Supporting Information

Effects of restrained sampling space and non-planar 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, University of Rochester, Rochester, New York 14627, Department of Chemistry and of Pediatrics, University of Rochester, Rochester, New York, 14627, Institute of Biophysics, Academy of Sciences of the Czech Republic, v.v.i., Královořecká 135, 612 65 Brno, Czech Republic.
yildirim@pas.rochester.edu, hstern@chem.rochester.edu, spackova@ncbr.chemi.muni.cz.
spackova@ncbr.chemi.muni.cz, turner@chem.rochester.edu
S.1. Rationale for modifying guanine parameters to allow a non-planar amino group

Quantum mechanical (QM) calculations on isolated guanine reveal partial sp\(^3\) hybridization of the amino group nitrogen atom and weakening of the C=N double bond with concomitant development of a lone electron pair above the amino group nitrogen atom.\(^1\) Due to this, the sum of the three valence angles around the amino group nitrogen drops below 360° which necessarily means that the overall architecture of the amino group becomes pyramidal with the hydrogens deviating from the plane of the aromatic ring in one direction. For isolated guanine there are two symmetrical pyramidal substates with the purely planar arrangement representing the transition state. The amino group pyramidalization is well-known for aniline but occurs, to variable extent, in all amino groups attached to aromatic rings.\(^2\) It has been verified by spectroscopic measurements.\(^3\)

When the guanine amino group participates in planar hydrogen bonds, its electronic structure changes to sp\(^2\) and becomes planar. Therefore, canonical GC base pairs are intrinsically planar. However, when the amino group is not involved in planar H-bonds, its intrinsic non-planarity can play a role. In such cases, the hydrogens can be either involved in out-of-plane interactions or the lone pair region above the nitrogen can serve as a weak H-bond acceptor. This happens in imino GA base pairs (but not in sheared GA base pairs where the amino group is involved in conventional H-bonds).

QM calculations show that to optimize local interactions at the minor groove side of an isolated imino GA pair, the bases adopt a propeller-twisted structure with rotational axis roughly coinciding with its major groove standard H-bond.\(^4,5\) There are three base – base contacts in the GA base pair. Two standard H-bonds and the contact between the guanine amino group and the C2-H2 group of adenine, which is localized on the minor groove edge of the base pair. The two standard H-bonds prefer the co-planar arrangement of bases because this minimizes the perturbation of the two standard H-bonds. On the other hand, the inter-hydrogen repulsion between the guanine amino group and the C2-H2 group of adenine can be reduced by propeller twisting of the base pair. Optimal balance of these two requirements is achieved by having the rotation axis for the propeller twist on the major groove side of the base pair. The C2-H2 group of adenine points above the guanine amino group nitrogen, where the
electron lone pair resides. Thus, a weak amino-acceptor interaction between the C2-H2 group of adenine and the lone pair above the amino group of guanine is possible.\textsuperscript{4,5} Furthermore, the out-of-plane hydrogen of the amino group of guanine is available to establish an out-of-plane H-bond, if a suitable partner is nearby.\textsuperscript{4} The energy gain associated with adopting a non-planar rather than a planar structure is \(\sim1.5\) kcal/mol for isolated base pair in gas phase, which is \(\sim10\%\) of the total stability of this base pair.\textsuperscript{6}

Crystal structures with imino GA base pairs show highly propeller-twisted geometries identical to the intrinsically preferred geometry.\textsuperscript{5,7,8} The crystal structures further show that in many cases the guanine amino group nitrogen is in close cross-strand contact with the O2 carbonyl group of either uracil (thymine) or cytosine of the preceding canonical base pair. In effect, there is an out-of-plane electrostatic H-bond between the non-planar amino group and the carbonyl group. This interaction is lost upon reversing the polarity of either the GA (GA \(\rightarrow\) AG) base pair or of the adjacent canonical pair (AU \(\rightarrow\) UA or GC \(\rightarrow\) CG substitution). Bioinformatics analysis has revealed that the out of plane H-bond affects the structure and stability of large functional RNAs. Formation of the out-of-plane contact (identified in available ribosomal X-ray structures) correlates with the lack of \(-G\ldots A\ldots G\) sequence covariation in equivalent positions in aligned ribosomal sequences of other species. In those positions when the guanine amino group is not involved in any interaction in the X-ray structure, the \(-G\ldots A\ldots G\) sequences alternate with \(-A\ldots G\ldots G\) ones, which is not surprising as imino GA and AG base pairs are isosteric.\textsuperscript{5}

As explained above, the amino group of guanine in imino GA base pairs substantially deviates from sp\textsuperscript{2} hybridization. Thus, for this single amino group, we modified the force field to provide a crude model of the partial sp\textsuperscript{3} hybridization effects. This modification of amino group parameters is not intended to be a general modification of the force field and should not be used for those guanines that are involved in other types of base pairs. The electronic structure of amino groups of “real” guanine is highly sensitive to its environment. Standard force fields do not allow modeling of flexible switches between sp\textsuperscript{2} and partially sp\textsuperscript{3} amino groups and the amino groups are typically fixed to be intrinsically purely planar, i.e., sp\textsuperscript{2}. Capturing the true properties of the amino groups would require a force field with
constantly updated parameters based on the instantaneous interactions the amino group is involved in. This task is not achievable with the presently available force field formalism.⁹

S.2. Modified force field parameters for guanine with non-planar amino group: Description, justification, and limitations of the force field modification frcmod_kg.

Guanine residue with non-planar amino group was prepared from the standard guanine residue (DG or RG) by changing the atom type of amino group hydrogens from H to HY. Atom types of other atoms remain unchanged. Charges can remain the same as in the original force field. The modified force field is defined according to Table S1.

General considerations. The main purpose of this force field modification is to allow the amino group of guanine to adopt non-planar geometry in a GA base pair, especially in the cis-WC GA (imino) arrangement, mimicking its genuine partial sp³ amino group hybridization. The amino group pyramidalization is primarily represented by reduction of the sum of the valence angles around the nitrogen below 360°¹,⁹ So, any force field modification should primarily include changes of the valence angle parameters further supplemented by torsional profile modifications. The amino group hydrogen dihedral angles, however, should not be primarily targeted by the parameterization as they do not reflect the physics of pyramidalization. Quantum chemical calculations show that the sum of the three valence bond angles of the amino group of isolated guanine is ~340° while the inversion barrier, i.e., the electronic energy difference between the planar and non-planar structure is ~1.1 kcal/mol.¹

Why the force field should not be fitted to the gas phase quantum chemical data and what are the requirements for the parameters? Naively, one should fit a force field to reproduce target gas phase quantum chemical data. However, in reality this approach is not justified for nucleic acids force fields and would lead to an unrealistic overestimation of the pyramidalization effects. In real systems, the amino group continuously adapts its electronic structure depending on the type of interactions it experiences. Full inclusion of the delicate sp² – sp³ electronic structure flexibility is not possible as it would require a sophisticated force field with parameters continuously updated based on the environment of the amino groups. However, the present simple force field form requires a fixed intrinsic
propensity of the amino group to be non-planar or planar. Therefore, tight fitting to reproduce the ab initio target data is not applied. Force fields with a fixed intrinsic preference for non-planarity may suffer from unbalances, e.g. exaggeration of the hydrogen flexibility or back and forth oscillations between two excessively rigid symmetrical non-planar minima. Both effects might, for example, destabilize canonical base pairs. The present force field modification is therefore primarily created to improve description of the intrinsically non-planar GA base pairs and is not designed to be used for other guanines in simulated structures. Thus we recommend using the force field exclusively for guanines of GA base pairs while standard parameters should be used for other amino groups.

Because exact parameterization of the amino group is not possible, there is no clear target value for the respective force field adjustment, and only qualified estimates (compromises) can be made based on the purpose of the computation. The purpose of the modification is to allow the amino group to adopt non-planar geometry in those situations when it is likely that non-planar geometry facilitates stabilizing out-of-plane interactions. From the biochemical point of view, rather than target one particular non-planar geometry of the amino group, it is much more important to give the amino group enough flexibility, so that its hydrogens can follow out-of-plane positions of H-bond acceptors. As evident from QM studies, for nucleic acid structure the primary importance of the amino group intrinsic non-planarity is that the amino group hydrogens can easily adopt non-planar geometries, i.e. have structural flexibility.

