Published in final edited form as:

J Chem Theory Comput. 2019 April 09; 15(4): 2734–2742. doi:10.1021/acs.jctc.9b00054.

Development and Testing of the OPLS-AA/M Force Field for RNA

Michael J. Robertson, Yue Qian, Matthew C. Robinson, Julian Tirado-Rives, and William L. Jorgensen*

Department of Chemistry, Yale University, New Haven, Connecticut 06520-8107, United States

Abstract

Significant improvements have been made to the OPLS-AA force field for modeling RNA. New torsional potentials were optimized based on DFT scans at the ω B97X-D/6–311++G(d,p) level for potential energy surfaces of the backbone α and γ dihedral angles. In combination with previously reported improvements for the sugar puckering and glycosidic torsion terms, the new force field was validated through diverse molecular dynamics simulations for RNAs in aqueous solution. Results for dinucleotides and tetranucleotides revealed both accurate reproduction of 3J couplings from NMR and the avoidance of several unphysical states observed with other force fields. Simulations of larger systems with noncanonical motifs showed significant structural improvements over the previous OPLS-AA parameters. The new force field, OPLSAA/M, is expected to perform competitively with other recent RNA force fields and to be compatible with OPLS-AA models for proteins and small molecules.

Keywords

Molecular Dynamics; Nucleoside; Nucleotide; Force Fields; RNA

Introduction

Within the central dogma of molecular biology, RNA performs a key role as intermediate in gene expression. Orthogonal to this role, RNA is also capable of folding into defined 3D-structures with functions similar to proteins. Given its varied purposes, RNA is becoming an increasingly important biomolecule for drug discovery. First, RNA itself can be targeted by small molecules, the typical example being the classes of antibiotics that target the ribosome. This continues to be a productive area, with many novel ribosome-targeting antibiotics in clinical trials.¹ Also riboswitches, natural RNA structural elements that bind small molecules, typically metabolites,² are also potential drug targets for small molecules. Compounds that bind to bacterial riboswitches and inhibit bacterial cell growth have been identified,³ opening a new avenue for antibiotic development. Moreover, RNA molecules

*Corresponding Author: william.jorgensen@yale.edu. Phone: 203-432-6278. Fax: 203-432-6299.

Supporting Information

A Word document containing a full description of the simulation details and the new parameters developed in this work. Spreadsheets of the energies for the α/γ potential energy surfaces and all original OPLS-AA parameters. These materials are available free of charge via the Internet at <http://pubs.acs.org>. CHARMM formatted topology and parameter files can be downloaded from <http://zarbi.chem.yale.edu/oplsaam.html>.

The authors declare no competing financial interest.

themselves can be used as therapeutics. Aptamers, RNA molecules that bind to another molecule, have been developed to target biomolecules, thereby providing therapeutic effects. For instance, pegaptanib is an FDA-approved drug of this type; it is an RNA aptamer that binds to vascular endothelial growth factor A (VEGFA) and is used in treating age-related macular degeneration.⁴

Given its importance, there is a need to be able to perform accurate computational modeling of RNA with molecular mechanics force fields. These are not only critical to study their structure, but ultimately to be able to apply computational structure-based drug discovery to RNA. While our group has recently updated our OPLS-AA force field for proteins^{5,6} and nucleotides/nucleosides,⁷ the parameters for RNA remain unchanged since their original development in the 1990s.⁸ Significant improvements in computational power, quantum chemical methods for generating potential energy surfaces, and solution-phase experimental data for benchmarking have been made since then. However, compared to proteins, force field development for RNA is far more challenging. This is likely the result from both its highly charged nature and the larger number of degrees of freedom in the backbone compared to proteins.

Numerous studies reporting improvements to the treatment of RNA in other force field families have been published,^{9–16} which has been reviewed extensively elsewhere.^{17,18} Briefly, several methodological approaches have been used to this end. Given the large number of dihedral angles in the backbone, performing a full multidimensional quantum-mechanical (QM) scan of the backbone dihedral potential energy surface would be prohibitively costly. This contrasts with proteins, where the backbone dihedrals ϕ and ψ can be scanned as a two-dimensional surface to capture the full conformational landscape. In the literature, different models are used for partitioning RNA dihedral angle QM scans into lower dimensional, computationally tractable, surfaces. The most common strategy is to evaluate two-dimensional surfaces for α/γ and ϵ/ζ (Figure 1), generally using multiple surfaces with different fixed values for other dihedrals.^{9–14} The β dihedral angle is generally treated as an uncoupled dihedral, as it is anti in almost every crystal structure of RNA. Empirical approaches have also been employed in parameterization, incorporating experimental data with traditional methods.^{14–16} Unlike protein force fields, where nonbonded parameters have remained largely the same for decades, RNA charges and Lennard Jones parameters have also been adjusted to improve the description of various secondary structures.^{19–21} Very recently great strides have also been made with polarizable force fields as an alternative approach to adjust interaction energies.^{22,23}

As an initial step in developing new torsional parameters for the OPLS-AA force field for RNA, molecular dynamics simulations of dinucleotides with the original force field were run first to identify deficiencies. High-level QM scans were then performed to generate α - γ two-dimensional potential energy surfaces. A Boltzmann-weighted parameterization scheme was used to generate new Fourier coefficients for the corresponding dihedral angles. Small adjustments were also made to existing ϵ and ζ dihedral parameters to accommodate the recently reported parameters for the ribose ring.⁷ Molecular dynamics simulations were then executed for dinucleotide monophosphates and tetranucleotide triphosphates with both the old OPLS-AA and new OPLS-AA/M force field.³ J couplings calculated from the dihedral

angle distributions in the simulations were compared to experimental values determined by NMR. Larger, more demanding validating simulations were also performed for loop E from *E. coli* 5S RNA, and the GAGA and UUCG tetraloops.

