JCTC Journal of Chemical Theory and Computation

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

Ilyas Yildirim,[†] Harry A. Stern,[‡] Jiri Sponer,[§] Nada Spackova,[§] and Douglas H. Turner*,[‡]

Department of Physics and Astronomy, Department of Chemistry and Department of Pediatrics, University of Rochester, Rochester, New York, 14627, and Institute of Biophysics, Academy of Sciences of the Czech Republic, v.v.i., Královopolská 135, 61265 Brno, Czech Republic

Received December 5, 2008

Abstract: Guanine-adenine (GA) base pairs play important roles in determining the structure, dynamics, and stability of RNA. In RNA internal loops, GA base pairs often occur in tandem arrangements and their structure is context and sequence dependent. Calculations reported here test the thermodynamic integration (TI) approach with the amber99 force field by comparing computational predictions of free energy differences with the free energy differences expected on the basis of NMR determined structures of the RNA motifs (5′-GCGGACGC-3′)$_2$, (5′-GCiGGAiCGC-3′)$_2$, (5′-GGCGAGCC-3′)$_2$, and (5′-GGiCGAiGCC-3′)$_2$. Here, iG and iC denote isoguanosine and isocytidine, which have amino and carbonyl groups transposed relative to guanosine and cytidine. The NMR structures show that the GA base pairs adopt either imino (*cis* Watson−Crick/Watson−Crick A-G) or sheared (*trans* Hoogsteen/Sugar edge A-G) conformations depending on the identity and orientation of the adjacent base pair. A new mixing function for the TI method is developed that allows alchemical transitions in which atoms can disappear in both the initial and final states. Unrestrained calculations gave ΔG° values 2−4 kcal/mol different from expectations based on NMR data. Restraining the structures with hydrogen bond restraints did not improve the predictions. Agreement with NMR data was improved by 0.7 to 1.5 kcal/mol, however, when structures were restrained with weak positional restraints to sample around the experimentally determined NMR structures. The amber99 force field was modified to partially include pyramidalization effects of the unpaired amino group of guanosine in imino GA base pairs. This provided little or no improvement in comparisons with experiment. The marginal improvement is observed when the structure has potential cross-strand out-of-plane hydrogen bonding with the G amino group. The calculations using positional restraints and a nonplanar amino group reproduce the signs of ΔG° from the experimental results and are, thus, capable of providing useful qualitative insights complementing the NMR experiments. Decomposition of the terms in the calculations reveals that the dominant terms are from electrostatic and interstrand interactions other than hydrogen bonds in the base pairs. The results suggest that a better description of the backbone is key to reproducing the experimental free energy results with computational free energy predictions.

1. Introduction

Agreement between experiments and computational predictions is a test of our understanding of intermolecular interactions. Molecular dynamics (MD) calculations are commonly used to provide insights into biological processes, including folding and dynamics of RNA.[1−4] It is now possible to calculate MD trajectories with durations ap-

* Corresponding author. Phone: (585) 275-3207. Fax: (585) 276-0205. E-mail: turner@chem.rochester.edu.

[†] Department of Physics and Astronomy, University of Rochester.

[‡] Department of Chemistry, University of Rochester.

[§] Academy of Sciences of the Czech Republic.

10.1021/ct800540c CCC: $40.75
Published on Web 07/02/2009

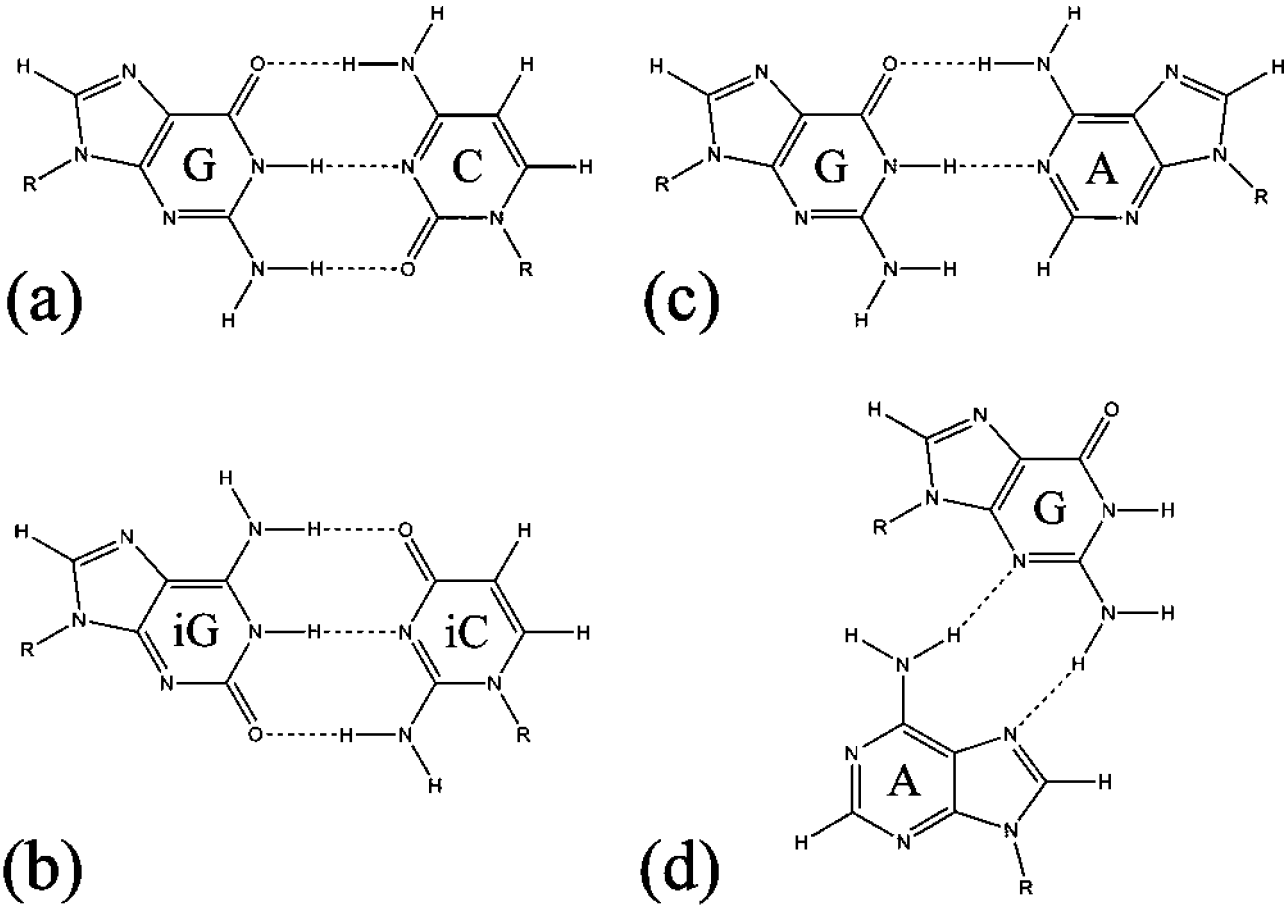

Figure 1. Structures of (a) Watson−Crick GC, (b) Watson−Crick iGiC, (c) Imino GA (*cis* Watson−Crick/Watson−Crick A-G), and (d) sheared GA (*trans* Hoogsteen/Sugar edge A-G). The difference of iG and iC from G and C, respectively, is that the amino and carbonyl groups are transposed. The amino group of G in an imino GA base pair is nonplanar,[31,32] and it has the potential to form an extra cross-strand out-of-plane hydrogen bond.

proaching the millisecond time scale.[5−9] These simulations, however, use approximate molecular mechanics (MM) potential energy functions or force fields,[10−12] which are not expected to reproduce the true potential energy surface as well as quantum mechanical (QM) methods.

Nucleic acid structures are stabilized by a range of molecular interactions such as base stacking, base pairing, solvation, and ionic effects, but the balance of these interactions is complex and poorly understood.[13] The structures and energetics of GA pairs in RNA oligonucleotides provide a particularly interesting case for testing computational methods. GA pairs occur in many RNAs.[14−17] They have roles in tertiary interactions,[18−22] metal binding,[18] and protein recognition.[23] Often, GA pairs occur in tandem.[14,24] NMR experiments[25,26] reveal that these tandem GA pairs have either imino (*cis* Watson−Crick/Watson−Crick A-G) or sheared (*trans* Hoogsteen/Sugar edge A-G) conformations (Figure 1) depending on the adjacent base pairs, and this has a dramatic effect on global 3D structure (Figure 2). The interactions responsible for this sequence dependence are not fully understood.[27,28] Thus, the sequence dependence of GA structures provides a benchmark for testing computational approaches.

