Modified Amber Force Field Correctly Models the Conformational Preference for Tandem GA pairs in RNA

Asaminew H. Aytenfisu,[†] Aleksandar Spasic,[†] Matthew G. Seetin,[†] John Serafini,[†] and David H. Mathews*,[†,§]

[†]Department of Biochemistry & Biophysics and Center for RNA Biology, University of Rochester Medical Center, 601 Elmwood Avenue, Box 712, Rochester, New York 14642, United States

[§]Department of Biostatistics and Computational Biology, University of Rochester Medical Center, 601 Elmwood Avenue, Box 712, Rochester, New York 14642, United States

Ⓢ Supporting Information

ABSTRACT: Molecular mechanics with all-atom models was used to understand the conformational preference of tandem guanine-adenine (GA) noncanonical pairs in RNA. These tandem GA pairs play important roles in determining stability, flexibility, and structural dynamics of RNA tertiary structures. Previous solution structures showed that these tandem GA pairs adopt either imino (cis Watson−Crick/Watson−Crick A-G) or sheared (trans Hoogsteen/sugar edge A-G) conformations depending on the sequence and orientation of the adjacent closing base pairs. The solution structures (GCGGACGC)$_2$ [*Biochemistry*, **1996**, *35*, 9677−9689] and (GCGGAUGC)$_2$ [*Biochemistry*, **2007**, *46*, 1511−1522] demonstrate imino and sheared conformations for the two central GA pairs, respectively. These systems were studied using molecular dynamics and free energy change calculations for conformational changes, using umbrella sampling. For the structures to maintain

their native conformations during molecular dynamics simulations, a modification to the standard Amber ff10 force field was required, which allowed the amino group of guanine to leave the plane of the base [*J. Chem. Theory Comput.*, **2009**, *5*, 2088−2100] and form out-of-plane hydrogen bonds with a cross-strand cytosine or uracil. The requirement for this modification suggests the importance of out-of-plane hydrogen bonds in stabilizing the native structures. Free energy change calculations for each sequence demonstrated the correct conformational preference when the force field modification was used, but the extent of the preference is underestimated.
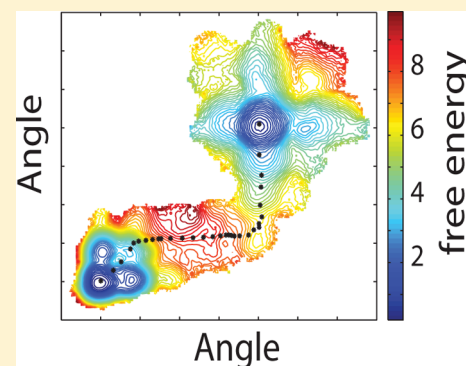
■ INTRODUCTION

RNA plays an important role in biological processes, including information storage, protein expression, catalysis, and regulation of gene expression.[1−12] RNA functions can involve conformational changes.[13−20] Therefore, understanding its secondary and tertiary structure and flexibility is important to understand its functions.

RNA internal loops are unpaired or noncanonically paired nucleotides closed on both sides by canonical base pairs, GU, GC, or AU. These loops provide one of the locations in RNA structures for tertiary or quaternary contacts. One of these internal loop motifs is the tandem guanine−adenine (GA) pair shown in Figure 1. Prior solution structures demonstrate that the conformational preference depends on the sequence identity of the closing base pairs.[21,22] Table 1 summarizes these findings. When adjacent to GC closing base pairs, an imino conformation (cis Watson−Crick/Watson−Crick A-G) for the tandem GA pairs is observed.[22] When the C is changed to U in the closing pairs, making GU pairs, a sheared conformation (trans Hoogsteen/sugar edge A-G) for the tandem GA pairs is observed.[21] It is unclear what interactions underlie this sequence-dependent conformational preference.

Molecular dynamics (MD) simulations can be used to study the structure and dynamics of RNA and other biomolecules. Some of the most commonly used software packages for all-atom MD are Amber, CHARMM, Gromacs, and NAMD.[23−26] The accuracy of these packages depends on the set of equations and parameters used to describe all the interactions in the system, called the force field. Currently, the force fields most commonly used for nucleic acids are the ones derived for the Amber and CHARMM packages, although most force fields can be used in any MD package with modifications. The latest version of the Amber force field for nucleic acids, distributed with Amber, was named ff10 (and was unchanged in the newest ff12SB force field).[27] This force field is based on the ff94 force field of Cornell et al. and its subsequent modifications.[28] Initially, this force field was improved by changing the sugar pucker and glycosidic torsions in force fields ff98 and ff99.[29,30] Later, a new set of torsion parameters called bsc0 was derived for α- and γ-torsions.[31] In 2010 and 2011, two new sets of glycosidic torsions were derived for each base, namely Chi_YIL

A)

B)
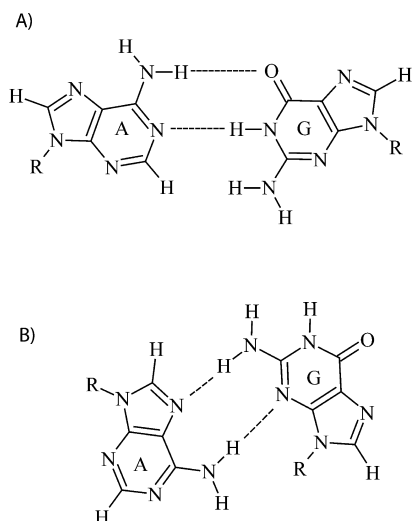
Figure 1. Hydrogen-bonding pattern for two conformations of a GA pair.[65] (A) The cis Watson−Crick/Watson−Crick (imino) interaction and (B) the trans-Hoogsteen/Sugar edge (sheared) A-G interaction. In each panel, one of two base pairs in the tandem pair is shown.

Table 1. Conformational Preference of Tandem GA pairs, Where Systems 1 and 4 (boldface) are Experimentally[21,22] Preferred Conformations, whereas Systems 2 and 3 Were Control Structures Based on Systems 4 and 1, Respectively[a]

Sequence	Imino	Sheared	Source
5'-GCG^{GA}**C**GC-3' 3'-CG**C**$_A$GGCG-5'	**1**	2	1MIS
5'-GCG^{GA}**U**GCU-3' 3'-UCG**U**$_A$GGCG-5'	3	**4**	2IRO

[a]The conformational preference is altered by changing closing GC pairs (system 1) with closing GU pairs (system 4). Source is the protein data bank accession number.[21,22]

and ChiOL, respectively, with ChiOL being incorporated into ff10.[32,33] Beyond ff10, an additional revision of the van der Waals interaction strengths for the Amber RNA force field was recently reported.[34] This work uses the ff10 force field as the starting point. The current CHARMM force field for nucleic acids, CHARMM36 is based on the CHARMM27 force field. Specifically, CHARMM36 improves the description of nucleic acids by reparametrizing the torsion of 2'-hydroxyl group of RNAs and sugar pucker and also several backbone torsions to better describe conformation equilibrium of DNAs.[35−37]