Methods

Quantum Mechanical Scans and Parameterization

Quantum mechanical scans of the α/γ potential energy surface were performed on the model compounds in Figure 1, with all other backbone and hydroxyl C-C-O-H dihedrals held fixed (values for these fixed angles can be found in SI Table 1). Additionally, the sugar puckering was fixed at either the C3' or C2' endo conformations. These scans were done with Gaussian 09 with the ω B97X-D/6-311++G(d,p) level of theory in vacuum.^{24,25} This combination of basis set and DFT-functional was selected due to its success for parameterizing protein torsions in OPLS-AA/M.⁵ All scans were performed in 15° increments. Equivalent molecular mechanics scans for parameterization were done with the OPLS-AA force field²⁶ and the BOSS software package.²⁷ Provided in eq 1 is the dihedral torsion portion of the force field, where ϕ is the dihedral angle and V_1 , V_2 , V_3 , V_4 are Fourier coefficients. V_4 is not needed for the present

$$E_{\text{torsion}} = \sum_i \frac{V_1^i}{2} [1 + \cos(\phi_i)] + \frac{V_2^i}{2} [1 - \cos(2\phi_i)] + \frac{V_3^i}{2} [1 + \cos(3\phi_i)] + \frac{V_4^i}{2} [1 - \cos(4\phi_i)] \quad (1)$$

cases and was set to zero, while V_1 , V_2 , and V_3 were fit in the torsion parameterization to minimize a Boltzmann-weighted error function (eq 2). A weighting temperature of 2000 K was chosen, as previous work demonstrated this to be the best choice for peptides.⁵ Here T is the weighting to bias the fitting towards low-energy regions, E_{MM} and E_{QM} are the molecular

$$\text{Error} = \sqrt{\frac{\sum_{i=0}^n (E_i^{\text{MM}} - E_i^{\text{QM}}) e^{-\frac{E_i^{\text{QM}}}{k_B T}}}{n}} \quad (2)$$

mechanics and quantum mechanics energy at that given scan point, and k_B is the Boltzmann constant. The V_i dihedral parameters were fit to both surfaces simultaneously, assigning equal weights to each surface (a comparison of the fit to these surfaces can be found in SI Table 3).

In the case of ϵ and ζ , the original OPLS-AA force field was found to perform well and accurately reproduce the 3J couplings for these dihedrals with dinucleotides. However, these parameters cannot be directly used with the new OPLS-AA/M force field, as the new angle

bending and torsion parameters for the ring are coupled to the ϵ - ζ potential energy surface, particularly for the ϵ portion. Hence, two-dimensional scans were performed with the original OPLS-AA force field for ϵ and ζ with the construct in Figure 1. New torsion parameters were then fit with the updated ribose parameters to these MM scans. The resulting parameters were adopted for OPLS-AA/M. In the case of β , quantum chemical scans were performed with the complex from Yildirim *et al.*;¹² however, these parameters performed significantly worse than the initial parameters, perhaps due to the differences between this construct and RNA. Instead, the parameters for the β dihedral angle were empirically adjusted to produce an almost entirely anti distribution.

Molecular Dynamics Simulations

All molecular dynamics simulations were performed with NAMD.²⁸ A constant temperature of 300K (with the exception of the tetranucleotides, which were simulated at 275K for compatibility with the experimental data) and a pressure of 1 atm were maintained with a Langevin thermostat using a damping coefficient of 1 ps⁻¹ and a Nose-Hoover Langevin piston barostat with a piston dampening timescale of 50 fs and period of 100 fs.^{29,30} Smoothing at 8.0–10 Å and a cutoff at 10Å were applied to the nonbonded interactions, with the long-range electrostatics treated with particle mesh Ewald. A 2 fs time-step was used with the SHAKE algorithm to constrain all bonds.³¹ Starting structures for di- and tetranucleotides were generated in an initial A-form geometry with Chimera.³² This was performed for AA, CA, AC, and CC dinucleotides, and AAAA, CCCC, and GACC tetranucleotides. The starting structure for the loop E of *E. coli* 5S RNA was the 1.5-Å crystal structure (PDB ID: 354D) with the two end residues removed;³³ while the GAGA tetraloop (PDB ID: 1ZIG)³⁴ and the UUCG tetraloop (PDB ID: 2KOC) were started from NMR structures.³⁵ All systems used cubic periodic water boxes; the numbers of explicit water molecules were *ca.* 1200, 2300, 3700, and 5100 for the dinucleotides, tetranucleotides, tetraloops, and loop E, respectively. Sensitivity of the structural results for polynucleotides to the choice of water model has been noted.³⁶ Though a thorough study of this issue is warranted with the present force field, results are reported here for the tetranucleotides in TIP4P water and the remainder in TIP3P water.³⁷ Sodium and chloride ions³⁸ were added to achieve charge neutrality. In the case of the *E. coli* 5S RNA Loop E, all the crystallographically determined magnesium ions were included in the simulation in addition to counter ions. For the dinucleotides, three MD runs were executed for 200 ns and one run for 1 μ s, each starting with different velocity assignments. For the remaining systems, one 1 μ s and two 200 ns simulations were performed.

³*J*couplings were calculated at each timestep based on the Karplus relationship using the parameters reported in Vokáčová, Z. *et al.*³⁹ from Wijimenga *et al.*,⁴⁰ Marino *et al.*,⁴¹ Yokoyama *et al.*,⁴² and Haasnoot *et al.*⁴³ The values from all parameterizations were averaged and the ensemble average over all trajectories was taken, although the results were independent of the Karplus parameters chosen (a comparison of individual values for AA is given in SI Tables 5 and 6). To calculate Nuclear Overhauser Effect distances (NOEs), for each experimentally measured atom pair, the calculated value *rMD* was obtained with eq 3, where *n* is the distance between the two atoms for a given molecular dynamics frame *i*.

$$r_{MD} = \langle r_i^{-6} \rangle^{-1/6} \quad (3)$$

Percentage of violation was calculated using the NMR observable ranges from AAAA, CCCC, and GACC.^{44,45} Clustering was performed using the K-means algorithm with the MMTSB tool set⁴⁶ at a radius between 4.1 to 5.4 Å depending upon the system.

