THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Effect of Guanine to Inosine Substitution on Stability of Canonical DNA and RNA Duplexes: Molecular Dynamics Thermodynamics Integration Study

Miroslav Krepl,[†] Michal Otyepka,[‡] Pavel Banáš,*[,†,‡] and Jiří Šponer*[,†,§]

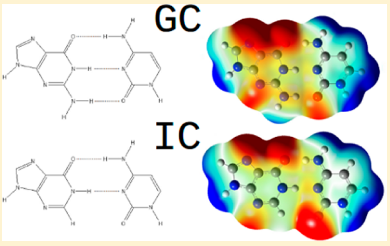
[†]Institute of Biophysics, Academy of Sciences of the Czech Republic, Kralovopolska 135, 612 65 Brno, Czech Republic

[‡]Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Faculty of Science, Palacky University, tr. 17 listopadu 12, 771 46 Olomouc, Czech Republic

[§]CEITEC—Central European Institute of Technology, Campus Bohunice, Kamenice 5, 625 00 Brno, Czech Republic

Supporting Information

ABSTRACT: Guanine to inosine (G → I) substitution has often been used to study various properties of nucleic acids. Inosine differs from guanine only by loss of the N2 amino group, while both bases have similar electrostatic potentials. Therefore, G → I substitution appears to be optimally suited to probe structural and thermodynamics effects of single H-bonds and atomic groups. However, recent experiments have revealed substantial difference in free energy impact of G → I substitution in the context of B-DNA and A-RNA canonical helices, suggesting that the free energy changes reflect context-dependent balance of energy contributions rather than intrinsic strength of a single H-bond. In the present study, we complement the experiments by free energy computations using thermodynamics integration method based on extended explicit solvent molecular dynamics simulations. The computations successfully reproduce the basic qualitative difference in free energy impact of G → I substitution in B-DNA and A-RNA helices although the magnitude of the effect is somewhat underestimated. The computations, however, do not reproduce the salt dependence of the free energy changes. We tentatively suggest that the different effect of G → I substitution in A-RNA and B-DNA may be related to different topologies of these helices, which affect the electrostatic interactions between the base pairs and the negatively charged backbone. Limitations of the computations are briefly discussed.



INTRODUCTION

Nucleic acids (RNA and DNA) play essential roles in a wide range of biological processes, including storage of genetic information, regulation, and catalytic function.[1] To properly understand functions of nucleic acids, it is necessary to know their three-dimensional (3D) structures.[2] 3D structures of nucleic acids are largely influenced by their secondary (2D) structures, i.e., the base pairing.[3,4] This calls for development of tools which can predict the 2D structure from a sequence. Successful prediction of 2D structure is obviously necessary for any attempts to model 3D structure. However, prediction of 2D structure of nucleic acids, especially of RNA, is difficult. For RNAs with length up to ∼700 nucleotides, contemporary methods for predicting the 2D structure by computational folding of a single sequence achieve only ∼70% accuracy.[5] This is calculated as the percentage of correctly predicted Watson−Crick (WC, canonical) base pairs minus predicted base pairs that do not occur.[5] The accuracy can be considerably improved when additional experimental data are available, for example, chemical or enzymatic probing data of the folded RNA molecule.[6]

Essential data for 2D prediction of nucleic acids are free energy parameters derived from benchmark thermodynamics (TD) experiments of short canonical duplexes and eventually small noncanonical nucleic acids elements.[7−11] The nearest-neighbor TD model treats a NA helix as a string of interactions between "neighboring" base pairs. The nearest-neighbor parameters for consecutive canonical (Watson−Crick) base pairs constitute the essence of free-energy based 2D predictions, deriving 2D structures from the sequence using minimization of the (predicted) free energy.[5] Obviously, especially for RNA, it is very vital to extend the available TD parameters also beyond the canonical base pairing, to include the noncanonical interactions. Unfortunately, extension of the free energy models beyond the canonical helices has been so far less successful, in part because of lack of sufficient knowledge of the balance of forces that govern the overall free energy data.[12]

An obvious disadvantage of TD experiments is that they do not provide any decomposition of the free energies on the individual energy contributions. There are two main forces between nucleic acid bases that affect their structure and stability, namely, base stacking and base pairing. Traditionally, it has been assumed that the sequence dependency of free energies measured in TD experiments should reflect the intrinsic strength of base pairing and stacking. If this view is correct, then it would be ultimately possible to derive plausible

Received: November 12, 2012
Revised: January 10, 2013
Published: January 27, 2013