Here, the sequence dependence of hydrogen-bonding patterns in GA pairs flanked by GC, CG, iGiC, or iCiG pairs is investigated with thermodynamic integration (TI)[29] calculations, which may be used to estimate free energy differences. NMR structures[25,26,30] show that the hydrogen-bonding patterns in tandem GA pairs change when the amino and carbonyl groups on flanking GC pairs are transposed to give flanking iGiC pairs (Figures 1and 2). We use explicit solvent MD combined with TI to calculate ΔG°_2 and ΔG°_3 for the alchemical steps in the thermodynamic cycles shown in Figures 3 and 4. In both cycles

$$\Delta G^\circ_1 + \Delta G^\circ_2 = \Delta G^\circ_3 + \Delta G^\circ_4 \qquad (1a)$$

$$\Delta G^\circ_3 - \Delta G^\circ_2 = \Delta G^\circ_1 - \Delta G^\circ_4 \qquad (1b)$$

The right-hand side of eq 1b, $\Delta G^\circ_1 - \Delta G^\circ_4$, is either positive (Figure 3) or negative (Figure 4) on the basis of structures determined by NMR.[25,26,30] For example, in Figure 3, ΔG°_1 and ΔG°_4 are positive and negative, respectively, because the NMR structures have imino and sheared GA pairs, respectively.[26,30] Thus, the sign calculated by TI for $\Delta G^\circ_3 - \Delta G^\circ_2$ in eq 1b can be compared to the positive sign determined by experiment for $\Delta G^\circ_1 - \Delta G^\circ_4$ in eq 1b. When the standard TI procedure with standard amber99 force field was applied, the free energy calculations did not reproduce the expected sign for $\Delta G^\circ_3 - \Delta G^\circ_2$ in the cycle of Figure 3. Thus, the protocol was modified to test two hypotheses to explain this lack of agreement: (1) the amino group of G

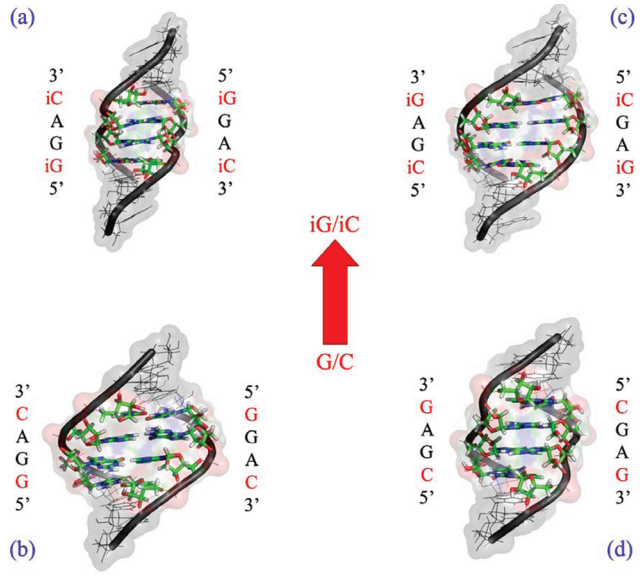

Figure 2. Comparison of structural motifs (a) (5′iGGAiC3′)$_2$ with (b) (5′GGAC3′)$_2$ and (c) (5′iCGAiG3′)$_2$ with (d) (5′CGAG3′)$_2$. Colored nucleotides in a−d are the loop regions of NMR structures of 2O81, 1MIS, 2O83, and 1YFV, respectively, from the PDB.[25,26,30] Dramatic changes in the backbone are seen when the adjacent canonical base pairs of GC in (5′GGAC3′)$_2$ and CG in (5′CGAG3′)$_2$ are replaced with iGiC (b → a) and iCiG (d → c), respectively.

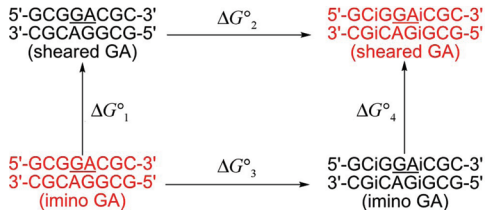

Figure 3. Thermodynamic cycle for GGAC → iGGAiC transformation. Structures in red are observed by NMR[26,30] so that $\Delta G_1 - \Delta G_4$ is positive. PDB IDs of (5′GCGGACGC3′)$_2$ and (5′GCiGGAiCGC3′)$_2$ are 1MIS and 2O81, respectively.

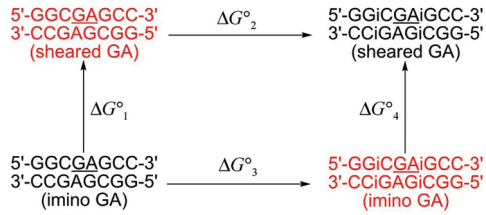

Figure 4. Thermodynamic cycle for CGAG → iCGAiG transformation. Structures in red are observed by NMR[25,30] so that $\Delta G_1 - \Delta G_4$ is negative. PDB IDs of (5′GGCGAGCC3′)$_2$ and (5′GGiCGAiGCC3′)$_2$ are 1YFV and 2O83, respectively.

in an imino GA pair can be nonplanar,[31−33] which is not considered by the force field, and (2) the conformational space sampled by MD simulations may include unrealistic conformations due to approximations in the force field.

The idea that the imino GA base pairs are involved in stabilizing out-of-plane molecular interactions via partial sp^3 hybridization was suggested by QM studies. High-level QM electronic structure calculations predict that GA base pairs are intrinsically nonplanar in isolation, in contrast to canonical base pairs.[31−33] Most striking is the nonplanarity of the *cis* Watson−Crick/Watson−Crick GA base pair (Figure 1c). In this base pair, the amino group of guanine is unpaired. In a planar base pair arrangement, it would face the H2 hydrogen of adenine, creating a repulsive contact. Instead, the guanine amino group adopts a substantially nonplanar (pyramidal) geometry due to partial sp^3 hybridization (see Supporting Information). This nonplanarity enables formation of weak out-of-plane hydrogen bonds by imino GA pairs. The nonplanarity is prevented by standard force fields. To test possible contributions of nonplanarity to determining structures of GA pairs, the intramolecular parameters for the amino group of the force field for guanine were modified and the effects of such modification were compared with results obtained with the standard force field.[10,34] With this modification applied exclusively to those guanines involved in GA base pairs, the calculated sign of $\Delta G°_3 - \Delta G°_2$ is consistent with the NMR experiments to within experimental error. Consideration of the nonplanar amino group suggests that the switch between imino and sheared GA base pairs is associated with the formation of an out-of-plane hydrogen bond between the G amino group and the flanking base pair.

The lack of agreement between computations and experiment could also reflect unphysical sampling of conformational space due to approximations in the force field. Thus, weak positional restraints were applied to limit the sampling to conformations resembling the NMR structures. The combination of positional restraints along with a nonplanar amino group of G in GA pairs provided the closest agreement between computational and experimental results.

2. Methods

2.1. Theory. The Helmholtz free energy difference between two systems, A and B, represented by Hamiltonians H_A and H_B, respectively, is given by

$$\Delta G° = G_B - G_A = -kT \ln\langle e^{-\Delta H/k_B T}\rangle_A \qquad (2)$$

Here $\Delta H = H_B - H_A$, T is temperature in kelvins, k_B is Boltzmann's constant, and $<...>_A$ denotes the canonical ensemble average for system A.[29] One way to calculate this free energy difference, ΔG, is to use the TI approach.[29] This approach can be applied to unphysical alchemical transformations, such as changing a GC pair to an iGiC pair.

In the TI approach, a hybrid Hamiltonian $H(\lambda)$, which is a mix of both states A and B, is defined as

$$H(\lambda) = f_1(\lambda)H_A + f_2(\lambda)H_B \qquad (3)$$

Here, $f_1(\lambda)$ and $f_2(\lambda)$ are mixing functions that depend on a mixing parameter, λ. The mixing parameter varies from 0 to 1 such that the hybrid Hamiltonian $H(\lambda) = H_A$ when $\lambda = 0$, and $H(\lambda) = H_B$ when $\lambda = 1$. For an effective calculation, it is desirable to use mixing functions that provide a smooth transformation from state A to state B.

With this approach, eq 2 can be rewritten as follows:

$$\Delta G^\circ = \int_0^1 \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda d\lambda \quad (4)$$

The transformation can be divided into discrete λ windows so that the integration in eq 4 may be performed numerically.

One choice of mixing functions, which is implemented in AMBER version 9,[35] is

$$\begin{aligned} f_1(\lambda, k) &= (1 - \lambda)^k \\ f_2(\lambda, k) &= 1 - (1 - \lambda)^k \end{aligned} \quad (5)$$

where k is an integer satisfying $k > 0$.[35] With these mixing functions, however, "dummy" atoms can only be used in the final state, B.[35] With this mixing rule, if dummy atoms are used in the initial state, A, then there is a convergence problem around $\lambda = 0$,[35] which does not allow a smooth transformation from A to B, thus precluding numerical integration to calculate ΔG in eq 4.[36−38]

A new mixing rule was derived to allow dummy atoms in both the initial and final states. Let the mixing functions, $f_1(\lambda,k)$ and $f_2(\lambda,k)$, be defined as follows:

$$\begin{aligned} f_1(\lambda, k) &= f(\lambda, k) \\ f_2(\lambda, k) &= 1 - f(\lambda, k) \end{aligned} \quad (6)$$

satisfying the following conditions:

$$\begin{aligned} &f(\lambda = 0, k) = 1 \\ &f(\lambda = 1, k) = 0 \\ &f^{(n)}(\lambda = 0, k) = f^{(n)}(\lambda = 1, k) = 0 \end{aligned} \quad (7)$$

Here k is a positive integer ($k > 0$), $f^{(n)}$ is the nth derivative of the mixing function $f(\lambda,k)$, and $n < k$. Assuming that $f(\lambda,k)$ is a $(2k - 1)$th-order polynomial in λ, the following mixing function can be derived with the help of Mathematica:[39]

$$f(\lambda, k) = (1 - \lambda)^k \sum_{i=0}^{k-1} \frac{(k - 1 + i)!}{i!(k - 1)!} \lambda^i \quad (8)$$

This new mixing rule makes it possible to smoothly transform an initial state, A, to a final state, B, when dummy atoms are present in initial, final, or both states (Figure 5). We implemented this new mixing function in AMBER version 9.[35] Other methods have been proposed to deal with convergence problems arising from alchemical transformations involving the appearance or disappearance of atoms with hard-core Lennard-Jones potentials.[40,41]

At each step of the simulation, the derivative of the hybrid Hamiltonian, $H(\lambda)$, with respect to the mixing parameter, λ, is calculated

$$\frac{\partial H(\lambda)}{\partial \lambda} = \frac{\partial f(\lambda)}{\partial \lambda}(H_A - H_B) \quad (9)$$

The derivatives of the old and new mixing functions are shown in Figure 5. The new mixing function improves convergence around $\lambda = 0$.