Results

Parameterization

Initial molecular dynamics simulations were performed for the Adenosine-Adenosine (AA) dinucleotide to identify any deficiencies with the existing OPLS-AA parameters for RNA (Table 1). While 3J coupling values for the dihedral angle ϵ were accurately reproduced, results for γ and β were found to have room for improvement. Examining the experimental dihedral distributions,⁴³ γ should almost entirely occupy a gauche-plus (g^+) conformation for this system; however, both the gauche-minus (g^-) and anti (a) states had significant populations. Two quantum chemical scans of the α/γ potential were performed, one for each sugar puckering (Figure 2), and new Fourier coefficients were fit to reproduce both surfaces simultaneously. For each of these surfaces, the global minima for γ was at g^+ , consistent with the prominence of this state. Dihedral angles that were held fixed during these scans are listed in Supplementary Information Table S1. β parameters were empirically modified to increase the favorability of anti over g^+ and g^- . These dihedral parameters, combined with modified ϵ/ζ parameters to accommodate the new ribose angle bending parameters, are the modifications yielding the new OPLS-AA/M force field (Table S2).

Simulations of Oligonucleotides

The simulations with the new OPLS-AA/M force field resulted in significantly reduced RMSD between the calculated and experimental 3J couplings for all dinucleotides (Figure 3). A higher population of β in the anti state has led to significantly improved 3J couplings for this dihedral (Table 1). The experimental value of the $J(4'C, P)$ coupling at ~10 Hz implies that β should be entirely anti. γ couplings benefited from the increased population of the g^+ state, with the couplings generally being too high with the original force field. The sugar puckering parameters from prior work⁷ led to improved δ 3J couplings for these systems, further demonstrating the accuracy of these force field terms. ϵ and ζ are treated well with both force fields, while the results for α and ζ are more difficult to assess. There is no rigorous parameterization of the Karplus eq for these two angles, and experimental data is sparse. While an empirical eq has been proposed for dinucleotides,³⁹ it is somewhat limited. In particular, the experimental coupling for ζ is generally ~5.3 Hz for all of the dinucleotides. However, with the given relation this can only be achieved if ζ is entirely anti, which is inconsistent with the distribution found in crystal structures in the PDB. The population of the $g^+/a/g^-$ states for these dihedrals can be compared over the course of the simulation to data from crystal structures.⁴⁷ This approach is often used for comparing dihedral angle distributions of short peptides to coil regions of the PDB.⁵ Presently, both α and ζ have large populations of the g^- state, with smaller but non-negligible populations of

the other two dihedral values. This is consistent with the distribution of dihedrals seen in the PDB. The RMSD of 1.36 Hz for the AA dinucleotide may also be compared to the result of 1.78 Hz from the C27 CHARMM force field.¹⁰

The new force field was then benchmarked for the AAAA, CCCC, and GACC tetranucleotides. Experimentally, the predominant state for all three tetranucleotides has been found to be that of the A-form helix.^{44,45} However, these test systems have been challenging for many force fields, it is often the case that these tetranucleotides collapse into intercalated or other incorrectly stacked states in molecular dynamics simulations.^{44,48} The original OPLS-AA force field was no exception; a higher percentage of non-A-form states occurred in all three 200-ns simulations of AAAA. In the 1.0 μ s simulation, a 1–2–4 stacked state was observed 33% of the time along with brief populations of extended structures where the terminal bases faced in opposite directions. Snapshots of the A-form and 1–2–4 states of the AAAA tetranucleotide with their relative populations with OPLS-AA and OPLS-AA/M are provided in Figure 4. With OPLS-AA/M, an A-form state or an A-form state with a single base flipped was occupied 87% of the time. The backbone RMSDs between the centroid structures for the A-form cluster and 1–2–4 stacked states and the A-form reference (residues 1656–1659 of PDB: 1JJ2)⁴⁹ were 1.20 Å and 2.74 Å, respectively. A full breakdown of the observed structures and their populations found by clustering is given in Supplementary Information Figure S1.

The RMSD between experimental and calculated 3J couplings for the AAAA tetranucleotide (Table 2, Figure 5) decreased significantly to 2.11 Hz (OPLS-AA-M) from 6.19 Hz (OPLS-AA). The simulations for CCCC and GACC produced states with relative populations that appeared qualitatively similar between the original OPLS-AA and new OPLS-AA/M force fields. However, significant improvement was still found for the calculated 3J couplings with OPLS-AA/M (Figure 5). Even though the overall geometries of the states look similar with OPLS-AA, key dihedral values still deviate from their experimental values to a degree that they significantly impact the calculated 3J couplings. Images of the centroid of each cluster for all tetranucleotides with both force fields, their relative populations, calculated couplings for that structure, and the RMSD to the A-form reference are provided in Figures S1 – S3.

The A-form state with one of the nucleotide bases slipped has been observed with OPLSAA/M and other force fields, e.g., CHARMM36 and AMBER ff14.^{48,50} While this represents a less perturbed conformation than intercalated states, experimental evidence suggests that this may not be accurate, and may be an area to target if improvements are made in the future. In the combined 4.2 μ s of simulation with OPLS-AA/M and 4.2 μ s with OPLS-AA for all the tetranucleotides, no populations of intercalated states were observed. While 1 μ s has been sufficiently long to observe intercalated states with other force fields,^{44,51} further simulation is necessary to definitively assess the propensity of OPLS-AA/M to form incorrect states. To assuage concerns of insufficient sampling not allowing their formation, independent simulations were begun with the AAAA tetranucleotide in the 2–4–3, 2–1–3, 1–3–2, and 3–1–4 intercalated states. In all cases, the intercalated state fell apart and a more extended state was recovered within the process of minimization, equilibration, and the first 200 ns of dynamics. These results further establish the successful improvement of OPLS-AA/M.