In this work, MD was used to understand the interactions that lead to the observed conformational preference of GA pairs. One focus was the role of the amino group in the G base. The GA imino pair has two hydrogen bonds per GA pair but leaves the guanine amino unpaired. Likewise, the GA sheared pair has two hydrogen bonds per GA pair, but one of which involves the guanine amino group.[38] Previous high level quantum calculations for the amino group were performed, and these showed that the amino groups do not need to be confined to the plane of the base.[39−42] A subsequent molecular dynamics study also found that the unpaired amino group in the imino pair stabilized the interaction by forming a weak out-of-plane hydrogen bond. The current Amber force field for RNA (referred to as ff10) prevents this interaction by requiring the amino group to stay in the plane of the base.[43] For this work, force field parameters of Yildirim et al. (referred to as modified ff10) were used to model this out-of-plane behavior.[38]

Yildirim et al. investigated the sequence-dependent stability of GA pairs using thermodynamic integration.[38] They studied tandem GA pairs flanked by GC, CG, iGiC, or iCiG base pairs, where iG and iC denote isoguanosine and iocytidine, which have amino and carbonyl groups transposed relative to guanosine and cytidine.[38] They used four solution structures (PDB numbers 1MIS, 2O83, 2O81, and 1YFV),[22,44,45] and they used thermodynamic integration to calculate free energy differences. These free energy differences were compared to experiments using thermodynamic cycles. Their results agreed with the experiments when they used the modified force field for the unpaired amino group of guanine in the tandem GA pairs. Their findings demonstrated that the out-of-plane hydrogen bonds were important to improve the accuracy of the result for the imino conformation.[38]

In this work, the conformational preference for GA pairs was investigated using molecular dynamics and two-dimensional umbrella sampling for both GC and GU closing base pairs. The system with GC closing pairs (1MIS) was previously studied as part of the free energy perturbation studied by Yildirim et al.[38] The system with GU closing pairs (2IRO) has not, to our knowledge, been studied by MD.

To examine conformational preference, control structures were constructed to represent conformations not observed (systems 2 and 3 in Table 1). The structures for both solution structures (systems 1 and 4 in Table 1) were maintained for unrestrained molecular dynamics when the modified ff10 force field was used. Umbrella sampling between sheared and imino conformations was performed along reaction coordinates suggested by nudged elastic band calculations of low potential energy conformational change pathways.[46,47] The free energy differences demonstrated that Amber can correctly model the conformational preference of tandem GA pairs.

■ METHODS

Force Fields and Software. All simulations were performed with the Amber 11 or Amber 12 simulation packages.[25,27] Two force fields were used for explicit solvent molecular dynamics simulations with the TIP3P water model.[48] The first set was with Amber ff10, which is based on the Amber ff99 force field,[28−30] with corrections to the α- and γ-torsions[31] and corrections to the glycosidic bond torsions.[33] The second set of simulations was with a modified ff10 that allows the exocyclic amine of guanine to leave the plane of the base.[38,43] The modified ff10 was used for guanines only in the loop regions of the structures. This modified ff10 applies the corrections for the amino dihedrals[38] to ff10, and these corrections had previously only been used in conjunction with ff99. The amino dihedral correction was assumed to be orthogonal to the backbone and gylcosidic corrections. The excellent performance of the modified ff10 force field, in terms of preserving the native loop structures for multiple 1 μs duration simulations, supports this assumption.

Model Structures. The starting structures were solution structures by Wu and Turner[22] (system 1) and Tolbert et al.[21]

(system 4), summarized in Table 1. The 3′ dangling uracils for system 4, present for the experiments,[21] were removed for this work. These are the preferred conformations, and additionally two more control conformations, named system 2 and system 3, were manually constructed. System 2 was generated by changing the two uracils in the closing base pair from system 4 to cytosines (by manually changing atom types). An unrestrained molecular dynamics simulation using Amber ff10 for 10 ns was sufficient to bring the two internal loop closing base pairs (GC) into the expected Watson−Crick pairing pattern.

System 3 was generated from system 1 by replacing the two cytosines of system 1 with uracils (by manually changing atom types). System 4 has GU closing base pairs with one hydrogen bond. To make the closing base pair in system 3 replicate this pair, a targeted molecular dynamics simulation on system 3, using a biasing potential force constant of 1 kcal/mol and performed in explicit solvent, was used on the two closing base pairs with the conformation of system 4 as the target. The simulation was 20 ps, and this was long enough to achieve the change in GU conformation for both closing pairs. The subsequent free molecular dynamics simulation showed a GU wobble interaction was preferred, instead of the single hydrogen bond GU pair observed in system 4.

Molecular Dynamics Setup. All four molecules described in Table 1 were neutralized with 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 ∼600 water molecules. This was 10 Na^+ and 10 Cl^- atoms, and this number of additional salt atoms was used for all simulations, including the umbrella sampling simulations (below). All simulations were performed with periodic boundary conditions using Particle Mesh Ewald[49,50] and a 10 Å direct space cutoff.

Explicit Solvent Simulations. All four structures were energy minimized in two steps. During the first step, RNA was held fixed with a restraint force of 500 kcal/(mol·Å²) and minimized for 2500 steps of steepest descent followed by a 2500 steps of conjugate gradient minimization. During the second step, the whole system was minimized without any restraints for 2500 steps of steepest descent followed by 2500 steps of conjugate gradient minimization. The final energy-minimized structure was heated at constant volume from 0 K to a final of 300 K over 50 ps. During this initial heating stage, the RNA was held fixed in space with a harmonic potential of 10 kcal/(mol·Å²) for 50 ps. The final step before production was to equilibrate at constant temperature of 300 K and a pressure of 1 atm for 100 ps. For the simulations, SHAKE[51,52] was used to restrain bonds involving hydrogen atoms and the time step was 2 fs.

Nudge Elastic Band (NEB). An implicit solvent NEB calculation was used to search the conformational change pathway for each sequence.[53,54] The Generalized Born implicit solvent was used with a 0.1 M NaCl concentration, a 25 Å cutoff for the solvation calculations and a 99 Å cutoff for nonbonded interactions.[55] To identify starting structures, all snapshots from the explicit solvent molecular dynamics were stripped of solvent, and energy-minimized with Generalized Born implicit solvent. These structures were energy minimized in two stages using the Amber ff10 force field. First, a steepest descent minimization for 2500 steps was performed, and this was followed by 2500 steps of conjugate gradient minimization. The lowest potential energy image was then subsequently used

in NEB as an end point structure. For NEB, 18 images copied from the imino conformation were followed by 18 images copied from the sheared conformation to have 36 images in total. Those 36 images were initially heated with a controlled increase of temperature from 0 to 300 K with an NEB spring constant of 10 kcal/(mol·Å²) over 50 ps. This followed by constant temperature equilibration over 100 ps. A simulated annealing calculation (detailed in Supporting Information Table 1) was performed to search the conformational change pathways.[53,54,56] During this pathway search, a time step of 1 fs and a Langevin dynamics collision frequency of 1000 ps⁻¹ were used. For the NEB simulated annealing, the modified ff10 force field was used.