2.2. RESP Charge Calculation for iG, iC, G, A, C, and U. The systems in Figures 3 and 4 contain the unnatural nucleotides iC and iG. In order to be consistent and complete, the restrained electrostatic potential (RESP) charges were derived for iG, iC, G, A, C, and U following the RESP protocol.[10,42−44] These RESP charges were used in all the simulations. The charges for G, A, C, and U are similar to those in amber99.[10]

Dimethylphosphate (DMP) and the nucleosides of iG, iC, G, A, C, and U with C3′-endo sugar pucker were created with the xleap program in the AMBER package.[35] The molecules were optimized and the electrostatic potential at a set of gridpoints were calculated at the HF level using the 6-31G* basis set. These calculations were performed with Gaussian03.[45] RESP charges for these nucleosides were then calculated with the RESP program.[44] The only difference between these RESP calculations and those of Cornell et al. (1995)[10] is that the sugar atoms were not made equivalent. That is, each of the nucleosides has different charges for the same sugar atoms.

Except for the RESP charges, the missing parameters for iG and iC were taken from the amber99 parameter set by analogy. The only difference of iG and iC from G and C is the transposition of the carbonyl and amino groups. In AMBER, this transposition does not change the type of carbonyl and amino group. Therefore, the missing bonded parameters of both the carbonyl and amino groups of iG and iC were taken to be identical to the parameters of G and C. As a result, this choice should not impact the final results.

A library of 18 residues was created (RX, RX3, and RX5 where X = A, U, G, C, iG, and iC, and 3 and 5 denote the 3′ and 5′ end version of residue RX, respectively). Dummy atom residues for G, C, iG, and iC (RGD, RCD, IGD, ICD) were also created to allow alchemical transformations from G to iG and C to iC. The order of the atoms of either RGD and IGD or RCD and ICD were kept the same, which is mandatory for the TI approach in AMBER version 9 (see Supporting Information).

2.3. Explicit Solvent Simulations. For each A → B alchemical transformation (Figures 3 and 4), the NMR structures[25,26,30] were used as the starting structures. These NMR structures are highlighted in red in Figures 3 and 4, and the 3D structures are shown in Figure 2. The structures were solvated with TIP3P water molecules[46] in an octahedral box with a buffer of 10 Å, which gave more than 3100 water molecules around the duplex. In each A → B alchemical transformation, states A and B had the same number of water molecules and box sizes. The system was neutralized with 14 Na^+ ions. The parameter/topology files for the A → B transformations were created with the xleap module.[35]

2.4. Force Fields. Two parameter sets were tested for each thermodynamic cycle: (1) the amber99 force field and (2) the amber99 force field with a modified frcmod file[35] (see Supporting Information), which allows nonplanarity of the amino group of G of the tandem GA base pairs.

Note that the modified force field does not fully capture the true energetics associated with amino group pyramidalization. Thus, the energies associated with amino group nonplanarity are most likely underestimated in this paper. Note also that when mimicking the pyramidalization effects for the purpose of condensed phase nucleic acid simulations, it is not advisible to fit the force field tightly to the gas-phase quantum chemical data (see Supporting Information).[47]

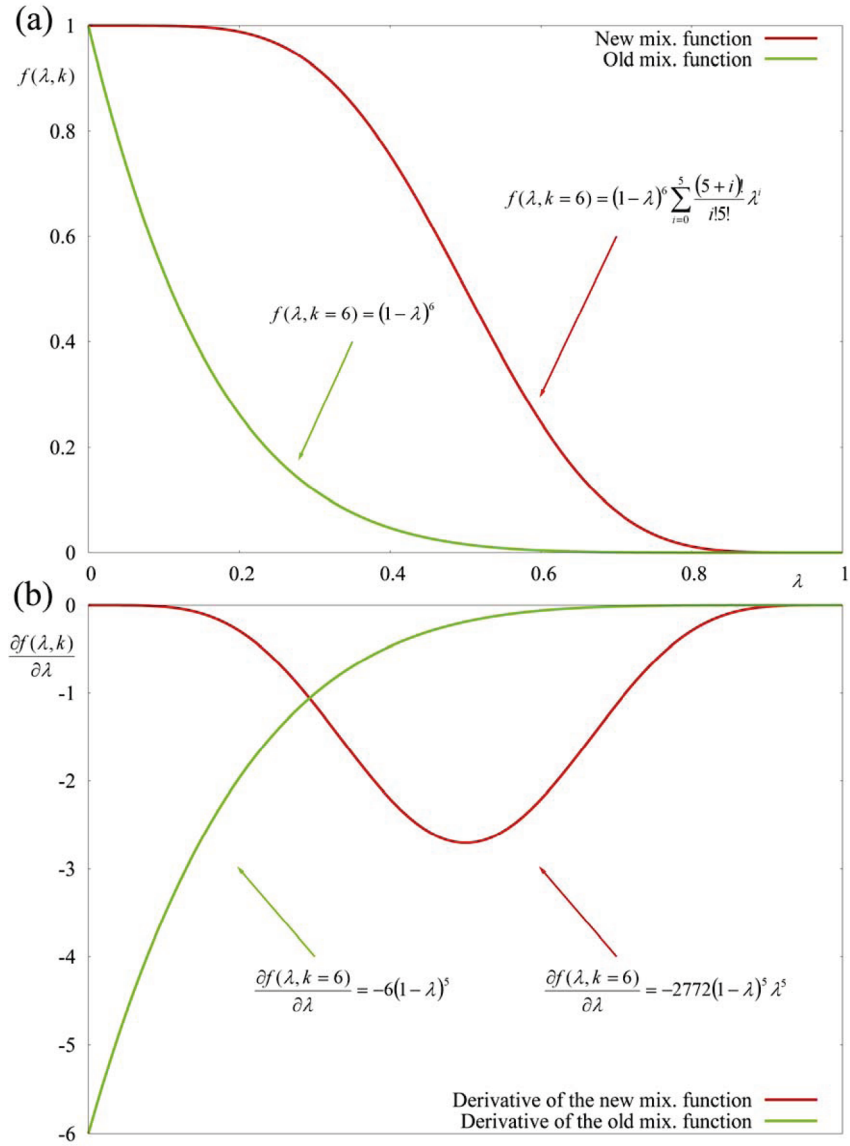

Figure 5. (a) Old (green) and new (red) mixing functions and (b) the derivatives of the old (green) and new (red) mixing functions for $k = 6$. In a simulation, eq 9 is calculated at each step, which depends on the derivative of the mixing function with respect to λ, $\partial f(\lambda,k)/\partial\lambda$. When dummy atoms are present in the initial state, there is a convergence problem around $\lambda = 0$ according to the old mixing rule, seen in (b).[35] The new mixing rule improves the convergence around $\lambda = 0$ by having a symmetric $\partial f(\lambda,k)/\partial\lambda$, as seen in (b). The plot was created with gnuplot (http://www.gnuplot.info/).

We also do not suggest using this modification for canonical base pairs where the amino group is planarized by planar hydrogen bonds.

2.5. Minimization. The structures were minimized in two steps. For each system, the same protocol was used: (1) With the RNA held fixed with a restraint force of 500 kcal/mol-Å^2, steepest descent minimization of 1000 steps was followed by a conjugate gradient minimization of 1500 steps. (2) With all restraints removed, steepest descent minimization of 1000 steps was followed by a conjugate gradient minimization of 1500 steps. During the minimization, an 8.0 Å long-range cutoff for nonbonded interactions was chosen.

2.6. Pressure Regulation. After the minimization, two steps of pressure equilibration were done on each system: (1) RNA structures were held fixed with a restraint force of 10 kcal/mol-Å^2. Constant volume dynamics with a long-range cutoff of 8.0 Å for nonbonded interactions was used. SHAKE[48] was turned on for bonds involving hydrogen atoms, except for the amino hydrogen and dummy atoms of GC and iGiC base pairs flanking the tandem GA base pairs. The temperature was raised from 0 to 300 K in 20 ps. Langevin dynamics with a collision frequency of 1 ps^{-1} was used. A total of 20 ps of MD were run with a 2 fs time step. (2) The same conditions as above were chosen, except that constant pressure dynamics with isotropic position scaling was turned on. Reference pressure was set to 1 atm with a pressure relaxation time of 2 ps. A total of 100 ps of MD were run with a 2 fs time step. The particle mesh Ewald (PME) method was on for all calculations (see Supporting Information).

