbb RNA force fields are consistently shifted to larger values than are optimal in the QM energy profile and measured in experimental RNA structures. This observation supports the hypothesis that the CH···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.[89, 90] 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···O energy profiles in **Figure 3**.

To overcome this short-range repulsion, we took the opportunity to empirically adjust the Lennard-Jones parameters for specific CH···O pairs in both the OL3 and LJbb force fields. 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. We decided to directly specify new "off-diagonal" Lennard-Jones parameters to be used only for CH···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···O repulsion can be reduced without any direct impact on the remainder of the RNA force field.

Here, we test the impact of adjusting Lennard-Jones R_{min} parameters for the interaction between defined pairs of hydrogen and oxygen atoms (see Methods for full details). While new

R_{min} (and ε) values could be trained directly against the QM energy profile, we believe that inconsistencies between the gas-phase QM data and MM that uses pre-polarized charges intended for solution simulations make the comparison qualitative (as we discussed in[86]). Instead, we made small empirical adjustments to the standard R_{min} values of ~3.1- 3.2 Å (see **Table S3**), testing two smaller R_{min} values of 2.8 Å and 2.7 Å throughout the remainder of this work.

We first recalculated the gas-phase MM potential energy profiles for UMP, using OL3 with reduced R_{min} values for CH···O interactions (**Figure 3B**). The adjustment resulted in the expected shift of the energy minimum to shorter values, approaching the distance of the QM energy minimum.

Next, we repeated the umbrella sampling in water for UMP, 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 3C**). 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 3B**).

MD Simulations of the Complete 7UCR RNA Structure. We next performed 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 (1.0-μsec 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 timescale, with histograms of RMSD values showing a single strong peak near 1.0 – 1.2 Å (**Figure S1**). 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 S2**).

We next calculated for each nucleotide the average RMSD value during the entire simulation, following best-fit of the entire RNA structure (**Figure S3**). This evaluates 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 S2**), 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, we focused on our goal for this model system of evaluating the ability of the force fields to reproduce the short distances between HC and O atoms seen in the crystal structure. We measured distances between H6 (U, C) atoms that are near O5' atoms 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 4**, the experimental distance distribution is strongly peaked near 2.2 Å, consistent with the analysis of diverse RNA crystal structures in **Figure 1**. Despite the more complex RNA structure environment in 7UCR, and inclusion of water and KCl, the MD results (**Figure 4**) are remarkably consistent with those obtained from the UMP model system (**Figure 3**). 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···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.

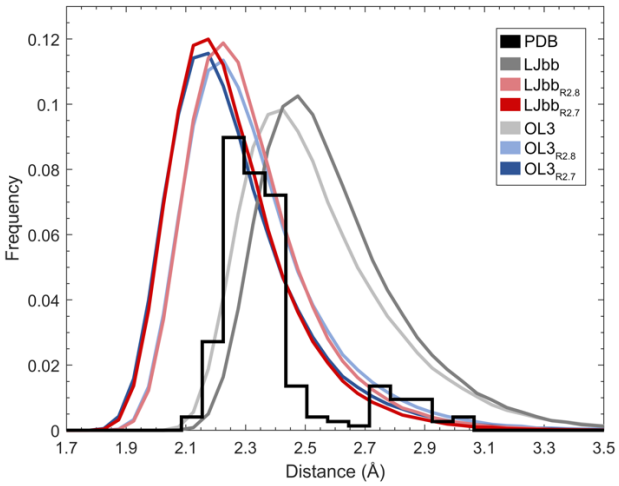

Figure 4. Histograms of distances between H6 and O5' atoms in the Sarcin/Ricin loop, including the PDB (black stairstep)and MD simulations (smooth curves). PDB data includes 26 high-resolution crystal structures of this system, and MD data for each parameter set aggregates the 3 independent runs of 1.0 μs.

Testing the Impact of Parameter Modification on the Stability of RNA Stem-loops in MD.