dx.doi.org/10.1021/jp311180u | J. Phys. Chem. B 2013, 117, 1872−1879

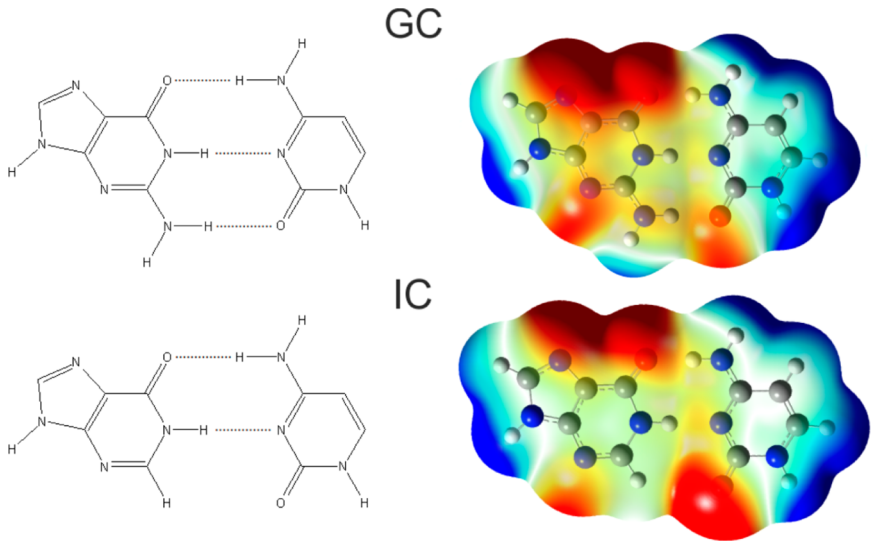


Figure 1. Schematic representation of guanine/cytosine (top) and inosine/cytosine (bottom) base pairs and electron density surface (isovalue = 0.0004 au) visualization according to the base pairs electrostatic potentials (right). Structures were optimized in Gaussian09 using PBE/cc-pVTZ. The Gaussview 5 program was used for visualization (scale: red ≤ −3.5e−2 au, blue ≥ −3.5e−2 au).

free energy models using accurate energy data of direct base pairing and stacking interactions, eventually supplemented by some simple models to consider solvent screening of the electrostatic interactions.[13] The intrinsic energies can be nowadays quite accurately derived by QM methods. However, there is mounting evidence that the above simple assumption is unrealistic.[14,15] Rather, the free energies ultimately result from a very complex mixture of numerous energy contributions, with the base pairing and stacking energies being just part of the overall balance.[4,16] In addition, the balance may differ from system to system, limiting the transferability of the TD data. Thus, the profound difference between free energy effect of guanine−cytosine (GC) to inosine−cytosine (IC) base pair substitution in the context of RNA and DNA canonical helices analyzed in the present study is one example of the complexity of the interactions. Another known example is the difference between incorporation of consecutive protonated cytosines into i-DNA tetraplex and Pyr-Pur-Pyr DNA triplex.[17−19] While stacking of protonated cytosines is a native canonical interaction in i-DNA, consecutive stacked protonated cytosines disrupt the triplex.

Computational methods, despite their numerous limitations, can provide useful complementary insight to the TD experiments,[20] especially when analyzing effects of small modifications of the studied systems. In this particular case, free energy computations based on explicit solvent atomistic simulations can be applied.[21] Although the force fields used in the simulations are in principle less accurate than QM calculations used for benchmark computations on model systems, the simulations can fully include the solvent effects and provide appropriate sampling of the conformational space.[4] Specifically, the thermodynamics integration (TI) method has been shown to furnish very promising results in evaluation of relative free energies of systems that are close to isosteric, i.e., are not associated with any significant geometrical changes.[22−28]

Since base mutations (even change of single functional group) can alter base pairing (secondary structure) and subsequently change shape of the entire system, it is necessary to understand TD implications of these interactions in the context of not only the interaction itself but of the entire system. The sequence dependence of Gibbs free energy (ΔG) of stability of nucleic acids structures can be measured by UV melting experiments.[9,10] The change of stability upon a given modification, for example, loss of a single H-bond, is often variable and depends not only on the modification but also on other contributions and the overall context. This phenomenon was termed nonadditivity.[29,30] One substitution/modification already present in the helix can modulate free energy effects of subsequent modifications.[31] This can be explained by the effect of the first modification on the strength of neighboring hydrogen bonds and stacking interactions. Additionally, since both major and minor grooves usually possess a specific system of dynamical ion and water mediated interactions,[32] any change of functional groups that perturbs the solvent network may have subsequent influence on free energies that cannot be deduced from the strength of the H-bond itself. The fact that nonadditivity is highly dependent upon the structural context supports this notion. For example, the nonadditivity is much greater at helix ends than in the middle of helices, indicating that substitution at helix ends may leave the terminal nucleotides unpaired and therefore insensitive to further mutations.[31] Nonadditivity of nucleic acids substitutions was extensively studied using double- and triple-cycle mutations.[16,31,33]

From the basic physical-chemistry point of view, interesting base substitution is the guanine (G) → inosine (I) substitution in the Watson−Crick GC base pair (Figure 1). This is one of the smallest changes that can occur in nucleic acids, since it includes just deletion of the N2 amino group of guanine occurring in the minor groove of canonical double helices. At first sight, such substitution should highlight free energy changes associated with a loss of single H-bond, since the GC and IC base pairs adopt the same WC geometry and differ only in the number of H-bonds. Further, the electrostatic potential of inosine resembles guanine, although the dipole moment of inosine is somewhat smaller (Figure 1).[34] It thus can be assumed that the G → I substitution can monitor the effect of altering a single functional group and, in terms of free energy, a loss of single H-bond. Indeed, the G → I substitution has been

dx.doi.org/10.1021/jp311180u | J. Phys. Chem. B 2013, 117, 1872−1879

many times used to study various properties of nucleic acids, both experimentally and theoretically.[35−40] Inosine nucleotide also occurs naturally in many RNA molecules such as tRNAs due to post-transcriptional modifications.[41]

Recent experiments revealed that the G → I substitution results in substantially different reduction of stability in B-DNA and A-RNA canonical helices (∼2.0 and ∼3.4 kcal/mol, respectively).[31] This shows that the free energy effect of G → I substitution cannot be rationalized by change of the intrinsic base pairing strength. It even is not straightforwardly explainable by stacking, since stacking in A-RNA and B-DNA is rather similar[15] and the electrostatic part of stacking of G and I is in general comparable, although the steric effect of the N2 amino group is lost with I.