NOE values can also be compared between experiment and the simulations. RMSDs between the experimental and computed interproton distances (SI Figure S4) as well as percent of NOE distances that violate the possible NMR range (Figure 6) were calculated. While the differences between the NOE predictions from OPLS-AA and OPLS-AA/M are not large, there are some reductions in both the percentage of NOEs falling outside the experimental ranges and in the RMSD between calculated and experimental distances. NOE data for this system is often studied in the literature, and, while informative, the relatively small differences in values despite the large differences in the dihedrals and occupied conformations suggests that other metrics of the quality of an RNA force field are more demonstrative of changes in different force fields. The percentages of NOE violations with OPLS-AA/M are similar to reported for other force fields.⁴¹

Noncanonical RNA Motifs

The next test for the OPLS-AA/M force field was examining noncanonical RNAs, which contain segments with unusual values for dihedral angles that are often difficult to reproduce in molecular dynamics simulations. The three systems studied were loop E of *E. coli* 5S RNA, and the GAGA and UUCG tetraloops (Figure 7). These systems were chosen from the work of Zgarbova *et al.*⁵² for the noted difficulty of recent AMBER force fields to reproduce their noncanonical dihedrals,^{53,54} particularly the $\alpha/\gamma = a/a$ state (a canonical A form helix has α/γ values of g^-/g^+).

Both OPLS-AA and OPLS-AA/M were used to model loop E of *E. coli* 5S RNA. While largely helical, this motif possesses two noncanonical $\alpha/\gamma g^-/a$ dihedral angles and several non-Watson Crick base pairs. Significant populations of $\alpha/\gamma g^-/a$ occurred with both the new and old force field in the last 50 ns of the simulations, although other canonical and non-canonical dihedral pairs were sampled. Complicating analysis of these dihedral populations is that with γ anti, the center of the dihedral angle distribution for the αg^- state is shifted to nearly -120° with OPLS-AA/M (a separate anti state still appears in a histogram plot centered at $\sim 170^\circ$). With the original OPLS-AA force field, there were often numerous subpopulations for α between -60° and -140° (Supplementary Figure SI 7 and 8). While performance for the non-canonical dihedrals was comparable for the two force fields, there were other significant differences in the overall structures. With OPLS-AA/M, the backbone RMSD values to the starting structure for the last 50 ns of the 1 μ s dynamics run varied between 1 and 3 Å, while with OPLS-AA a much higher range, 3 to 4 Å, was sampled yielding larger average errors (Figure 8). It should be noted that in the 1 μ s run, the conformation of loop E exchanges between states with lower and higher RMSDs, as seen in Figure 8, with transitions from the higher RMSD state back to the lower RMSD state and vice versa.

Examining the GAGA tetraloop, OPLS-AA/M performed well. It retained a stable stem region and the a/g^+ combination for α/γ between guanosine five (G5) and adenosine six (A6) over the course of 1 μ s. Like some recent versions of AMBER,⁵⁰ the a/a state between residues A8 and G9 was not stable. This leads to a slight repositioning of the ribose sugar at this point, which results in some twist for the A-form stem. Several of the γ dihedrals did briefly convert to the anti state at various points in the trajectories, but converted back to a g

+ state. From this it can be inferred that the penalty for the γ anti conformer is not prohibitively large so as to prevent any population being observed. It may require a very fine balance of angle, dihedral, and nonbonded terms to improve the stability of the non-canonical a/a α/γ dihedrals of the GAGA tetraloop without detriment of other conformations. With OPLS-AA/M, the A6 base flips between the experimental structure and a position pointing out to the solvent, the latter being preferred. Changes in this population, should they be necessary, would most likely need to result from adjustments in nonbonded parameters to enhance the base stacking energy. This approach has proven successful previously for other force fields.⁵⁴

Clustering was performed on the structures from the last 50 ns of the 1 μ s trajectory, and the centroids of the clusters were compared to the first structure in the NMR ensemble, although the results were similar regardless of which structure was examined. The RMSD between backbone atoms over the whole structure was 2.0 ± 0.1 Å. Alignment of the backbone atoms for just the four residues in the tetraloop yields an RMSD of 0.8 ± 0.1 Å. The original OPLS-AA force field performed much worse, with the entire GAGA tetraloop portion rearranging to a configuration that no longer resembles the correct fold. At the bottom of the stem there is a loss of base pairing and formation of incorrect base stacking interactions. The backbone RMSD was significantly higher than OPLS-AA/M, at 4.2 ± 0.2 Å for the whole structure and 2.1 ± 0.3 Å for the GAGA loop. However, in the case of OPLS-AA the alignment of only the loop region is nonsensical as it results in exceedingly poor alignment of the stem region. A comparison of the starting NMR structure and the centroid of the largest cluster from the last 50 ns of the 1 μ s trajectory for OPLS-AA/M and OPLS-AA are provided in Figure 9. Additional alignments of the structures can be found in Figure S5. Average Backbone RMSD values from the last 50 ns of the two shorter 200 ns runs were 2.2 Å with OPLS-AA/M and 2.9 Å with OPLS-AA. In the case of OPLS-AA/M, the state shown in Figure 9 is reached fairly quickly in more than one trajectory and, as shown in the 1 μ s simulation, is stable on this timescale. OPLS-AA, however, generally does not reach its fully distorted state after only 200 ns, requiring the full 1 μ s.

For the UUCG tetraloop both non-canonical motifs were stable over 1 μ s of simulation with the OPLS-AA/M force field. These motifs feature α/γ as a/g^+ between residues U6 and U7 and α/γ as g^+/a between residues C8 and G9. In fact, the UUCG motif was essentially stable over this timescale. The average backbone RMSD between the initial NMR structure and the clustered structures of the last 50 ns from 1 μ s of dynamics averaged 0.3 ± 0.1 Å for the four residues in the UUCG motif. The whole system, however, had a significantly higher backbone RMSD at 3.2 ± 0.3 Å due to fraying at the end of the stem region. Fraying of RNA and DNA is a phenomenon observed both experimentally and computationally, although it has been noted that there are both accurate and inaccurate states of fraying observed in simulations.⁵⁶ The deviation begins with the third base pair from the bottom of the stem, which has slipped out of perfect Watson-Crick base pairing. The rest of the 5' end of the structure frayed only slightly, with the bottom base becoming fully exposed to solvent (Figure S6). The 3' end of the RNA strand remains largely A form, with the exception of the final sugar, which inverts its puckering. While the unphysical states described in Zgarbová *et al.* were generally not observed, it is not clear if the observed states are indeed correct conformations. In both 200 ns trajectories no fraying is observed. With the original

OPLSAA force field, this motif was far less stable. For the trajectory in Figure 10, rather than a slight shift for the second uracil, both the base and the ribose sugar deviate significantly in position, due to shifts in the dihedral angles in the backbone. The RMSD for the four residues in the loop is 1.4 Å, nearly five times that with the new force field. Furthermore, the stem region is again distorted by the conversion of γ dihedral angles to the anti state, resulting in an average overall backbone RMSD of 2.87 Å. An image of this alignment for the 1 μ s-trajectory is provided in Figure 10.