We carried out simulations for three additional RNA stem-loop systems (**Figure 5**), including

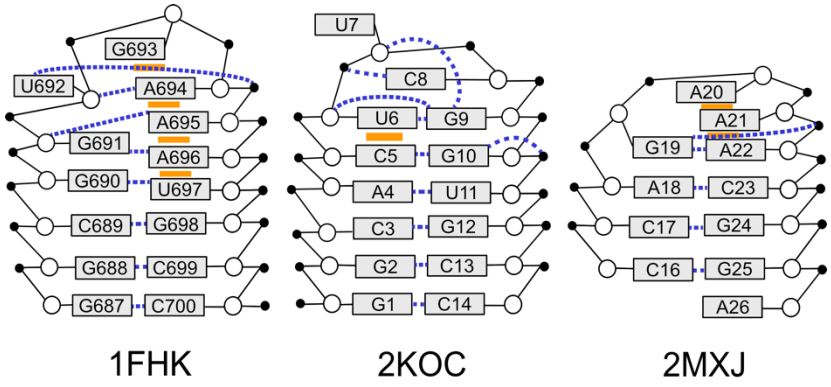

Figure 5. Sequence and topology of the RNA stem-loops studied herein. Numbering reflects biological numbering in the original structures (= hydrogen bonds; = stacking interactions; O=ribose; ●= phosphate; = base).

the solution structure of the *Escherichia coli* 16S ribosomal RNA 690 hairpin (PDBID: 1FHK[73]), a model thermostable 14-mer cUUCGg stem-loop (PDBID: 2KOC[74]) and the 3' splice site in influenza A (PDBID: 2MXJ[75]). 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. We carried out simulations of each hairpin system in each of the six force field variations used above. For each, eight independent simulations of 2.0 μsec in length were generated. It remains intractable to generate fully converged equilibrium ensembles for RNA systems of this size; therefore we focus on 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). For example, it has been reported that it is exceptionally challenging to generate stable MD simulations for 2KOC,[91] despite the high stability of the UUCG tetraloop in experiments.[92] Here, we show that adjustment of the R_{min} for CH···O pairs to better match QM reference data leads to significant improvement in the stability of the RNA hairpins during MD simulation.

We first analyzed the overall stability of each system using RMSD calculations. For each stem-loop and force field modification, RMSDs from the experimental structure over time for the eight independent simulations (**Figure 6, left**) were monitored. These indicate how many independent simulations maintained the experimental structure for each force field. For each stem-loop, **Figure 6, 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 we discuss below, this instability was markedly improved when the CH···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 6**.

1FHK. Five of the eight simulations of 1FHK using OL3 unfolded within 2.0 μsec (RMSD >5 Å), and the others showed high flexibility, sampling RMSD values near 1 Å, but also reversibly sampling peaks near 1.5 Å 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 sixteen 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