In this work we introduce GC → IC base pair substitution into A-RNA and B-DNA helix and use the TI method[42] with molecular dynamics simulation in explicit solvent to compute the destabilization energy associated with this substitution. The calculations are compared with previous TD experiments.[31] The TI method excellently captures the different effect of G → I substitution in DNA and RNA canonical helices, although the effect is underestimated. However, the computations do not reproduce the experimental salt dependence of the free energy changes. Possible reasons for this limitation are discussed. In addition, we tried to use the energy decomposition method to elucidate the nature of energy contributions associated with the G → I transformation. Although such decompositions suffer from sizable approximations,[43] we obtained some clues regarding the balance of the free energy contributions and propose a hypothesis about possible origin of the free energy differences between A-RNA and B-DNA helices with the G → I substitution.

METHODS

Starting Structures. We used the web-based Nucleic Acid Builder to create an initial set of starting coordinates for canonical A-RNA and B-DNA helices simulations. In both cases, we used the sequence of the helix part from RNA hairpin from ref 31 extended by two additional GC base pairs at both ends to prevent end effects (Figure 2). Simulations may be more sensitive to end effects than experiments due to force field limitations. For this reason, we have performed the calculations only for internal G → I substitution where the experimental data is available. In all simulations, guanine 6 (GC pair) has been transformed into inosine (IC pair) (Figure 1). Structures for reference single-strand simulations were obtained by removal of the strand without mutation. Gaussian09[44] was used to compute electrostatic potentials of GC/IC base pairs and Gaussview[45] for visualization of electron density surface in Figure 1.

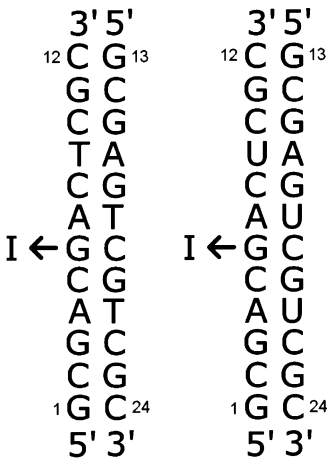


Figure 2. Sequence of DNA and RNA helices used in simulations. Arrow indicates site of the guanine → inosine substitution.

System Building. We used the xLEaP program from the AMBER 12[46] program suite to prepare starting coordinates of all simulations. Bsc0 force field[47,48] was used for DNA and bsc0χ_{OL3} for RNA.[47−50] Different ion concentrations of Na^+—100, 200, and 500 mM—were used. We used the TIP3P water model[51] with minimal distance of 18, 15, and 10 Å from the solute border for 100, 200, and 500 mM simulations. The concentrations refer to Na^+ concentrations and are calculated from the volume of the simulation box and the number of the ions. In the case of 100 and 200 mM Na^+ simulations, we had to significantly extend the size of the water box as even net-neutral number of Na^+ ions would yield more than the desired Na^+ concentration using the standard box size (which is 10 Å from the solute border). The size of the water box (in terms of the initial distance from the solute border) and concentration of Na^+ ions were the same for both double-strand and single-strand simulations of the respective ion strength. However, in order to compensate for the charge of the missing strand in the single-strand simulations, there has been a difference in Cl^- concentrations: 0, 24, and 167 mM for double-strand simulations and 48, 104, and 323 mM for single-strand simulations. We decided not to test the low-salt conditions (∼14 mM) which were also used for some experimental measurements[31] as they would require prohibitively large water boxes in the simulation setup. Note again that, with presently used simulation boxes, the cation concentration is pretty large even upon the basic net-neutralizing condition (as noted above, the ion concentration is given by the number of ions and size of the simulation box—see more detailed discussion below). Nevertheless, we make an assumption that the salt concentration dependency trends should be observable even in the 100−500 mM range we use in our simulations. For further comment on the substantial difference of ion treatment in experimental and simulations setups, see Results and Discussion.

We used Aqvist parameters for sodium (radius 1.868 Å and well depth of 0.002 77 kcal/mol)[52] and Smith parameters (radius 2.47 Å and well depth of 0.1 kcal/mol)[53] for chloride ions. We have monitored the simulations for possible ion-pair artifacts which might occur with this ion parameter combination (though rather for KCl than NaCl), but we did not observe any problems.[54] Further, we conducted one control simulation (RNA, 500 mM) using Joung et al. parameters (radius/well depth of 1.369 Å/0.0874 kcal/mol for sodium and 2.513 Å/0.0356 kcal/mol for chloride)[55] adapted for the TIP3P water model. The ion model does not appear to have any effect on the calculated free energies (see Results and Discussion).

We first used xLEaP to create topology and coordinates of the system with guanine. It was then again loaded in xLEaP and manually modified to obtain topology with inosine. It is necessary to obtain an identical set of coordinates for TI calculations with the only unique atoms being in the transformation region, in this case C2 atom and N2 amino group of guanine versus C2 and H2 atoms of inosine. Note that we included the C2 base atom in the scheme, as the method we used (see below) prohibits bonds with applied SHAKE

dx.doi.org/10.1021/jp311180u | J. Phys. Chem. B 2013, 117, 1872−1879

restrains to cross into the transformation region, which would be the case with the inosine C2−H2 bond. To obtain coordinates with inosine, we used the system with guanine as a basis. Since we assume that the guanine → inosine mutation has only a small impact on the rest of the atom positions, we used a coordinates file containing a previously minimized system with guanine and removed coordinates of hydrogen atoms from the base N2 amino group. The C2−N2 bond length was then changed to the length of C2−H2 bond in inosine. In this way, we obtained minimized starting coordinates for both systems, one with guanine and one with inosine with identical coordinates of individual atoms (apart from the transformation region).