The above considerations indicate that the target degree of the force field amino group non-planarity must be smaller than the gas phase value predicted by QM calculations for isolated guanine. Especially the inversion barrier for the pyramidalization must be much smaller to avoid enforcing the gas phase non-planarity irrespective of the environment.

**Amino group parameters for the guanines involved in GA base pairs.** A new atom type, HY, was defined for hydrogens in the amino group. Target force field values of valence angles around the N2 nitrogen were reduced from 120° to 116° to support non-planarity. The basic philosophy was to mimic the partial pyramidalization and to be able to properly switch between the non-planar geometry (when
they participate in the non-planar GA pair) and almost planar geometry. As noted above, it is more important to give the amino group enough flexibility (to reach out of plane H-bonds) than to try to reproduce tightly the gas phase quantum chemical data.\textsuperscript{1,10} Thus, the force constants for the valence angles were reduced to 30 kcal/mol.rad\textsuperscript{2} (from 35 and 50 kcal/mol.rad\textsuperscript{2} for H-N2-H and CA-N2-H, respectively) and the respective constants for dihedral angles were reduced to 7 kcal/mol.rad\textsuperscript{2} (from 9.6 kcal/mol.rad\textsuperscript{2} for X-CA-N2-X). The improper torsion applied to the amino group was entirely removed, as it would block the out of plane deviations of the hydrogens.

This force field modification gives an energy difference (inversion barrier) of 0.27 kcal/mol between optimized planar geometry of isolated guanine (optimization in vacuum, start from the planar geometry, calculated in Sander module of AMBER) and optimized non-planar geometry (start from the non-planar coordinates). This is \~25\% of the inversion barrier of guanine in the gas phase predicted by QM calculations (see above). As explained above, we intentionally underestimate the gas phase value of the guanine inversion barrier predicted by quantum chemistry calculations.\textsuperscript{1} The sum of amino group valence angles drops to 345°, which is modestly less pyramidal than the gas phase optimum QM value of 340°. Thus, the amino group is flexible enough but not rigidly non-planar. The force field was selected as a reasonable compromise after a careful analysis of multiple modifications, but the amino group pyramidalization effects are probably underestimated.

When preparing these parameters, we tested \~40 combinations of modified parameters using optimizations of guanine and GC and GA base pairs. The amino group valence angle target values were considered in the range of 113-118°, the valence angle force constant in the range of 17.5-70 kcal/mol.rad\textsuperscript{2}, with the ultimately targeted (meaningful) range 25-35 kcal/mol.rad\textsuperscript{2}, the dihedral torsion force constants in the range 3.5-10.0 kcal/mol.rad\textsuperscript{2}, with the meaningful range identified as 7.0-9.5 kcal/mol.rad\textsuperscript{2}. The improper dihedral was switched off in most cases, or substantially weakened.

We also performed simulation tests for the B-DNA DAPI complex\textsuperscript{12} where the DAPI amidinium group binds to a GC base pair, mismatch self-complementary B-DNA duplex, d(CCAAGATTTGG), where the amino group of the underlined G is assumed to be non-planar because the two tandem imino
GA base pairs are involved in out-of-plane H-bonds and the standard B-DNA duplex, d(CGCGAATTCGC). These tests (not shown) confirmed that the non-planar amino group improves the DAPI binding in the crystallographically observed mode with close contact between the DAPI amidinium group and guanine amino group and also improves the geometry of the decamer’s GA base pairs, similarly to the RNA imino GA base pairs reported in the main text of this paper. On the other hand, the modified amino group, when used for GC Watson-Crick base pairs, increased the likelihood of fraying of terminal base pairs, while knocking off the improper dihedral caused infrequent amino group rotations (one event per ~5 ns long simulation). This behavior was entirely expected, as explained in the paragraphs above and confirms that the force field should be used for the GA base pairs but not for the GC Watson-Crick pairs.

S.3. Modified force field parameters for isoguanosine, isocytidine, and dummy atoms:

Residues that are transformed from C → iC, iC → C, G → iG, and iG → G have dummy atoms to allow these alchemical transformations (Figure S1). The missing parameters are taken from the amber99 parameter set by analogy. Table S2 shows the parameters used to describe the missing parameters for isoguanosine, isocytidine, and dummy atoms. The last column in Table S2 shows the parameters taken from amber99 that are used for the parameters defined in the first column. Residue library, modified force fields, and structures used in the calculations can be found at http://rna.chem.rochester.edu/ga_imino_sheared.

S.4 Consideration of parmbsc0 force field:

We did not test a recent revision of the AMBER nucleic acid force field, parmbsc0. Compared to parm99, the parmbsc0 force field has substantially modified parameterization of the torsional profiles for the α and γ backbone torsions. This change was evoked by major problems in longer (>10 ns) simulations of B-DNA with the earlier variants of the AMBER force field, which resulted in accumulation of irreversible α/γ backbone substates with the γ torsional angle (around the C4'-C5' bond) in trans. These substates cause severe structural deformations in simulated B-DNA which can be considered as entire degradation of the B-DNA structure. In other words, the canonical B-DNA double
helix is not predicted as global minimum by the parm99 force field. Parmbsc0 successfully stabilizes the DNA simulations. The situation with RNA is different and both force fields appear to have similar and satisfactory performance for RNA. In contrast to B-DNA the pathological trans-$\gamma$ substates do not accumulate in RNA simulations with the 99 version of the force field. There are short-living trans-$\gamma$ substates occurring in A-RNA simulations which do not accumulate and have equilibrium population of ~10%. As discussed elsewhere, it appears to be in agreement with occurrence of similar secondary A-RNA backbone conformation in A-RNA X-ray structures.\(^1\)! Likewise, simulations for an RNA internal loop (the Sarcin-Ricin loop), where high-resolution X-ray data is available, show stable trajectories with parm99 with no pathological backbone substates.\(^2\) Consistent with this, test simulations performed during the development of the parmbsc0 force field do not indicate any large difference between the parm99 and parmbsc0 force fields for RNA molecules.\(^3\) Both force fields can be recommended for RNA. Specifically for an internal loop studied in the present paper, which is an A-RNA duplex with two simple non-Watson-Crick base pairs, no substantial differences between these two force fields which could change the results are expected.

**S.5. Convergence of the systems:**

Figures S3 to S11 show the convergence of $\Delta G^\circ$ of the simulations reported in this paper. Tables S10 and S11 show $\Delta G^\circ_2$ and $\Delta G^\circ_3$ values at times $t$ and $t/2$. The results show that $\Delta G^\circ$ values for individual simulations are well converged.

**S.6. Energy Decomposition – Method 1:**

\[
\Delta G^\circ = \Delta G^\circ_{\text{bond}} + \Delta G^\circ_{\text{angle}} + \Delta G^\circ_{\text{dihedral}} + \Delta G^\circ_{\text{es}} + \Delta G^\circ_{\text{vdw}} \tag{S1}
\]

The free energy of any transformation can be written as eq S1, where $\Delta G^\circ_{\text{bond}}$, $\Delta G^\circ_{\text{angle}}$, and $\Delta G^\circ_{\text{dihedral}}$ are the free energy contributions of the bonded interactions, and $\Delta G^\circ_{\text{es}}$ and $\Delta G^\circ_{\text{vdw}}$ are the free energy contributions of electrostatic and van der Waals interactions, respectively.

Thermodynamic cycles of the restrained transformations with positional restraints were analyzed. Trajectory files of the restrained $\lambda$ simulation with positional restraints were used in order to get the individual contributions. The first 250 ps of the simulations were omitted, and the rest were used in the
analysis. There were 250 structures for each \( \lambda \) simulation, except for the \( \lambda = 0.15 \) to \( \lambda = 0.85 \) simulations of CGAG → iCGAiG (\( \Delta G^o_3 \), imino) transformations, which had 750 structures.

For each structure, a single-point energy calculation was done and individual terms were extracted from these calculations. A sample data set for each \( \lambda \) simulation was created this way. Mean and standard error of mean were calculated for each sample set. The sample data set was divided into 3 parts, and the group averaging method was used to get the standard error of mean. Table 4 of the paper shows the results of this decomposition method.