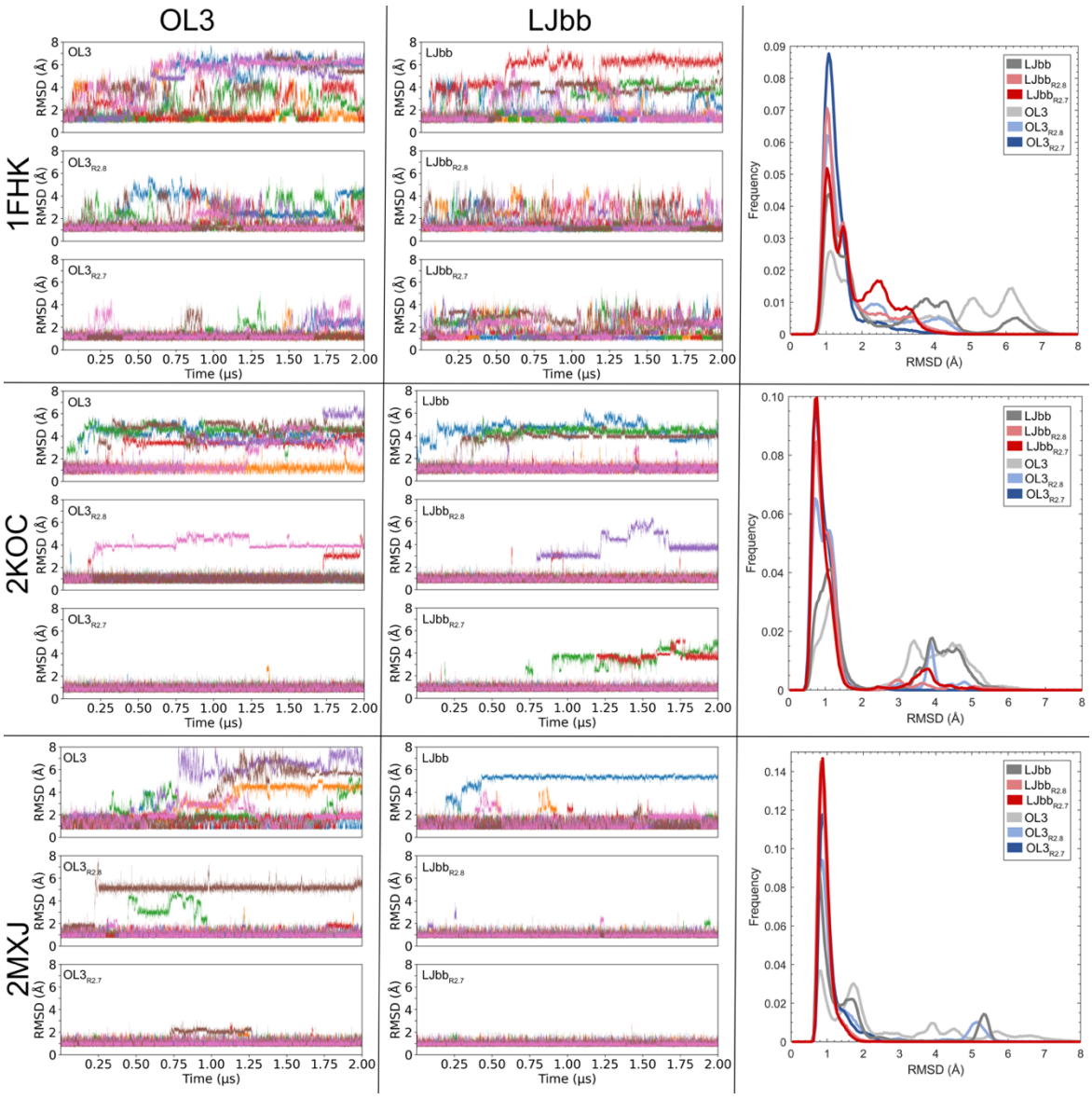

Figure 6. Results of RNA stem-loop simulations. Columns represent the parent RNA force field (OL3 and LJbb), and rows represent the 3 RNA stem-loops (PDBID: 1FHK, 2KOC, and 2MXJ). In each, 3 panels provide results from the force field using different R_{min} for CH···O pairs: default, 2.8 Å and 2.7 Å (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.

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 Å 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 Å 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 7** 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 modifications, **Figure 6**), 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: 1FHK[73]) is striking due to 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,