AMBER Simulations. Prior to production simulations, all systems were minimized and then slowly heated to 298.16 K over 100 ps using 2 ps integration step and Langevin thermostat.[56] Production simulation runs were performed with the Sander module of AMBER 12.[46] The particle mesh Ewald (PME)[57,58] was used for calculation of electrostatic interactions. Covalent bonds involving hydrogen atoms were constrained using the SHAKE algorithm.[59] Periodic boundary conditions (PBC), a 2 fs integration step, and 298.16 K temperature (Langevin Thermostat) were used. We used 9 Å as a cutoff distance for Lennard-Jones interactions. Trajectories error checkings and analyses not related to TI calculations were performed using the ptraj module of AMBER and VMD program.[60]

Thermodynamic Integration. Two sets of topologies are necessary for TI calculations.[42] The first is for the starting state (V_0), and the second corresponds to the final state (V_1). We used soft core potentials[61] to handle different numbers of atoms in guanine and inosine. Using this method, appearing and disappearing atoms can be present in the system at the same time. This is implemented through the modified equation of van der Waals (vdW) interactions which ensures that vdW interactions of unique atoms with the rest of the system are smoothly switched off. Similarly, soft core electrostatics[62] were used to handle the electrostatic component. The Sander program of AMBER collects statistics ($\partial V/\partial\lambda$ values) which we used to numerically estimate the ΔG energy integral. We used nine-point Gaussian quadrature[46] for integral estimation (Table 1). TI calculations were carried out in parallel runs for nine

Table 1. Values of Weights and Quadrature Nine-Points and the Equation Used To Estimate the ΔG Energy Integral[a]

nine points quadrature			
λ_i	λ_i	w_i	
0.01592	0.98408	0.04064	$\Delta G_i \approx \sum_n^{\lambda_i} w_i \left(\frac{\delta V}{\delta\lambda}\right)_i$
0.08198	0.91802	0.09032	
0.19331	0.80669	0.13031	
0.33787	0.66213	0.15617	
0.5		0.16512	

[a] λ_i are lambdas of specific simulation runs. w_i are weights of the mean.

lambdas. The length of each simulation run was 20 ns, and $\partial V/\partial\lambda$ values were collected every integration step. The number of collected $\partial V/\partial\lambda$ values was then 10 millions for every lambda value. Energy for each step of the G → I transformation was calculated from these values by nine-point Gaussian quadrature (Table 1) using in-house scripts. Standard deviations of the results were then calculated according to common statistical principles using in-house scripts. One of the problems we had to overcome was strong autocorrelation of the 10 millions $\partial V/\partial\lambda$ values. This was done by standard decorrelation procedure, namely, calculating 10 average values per every million of $\partial V/\partial\lambda$ values and then calculating standard deviations of these 10 averages. Final energy difference of the G → I transformation was calculated according to the equation (thermodynamics cycle, Figure 3) $\Delta\Delta G = \Delta G_{GC} - \Delta G_{IC} = \Delta G_{ds} - \Delta G_{ss}$.

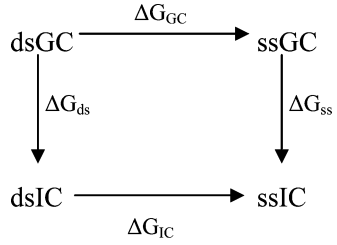


Figure 3. Thermodynamic cycle of guanine ↔ inosine transformation and duplex formation ↔ melting processes. ΔG_{ds} and ΔG_{ss} represent sums over steps from G → I transformation obtained by TI in the context of the double strand and single strand. ΔG_{GC} and ΔG_{IC} represent dissociation free energies of helix with GC and IC base pair. Using a thermodynamics cycle, we can derive $\Delta\Delta G = \Delta G_{GC} - \Delta G_{IC} = \Delta G_{ds} - \Delta G_{ss}$. The TI method divides the "alchemical" G ↔ I transformation, which, however, can be calculated, into a series of small energy windows (lambdas). For each energy window, derivatives of the systems energy are computed. The integral of energy for the entire step is then estimated using a sufficient number of energy windows and their reasonable weights (Table 1). Note that although the G ↔ I process is unphysical, due to the thermodynamics cycle we can use this transformation to calculate the $\Delta G_{ds} - \Delta G_{ss}$ values of real duplex dimerization processes which cannot be evaluated directly via computations.

ΔG_{ds} and ΔG_{ss} represent sums over steps from G → I transformation in double helix and single strand, respectively, i.e., the processes that have been calculated. ΔG_{GC} and ΔG_{IC} represent dissociation free energies of helix without and with inosine mutation, respectively. These energy differences are directly equivalent to the process measured in the experimental setting (Figure 3). In this way, we obtain difference of dissociation energies between double-stranded nucleic acids with GC pair and IC pair ($\Delta\Delta G$).

Energy Decomposition. We used the Anal. module of AMBER 5[63] for postsimulation energy decomposition. Several representative atom groups were defined for both double-stranded and single-stranded system (see the Supporting Information).