Umbrella Sampling. For each sequence described in Table 1, three umbrella sampling calculations were performed. For system 1 and system 2, a total of 153 windows were used. For system 4 and system 3, 130 windows were used. Each window was sampled for 13 ns. A two-dimension reaction coordinate involving angles between A13 (center mass of N3, C4, and C2), G4 (center mass of N1, C6, and C5), and G4 (atom N3) and between A5 (center mass of N3, C4 and C5), A5 (center mass of N1, C5, and C6), and G12 (atom N3) was chosen to model the conformational change pathway. The MD setup was the same as the unrestrained molecular dynamics setup explained in the above, except a restraining harmonic potential was added during the second step of minimization, and this restraining potential was kept for the rest of the simulation time. This harmonic potential was an angle restraint with a parabolic side with a force constant of 100 kcal/(mol·rad²). Coordinates were written to disk every 0.5 ps.

Free Energy Difference Calculation. The Weighted Histogram Analysis Method[57,58] (WHAM) was used to merge potential of mean force (PMF) surfaces of individual windows into a single PMF profile. This was done using the WHAM software package, version 2.0.7.[59] In the WHAM calculation, 1 ns of equilibration was used, 12 ns of sampling for each umbrella simulation were used, periodicity was off in both dimensions, and the bin size was one degree in each dimension.

Based on the angles observed in unrestrained molecular dynamics, regions on the PMF were chosen to represent the imino and sheared conformations. Then, the free energy difference was calculated using

$$\Delta G_{\text{reactant}\rightarrow\text{product}} = -RT \ln\left(\frac{\sum_{\text{product}} p_{\text{bin}}}{\sum_{\text{reactant}} p_{\text{bin}}}\right) \quad (1)$$

where the terms $\Delta G_{\text{reactant}\rightarrow\text{product}}$ is the free energy difference going from the nonobserved structure to the native structure and p_{bin} was probability of being inside a bin of size 1 degree by 1 degree area. The sums were over all bins attributed to the imino or sheared conformations.

General Analysis. Distance and root-mean-square deviation (RMSD) analyses were calculated using the ptraj module in the AmberTools package.[27] Nucleotides structures were diagrammed with ChemDraw Ultra 12.0 and the PMF plot was generated using Matlab.[60,61]

■ RESULTS

Molecular Dynamics and the Importance of Modified ff10. The initial molecular dynamics simulations were run with the standard Amber ff10 force field.[25] For system 1 and system 4, the native structures shown in Table 1, three independent simulations were run for 1 μs each. Three independent
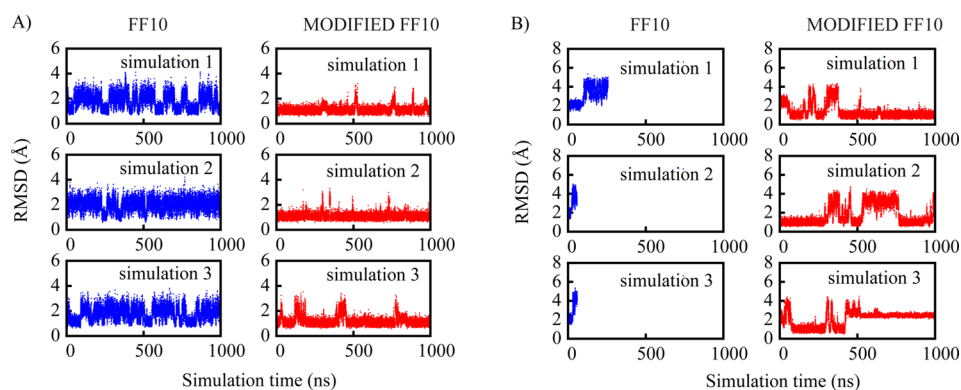
Figure 2. Mass-weighted root-mean-square deviations (RMSD) for the imino-paired structures, system 1 (native structure, panel A) and system 3 (non-native structure, panel B), for nucleotides in the internal loop and the loop-closing base pairs as a function of simulation time. Blue and red plots were the RMSD with ff10 and modified ff10 force fields, respectively. For system 3, the non-native structure, the ff10 simulations were stopped because of the evidence that the modified ff10 could better model the native structure.
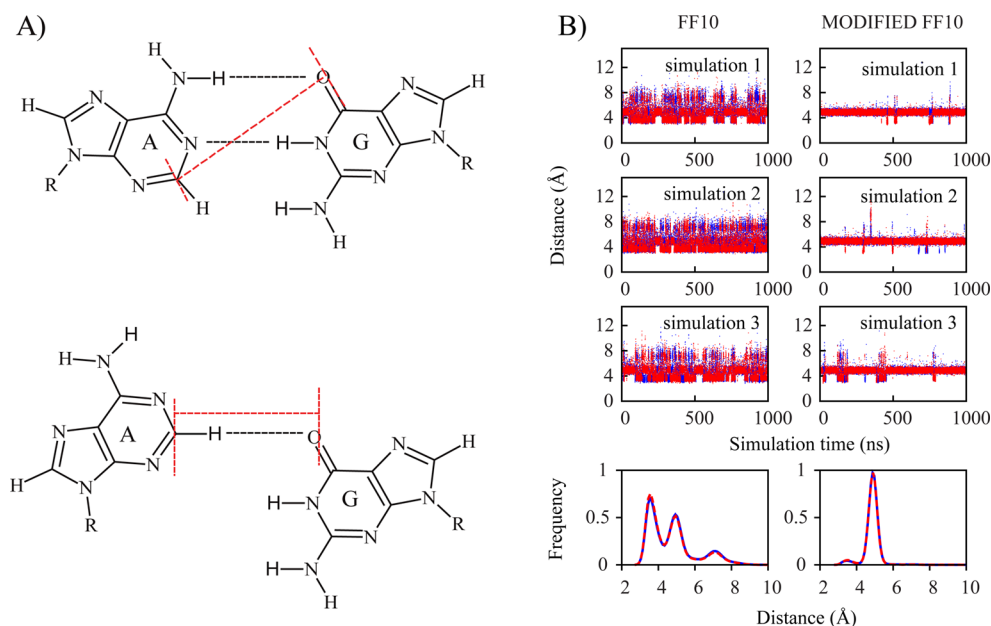