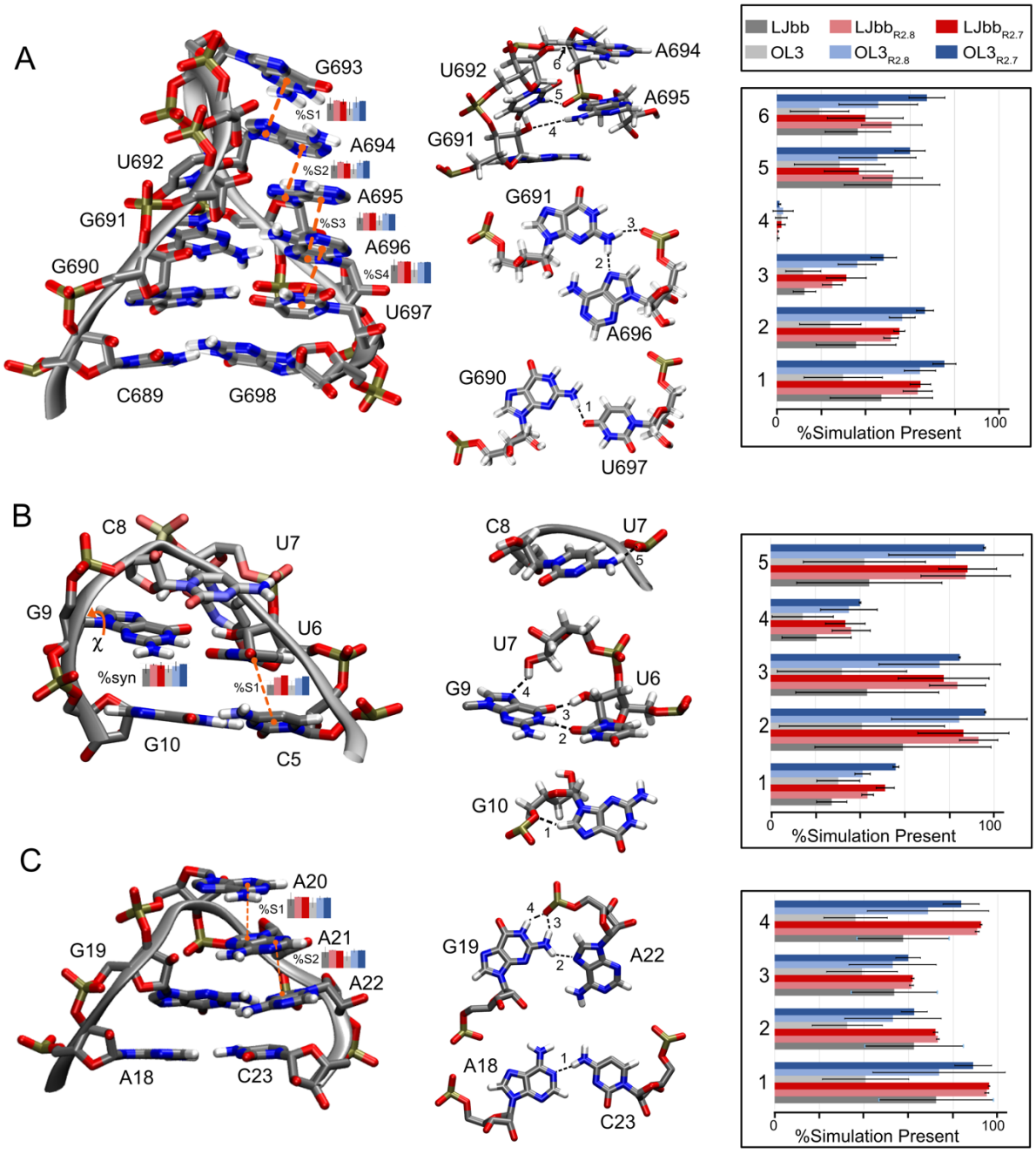

Figure 7. Presence of key structural features for the three RNA stem-loop systems during MD simulations. A) Structure of the *Escherichia coli* 690 loop of 16S rRNA (PDBID: 1FHK[73]) including S1-S4 stacking interactions, and hydrogen bonds in the loop region (1-6). B) Structure of the UUCG tetraloop (PDBID: 2KOC[74]), 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: 2MXJ[75]) 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.

Figure 7A) 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 7A**), simulations with OL3 did sample structures with the characteristic hydrogen bonds present between 12-30% of the simulation time, while those with LJbb were a bit better with hydrogen bond occupancy between 13-52%. Relatively low values are not surprising; as seen in the RMSD plots (**Figure 6**), the stem-loop unfolded in about a third of the simulations. With the modifications, presence of hydrogen bonds (1-3 and 5-6, **Figure 7A**) 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 simulation time. However, they also universally markedly improved with modifications, with MD using $OL3_{R2.7}$ being the most stable (70-90%±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 6**), indicate that $OL3_{R2.7}$ results in the most stable simulation of 1FHK.