RESULTS AND DISCUSSION

The Calculations Reproduce the Basic Experimental Trends. Molecular dynamics simulations (2.52 μs) (126 runs, each 20 ns long) were conducted. The results of the TI calculations are presented in Table 2 and compared with the relevant experimental studies. The main reference data were taken from Siegfried et al.[31] In this study, a G → I mutation was synthetically introduced into RNA and DNA. A difference between melting temperature of helices with and without this mutation was then measured by UV spectrometry. Note that the structure we used was somewhat different from Siegfried et al.[31] (missing the capping tetraloop and having the helix ends extended—see Methods). However, at least three base pairs (four base pairs on one side) adjacent to the point of mutation were of identical sequence. We did not investigate substitutions

dx.doi.org/10.1021/jp311180u | J. Phys. Chem. B 2013, 117, 1872−1879

Table 2. Results of TI Calculations of Guanine ↔ Inosine Transformation (expressed as $\Delta\Delta G$) and Their Comparison with Experimental Values

data	ion strength (mM)	$\Delta\Delta G$ (kcal/mol)	error of $\Delta\Delta G$
	A-RNA		
experiment[31]	14	3.44	0.06
experiment[31]	114	3.11	0.09
simulation	100	2.6	0.4
simulation	200	2.7	0.6
simulation	500	2.7	0.5
simulation[a]	500	2.5	0.4
	B-DNA		
experiment[31]	14	2.01	0.06
experiment[31]	114	1.30	0.08
simulation	100	1.9	0.4
simulation	200	2.1	0.6
simulation	500	2.0	0.7

[a]Control simulation conducted with Joung et al. set of ion parameters[55] (see Methods).

at the helix ends, as the quality of force field and affordable sampling may be insufficient to guarantee convergence of the results, since sampling of unpaired geometries may be needed to analyze substitutions at the helix ends. In another study, Sun et al.[33] also reported energy values for two simultaneous GC → IC mutations in a DNA helix; however, we were unable to effectively extract the value for one mutation due to nonadditivity (see Introduction) from this particular work.

Our TI computations qualitatively reproduce the main experimental result (Table 2). Namely, the destabilization effect of GC → IC transformation is visibly greater in RNA helix than in DNA helix, although the simulations slightly underestimate the difference (~1.6 kcal/mol in experiment and ~0.7 kcal/mol in computations). Also the absolute values of free energy changes are in a reasonable agreement with the experimental data. The experimental values are essentially all within internal error limits of the TI procedure. Considering all the approximations in the computations (including the substantial force field approximations) and many differences between the simulation and experimental setups (see below and ref 64), the results indicate satisfactory performance of the simulation force field as well as of the TI procedure. In fact, our agreement between theory and experiment is likely the maximum for what we could hope. For example, uncertainty of the best reference gas phase QM calculations of base pairing energies remains around ~0.5 kcal/mol.[65,66]

Additional Computations. To further assess validity of our results, we did some additional calculations. We computed free energy difference at one ion strength for a second position of the GC → IC mutation in our RNA helix (bp. C8/G17, see Figure 2) for which there is an experimental value (albeit only for one ion strength).[31] The value of 2.1 kcal/mol that we obtained by our TI calculations is in good agreement with the 2.6 kcal/mol obtained in the experiment.[31] We also tried to estimate the sequence effect, by investigating the impact of eliminating the C5pG6 step in B-DNA, since CpG B-DNA steps have specific stacking properties due to the minor groove steric clash between the guanine amino groups.[67] We thus change the C5/G20 base pair to A5/U20. The C5/G20 → A5/U20 substitution slightly reduced (from 2.0 to 1.6 kcal/mol) the energy penalty for the G6 → I6 substitution, which indicates that the results are not dramatically sequence/dependent.

Free Energy Decomposition. The TI procedure, in principle, allows decomposition of the calculated numbers into some individual energy contributions, and we indeed attempted such a procedure. However, we faced several problems. Namely, it is impossible to realistically separate the contributions of the missing hydrogen bond (caused by the GC → IC mutation) and change in solvation of the minor groove. We also noticed that the solvent in one periodic box does not provide a complete shielding for the nucleic acid for the decomposition purposes, which may lead to minor influence of the box size on the energy calculations (see below). Due to these uncertainties in the available decomposition procedure, we decided to not report any specific numbers. On the other hand, despite the above-mentioned limitations, the decomposition showed that long-range electrostatic interactions between the mutated base pair and phosphates, ions, and waters do matter. Thus the decomposition could tentatively suggest that the difference between RNA and DNA could be attributed to basic geometrical differences between the helices that lead to different long-range electrostatic interactions between GC/IC base pair and phosphate groups and ions (all long-range interactions being partially shielded by waters). Although the difference between dipole moment of GC and IC base pairs is small,[34] it is still significant, and thus the most relevant part of long-range electrostatic contribution to calculated $\Delta\Delta G$ is dipole−monopole interaction between dipole of GC/IC base pair and monopoles of phosphates and ions. Figure 4 illustrates the different topologies of A-form and B-form helices. While B-DNA has all bases on its helical axis with phosphates and ions spread all around, the displacement of the base pairs with respect to the helical axis in A-RNA leads to rather significant anisotropy of phosphate and ions arrangement with respect to the given base pair (Figure 4). As a consequence, we suggest that the A-form helix is more susceptible to change of dipole moment of mutated base pair than the B-form. In order to qualitatively support such a hypothesis, we calculated the sum of dipole−monopole interactions between the dipole of GC/IC base pair and monopoles of negatively charged phosphates. We indeed found that magnitude of the difference in dipole−monopole interaction energies between GC and IC base pairs was about twice higher in A-RNA than in B-DNA. Furthermore, these interactions seem to prefer different base pair step in different helix type albeit it is not possible to quantitatively interpret such information as these interactions represent only part of total long-range electrostatic contribution that includes also other two important contributions, namely interactions to ions and waters. Nonetheless, we believe it is sufficient to support the statement about higher sensitivity of A-form helix to change of dipole moment of the mutated base pair (see Figure 4).