Figure 3. Distance plots of GA pairs of system 1. The distance between C2 of adenine and O6 of guanine was followed as a function of time for each of the GA pairs. (A) The top structure shows the imino pairing as observed in the solution structure. In the solution structure, this distance is 3.93 and 3.92 Å for the two pairs. The bottom structure shows a structure sampled frequently in the simulations using ff10. This structure corresponds to a distance of approximately 3 Å. (B) The top plots show the distance as a function of time for all three simulations. The bottom plots show histograms for distances, aggregated from all three simulations. The left plots are for ff10 and the right plots are for modified ff10. Blue and red colors are for the distance between A13 (atom C2) and G4 (atom O6) and the distance between A5 (atom C2) and G12 (atom O6), respectively. The histogram plots show excellent convergence of the distance for the two pairs because the two plots nearly superimpose. The histograms also show clear differences between ff10 and modified ff10.

trajectories were run for systems 2 and 3, where system 2 was run for 1 μs each, but the system 3 simulations were run for shorter durations. Figure 2 shows the mass-weighted RMSD of the loop regions for the imino-paired systems, including closing base pairs. For ff10, the plots show the native loop structure (system 1) was consistently lost. The structure oscillated between the native structure and a non-native structure, spending a majority of the simulation time at the non-native structure.

Previous work on imino GA pairs had demonstrated the importance of allowing amino groups to leave the plane of the base.[38] Using these modified ff10 parameters, a separate set of simulations were run. For each system, shown in Table 1, a total of three independent simulations of one microsecond duration were run. The native imino pairing structure was preserved for the duration of the simulations (Figure 2). In addition, the structures of the sheared-paired systems, 2 and 4, were maintained using the modified ff10 as shown by RMSD (Supporting Information Figure S1). Because the native structures were more consistently maintained for MD using modified ff10 than for ff10, MD simulations for the imino-paired control structure (system 3) were completed only with modified ff10.

In addition to following the simulations of structures by RMSD, atomic distances were also followed as a function of time. One instructive distance is that between C2 of adenine and O6 of guanine (Figure 3A), which is at about 4 Å in the solution structure. For system 1, this distance demonstrated the
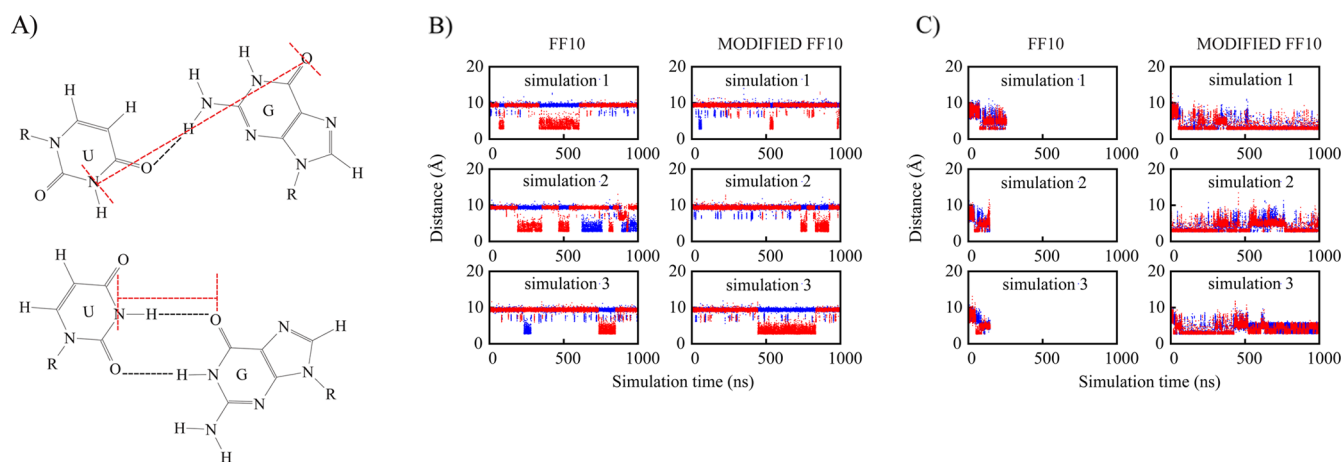
Figure 4. Distance plots for closing GU pairs. The guanine O6 and uracil N3 distance was measured for both closing GU pairs in systems 3 and 4. (A) The top structure is the solution structure (system 4), which has a distance of 9.11 Å and 9.02 Å for the two pairs, and the bottom structure is a wobble GU, with a distance of approximately 3 Å. (B) The distance as a function of time for system 4, the native sheared structure. (C) The distance as a function of time for the imino control structure (system 3). In panels B and C, the ff10 force field is to the left and the modified ff10 force field is to the right. Blue and red colors are for the distance between U14 (atom N3) and 3G (atom O6) and the distance between U6 (atom N3) and G11 (atom O6), respectively.
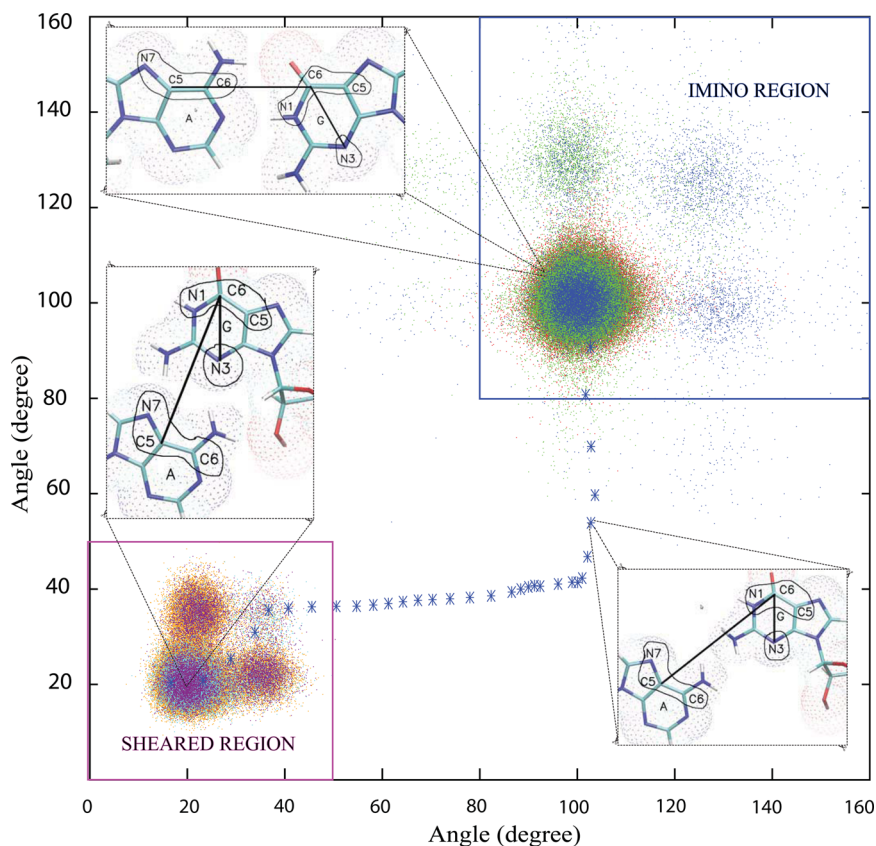
Figure 5. NEB images and unrestrained molecular dynamics as a function of reaction coordinate for molecule (GCGGACGC)$_2$.[22] Blue stars plot the NEB images for one calculation and the dots are snapshots from the unrestrained molecular dynamic simulations. The five NEB calculations showed a consistent path. Snapshots were taken every 10 ps from three unrestrained molecular dynamics calculations (shown in red, green, and blue) for imino (system 1) and (shown in purple, orange and cyan) for sheared (system 2). The reaction coordinate involves angles between A13 (center of mass of N3, C4, and C2), G4 (center of mass of N1, C6, and C5), and G4 (atom N3) on the x-axis and the angle between A5 (center of mass of N3, C4, and C5), G12 (center of mass of N1, C5, and C6), and G12 (atom N3) on the y-axis.