The thermostable UUCG[92] tetraloop, represented here as 2KOC,[74] contains a series of mostly backbone-base hydrogen bonds and one U-G *trans*-wobble base pair (1-5 in **Figure 7B**). 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 10-40 percentage points more often. The G9 *syn* conformation is observed for a significant proportion of simulations conducted with both OL3 (76 ±15%) and LJbb (76±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 stable[93] and prolific tetraloop motif is the GNRA class of tetraloops[94] in which N is any base and R is A or G. The GAAA tetraloop (PDBID: 2MXJ[75]) 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 7C**). 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 ≤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, we revisited the role of weak CH···O interactions in RNA structure [54] and sought to improve the model. We built on the insights from that work by analyzing a set of high-resolution RNA crystal structures, demonstrating that RNA CH···O distances frequently adopt values that lie in the repulsive region of the Lennard-Jones function for these atoms. Short CH···O pairs are observed for common base-phosphate interactions, but short CH···O pairs involving the ribose ring are also abundant. We next built a simple model using a single nucleotide (UMP), with a conformation extracted from a high-quality experimental RNA structure. Comparing locations of energy minima for the RNA force fields to those obtained from QM calculations supported the hypothesis that the Lennard-Jones function for these interactions is overly repulsive at distances seen in experimental RNA structures. We altered only the energy profile for these atom pairs,

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. Altering only the CH···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···O alteration is not convoluted with changes to any other interactions.

We examined the effects of decreasing the repulsion of hydrogens involved in the RNA CH···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. Results with the unmodified force fields recapitulate other reports in the literature,[95-97] 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···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. 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···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···O minimum energy distance can lead to notably 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 simulations.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org.

- 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 7, histograms of RMSD values for 7UCR simulations.

Acknowledgements

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 CS), the Panaphil Foundation, and URECA fellowships (S.G. and J.T.L.). 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, DOI: 10.1038/227561a0.

(3) Woolhouse, M. E. J.; Brierley, L. Epidemiological characteristics of human-infective RNA viruses. *Sci. Data* **2018**, *5*, 180017, DOI: 10.1038/sdata.2018.17.

(4) Morris, K. V.; Mattick, J. S. The rise of regulatory RNA. *Nat. Rev. Genet.* **2014**, *15* (6), 423-437, DOI: 10.1038/nrg3722.

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

(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, DOI: 10.1038/35888.

(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.; Gingeras, T. R.; Guttman, M.; Hirose, T.; Huarte, M.; Johnson, R.; Kanduri, C.; Kapranov, P.; Lawrence, J. B.; Lee, J. T.; Mendell, J. T.; Mercer, T. R.; Moore, K. J.; Nakagawa, S.; Rinn, J. L.; Spector, D. L.; Ulitsky, I.; Wan, Y.; Wilusz, J. E.; Wu, M. Long non-coding RNAs: Definitions, functions, challenges and recommendations. *Nat. Rev. Mol. Cell Biol.* **2023**, *24* (6), 430-447, DOI: 10.1038/s41580-022-00566-8.

(8) Kavita, K.; Breaker, R. R. Discovering riboswitches: The past and the future. *Trends Biochem. Sci.* **2023**, *48* (2), 119-141, DOI: 10.1016/j.tibs.2022.08.009.

(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, DOI: 10.1038/nature01145.

(10) Cruz, J. A.; Westhof, E. The dynamic landscapes of RNA architecture. *Cell* **2009**, *136* (4), 604-609, DOI: 10.1016/j.cell.2009.02.003.

(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, DOI: 10.1021/ar200098t.

(12) 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.

(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) 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, DOI: 10.1093/nar/28.1.375.

(15) Westhof, E.; Fritsch, V. RNA folding: Beyond watson-crick pairs. *Structure* **2000**, *8* (3), R55-R65, DOI: 10.1016/S0969-2126(00)00112-X.

(16) Durant, P. C.; Davis, D. R. Stabilization of the anticodon stem-loop of tRNALys,3 by an A+-C base-pair and by pseudouridine. *J. Mol. Biol.* **1999**, *285* (1), 115-131, DOI: 10.1006/jmbi.1998.2297.