**S.7. Energy Decomposition – Method 2:**

\[
\Delta G^o = \Delta G^o_{RNA-env} + \Delta G^o_{HB} + \Delta G^o_{cross} + \Delta G^o_{ss1} + \Delta G^o_{ss2} + \Delta G^o_{other}
\]  

(S2)

To give a physical sense to the alchemical transformations, the free energy was decomposed according to eq S2. The terms, \( \Delta G^o_{RNA-env} \), \( \Delta G^o_{HB} \), \( \Delta G^o_{cross} \), \( \Delta G^o_{ss1} \), and \( \Delta G^o_{ss2} \), represent free energy changes due to the RNA-environment, hydrogen bonds within base pairs, cross-strand stacking, single strand 1 stacking, and single strand 2 stacking, respectively. \( \Delta G^o_{other} \) is the total free energy change due to the alchemical transformations of the individual bases. Table S9 shows the detailed results of this decomposition method.

Each of these free energy terms was calculated with the TI Approach using the same structures extracted from the trajectory files as used for Energy Decomposition Method 1. This means that for each \( \lambda \) simulation,

\[
\left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda} = \frac{\partial f(\lambda)}{\partial \lambda} \left\langle H_A - H_B \right\rangle_{\lambda}
\]  

(S3)

was calculated, where \( < \ldots >_{\lambda} \) is the ensemble average of the derivative of the hybrid Hamiltonian \( H(\lambda) \) with respect to \( \lambda \), \( H_A \) and \( H_B \) are the Hamiltonians of the initial and final states, respectively, and \( f(\lambda) \) is the new mixing function defined in eq 8. For simplicity, the following equation is used in the following descriptions:

\[
\Delta H = H_A - H_B
\]  

(S4)

**S.7.1 Calculation of \( \Delta G^o_{RNA-env} \):**
The term $\Delta G^{\circ}_{RNA-env}$ gives a rough estimate of the free energy change due to the interactions of the RNA with both water and sodium ions in the alchemical transformations. The following initial and final Hamiltonians were used to describe the free energy change due to the RNA-environment interactions:

$$H_A = H_{all,initial} - H_{vac,initial}$$
$$H_B = H_{all,final} - H_{vac,final}$$  \hspace{1cm} (S5)

Here, $H_{all,initial}$ and $H_{all,final}$ are the Hamiltonians of the initial and final states, including RNA, solvent, and ions, and $H_{vac,initial}$ and $H_{vac,final}$ are the Hamiltonians of the initial and final states of the RNA in vacuum. As a result, eq S4 becomes

$$\Delta H_{RNA-env} = (H_{all,initial} - H_{all,final}) - (H_{vac,initial} - H_{vac,final})$$  \hspace{1cm} (S6)

In the single point energy calculations, the boundaries for $H_{all}$ and $H_{vac}$ calculations were the same.

The new mixing rule with $k=6$ (icfe=2 and klambd=6) was used in all the TI calculations. Therefore, the derivative of $f(\lambda)$ with respect to $\lambda$ is as follows:

$$\frac{\partial f(\lambda)}{\partial \lambda} = (-2772)(1-\lambda)^5 \lambda^5$$  \hspace{1cm} (S7)

**S.7.2 Calculation of $\Delta G^{\circ}_{HB}$:**

HB stands for Hydrogen Bond and this term gives a rough estimate of the free energy changes due to the change in the base-pair hydrogen bonding for the base pairs involved in alchemical transformations.

The following equations are used in eq S4:

$$H_A = H_{bp,initial} - H_{b1,initial} - H_{b2,initial}$$
$$H_B = H_{bp,final} - H_{b1,final} - H_{b2,final}$$  \hspace{1cm} (S8)

Here, $H_{bp,initial}$ and $H_{bp,final}$ are the Hamiltonians of the initial and final base pairs that are transformed, $H_{b1,initial}$ and $H_{b2,initial}$ are the Hamiltonians of the individual bases of the initial base pair, and $H_{b1,final}$ and $H_{b2,final}$ are the Hamiltonians of the individual bases of the final base pair. As a result, eq S4 becomes as follows:

$$\Delta H_{HB} = (H_{bp,initial} - H_{bp,final}) - (H_{b1,initial} + H_{b2,initial} - H_{b1,final} - H_{b2,final})$$  \hspace{1cm} (S9)

That is, to calculate $\Delta G^{\circ}_{HB}$ of the alchemical base pair transformations of GC $\rightarrow$ iGiC. $H_A$ and $H_B$ are going to be as follows:
\[
H_A = H_{GC} - H_{G} - H_{C} \\
H_B = H_{iGC} - H_{iG} - H_{iC}
\] (S10)

For reverse transformation, \(H_A\) and \(H_B\) are reversed. All calculations were done in vacuum.

**S.7.3 Calculations of \(\Delta G^\circ\) cross, \(\Delta G^\circ\) ss1, and \(\Delta G^\circ\) ss2:**

\(\Delta G^\circ\) cross is the free energy change due to the cross strand interactions excluding hydrogen bonding in base pairs. In a duplex there are two strands. As a result, \(\Delta G^\circ\) ss1 and \(\Delta G^\circ\) ss2 are the free energy changes of single strand 1 and single strand 2 due to the intra-strand stacking interactions in the alchemical transformations. Both strands have the same sequence, so in principle, \(\Delta G^\circ\) ss1 and \(\Delta G^\circ\) ss2 would be equal for infinitely long simulations. Thus, they were added to provide the estimate for \(\Delta G^\circ\) ss listed in Table 5 (\(\Delta G^\circ\) ss = \(\Delta G^\circ\) ss1 + \(\Delta G^\circ\) ss2).

In order to calculate \(\Delta G^\circ\) cross, the following equation for eq S4 is used:

\[
\Delta H_{cross} = \left( H_{\text{duplex, initial}} - H_{ss1, initial} - H_{ss2, initial} \right) - \left( H_{\text{duplex, final}} - H_{ss1, final} - H_{ss2, final} \right) - \Delta H_{HB}
\] (S11)

Here, \(H_{\text{duplex, initial}}\) and \(H_{\text{duplex, final}}\) are the Hamiltonians of the duplexes of the initial and final systems, \(H_{ss1, initial}\) and \(H_{ss2, initial}\) are the Hamiltonians of the single strand 1 and 2 of the initial system, and \(H_{ss1, final}\) and \(H_{ss2, final}\) are the Hamiltonians of the single strand 1 and 2 of the final system.

In order to calculate \(\Delta G^\circ\) ss1 and \(\Delta G^\circ\) ss2, the following equations for eq S4 were used:

\[
\Delta H_{ss1} = \left( H_{ss1, initial} - \sum_i H_{\text{base, } i, initial} \right) - \left( H_{ss1, final} - \sum_i H_{\text{base, } i, final} \right)
\]
\[
\Delta H_{ss2} = \left( H_{ss2, initial} - \sum_i H_{\text{base, } i, initial} \right) - \left( H_{ss2, final} - \sum_i H_{\text{base, } i, final} \right)
\] (S12)

Here, \(H_{\text{base, } i, initial}\) and \(H_{\text{base, } i, final}\) are the Hamiltonians of the base number \(i\) of the initial and final states.

The summations include all the bases in the single strands.