differences in the conformations visited by simulations run with ff10 and modified ff10. Figure 3B provides the distance as a function of time and a histogram plot. For ff10, the distance fluctuated, demonstrating both an alternative base pairing (at a distance of approximately 3 Å) and a loss of base pairing (at a distance of approximately 7 Å). For modified ff10, the distance was more consistent with the imino pairing observed in the solution structure.
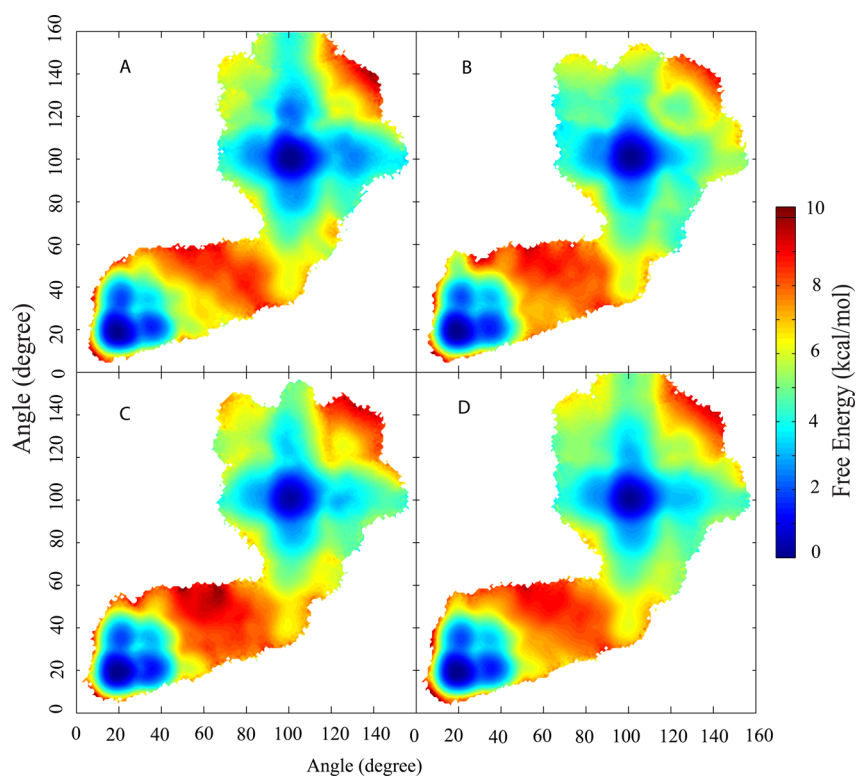
Figure 6. Potential of mean force plots for the conformational change from system 1 to system 2 calculated using umbrella sampling and WHAM. Plots A, B, and C were the PMF corresponding to three independent calculations, whereas plot D was the PMF calculated with the aggregate data. The plots show the angle between A13 (center of mass of N3, C4, and C2), G4 (center of mass of N1, C6, and C5) and G4 (atom N3) on the x-axis and the angle between A5 (center of mass of N3, C4, and C5), G12 (center of mass of N1, C5, and C6) and G12 (atom N3) on the y-axis.

Preference of GU Conformation of System 3 and System 4. The modified ff10 parameters were previously thought to stabilize the imino pairing pattern by allowing the exocyclic amine to participate in other interactions. In this work, the stability of system 4, the sheared conformation, was also improved because modified ff10 stabilized the closing GU base pairs. As shown in the top of Figure 4A, the solution structure has a single hydrogen bond to the uracil O4 (called SHB4 by Tolbert et al.[21]). The loop RMSD also showed the switching from SHB4 to GU wobble interaction as shown by Figure 4B (right column) and Supporting Information Figure S1 (right column). Simulations with ff10 (left of Figure 4B) often visited the wobble GU pair (bottom of Figure 4A), as demonstrated by the distance of uracil N3 to guanine O6. This pairing was also observed to a lesser extent in the modified ff10 simulations (right of Figure 4B). In contrast, the non-native imino pairing (system 3) preferred the GU wobble interaction (bottom of Figure 4A and Figure 4C).

Conformational Preference of Tandem GA Pairs. The simulations provide insight into the atomic interactions that drive the conformational preference of tandem RNA GA pairs. For GA pairs closed by either GC or GU, the out-of-plane[38] interactions between the guanine N2 and O2 of the adjacent paired uracil or cytosine stabilized the imino conformation of the GA pair. For GC closure, however, the additional stabilization is to a greater extent because the average distance between the amino group of G and the O2 of the pyrimidine in the closing pair is shorter when the pyrimidine in the closing pair is C (Supporting Information Figure S2A and B).

The MD snapshots provide an atomic level detail how the interaction of the guanine amino group was affected by closing with either GC or GU pairs. In the case of GU closure (system 3), the O2 of U14 was involved in GU wobble interactions as shown in Supporting Information Figure S2A and the out-of-plane interaction distance between O2 and N2 was generally longer than hydrogen bonding distance. Among all trajectories with modified ff10, 96% of snapshots with GC closure (system 1, native structure) had a distance between N2 of G4 and O2 of C14 of less than 4 Å. For GU closure (system 3, non-native), however, only 64% of snapshots had an O2 of U14 to N2 of G4 distance of less than 4 Å. As shown in panel C of Supporting Information Figure S2, in system 1 the distance between O2 of C14 and N2 of G3, a hydrogen bond for the Watson−Crick pair, is generally less than 3 Å, while the out-of-plane interaction between O2 of C14 and N2 of G4 is simultaneously less than 3 Å. In contrast, in panel D of Supporting Information Figure S2, system 3 shows a significant number of snapshots at hydrogen bonding distance for O2 of U14 and N1 of G3, a hydrogen bond for the wobble pair, but with a long distance between O2 of U14 and N2 of G4. Based on these findings, it appears that a GA pair closed by GU pair prefers the sheared interactions because the GU wobble interaction sequesters the O2 of the U and reduces its out-of-plane interactions with the amino of guanine.