(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, DOI: 10.1038/353579a0.

(18) Auffinger, P. Molecular dynamics simulations of RNA systems. In *Handbook of RNA biochemistry*, 2014; 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.; Liang, E.; Saju, S.; Vohr, M.; Wu, C.; Xu, Y.; Xue, E. Biomolecular modeling and simulation: A prospering multidisciplinary field. *Annu. Rev. Biophys.* **2021**, *50* (1), 267-301, DOI: 10.1146/annurev-biophys-091720-102019.

(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, DOI: 10.1021/ja00124a002.

(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.* **1993**, *97* (40), 10269-10280, DOI: 10.1021/J100142a004.

(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, DOI: 10.1080/07391102.1999.10508297.

(23) 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 α/γ conformers. *Biophys. J.* **2007**, *92* (11), 3817-3829, DOI: 10.1529/biophysj.106.097782.

(24) 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.

(25) 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.

(26) Steinbrecher, T.; Latzer, J.; Case, D. A. Revised amber parameters for bioorganic phosphates. *J. Chem. Theory Comput.* **2012**, *8* (11), 4405-4412, DOI: 10.1021/ct300613v.

(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, DOI: 10.1021/ct900604a.

(28) Aytenfisu, 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, DOI: 10.1021/acs.jctc.6b00870.

(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, DOI: 10.1021/acs.jctc.6b00808.

(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, DOI: 10.1021/acs.jctc.3c01164.

(31) Kuhrova, 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, DOI: 10.1021/acs.jctc.8b00955.

(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, DOI: 10.1073/pnas.1713027115.

(33) 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.

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

(35) 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.

(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 54a7 and 54b7. *Eur. Biophys. J.* **2011**, *40* (7), 843-856, DOI: 10.1007/s00249-011-0700-9.

(37) 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.

(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, DOI: 10.1002/jcc.20193.

(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. Chem. Theory Comput.* **2022**, *21* (4), 461-471, DOI: 10.1142/s2737416521420060.

(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, DOI: 10.1021/acs.jctc.7b01169.

(41) Vangaveti, S.; Ranganathan, S. V.; Chen, A. A. Advances in RNA molecular dynamics: A simulator's guide to RNA force fields. *Wiley Interdiscip. Rev. RNA* **2017**, *8* (2), e1396, DOI: 10.1002/wrna.1396.

(42) Wahl, M. C.; Sundaralingam, M. C-H...O hydrogen bonding in biology. *Trends Biochem. Sci.* **1997**, *22* (3), 97-102, DOI: 10.1016/s0968-0004(97)01004-9.

(43) 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, DOI: 10.1007/s002140050415.

(44) 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, DOI: 10.1007/s11224-012-0108-x.

(45) Louise-May, S.; Auffinger, P.; Westhof, E. Calculations of nucleic acid conformations. *Curr. Opin. Struc. Biol.* **1996**, *6* (3), 289-298, DOI: 10.1016/S0959-440X(96)80046-7.

(46) 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, DOI: 10.1039/d0md00167h.

(47) Hermann, T. Small molecules targeting viral RNA. *Wiley Interdiscip. Rev. RNA* **2016**, *7* (6), 726-743, DOI: 10.1002/wrna.1373.

(48) Horowitz, S.; Trievel, R. C. Carbon-oxygen hydrogen bonding in biological structure and function. *J. Biol. Chem.* **2012**, *287* (50), 41576-41582, DOI: 10.1074/jbc.R112.418574.

(49) Newberry, R. W.; Raines, R. T. Secondary forces in protein folding. *ACS Chem. Biol.* **2019**, *14* (8), 1677-1686, DOI: 10.1021/acschembio.9b00339.

(50) 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, DOI: 10.1002/chem.201602905.

(51) 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, DOI: 10.1021/acs.biochem.8b00065.

(52) Derewenda, Z. S.; Hawro, I.; Derewenda, U. C─H···O hydrogen bonds in kinase-inhibitor interfaces. *IUBMB Life* **2020**, *72* (6), 1233-1242, DOI: 10.1002/iub.2282.