Conclusions

The new OPLS-AA/M force field for RNA has incorporated careful application of quantum chemical scans to improve the conformational energetics. This resulted in significantly better performance over the previous OPLS-AA force field for both short nucleotides and longer sequences featuring non-canonical RNA motifs. The most common fitting construct for the two-dimensional α/γ QM surface proved to be an appropriate choice for the refinement of the torsional parameters. At this point, changes to the nonbonded terms were unnecessary to achieve significant advances, although some improvements might be possible by optimization of the base stacking energetics.

The application of OPLS-AA/M improved significantly the description of short nucleotides in solution. More striking was the improvement for noncanonical motifs, especially the notoriously difficult tetraloops. Use of OPLS-AA/M demonstrated the stability of most but not every noncanonical backbone dihedral for the GAGA and UUCG tetraloops. It is curious to note the shortage of reported force fields capable of producing a stable α/γ anti/anti state.⁴⁰ Many force fields, including OPLS-AA/M, are capable of producing α and γ anti individually in noncanonical dihedral pairs, but the combination proves elusive. $\alpha/\beta/\gamma$ all occupying anti conformations provides a very extended state, lacking any inherent steric clashes or other obvious issues with accurate reproduction by a molecular mechanics force field. It is thus not clear why this pairing is such an issue. Regardless, there was still high stability of the original motifs after 1 μ s of simulation, with low RMSD values between the results of the simulations and the original structures.

In the OPLS-AA/M dinucleotide simulations, there is still some room for improvement in the 3J couplings for the χ angle of the first base. However, further tuning of the torsional parameters for χ was not pursued since they were well optimized for nucleosides.⁷ The base-pair stacking energy could potentially be tuned to address the terminal nucleotide flipping preference, although the OPLS-AA non-bonded parameters of base pairs have previously shown consistency in gas-phase interaction energy with QM calculations and free energy of solvation in chloroform with experimental values respectively.⁸ While more validation work is necessary to compare OPLSAA/M to other force fields, it represents a significant improvement the OPLS-AA force field.

The development of the OPLS-AA/M force field for RNA complements the earlier work for proteins to extend the treatment of biomolecular systems. This should enhance the abilities to perform accurate modeling in the context of structure-based drug discovery for targeting RNA, proteins, and protein-RNA complexes. Combined with OPLS-AA parameterization

for lipids⁵⁷ and the recently developed LigParGen server for small molecules,⁵⁸ applications to widespread problems in chemistry and biology are now possible.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

Gratitude is expressed to the National Institutes of Health (GM32136) for support, to the National Science Foundation for a Graduate Research Fellowship under Grant No. DGE-1122492 (M.J.R.), and to the Yale Center of Research Computing for computational resources.

REFERENCES

- (1). Poehlsgaard J; Douthwait S The Bacterial Ribosome as a Target for Antibiotics. *Nature Reviews. Microbiology* 2005, 870–881. [PubMed: 16261170]
- (2). Breaker RR; Riboswitches and the RNA world. *Cold Spring Harb. Perspect. Biol* 2012, 4, a003566. [PubMed: 21106649]
- (3). Deigan KE; Ferre-D'Amare AR; Riboswitches: discovery of drugs that target bacterial gene-regulatory RNAs. *Acc. Chem. Res* 2011, 44, 1329–1339. [PubMed: 21615107]
- (4). Kang KN; Lee YS, RNA Aptamers: A Review of Recent Trends and Applications In *Future Trends in Biotechnology*, Zhong JJ, Ed. Springer-Verlag Berlin: Berlin, 2013; Vol. 131, pp 153–169.
- (5). Robertson MJ; Tirado-Rives J; Jorgensen WL Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J. Chem. Theory Comput* 2015, 11, 3499–3509. [PubMed: 26190950]
- (6). Robertson MJ; Tirado-Rives J; Jorgensen WL Performance of Protein-Ligand Force Fields for the Flavodoxin-Flavin Mononucleotide System. *J. Phys. Chem. Lett* 2016, 7, 3032–3036. [PubMed: 27441982]
- (7). Robertson MJ; Tirado-Rives J; Jorgensen WL Improved Treatment of Nucleosides and Nucleotides in the OPLS-AA Force Field. *Chem. Phys. Lett* 2017, 683, 276–280. [PubMed: 29479109]
- (8). Pranata J; Wierschke SG; Jorgensen WL OPLS Potential Functions for Nucleotide Bases. Relative Association Constants of Hydrogen-Bonded Base Pairs in Chloroform. *J. Am. Chem. Soc* 1991, 113, 2810–2819.
- (9). Pérez A; Marchán I; Svozil D; Sponer J; Cheatham TE; Laughton CA; Orozco M Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J* 2007, 92, 3817–3829. [PubMed: 17351000]
- (10). Foloppe N; MacKerell AD Jr. All-atom empirical force field for nucleic acids: 1. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comp. Chem* 2000, 21, 86–104.
- (11). Wales DJ; Yildirim I; Improving Computational Predictions of Single-Stranded RNA Tetramers with Revised α/γ Torsional Parameters for the Amber Force Field. *J. Phys. Chem. B* 2017, 121, 2989–2999. [PubMed: 28319659]
- (12). Yildirim I; Kennedy SD; Stern HA; Hart JM; Kierzek R; Turner DH; Revision of AMBER Torsional Parameters for Tetramer Duplexes with GC and iGiC Base Pairs. *J. Chem. Theory Comput* 2012, 8, 172–181. [PubMed: 22249447]
- (13). Denning EJ; Priyakumar UD; Nilsson L; MacKerell AD Jr. Impact of 2'-hydroxyl sampling of the conformational properties of RNA: Update of the CHARMM all-atom additive force field for RNA. *J. Comput. Chem* 2011, 32, 1929–1943. [PubMed: 21469161]
- (14). Gil-Ley A; Bottaro S; Bussi G Empirical Corrections to the Amber RNA Force Field with Target Metadynamics. *J. Chem. Theory Comput* 2016, 12, 2790–2798. [PubMed: 27153317]

