

# Revised RNA Dihedral Parameters for the Amber Force Field Improve RNA Molecular Dynamics

Asaminew H. Aytenfisu,<sup>†,‡,§</sup> Aleksandar Spasic,<sup>†,‡,§</sup> Alan Grossfield,<sup>†</sup> Harry A. Stern,<sup>§</sup> and David H. Mathews<sup>\*,†,‡,§,||</sup><sup>ID</sup>

<sup>†</sup>Department of Biochemistry & Biophysics, University of Rochester Medical Center, Rochester, New York 14642, United States

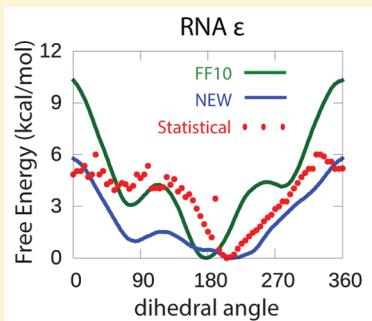
<sup>‡</sup>Center for RNA Biology, University of Rochester Medical Center, Rochester, New York 14642, United States

<sup>§</sup>Center for Integrated Research Computing, University of Rochester, Rochester, New York 14642, United States

<sup>||</sup>Department of Biostatistics & Computational Biology, University of Rochester Medical Center, Rochester, New York 14642, United States

## Supporting Information

**ABSTRACT:** The backbone dihedral parameters of the Amber RNA force field were improved by fitting using multiple linear regression to potential energies determined by quantum chemistry calculations. Five backbone and four glycosidic dihedral parameters were fit simultaneously to reproduce the potential energies determined by a high-level density functional theory calculation (B97D3 functional with the AUG-CC-PVTZ basis set). Umbrella sampling was used to determine conformational free energies along the dihedral angles, and these better agree with the population of conformations observed in the protein data bank for the new parameters than for the conventional parameters. Molecular dynamics simulations performed on a set of hairpin loops, duplexes and tetramers with the new parameter set show improved modeling for the structures of tetramers CCCC, CAAU, and GACC, and an RNA internal loop of noncanonical pairs, as compared to the conventional parameters. For the tetramers, the new parameters largely avoid the incorrect intercalated structures that dominate the conformational samples from the conventional parameters. For the internal loop, the major conformation solved by NMR is stable with the new parameters, but not with the conventional parameters. The new force field performs similarly to the conventional parameters for the UUCG and GCAA hairpin loops and the  $[U(UA)_6A]_2$  duplex.



## INTRODUCTION

As longer trajectories became possible with the availability of modern computers, limitations in the force fields for nucleic acids became apparent.<sup>1,2</sup> These limitations remain for simulations of RNA, where recent benchmarks of the Amber force field reveal serious flaws in its modeling of tetramers, noncanonical pairs in internal loops, and tetraloops. Tetramers were characterized by NMR, and simulations for those sequences do not match the experimental data.<sup>3–8</sup> Moreover, free energy calculations on small systems involving noncanonical pairs do not match the conformational preferences demonstrated in solution structures.<sup>9–11</sup> Banáš et al. demonstrated signature interactions in tetraloops that are lost during state-of-the-art simulations.<sup>12</sup> Longer simulations using enhanced sampling show that problems simulating tetraloops persist, and that CHARMM36<sup>13</sup> and an Amber force field with revised van der Waals parameters<sup>14</sup> are also unable to correctly model the solution structures of tetramers and tetraloops.<sup>4,15,16</sup>

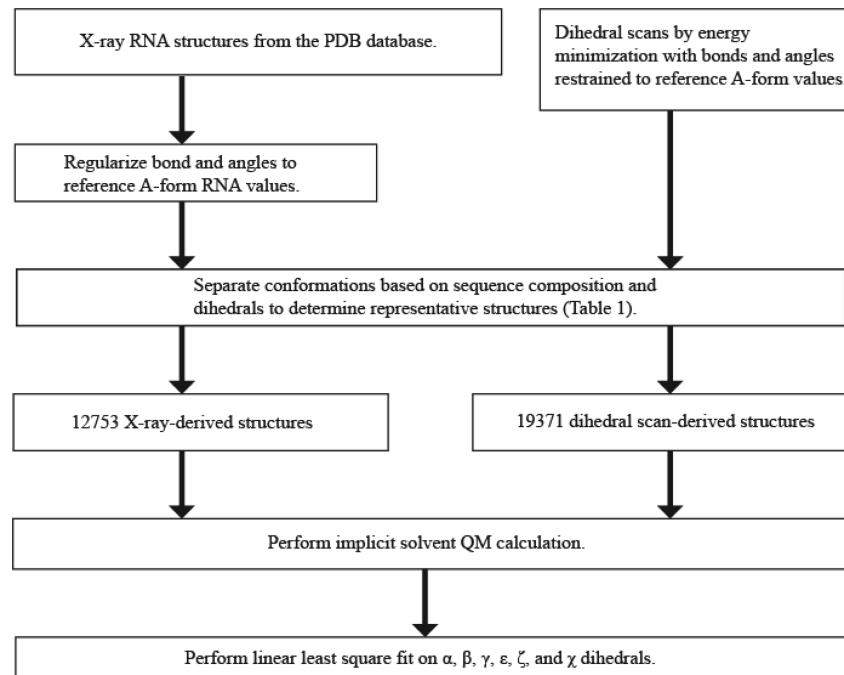
To address the limitations of the Amber nucleic acids force field, work has largely focused on revising subsets of dihedral parameters by fitting a single dihedral at a time using model systems.<sup>13,17–19</sup> Development of protein force fields generally followed a similar approach,<sup>20–22</sup> although the simpler

backbone has allowed for simultaneous fitting of both backbone dihedrals.<sup>23</sup> The version of the Amber force field for nucleic acids, distributed with Amber 14, was named ff10 (and was unchanged in the ff14SB force field).<sup>24</sup> This force field is a refinement of the ff94 force field of Cornell et al.<sup>25</sup> Initially, this force field was improved by changing the sugar pucker and glycosidic torsions in force fields ff98 and ff99.<sup>26,27</sup> Later, a new set of parameters called bsc0 was derived for  $\alpha$  and  $\gamma$  dihedrals.<sup>17</sup> The bsc0 force field was derived by fitting to quantum mechanical (QM) energies on an extended model of backbone atoms derived from RNA and DNA molecules based on B3LYP and LMP2 calculations.<sup>17</sup> In 2010 and 2011, two new sets of glycosidic torsions were derived for each base, namely Chi\_YIL<sup>19</sup> and ChiOL3,<sup>18</sup> respectively, with ChiOL3 being incorporated into ff10. Additionally, Chen & Garcia revised the van der Waals parameters in addition to the dihedrals,<sup>14</sup> although that parameter set did not solve the problems with tetraloop or tetramer conformational preferences<sup>4</sup> and introduced problems that can be observed in the loss of stacking in an HIV-1 kissing loop complex.<sup>28</sup>

Received: September 1, 2016

Published: January 3, 2017





**Figure 1.** Procedure for fitting force field parameters. A diverse database of conformations was generated from X-ray structures and from dihedral scanning. For X-ray structures, bonds and angles were fixed to an A-form reference value using sander from the Amber software package.<sup>24</sup> The dihedral scan structures were generated by energy minimization with restraints on selected torsions. Additional restraints were applied on bond and angles to set them to the A-form reference value. The database was then reduced in size by considering sequence identity and dihedrals to remove redundancies. Structures generated by scanning and taken from the PDB were both used for the linear least-square fit.

In this work, the Amber force field for RNA was improved by simultaneously refitting the  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\epsilon$ ,  $\zeta$ , and glycosidic dihedral ( $\chi$ ) parameters by linear regression to potentials determined by QM calculations. In place of the conventional scan of each individual dihedral, a diverse range of nucleotide and dinucleotide conformations were chosen, combining structures from the PDB with a set of dihedral scans to sample the high potential energy barriers. An overview of the procedure is shown in Figure 1.

The newly fitted parameters were tested in several ways. Free energy profiles for dihedral angles of mono- and dinucleotides were calculated using the new dihedral parameters with umbrella sampling. These profiles better resemble the population of conformations observed in the protein data bank than do conformational free energies using the conventional Amber force field. Additionally, molecular dynamics trajectories were calculated for tetramers, internal loops, and the UUCG tetraloop.<sup>3,5,29</sup> Improved agreement with solution structures was found for many systems, although the new parameters did not resolve all the anomalies.

## METHODS

**Assembling a Database of Structures for Fitting.** The first step was creating a database of diverse dinucleotide conformations. It was important that the database cover a variety of feasible conformations to ensure that the fitted dihedral parameters are transferable to diverse structures. The database was then subsequently reduced in size to remove redundant conformations (Figure 1). A database of 23 399 dinucleotide systems was generated by extracting all RNA dinucleotide fragments from a diverse set of X-ray structures with resolution of 4 Å or better from the Protein Data Bank (PDB).<sup>30,31</sup> All dinucleotides were terminated with OH groups at the 5' and

**Table 1. Sequence Composition of All Dinucleotides and Nucleosides Used in Fitting**

sequence	no. conformations	sequence	no. conformations
AA	2198	CA	1913
AC	1832	CC	1756
AG	2152	CG	1782
AU	1903	CU	1610
GA	2048	UA	1737
GC	1714	UC	1700
GG	1753	UG	1728
GU	1785	UU	1636
A	717	G	724
C	718	U	718

3' carbons. The list of structures is provided in Table S1. When the crystal structure contained alternative coordinates, only the primary structure was kept.

To expand the diversity of conformations beyond those found in crystal structures, an additional 16 499 dinucleotide and 2877 nucleoside conformations were generated using one-dimensional dihedral scans by energy minimization with the sander program from Amber.<sup>24,32</sup> The dihedral scans were performed in two ways. Both ways restrained the target dihedral angle with a force constant of 1000 kcal / (mol × rad<sup>2</sup>), while bonds and angles were kept close to those found in A-form RNA by applying restraints using a force constant of 500 kcal / (mol × Å<sup>2</sup>) and 500 kcal / (mol × rad<sup>2</sup>), respectively. The structural parameters for A-form RNA were derived from the NAB module from Amber,<sup>33</sup> which determined them using fiber diffraction data.<sup>34</sup> The first set of scans additionally restrained the other dihedrals in the backbone to their starting configuration, with a force constant of 50 kcal / (mol × rad<sup>2</sup>). The second set of scans was performed without dihedral

restraints other than the dihedral of the scan. This provided a diversity of conformations in addition to the diversity along the scanned dihedral. The starting structures for the dihedral scan were A-form helices built using NAB;<sup>33</sup> restraints for bonds and angles were generated using “rdparm” from Amber.<sup>24</sup> Backbone dihedral restraints for the first set were generated using cpptraj by calculating the dihedral angles of each X-ray structure from the database.<sup>24,35</sup> Each dihedral scan was done with 5° intervals between structures. Each structure was minimized for 2500 steps using the steepest descent algorithm, followed by 2500 steps of conjugate gradient minimization. During the scans, the solvent was represented using the Generalized Born implicit solvent model as implemented in the sander program.<sup>24</sup>

**Regularization of Bonds and Angles.** The next step was to fix bonds and angles of all the X-ray-structure-derived configurations to those observed in reference A-form helices; for those structures generated above by scanning, the bonds and angles were regularized as part of the procedure for generating the coordinates.<sup>34</sup> This was done to ensure the bonded terms of the force field would have the same values in order to prevent excessive noise in the potential energies of the configurations being introduced by variations in these terms, and was accomplished using energy minimization with the GB implicit solvent<sup>36–38</sup> model in sander.<sup>24</sup> Restraints for bonds and angles to the reference A-form helices were ramped slowly from an initial 100 kcal/(mol × Å<sup>2</sup>) and 100 kcal/(mol × rad<sup>2</sup>), respectively, to 500 kcal/(mol × Å<sup>2</sup>) and 500 kcal/(mol × rad<sup>2</sup>), respectively, over 1000 steps of steepest descent minimization and then kept at 500 kcal/(mol × rad<sup>2</sup>) for the rest of 1500 steps of steepest descent followed by a 2500 steps of conjugate gradient minimization. To keep the X-ray structure close to the original structure during bond and angle regularization, dihedral restraints with 30 kcal/(mol × rad<sup>2</sup>) were applied in one set of 10668 configurations. A second set of 2085 configurations were energy minimized without any dihedral restraint regularization.