**S.7.4 Calculation of \(\Delta G^\circ\) other:**

Because we are decomposing the free energies of alchemical transformations, the free energy changes of the individual bases, which are the components of smallest size in the system, will dominate the free energy \(\Delta G^\circ\) defined in eq S2 (see Table S9). As an example, let’s say that we want to calculate the free energy change of the alchemical transformation of \(G \rightarrow iG\). Because this is an unnatural transformation,
the free energy change representing this transformation is unphysical. As a result, we calculated the free energy changes of the transformations of these individual bases. Because only the closing base pairs of GC of the tandem GA loops are alchemically transformed to iG\( \text{iC} \), there are only 4 such free energy components (residue numbers of 3, 6, 11, and 14) (see Table S9).
Figure S1. Residues of C, iC, G, and iG, which have dummy atoms in the alchemical transformations.
Figure S2. RMSD as a function of time of 19 $\lambda$ simulations of CGAG $\rightarrow$ iCGAiG transformation (sheared GA – unrestrained $\lambda$ simulations). This plot shows that the RMSD fluctuations of the $\lambda$ simulations are around 1.5 Å as calculated with respect to the initial structure.
Figure S3. Convergence of $\Delta G^\circ_2$ (sheared – amber99) (green), $\Delta G^\circ_3$ (imino – amber99) (black), $\Delta G^\circ_2$ (sheared – modified amber99) (blue), and $\Delta G^\circ_3$ (imino – modified amber99) (red) of CGAG thermodynamic cycle as a function of time (unrestrained simulations).
**Figure S4.** Convergence of $\Delta G^\circ_2$ (sheared – amber99) (red), and $\Delta G^\circ_3$ (imino – amber99) (black) of GGAC thermodynamic cycle as a function of time (forward transformations - unrestrained simulations).
Figure S5. Convergence of $\Delta G^\circ_2$ (sheared – amber99) (red), and $\Delta G^\circ_3$ (imino – amber99) (black) of GGAC thermodynamic cycle as a function of time (backward transformations - unrestrained simulations). Note that the $\Delta G^\circ$’s are negative because of the directions of the transformations.
Figure S6. Convergence of $\Delta G^\circ_3$ (imino – modified amber99) of GGAC thermodynamic cycle as a function of time of i) forward transformation (black), ii) backward transformation (green), and iii) forward transformation with a different starting structure (red) (unrestrained simulations).
Figure S7. Convergence of $\Delta G^\circ_2$ (sheared – modified amber99) of GGAC thermodynamic cycle as a function of time of i) forward transformation (black), ii) backward transformation (green), and iii) forward transformation with a different starting structure (red) (unrestrained simulations).
Figure S8. Convergence of $\Delta G^\circ_3$ (imino GA) of CGAG thermodynamic cycle as a function of time with amber99 (black), and modified amber99 force fields (red) (restrained simulations with positional restraints).
Figure S9. Convergence of $\Delta G^\circ_2$ (sheared GA) of CGAG thermodynamic cycle as a function of time with amber99 (black), and modified amber99 (red) force fields (restrained simulations with positional restraints).
Figure S10. Convergence of $\Delta G^\circ_3$ (imino GA) of GGAC thermodynamic cycle as a function of time with amber99 (black), and modified amber99 (red) force fields (restrained simulations with positional restraints).
Figure S11. Convergence of $\Delta G^\circ_2$ (sheared GA) of GGAC thermodynamic cycle as a function of time with amber99 (black), and modified amber99 (red) force fields (restrained simulations with positional restraints).
Figure S12. Convergence of $\Delta G^o_2$ (sheared – amber99) (green), $\Delta G^o_3$ (imino – amber99) (black), $\Delta G^o_2$ (sheared – modified amber99) (blue), and $\Delta G^o_3$ (imino – modified amber99) (red) of CGAG thermodynamic cycle as a function of time (restrained simulations with H-Bond restraints).
Figure S13. Convergence of $\Delta G^\circ_2$ (sheared – amber99) (green), $\Delta G^\circ_3$ (imino – amber99) (black), $\Delta G^\circ_2$ (sheared – modified amber99) (blue), and $\Delta G^\circ_3$ (imino – modified amber99) (red) of GGAC thermodynamic cycle as a function of time (restrained simulations with H-Bond restraints).
Table S1. Modified force field file frcmod_kg. These parameters allow the amino group of guanine to be non-planar.

| Parameters for modified guanine with non-planar amino group |
|------------------------------------------------------------|
| **MASS**                                                  |
| HY             | 1.008 |
| BOND           |       |
| HY-N2          | 434.0 | 1.010 |
| ANGL           |       |
| HY-N2-HY       | 30.0  | 116.00 |
| CA-N2-HY       | 30.0  | 116.00 |
| DIHE           |       |
| NA-CA-N2-HY    | 4     | 7.00  | 180.0 | 2. |
| NC-CA-N2-HY    | 4     | 7.00  | 180.0 | 2. |
| NONBOND        |       |
| HY             | 0.6000 | 0.0157 |
### Table S2. Parameters to define isoguanosine, isocytidine and dummy atoms.

| Parameters | MASS   | BOND          | ANGL             | DIHE            | IMPR             | NONB      |
|------------|--------|---------------|------------------|-----------------|------------------|-----------|
| **MASS**   |        | O -DH  434.0  | CA-N*  424.0     | DH-O -C -CB  1.0  | DH-DH-O -C -CM 1.0 | 0.00000  |
|            |        |               |                  |                 |                  |           |
| **BOND**   |        |               |                  |                 |                  |           |
|            |        |               |                  |                 |                  |           |
| **ANGL**   |        |               |                  |                 |                  |           |
|            |        |               |                  |                 |                  |           |
| **DIHE**   |        |               |                  |                 |                  |           |
|            |        |               |                  |                 |                  |           |
| **IMPR**   |        |               |                  |                 |                  |           |
|            |        |               |                  |                 |                  |           |
| **NONB**   |        |               |                  |                 |                  |           |
|            |        |               |                  |                 |                  |           |

**Missing params of dummy atoms and c,ic,g,ig residues**
Table S3. RESP charges of IC5, IC, IC3

|        | IC5      | IC         | IC3        |
|--------|----------|------------|------------|
| H5T    | 0.448031 |            |            |
| P      |          | 1.220980   | 1.220980   |
| O1P    | -0.792102| -0.792102  | -0.792102  |
| O2P    | -0.792102| -0.792102  | -0.792102  |
| O5'    | -0.637700| -0.506956  | -0.506956  |
| C5'    | 0.037653 | 0.037653   | 0.037653   |
| H5'1   | 0.077690 | 0.077690   | 0.077690   |
| H5'2   | 0.077690 | 0.077690   | 0.077690   |
| C4'    | 0.037451 | 0.037451   | 0.037451   |
| O4'    | -0.325306| -0.325306  | -0.325306  |
| C1'    | 0.025342 | 0.025342   | 0.025342   |
| H1'    | 0.166911 | 0.166911   | 0.166911   |
| C2'    | 0.065525 | 0.065525   | 0.065525   |
| H2'1   | 0.055914 | 0.055914   | 0.055914   |
| O2'    | -0.606921| -0.606921  | -0.606921  |
| HO'2   | 0.425760 | 0.425760   | 0.425760   |
| N1     | -0.052713| -0.052712  | -0.052712  |
| C2     | 0.779754 | 0.779754   | 0.779754   |
| N2     | -0.942735| -0.942735  | -0.942735  |
| H21    | 0.401568 | 0.401568   | 0.401568   |
| H22    | 0.401568 | 0.401568   | 0.401568   |
| N3     | -0.810455| -0.810455  | -0.810455  |
| C4     | 0.960425 | 0.960425   | 0.960425   |
| O4     | -0.650446| -0.650446  | -0.650446  |
| C5     | -0.507673| -0.507673  | -0.507673  |
| H5     | 0.193524 | 0.193524   | 0.193524   |
| C6     | -0.084938| -0.084938  | -0.084938  |
| H6     | 0.214206 | 0.214206   | 0.214206   |
| H4'    | 0.133183 | 0.133183   | 0.133183   |
| C3'    | 0.271957 | 0.271957   | 0.271957   |
| H3'    | 0.062677 | 0.062677   | 0.062677   |
| O3'    | -0.537432| -0.537432  | -0.665874  |
| H3T    |          |            | 0.447932   |
|     | RC5             | RC              | RC3            |
|-----|-----------------|-----------------|----------------|
| H5T | 0.433837        | 1.220980        | 1.220980       |
| P   |                 | 0.792102        | 0.792102       |
| O1P |                | -0.792102       | -0.792102      |
| O2P |                | -0.792102       | -0.792102      |
| O5' | -0.623506       | -0.506956       | -0.506956      |
| C5' | 0.038747        | 0.038747        | 0.038747       |
| H5' | 0.075663        | 0.075663        | 0.075663       |
| H5' | 0.075663        | 0.075663        | 0.075663       |
| C4' | 0.066359        | 0.066359        | 0.066359       |
| H4' | 0.133662        | 0.133662        | 0.133662       |
| O4' | -0.358640       | -0.358640       | -0.358640      |
| C1' | 0.021717        | 0.021717        | 0.021717       |
| H1' | 0.177559        | 0.177559        | 0.177559       |
| C2' | 0.116810        | 0.116810        | 0.116810       |
| H2' | 0.088089        | 0.088089        | 0.088089       |
| O2' | -0.603726       | -0.603726       | -0.603726      |
| H0' | 0.417221        | 0.417221        | 0.417221       |
| N1  | -0.025987       | -0.025986       | -0.025986      |
| C2  | 0.783130        | 0.783130        | 0.783130       |
| O2  | -0.624282       | -0.624282       | -0.624282      |
| N3  | -0.767443       | -0.767443       | -0.767443      |
| C4  | 0.826745        | 0.826745        | 0.826745       |
| C5  | -0.522460       | -0.522460       | -0.522460      |
| H5  | 0.192144        | 0.192144        | 0.192144       |
| N4  | -0.957688       | -0.957688       | -0.957688      |
| H41 | 0.411809        | 0.411809        | 0.411809       |
| H42 | 0.411809        | 0.411809        | 0.411809       |
| C6  | -0.040620       | -0.040620       | -0.040620      |
| H6  | 0.225025        | 0.225025        | 0.225025       |
| C3' | 0.176434        | 0.176434        | 0.176434       |
| H3' | 0.069871        | 0.069871        | 0.069871       |
| O3' | -0.537432       | -0.537432       | -0.660896      |
| H3T |                 |                 | 0.442954       |
Table S5. RESP charges of IG5, IG, IG3