Single-Strand Simulations. An important point to understand the limits of our computations is the behavior of the single-strand simulations, which served as a reference state for the TI calculations. The single-stranded structures were more dynamical in the simulations than the double strand. As such, proper sampling was of some concern, and it led us to the extension of all simulations from the initial length up to 20 ns. During the simulations the single strand underwent some reversible structural changes including formation of various short-lived intramolecular contacts (especially at the strand ends). Apart from that, the single-strand dynamics were

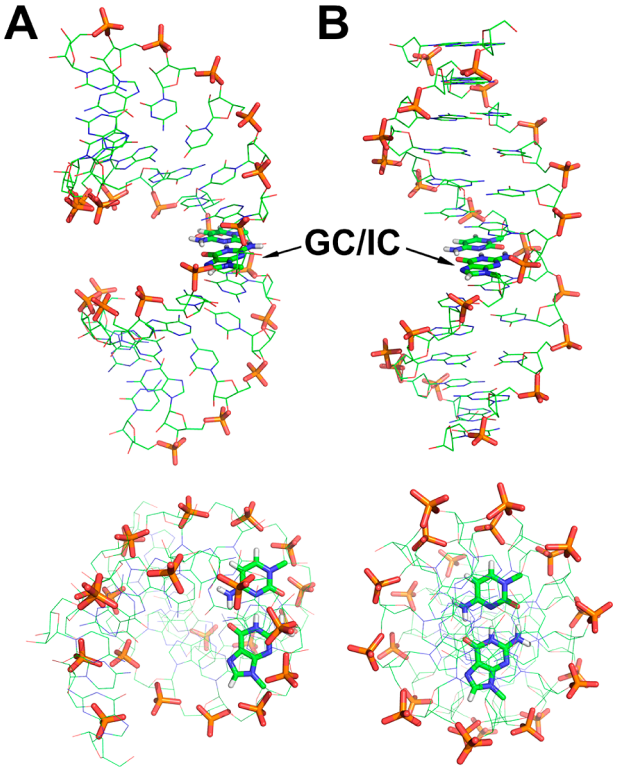


Figure 4. Sideway (top) and vertical (bottom) view of A-form (panel A) and B-form (panel B) RNA and DNA helices. While the mutated GC/IC base pair (shown in stick and highlighted by arrows) in B-DNA lies on the helical axis with phosphates (also shown in stick) and ions (not shown) spread around this base pair, in the case of A-RNA the GC/IC base pair is displaced from the helical axis. Thus the phosphates are located (and consequently also the ions that are supposed to be diffused around them) mostly on the major groove (Hoogsteen) side of the GC/IC base pair. As a consequence, the A-form helix could be more sensitive to change of the dipole moment of the mutated base pair, which is also indicated by the energy decomposition.

dominated by fluctuations of the bases, which formed continuous stack for most of the time. Random changes in χ dihedrals of the bases were then highly correlated and propagated through the whole strand. Note that due to lack of sufficient experimental data for the behavior of the single strands,[68] we are unable to confidently decide if these temporarily sampled substates are realistic or not. Fortunately, the substates are populated independently on the λ value of the respective simulations and as such are assumed to not influence the $\partial V/\partial\lambda$ values dramatically. We also include an example of $\partial V/\partial\lambda$ development (Figure S1) and a brief discussion on simulation convergence for double-strand and single-strand simulations in the Supporting Information. Nevertheless, the single-strand behavior may be limiting the accuracy of our TI calculations, from both the point of view of sampling and the force field.

Ionic Strength Dependence. The experimental results suggest a visible reduction of the G → I destabilization effects upon increased salt concentration. In contrast, the computed results are more or less insensitive to the ion concentration. However, this result is not so surprising. There is a substantial difference between the salt treatment in simulations and in experiment. Besides the obvious force field approximation (real ions vs Lennard-Jones probes with a centrally positioned fixed charge), there are other differences. The simulations are done with very high solute molecule concentration (~0.01 M). Although any direct interactions between solute molecules are prevented by the periodic boundary condition, the details of ion exchange properties between bound and bulk ions are likely to be affected.[69] The presently used simulation boxes are too small to achieve a real bulk behavior far from the solute molecule (when departing from the solute molecule, the ion enters a nearby periodic box before reaching the bulk regime of behavior).[70] Therefore, if the experimentally observed salt dependence is related to exchange of bound ions with the bulk, the simulations cannot capture it. The simulations are still expected to provide very meaningful behavior of ions near the solute molecule.[71] Additionally, as we had to extend size of water box for 100 and 200 mM simulations (see Methods), it may obscure the salt dependence for lower ion concentrations. It is possible that the size of the water box itself has larger influence on the calculated free energies than different ion concentrations in a standard size water box. To this date, due to limited computational power, almost all MD simulations of nucleic acids in explicit solvents were done with small sizes of periodic (water) boxes, leading to the above-noted high effective solute concentrations and lack of true bulk background of ions. All processes that require accurate physical description of ion exchanges such as trends associated with ion strength are then likely not fully reproduced in MD simulations. While direct binding of ions to solute is correctly (within the force field approximation) described, contemporary simulations may exaggerate "glassy" behavior of ions further away from the solute (for a recent review of ion binding to nucleic acids, see ref 72).

CONCLUSIONS

Recent experimental studies revealed surprisingly different free energy impact of G → I substitution in canonical B-DNA (~2 kcal/mol) and A-RNA (~3.4 kcal/mol). It shows that the effect of G → I substitution on double-helix stability is driven not only by a loss of an H-bond but also by other energy contributions. We have attempted to reproduce this experimental observation using extended thermodynamics integration (TI) free energy computations. The computation very well captured the basic experimental trend, although the difference has been underestimated, with the computed difference being ~0.7 kcal/mol. In other words, the destabilization effect of GC → IC transformation on helix stability and its differential effect in RNA and DNA were both successfully reproduced by TI calculations. The computations, however, do not capture the experimentally observed effect of salt strength on the free energy effect of the G → I substitution.