Pressure relaxation decreases the box size of the system to bring the pressure to about 1 atm. The final restart file was used as the initial coordinate file for the λ simulations.

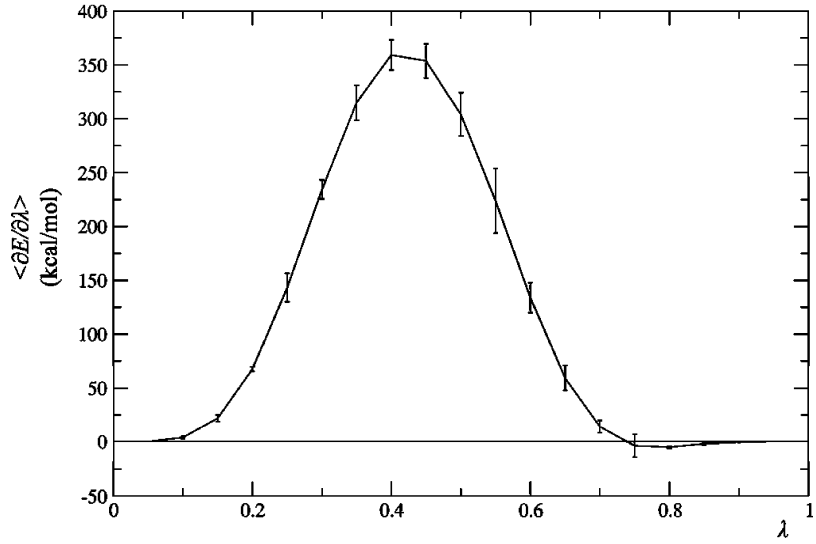

Figure 6. $<\partial E/\partial\lambda>_\lambda$ as a function of λ for the CGAG → iCGAiG (sheared GA) transformation (unrestrained). For each λ simulation, the group-averaging method was used for error estimates by dividing each simulation into three parts. The plotted error bars, which are standard error of the mean, are enlarged 30-fold. The plot was created with GRACE (http://plasma-gate.weizmann.ac.il/Grace/).

2.7. λ Simulations. A total of 19 λ values were used, $\lambda = 0.05$ to $\lambda = 0.95$, with an increment of 0.05. The new mixing rule (with $k = 6$) was used in all λ simulations. For each λ simulation, the last structure of pressure regulation was taken as the initial structure. First, the structures were minimized following the minimization protocol described above. The production run was similar to the first step of the pressure equilibration described above. Constant volume dynamics was chosen with a long-range cutoff of 8.0 Å for nonbonded interactions. SHAKE was turned on for bonds involving hydrogen atoms, except the amino hydrogen and dummy atoms of GC and iGiC flanking the tandem GA base pairs. A total of 1 ns of MD was run at 300 K with a 1 fs time step. Another 1×10^6 time steps for $\lambda = 0.2$ to 0.8 values were run with a 1 fs time step, yielding a total of 2 ns of MD for these particular λ values. Because the TI calculations are done at constant volume, the calculated free energy changes are Helmholtz free energy changes. Because the *PV* contribution is small in the condensed phase, the difference between Helmholtz and Gibbs free energy differences is expected to be small.

2.8. Explicit Solvent Simulations with Positional Restraints. The same methodology described above was used with positional restraints to calculate each A → B alchemical transformation (Figures 3 and 4) for both the original amber99 and modified amber99 force fields. Harmonic restraints[35] with a force constant of 0.1 kcal/ mol·Å^2 were applied to the backbone heavy atoms and to all atoms of the first two and last two base pairs of the structures. A total of 500 ps of MD was run with a 1 fs time step for all λ simulations. Another 5×10^5 time steps for $\lambda = 0.15$ to 0.85 of CGAG → iCGAiG ($\Delta G°_3$, imino) transformations were run with a 1 fs time step to attain better convergence, which yielded a total of 1 ns of MD for these particular λ values (see Supporting Information).

2.9. Explicit Solvent Simulations with Hydrogen Bond Restraints. The same methodology described above was used with distance restraints on hydrogen bonds to calculate each A → B alchemical transformation (Figures 3 and 4) using both the original amber99 and modified amber99 force fields. Distance restraints of 25 kcal/mol·Å^2 were applied to all base pair hydrogen bonds (see Supporting Information). A total of 1 ns of MD was run with a 1 fs time step for all λ simulations. Another 5×10^5 MD steps for $\lambda = 0.15$ to 0.85 of CGAG → iCGAiG ($\Delta G°_2$, sheared, amber99), CGAG → iCGAiG ($\Delta G°_3$, imino, modified amber99), and GGAC → iGGAiC ($\Delta G°_3$, imino, modified amber99) transformations were run with a 1 fs time step to attain better convergence, which yielded a total of 1.5 ns of MD for these particular λ values (see Supporting Information).

2.10. Analysis. The first 250 ps of each λ simulation were omitted from the calculations to allow equilibration. For each λ simulation, $<\partial E/\partial\lambda>_\lambda$ was calculated, where E is potential energy. The group-averaging method was used for error estimates by dividing each λ simulation into three parts. Error estimates are the standard error of the mean.

The trapezoidal rule was used to numerically integrate $\Delta G°$ in eq 4. A typical λ vs $<\partial E/\partial\lambda>_\lambda$ plot is shown in Figure 6. For calculation of the error in $\Delta G°$, $<\partial E/\partial\lambda>_\lambda$ values were assumed to be independent.

Rmsd calculations were done with the ptraj module of AMBER 9.[35] Water molecules and Na^+ ions were stripped out of the trajectory files, and all atoms of the RNA were included in the rmsd calculations. Each snapshot in the trajectory file was aligned with the initial starting structure (see Supporting Information for a typical time vs rmsd plot). For unrestrained simulations, the rmsd typically fluctuates between 1 and 2 Å, which implies that the simulation is sampling configurations close to the initial structure. For one or two unrestrained λ simulations, however, configurations with rmsd values greater than 3 Å are sampled. For the λ simulations with positional restraints, rmsd values are around 0.7 Å, which implies that only configurations very close to the initial structure are sampled. For the λ simulations with

Table 1. TI Method with Unrestrained and Restrained Molecular Dynamics with (or in parentheses without) Nonplanar G Amino Group in GA Base Pairs[a,b]

alchemical transformation	ΔG°_2	ΔG°_3	$\Delta G^\circ_3 - \Delta G^\circ_2$	$\Delta G^\circ_1 - \Delta G^\circ_4$ (experimental)
		No Restraints		
GGAC → iGGAiC	111.227 ± 0.100	111.155 ± 0.110	−0.1 ± 0.2 (−0.7 ± 0.1)	≥3.1
CGAG → iCGAiG	111.306 ± 0.086	110.755 ± 0.073	−0.6 ± 0.1 (−0.6 ± 0.1)	≤−2.2
		Positional Restraints		
GGAC → iGGAiC	109.96 ± 0.16	110.45 ± 0.15	0.5 ± 0.2 (0.0 ± 0.2)	≥3.1
CGAG → iCGAiG	112.23 ± 0.18	110.41 ± 0.11	−1.8 ± 0.2 (−2.1 ± 0.2)	≤−2.2
		Hydrogen Bond Restraints		
GGAC → iGGAiC	111.31 ± 0.10	111.30 ± 0.07	0.0 ± 0.1 (−1.0 ± 0.2)	≥3.1
CGAG → iCGAiG	111.15 ± 0.10	111.12 ± 0.09	0.0 ± 0.1 (−0.6 ± 0.1)	≤−2.2

[a] Results in kcal/mol. [b] See Supporting Information for details.

distance restraints on hydrogen bonds, rmsd values are around 1.5−2.0 Å.

2.11. Energy Decomposition. Free energy, ΔG°, can be decomposed into contributions resulting from bond, angle, dihedral, electrostatic, and van der Waals interactions:

$$\Delta G^\circ = \Delta G^\circ_{\text{bond}} + \Delta G^\circ_{\text{angle}} + \Delta G^\circ_{\text{dihedral}} + \Delta G^\circ_{\text{es}} + \Delta G^\circ_{\text{vdw}} \quad (10a)$$

where

$$\Delta G^\circ_{\text{bond}} = \int_0^1 \left\langle \frac{\partial E_{\text{bond}}}{\partial \lambda} \right\rangle_\lambda \mathrm{d}\lambda \quad (10b)$$

etc.

Another way of decomposing the free energy is

$$\Delta G^\circ = \Delta G^\circ_{\text{RNA-env}} + \Delta G^\circ_{\text{HB}} + \Delta G^\circ_{\text{cross}} + \Delta G^\circ_{\text{ss}} + \Delta G^\circ_{\text{other}} \quad (11)$$

where $\Delta G^\circ_{\text{RNA-env}}$, $\Delta G^\circ_{\text{HB}}$, $\Delta G^\circ_{\text{cross}}$, and $\Delta G^\circ_{\text{ss}}$ represent the RNA-environment, hydrogen bonds within base pairs, cross-strand stacking, and single-strand stacking, respectively. $\Delta G^\circ_{\text{other}}$ includes the alchemical transformations of the individual bases. Both types of free energy decompositions were done on the alchemical transformations with positional restraints, ΔG°_2 and ΔG°_3 in Figures 3 and 4, to examine qualitatively the dominant factors in the TI calculations.