(53) 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), 3436, DOI: 10.1038/s41467-021-23693-w.

(54) 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, DOI: 10.1021/jp400084p.

(55) 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.

(56) Humphrey, W.; Dalke, A.; Schulten, K. Vmd: Visual molecular dynamics. *J. Mol. Graph. Model.* **1996**, *14* (1), 33-38, DOI: 10.1016/0263-7855(96)00018-5.

(57) 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, DOI: 10.1021/ct400341p.

(58) 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.

(59) 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, DOI: 10.1093/nar/gkac577.

(60) Neese, F. The orca program system. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2* (1), 73-78, DOI: 10.1002/wcms.81.

(61) 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), 154104, DOI: 10.1063/1.3382344.

(62) 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, DOI: 10.1002/jcc.21759.

(63) 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, DOI: 10.1021/acs.jpclett.6b00780.

(64) 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, DOI: 10.1063/1.456153.

(65) Burns, L. A.; Mayagoitia, Á. 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), 084107, DOI: 10.1063/1.3545971.

(66) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. MP2 energy evaluation by direct methods. *Chem. Phys. Lett.* **1988**, *153* (6), 503-506, DOI: 10.1016/0009-2614(88)85250-3.

(67) 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, DOI: 10.1002/jcc.540040303.

(68) 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, DOI: 10.1063/1.438955.

(69) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The weighted histogram analysis method for free-energy calculations on biomolecules .1. The method. *J. Comput. Chem.* **1992**, *13* (8), 1011-1021, DOI: 10.1002/jcc.540130812.

(70) Grossfield, A. *WHAM: An implementation of the weighted histogram analysis method. Version 2.0.10*. http://membrane.urmc.rochester.edu/content/wham/.

(71) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building water models: A different approach. *J. Phys. Chem. Lett.* **2014**, *5* (21), 3863-3871, DOI: 10.1021/jz501780a.

(72) 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, DOI: 10.1021/acs.jcim.0c01390.

(73) 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.

(74) 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.

(75) 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, DOI: 10.1021/acs.biochem.5b00012.

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

(77) 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, DOI: 10.1063/1.465608.

(78) Langevin, P. Sur la the âorie du mouvement brownien. *C. R. Acad. Sci.* **1908**, *146*, 530-533,

(79) 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, DOI: 10.1016/j.cplett.2003.12.039.

(80) 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, DOI: 10.1016/0021-9991(77)90098-5.

(81) 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, DOI: 10.1021/ct800573m.

(82) 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, DOI: 10.1021/ct5010406.

(83) Donohue, J.; Trueblood, K. N. Base pairing in DNA. *J. Mol. Biol.* **1960**, *2*, 363-371, DOI: 10.1016/s0022-2836(60)80047-2.

(84) 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, DOI: 10.1016/0022-2836(67)90026-5.

(85) 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.

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

(87) 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, DOI: 10.1002/prot.21123.

(88) 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, DOI: 10.1002/jcc.540161106.

(89) 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, DOI: 10.1021/acs.jctc.2c01170.

(90) 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, DOI: 10.1021/acs.jpclett.2c00121.

(91) 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.

(92) Sakata, T.; Hiroaki, H.; Oda, Y.; Tanaka, T.; Ikehara, M.; Uesugi, S. Studies on the structure and stabilizing factor of the cUUCGg hairpin RNA using chemically synthesized oligonucleotides. *Nucleic Acids Res.* **1990**, *18* (13), 3831-3839, DOI: 10.1093/nar/18.13.3831.

(93) Heus, H. A.; Pardi, A. Structural features that give rise to the unusual stability of RNA hairpins containing GRNA loops. *Science* **1991**, *253* (5016), 191-194, DOI: 10.1126/science.1712983.

(94) 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.

(95) Banas, 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, DOI: 10.1021/ct100481h.

(96) 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.

(97) Kuhrova, 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, DOI: 10.1021/acs.jctc.6b00300.

TOC Graphic