Restraints for bonds and angles were generated using “rdparm” from AmberTools15.<sup>24</sup> “rdparm” prints all bonds, angles and dihedral angles in a file along with atom pair lists that form bonds, three atom pair lists that form angles and four atom pair lists that form dihedral angle. Sample output from “rdparm” is given in Tables S3–S5. Restraints for the torsions were generated using cpptraj by calculating dihedral angle value of each X-ray structure from the database.<sup>24,35</sup>

Next, clustering (in 5° wide bins on dihedral value as the selection criteria) was performed for systems of identical sequence. A set of diverse structures were generated (shown in Table 1), including 12 753 X-ray structure derived and 19 371 dihedral-scan derived structures, equally divided by configuration and sequence identity. Of the dihedral-scan derived structures, 2877 were nucleosides.

**Potential Energy Calculation.** For each structure, the energy was calculated using dispersion corrected density functional B97 with dampening function D3(BJ) and the triple- $\zeta$  Dunning correlation consistent basis set with diffuse functions, AUG-CC-PVTZ.<sup>39,40</sup> This method was previously tested against a databases of high level QM calculations, the S66 database,<sup>41</sup> which includes CCSD(T)/CBS energies of 66 complexes stabilized by hydrogen bonding and dispersion interactions. B97D3 produces results that are close to the reference values; the mean absolute deviation between B97D3/def2-QZVP calculated energies and the reference S66 energies is only 0.29 kcal/mol.<sup>42</sup> A more recent benchmark with a database containing all possible RNA backbone conformations

using various QM methods also found that similar functionals based on generalized gradient approximation (GGA) with dampening correction perform well in describing the RNA backbone.<sup>43</sup> B97D3 with this relatively large basis set was chosen as a balance between accuracy and computational efficiency, since it had to be run for a large number of conformations. These QM calculations were performed with Gaussian 09.<sup>44</sup> An implicit water model, solvation model based on density (SMD),<sup>45</sup> was used for the QM.

**Regression Procedure.** The next step was to perform a linear least-squares (LLS) fit of dihedral parameters so as to best reproduce the QM potential energies. The composition of the database, in terms of sequences, is shown in Table 1.<sup>46</sup> Because bond lengths and angles were regularized, relative energies exclude the undesired noisy contributions that would have resulted from changes in bond lengths and angles.

$\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\epsilon$ ,  $\zeta$ , and four  $\chi$  dihedrals (one for each base) were fit simultaneously. The delta torsion was excluded at this time because its energy is correlated with the other sugar ring torsions. To avoid conflict between the  $\epsilon$  dihedral, defined by torsion angle C4' $_i$ –C3' $_i$ –O3' $_i$ –P $_{i+1}$ , and the sugar ring torsion, defined by C2' $_i$ –C3' $_i$ –O3' $_i$ –P $_{i+1}$ , the atom type of C4' was changed from CT to CF; in Amber ff10, these chemically distinct dihedrals have the same definition. Details of all fit dihedrals are given in Table 2.

The equation used for the dihedral potential in Amber is

$$E_{\text{Amber}, \text{dihedral}} = \sum_{n=1}^4 v_n (1 + \cos(n\varphi - \gamma_n)) \quad (1)$$

where  $\varphi$  is the dihedral angle,  $n$  is the multiplicity,  $\gamma_n$  is a phase offset, and  $v_n$  is a barrier height.<sup>25</sup> Although the energy is nonlinear in the phase,  $\gamma_n$ , a predetermined phase can be introduced, such that the cosine can be rewritten as eqs 2 and 3<sup>46,47</sup>

$$E_{\text{Amber}, \text{dihedral}} = E_{\text{fit}, \text{dihedral}} + \sum_{n=1}^4 (1 - \cos(\gamma_n) - \sin(\gamma_n)) \quad (2)$$

$$E_{\text{fit}, \text{dihedral}} = \sum_{n=1}^4 [v_{1n}(1 + \cos(n\varphi)) + v_{2n}(1 + \sin(n\varphi))] \quad (3)$$

where the barrier height,  $v_n$ , and phase,  $\gamma_n$ , are

$$v_n = \sqrt{v_{1n}^2 + v_{2n}^2} \quad (4)$$

$$\gamma_n = \tan^{-1}\left(\frac{v_{2n}}{v_{1n}}\right) \quad (5)$$

The linear function for each dihedral included four sine and cosine terms of differing multiplicity, which can be subsequently transformed to a sum of cosines (plus an arbitrary offset) as used in the Amber force field.<sup>46,47</sup> Table S7 in the Supplement provides the values and standard errors for the fit parameters.

The total Amber energy, including the fitting part, is given by

$$\begin{aligned} E_{\text{Amber}}(v_1, v_2, \dots, v_n) &= E_{\text{bonds}} + E_{\text{angles}} + E_{\text{dihed-const}} \\ &+ E_{\text{dihed-fit}}(v_1, v_2, \dots, v_n) + E_{\text{VDW}} + E_{\text{elect}} + E_{\text{solvation}} \\ &= E_{\text{const}} + E_{\text{dihed-fit}}(v_1, v_2, \dots, v_n) \end{aligned} \quad (6)$$

**Table 2.** Fitting Torsion Names, Atom Names, and Atom Types According to Amber Force Field Nomenclature<sup>a</sup>

torsion name	dihedral angle definition			torsion name	dihedral angle definition		
	atom name	atom type	alternative paths		atom name	atom type	alternative paths
1 $\alpha$	O3' <sub>i-1</sub> —P <sub>i</sub> —OS5' <sub>i</sub> —CS5' <sub>i</sub>	OS/OH—P—OS—CI	OP1 <sub>i</sub> —P <sub>i</sub> —OS5' <sub>i</sub> —CS5' <sub>i</sub> OP2 <sub>i</sub> —P <sub>i</sub> —OS5' <sub>i</sub> —CS5' <sub>i</sub>	7 $\chi$ cytosine	O4' <sub>i</sub> —C1' <sub>i</sub> —N1 <sub>i</sub> —C6 <sub>i</sub>	OS—CT—N*—C4	O4' <sub>i</sub> —C1' <sub>i</sub> —N1 <sub>i</sub> —C2 <sub>i</sub> C2' <sub>i</sub> —C1' <sub>i</sub> —N1 <sub>i</sub> —C6 <sub>i</sub> C2' <sub>i</sub> —C1' <sub>i</sub> —N1 <sub>i</sub> —C2 <sub>i</sub> H1' <sub>i</sub> —C1' <sub>i</sub> —N1 <sub>i</sub> —C2 <sub>i</sub>
2 $\beta$	P <sub>i</sub> —OS5' <sub>i</sub> —CS5' <sub>i</sub> —C4' <sub>i</sub>	P—OS—CI—CF	P <sub>i</sub> —OS5' <sub>i</sub> —CS5' <sub>i</sub> —H5' <sub>i</sub> P <sub>i</sub> —OS5' <sub>i</sub> —CS5' <sub>i</sub> —H5'' <sub>i</sub>	8 $\chi$ guanine	O4' <sub>i</sub> —C1' <sub>i</sub> —N9 <sub>i</sub> —C8 <sub>i</sub>	OS—CT—N*—CP	O4' <sub>i</sub> —C1' <sub>i</sub> —N9 <sub>i</sub> —C4 <sub>i</sub> C2' <sub>i</sub> —C1' <sub>i</sub> —N9 <sub>i</sub> —C8 <sub>i</sub> C2' <sub>i</sub> —C1' <sub>i</sub> —N9 <sub>i</sub> —C4 <sub>i</sub> H1' <sub>i</sub> —C1' <sub>i</sub> —N9 <sub>i</sub> —C8 <sub>i</sub> H1' <sub>i</sub> —C1' <sub>i</sub> —N9 <sub>i</sub> —C4 <sub>i</sub> H1' <sub>i</sub> —C1' <sub>i</sub> —N9 <sub>i</sub> —C4 <sub>i</sub>
3 $\gamma$	OS5' <sub>i</sub> —CS5' <sub>i</sub> —C4' <sub>i</sub> —C3' <sub>i</sub>	OS/OH—CI—CF—CT	HS5' <sub>i</sub> —C5' <sub>i</sub> —C4' <sub>i</sub> —C3' <sub>i</sub> HS5' <sub>i</sub> —C5' <sub>i</sub> —C4' <sub>i</sub> —H4' <sub>i</sub> HS5'' <sub>i</sub> —CS5' <sub>i</sub> —C4' <sub>i</sub> —C3' <sub>i</sub> HS5'' <sub>i</sub> —CS5' <sub>i</sub> —C4' <sub>i</sub> —H4' <sub>i</sub> HS5'' <sub>i</sub> —CS5' <sub>i</sub> —C4' <sub>i</sub> —O4' <sub>i</sub> OS5' <sub>i</sub> —CS5' <sub>i</sub> —C4' <sub>i</sub> —H4' <sub>i</sub> OS5' <sub>i</sub> —CS5' <sub>i</sub> —C4' <sub>i</sub> —O4' <sub>i</sub>	9 $\chi$ uracil	O4' <sub>i</sub> —C1' <sub>i</sub> —N1 <sub>i</sub> —C6 <sub>i</sub>	OS—CT—N*—CS	O4' <sub>i</sub> —C1' <sub>i</sub> —N1 <sub>i</sub> —C2 <sub>i</sub> C2' <sub>i</sub> —C1' <sub>i</sub> —N1 <sub>i</sub> —C6 <sub>i</sub> C2' <sub>i</sub> —C1' <sub>i</sub> —N1 <sub>i</sub> —C2 <sub>i</sub> H1' <sub>i</sub> —C1' <sub>i</sub> —N1 <sub>i</sub> —C2 <sub>i</sub> H1' <sub>i</sub> —C1' <sub>i</sub> —N1 <sub>i</sub> —C2 <sub>i</sub>
4 $\varepsilon$	C4' <sub>i</sub> —C3' <sub>i</sub> —O3' <sub>i</sub> —P <sub>i+1</sub>	CF—CT—OS—P	H3' <sub>i</sub> —C3' <sub>i</sub> —O3' <sub>i</sub> —P <sub>i+1</sub> C2' <sub>i</sub> —C3' <sub>i</sub> —O3' <sub>i</sub> —P <sub>i+1</sub>				
5 $\zeta$	C3' <sub>i</sub> —O3' <sub>i</sub> —P <sub>i+1</sub> —OS5' <sub>i+1</sub>	CT—OS—P—OS	C3' <sub>i</sub> —O3' <sub>i</sub> —P <sub>i+1</sub> —OP1 <sub>i+1</sub> C3' <sub>i</sub> —O3' <sub>i</sub> —P <sub>i+1</sub> —OP2 <sub>i+1</sub>				
6 $\chi$ adenine	O4' <sub>i</sub> —C1' <sub>i</sub> —N9 <sub>i</sub> —C8 <sub>i</sub>	OS—CT—N*—C5	O4' <sub>i</sub> —C1' <sub>i</sub> —N9 <sub>i</sub> —C4 <sub>i</sub> C2' <sub>i</sub> —C1' <sub>i</sub> —N9 <sub>i</sub> —C8 <sub>i</sub> C2' <sub>i</sub> —C1' <sub>i</sub> —N9 <sub>i</sub> —C4 <sub>i</sub> H1' <sub>i</sub> —C1' <sub>i</sub> —N9 <sub>i</sub> —C8 <sub>i</sub> H1' <sub>i</sub> —C1' <sub>i</sub> —N9 <sub>i</sub> —C4 <sub>i</sub>				

<sup>a</sup>The atom type “OS/OH” is to include the atom type for terminal O3' and internal O3' atom type.