For decompositions, the trajectory files of each restrained λ simulation were used. The first 250 ps of the trajectories were omitted. The remaining structures generated every 1 ps were used to calculate the individual terms in eqs 10a and 11 using the TI approach. Details of the decompositions are described in Supporting Information.

Because the decompositions were done on alchemical transformations, they do not have a direct physical meaning. Comparisons of $(\Delta G^\circ_3 - \Delta G^\circ_2)$ values, however, give an idea of the contributions of individual terms.

3. Results

Two thermodynamic cycles are analyzed, $(GCG\underline{GA}CGC)_2 \rightarrow (GCiG\underline{GA}iCGC)_2$ (Figure 3) and $(GGC\underline{GA}GCC)_2 \rightarrow (GGiC\underline{GA}iGCC)_2$ (Figure 4). The TI approach was used to calculate ΔG°_2 and ΔG°_3 in eq 1a to see if the amber99 force field or its modified version, that allows a nonplanar amino group of G in GA pairs, is consistent with the experimental results for $\Delta G^\circ_1 - \Delta G^\circ_4$ (see eq 1b). For both unrestrained and restrained simulations, convergence analysis shows that the free energies are well converged (see Supporting Information).

3.1. GGAC → iGGAiC Transformation (Unrestrained Simulations). NMR structures (Figure 2a and b) show that tandem GA base pairs flanked by GC have primarily imino GA base pairs,[26] while when flanked by iGiC, they have primarily sheared GA base pairs (Figure 3).[30] On the basis of the NMR spectra,[26,30] there could be as much as 5% sheared GA for $(GCG\underline{GA}CGC)_2$ and 10% imino GA for $(GCiG\underline{GA}iCGC)_2$. At 300 K, these limits amount to $\Delta G^\circ_1 \approx -R(300)\ln(1/19) = 1.8$ kcal/mol and $\Delta G^\circ_4 \approx -R(300)\ln(9) = -1.3$ kcal/mol. Smaller populations of the minor species would increase the magnitude of each ΔG°. Thus eq 1b becomes

$$\Delta G^\circ_3 - \Delta G^\circ_2 = \Delta G^\circ_1 - \Delta G^\circ_4 \geq 3.1 \text{ kcal/mol} \quad (12)$$

Table 1 shows the results of the TI calculations for the $G\underline{GA}C \rightarrow iG\underline{GA}iC$ transformation when unrestrained (see also Supporting Information). The amber99 force field gives a free energy difference of -0.7 ± 0.1 kcal/mol for $\Delta G^\circ_3 - \Delta G^\circ_2$, while the modified amber99 force field yields a free energy difference of -0.1 ± 0.2 kcal/mol. Neither force field predicts the experimental result of ≥3.1 kcal/mol, but modified amber99 improves the predictions by 0.6 kcal/mol.

In an imino GA base pair, the amino group of G is not hydrogen bonded with adenosine (Figure 1). As a result, this amino group favors a conformation[32] where it can form an out-of-plane hydrogen bond.[49] When GA base pairs are in an imino conformation in the $G\underline{GA}C$ system, there are two potential cross-strand out-of-plane hydrogen bonds between the amino groups of G of GA base pairs with the carbonyl groups of C of the flanking Watson−Crick GC base pairs. The inclusion of nonplanarity of the amino group of G in an imino GA base pair strengthens the effects of this cross-strand interaction. Figure 7 shows the cross-strand distance of H21−O2 after minimization with (a) amber99 and (b) modified amber99. This distance is around 3.0 Å in amber99, while it is around 1.8 Å in modified amber99. As a result, this out-of-plane hydrogen bond is strengthened with modified amber99. The nonplanar amino group of G in an imino GA base pair is a quantum mechanical effect[31,32] that improves the free energy prediction, but not by enough to

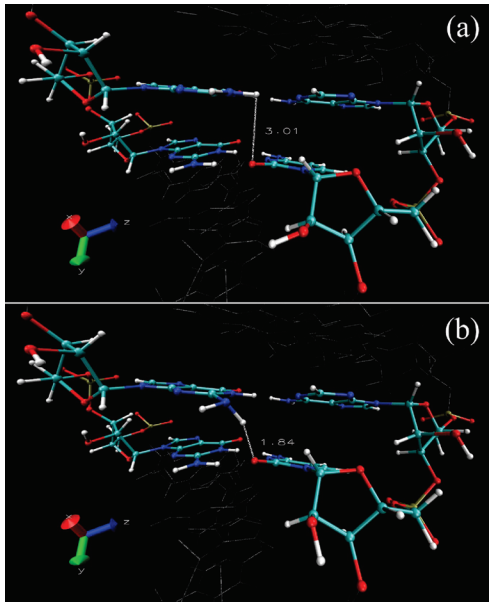

Figure 7. 5′GG3′/3′CA5′, imino GA stacked on GC in 5′GCGGACGC3′/3′CGCAGGCG5′ (1MIS) after minimization with (a) amber99 force field and (b) modified amber99 force field. The cross-strand hydrogen bond of H2−O2 distance is reduced from 3.01 to 1.84 Å when nonplanarity on amino group of G in a GA base pair is imposed on the structure. The figures were created with VMD.[56]

be consistent with the experimental results. Evidently, neither amber99 nor the modified amber99 force fields accurately take into account all the interactions responsible for the observed structures.

3.2. CGAG → iCGAiG Transformation (Unrestrained Simulations). NMR structures (Figure 2c and d) show that tandem GA base pairs flanked by CG have primarily sheared GA base pairs,[25] while when flanked by iCiG, they have primarily imino GA base pairs (Figure 4).[30] On the basis of the NMR spectra, minor species are possible at roughly 5 and 33%, respectively, for $(GGCGAGCC)_2$ and $(GGiCGAiGCC)_2$, so $\Delta G^\circ_1 \approx -R(300)\ln(19/1) = -1.8$ kcal/mol and $\Delta G^\circ_4 \approx -R(300)\ln(1/2) = 0.4$ kcal/mol. Thus eq 1b becomes

$$\Delta G^\circ_3 - \Delta G^\circ_2 = \Delta G^\circ_1 - \Delta G^\circ_4 \leq -2.2 \text{ kcal/mol} \quad (13)$$

Both the amber99 and modified amber99 force fields give a free energy difference of -0.6 ± 0.1 kcal/mol for $\Delta G^\circ_3 - \Delta G^\circ_2$ (Table 1). In the iCGAiG system, the cross-strand distance between the amino group (H21) of G of the GA base pair and the carbonyl group (O4) of iC of the iCiG base pair is longer than in iGGAiC (compare Figures 7 and 8). As a result, the prediction of modified amber99 is similar to the prediction of amber99. The magnitude of $\Delta G^\circ_3 - \Delta G^\circ_2$ is smaller than suggested by experiments. Again, neither amber99 nor the modified amber99 force fields accurately take into account all the interactions responsible for the observed structures.

3.3. GGAC → iGGAiC Transformation (Simulations with Positional Restraints). Table 1 shows the results of the TI calculations with positional restraints. The amber99 and modified amber99 force fields, respectively, give free energy

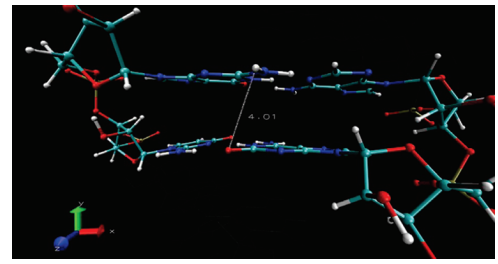

Figure 8. 5′iCG3′/3′iGA5′, imino GA stacked on iCiG in 5′GGiCGAiGCC3′/3′CCiGAGiCGG5′ (2O83) after minimization with amber99 force field. The distance between the cross-strand hydrogen bond of H21−O6 is 4.01 Å. With the modified amber99 force field, this distance is 2.60 Å. The figure was created with VMD.[56]

differences, $\Delta G^\circ_3 - \Delta G^\circ_2$, of 0.0 ± 0.2 and 0.5 ± 0.2 kcal/mol for the GGAC → iGGAiC transformation. Thus, restraining the backbone gives a value about 0.6 kcal/mol closer to the lower limit of 3.1 kcal/mol for the experimental free energy difference (eq 12 and Table 1). The λ simulations are unphysical with hybrid Hamiltonians defined in eq 3. Some of the λ simulations of unrestrained TI calculations have rmsd values greater than 3 Å for more than half the time, meaning that most configurations sampled do not resemble the NMR structures. The largest rmsd values are reduced to about 0.7 Å when restraints are applied, which reduces sampling of conformations that do not resemble the NMR structures and, most likely, provides a smoother transformation from state A to state B. Evidently, forcing sampling around the experimentally determined structures gives better results.