- (15). Cesari A; Gil-Ley A; Bussi G Combining Simulations and Solution Experiments as a Paradigm for RNA Force Field Refinement. *J. Chem. Theory Comp* 2016, 12, 6192–6200.
- (16). Aytenfisu AH; Spasic A; Grossfield A; Stern HA; Mathews DH Revised RNA Dihedral Parameters for the Amber Force Field Improved RNA Molecular Dynamics. *J. Chem. Theory. Comput* 2017, 13, 900–915. [PubMed: 28048939]
- (17). Šponer J; Bussi G; Krepl M; Banáš P; Bottaro S; Cunha RA; Gil-Ley A; Pinamonti G; Poblete S; Jurek P; Walter NG; Otyepka M RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview. *Chem. Rev* 2018, 119, 4177–4338.
- (18). Dans PD; Gallego D; Balaceanu A; Darré L; Gómez H; Orozco M Modeling, simulations, and bioinformatics at the service of RNA structure. *Chem* 2018, 5, 51–73.
- (19). Chen AA; Garcia AE; High-Resolution Reversible Folding of Hyperstable RNA Tetraloops Using Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. USA* 2013, 110, 16820–16825. [PubMed: 24043821]
- (20). 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, 395–399.
- (21). Steinbrecher T; Latzer J; Case DA Revised AMBER Parameters for Bioorganic Phosphates. *J. Chem. Theory Comput* 2012, 8, 4405–4412. [PubMed: 23264757]
- (22). Lemkul JA; MacKerell AD Jr. Polarizable force field for RNA based on the classical drude oscillator. *J. Comp. Chem* 2018, 39, 2624–2646. [PubMed: 30515902]
- (23). Zhang C; Lu C; Jing Z; Wu C; Piquemal J-P; Ponder JW; Ren P AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids. *J. Chem. Theory. Comput* 2018, 14, 2084–2108. [PubMed: 29438622]
- (24). Frisch MJ; Trucks GW; Schlegel HB; Scuseria GE; Robb MA; Cheeseman JR; Scalmani G; Barone V; Mennucci B; Petersson GA; Nakatsuji H; Caricato M; Li X; Hratchian HP; Izmaylov AF; Bloino J; Zheng G; Sonnenberg JL; Hada M; Ehara M; Toyota K; Fukuda R; Hasegawa J; Ishida M; Nakajima T; Honda Y; Kitao O; Nakai H; Vreven T; Montgomery JA Jr.; Peralta JE; Ogliaro F; Bearpark M; Heyd JJ; Brothers E; Kudin KN; Staroverov VN; Kobayashi R; Normand J; Raghavachari K; Rendell A; Burant JC; Iyengar SS; Tomasi J; Cossi M; Rega N; Millam NJ; Klene M; Knox JE; Cross JB; Bakken V; Adamo C; Jaramillo J; Gomperts R; Stratmann RE; Yazyev O; Austin AJ; Cammi R; Pomelli C; Ochterski JW; Martin RL; Morokuma K; Zakrzewski VG; Voth GA; Salvador P; Dannenberg JJ; Dapprich S; Daniels AD; Farkas Ö; Foresman JB; Ortiz JV; Cioslowski J; Fox DJ Gaussian 09, revision D.01; Gaussian, Inc.: Wallingford, CT, 2009.
- (25). Chai J; Head-Gordon M Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys* 2008, 10, 6615–6620. [PubMed: 18989472]
- (26). Jorgensen WL; Maxwell DS; Tirado-Rives J Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc* 1996, 118, 11225–11236.
- (27). Jorgensen WL; Tirado-Rives J Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J. Comput. Chem* 2005, 26, 1689–1700. [PubMed: 16200637]
- (28). Phillips JC; Braun R; Wang W; Gumbart J; Tajkhorshid E; Villa E; Chipot C; Skeel RD; Kale L; Schulten K Scalable molecular dynamics with NAMD. *J. Comput. Chem* 2005, 26, 1781–1802. [PubMed: 16222654]
- (29). Martyna GJ; Tobias DJ; Klein ML Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys* 1994, 101, 4177–4189.
- (30). Feller SE; Zhang Y; Pastor RW Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *J. Chem. Phys* 1995, 103, 4613–4621.
- (31). Miyamoto S; Kollman PA Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comp. Chem* 1992, 13, 952–962.
- (32). Pettersen EF; Goddard TD; Huang CC; Couch GS; Greenblatt DM; Meng EC; and Ferrin TE UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem* 2004, 15, 1605–1612.