**Table 3.** RNA Molecules Used for Molecular Dynamics Simulations

sequence	source	model no.	PDB ID	no. NaCl
5' GGUGAAGGC 3'	NMR	22	2DD2 <sup>71</sup>	7
3' CCGAACCG 5'				
5' GGCACUUCGGUGCC 3'	NMR	11	2KOC <sup>29</sup>	6
5' U(UA) <sub>6</sub> A 3'	X-ray	1	1RNA <sup>66</sup>	11
5' A(AU) <sub>6</sub> U 5'				
5' GCGCAAGC 3' <sup>a</sup>	NMR	6	1ZIH <sup>77</sup>	5
5' AAAA 3'	single strand of A-form RNA built using NAB module of Amber <sup>24</sup>			3
5' CAAU 3'				
5' CCCC 3'				
5' GACC 3'				
5' UUUU 3'				
5' CCCC 3'	C2' endo anti C2' endo syn C3' endo syn	taken from the work of Condon et al. <sup>3,5</sup>		

<sup>a</sup>Two base pairs were removed from the 1ZIH sequence for simulations relative to the solution structure.

where the total Amber potential energy ( $E_{\text{Amber}}$ ) is the sum of terms for bond length ( $E_{\text{bonds}}$ ), bond angles ( $E_{\text{angles}}$ ), the dihedrals that are not being fit ( $E_{\text{dihed-const}}$ ), the dihedrals being fit ( $E_{\text{dihed-fit}}$ ), van der Waals ( $E_{\text{VDW}}$ ), electrostatics ( $E_{\text{elect}}$ ), and implicit solvation ( $E_{\text{solvation}}$ ). All of these terms are constant except for the  $E_{\text{dihed-fit}}$  and the goal of the fitting is to improve  $E_{\text{dihed-fit}}$  to best match the total potential energy calculated by QM, including implicit solvation ( $E_{\text{QM}}$ )

$$E_{\text{const}} + E_{\text{dihed-fit}}(\nu_1, \nu_2, \dots, \nu_n) = E_{\text{QM}} \quad (7)$$

The potential energies calculated by Amber (in  $E_{\text{const}}$ ) and in the QM as part of total potential energy ( $E_{\text{QM}}$ ) both include implicit solvation to account for fact that most simulations are run in an aqueous environment. This is consistent with fitting done previously.<sup>48</sup> Only the electrostatic part of the solvation energy<sup>48</sup> was included.<sup>34</sup>

Equation 7 reduces to the linear equation

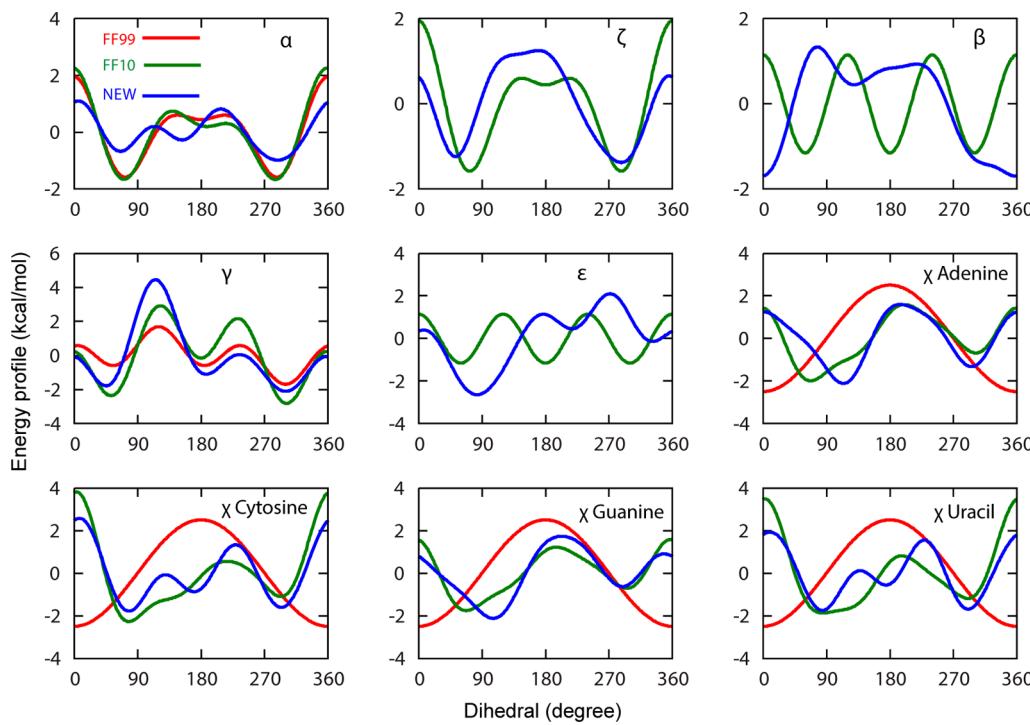
$$AX = B; X = [\nu_1, \nu_2, \dots, \nu_n] \quad (8)$$

where  $X$  is vector of fitting parameters  $\nu_i$ ,  $B$  is a vector of energy differences between QM and MM,  $A$  is a  $p \times n$  coefficient matrix with  $p$  being number of fitting parameters, and  $n$  is the number of structures. The derivation can be found in recent publications.<sup>46,47</sup> All constant terms are precalculated using the conventional Amber parameters before fitting.

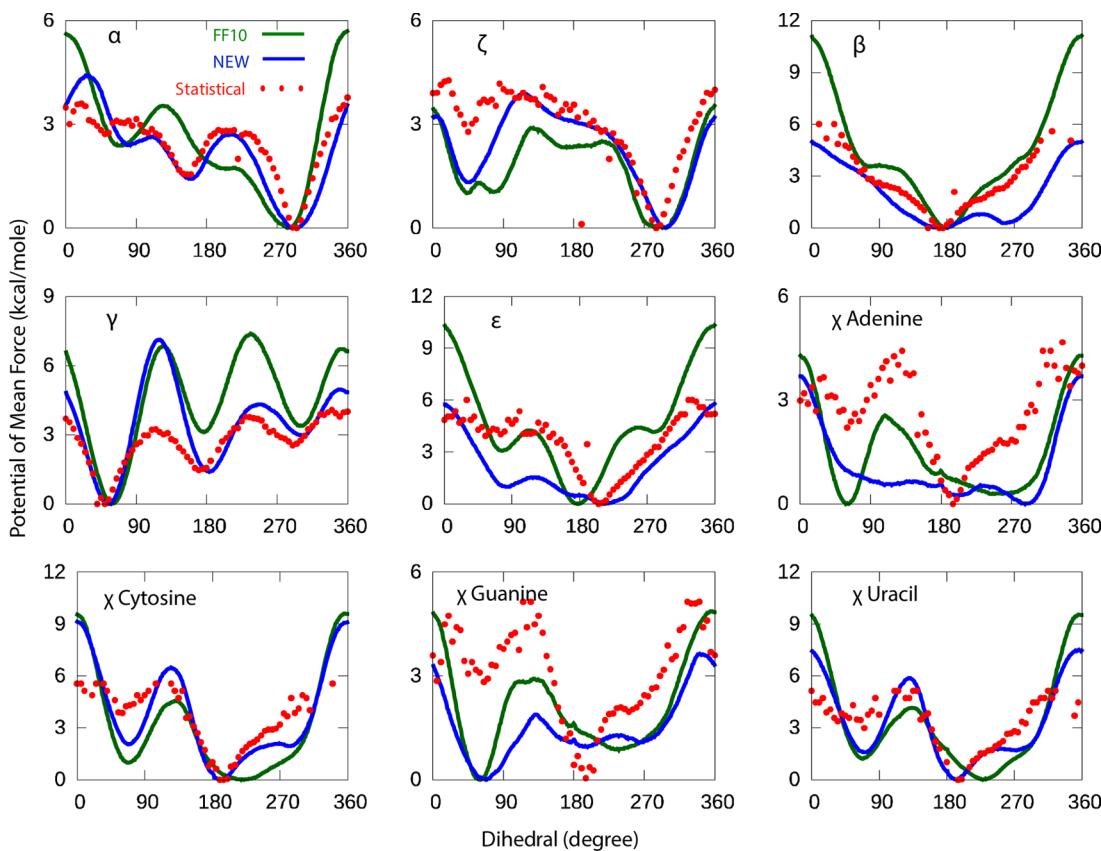
The solvation free energy in eq 6 was estimated using the analytical linearized Poisson–Boltzmann method from the Amber software package via sander with the effective atomic radii from Tsui and Case.<sup>51,52</sup> The solvent permittivity was set to 78.4 (water), ionic strength was set to 150 mM, and all other parameters were default.

**Fitting Software.** The potential energy, bond, angle, electrostatic, van der Waals, and the constant portion of dihedrals ( $E_{\text{const}}$  in eq 6), were computed using a custom FORTRAN code. Linear least square (LLS) fitting was by singular value decomposition (SVD) using the GELSS routine from the Intel FORTRAN 95 LAPACK library.<sup>53</sup>

**Umbrella Sampling Calculation to Validate Results.** To test the new dihedral parameters, the free energy profiles along  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\varepsilon$ ,  $\zeta$  and the four  $\chi$  dihedrals were determined using umbrella sampling on dinucleotides in TIP3P water. Free energies were determined using the conventional Amber parameters and using the new dihedral parameters. Sixteen RNA dinucleotide molecules (one molecule for each possible dinucleotide sequence) were used. For each dihedral, 1 ns of restrained molecular dynamics was performed for each of 72 windows at 5° increments. In total, per molecule, 1152 ns total sampling was performed. The initial conformation for each window was generated with a separate MD simulation prior to the sampling (see below). For each backbone torsion, the PMF was calculated by combining the sampled conformations of all 16 RNA molecules using WHAM.<sup>54–56</sup>



**Figure 2.** Dihedral term potential energy (kcal/mol) as a function of dihedral angle for Amber99 (red),<sup>25</sup> Amber ff10 (Zgarbova et al. and  $\alpha/\gamma$  bsc0; green),<sup>17,18</sup>  $\chi$ , and this work (blue). For comparison, the average energy of each curve was set to zero.



**Figure 3.** Dihedral potentials of mean force (PMF) for nine torsions. Shown are the new dihedrals from this work (blue), Amber ff10<sup>17,18</sup> with  $\chi$  and  $\alpha/\gamma$  bsc0 correction (green), and a statistical potential derived from a set of crystal structures in the PDB (red). For bins of statistical data where there were no representatives in the pdb, the points are not plotted. The PMF is an average of 16 RNA dinucleotide molecules in explicit solvent (TIP3P) water model.

For the glycosidic dihedral, the umbrella sampling was performed for each base of the dinucleotide, one at a time. The  $\chi$

PMF for each base type was determined by adding all sampling for that base type.

Each of the 16 RNA molecules was neutralized with a  $\text{Na}^+$  and then solvated with a cubic TIP3P water box. The box dimensions were set so that the RNA was at least 10 Å from the edge of the box. Finally, 0.1 M NaCl was added based on number of water molecules. Each starting structure was energy minimized for 2500 steps of steepest descent followed by 2500 steps of conjugate gradient minimization. The minimized structure was heated at constant volume from 0 to 300 K over 20 ps with a time step of 1 fs, then equilibrated at 300 K and 1 atm for 100 ps with a time step of 1 fs. Next, a series of 200 ns restrained MD simulations was performed using the previous restart file as a starting structure. The dihedral angle restraint had a force constant of 200 kcal/(mol × rad<sup>2</sup>), with a width of  $\pm 100^\circ$ , where the potential was flat outside this width. Simulations were at constant temperature of 300 K and a pressure of 1 atm. For both the scanning and umbrella sampling simulations, SHAKE<sup>57,58</sup> was used to constrain bonds involving hydrogen atoms, allowing a 2 fs time step. All simulations were performed with periodic boundary conditions, using particle mesh Ewald<sup>59,60</sup> with an 8 Å direct space cutoff with the default grid spacing of 1 Å. The Langevin integrator was used with a collision frequency of 1 ps<sup>-1</sup>.

**Statistical Potential.** The statistical potentials along each dihedral were generated by binning dihedrals observed in all crystal structures of 3 Å or better resolution from the PDB, which included 69 523 dinucleotides.<sup>30</sup> The statistical free energy profile was computed for dihedrals in 5° bins according to

$$P_{\text{stat}}(i) = -RT \times \ln\left(\frac{\text{bin}(i)}{\max(\text{bin})}\right) \quad (9)$$

where  $R$  is the gas constant,  $T = 300$  K,  $\text{bin}(i)$  is number of structures found with torsions in bin  $i$ , and  $\max(\text{bin})$  is the value of the most populated bin for a given dihedral angle. Normalizing using  $\max(\text{bin})$  ensures that the largest bin value (the most probable/lowest free energy state) will have a free energy of 0 kcal/mol.