Free energy decompositions of the TI calculations with positional restraints according to eq 10a show that the dominant contributions to the ΔG° for the alchemical transformations are due to electrostatics (Table 2). Decompositions according to eq 11 (Table 3) show that the cross-strand interactions, excluding the base pair hydrogen bonding, provide the largest contributions toward a free energy difference consistent with the experimental results. For calculations using both the original amber99 and modified amber99 force fields, $(\Delta G^\circ_3 - \Delta G^\circ_2)_{cross}$ for GGAC → iGGAiC is positive, 44.9 ± 0.3 and 51.7 ± 0.3 kcal/mol, respectively (Table 3). Thus, inclusion of nonplanar amino groups of G in GA base pairs gives a result more in line with experiment ($\Delta G^\circ_3 - \Delta G^\circ_2 = \Delta G^\circ_3 - \Delta G^\circ_2 \geq 3.1$ kcal/mol), which is expected due to the improved cross-strand interaction of the amino group of G of the GA base pair with the adjacent base pair's carbonyl group.

It is noteworthy that the RNA−environment interaction, $(\Delta G^\circ_3 - \Delta G^\circ_2)_{RNA-env}$, which includes the RNA−solvent and RNA−counterion interactions, has almost no effect on the free energy differences for GGAC → iGGAiC both in amber99 and modified amber99 force fields.

One interesting result of the decomposition of GGAC → iGGAiC transformations is that the base pair hydrogen bond interactions of $\Delta G^\circ_{3,HB}$ and $\Delta G^\circ_{2,HB}$ (Table 3) always favor the structures with tandem GA base pairs closed by iGiC base pairs (by −1.1 and −3.0 kcal/mol in amber99 force field and by −1.3 and −2.7 kcal/mol in modified amber99

Table 2. Decomposition of Free Energy Differences (kcal/mol) to Bond, Angle, Dihedral, Electrostatic, and van der Waals Terms from TI Method using Molecular Dynamics with Positional Restraints[a]

alchemical transformation	ΔG°	ΔG°_{bond}	ΔG°_{angle}	$\Delta G^\circ_{dihedral}$	ΔG°_{es}	ΔG°_{vdw}
			amber99			
GGAC → iGGAiC (ΔG°_3, imino GA)	110.2 ± 0.2	−5.0 ± 0.1	0.4 ± 0.0	0.2 ± 0.0	113.4 ± 0.2	1.0 ± 0.0
GGAC → iGGAiC (ΔG°_2, sheared GA)	110.0 ± 0.3	−4.7 ± 0.1	0.3 ± 0.0	0.2 ± 0.0	112.7 ± 0.3	1.4 ± 0.0
($\Delta G^\circ_3 - \Delta G^\circ_2$)	0.2 ± 0.4	−0.3 ± 0.1	0.1 ± 0.0	0.0 ± 0.0	0.7 ± 0.4	−0.4 ± 0.0
CGAG → iCGAiG (ΔG°_3, imino GA)	110.6 ± 0.1	−5.2 ± 0.1	0.6 ± 0.0	0.2 ± 0.0	114.3 ± 0.1	0.7 ± 0.0
CGAG → iCGAiG (ΔG°_2, sheared GA)	112.5 ± 0.3	−5.4 ± 0.2	0.4 ± 0.0	0.2 ± 0.0	115.9 ± 0.3	1.4 ± 0.0
($\Delta G^\circ_3 - \Delta G^\circ_2$)	−1.9 ± 0.3	0.2 ± 0.2	0.2 ± 0.0	0.0 ± 0.0	−1.6 ± 0.3	−0.7 ± 0.0
			amber99 with nonplanar G amino group in GA base pairs			
GGAC → iGGAiC (ΔG°_3, imino GA)	110.6 ± 0.3	−4.7 ± 0.2	0.3 ± 0.0	0.2 ± 0.0	113.5 ± 0.2	1.2 ± 0.0
GGAC → iGGAiC (ΔG°_2, sheared GA)	110.0 ± 0.3	−4.8 ± 0.1	0.3 ± 0.0	0.2 ± 0.0	112.9 ± 0.2	1.3 ± 0.0
($\Delta G^\circ_3 - \Delta G^\circ_2$)	0.6 ± 0.4	0.1 ± 0.2	0.0 ± 0.0	0.0 ± 0.0	0.6 ± 0.3	−0.1 ± 0.0
CGAG → iCGAiG (ΔG°_3, imino GA)	110.8 ± 0.1	−5.0 ± 0.0	0.6 ± 0.0	0.3 ± 0.0	114.1 ± 0.1	0.8 ± 0.0
CGAG → iCGAiG (ΔG°_2, sheared GA)	111.9 ± 0.2	−5.7 ± 0.2	0.5 ± 0.0	0.2 ± 0.0	115.6 ± 0.3	1.4 ± 0.0
($\Delta G^\circ_3 - \Delta G^\circ_2$)	−1.1 ± 0.2	0.7 ± 0.2	0.1 ± 0.0	0.1 ± 0.0	−1.5 ± 0.3	−0.6 ± 0.0

[a] The values for ΔG° differ somewhat from those in Table 1 because the data set for the table includes structures generated every 50 fs, while the data set for this table contained only structures generated every 1 ps.

Table 3. Decomposition of Free Energy Differences (kcal/mol) to RNA Environment, Hydrogen Bond, Cross-Strand Stacking, Single-Strand Stacking, and Other Interactions[a,b]

alchemical transformation	ΔG°	$\Delta G^\circ_{RNA-env}$	ΔG°_{HB}	ΔG°_{cross}	ΔG°_{ss}	ΔG°_{other}[c]
			amber99			
GGAC → iGGAiC (ΔG°_3, imino GA)	109.5 ± 0.5	1.7 ± 0.3	−1.1 ± 0.1	−10.5 ± 0.2	−18.2 ± 0.2	137.6 ± 0.3
GGAC → iGGAiC (ΔG°_2, sheared GA)	109.6 ± 0.6	1.7 ± 0.4	−3.0 ± 0.1	−55.4 ± 0.2	27.1 ± 0.3	139.2 ± 0.3
($\Delta G^\circ_3 - \Delta G^\circ_2$)	−0.1 ± 0.8	0.0 ± 0.5	1.9 ± 0.1	44.9 ± 0.3	−45.3 ± 0.4	−1.6 ± 0.4
CGAG → iCGAiG (ΔG°_3, imino GA)	111.4 ± 0.3	0.3 ± 0.1	−1.8 ± 0.1	−13.8 ± 0.1	−9.6 ± 0.1	136.3 ± 0.2
CGAG → iCGAiG (ΔG°_2, sheared GA)	113.0 ± 0.6	−8.3 ± 0.4	0.2 ± 0.1	16.5 ± 0.2	−30.6 ± 0.2	135.2 ± 0.3
($\Delta G^\circ_3 - \Delta G^\circ_2$)	−1.6 ± 0.7	8.6 ± 0.4	−2.0 ± 0.1	−30.3 ± 0.2	21.0 ± 0.2	1.1 ± 0.4
			amber99 with nonplanar G amino group in GA base pairs			
GGAC → iGGAiC (ΔG°_3, imino GA)	110.0 ± 0.6	0.9 ± 0.3	−1.3 ± 0.1	−6.6 ± 0.2	−20.8 ± 0.1	137.9 ± 0.4
GGAC → iGGAiC (ΔG°_2, sheared GA)	109.7 ± 0.6	1.3 ± 0.3	−2.7 ± 0.1	−58.3 ± 0.2	30.7 ± 0.3	138.7 ± 0.4
($\Delta G^\circ_3 - \Delta G^\circ_2$)	0.3 ± 0.8	−0.4 ± 0.4	1.4 ± 0.1	51.7 ± 0.3	−51.5 ± 0.3	−0.8 ± 0.6
CGAG → iCGAiG (ΔG°_3, imino GA)	111.7 ± 0.3	−1.2 ± 0.2	−2.1 ± 0.1	−23.7 ± 0.1	1.7 ± 0.1	137.0 ± 0.2
CGAG → iCGAiG (ΔG°_2, sheared GA)	112.6 ± 0.5	−7.6 ± 0.3	−0.2 ± 0.1	13.4 ± 0.2	−30.1 ± 0.2	137.1 ± 0.3
($\Delta G^\circ_3 - \Delta G^\circ_2$)	−0.9 ± 0.6	6.4 ± 0.4	−1.9 ± 0.1	−37.1 ± 0.2	31.8 ± 0.2	−0.1 ± 0.4

[a] Other interactions include alchemical transformations of the individual bases from the TI method with MD using positional restraints. [b] The sum of the ΔG°'s for each transformation differ from those in Table 2 even though the same structures were used because the interactions are decomposed in a totally different way as described in Supporting Information. [c] See Supporting Information.

force field). The results of an Individual Nearest-Neighbor Hydrogen-Bonding (INN-HB) model analysis of thermodynamic data for duplexes with GC and iGiC base pairs predict that the hydrogen bonding in two iGiC pairs should favor duplex formation by $4(-0.24) = -0.96$ kcal/mol at 300 K relative to that from two GC pairs.[50] Here, −0.24 kcal/mol is the favorable free energy increment at 300 K "per terminal iG-iC", which is half the increment per internal iG-iC.