|      | IG5     | IG       | IG3      |
|------|---------|----------|----------|
| H5T  | 0.430457|          |          |
| P    | 1.220980| 1.220980 |          |
| O2P  | -0.792102| -0.792102| -0.792102|
| O1P  | -0.792102| -0.792102| -0.792102|
| O5'  | -0.620126| -0.506956| -0.506956|
| C5'  | 0.078194 | 0.078194 | 0.078194 |
| H5'2 | 0.059901 | 0.059901 | 0.059901 |
| H5'1 | 0.059901 | 0.059901 | 0.059901 |
| C4'  | 0.087820 | 0.087820 | 0.087820 |
| H4'  | 0.092671 | 0.092671 | 0.092671 |
| O4'  | -0.353925| -0.353925| -0.353925|
| C1'  | 0.068202 | 0.068202 | 0.068202 |
| H1'  | 0.142092 | 0.142092 | 0.142092 |
| C2'  | 0.008437 | 0.008437 | 0.008437 |
| H2'1 | 0.123321 | 0.123321 | 0.123321 |
| O2'  | -0.581986| -0.581986| -0.581986|
| H0'2 | 0.414800 | 0.414800 | 0.414800 |
| N9   | 0.030161 | 0.030162 | 0.030162 |
| C4   | 0.197066 | 0.197066 | 0.197066 |
| C5   | 0.191948 | 0.191948 | 0.191948 |
| C6   | 0.330497 | 0.330497 | 0.330497 |
| N1   | -0.427296| -0.427296| -0.427296|
| H1   | 0.336019 | 0.336019 | 0.336019 |
| C2   | 0.765782 | 0.765782 | 0.765782 |
| O2   | -0.635949| -0.635949| -0.635949|
| N6   | -0.812476| -0.812476| -0.812476|
| H61  | 0.406693 | 0.406693 | 0.406693 |
| H62  | 0.406693 | 0.406693 | 0.406693 |
| N7   | -0.583932| -0.583932| -0.583932|
| N3   | -0.637384| -0.637384| -0.637384|
| C8   | 0.156297 | 0.156297 | 0.156297 |
| H8   | 0.157485 | 0.157485 | 0.157485 |
| C3'  | 0.229632 | 0.229632 | 0.229632 |
| H3'  | 0.096947 | 0.096947 | 0.096947 |
| O3'  | -0.537432| -0.537432| -0.632508|
| H3T  |        |          | 0.414566 |
|       | RG5         | RG          | RG3          |
|-------|-------------|-------------|-------------|
| H5T  | 0.436704    | 1.220980    | 1.220980    |
| P    |             | 1.220980    | 1.220980    |
| O1P  | -0.792102   | -0.792102   | -0.792102   |
| O2P  | -0.792102   | -0.792102   | -0.792102   |
| O5'  | -0.626373   | -0.506956   | -0.506956   |
| C5'  | 0.084286    | 0.084286    | 0.084286    |
| H5'1 | 0.055150    | 0.055150    | 0.055150    |
| H5'2 | 0.055150    | 0.055150    | 0.055150    |
| C4'  | 0.133343    | 0.133343    | 0.133343    |
| H4'  | 0.104390    | 0.104390    | 0.104390    |
| O4'  | -0.412578   | -0.412578   | -0.412578   |
| C1'  | 0.203435    | 0.203435    | 0.203435    |
| H1'  | 0.134663    | 0.134663    | 0.134663    |
| C2'  | 0.059602    | 0.059602    | 0.059602    |
| H2'1 | 0.098840    | 0.098840    | 0.098840    |
| O2'  | -0.624700   | -0.624700   | -0.624700   |
| HO'2 | 0.428604    | 0.428604    | 0.428604    |
| N9   | -0.024617   | -0.024616   | -0.024615   |
| C4   | 0.137267    | 0.137267    | 0.137267    |
| C5   | 0.170451    | 0.170451    | 0.170451    |
| C6   | 0.524469    | 0.524469    | 0.524469    |
| O6   | -0.548027   | -0.548027   | -0.548027   |
| N1   | -0.526159   | -0.526159   | -0.526159   |
| H1   | 0.354636    | 0.354636    | 0.354636    |
| C2   | 0.725074    | 0.725074    | 0.725074    |
| N2   | -0.904385   | -0.904385   | -0.904385   |
| H21  | 0.395915    | 0.395915    | 0.395915    |
| H22  | 0.395915    | 0.395915    | 0.395915    |
| N7   | -0.570593   | -0.570593   | -0.570593   |
| N3   | -0.598347   | -0.598347   | -0.598347   |
| C8   | 0.147708    | 0.147708    | 0.147708    |
| H8   | 0.156992    | 0.156992    | 0.156992    |
| C3'  | 0.194800    | 0.194800    | 0.194800    |
| H3'  | 0.056327    | 0.056327    | 0.056327    |
| O3'  | -0.537432   | -0.537432   | -0.665597   |
| H3T  |             | 0.447654    |             |
### Table S7. RESP charges of RA5, RA, RA3

|       | RA5  | RA    | RA3   |
|-------|------|-------|-------|
| H5T   | 0.435778 | 1.220980 | 1.220980 |
| P     |       | 0.792102 | 0.792102 |
| O1P   |       | -0.792102 | -0.792102 |
| O2P   |       | -0.792102 | -0.792102 |
| O5'   | -0.625447 | -0.506956 | -0.506956 |
| C5'   | 0.056379  | 0.056379  | 0.056379  |
| H5'1  | 0.067207  | 0.067207  | 0.067207  |
| H5'2  | 0.067207  | 0.067207  | 0.067207  |
| C4'   | 0.097730  | 0.097730  | 0.097730  |
| H4'   | 0.117234  | 0.117234  | 0.117234  |
| O4'   | -0.386099 | -0.386099 | -0.386099 |
| C1'   | 0.112729  | 0.112729  | 0.112729  |
| H1'   | 0.177137  | 0.177137  | 0.177137  |
| C2'   | 0.054459  | 0.054459  | 0.054459  |
| H2'1  | 0.103963  | 0.103963  | 0.103963  |
| O2'   | -0.611327 | -0.611327 | -0.611327 |
| HO'2  | 0.423845  | 0.423845  | 0.423845  |
| N9    | -0.048865 | -0.048864 | -0.048864 |
| C4    | 0.307976  | 0.307976  | 0.307976  |
| C5    | 0.085758  | 0.085758  | 0.085758  |
| C6    | 0.697913  | 0.697913  | 0.697913  |
| N1    | -0.767773 | -0.767773 | -0.767773 |
| C2    | 0.605238  | 0.605238  | 0.605238  |
| H2    | 0.041189  | 0.041189  | 0.041189  |
| N6    | -0.920677 | -0.920677 | -0.920677 |
| H61   | 0.406843  | 0.406843  | 0.406843  |
| H62   | 0.406843  | 0.406843  | 0.406843  |
| N7    | -0.608493 | -0.608493 | -0.608493 |
| N3    | -0.706538 | -0.706538 | -0.706538 |
| C8    | 0.192718  | 0.192718  | 0.192718  |
| H8    | 0.159416  | 0.159416  | 0.159416  |
| C3'   | 0.215041  | 0.215041  | 0.215041  |
| H3'   | 0.060558  | 0.060558  | 0.060558  |
| O3'   | -0.537432 | -0.537432 | -0.659907 |
| H3T   |       | 0.441965  |       |
Table S8. RESP charges of RU5, RU, RU3