**Estimating RMSD for Umbrella Sampling.** The umbrella sampling PMF and statistical potential were compared using RMSD

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^N (\text{PMF}(i) - P_{\text{stat}}(i))^2}{N}} \quad (10)$$

where  $\text{PMF}(i)$  is umbrella sampling PMF value and  $P_{\text{stat}}(i)$  is the statistical potential. The sum is over all bins.

**Force Fields and Simulation Software.** All simulations were performed with the Amber 14 molecular dynamic simulation package.<sup>24,32</sup> All simulations used the TIP3P water model.<sup>61</sup> Two sets of Amber force field parameters were used. The first set was with Amber ff10 (specified in the Amber leap programs using leaprc.RNA.OL3).<sup>25–27</sup> The second set of simulations was with the new backbone and glycosidic torsion dihedrals from this work.

**Molecular Dynamics Setup.** MD was performed on two duplexes, two hairpin loops, and five tetramers. For all MD simulations, four independent simulations were performed. All molecules were neutralized with  $\text{Na}^+$  ions and then solvated with a TIP3P water box of equal length sides. The box dimensions were set so that the RNA was no less than 10 Å from the edge of the box. Finally, 0.1 M NaCl was added based on number of water molecules. Starting structures for five tetramers (namely AAAA, CCCC, CAAU, GACC, and UUUU) were built as starting structures using the NAB module of

**Table 4. RMSD of Umbrella Sampling of Torsion Angles with Respect to Experimental Statistical Potential<sup>a</sup>**

torsion	ff10	this work
$\alpha$	1.12	0.60
$\beta$	1.36	1.41
$\gamma$	2.24	1.41
$\epsilon$	2.31	1.86
$\zeta$	1.37	0.92
$\chi$ adenine	1.54	1.95
$\chi$ guanine	1.56	2.02
$\chi$ cytosine	2.14	1.52
$\chi$ uracil	2.02	1.37

<sup>a</sup>RMSD is calculated in 5° bin interval. Unit of RMSD is kcal/mol. Bins that were empty because there were no examples in the PDB were excluded from the RMSD calculation.

AmberTools.<sup>24</sup> In the case of NMR solution structures, the selected starting structures were the reported model found to have the lowest potential energy after energy minimization using ff10 with TIP3P water. Details are shown in Table 3.

**Explicit Solvent Simulations.** Tetramers, hairpins, and duplex RNA molecules were energy minimized, with the RNA atoms restrained with a force constant of 25 kcal/(mol × Å<sup>2</sup>) for 1000 steps of steepest descent, followed by a 1000 steps of conjugate gradient minimization. The final minimized structure was heated at constant volume from 0 K to a final of 150 K over 50 ps followed by a second heating at constant pressure from 150 to 300 K in another 50 ps. During the initial heating stage, the RNA was held fixed in space with a harmonic potential of 25 kcal/(mol × Å<sup>2</sup>); the spring constant was reduced to 5 kcal/(mol × Å<sup>2</sup>) for the second stage.

The final step before production was to equilibrate at constant temperature of 300 K and a pressure of 1 atm for 5 ns with a time step of 2 fs. During the equilibration, the RNA was held fixed in space with a harmonic potential of 0.5 kcal/(mol × Å<sup>2</sup>). For the simulations, SHAKE<sup>57,58</sup> was used to constrain bonds involving hydrogen atoms and the time step was 2 fs. Temperature was controlled with Langevin dynamics with a collision frequency of 2 ps<sup>-1</sup>. All simulations were performed with periodic boundary conditions using Particle Mesh Ewald<sup>59,60</sup> and an 8 Å direct space cutoff.

**Comparison to NMR Distance Measurements.** To estimate how consistent the sampled structures are, compared to NMR estimates of proton distances from NOEs, the RMSD was calculated

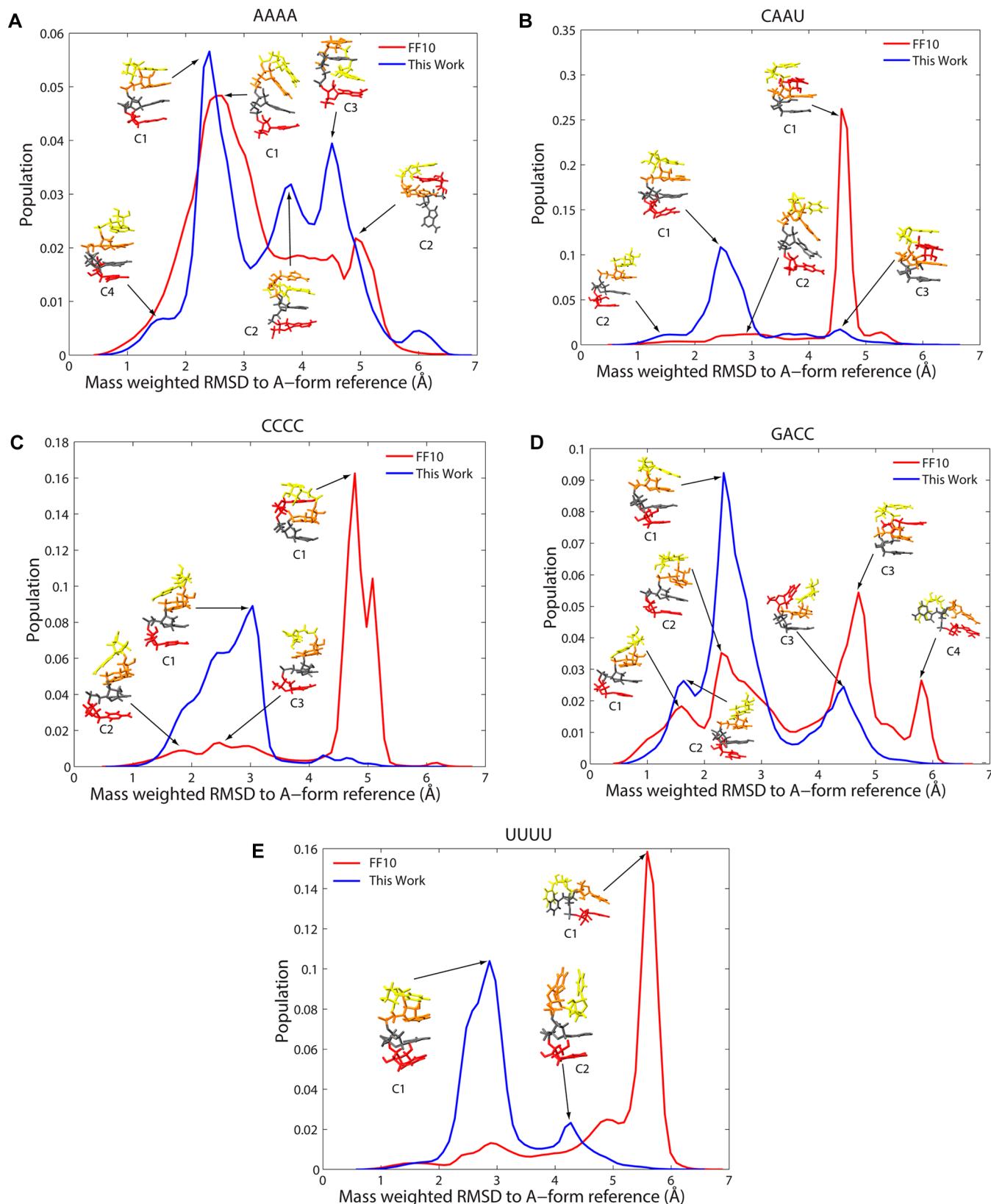
$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^d (D_{i\text{MD}} - D_{i\text{NMR}})^2}{d}} \quad (11)$$

where  $d$  is the number of distances estimated by NMR,  $D_{i\text{MD}}$  is the distance estimated by the simulation, and  $D_{i\text{NMR}}$  is the distance estimated by NMR.  $D_{i\text{MD}}$  is an average distance that is weighted by the negative sixth power of the distance to reflect that NMR NOEs decay by the sixth power of the nuclear distance

$$D_{i\text{MD}} = \left( \sum_{i=1}^N \frac{r_i^{-6}}{N} \right)^{-1/6} \quad (12)$$

where  $r_i$  is the distance in a single snapshot of the trajectory and  $N$  is the number of snapshots.

**Clustering of Tetramer Conformations.** Conformations were sampled every 100 ps from the trajectories and clustered by



**Figure 4.** Histogram of mass-weighted RMSD to A-form-like reference. Histograms are provided for AAAA (panel A), CAAU (panel B), CCCC (panel C), GACC (panel D), and UUUU (panel E). Each histogram was generated by merging the conformations for four independent simulations. The bin widths are 0.01 Å. For major peaks in the histogram, corresponding centroid structures from clustering (Table 5) are labeled.

RMSD. Dbscan clustering<sup>62</sup> was used as implemented in the cpptraj<sup>35</sup> module of Amber. The minimum distance between

points required for forming a cluster was 1.2 Å and the minimum number of points required for forming a cluster was 100.

**Table 5. Clustering Results from the Combined Trajectories of All Five Tetramers Run Using the Current Amber Force Field (ff10) and the Force Field Derived in This Work<sup>a</sup>**

sequence	force field	% noise	cluster no.	% of frames	average distance within cluster (Å)	standard deviation within cluster (Å)	average distance between clusters (Å)	RMSD of centroid to A-form (Å)
AAAA	ff10	32.7	1	55.9	2.67	0.72	4.49	1.89
			2	4.2	1.08	0.38	4.02	5.27
			3	2.5	0.78	0.24	3.73	5.01
			4	2	1.39	0.35	4.33	4.42
	this work	39.7	1	28.1	1.67	0.54	3.89	2.42
			2	9.2	2.18	0.67	3.80	3.68
			3	5.5	1.53	0.40	3.63	4.74
			4	5.4	1.68	0.43	4.16	1.33
CAAU	ff10	13.3	1	73.3	0.97	0.48	4.57	4.60
			2	8.8	2.00	0.53	3.26	2.65
			3	3.5	1.94	0.65	3.79	1.00
			4	0.7	0.96	0.36	4.04	4.16
	this work	19.1	1	66.0	1.43	0.52	3.04	2.52
			2	10.0	1.42	0.45	3.47	1.46
			3	3.6	1.06	0.36	3.39	4.56
			4	1.0	0.88	0.28	3.13	4.00
CCCC	ff10	6.2	1	72.7	1.13	0.43	3.75	4.87
			2	11.4	1.76	0.76	4.02	1.91
			3	7.7	1.88	0.64	3.97	2.78
			4	0.8	0.99	0.42	3.81	4.40
	this work	3.3	1	93.9	1.71	0.64	4.11	2.25
			2	1.6	1.25	0.42	4.96	4.76
			3	1.1	0.83	0.27	5.21	4.40
			4	1.1	0.83	0.27	5.21	4.40
GACC	ff10	20.0	1	21.3	1.66	0.77	3.93	1.46
			2	19.6	1.76	0.61	3.98	2.61
			3	7.7	1.34	0.48	3.85	4.83
			4	6.9	1.11	0.35	5.43	5.89
	this work	15.0	1	45.0	1.43	0.53	3.78	2.47
			2	28.0	1.18	0.41	3.78	1.74
			3	3.8	0.79	0.24	4.23	4.44
			4	2.6	0.92	0.30	4.38	4.75
UUUU	ff10	28.5	1	39.8	1.52	0.61	4.41	5.64
			2	9.9	1.27	0.40	4.43	5.79
			3	8.8	2.04	0.57	4.52	2.67
			4	4.7	1.27	0.40	4.29	5.04
	this work	18.5	1	72.8	1.43	0.60	3.06	2.78
			2	3.7	0.88	0.33	4.33	4.28
			3	3.3	1.41	0.43	3.23	1.53
			4	1.0	1.22	0.29	3.67	3.71

<sup>a</sup>Only the top 4 clusters are presented. The sizes of the remaining clusters (if they existed) were always less than 5% of total frames in trajectory. Starting from the left, columns denote molecule type, force field used, % of noise frames (those that did not get placed in a cluster), cluster number, % of frames within that cluster, average distance and standard deviation between elements of that cluster, average distance between that and all other clusters, and finally mass-weighted RMSD of the centroid of the cluster to the A-form conformation of the respective molecule.