3.4. CGAG → iCGAiG Transformation (Simulations with Positional Restraints). Restrained simulations for the CGAG → iCGAiG transformation with positional restraints give free energy differences of -2.1 ± 0.2 and -1.8 ± 0.2 kcal/mol for $\Delta G^\circ_3 - \Delta G^\circ_2$ with the amber99 and modified amber99 force fields, respectively (Table 1). Both results are consistent with NMR experiments because $\Delta G^\circ_3 - \Delta G^\circ_2 \leq -2.2$ kcal/mol (eq 13 and Table 1). Rmsd values in the λ simulations are around 0.7 Å, similar to those obtained in the restrained GGAC → iGGAiC transformation with positional restraints. Evidently, imposing restraints restricts sampling to conformations resembling the NMR structures in the TI calculations. Compared to the unrestrained TI calculation results for the CGAG → iCGAiG transformation, the free energy value is closer to that expected from NMR experiments by around 1 kcal/mol for both the original amber99 and modified amber99 force fields.

Similar to the GGAC → iGGAiC transformation, the dominant terms contributing to the free energy differences for the CGAG → iCGAiG transformation are due to electrostatics (Table 2) and cross-strand interactions (Table 3). Values for $(\Delta G^\circ_3 - \Delta G^\circ_2)_{cross}$ for amber99 and modified amber99 are -30.3 ± 0.2 and -37.1 ± 0.2 kcal/mol, respectively (Table 3), which provide the largest contribution toward a free energy difference consistent with the experimental results. Again, inclusion of nonplanar amino groups of G in GA base pairs results in cross-strand interactions more in line with experiment ($\Delta G^\circ_3 - \Delta G^\circ_2 = \Delta G^\circ_1 - \Delta G^\circ_4 \leq -2.2$ kcal/mol).

The contributions of the RNA−environment interaction, $(\Delta G^\circ_3 - \Delta G^\circ_2)_{RNA-env}$, are 8.6 ± 0.5 and 6.4 ± 0.4 kcal/mol in amber99 and modified amber99 force fields, respectively (Table 3). This result is different from that obtained for the GGAC → iGGAiC transformation. The result suggests that the free energy contribution of the RNA−environment interaction disfavors the experimental results. Most of this contribution is due to the CGAG → iCGAiG sheared GA transformation. The values of $\Delta G^\circ_{2,RNA-env}$ from am-

ber99 and modified amber99 calculations for the CGAG → iCGAiG alchemical transformations are −8.3 ± 0.4 and −7.6 ± 0.3 kcal/mol, respectively. Even though this quantity does not have any physical meaning, it implies that replacing the adjacent GC base pairs of the tandem sheared GA base pairs in (5′-GGCGAGCC-3′)$_2$ with iGiC base pairs results in a more favorable RNA−environment interaction. This is probably due to the different electronic structures of GC and iGiC base pairs, which will interact with water and counterions differently.

Similar to the GGAC → iGGAiC transformation, the hydrogen-bonding interactions of $\Delta G°_{3,HB}$ favor the structures with tandem GA base pairs closed by iCiG base pairs by −1.8 and −2.1 kcal/mol in amber99 and modified amber99 force fields, respectively, (Table 3) which is consistent with the INN-HB analysis of Chen et al. (2001).[50] Values of $\Delta G°_{2,HB}$ from amber99 and modified amber99 force fields, however, are almost zero. The eight calculations give an average predicted enhancement of 1.5 kcal/mol in stability from substituting two GC pairs with iGiC pairs. This value is similar to the enhancement of 0.96 kcal/mol at 300 K expected on the basis of the INN-HB model.[50]

3.5. GGAC → iGGAiC and CGAG → iCGAiG Transformation (Simulations with Hydrogen Bond Restraints). Table 1 shows the results of the restrained TI calculations with hydrogen bond restraints. The amber99 and modified amber99 force fields, respectively, give free energy differences, $\Delta G°_3 - \Delta G°_2$, of −1.0 ± 0.2 and 0.0 ± 0.1 kcal/mol for the GGAC → iGGAiC transformation and −0.6 ± 0.1 and 0.0 ± 0.1 kcal/mol for the CGAG → iCGAiG transformation. Compared to the unrestrained simulations, there is no improvement in the free energy predictions. Yet, inclusion of nonplanarity in the amino groups of G in GA base pairs gives a free energy difference for the GGAC → iGGAiC transformation that is closer to the experimental value by 1 kcal/mol. TI calculations with hydrogen bond restraints imply that the sampling problem is due to the improper description of the backbone torsions.

4. Discussion

Non-Watson−Crick base pairs are common in RNA and each has more than one possible conformation.[51] Predicting the dependence of these conformations on sequence context provides a test of force fields.[28] A new mixing function allows TI calculations that can be compared with NMR results for the structural change of GA base pairs associated with replacing adjacent GC pairs with iGiC pairs. Allowing nonplanar G amino groups in GA pairs and applying weak positional restraints to sample around NMR determined structures improves agreement with experiment by 1.2 kcal/mol for both systems studied and provides the same signs for the experimental and computational free energy results (Table 1). Distance restraints on hydrogen bonds alone or in combination with nonplanar G amino groups do not result in agreement between experiment and computations. This result suggests that the backbone torsions in the amber99 force field may allow sampling of unphysical conformations. The amber99 force field provides good predictions for structures of canonical base pairs.[34] The backbone torsions of noncanonical base pairs (e.g., Figure 2), however, can be different from canonical base pairs. A better description of backbone torsions to include the characteristics of noncanonical base pairs might improve sampling.

While TI calculations give values for $\Delta G°_3 - \Delta G°_2$ with the same sign as experimental values, they do not reproduce the magnitudes expected from experimental results (Table 1). The error estimates in Table 1 reflect the statistical uncertainties of the individual TI runs but do not include systematic errors resulting from choosing starting structures, structural rearrangements, substates, and relaxations that occur on time scales longer than the simulation runs. Also, they do not include errors due to the force field approximations (see Supporting Information). Given these uncertainties and the number and magnitudes of individual terms (Tables 1−3), even agreement between the calculated and experimental signs of free energy differences is notable (Table 1).

The inclusion of a nonplanar amino group of G in GA base pairs improves the predictions of free energy differences between sheared and imino GA pair conformations if there are potential out-of-plane hydrogen-bonding sites for the free amino group of G in an imino GA pair (Table 1 and Figures 7 and 8). The nonplanarity can strengthen this hydrogen bond by reducing the distance between the partial positive charge on the amino hydrogen and the partial negative charge on other atoms. Allowing nonplanarity has little effect on the free energy predictions of systems that have sheared GA base pairs because the amino group of G in a sheared GA base pair already forms a hydrogen bond with adenosine (Figure 1). A force field cannot fully capture the true electronic structure effects associated with the amino group's delicate sp^2-sp^3 balance (see Supporting Information). The force field with its few parameters and constant atom-centered point charges merely allow the amino group to be sufficiently flexible to allow for nonplanarity. It cannot reflect the true flexibility of the amino group electronic structure upon continually changing the amino group geometry and environment, charge redistributions, lone pair formation, etc. Thus, the free energy effects of the nonplanarity are most likely underestimated. Even so, the force field with nonplanarity leads to a local structure consistent with expectations from crystal structures and sequence comparisons.[32] Note that the modified force field was applied only for guanines involved in GA base pairs (see Supporting Information). For all other amino groups, standard parameters assuming purely planar amino groups were used, as amino groups in canonical base pairs exhibit sp^2 hybridization.

Additional interactions may also be important for accurately describing the sequence dependence of structures of GA pairs. For example, the current force fields use fixed atom-centered charges to mimic the electronic environment of the molecules. Polarizable force fields,[52−55] when completely developed for nucleic acid simulations, with geometry-dependent electrostatics, including flexible lone pairs, might improve the free energy predictions for the types of alchemical transformations presented in this paper. Nevertheless, the calculations come close to reproducing the conformational trends revealed by NMR structures. Thus, such computations can provide qualitative insights to complement experiments.

Qualitative insights into the effects of different interactions are provided by the energy decompositions in Tables 2 and 3. The results in Table 2 show that the largest free energy term is due to electrostatic interactions. The second decomposition method (Table 3) shows that all the cross-strand interactions favor the NMR results. Moreover, replacement of GC with iGiC base pairs enhanced hydrogen bonding in the base pairs, consistent with the interpretation of experimental thermodynamic results.[50]

Acknowledgment. We thank Prof. Scott Kennedy for help with Figure 2 and Dr. Daniel Svozil for helpful discussions. This work was supported by NIH grant GM22939 (D.H.T.), by the Academy of Sciences of the Czech Republic, grants no. AV0Z50040507 and AV0Z50040702, grant IAA400040802 by Grant Agency of the Academy of Sciences of the Czech Republic, grant LC06030, Ministry of Education of the Czech Republic and grant 203/09/1476, Grant Agency of the Czech Republic (J.S and N.S.). Preliminary calculations utilized the NCSA Xeon IA-32 Linux Cluster supported by the National Center for Supercomputing Applications under CHE060041T.

Supporting Information Available: Modified force field parameters for guanine with nonplanar amino group; description of the force field modification; a typical plot of rmsd as a function of time; modified force field parameters for isoguanosine, isocytidine and dummy atoms; RESP charges for C, iC, G, iG, A, and U; convergence analysis of the unrestrained alchemical transformations; decomposition of the free energies of the restrained alchemical transformations; details of individual TI calculations with unrestrained simulations of GGAC → iGGAiC; examples of parameters used in the λ simulations; example hydrogen bond restraint file; detailed version of Table 1. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Krasovska, M. V.; Sefcikova, J.; Spackova, N.; Sponer, J.; Walter, N. G. Structural dynamics of precursor and product of the RNA enzyme from the hepatitis delta virus as revealed by molecular dynamics simulations. *J. Mol. Biol.* **2005**, *351*, 731.