- (33). Correll CC; Freeborn B; Moore PB; Steitz TA Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell*, 1997, 28, 705–712.
- (34). Jucker FM; Heus HA; Yip PF; Moors EH; Pardi A A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol* 1996, 20, 968–980.
- (35). Nozinovic S; Fürtig B; Jonker HR; Richter C; Schwalbe H High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA. *Nucleic Acids Res* 2010, 38, 683–694. [PubMed: 19906714]
- (36). Bergonzo C; Cheatham TE III Improved Force Field Parameters Lead to a Better Description of RNA Structure. *J. Chem.Theory Comput* 2015, 11, 3969–3972. [PubMed: 26575892]
- (37). Jorgensen WL; Chandrasekhar J; Madura JD; Impey RW; Klein ML Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys* 1983, 79, 926–935.
- (38). Jensen KP; Jorgensen WL Halide, Ammonium, and Alkali Metal Ion Parameters for Modeling Aqueous Solutions. *J. Chem. Theory Comput* 2006, 2, 1499–1509. [PubMed: 26627020]
- (39). Vokálová Z; Buděšínský M; Rosenberg I; Schneider B; Šponer J; Sychrovský V Structure and Dynamics of the ApA, ApC, CpA, and CpC RNA Dinucleoside Monophosphates Resolved with NMR Scalar Spin-Spin Couplings. *J. Phys. Chem. B* 2009, 113, 1182–1191. [PubMed: 19128019]
- (40). Wijmenga SS; van Buuren BNM; The use of NMR methods for conformational studies of nucleic acids. *Prog. Nucl. Magn. Reson. Spectrosc* 1998, 32, 287–387.
- (41). Marino JP; Schwalbe H; Griesinger C J-Coupling Restraints in RNA Structure Determination. *Acc. Chem. Res* 1999, 32, 614–623.
- (42). Yokoyama S; Inagaki F; Miyazawa T Advanced nuclear magnetic resonance lanthanide probe analyses of short-range conformational interrelations controlling ribonucleic acid structures. *Biochemistry* 1981, 20, 2981–2988. [PubMed: 6166319]
- (43). Haasnoot CAG; de Leeuw F; Altona C The relationship between proton-proton NMR coupling constants and substituent electronegativities-I: An empirical generalization of the Karplus eq. *Tetrahedron* 1980, 36, 2783–2792.
- (44). Condon DE; Kennedy SD; Mort BC; Kierzek R; Yildirim I; Turner DH Stacking in RNA: NMR of Four Tetramers Benchmark Molecular Dynamics. *J. Chem. Theory Comput* 2015, 11, 2729–2742. [PubMed: 26082675]
- (45). Tubbs JD; Condon DE; Kennedy SD; Hauser M; Bevilacqua PC; Turner DH The Nuclear Magnetic Resonance of CCCC RNA Reveals a Right-Handed Helix, and Revised Parameters for AMBER Force Field Torsions Improve Structural Predictions from Molecular Dynamics. *Biochemistry* 2013, 52, 996–1010. [PubMed: 23286901]
- (46). Feig M; Karanicolas J; Brooks CL MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology. *J. Mol. Graph. Model* 2004, 22, 377–395. [PubMed: 15099834]
- (47). Murray LJW; Arendall WB III; Richardson DC; Richardson JS RNA Backbone is Rotameric. *Proc. Natl. Acad. Sci. USA* 2003, 100, 13904–13909. [PubMed: 14612579]
- (48). Bergonzo C; Henriksen NM; Roe RD; Cheatham TE III Highly sampled tetranucleotide and tetraloop motifs enable evaluation of common RNA force fields. *RNA*, 2015, 21, 1578–1590. [PubMed: 26124199]
- (49). Klein DJ; Schmeing TM; Moore PB; Steitz TA; The kink-turn: a new RNA secondary structure motif. *EMBO J* 2001, 20, 4214–4221. [PubMed: 11483524]
- (50). Tan D; Piana S; Dirks RM; Shaw DE RNA Force Field with Accuracy Comparable to State-of-the-Art Protein Force Fields. *Proc. Natl. Acad. Sci. USA* 2018, 115, E1346–E1355. [PubMed: 29378935]
- (51). Schrodtr MV; Andrews CT; Elcock AH Large-scale analysis of 48 DNA and 48 RNA tetranucleotides studied by 1 μ s explicit-solvent molecular dynamics simulations. *J. Chem. Theory Comput* 2015, 11, 5906–5917. [PubMed: 26580891]
- (52). Zgarbová M; Jurek P; Banáš P; Varila M; Šponer J; Otyepka M Noncanonical α/γ Backbone Conformations in RNA and the Accuracy of Their Description by the AMBER Force Field. *J. Phys. Chem. B* 2017, 121, 2420–2433. [PubMed: 28290207]

- (53). Kührová P; Best RB; Bottaro S; Bussi G; Šponer J; Otyepka M; Banáš P Computer Folding of RNA Tetraloops: Identification of Key Force Field Deficiencies. *J. Chem. Theory Comput* 2016, 12, 4534–4548. [PubMed: 27438572]
- (54). Bottaro S; Banáš P; Šponer J; Bussi G Free Energy Landscape of GAGA and UUCG RNA Tetraloops. *J. Phys. Chem. Lett* 2016, 7, 4032–4038. [PubMed: 27661094]
- (55). Zgarbová M; Otyepka M; Šponer J; Mládek A; Banáš P; Cheatham TE III; Jurek 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, 2886–2902. [PubMed: 21921995]
- (56). Zgarbová M; Otyepka M; Šponer J; Lankaš F; Jurek P Base Pair Fraying in Molecular Dynamics Simulations of DNA and RNA. *J. Chem. Theory Comput* 2014, 10, 3177–3189. [PubMed: 26588288]
- (57). Kulig W; Pasenkiewicz-Gierula M; Róg T Topologies, structures and parameter files for lipid simulations in GROMACS with the OPLS-AA force field: DPPC, POPC, DOPC, PEPC, and cholesterol. *Data in brief*, 2015, 5, 333–336. [PubMed: 26568975]
- (58). Dodda LS; Cabeza de Vaca I; Tirado-Rives J; Jorgensen WL LigParGen web server: An automatic OPLS-AA parameter generator for organic liquids. *Nucleic Acids Research* 2017, 45, W331–W336. [PubMed: 28444340]

Figure 1.
Constructs used for (A) the α - γ and (B) the ϵ - ζ potential energy surfaces.

Figure 2.

Quantum mechanical potential energy maps for variation of α and γ with the ribose held in the (A) C2' endo and (B) C3' endo conformation. (C) and (D) correspond to the OPLS-AA/M potential energy surfaces for C2' and C3' endo. Energies are in kcal/mol.

Figure 3. RMSD between calculated and experimental 3J couplings for dinucleotide monophosphates with the OPLS-AA and OPLS-AA/M force field.

Figure 4.
Structures of the centroids of the clusters for A-form like (left) and 1–2–4 stacked (right).
Relative populations in OPLS-AA/M and OPLS-AA simulations are shown.

Figure 5. RMSD between calculated and experimental 3J couplings for tetranucleotide triphosphates with the OPLS-AA and OPLS-AA/M force fields.