We have attempted decomposition of the computed free energy data, but the available procedure is not robust enough to give unambiguous detailed decomposition of the computed free energies into the individual terms. Still, we were able to identify that long-range electrostatics plays a role in the destabilization effect of the GC → IC mutation. Subsequently we suggested that A-RNA free energies of duplex formation might be more sensitive to change of dipole moment of the base pair. It is due to displacement of A-RNA base pairs out of helical axis and corresponding strongly anisotropic distribution of phosphates and counterions around the mutated base pair. We thus hypothesized that different free energy impact of GC → IC mutation in DNA and RNA might stem from different topology of B-form and A-form helices and long-range electrostatic

dx.doi.org/10.1021/jp311180u | J. Phys. Chem. B 2013, 117, 1872−1879

interactions between dipole moment of base pair, phosphates, and counterions.

The present results indicate that the MD simulations with TI calculations are likely capable to reflect the basic physical-chemistry properties that determine the free energy changes associated with the G → I substitution in real systems and can be applied also for other contexts. Limitations of the TI procedure have been discussed.

■ ASSOCIATED CONTENT

Supporting Information

Details of energy decomposition calculations; discussion about simulation convergence rate. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: pavel.banas@upol.cz (P.B.); sponer@ncbr.muni.cz (J.S.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Grant Agency of the Czech Republic (grants P305/12/G034 and P301/11/P558), "CEITEC—Central European Institute of Technology" (CZ.1.05/1.1.00/02.0068) from the European Regional Development Fund, by the Operational Program Research and Development for Innovations—European Social Fund (CZ.1.05/2.1.00/03.0058 and CZ.1.07/2.3.00/20.0017), and by the Academy of Sciences of the Czech Republic project AV0Z50040702.

■ REFERENCES

(1) Breaker, R. R. *Nature* **2004**, *432*, 838−845.
(2) *RNA 3D Structure Analysis and Prediction*; Leontis, N. B., Westhof, E., Eds; Springer-Verlag: Berlin, Heidelberg, Germany, 2012.
(3) Brion, P.; Westhof, E. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 113−137.
(4) Sponer, J.; Sponer, J. E.; Petrov, A. I.; Leontis, N. B. *J. Phys. Chem. B* **2010**, *114*, 15723−15741.
(5) Mathews, D. H.; Turner, D. H. *Curr. Opin. Struct. Biol.* **2006**, *16*, 270−278.
(6) Deigan, K. E.; Li, T. W.; Mathews, D. H.; Weeks, K. M. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 97−102.
(7) Mathews, D. H.; Sabina, J.; Zuker, M.; Turner, D. H. *J. Mol. Biol.* **1999**, *288*, 911−940.
(8) SantaLucia, J. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1460−1465.
(9) Turner, D. H.; Sugimoto, N.; Kierzek, R.; Dreiker, S. D. *J. Am. Chem. Soc.* **1987**, *109*, 3783−3785.
(10) Kierzek, R.; Burkard, M. E.; Turner, D. H. *Biochemistry* **1999**, *38*, 14214−14223.
(11) Chen, J. L.; Dishler, A. L.; Kennedy, S. D.; Yildirim, I.; Liu, B.; Turner, D. H.; Serra, M. J. *Biochemistry* **2012**, *51*, 3508−3522.
(12) Yildirim, I.; Turner, D. H. *Biochemistry* **2005**, *44*, 13225−13234.
(13) Morgado, C. A.; Svozil, D.; Turner, D. H.; Sponer, J. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12580−12591.
(14) Sponer, J.; Morgado, C. A.; Svozil, D. *J. Phys. Chem. B* **2012**, *116*, 8331−8332.
(15) Svozil, D.; Hobza, P.; Sponer, J. *J. Phys. Chem. B* **2010**, *114*, 1191−1203.
(16) Siegfried, N. A.; Kierzek, R.; Bevilacqua, P. C. *J. Am. Chem. Soc.* **2010**, *132*, 5342−5344.
(17) Spackova, N.; Berger, I.; Egli, M.; Sponer, J. *J. Am. Chem. Soc.* **1998**, *120*, 6147−6151.
(18) Gallego, J.; Golden, E. B.; Stanley, D. E.; Reid, B. R. *J. Mol. Biol.* **1999**, *285*, 1039−1052.
(19) Soliva, R.; Laughton, C. A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **1998**, *120*, 11226−11233.
(20) Spasic, A.; Serafini, J.; Mathews, D. H. *J. Chem. Theory Comput.* **2012**, *8*, 2497−2505.
(21) Koller, A. N.; Bozilovic, J.; Engels, J. W.; Gohlke, H. *Nucleic Acids Res.* **2010**, *38*, 3133−3146.
(22) Kopitz, H.; Zivkovic, A.; Engels, J. W.; Gohlke, H. *Chembiochem* **2008**, *9*, 2619−2622.
(23) Yildirim, I.; Stern, H. A.; Sponer, J.; Spackova, N.; Turner, D. H. *J. Chem. Theory Comput.* **2009**, *5*, 2088−2100.
(24) Beierlein, F. R.; Kneale, G. G.; Clark, T. *Biophys. J.* **2011**, *101*, 1130−1138.
(25) Cubero, E.; Laughton, C. A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2000**, *122*, 6891−6899.
(26) Florian, J.; Goodman, M. F.; Warshel, A. *J. Phys. Chem. B* **2000**, *104*, 10092−10099.
(27) Hernandez, B.; Soliva, R.; Luque, F. J.; Orozco, M. *Nucleic Acids Res.* **2000**, *28*, 4873−4883.
(28) Spackova, N.; Cubero, E.; Sponer, J.; Orozco, M. *J. Am. Chem. Soc.* **2004**, *126*, 14642−14650.
(29) Di Cera, E. *Chem. Rev.* **1998**, *98*, 1563−1591.
(30) Klostermeier, D.; Millar, D. P. *Biochemistry* **2002**, *41*, 14095−14102.
(31) Siegfried, N. A.; Metzger, S. L.; Bevilacqua, P. C. *Biochemistry* **2007**, *46*, 172−181.
(32) Anderson, C. F.; Record, M. T. *Annu. Rev. Phys. Chem.* **1995**, *46*, 657−700.
(33) Sun, Z. H.; McLaughlin, L. W. *Biopolymers* **2007**, *87*, 183−195.
(34) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965−1974.
(35) Lankas, F.; Cheatham, T. E.; Spackova, N.; Hobza, P.; Langowski, J.; Sponer, J. *Biophys. J.* **2002**, *82*, 2592−2609.
(36) Bailly, C.; Mollegaard, N. E.; Nielsen, P. E.; Waring, M. J. *EMBO J.* **1995**, *14*, 2121−2131.
(37) Sherer, E. C.; Harris, S. A.; Soliva, R.; Orozco, H.; Laughton, C. A. *J. Am. Chem. Soc.* **1999**, *121*, 5981−5991.
(38) ShatzkySchwartz, M.; Arbuckle, N. D.; Eisenstein, M.; Rabinovich, D.; BareketSamish, A.; Haran, T. E.; Luisi, B. F.; Shakked, Z. *J. Mol. Biol.* **1997**, *267*, 595−623.
(39) Gros, J.; Rosu, F.; Amrane, S.; De Cian, A.; Gabelica, V.; Lacroix, L.; Mergny, J. L. *Nucleic Acids Res.* **2007**, *35*, 3064−3075.
(40) Strobel, S. A.; Shetty, K. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2903−2908.
(41) Rao, S. T.; Sundaralingam, M. *J. Am. Chem. Soc.* **1969**, *91*, 1210−1217.
(42) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395−2417.
(43) Beutler, T. C.; Mark, A. E.; Vanschaik, R. C.; Gerber, P. R.; Vangunsteren, W. F. *Chem. Phys. Lett.* **1994**, *222*, 529−539.
(44) Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09*, revision A.2; Gaussian, Inc.: Wallingford CT, 2009.
(45) Dennington, R. K., T.; Millam, J. *GaussView*, version 5; Semichem Inc.: Shawnee Mission, KS, 2009.
(46) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; et al. *AMBER 12*; University of California: San Francisco, CA, 2012.
(47) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817−3829.
(48) 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. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(49) Banas, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham, T. E.; Sponer, J.; Otyepka, M. *J. Chem. Theory Comput.* **2010**, *6*, 3836−3849.