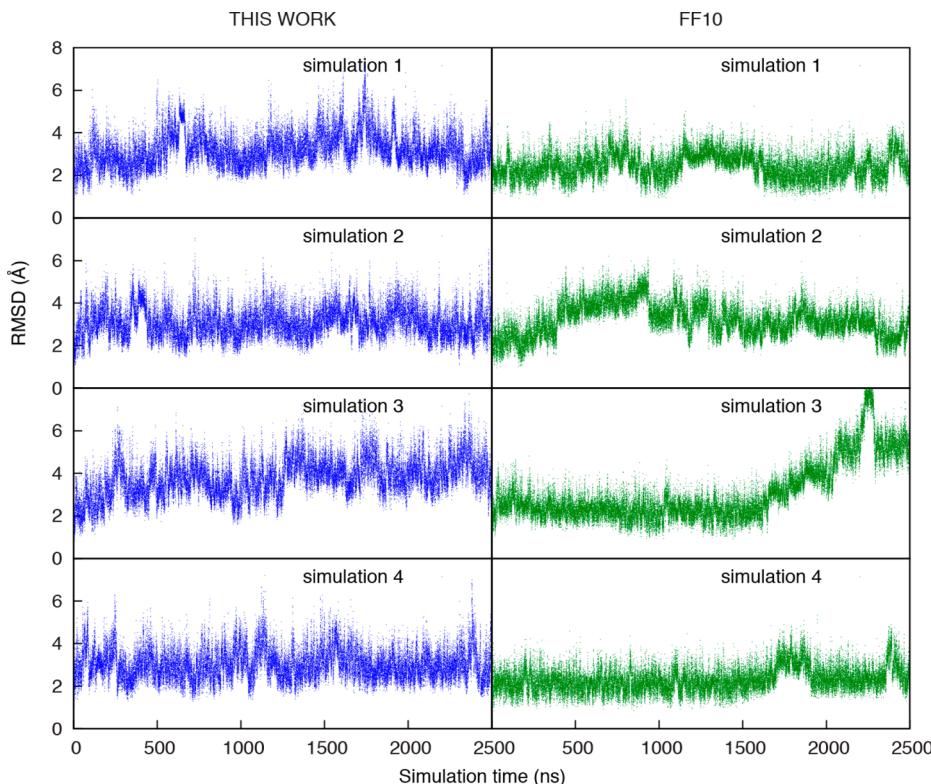
**Analysis of A-Form Helices.** Calculation of structural parameters for the A-RNA double helix were performed using cpptraj module of AmberTools<sup>35</sup> and 3DNA software.<sup>63</sup> For exact definitions of structural parameters presented in Table 6, see Lu et al.<sup>64</sup>

## RESULTS

**Dihedral Energy Profiles.** The dihedral potential energy for  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\epsilon$ ,  $\zeta$ , and  $\chi$  of the four bases as a function of dihedral angle are shown in Figure 2. For comparison, Amber ff99, Amber ff10 (the conventional parameters), and this work are included in the plot. Each dihedral potential energy is sum of all paths for that particular torsion. The energy is offset such that the average energy of an individual energy profile is zero.

Following the convention of Zgarbová et al.,<sup>18</sup> the glycosidic dihedrals were defined with a phase difference of  $180^\circ$  as compared to the standard convention; to be compatible with standard definition of  $\chi$ , all four glycosidic torsions were shifted back by  $180^\circ$ .<sup>25</sup>

**Potentials of Mean Force Along Fitted Dihedrals.** Umbrella sampling was performed in explicit solvent to calculate the potential of mean force (PMF) for each of the nine torsions with conventional Amber ff10 and the new parameters from this work. A statistical potential was also estimated from the population of torsions from the X-ray structures found in the PDB. The PMFs shown in Figure 3 are offset such that each has a minimum value of zero, and the glycosidic torsion PMF was shifted by  $180^\circ$  to be consistent with the Amber force field.<sup>25</sup>



**Figure 5.** Comparison of Amber ff10 (right panels, green) to the dihedral parameters fit in this work (left panels, blue) for dynamics of the Watson–Crick duplex, 5' U(UA)<sub>6</sub>A 3'.<sup>66</sup> Mass-weighted atomic RMSD to the solution structure is shown as a function of time for four independent simulations. The higher RMSD for ff10 of simulation 3 is an unfolding of four nucleotides of the 5' terminal pairs. For comparison, these RMSDs as a function of time when excluding terminal base pairs are provided in Figure S5. The same trends are seen in both plots.

As shown in Figure 3, the PMFs calculated with the new parameters from this work are closer to the statistical potential than Amber ff10 (shown using with RMSD in Table 4), with the exception of the adenine and guanine  $\chi$  dihedrals. The disagreement with the statistical potential is partly because adenine and guanine are known to not be at their global energy minima with respect to the  $\chi$  in A-form helices; in other words, the population of  $\chi$  values does not reflect the local energetics of this torsion, but rather frustration with other favorable interactions that stabilize A-form.<sup>18</sup> The PMF calculated from the umbrella sampling is an average of all 16 RNA dinucleotide sequences. Overall, the results in this work show the same trends as the ff10<sup>25–27</sup> parameters in terms of positions of minima and barrier heights in free energy, with a few exceptions. For example, the new  $\alpha$  PMF has a similar overall shape and local minimum at about 130° as the statistical potential; by contrast, the PMF calculated with ff10 has a local maximum at 150° that is not reflected in the statistical potential (Figure 3). Moreover, the new  $\gamma$  PMF is similar to the statistical potential except around 120°, where the barrier height of the new dihedrals (and the conventional ff10 dihedrals) is higher than that suggested by the statistical potential.

**Testing the New Dihedral Parameters Using Single-Stranded Tetramers.** Tetramers are an excellent model to test an RNA force field, especially for backbone and glycosidic torsions, because their stability depends on the accuracy of the backbone parameters and water model. Moreover, their short length and fast dynamics in solution make adequate statistical sampling computationally tractable. A collection of five tetra-nucleotides r(GACC),<sup>65</sup> r(CCCC),<sup>5</sup> r(UUUU),<sup>3</sup> r(AAAA),<sup>3</sup> and r(CAAU)<sup>3</sup> were previously studied by NMR spectroscopy.

Stacking, puckers, and backbone torsions for these sequences are generally consistent with A-form-like conformations but with some deviations. Spectra for r(GACC), r(AAAA), and r(CAAU) suggest minor conformations with C2'-endo 3'-end sugars.<sup>3,65</sup> r(GACC) and r(CCCC) have minor populations where the sugar in the C4 position is inverted.<sup>5,65</sup> r(UUUU) data indicates that all sugars are mostly in C2'-endo orientation and the nucleotides are dynamic.<sup>3</sup> Sampling with the current ff10 RNA force field, using either conventional MD or replica exchange,<sup>3,4,65</sup> yields structures with the intercalated bases that are not observed in the NMR solution experiment.<sup>3,65</sup>

In this work, four independent MD trajectories were run for each of AAAA, CAAU, CCC, GACC, and UUUU, starting from A-form-like configurations. Additionally, three other starting structures for CCC were used for 12 additional trajectories (4 for each starting structure) to test convergence. Table 3 details the simulations that were run. Simulations were run using both the force field from this work and the conventional Amber ff10.

The tetramers sample a number of conformations in the simulations as they do in solution. Following prior practice,<sup>4</sup> the trajectories were clustered and the dominant clusters compared to A-form-like structures and to the NMR data. Figure 4 shows histogram plots of RMSD to A-form-like structures, annotated with structures of cluster centroids, and Table 5 details each cluster.

A major finding of these simulations is that the intercalated structure, where a base intercalates between other bases in the sequence in an unexpected order in an A-form-like structure, is not found in the simulations using the newly fit parameters. This intercalated structure was observed in prior simulations of ff10, including the CCC tetramer in 2013 by Tubbs et al.,<sup>5</sup>

**Table 6.** Mean Structural Parameters for Watson–Crick Duplex  $[U(UA)_6A]_2^a$ 

	this work	ff10	X-ray
local base-pair parameters			
shear [Å]	0.02 (0.9)	0.0 (0.5)	-0.1 (0.4)
stretch [Å]	-0.05 (0.5)	0.01 (0.3)	-0.2 (0.1)
stagger [Å]	0.08 (0.5)	-0.04 (0.6)	-0.01 (0.2)
buckle [deg]	0.03 (13)	0.5 (15)	1.0 (5)
propeller [deg]	-12.2 (12)	-14.1 (14)	-18.8 (2)
opening [deg]	-0.3 (12)	1.9 (9)	0.1 (3)
local base-pair step parameters			
shift [Å]	0.0 (0.8)	-0.01 (0.7)	0.03 (0.4)
slide [Å]	-1.4 (0.7)	-1.2 (0.6)	-1.3 (0.1)
rise [Å]	3.1 (0.8)	3.1 (0.9)	3.3 (0.2)
tilt [deg]	0.0 (6)	0.07 (6)	-0.2 (3)
roll [deg]	8.5 (8)	12.2 (9)	10.7 (5)
twist [deg]	29.1 (10)	27.7 (8)	31.1 (5)
local base-pair helical parameters			
X-displacement [Å]	-3.7 (3)	-3.8 (2)	-4.1 (1)
Y-displacement [Å]	0.01 (2)	0.03 (2)	-0.1 (0.7)
helical rise [Å]	2.6 (1)	2.3 (1)	2.7 (0.3)
inclination [deg]	14.5 (14)	21.1 (15)	19.5 (10)
tip [deg]	-0.03 (11)	-0.1 (10)	0.02 (6)
helical twist [deg]	31.4 (12)	31.6 (10)	33.3 (4)

<sup>a</sup>Closing base-pairs have been left out of analysis to avoid the noise from end fraying. Data presented is the average over all other nucleotides and over all four trajectories. The second column represents results using the revised force field from this work, and the third column represents the simulations with the conventional Amber force field. The last column contains the values from the starting X-ray structure (PDB 1RNA). The values in parentheses are standard deviations.

dynamics of four RNA tetramers in 2015 by Condon et al.<sup>3</sup> and the highly sampled replica exchange dynamics on GACC and CCCC by Bergonzo et al.<sup>4</sup> There is no evidence for this intercalated structure in the NMR experiments, suggesting that this structure is at most a small component of the ensemble (if it is present at all). In the simulations here using ff10, intercalated structures were frequently observed. For example, the dominant clusters for ff10 for CAAU (Figure 4B), CCCC (Figure 4C), and GACC (Figure 4D) are intercalated structures.

Figure S1 shows RMSD to NMR distances, estimated from NOEs, as a function of time for ff10 (right panels) and the new parameters from this work (left panels) for the tetramers. Figure S3 compares the RMSD for NMR proton distances as a function of time for alternative starting structures for CCCC, and these confirm that the A-form-like structures are sampled when starting from alternative conformations. Because NOEs decay as a function of  $1/r^6$ , where  $r$  is atomic distance, Figure S4 compares the  $r^{-6}$ -weighted average of proton distances for protons with distances estimated from the NMR experiments. The figure shows the weighted average distance for the simulations using the new parameters, for simulations using the conventional ff10 parameters, and for the NMR experiments.