Conformational Change Pathway Search. Reaction coordinates were needed to compute the free energy difference between the imino and sheared conformations. In order to find reasonable pathways, Nudged Elastic Band (NEB) calculations were used. These calculations, detailed in the Methods, were repeated five times for each of the sequences.

The NEB calculations suggested a two-dimensional reaction coordinate (one dimension for each of the tandem pairs),
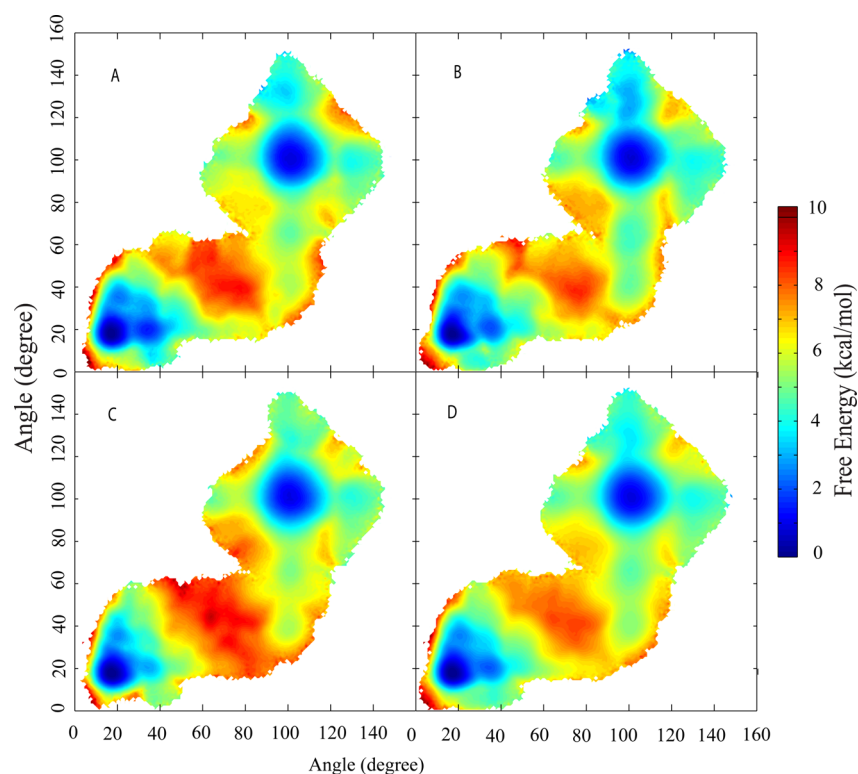
Figure 7. Potential of mean force plot for the conformational change from system 4 to system 3 calculated using umbrella sampling and WHAM. Plots A, B, and C were the PMF corresponding to three independent simulations, whereas plot D was the PMF calculated from the aggregate data. The plots show the angle between A13 (center of mass of N3, C4, and C2), G4 (center of mass of N1, C6, and C5), and G4 (atom N3) on the x-axis and the angle between A5 (center of mass of N3, C4, and C5), G12 (center of mass of N1, C5, and C6), and G12 (atom N3) on the y-axis.
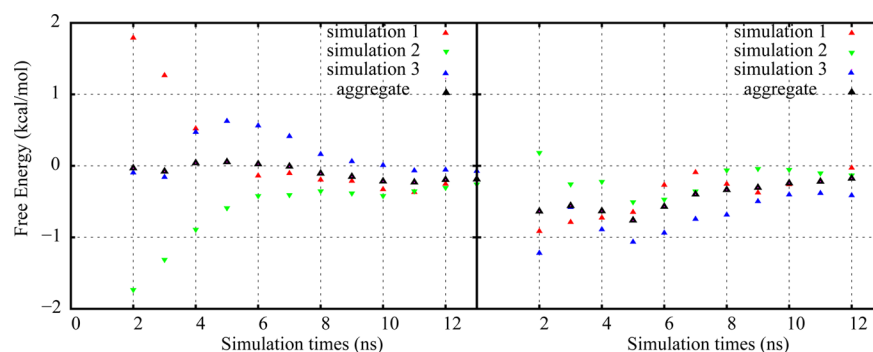
Figure 8. Convergence of free energy change. The points annotated with red, green, and blue were three independent simulations and the black was an average over all three simulations. The free energy difference was calculated at 1 ns intervals, using data prior to that simulation time. Left and right plots correspond to the conformational change of system 2 to system 1 and system 3 to system 4, respectively.

involving the angles between A13 (center of mass of N3, C4 and C2), G4 (center of mass of N1, C6, and C5), and G4 (atom N3) and between A5 (center of mass of N3, C4, and C5), G12 (center of mass of N1, C5, and C6), and G12 (atom N3). On the plane defined by these reaction coordinates (Figure 5), the imino conformation is the right top corner centered at (100, 100) and the sheared conformation is the left bottom corner centered at (20, 20). The NEB pathway involves a stepwise movement of the pairs as shown in Figure 5 for sequence (GCGGACGC)$_2$.[22] A similar path was also found for sequence (GCGGAUGC)$_2$[21] as shown in Supporting Information Figure S3. This path shows only one GA pair changing conformation at a time. This was far from the diagonal of the reaction coordinate, where both pairs would be moving through transition states simultaneously.

Umbrella Sampling. Umbrella sampling and WHAM were applied to calculate the potential of mean force (PMF) profile along the reaction coordinate. To calculate the PMF of the transition from system 2 to system 1, a total of 153 windows were used. That transition was studied with three separate calculations with 1 ns of equilibration, followed by 12 ns of sampling in each window, for a total of approximately 6 μs aggregate simulation time. The location and number of windows was selected in conjunction with the restraint force constant to ensure sampling overlap of adjacent windows. Supporting Information Figure S4 demonstrates the sampling overlap for each of the three calculations. Additionally, 130 windows were used to calculate the PMF of the transition from system 3 to system 4 with 12 ns of sampling after 1 ns of equilibration for each for a total approximately 5 μs simulation

time. Supporting Information Figure S4 demonstrates the sampling overlap for each of the three calculations.

The PMF profiles (Figures 6 and 7 for GC closed loops and GU closed loops, respectively) had two major minima located on the top right corner, the imino conformation, and on the left bottom corner, the sheared conformation. An additional fourth PMF profile was also generated for each sequence by aggregating the data from all three umbrella sampling simulations. Overall, the locations of the global minima are similar for the individual simulations.