|     | RU5  | RU               | RU3               |
|-----|------|------------------|-------------------|
| H5T | 0.437155 | 1.220980   | 1.220980         |
| P   |       | -0.792102  | -0.792102        |
| O1P |       | -0.792102  | -0.792102        |
| O2P |       | -0.792102  | -0.792102        |
| O5' | -0.626824 | -0.506956  | -0.506956        |
| C5' | 0.062377  | 0.062377   | 0.062377         |
| H5'1| 0.071084  | 0.071084   | 0.071084         |
| H5'2| 0.071084  | 0.071084   | 0.071084         |
| C4' | 0.057104  | 0.057104   | 0.057104         |
| H4' | 0.134347  | 0.134347   | 0.134347         |
| O4' | -0.353866 | -0.353866  | -0.353866        |
| C1' | 0.027969  | 0.027969   | 0.027969         |
| H1' | 0.189321  | 0.189321   | 0.189321         |
| C2' | 0.055048  | 0.055048   | 0.055048         |
| H2'1| 0.097072  | 0.097072   | 0.097072         |
| O2' | -0.597239 | -0.597239  | -0.597239        |
| H0'2| 0.425523  | 0.425523   | 0.425523         |
| N1  | 0.057895  | 0.057896   | 0.057897         |
| C2  | 0.496634  | 0.496634   | 0.496634         |
| O2  | -0.548680 | -0.548680  | -0.548680        |
| N3  | -0.402932 | -0.402932  | -0.402932        |
| H3  | 0.313275  | 0.313275   | 0.313275         |
| C4  | 0.634795  | 0.634795   | 0.634795         |
| O4  | -0.569910 | -0.569910  | -0.569910        |
| C5  | -0.378463 | -0.378463  | -0.378463        |
| H5  | 0.179166  | 0.179166   | 0.179166         |
| C6  | -0.113585 | -0.113585  | -0.113585        |
| H6  | 0.221782  | 0.221782   | 0.221782         |
| C3' | 0.219375  | 0.219375   | 0.219375         |
| H3' | 0.058435  | 0.058435   | 0.058435         |
| O3' | -0.537432 | -0.537432  | -0.666178        |
| H3T |       | 0.448235    |                   |
Table S9. Decomposition of the free energy (kcal/mol) to RNA-environment, hydrogen bond, cross-strand stacking, single strand 1 stacking, single strand 2 stacking, and other interactions, which include alchemical transformations of the individual bases from Thermodynamic Integration Method with Restrained Molecular Dynamics with positional restraints.

|                  | S.E.       | HB        | cs.S.     | ss1.S. | ss2.S. | B.3 | B.6 | B.11 | B.14 | Total |
|------------------|------------|-----------|-----------|--------|--------|-----|-----|------|------|-------|
| A.i.99           | 1.7±0.3    | -11.1±0.1 | -10.5±0.2 | -10.2±0.1 | -8.0±0.2 | 46.9±0.1 | 22.5±0.2 | 46.4±0.1 | 21.8±0.1 | 109.5±0.5 |
| A.s.99           | 1.7±0.4    | -3.0±0.1  | -55.4±0.2 | 22.4±0.2 | 4.7±0.2  | 46.2±0.1 | 23.2±0.2 | 46.2±0.1 | 23.6±0.2 | 109.6±0.6 |
| (Diff)           | 0.0±0.5    | 1.9±0.2   | 44.9±0.3  | -32.6±0.2 | -12.7±0.3 | 0.7±0.2  | -0.7±0.3 | 0.2±0.2  | -1.8±0.2 | -0.1±0.8 |
| B.i.99           | 0.3±0.1    | -1.8±0.1  | -13.8±0.1 | -2.8±0.1 | -6.8±0.1 | 21.8±0.1 | 46.3±0.1 | 21.5±0.1 | 46.7±0.1 | 111.4±0.3 |
| B.s.99           | -8.3±0.4   | 0.2±0.1   | 16.5±0.2  | -12.3±0.1 | -18.3±0.2 | 21.7±0.2 | 45.7±0.1 | 22.0±0.2 | 45.8±0.1 | 113.0±0.6 |
| (Diff)           | 8.6±0.4    | -2.0±0.2  | -30.3±0.2 | 9.5±0.2  | 11.5±0.2 | 0.1±0.2  | 0.6±0.2  | -0.5±0.2 | 0.9±0.2  | -1.6±0.7 |
| A.i.S            | 0.9±0.3    | -1.3±0.1  | -6.6±0.2  | -12.4±0.1 | -8.4±0.1 | 47.0±0.1 | 23.0±0.3 | 46.4±0.1 | 21.5±0.2 | 110.0±0.6 |
| A.s.S            | 1.3±0.3    | -2.7±0.1  | -58.3±0.2 | 24.3±0.2 | 6.4±0.2  | 46.3±0.2 | 22.4±0.2 | 46.4±0.1 | 23.6±0.3 | 109.7±0.6 |
| (Diff)           | -0.4±0.4   | 1.4±0.2   | 51.7±0.3  | -36.7±0.2 | -14.8±0.2 | 0.7±0.2  | 0.6±0.4  | 0.0±0.2  | -2.1±0.4 | 0.3±0.8 |
| B.i.S            | -1.2±0.2   | -2.1±0.1  | -23.7±0.1 | 1.3±0.1  | 0.4±0.1  | 47.3±0.1 | 21.6±0.1 | 46.3±0.0 | 111.7±0.3 |
| B.s.S            | -7.6±0.3   | -0.2±0.1  | 13.4±0.2  | -15.8±0.1 | -14.3±0.2 | 22.7±0.2 | 46.3±0.1 | 22.5±0.1 | 45.6±0.1 | 112.6±0.5 |
| (Diff)           | 6.4±0.3    | -1.9±0.2  | -37.1±0.2 | 17.1±0.2 | 14.7±0.2 | -0.9±0.2 | 1.0±0.2  | -0.9±0.2 | 0.7±0.1  | -0.9±0.6 |

*In Table 5, ΔG°_ss is reported, which is the sum of ss1.S. and ss2.S (ΔG°_ss = ΔG°_ss1 + ΔG°_ss2).*

**Note:**
- S.E. - RNA-Environment Interaction (Environment = Solvent + ions)
- HB - hydrogen bonding between basepaired residues
- cs.S. - cross-strand stacking (includes overlap and cross-strand un-basepaired interactions)
- ss1.S. - single-strand # 1 stacking
- ss2.S. - single-strand # 2 stacking
- B.3 - Base # 3 alchemical transformation (G to iG or C to iC)
- B.6 - Base # 6 alchemical transformation (G to iG or C to iC)
- B.11 - Base # 11 alchemical transformation (G to iG or C to iC)
- B.14 - Base # 14 alchemical transformation (G to iG or C to iC)
- Diff - (imino - sheared)
Table S10. Error analysis of $\text{GGAC} \rightarrow \text{iGGAiC}$ thermodynamic cycle for amber99 and modified amber99 force field unrestrained calculations (values are in kcal/mol). Figures S4, S5, S6, and S7 show the convergence of each alchemical transformations of the $\text{GGAC} \rightarrow \text{iGGAiC}$ thermodynamic cycle.

|                      | amber99                                                                 | modified amber99                                                                 |
|----------------------|-------------------------------------------------------------------------|---------------------------------------------------------------------------------|
|                      | $\Delta G^\circ_2$ (sheared)                                            | $\Delta G^\circ_2$ (sheared)                                                   |
|                      | $\Delta G^\circ_3$ (imino)                                             | $\Delta G^\circ_3$ (imino)                                                   |
|                      | 0.875 ns 1.750 ns 0.875 ns 1.750 ns                                      | 0.875 ns 1.750 ns 0.875 ns 1.750 ns                                           |
| $\text{GGAC} \rightarrow \text{iGGAiC}^a$ | 110.99 111.05 110.38 110.45                                           | 110.952 111.214 110.99 111.07                                               |
| $\text{iGGAiC} \rightarrow \text{GGAC}^b$ | 110.75 111.00 110.52 110.28                                           | 111.346 111.066 111.41 111.37                                               |
| Mean and St. Error   | 110.87 ± 0.12 111.02 ± 0.03 110.45 ± 0.07 110.36 ± 0.09               | 111.18 ± 0.10 111.23 ± 0.10 111.18 ± 0.10 111.15 ± 0.09                      |