(50) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E.; Jurecka, P. *J. Chem. Theory Comput.* **2011**, *7*, 2886−2902.
(51) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(52) Aqvist, J. *J. Phys. Chem.* **1990**, *94*, 8021−8024.
(53) Smith, D. E.; Dang, L. X. *J. Chem. Phys.* **1994**, *100*, 3757−3766.
(54) Auffinger, P.; Cheatham, T. E.; Vaiana, A. C. *J. Chem. Theory Comput.* **2007**, *3*, 1851−1859.
(55) Joung, I. S.; Cheatham, T. E. *J. Phys. Chem. B* **2008**, *112*, 9020−9041.
(56) Adelman, S. A.; Doll, J. D. *J. Chem. Phys.* **1976**, *64*, 2375−2388.
(57) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(58) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(59) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(60) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(61) Steinbrecher, T.; Mobley, D. L.; Case, D. A. *J. Chem. Phys.* **2007**, *127*, 214108.
(62) Steinbrecher, T.; Joung, I.; Case, D. A. *J. Comput. Chem.* **2011**, *32*, 3253−3263.
(63) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1−41.
(64) Sponer, J.; Cang, X. H.; Cheatham, T. E. *Methods* **2012**, *57*, 25−39.
(65) Sponer, J.; Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142−10151.
(66) Takatani, T.; Hohenstein, E. G.; Malagoli, M.; Marshall, M. S.; Sherrill, C. D. *J. Chem. Phys.* **2010**, *132*, 144104.
(67) Sponer, J.; Kypr, J. *Int. J. Biol. Macromol.* **1994**, *16*, 3−6.
(68) Yildirim, I.; Stern, H. A.; Tubbs, J. D.; Kennedy, S. D.; Turner, D. H. *J. Phys. Chem. B* **2011**, *115*, 9261−9270.
(69) Besseova, I.; Banas, P.; Kuhrova, P.; Kosinova, P.; Otyepka, M.; Sponer, J. *J. Phys. Chem. B* **2012**, *116*, 9899−9916.
(70) Chen, A. A.; Draper, D. E.; Pappu, R. V. *J. Mol. Biol.* **2009**, *390*, 805−819.
(71) Krasovska, M. V.; Sefcikova, J.; Reblova, K.; Schneider, B.; Walter, N. G.; Sponer, J. *Biophys. J.* **2006**, *91*, 626−638.
(72) Bowman, J. C.; Lenz, T. K.; Hud, N. V.; Williams, L. D. *Curr. Opin. Struct. Biol.* **2012**, *22*, 262−272.