As detailed in the Methods, the unrestrained MD suggested regions of the PMF to consider as representing the shared or imino structures (Figure 5). The free energy differences were then calculated between these areas on the plots. To test convergence, the free energy difference between the two conformations was generated using eq 1 as a function of time for all the individual runs and for the average PMF at 1 ns intervals (Figure 8). For all six umbrella sampling calculations, 12 ns of sampling (after 1 ns of equilibration) was sufficient so that the PMFs were no longer changing as a function of sampling time.

The average free energy difference, using all the sampling data, for the conformational change from system 2 (sheared) to system 1 (imino) was -0.16 ± 0.10 kcal/mol. For the conformational change for system 3 (imino) to system 4 (sheared), the free energy difference was -0.24 ± 0.11 kcal/mol. The direction of the free energy difference was in agreement with the experimentally determined conformational preference, although the magnitudes are probably underestimating the extent of conformational preference because no characteristics of the control structures are observed in the NMR spectra.[21,22]

■ DISCUSSION

In this work, the conformational preference of tandem GA pairs was studied with explicit solvent molecular dynamics. The native imino structure, system 1, was not preserved during MD simulations with ff10. A modified Amber force field, modified ff10,[38] was used and force field was shown to preserve the native structure in MD and it was also capable of predicting the conformational preference.

Recent studies reported limitations in the available Amber force fields for RNA when modeling loop structures. ff99 with the parmbsc0 modification was shown to not correctly stabilize tetraloop structures,[62] an AA noncanonical pair[46] or the preferred conformations of GACC.[63] The recent ff10 force field, which includes revised torsion parameters for α, γ, and χ, however, was shown to correctly model tetraloop structures.[62] Also, ff99 with the parmbsc0 modification was shown to correctly model the relative stability of helices, within the sampling error.[64]

This work shows that molecular mechanics can also correctly model the conformational preference of tandem GA noncanonical pairs, once a simple modification of the force field is applied.[38] This modified force field was implemented for both imino and sheared conformations despite a previous recommendation to use the parameters for GA imino interactions only. Apparently, the relaxation of the improper dihedral that keeps the amino group atoms in the plane of the base does not adversely affect the sheared interaction when closed by GC pairs. This is also reinforced by unrestrained molecular dynamics simulations on system 4 (Figure 4B).

The extent of the conformational preference for the native structures is underestimated for both sequences. As shown in Table 2, the free energy differences between the native

Table 2. Free Energy Difference Calculated from the PMF (Figures 6 and 7) Using Equation 1

system 2 ⇌ 1	$\Delta\Delta G$ (kcal/mol)	system 3 ⇌ 4	$\Delta\Delta G$ (kcal/mol)
simulation 1	−0.17	simulation 1	−0.16
simulation 2	−0.25	simulation 2	−0.19
simulation 3	0.06	simulation 3	−0.36
aggregate[a]	−0.16	aggregate	−0.23
average[b]	−0.16 ± 0.10	average	−0.24 ± 0.11

[a]$\Delta\Delta G$ calculated based on the aggregate data of all three simulations. [b]Mean $\Delta\Delta G$ calculated from the individual calculations. Errors were the standard deviation of three simulations.

structure and the non-native, decoy structure are less than RT, although the decoys were not observed in solution. This suggests that there are other interactions that are not yet correctly modeled by the modified ff10 force field. This shallow conformational preference also highlights the importance of the free energy calculations performed here. In each sequence, unrestrained MD starting from the solution structure does not sample the control (non-native) conformation because the barrier between the states is larger than 6 kcal/mol (Figures 6 and 7). It would not be apparent that the native conformations are understabilized by the force field if the umbrella sampling, or other enhanced sampling, was not performed.

This work shows the promise of currently available RNA force fields, but it also shows the importance of continued development and testing of force field parameters. The free energy calculations performed here allow direct comparison between simulation and experiment and provide a benchmark for force fields. This work also shows the importance of broadening the set of RNA molecules used to test force field performance.

■ ASSOCIATED CONTENT

ⓢ Supporting Information
Plots of mass-weighted RMSD for trajectories of systems 2 and 4, figures and plots demonstrating the interactions that favor imino pairing or sheared pairing, NEB pathway for conformational change between systems 3 and 4, overlap of snapshots from umbrella sampling, and a table showing the NEB simulated annealing protocol. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: David_Mathews@urmc.rochester.edu.
Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Altman, S. A view of RNase P. *Mol. BioSyst.* **2007**, 3, 604−607.

(2) Doudna, J. A.; Cech, T. R. The chemical repertoire of natural ribozymes. *Nature* **2002**, *418*, 222−228.

(3) Eddy, S. R. Noncoding RNA genes and the modern RNA world. *Nat. Rev. Genet.* **2001**, *2*, 919−929.

(4) Nissen, P.; Hansen, J.; Ban, N.; Moore, P. B.; Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **2000**, *289*, 920−930.

(5) Scott, W. G. Ribozymes. *Curr. Opin. Struct. Biol.* **2007**, *17*, 280−286.

(6) Wu, L.; Belasco, J. G. Let me count the ways: Mechanisms of gene regulation by miRNAs and siRNAs. *Mol. Cell* **2008**, *29*, 1−7.

(7) Bachellerie, J.-P.; Cavaillé, J.; Hüttenhofer, A. The expanding snoRNA world. *Biochimie* **2002**, *84*, 775−790.

(8) Serganov, A.; Patel, D. J. Ribozymes, riboswitches and beyond: Regulation of gene expression without proteins. *Nat. Rev. Genet.* **2007**, *8*, 776−790.

(9) Tucker, B. J.; Breaker, R. R. Riboswitches as versatile gene control elements. *Curr. Opin. Struct. Biol.* **2005**, *15*, 342−348.

(10) Wahl, M. C.; Will, C. L.; Lührmann, R. The spliceosome: Design principles of a dynamic RNP machine. *Cell* **2009**, *136*, 701−718.

(11) Walter, P.; Blobel, G. Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. *Nature* **1982**, *299*, 691−698.

(12) Gottesman, G. S. a. S. Versatile roles of small RNA regulators in bacteria. In *The RNA World*, 3rd ed.; Gesteland, R. F., Cech, T. R., Atkins, J. F., Eds., Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 2006; pp 567−594.

(13) Lilley, D. M. J. Analysis of global conformational transitions in ribozymes. *Methods Mol. Biol.* **2004**, *252*, 77−108.

(14) Tama, F.; Valle, M.; Frank, J.; Brooks, C. L. Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy. *Proc. Natl. Acad. Sci. U.S.A* **2003**, *100*, 9319−9323.

(15) Rodnina, M. V.; Savelsbergh, A.; Wintermeyer, W. Dynamics of translation on the ribosome: Molecular mechanics of translocation. *FEMS Microbiol. Rev.* **1999**, *23*, 317−333.

(16) Peske, F.; Savelsbergh, A.; Katunin, V. I.; Rodnina, M. V.; Wintermeyer, W. Conformational changes of the small ribosomal subunit during elongation factor G-dependent tRNA−mRNA translocation. *J. Mol. Biol.* **2004**, *343*, 1183−1194.