$^a$ $\text{GGAC} \rightarrow \text{iGGAiC}$ alchemical transformation (forward); $^b$ $\text{iGGAiC} \rightarrow \text{GGAC}$ alchemical transformation (backward); $^c$ another $\text{GGAC} \rightarrow \text{iGGAiC}$ alchemical transformation with a different starting structure (forward).
Table S11. Error analysis of CGAG → iCGAiG thermodynamic cycle for amber99 and modified amber99 force field unrestrained calculations (values are in kcal/mol). Figure S3 shows the convergence of each alchemical transformation of the CGAG → iCGAiG thermodynamic cycle.

|                | amber99        | modified amber99 |
|----------------|----------------|------------------|
|                | ΔG°<sub>2</sub> (sheared) | ΔG°<sub>3</sub> (imino) |
| CGAG → iCGAiG | 0.875 ns 1.750 ns | 0.875 ns 1.750 ns |
|                | 111.296 111.160 | 110.585 110.602 |
|                | 111.440 111.306 | 110.669 110.755 |
Table S12. Details of Individual Thermodynamic Integration Calculations with Unrestrained Molecular Dynamics of GGAC → iGGAiC (in kcal/mol)

|                      | amber99                        | modified amber99                  |
|----------------------|--------------------------------|-----------------------------------|
|                      | ΔG°₂ (sheared) | ΔG°₃ (imino) | ΔG°₂ (sheared) | ΔG°₃ (imino) |
| GGAC → iGGAiC<sup>a</sup> | 111.048         | 110.449     | 111.214         | 111.072       |
| iGGAiC → GGAC<sup>b</sup>  | 110.995         | 110.278     | 111.402         | 111.019       |
| Mean and Error       | 111.022 ± 0.027 | 110.364 ± 0.086 | 111.227 ± 0.100 | 111.155 ± 0.110 |

<sup>a</sup> GGAC → iGGAiC alchemical transformation (forward);  <sup>b</sup> iGGAiC → GGAC alchemical transformation (backward);  <sup>c</sup> another GGAC → iGGAiC alchemical transformation with a different starting structure (forward).
Table S13. An example of the parameters used in the CGAG $\rightarrow$ iCGAiG imino transformation for $\lambda=0.5$ Molecular Dynamics simulation.

| 2. CONTROL DATA FOR THE RUN |
|------------------------------|
| General flags:               |
| imin = 0, nmropt = 0         |
| Nature and format of input:  |
| ntx = 1, irest = 0, ntrx = 1 |
| Nature and format of output:|
| ntxo = 1, ntpr = 50, ntrx = 1, ntwr = 1000 |
| iwrap = 0, ntwx = 1000, ntwv = 0, ntwe = 0 |
| ioutfm = 0, ntwprt = 0, idecomp = 0, rbornstat = 0 |
| Potential function:          |
| ntf = 2, ntb = 1, igb = 0, nsnb = 25 |
| ipol = 0, gbsa = 0, iesp = 0 |
| dielc = 1.00000, cut = 8.00000, intdiel = 1.00000 |
| scnb = 2.00000, scee = 1.20000 |
| Frozen or restrained atoms:  |
| ibelly = 0, ntr = 0          |
| Molecular dynamics:          |
| nstlim = 1000000, nscm = 5000, nrespa = 1 |
| t = 0.00000, dt = 0.00100, vlimit = 20.00000 |
| Langevin dynamics temperature regulation: |
| ig = 233                    |
| temp0 = 300.00000, tempi = 300.00000, gamma_ln = 1.00000 |
| SHAKE:                       |
| ntc = 2, jfastw = 0          |
| tol = 0.000001              |
| Free energy options:        |
| icfe = 2                    |
| klambda = 6                 |
| clambda = 0.500000          |
| Ewald parameters:           |
| verbose = 0, ew_type = 0, nbflag = 1, use_pme = 1 |
| vdmeth = 1, eedmeth = 1, netfrc = 1 |
| Box X = 51.905, Box Y = 51.905, Box Z = 51.905 |
| Alpha = 109.471, Beta = 109.471, Gamma = 109.471 |
| NFFT1 = 54, NFFT2 = 54, NFFT3 = 54 |
| Cutoff = 8.000, Tol = 0.100E-04 |
| Ewald Coefficient = 0.34864 |
| Interpolation order = 4     |
| Setting ifbox to 3 for non-orthogonal unit cell |
| Noshake mask @%DH | (:RCD,RGD,ICD,IGD@%H & !@H1); matches 16 atoms |
| Removing shake constraints from N4 RCD 3 -- H41 RCD 3 |
| Removing shake constraints from N4 RCD 3 -- H42 RCD 3 |
| Removing shake constraints from N2 RGD 6 -- H21 RGD 6 |
| Removing shake constraints from N2 RGD 6 -- H22 RGD 6 |
| Removing shake constraints from N4 RCD 11 -- H41 RCD 11 |
| Removing shake constraints from N4 RCD 11 -- H42 RCD 11 |
| Removing shake constraints from N2 RGD 14 -- H21 RGD 14 |
| Removing shake constraints from N2 RGD 14 -- H22 RGD 14 |
| Setting ntf to 1             |
Table S14. The hydrogen bond restraint file used in the CGAG → iCGAiG imino transformations.

```plaintext
# 1 RG5 N1 16 RC3 N3 2.5 3.5
&rst
  ixpk=0, nxpk=0, iat=20, 518, r1=2.00, r2=2.50, r3=3.50, r4=4.00,
  rk2=25.0, rk3=25.0, ir6=1, iatld=0,
&end
#
# 1 RG5 N2 16 RC3 O2 2.5 3.5
&rst
  ixpk=0, nxpk=0, iat=23, 517, r1=2.00, r2=2.50, r3=3.50, r4=4.00,
&end
#
# 1 RG5 O6 16 RC3 N4 2.5 3.5
&rst
  ixpk=0, nxpk=0, iat=19, 522, r1=2.00, r2=2.50, r3=3.50, r4=4.00,
&end
#
# 2 RG N1 15 RC N3 2.5 3.5
&rst
  ixpk=0, nxpk=0, iat=54, 487, r1=2.00, r2=2.50, r3=3.50, r4=4.00,
&end
#
# 2 RG N2 15 RC O2 2.5 3.5
&rst
  ixpk=0, nxpk=0, iat=57, 486, r1=2.00, r2=2.50, r3=3.50, r4=4.00,
&end
#
# 2 RG O6 15 RC N4 2.5 3.5
&rst
  ixpk=0, nxpk=0, iat=53, 491, r1=2.00, r2=2.50, r3=3.50, r4=4.00,
&end
#
# 3 RCD N3 14 RGD N1 2.5 3.5
&rst
  ixpk=0, nxpk=0, iat=85, 452, r1=2.00, r2=2.50, r3=3.50, r4=4.00,
&end
#
# 3 RCD O2 14 RGD N2 2.5 3.5
&rst
  ixpk=0, nxpk=0, iat=92, 455, r1=2.00, r2=2.50, r3=3.50, r4=4.00,
&end
#
# 3 RCD N4 14 RGD O6 2.5 3.5
&rst
  ixpk=0, nxpk=0, iat=89, 458, r1=2.00, r2=2.50, r3=3.50, r4=4.00,
&end
#
# 4 RG N1 13 RA N1 2.5 3.5
&rst
  ixpk=0, nxpk=0, iat=121, 419, r1=2.00, r2=2.50, r3=3.50, r4=4.00,
&end
#
# 4 RG N6 13 RA N6 2.5 3.5
&rst
  ixpk=0, nxpk=0, iat=120, 422, r1=2.00, r2=2.50, r3=3.50, r4=4.00,
&end
#
# 5 RA N1 12 RG N1 2.5 3.5
&rst
  ixpk=0, nxpk=0, iat=154, 386, r1=2.00, r2=2.50, r3=3.50, r4=4.00,
&end
#
# 5 RA N6 12 RG O6 2.5 3.5
&rst
  ixpk=0, nxpk=0, iat=157, 385, r1=2.00, r2=2.50, r3=3.50, r4=4.00,
&end
#
# 6 RGD N1 11 RCD N3 2.5 3.5
&rst
  ixpk=0, nxpk=0, iat=187, 350, r1=2.00, r2=2.50, r3=3.50, r4=4.00,
&end
#
# 6 RGD N2 11 RCD O2 2.5 3.5
&rst
  ixpk=0, nxpk=0, iat=190, 357, r1=2.00, r2=2.50, r3=3.50, r4=4.00,
&end
#
# 6 RGD O6 11 RCD N4 2.5 3.5
&rst
```