An additional set of four simulations was performed using CAAU to test convergence. Four snapshots of intercalated structures were manually chosen from the ff10 simulations, and these structures were used as starting conformations. Figure S2 shows the RMSD for the NMR proton distances, and the results confirm simulations with the new dihedral parameter are

**Table 7.** Mean Values of Backbone and Glycosidic Torsions, as Well as Sugar Puckers, from the Simulations of Watson–Crick Duplex  $[U(UA)_6A]_2^a$ 

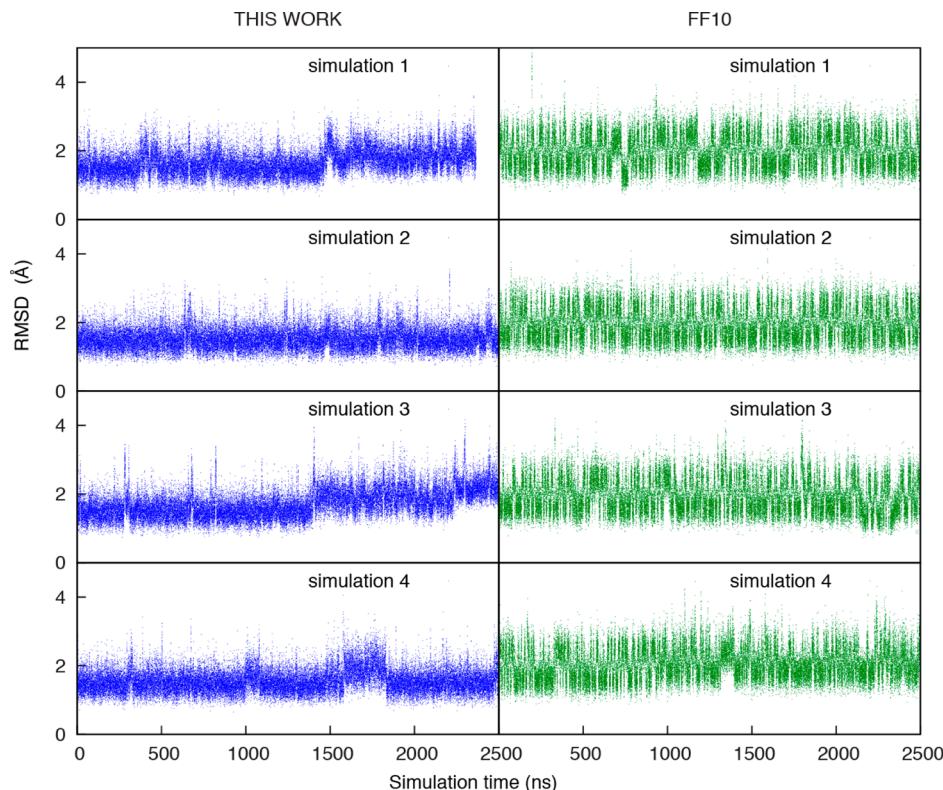
	this work	ff10	X-ray	A-RNA PDB average
$\alpha$ [deg]	263 (66) <sup>b</sup>	277 (33)	287 (36) <sup>c</sup>	295 (8)
$\beta$ [deg]	167 (32)	174 (11)	171 (14)	174 (8)
$\gamma$ [deg]	86 (54) <sup>d</sup>	64 (16)	65 (31) <sup>e</sup>	54 (6)
$\delta$ [deg]	83 (15)	81 (13)	80 (5)	81 (3)
$\epsilon$ [deg]	217 (18)	203 (16)	213 (17)	212 (10)
$\zeta$ [deg]	268 (54)	284 (40)	280 (12)	289 (7)
$\chi$ [deg]	200 (20)	208 (17)	201 (9)	NA
pucker (% C3'-endo)	95.7	97.1	100	NA

<sup>a</sup>Closing base-pairs have been left out to avoid noise from fraying of ends. The numbers are values in degrees averaged over all internal nucleotides and over all trajectories, except for pucker where % of C3'-endo conformation is indicated. The second column are the results using the parameters derived in this work, and the third column are the results from the conventional Amber current force field. The fourth column contains average values taken from the starting X-ray structure (PDB ID 1RNA). The fifth column contains averages from analysis of many structures with the A-RNA conformation taken from work of Richardson et al.<sup>5</sup> The values in parentheses are standard deviations. The notes indicate angles were the average is affected by the presence of secondary peak. See Figure S6 for more details.<sup>b</sup>20% of population is in a trans conformation with a peak at 150°, 76.4% is in a gauche- population with a peak at 295°, and the remaining 3.6% is in gauche+ confirmation.<sup>c</sup>The X-ray structure has 2 out of 24 nucleotides in trans position, that is, 8%; the rest is gauche-.<sup>d</sup>21.5% of population is in a trans conformation with a peak at 180°; the remaining 78.5% is in a gauche+ population with a peak at 58°.<sup>e</sup>The X-ray structure has 2 out of 24 nucleotides in trans position, that is, 8%; the rest is gauche+.

able to leave the intercalated state and return to the A-form-like conformation.

**Simulations of Duplexes.** To further test the revised dihedral parameters, long ( $>1 \mu s$ ) simulations were run on RNA duplexes and hairpin loops, where solution structures are available for comparison. For example, the Watson–Crick duplex,  $[U(UA)_6A]_2$ ,<sup>66</sup> was studied because of its tendency to form a “ladder-like” structure with the ff99 force field.<sup>12,18,48,67</sup> The atomic RMSD as compared to the starting structure was calculated for both the current Amber force field (ff10) and the parameters from this work (Figure 5); both force field parameters reproduced the structural well, and did not show any tendency to form the “ladder-like” structure. The third simulation of the Amber force field (ff10) has two terminal base pairs unpaired at one end of the helix and four terminal base pairs unpaired at the other end, as shown by high RMSD on Figure 5.

Additionally, the helical parameters were calculated across the trajectories of  $[U(UA)_6A]_2$ , and these are provided in Table 6 as mean values. The terminal base pairs were excluded from the analysis. Both the current Amber parameters and the parameters from this work show good agreement with the X-ray values, with the mean values falling between bounds set by standard deviations of values from the X-ray structure. The exceptions are the mean helical rise for simulations with the current Amber force field and the propeller twist calculated with both parameters sets. The propeller twists for both simulations with both force field parameters are outside X-ray bounds, but the distributions of observed conformations in the simulations are also broad.



**Figure 6.** Comparison of Amber ff10 (right panel, green) and the dihedral parameters fit in this work (left panels, blue) for dynamics of 5'GGUGAAGGC3'/3'CCGAAGCCG5' (major conformation).<sup>71</sup> Mass-weighted atomic RMSD to the solution structure is shown as a function of time for four independent simulations. The higher RMSD for ff10 is a result of stem nucleotides U3 and C17 frequently moving from their base pairing partner and flipping out into solution.

Table 7 presents mean values of backbone and glycosidic dihedrals and sugar pucker from the simulations of  $[U(UA)_6A]_2$ . Again, the terminal base-pairs were not included to avoid noise because those pairs have a more conformational freedom than the base pairs on the interior of the helix. The sugar pucker values are the percentage of structures in C3'-endo conformation, where the remaining conformations are C2'-endo pucksers. Using the dihedral parameters from this work,  $\alpha$  and  $\gamma$  the distributions have two peaks. Approximately 15–20% of the population is in a trans position, of approximately 150° for  $\gamma$  and 180° degrees for  $\alpha$ . This roughly matches the X-ray structure, as nucleotides 11 and 24 both have alpha and gamma in trans positions. In contrast, the conventional Amber force field includes  $\alpha/\gamma$  parameters derived to prevent trans positions in B-form DNA,<sup>17</sup> which completely prevent trans conformations. In view of the observation that  $\alpha/\gamma$  trans/trans conformation is commonly seen in RNA structures,<sup>68–70</sup> the ability of the force field derived in this work to visit these conformations is possibly an advantage. Figure S6 contains normalized populations of each dihedral for simulations with the force field from this work and also the conventional Amber force field.

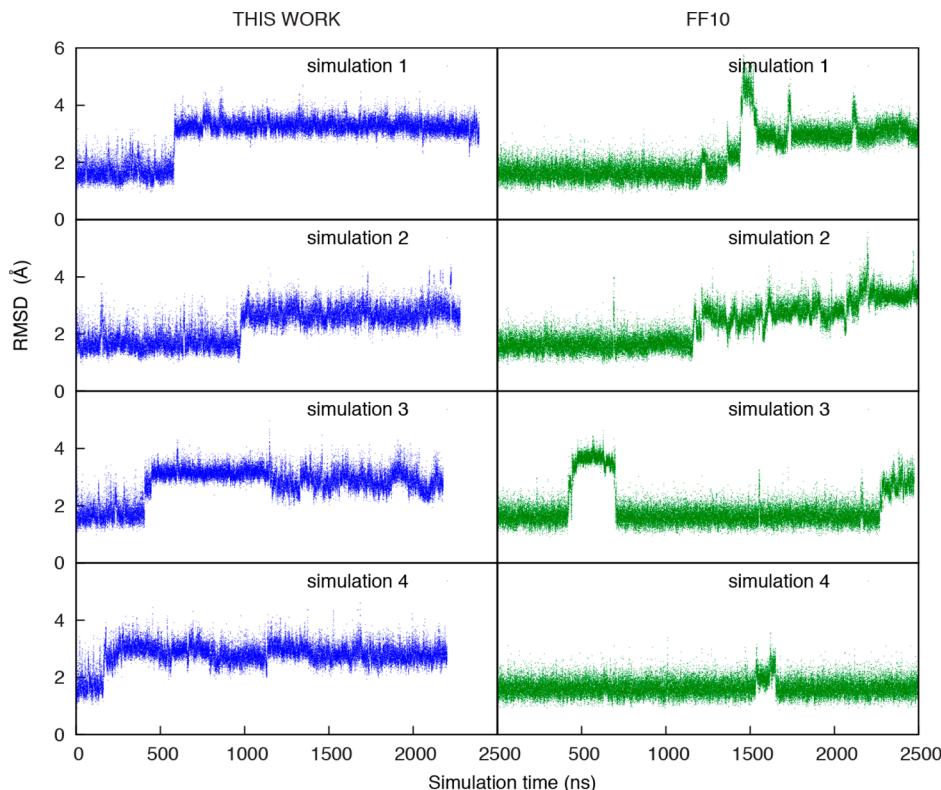
Another duplex with an internal loop (5'GGUGAAGGC3'/3'CCGAAGCCG5'), which is known by NMR to populate two conformations, was also studied.<sup>71</sup> In prior work, it was shown<sup>11</sup> that ff99 + bsc0<sup>17</sup> incorrectly stabilized the minor conformations over the major conformation.<sup>71</sup> For this work, simulations were run starting from the major NMR conformation. Figure 6 shows RMSD as a function of time for four separate 2.5  $\mu$ s simulations with ff10 (right panels) and the parameters fit in this work (left panels). The new parameters perform

better than the conventional ff10 parameters. Using ff10, nucleotides U3 and C17 frequently leave their base-pairing partner to flip out into the solvent.

**Simulations of Tetraloops.** There are solution structures for a number of tetraloops, and prior work found that sets of signature interactions are lost in tetraloop simulations using ff99, ff10, CHARMM36, and f99 + Chen–Garcia.<sup>12–14,72–75</sup> Even though the UUCG hairpin loop 2KOC (5' GGCACU-UCGGUGCC 3')<sup>29</sup> is not stable for microsecond-time scale simulations, in this work a 2.5  $\mu$ s simulation was used, as shown in Figure 7 to test the new parameters.<sup>4,14,76</sup> The result (Figure 7, left panel) was a higher RMSD compared to ff10 starting at 200–500 ns and remaining for the duration of the simulation. This is because U7 and C8 tend to be more dynamic with the new parameters, as compared to conventional ff10. On the other hand, the GCAA tetraloop with eight nucleotides, derived from 1ZIH (5' GCGCAAGC 3'),<sup>77</sup> is more stable and stays close to the solution structure using the parameters from this work. In contrast, with ff10 the structure in the loop is unstable, and subsequently the whole structure denatures within 2.5  $\mu$ s (Figure 8).