(2) Csaszar, K.; Spackova, N.; Stefl, R.; Sponer, J.; Leontis, N. B. Molecular dynamics of the frame-shifting pseudoknot from beet western yellows virus: The role of non- Watson-Crick base-pairing, ordered hydration, cation binding and base mutations on stability and unfolding. *J. Mol. Biol.* **2001**, *313*, 1073.

(3) Auffinger, P.; Hashem, Y. Nucleic acid solvation: from outside to insight. *Curr. Opin. Struct. Biol.* **2007**, *17*, 325.

(4) McDowell, S. E.; Spackova, N.; Sponer, J.; Walter, N. G. Molecular dynamics simulations of RNA: An in silico single molecule approach. *Biopolymers* **2007**, *85*, 169.

(5) Trylska, J.; Tozzini, V.; McCammon, J. A. Exploring global motions and correlations in the ribosome. *Biophys. J.* **2005**, *89*, 1455.

(6) Zagrovic, B.; Pande, V. Solvent viscosity dependence of the folding rate of a small protein: Distributed computing study. *J. Comput. Chem.* **2003**, *24*, 1432.

(7) Russell, R.; Millettt, I. S.; Tate, M. W.; Kwok, L. W.; Nakatani, B.; Gruner, S. M.; Mochrie, S. G. J.; Pande, V.; Doniach, S.; Herschlag, D.; Pollack, L. Rapid compaction during RNA folding. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 4266.

(8) Zagrovic, B.; Sorin, E. J.; Pande, V. Beta-hairpin folding simulations in atomistic detail using an implicit solvent model. *J. Mol. Biol.* **2001**, *313*, 151.

(9) Duan, Y.; Kollman, P. A. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* **1998**, *282*, 740.

(10) 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.

(11) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586.

(12) Scott, W. R. P.; Hunenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Kruger, P.; van Gunsteren, W. F. The GROMOS biomolecular simulation program package. *J. Phys. Chem. A* **1999**, *103*, 3596.

(13) Sponer, J.; Riley, K. E.; Hobza, P. Nature and magnitude of aromatic stacking of nucleic acid bases. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2595.

(14) Gautheret, D. F.; Konings, D.; Gutell, R. R. A major family of motifs involving G·A mismatches in ribosomal RNA. *J. Mol. Biol.* **1994**, *242*, 1.

(15) Gutell, R. R.; Gray, M. W.; Schnare, M. N. A compilation of large subunit (23S- and 23S-like) ribosomal-RNA structures: 1993. *Nucleic Acids Res.* **1993**, *21*, 3055.

(16) Gutell, R. R.; Weiser, B.; Woese, C. R.; Noller, H. F. Comparative anatomy of 16S-like ribosomal-RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **1985**, *32*, 155.

(17) Cannone, J. J.; Subramanian, S.; Schnare, M. N.; Collett, J. R.; D'Souza, L. M.; Du, Y. S.; Feng, B.; Lin, N.; Madabusi, L. V.; Muller, K. M.; Pande, N.; Shang, Z. D.; Yu, N.; Gutell, R. R. The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* **2002**, *3*, 2.

(18) Pley, H. W.; Flaherty, K. M.; McKay, D. B. 3-Dimensional structure of a hammerhead ribozyme. *Nature* **1994**, *372*, 68.

(19) Pley, H. W.; Flaherty, K. M.; McKay, D. B. Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature* **1994**, *372*, 111.

(20) Biou, V.; Yaremchuk, A.; Tukalo, M.; Cusack, S. The 2.9 Angstrom crystal structure of T. thermophilus Seryl-tRNA synthetase complexed with tRNA(Ser). *Science* **1994**, *263*, 1404.

(21) Murphy, F. L.; Cech, T. R. GAAA tetraloop and conserved bulge stabilize tertiary structure of a Group-I Intron domain. *J. Mol. Biol.* **1994**, *236*, 49.

(22) Michel, F.; Westhof, E. Modeling of the 3-Dimensional architecture of Group-I Catalytic Introns based on comparative sequence-analysis. *J. Mol. Biol.* **1990**, *216*, 585.

(23) Zwieb, C. Recognition of a tetranucleotide loop of signal recognition particle RNA by protein-Srp19. *J. Biol. Chem.* **1992**, *267*, 15650.

(24) SantaLucia, J.; Kierzek, R.; Turner, D. H. Effects of GA mismatches on the structure and thermodynamics of RNA internal loops. *Biochemistry* **1990**, *29*, 8813.

(25) SantaLucia, J., Jr.; Turner, D. H. Structure of r(GGCGAGCC)$_2$ in solution from NMR and restrained molecular dynamics. *Biochemistry* **1993**, *32*, 12612.

(26) Wu, M.; Turner, D. H. Solution structure of r(GCGGACGC)$_2$ by two-dimensional NMR and the iterative relaxation matrix approach. *Biochemistry* **1996**, *35*, 9677.

(27) Villescas-Diaz, G.; Zacharias, M. Sequence context dependence of tandem Guanine-Adenine mismatch conformations in RNA: A continuum solvent analysis. *Biophys. J.* **2003**, *85*, 416.

(28) Yildirim, I.; Turner, D. H. RNA challenges for computational chemists. *Biochemistry* **2005**, *44*, 13225.

(29) Kollman, P. A. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395.

(30) 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.

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

(32) Sponer, J.; Mokdad, A.; Sponer, J. E.; Spackova, N.; Leszczynski, J.; Leontis, N. B. Unique tertiary and neighbor interactions determine conservation patterns of cis Watson-Crick A/G base-pairs. *J. Mol. Biol.* **2003**, *330*, 967.

(33) Sponer, J.; Florian, J.; Hobza, P.; Leszczynski, J. Nonplanar DNA base pairs. *J. Biomol. Struct. Dyn.* **1996**, *13*, 827.

(34) Wang, J. M.; 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.

(35) Case, D. A.; Darden, T. A.; Cheatham, T. E. I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. *AMBER 9*, University of California: San Francisco, CA; 2006.

(36) Simonson, T.; Carlsson, J.; Case, D. A. Proton binding to proteins: pK(a) calculations with explicit and implicit solvent models. *J. Am. Chem. Soc.* **2004**, *126*, 4167.

(37) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J. Chem. Phys.* **2003**, *119*, 5740.

(38) Simonson, T. Free-Energy of particle insertion - an exact analysis of the origin singularity for simple liquids. *Mol. Phys.* **1993**, *80*, 441.

(39) Wolfram Research, Inc. *Mathematica Edition*, Version 5.2; Wolfram Research, Inc.: Champaign, IL, 2005.

(40) Pitera, J. W.; Van Gunsteren, W. F. A comparison of non-bonded scaling approaches for free energy calculations. *Mol. Simul.* **2002**, *28*, 45.

(41) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J. Chem. Phys.* **1994**, *100*, 9025.

(42) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the multimolecule and multiconformational RESP Methodology to biopolymers - Charge derivation for DNA, RNA, and Proteins. *J. Comput. Chem.* **1995**, *16*, 1357.

(43) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. Application of RESP charges to calculate conformational energies, hydrogen-bond energies, and free-energies of solvation. *J. Am. Chem. Soc.* **1993**, *115*, 9620.

(44) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges - the Resp Model. *J. Phys. Chem.* **1993**, *97*, 10269.

(45) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; 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.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C. ; Pople, J. A. *Gaussian 03*, Revision C.02; Gaussian, Inc.: Wallingford CT, 2004.

(46) 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.

(47) Ryjacek, F.; Kubar, T.; Hobza, P. New parameterization of the Cornell et al. empirical force field covering amino group nonplanarity in nucleic acid bases. *J. Comput. Chem.* **2003**, *24*, 1891.

(48) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-Integration of cartesian equations of motion of a system with constraints: Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327.

(49) Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Oxford University Press: New York, 1997; pp 11−32.

(50) Chen, X.; Kierzek, R.; Turner, D. H. Stability and structure of RNA duplexes containing isoguanosine and isocytidine. *J. Am. Chem. Soc.* **2001**, *123*, 1267.

(51) Leontis, N. B.; Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **2001**, *7*, 499.

(52) Grossfield, A.; Ren, P. Y.; Ponder, J. W. Ion solvation thermodynamics from simulation with a polarizable force field. *J. Am. Chem. Soc.* **2003**, *125*, 15671.

(53) Ren, P. Y.; Ponder, J. W. Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B* **2003**, *107*, 5933.

(54) Babin, V.; Baucom, J.; Darden, T. A.; Sagui, C. Molecular dynamics simulations of DNA with polarizable force fields: Convergence of an ideal B-DNA structure to the crystallographic structure. *J. Phys. Chem. B* **2006**, *110,* 11571.

(55) Gresh, N.; Sponer, J. E.; Spackova, N.; Leszczynski, J.; Sponer, J. Theoretical study of binding of hydrated Zn(II) and Mg(II) cations to 5′-guanosine monophosphate. Toward polarizable molecular mechanics for DNA and RNA. *J. Phys. Chem. B* **2003**, *107*, 8669.

(56) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33.

CT800540C