| 7  | RC N3 | 10  | RG N1 | r1 | 2.00 | r2 | 2.50 | r3 | 3.50 | r4 | 4.00 | &end |
|----|-------|-----|-------|----|------|----|------|----|------|----|------|------|
| 7  | RC O2 | 10  | RG N2 | r1 | 2.00 | r2 | 2.50 | r3 | 3.50 | r4 | 4.00 | &end |
| 7  | RC N4 | 10  | RG O6 | r1 | 2.00 | r2 | 2.50 | r3 | 3.50 | r4 | 4.00 | &end |
| 8  | RC3 N3| 9   | RG5 N1| r1 | 2.00 | r2 | 2.50 | r3 | 3.50 | r4 | 4.00 | &end |
| 8  | RC3 O2| 9   | RG5 N2| r1 | 2.00 | r2 | 2.50 | r3 | 3.50 | r4 | 4.00 | &end |
| 8  | RC3 N4| 9   | RG5 O6| r1 | 2.00 | r2 | 2.50 | r3 | 3.50 | r4 | 4.00 | &end |
Table S15. Results in kcal/mol from Thermodynamic Integration with Unrestrained and Restrained Molecular Dynamics

|                | No Restraints |              | Experimental |
|----------------|---------------|--------------|--------------|
|                |               | Amber99      |              |
|                | alchemical transformation | ΔG°₂ | ΔG°₃ | ΔG°₃ – ΔG°₂ | ΔG°¹ – ΔG°₄ |
| GGAC → iGGAiC | 111.022 ± 0.027 | 110.364 ± 0.086 | -0.7 ± 0.1 | ≥ 3.1 |
| CGAG → iCGAiG | 111.158 ± 0.087 | 110.602 ± 0.081 | -0.6 ± 0.1 | ≤ -2.2 |
|                | amber99 with non-planar G amino group in GA base pairs | | |
|                |              | ΔG°₂ | ΔG°₃ | ΔG°₃ – ΔG°₂ | ΔG°¹ – ΔG°₄ |
| GGAC → iGGAiC | 111.227 ± 0.100 | 111.155 ± 0.110 | -0.1 ± 0.2 | ≥ 3.1 |
| CGAG → iCGAiG | 111.306 ± 0.086 | 110.755 ± 0.073 | -0.6 ± 0.1 | ≤ -2.2 |
|                | Positional Restraints | Amber99 | | |
|                | alchemical transformation | ΔG°₂ | ΔG°₃ | ΔG°₃ – ΔG°₂ | ΔG°¹ – ΔG°₄ |
| GGAC → iGGAiC | 110.09 ± 0.18 | 110.13 ± 0.15 | 0.0 ± 0.2 | > 3.1 |
| CGAG → iCGAiG | 112.67 ± 0.16 | 110.58 ± 0.06 | -2.1 ± 0.2 | < -2.2 |
|                | amber99 with non-planar G amino group in GA base pairs | | |
|                |              | ΔG°₂ | ΔG°₃ | ΔG°₃ – ΔG°₂ | ΔG°¹ – ΔG°₄ |
| GGAC → iGGAiC | 109.96 ± 0.16 | 110.45 ± 0.15 | 0.5 ± 0.2 | > 3.1 |
| CGAG → iCGAiG | 112.23 ± 0.18 | 110.41 ± 0.11 | -1.8 ± 0.2 | < -2.2 |
|                | H-Bond Restraints | Amber99 | | |
|                | alchemical transformation | ΔG°₂ | ΔG°₃ | ΔG°₃ – ΔG°₂ | ΔG°¹ – ΔG°₄ |
| GGAC → iGGAiC | 111.13 ± 0.14 | 110.13 ± 0.11 | -1.0 ± 0.2 | > 3.1 |
| CGAG → iCGAiG | 110.93 ± 0.08 | 110.30 ± 0.11 | -0.6 ± 0.1 | < -2.2 |
|                | amber99 with non-planar G amino group in GA base pairs | | |
|                |              | ΔG°₂ | ΔG°₃ | ΔG°₃ – ΔG°₂ | ΔG°¹ – ΔG°₄ |
| GGAC → iGGAiC | 111.31 ± 0.10 | 111.30 ± 0.07 | 0.0 ± 0.1 | > 3.1 |
| CGAG → iCGAiG | 111.15 ± 0.10 | 111.12 ± 0.09 | 0.0 ± 0.1 | < -2.2 |

a Details of these calculations are described in Table S12.
References:

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

2. Bludsky, O.; Sponer, J.; Leszczynski, J.; Spirko, V.; Hobza, P. Amino groups in nucleic acid bases, aniline, aminopyridines, and aminotriazine are nonplanar: Results of correlated ab initio quantum chemical calculations and anharmonic analysis of the aniline inversion motion. *J. Chem. Phys.* 1996, 105, 11042.

3. Dong, F.; Miller, R. E. Vibrational transition moment angles in isolated biomolecules: A structural tool. *Science* 2002, 298, 1227.

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

5. 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.

6. Sponer, J.; Jurecka, P.; Hobza, P. Accurate interaction energies of hydrogen-bonded nucleic acid base pairs. *J. Am. Chem. Soc.* 2004, 126, 10142.

7. Rupert, P. B.; Ferre-D'Amare, A. R. Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature* 2001, 410, 780.

8. Prive, G. G.; Heinemann, U.; Chandrasegaran, S.; Kan, L. S.; Kopka, M. L.; Dickerson, R. E. Helix geometry, hydration, and GA mismatch in a B-DNA decamer. *Science* 1987, 238, 498.

9. Meier, R. J.; Coussens, B. The molecular-structure of hydrazine and melamine - Rotational barriers and hybridization. *Theochem-Journal of Molecular Structure* 1990, 68, 303.

10. Sponer, J.; Hobza, P. Nonplanar geometries of DNA bases. Second order Moller-Plesset study. *J. Phys. Chem.* 1994, 98, 3161.
11. Sponer, J.; Hobza, P. Bifurcated hyrdogen-bonds in DNA crystal-structures - An ab-initio Quantum-Chemical study. J. Am. Chem. Soc. 1994, 116, 709.

12. Vlieghe, D.; Sponer, J.; Van Meervelt, L. Crystal structure of d(GGCCAATTGG) complexed with DAPI reveals novel binding mode. Biochemistry 1999, 38, 16443.

13. Luisi, B.; Orozco, M.; Sponer, J.; Luque, F. J.; Shakked, Z. On the potential role of the amino nitrogen atom as a hydrogen bond acceptor in macromolecules. J. Mol. Biol. 1998, 279, 1123.

14. Sponer, J.; Hobza, P. DNA base amino groups and their role in molecular interactions: Ab initio and preliminary density functional theory calculations. Int. J. Quantum Chem. 1996, 57, 959.

15. Spackova, N.; Cheatham, T. E.; Ryjacek, F.; Lankas, F.; van Meervelt, L.; Hobza, P.; Sponer, J. Molecular dynamics simulations and thermodynamics analysis of DNA-drug complexes. Minor groove binding between 4 ’,6-diamidino-2-phenylindole and DNA duplexes in solution. J. Am. Chem. Soc. 2003, 125, 1759.

16. Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. Biophys. J. 2007, 92, 3817.

17. Reblova, K.; Lankas, F.; Razga, F.; Krasovska, M. V.; Koca, J.; Sponer, J. Structure, dynamics, and elasticity of free 16S rRNA helix 44 studied by molecular dynamics simulations. Biopolymers 2006, 82, 504.

18. Spackova, N.; Sponer, J. Molecular dynamics simulations of sarcin-ricin rRNA motif. Nucleic Acids Res. 2006, 34, 697.