## ■ DISCUSSION AND CONCLUSION

In this work, the  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\varepsilon$ ,  $\zeta$ , and  $\chi$  dihedral parameters for the Amber RNA force field were reparametrized simultaneously using multiple linear regression on full nucleotide and dinucleotide models. This differs substantially from prior practice for nucleic acids, where the parameters were fit for one dihedral at a time for a scan around that dihedral.<sup>17,18</sup> Also, prior work for nucleic acids used simplified models for dihedral scans of



**Figure 7.** Comparison of Amber ff10 (right panel; green) and the dihedral parameters fit in this work (left panel; blue) for dynamics of the UUCG tetraloop, 2KOC.<sup>29</sup> Mass-weighted atomic RMSD to the solution structure is shown as a function of time for four independent simulations. The higher RMSD for simulations with parameters derived in this work is a reflection of C8 (the loop C) leaves the conformation of the solution structure and becomes either exposed to solvent or extends into the helix major groove.

sugar–phosphate backbones.<sup>17</sup> The work of Vanommeslaeghe et al.<sup>46</sup> and Hopkins and Roitberg<sup>47</sup> showed an approach of force field fitting using linear least-squares fitting as a promising tool for generating dihedral parameters for molecular mechanics force fields. An advantage of this approach is that the optimal set of parameters, i.e. those that minimize the least squared deviation, is guaranteed from the fit. This work used the linear least squared approach to fit multiple dihedral parameters for RNA. Additionally, structures for fitting were generated with more diverse conformations than those from traditional potential energy scan-fits, sampling combinations of multiple dihedrals that would not be sampled with one-at-a-time fitting.<sup>24</sup> The diverse configurations here were obtained from X-ray and dihedral scans, without geometry optimization other than regularization of the bond lengths and angles. The improvement in RNA parameters was clearly seen by performance of simulations of tetramers that failed with the conventional Amber force field.<sup>3–5</sup>

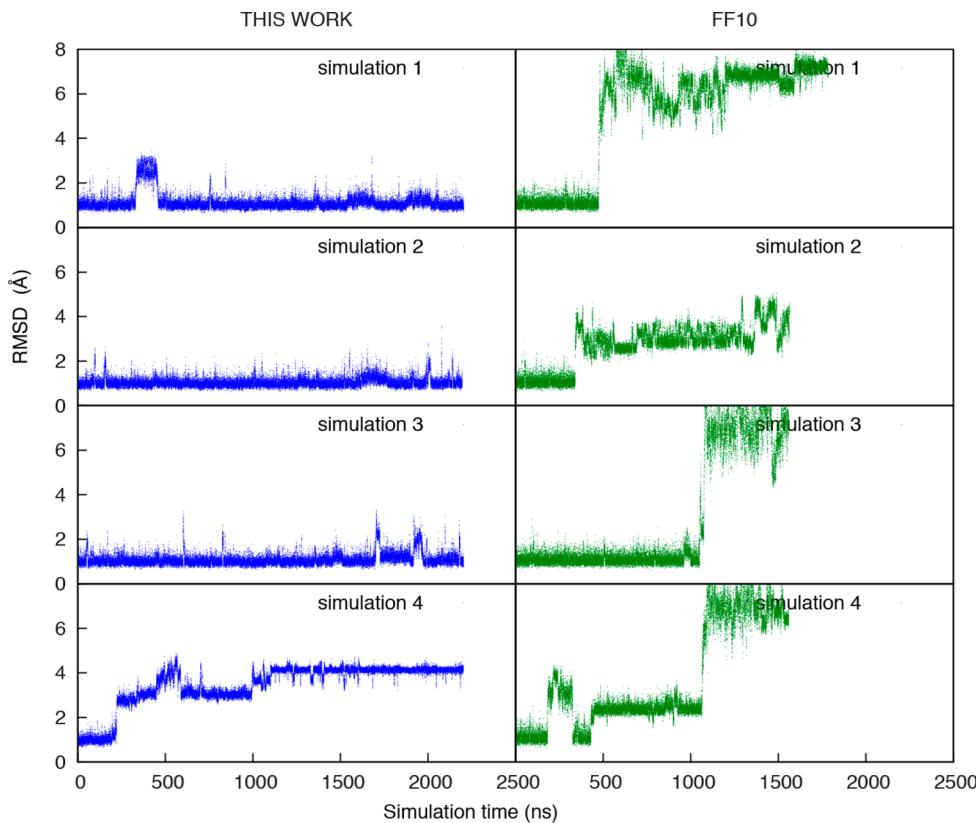
An alternative approach for improving force field dihedral parameters also recently appeared.<sup>78</sup> In that work, the  $\alpha$ ,  $\beta$ ,  $\varepsilon$ , and  $\zeta$  dihedrals of Amber14 were revised using metadynamics<sup>79,80</sup> for dinucleotides to match dihedral distributions derived from high-resolution crystal structures. Implicitly, this changes the parameters so that the simulation free energies match the statistical potentials for dihedrals. This work shares a similarity in that multiple dihedrals were revised simultaneously, without freezing all but one degree of freedom. In contrast, Gil-Ley et al. revise dihedrals to explicitly match the dihedrals observed in the database of crystal structures.<sup>78</sup> As they point out, there is likely a bias toward A-form helices in the database, which might prevent their parameters from being

transferrable to all RNA systems. Tests using the AAAA, CCCC, and GACC tetramers did, however, show improved agreement with NMR data as compared to Amber14. In this work, improved agreement with quantum-calculated potential energies is the figure of merit for the fit. This should result in parameters that are transferable widely, given the Amber force field functional form.

The potential energy landscapes of the newly fitted dihedrals (Figure 2) show a large number of differences in barrier and minima positions as compared to the conventional Amber force field. In the conventional force field,  $\beta$  and  $\varepsilon$  are simple sinusoidal functions, but show a lot more structure in the newly fitted parameters. Interestingly,  $\varepsilon$  has much higher potential energy barriers, although the free energy barriers (Figure 3) are lower for the newly fitted parameters. This arises because the free energy surface is a function of the other force field potentials, and the high dimensional correlation with other dihedrals.<sup>81</sup> This highlights the importance of calculating free energy surfaces for comparison of dihedrals, rather than simply comparing potential energy surfaces.

The potential of mean force (PMF) calculations showed good agreement with the positions of local minima as well as the global minimum as compared to the statistical potential for most of backbone dihedrals as shown in Figure 3. Because PMFs were calculated in explicit water using dinucleotides, the surfaces should be comparable with the statistical population in the PDB because it includes the influence of stacking, which should favor helix formation.

This work used the existing Amber force field functional form for dihedrals, which uses the same parameter values for terminal nucleotides as for nucleotides interior to a sequence.



**Figure 8.** Comparison of Amber ff10 (right panel; green) and the dihedral parameters fit in this work (left panel; blue) for dynamics of the GCAA tetraloop, 1ZIH.<sup>77</sup> Mass-weighted atomic RMSD to the solution structure is shown as a function of time for four independent simulations. The higher RMSD for both ff10 and this work is due to unfolding of the loop region away from the solution structure.

The QM potential energy calculations were performed on dinucleotides, and it was assumed that terminal nucleotides will have the same dihedral potential energies as interior nucleotides. It is possible this assumption is not correct; in particular, the terminal nucleotides have terminal OH groups on the 5' and 3' carbons instead of connections to phosphate groups. The good performance of the newly fitted parameters for simulations of duplexes supports the use of this assumption, but future work might test the assumption and improve the framework.

From the simulation benchmarks, the most striking improvement is in the simulations of tetramers. Prior work showed that the conventional ff10 tended to intercalate the fourth nucleotide between other nucleotides,<sup>3–6,65</sup> although NMR data suggested the major conformations are largely A-form-like.<sup>3,5,65</sup> These new parameters appear to correct this artifact. On the other hand, some problems persist. Most notably, the UUCG tetraloop solution structure<sup>29</sup> is not stable with these new torsions in ff10, conventional ff10, the Chen & Garcia force field, or CHARMM27.<sup>4</sup> One promising direction for continued force field development is to continue the focus on optimizing van der Waals parameters,<sup>13–15,82</sup> as these affect base stacking, hydrogen bonding, and nucleotide-solvent interactions.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.jctc.6b00870](https://doi.org/10.1021/acs.jctc.6b00870).

Summary of the revised parameters, plots of RMSD of NMR distance for trajectories of the five tetramers as a

function of time, RMSD of NMR distances for the CAAU tetramer starting from an intercalated structure, RMSD of CCCC tetramer with three different starting structures,  $r^6$ -weighted proton distances compared to NOE-estimated distances for all five tetramers, RMSD as a function of time for the A-form duplex,  $[U(UA)_6A]_2$ , histograms of dihedral populations from the simulations of  $[U(UA)_6A]_2$ , crystal structures used for generating the dinucleotide conformations for fitting, example output and restraint result from “rdparm” for the AA dinucleotide, restraint input for regularization, and values of the fitted dihedral parameters ([PDF](#))

Force field parameters and input files compatible with tleap ([ZIP](#))

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [David\\_Mathews@urmc.rochester.edu](mailto:David_Mathews@urmc.rochester.edu).

### ORCID

David H. Mathews: [0000-0002-2907-6557](https://orcid.org/0000-0002-2907-6557)

### Author Contributions

<sup>#</sup>A.H.A. and A.S. contributed equally to this work.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was supported by NIH grant R01 GM076485 to D.H.M. and R01 GM095496 to A.G. Computer time was provided by the University of Rochester Center for Integrated

Research Computing. The authors thank our anonymous reviewers for their careful reviews and specific suggestions.

## ■ REFERENCES

- (1) Dixit, S. B.; Beveridge, D. L.; Case, D. A.; Cheatham, T. E., III; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Sklenar, H.; Thayer, K. M.; Varnai, P. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II: sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. *Biophys. J.* **2005**, *89*, 3721–3740.
- (2) Beveridge, D. L.; Barreiro, G.; Suzie Byun, K.; Case, D. A.; Cheatham, T. E., III; Dixit, S. B.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Seibert, E.; Sklenar, H.; Stoll, G.; Thayer, K. M.; Varnai, P.; Young, M. A. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. research design and results on d(CpG) steps. *Biophys. J.* **2004**, *87*, 3799–3813.
- (3) Condon, D. E.; Kennedy, S. D.; Mort, B. C.; Kierzek, R.; Yildirim, I.; Turner, D. H. Stacking in RNA: NMR of four tetramers benchmark molecular dynamics. *J. Chem. Theory Comput.* **2015**, *11*, 2729–2742.
- (4) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Cheatham, T. E. Highly sampled tetranucleotide and tetraloop motifs enable evaluation of common RNA force fields. *RNA* **2015**, *21*, 1578–1590.
- (5) Tubbs, J. D.; Condon, D. E.; Kennedy, S. D.; Hauser, M.; Bevilacqua, P. C.; Turner, D. H. 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.
- (6) Bergonzo, C.; Cheatham, T. E., III Improved force field parameters lead to a better description of RNA structure. *J. Chem. Theory Comput.* **2015**, *11*, 3969–3972.
- (7) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Swails, J. M.; Roitberg, A. E.; Cheatham, T. E. Multidimensional replica exchange molecular dynamics yields a converged ensemble of an RNA tetranucleotide. *J. Chem. Theory Comput.* **2014**, *10*, 492–499.
- (8) Henriksen, N. M.; Roe, D. R.; Cheatham, T. E. Reliable oligonucleotide conformational ensemble generation in explicit solvent for force field assessment using Reservoir replica exchange molecular dynamics simulations. *J. Phys. Chem. B* **2013**, *117*, 4014–4027.
- (9) Yildirim, I.; Stern, H. A.; Sponer, J.; Spackova, N.; Turner, D. H. Effects of restrained sampling space and nonplanar amino groups on free-energy predictions for RNA with imino and sheared tandem GA base pairs flanked by GC, CG, iGiC or iCiG base pairs. *J. Chem. Theory Comput.* **2009**, *5*, 2088–2100.
- (10) Aytenfisu, A. H.; Spasic, A.; Seetin, M. G.; Serafini, J.; Mathews, D. H. Modified Amber force field correctly models the conformational preference for tandem GA pairs in RNA. *J. Chem. Theory Comput.* **2014**, *10*, 1292–1301.
- (11) Van Nostrand, K. P.; Kennedy, S. D.; Turner, D. H.; Mathews, D. H. Molecular mechanics investigation of an adenine-adenine non-canonical pair conformational change. *J. Chem. Theory Comput.* **2011**, *7*, 3779–3792.
- (12) Banáš, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham, T. E.; Šponer, J.; Otyepka, M. Performance of molecular mechanics force fields for RNA simulations: Stability of UUCG and GNRA hairpins. *J. Chem. Theory Comput.* **2010**, *6*, 3836–3849.
- (13) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell, A. D. Impact of 2'-hydroxyl sampling on the conformational properties of RNA: Update of the CHARMM all-atom additive force field for RNA. *J. Comput. Chem.* **2011**, *32*, 1929–1943.
- (14) Chen, A. A.; García, A. E. High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 16820.
- (15) Kührová, P.; Best, R. B.; Bottaro, S.; Bussi, G.; Šponer, J.; Otyepka, M.; Banas, P. Computer Folding of RNA Tetraloops: Identification of Key Force Field Deficiencies. *J. Chem. Theory Comput.* **2016**, *12*, 4534–4548.
- (16) Bottaro, S.; Banas, P.; Šponer, J.; Bussi, G. Free Energy Landscape of GAGA and UUCG RNA Tetraloops. *J. Phys. Chem. Lett.* **2016**, *7*, 4032–4038.
- (17) Pérez, A.; Marchán, I.; Svozil, D.; Šponer, J.; Cheatham, T. E., III; Laughton, C. A.; Orozco, M. Refinement of the Amber force field for nucleic acids: Improving the description of  $\alpha/\gamma$  conformers. *Biophys. J.* **2007**, *92*, 3817–3829.
- (18) Zgarbova, M.; Otyepka, M.; Šponer, J.; Mladek, A.; Banáš, P.; Cheatham, T. E.; Jurecka, P. Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.* **2011**, *7*, 2886–2902.
- (19) Yildirim, I.; Stern, H. A.; Kennedy, S. D.; Tubbs, J. D.; Turner, D. H. Reparameterization of RNA  $\chi$  torsion parameters for the Amber force field and comparison to NMR spectra for cytidine and uridine. *J. Chem. Theory Comput.* **2010**, *6*, 1520–1531.
- (20) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; Mackerell, A. D., Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.
- (21) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 712–25.
- (22) 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*, 3696–713.
- (23) Mackerell, A. D., Jr.; Feig, M.; Brooks, C. L., 3rd Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400–15.
- (24) Case, D. A.; Berryman, J. T.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Monard, G.; Needham, P.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Salomon-Ferrer, R.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; York, D. M.; Kollman, P. A. *Amber 2015*; University of California: San Francisco, CA, 2015.
- (25) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (26) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845–862.
- (27) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (28) Havrlá, M.; Zgarbova, M.; Jurecka, P.; Banas, P.; Krepl, M.; Otyepka, M.; Šponer, J. Microsecond-Scale MD Simulations of HIV-1 DIS Kissing-Loop Complexes Predict Bulged-In Conformation of the Bulged Bases and Reveal Interesting Differences between Available Variants of the AMBER RNA Force Fields. *J. Phys. Chem. B* **2015**, *119*, 15176–90.
- (29) Nozinovic, S.; Fürtg, 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.* **2010**, *38*, 683–694.
- (30) Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Feng, Z.; Gilliland, G. L.; Iype, L.; Jain, S.; Fagan, P.; Marvin, J.; Padilla, D.; Ravichandran, V.; Schneider, B.; Thanki, N.; Weissig, H.; Westbrook, J. D.; Zardecki, C. The Protein