Figure 6.
Percent of simulated NOE distances that fall outside of the experimentally acceptable range for each tetranucleotide with the OPLS-AA and OPLS-AA/M force fields.

Figure 7.
Schematic structures of non-canonical RNA motifs studied in this work. Sites of non-canonical dihedrals are colored in blue while mismatched base-pairing is in red.

Figure 8.
RMSD values to the starting structure of the E loop of *E. coli* 5S RNA with the OPLS AA and OPLS-AA/M force fields in the last 50 ns of 1- μ s runs.

Figure 9.

Comparison of the NMR structure for the GAGA tetraloop (1ZIG) with the centroid structure of the largest cluster for the last 50 ns of a 1 μ s trajectory with the new OPLS-AA/M and old OPLS-AA force fields. The centroid structures have been aligned to the NMR structure to minimize the backbone atom RMSD.

Figure 10.

Comparison of the NMR structure for the UUCG tetraloop (2KOC) with the centroid structure of the largest cluster for the last 50 ns of 1- μ s trajectories with the new OPLS-AA/M and old OPLS-AA force fields. The centroid structures have been aligned to the NMR structure to minimize the backbone atom RMSD.

Table 1.

³JCouplings for AA Dinucleotide Monophosphate

	Expt	AA	
		OPLS-AA	OPLS-AA/M
γ 1 ^a	3.6	8.0±0.9	4.2±0.3
	2.5	2.8±0.6	4.6±0.3
δ 1	5	6.9±0.2	3.0±1.0
	1.9	6.1±0.1	5.6±0.2
χ 1 ^b	4.2	3.8±0.0	3.6±0.1
	5.3	4.6±0.4	3.6±0.5
ϵ 1 ^c	3.7	4.1±0.5	4.0±0.4
	9.0	8.3±0.2	10.3±0.2
ζ 1	5.4	7.0±0.1	7.0±0.0
α 2	5.4	5.6±0.1	5.8±0.1
	9.4	5.7±0.9	9.3±0.0
β 2 ^d	3.8	6.1±0.9	4.1±0.7
	3.2		
γ 2 ^a	~	5.8±0.9	4.0± 1.3
	3.7	4.4±0.6	2.7±0.1
δ 2	5.5	7.9±0.1	4.2±2.1
	2.5	4.4±0.7	2.5±0.9
χ 2 ^e	3.1	3.1±0.3	3.1±0.1

^a $J(\text{H4}'5'')$, $J(\text{H4}'5')$.^b $J(\text{H1}'\text{C2})$, $J(\text{H1}'\text{C6})$.^c $J(\text{C4}',\text{P})$, $J(\text{C2}',\text{P})$, $J(\text{H3}',\text{P})$.^d $J(\text{C4}',\text{P})$, $J(\text{H5}',\text{P})$, $J(\text{H5}'',\text{P})$.^e $J(\text{H1}',\text{C4})$, $J(\text{H1}',\text{C8})$.^f Ref. 39. Errors are the standard deviation from triplicate simulations.

Table 2.

³JCouplings for Tetranucleotides

	AAAA			CCCC			GACC		
	Expt ^d	OPLS-AA	OPLS-AA/M	Expt ^d	OPLS-AA	OPLS-AA/M	Expt ^d	OPLS-AA	OPLS-AA/M
β 2 ^a	3.8	5.2±0.4	3.7±1.5	3.8	6.3±0.2	6.7±0.2	3.7	4.5±0.6	3.7±0.1
	1	15.0±2.8	2.3±0.6	1.2	7.2±2.9	1.3±0.0	0.9	6.9±0.5	2.2±0.4
β 3 ^a	3	4.7±0.5	2.9±0.5	3.9	5.3±0.1	4.5±0.3	4.0	4.4±0.4	4.4±0.3
	1	13.5±0.8	2.6±0.6	0.5	7.6±3.5	2.0±0.5	2.0	5.8±3.7	2.3±1.1
β 4 ^a	3.2	4.8±0.4	3.1±0.4	3.8	6.8±0.9	4.7±0.1	4.4	5.5±0.2	4.5±0.1
	1	10.6±3.0	2.4±0.4	1.1	6.5±0.8	1.4±0.1	2.0	5.2±0.7	1.5±0.0
γ 1 ^b	3.8	2.7±0.2	4.0±0.5	-	2.8±0.3	2.5±1.0	4.6	2.4±0.1	4.4±1.0
	2	8.3±0.1	4.5±1.0	-	8.0±0.3	5.0±0.1	2.3	8.5±0.2	5.1±0.1
γ 2 ^b	2	4.1±0.3	2.5±0.2	1	4.1±0.2	3.5±0.0	~2	3.7±0.1	2.6±0.2
	1	9.4±1.1	2.9±1.8	1	4.6±1.6	1.0±0.0	~2	4.1±0.3	3.1±2.2
γ 3 ^b	2	4.0±0.2	2.6±0.4	2.1	4.1±0.2	2.7±0.1	1.5	3.5±0.2	2.6±0.1
	2	8.3±0.3	3.1±2.5	1	5.5±2.0	2.4±0.7	<1	4.1±2.2	2.5±1.3
γ 4 ^b	2	3.6±0.3	2.8±0.3	2	4.5±0.5	2.7±0.1	1.8	3.8±0.1	2.6±0.1
	2	6.3±1.2	2.7±1.7	1.4	4.6±0.4	1.5±0.1	1.3	4.1±0.3	1.5±0.1
ϵ 1 ^c	8.45	8.1±0.5	7.0±3.0	8.8	9.5±0.5	11.1±0.0	9.3	9.2±0.3	10.5±0.2
ϵ 2 ^c	8.7	8.2±0.4	5.1±2.1	9.3	8.7±0.8	10.5±0.1	9.1	9.5±0.5	10.4±0.1
ϵ 3 ^c	8.35	8.8±0.6	5.2±1.2	9.3	9.2±0.2	10.7±0.0	9.0	9.4±0.2	10.6±0.0

^a³J(P5-H5'), ³J(P5-H5'').^b³J(H4'-H5'), J(H4'-H5'').^c³J(H3'-P).^dRefs. 44, 45. Errors are the standard deviation from triplicate simulations.