(17) Matassova, A. B.; Rodnina, M. V.; Wintermeyer, W. Elongation factor G-induced structural change in helix 34 of 16S rRNA related to translocation on the ribosome. *RNA* **2001**, *7*, 1879−1885.

(18) Fourmy, D.; Recht, M. I.; Blanchard, S. C.; Puglisi, J. D. Structure of the A site of *Escherichia coli* 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science* **1996**, *274*, 1367−1371.

(19) Whitford, P. C.; Blanchard, S. C.; Cate, J. H. D.; Sanbonmatsu, K. Y. Connecting the kinetics and energy landscape of tRNA translocation on the ribosome. *PLoS Comput. Biol.* **2013**, *9*, DOI: 10.1371/journal.pcbi.1003003.

(20) Sanbonmatsu, K. Y. Computational studies of molecular machines: The ribosome. *Curr. Opin. Struct. Biol.* **2012**, *22*, 168−174.

(21) Tolbert, B. S.; Kennedy, S. D.; Schroeder, S. J.; Krugh, T. R.; Turner, D. H. NMR structures of (rGCUGAGGCU)2 and (rGCGGAUGCU)2: Probing the structural features that shape the thermodynamic stability of GA pairs. *Biochemistry* **2007**, *46*, 1511−1522.

(22) Wu, M.; Turner, D. H. Solution structure of (rGCGGACGC)2 by two-dimensional NMR and the iterative relaxation matrix approach. *Biochemistry* **1996**, *35*, 9677−9689.

(23) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.

(24) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.;

Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.

(25) 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.

(26) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; Spoel, D. v. d.; Hess, B.; Lindahl, E. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845−854.

(27) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J. ; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 12*, University of California: San Francisco, 2012.

(28) 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.

(29) 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.

(30) 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.

(31) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham Iii, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophys. J.* **2007**, *92*, 3817−3829.

(32) Yildirim, I.; Stern, H. A.; Kennedy, S. D.; Tubbs, J. D.; Turner, D. H. Reparameterization of RNA χ torsion parameters for the AMBER force field and comparison to NMR spectra for cytidine and uridine. *J. Chem. Theory Comput.* **2010**, *6*, 1520−1531.

(33) Zgarbová, M.; Otyepka, M.; Sponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E., 3rd; Jurečka, 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.

(34) 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**, DOI: 10.1073/pnas.1309392110.

(35) Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; MacKerell, A. D. Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. *J. Chem. Theory Comput.* **2012**, *8*, 348−362.

(36) Foloppe, N.; MacKerell, J. A. D. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* **2000**, *21*, 86−104.

(37) 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.

(38) 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.

(39) Šponer, J.; Leszczynski, J.; Hobza, P. Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases. *Biopolymers* **2001**, *61*, 3−31.

(40) Krepl, M.; Otyepka, M.; Banáš, P.; Šponer, J. Effect of guanine to inosine substitution on stability of canonical DNA and RNA duplexes: Molecular dynamics thermodynamics integration study. *J. Phys. Chem. B* **2013**, *117*, 1872−1879.

(41) Sponer, J.; Florián, J.; Hobza, P.; Leszczynski, J. Nonplanar DNA base pairs. *J. Biomol. Struct. Dyn.* **1996**, *13*, 827−833.

(42) Hobza, P.; Šponer, J. Structure, energetics, and dynamics of the nucleic acid base pairs: Nonempirical ab initio calculations. *Chem. Rev.* **1999**, *99*, 3247−3276.

(43) 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.

(44) Santa Lucia, J.; Turner, D. H. Structure of (rGGCGAGCC)2 in solution from NMR and restrained molecular dynamics. *Biochemistry* **1993**, *32*, 12612−12623.

(45) Chen, G.; Kierzek, R.; Yildirim, I.; Krugh, T. R.; Turner, D. H.; Kennedy, S. D. Stacking effects on local structure in RNA: Changes in the structure of tandem GA pairs when flanking GC pairs are replaced by isoG−isoC Pairs. *J. Phys. Chem. B* **2007**, *111*, 6718−6727.

(46) Van Nostrand, K. P.; Kennedy, S. D.; Turner, D. H.; Mathews, D. H. Molecular mechanics investigation of an adenine−adenine noncanonical pair conformational change. *J. Chem. Theory Comput.* **2011**, *7*, 3779−3792.

(47) *Computational Biochemistry and Biophysics*; Simonson, T. I. B., O. M., Alexander D. MacKerell, J., Roux, B., Watanabe, M., Eds.; Marcel Dekker, Inc.: New York, pp 169−197.

(48) 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.

(49) 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.

(50) 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.

(51) 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.

(52) 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.

(53) Mathews, D. H.; Case, D. A. Nudged elastic band calculation of minimal energy paths for the conformational change of a GG non-canonical pair. *J. Mol. Biol.* **2006**, *357*, 1683−1693.

(54) Bergonzo, C.; Campbell, A. J.; Walker, R. C.; Simmerling, C., A partial nudged elastic band implementation for use with large or explicitly solvated systems. *Int. J. Quantum Chem.* **2009**, *109*.

(55) Tsui, V.; Case, D. A. Theory and applications of the generalized born solvation model in macromolecular simulations. *Biopolymers* **2000**, *56*, 275−291.

(56) Réblová, K.; Střelcová, Z.; Kulhánek, P.; Beššeová, I.; Mathews, D. H.; Van Nostrand, K.; Yildirim, I.; Turner, D. H.; Šponer, J. An RNA molecular switch: Intrinsic flexibility of 23S rRNA Helices 40 and 68 5′-UAA/5′-GAN internal loops studied by molecular dynamics methods. *J. Chem. Theory Comput.* **2010**, *6*, 910−929.

(57) 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.

(58) Roux, B. The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* **1995**, *91*, 275−282.

(59) Grossfield, A. WHAM: The Weighted Histogram Analysis Method, version 2.0.7; http://membrane.urmc.rochester.edu/content/wham (accessed May 10, 2013).

(60) Cousins, K. R. Computer review of ChemDraw Ultra 12.0. *J. Am. Chem. Soc.* **2011**, *133*, 8388−8388.

(61) *MATLAB and Statistics Toolbox Release 2012b*; T. M., Inc.: Natick, MA, 2012

(62) 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.

(63) 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 χ torsions. *J. Phys. Chem. B* **2011**, *115*, 9261−9270.

(64) Spasic, A.; Serafini, J.; Mathews, D. H. The Amber ff99 force field predicts relative free energy changes for RNA helix formation. *J. Chem. Theory Comput.* **2012**, *8*, 2497−2505.

(65) Leontis, N. B.; Stombaugh, J.; Westhof, E. The non-Watson−Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.* **2002**, *30*, 3497−3531.