- Data Bank. *Acta Crystallogr, Sect. D: Biol. Crystallogr.* **2002**, *58*, 899–907.
- (31) Berman, H. M.; Westbrook, J.; Feng, Z.; Iype, L.; Schneider, B.; Zardecki, C. The Nucleic Acid Database. *Acta Crystallogr, Sect. D: Biol. Crystallogr.* **2002**, *58*, 889–898.
- (32) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (33) Macke, T. J.; Case, D. A., Modeling unusual nucleic acid structures. In *Molecular Modeling of Nucleic Acids*; American Chemical Society: Washington, DC, 1997; Vol. 682, pp 379–393.
- (34) Arnott, S.; Hukins, D. W.; Dover, S. D.; Fuller, W.; Hodgson, A. R. Structures of synthetic polynucleotides in the A-RNA and A'-RNA conformations: x-ray diffraction analyses of the molecular conformations of polyadenylic acid–polyuridylic acid and polyinosinic acid–polycytidylc acid. *J. Mol. Biol.* **1973**, *81*, 107–122.
- (35) 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*, 3084–3095.
- (36) Anandakrishnan, R.; Drozdetski, A.; Walker, R. C.; Onufriev, A. V. Speed of conformational change: comparing explicit and implicit solvent molecular dynamics simulations. *Biophys. J.* **2015**, *108*, 1153–1164.
- (37) Onufriev, A. V.; Sigalov, G. A strategy for reducing gross errors in the generalized Born models of implicit solvation. *J. Chem. Phys.* **2011**, *134*, 164104.
- (38) Onufriev, A. Implicit solvent models in molecular dynamics simulations: A brief overview. In *Annual Reports in Computational Chemistry*; Spellmeyer, R. A. W.; David, C., Eds.; Elsevier: 2008; Vol. 4, Chapter 7, pp 125–137.
- (39) Antony, J.; Grimme, S. Density functional theory including dispersion corrections for intermolecular interactions in a large benchmark set of biologically relevant molecules. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5287.
- (40) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.
- (41) Rezac, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427–2438.
- (42) Goerigk, L.; Kruse, H.; Grimme, S. Benchmarking density functional methods against the S66 and S66 × 8 datasets for non-covalent interactions. *ChemPhysChem* **2011**, *12*, 3421–33.
- (43) Kruse, H.; Mladek, A.; Gkionis, K.; Hansen, A.; Grimme, S.; Sponer, J. Quantum chemical benchmark study on 46 RNA backbone families using a dinucleotide unit. *J. Chem. Theory Comput.* **2015**, *11*, 4972–91.
- (44) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision D.01; Gaussian, Inc.: Wallingford, CT, 2009.
- (45) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal solvation model based on solute Electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.
- (46) Vanommeslaeghe, K.; Yang, M.; MacKerell, A. D. Robustness in the fitting of molecular mechanics parameters. *J. Comput. Chem.* **2015**, *36*, 1083–1101.
- (47) Hopkins, C. W.; Roitberg, A. E. Fitting of dihedral terms in classical force fields as an analytic linear least-squares problem. *J. Chem. Inf. Model.* **2014**, *54*, 1978–1986.
- (48) Zgarbová, M.; Luque, F. J.; Šponer, J.; Otyepka, M.; Jurečka, P. A novel approach for deriving force field torsion angle parameters accounting for conformation-dependent solvation effects. *J. Chem. Theory Comput.* **2012**, *8*, 3232–3242.
- (49) Sigalov, G.; Fenley, A.; Onufriev, A. Analytical electrostatics for biomolecules: beyond the generalized Born approximation. *J. Chem. Phys.* **2006**, *124*, 124902.
- (50) Sigalov, G.; Scheffel, P.; Onufriev, A. Incorporating variable dielectric environments into the generalized Born model. *J. Chem. Phys.* **2005**, *122*, 094511.
- (51) Tsui, V.; Case, D. A. Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers* **2000**, *56*, 275–91.
- (52) Tsui, V.; Case, D. A. Molecular dynamics simulations of nucleic acids with a generalized born solvation model. *J. Am. Chem. Soc.* **2000**, *122*, 2489–2498.
- (53) Anderson, E.; Bai, Z.; Bischof, C.; Blackford, L.; Demmel, J.; Dongarra, J.; Du Croz, J.; Greenbaum, A.; Hammarling, S.; McKenney, A.; Sorensen, D. *LAPACK Users' Guide*; Society for Industrial and Applied Mathematics, 1999; p 424.
- (54) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Comput. Chem.* **1995**, *16*, 1339–1350.
- (55) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
- (56) Grossfield, A. WHAM: The Weighted Histogram Analysis Method, version 2.0.8. <http://membrane.urmc.rochester.edu/content/wham>.
- (57) 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.
- (58) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952–962.
- (59) Toukmaji, A.; Sagui, C.; Board, J.; Darden, T. Efficient particle-mesh Ewald based approach to fixed and induced dipolar interactions. *J. Chem. Phys.* **2000**, *113*, 10913–10927.
- (60) Sagui, C.; Pedersen, L. G.; Darden, T. A. Towards an accurate representation of electrostatics in classical force fields: Efficient implementation of multipolar interactions in biomolecular simulations. *J. Chem. Phys.* **2004**, *120*, 73–87.
- (61) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (62) Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. In *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*, KDD-96, Portland, Oregon; Simoudis, V.; Han, J.; Fayyad, U., Eds. AAAI: Portland, OR, 1996; pp 226–231.
- (63) Lu, X. J.; Olson, W. K. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.* **2008**, *3*, 1213–27.
- (64) Lu, X.-J.; Olson, W. K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **2003**, *31*, 5108–5121.
- (65) Yildirim, I.; Stern, H. A.; Tubbs, J. D.; Kennedy, S. D.; Turner, D. H. Benchmarking AMBER force fields for RNA: Comparisons to NMR spectra for single-stranded r(GACC) are improved by revised  $\chi$  torsions. *J. Phys. Chem. B* **2011**, *115*, 9261–9270.

- (66) Dock-Bregeon, A. C.; Chevrier, B.; Podjarny, A.; Johnson, J.; de Bear, J. S.; Gough, G. R.; Gilham, P. T.; Moras, D. Crystallographic structure of an RNA helix: [U(UA)<sub>6</sub>A]<sub>2</sub>. *J. Mol. Biol.* **1989**, *209*, 459–474.
- (67) Zgarbová, M.; Otyepka, M.; Šponer, J.; Lankáš, F.; Jurečka, P. Base pair fraying in molecular dynamics simulations of DNA and RNA. *J. Chem. Theory Comput.* **2014**, *10*, 3177–3189.
- (68) Besseova, I.; Otyepka, M.; Reblova, K.; Sponer, J. Dependence of A-RNA simulations on the choice of the force field and salt strength. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10701–11.
- (69) Besseova, I.; Banas, P.; Kuhrova, P.; Kosinova, P.; Otyepka, M.; Sponer, J. Simulations of A-RNA duplexes. The effect of sequence, solute force field, water model, and salt concentration. *J. Phys. Chem. B* **2012**, *116*, 9899–916.
- (70) Richardson, J. S.; Schneider, B.; Murray, L. W.; Kapral, G. J.; Immormino, R. M.; Headd, J. J.; Richardson, D. C.; Ham, D.; Hershkovits, E.; Williams, L. D.; Keating, K. S.; Pyle, A. M.; Micallef, D.; Westbrook, J.; Berman, H. M. RNA backbone: consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA* **2008**, *14*, 465–81.
- (71) Chen, G.; Kennedy, S. D.; Qiao, J.; Krugh, T. R.; Turner, D. H. An alternating sheared AA pair and elements of stability for a single sheared purine-purine pair flanked by sheared GA pairs in RNA. *Biochemistry* **2006**, *45*, 6889–6903.
- (72) Butcher, S. E. Solution structure of a GAAA tetraloop receptor RNA. *EMBO J.* **1997**, *16*, 7490–7499.
- (73) Jucker, F. M.; Pardi, A. Solution structure of the CUUG hairpin loop: a novel RNA tetraloop motif. *Biochemistry* **1995**, *34*, 14416–14427.
- (74) Varani, G.; Cheong, C.; Tinoco, I. Structure of an unusually stable RNA hairpin. *Biochemistry* **1991**, *30*, 3280–3289.
- (75) Foloppe, N.; MacKerell, A. D., Jr. 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*, 86–104.
- (76) Kührová, P.; Banáš, P.; Best, R. B.; Šponer, J.; Otyepka, M. Computer folding of RNA tetraloops? Are we there yet? *J. Chem. Theory Comput.* **2013**, *9*, 2115–2125.
- (77) Jucker, F. M.; Heus, H. A.; Yip, P. F.; Moors, E. H. M.; Pardi, A. A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol.* **1996**, *264*, 968–980.
- (78) Gil-Ley, A.; Bottaro, S.; Bussi, G. Empirical Corrections to the Amber RNA Force Field with Target Metadynamics. *J. Chem. Theory Comput.* **2016**, *12*, 2790–8.
- (79) White, A. D.; Dama, J. F.; Voth, G. A. Designing free energy surfaces that match experimental data with metadynamics. *J. Chem. Theory Comput.* **2015**, *11*, 2451–60.
- (80) Marinelli, F.; Faraldo-Gomez, J. D. Ensemble-Biased Metadynamics: A Molecular Simulation Method to Sample Experimental Distributions. *Biophys. J.* **2015**, *108*, 2779–82.
- (81) Tabei, Y.; Kiryu, H.; Kin, T.; Asai, K. A fast structural multiple alignment method for long RNA sequences. *BMC Bioinf.* **2008**, *9*, 33.
- (82) Steinbrecher, T.; Latzer, J.; Case, D. A. Revised AMBER parameters for bioorganic phosphates. *J. Chem. Theory Comput.* **2012**, *8*, 4405–4412.