Free Energy Simulations of Uncatalyzed DNA Replication Fidelity: Structure and Stability of T·G and dTTP·G Terminal DNA Mismatches Flanked by a Single Dangling Nucleotide

Urban Bren,[†] Václav Martínek,[‡] and Jan Florián*

Department of Chemistry, Loyola University Chicago, Chicago, Illinois 60660, USA

Received: January 15, 2006; In Final Form: February 24, 2006

A reference system for DNA replication fidelity was studied by free energy perturbation (FEP) and linear interaction energy (LIE) methods. The studied system included a hydrated duplex DNA with the 5′-CG dangling end of the templating strand, and $dCTP^{4-}\cdot Mg^{2+}$ or $dTTP^{4-}\cdot Mg^{2+}$ inserted opposite the dangling G to form a correct (i.e., Watson−Crick) or incorrect (i.e., wobble) base pair, respectively. The average distance between the 3′-terminal oxygen of the primer strand and the α-phosphorus of dNTP was found to be 0.2 Å shorter for the correct base pair than for the incorrect base pair. Binding of the incorrect dNTP was found to be disfavored by 0.4 kcal/mol relative to the correct dNTP. We estimated that improved binding and more near-attack configurations sampled by the correct base pair should translate in aqueous solution and in the absence of DNA polymerase into a six times faster rate for the incorporation of the correct dNTP into DNA. The accuracy of the calculated binding free energy difference was verified by examining the relative free energy for melting duplex DNA containing GC and GT terminal base pairs flanked by a 5′ dangling C. The calculated LIE and FEP free energies of 1.7 and 1.1 kcal/mol, respectively, compared favorably with the experimental estimate of 1.4 kcal/mol obtained using the nearest neighbor parameters. To decompose the calculated free energies into additive electrostatic and van der Waals contributions and to provide a set of rigorous theoretical data for the parametrization of the LIE method, we suggested a variant of the FEP approach, for which we coined a binding-relevant free energy (BRFE) acronym. BRFE approach is characterized by its unique perturbation pathway and by its exclusion of the intramolecular energy of a rigid part of the ligand from the total potential energy.

1. Introduction

Accurate prediction of ligand binding free energies (ΔG_{bind}) and a decomposition of these energies into components are important for understanding and manipulating enzyme catalysis.[1−4] Presently, free energy perturbation (FEP) calculations using energies sampled on molecular dynamics (MD) trajectories represent the most robust methodology for ΔG_{bind} calculations.[5] However, this approach is practically applicable only for calculating free energy differences between structurally similar ligands. This limitation prevents successful application of the FEP methodology to important biochemical systems. For example, DNA replication fidelity is partially determined by relative binding affinities of the four natural deoxynucleoside triphosphates (dNTPs) to a DNA polymerase−DNA complex.[6,7] Here, the large structural change between pyrimidine and purine containing nucleotides makes it more convenient to calculate relative ΔG_{bind} using less rigorous linear interaction energy (LIE) [8,9] or linear response approximation (LRA)[10,11] calculations rather than FEP. Moreover, free energy components obtained using LRA and LIE methods have the advantage of being additive. On the other hand, the accuracy of the LRA and LIE methods is limited by the need to use empirical parameters that may differ for structurally dissimilar ligands and/or for different host molecules.[12]

To bridge the gap between the FEP and LRA or LIE methods, we propose in this paper a modified FEP approach called binding-relevant free energy (BRFE), which includes the definition of a unique path for which the total free energy can be accurately and meaningfully dissected into its components. The dissection of ΔG_{bind} into additive contributions originating from groups of atoms or force field terms has the potential to provide free energy-based relationships between the structure and biological activity of biomolecules. This concept has been addressed [13−19] but deserves further attention. In light of the promise of the BRFE method to provide rigorous separation of ΔG_{bind} into the van der Waals and electrostatic components, this method appears to be an ideal source of benchmark data for adjusting empirical constants α and β in the LIE and LRA methods.

Theoretical and practical aspects of the BRFE approach are examined in this paper by calculating the relative stability of a duplex DNA containing terminal Watson−Crick and non-Watson−Crick base pairs, and a 5′ dangling nucleoside attached to the templating base (Figure 1, model A). Biochemically, this structure corresponds to the product of the DNA polymerization reaction in the absence of the DNA polymerase. Thus, the relative stability of the Watson−Crick and mismatched base pairs in this "product" structure is conceptually important for understanding the proofreading selectivity of DNA polymerases.

We also used free energy simulations to study deoxyribonucleoside triphosphate (dNTP) binding in a model system that was derived from the crystal structure of the ternary complex of a DNA, dNTP, and DNA polymerase β,[20] the protein part of which was removed (Figure 1, model B). The relative stabilities of the guanine−cytosine and guanine−thymine base pairs calculated for this model system provide an important reference point for the examination of the contribution of the DNA

* To whom correspondence should be addressed.

[†] Permanent address: National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana, Slovenia.

[‡] Permanent address: Charles University in Prague, Faculty of Science, Department of Biochemistry, Praque, Czech Republic.

10.1021/jp060292b CCC: $33.50
Published on Web 05/05/2006

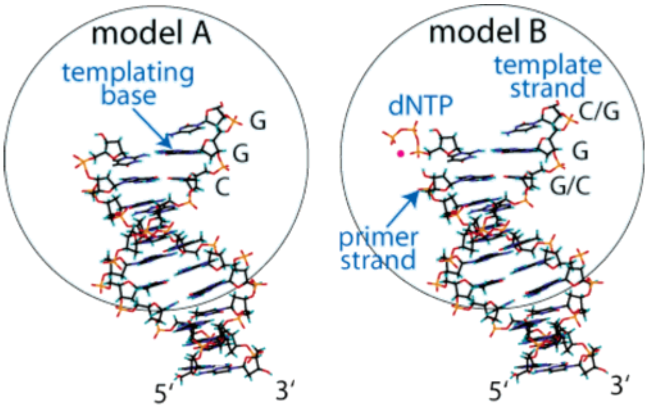

Figure 1. Schematic structure of the model systems used to calculate relative free energy for binding of 3′ terminal primer nucleotides C and T in duplex DNA opposite a G nucleotide in the template strand (model A), and for binding of dCTP^{4-}·Mg^{2+} and dTTP^{4-}·Mg^{2+} complexes opposite G in the template strand (model B). Studied base sequences are indicated by C or G letters printed alongside the template strand. DNA base pairs are of the Watson−Crick type except for the base pair involving the templating base, which is either CG or TG. In both systems, the 5′ template end has one dangling nucleoside. The blue sphere indicates water droplet (with sodium counterions) that solvates DNA. The position of the magnesium ion is indicated by a purple dot.

polymerase β to the fidelity of DNA replication. The current consensus is that, in the absence of DNA polymerase, the correct base pair is preferred by 0.2−4 kcal/mol (see, e.g., ref 21 and references therein).

To assess relative dNTP binding, we carried out extensive FEP calculations that included the examination of various perturbation paths for the dTTP → dCTP and dCTP → dTTP mutations. Nevertheless, LIE calculations provide a more stable approach for determining the relative free energies for this reference system. This is because more rigorous FEP, BRFE, and LRA calculations are hampered by a large statistical uncertainty due to the diffusion of the base moiety of dNTPs away from DNA. This diffusion occurs as an artifact during the simulation of the uncharged ligand state, which is an essential part of these methods. To further improve the accuracy of the LIE method, we adjusted its empirical parameters by comparing binding free energies for model A calculated at the LIE, BRFE, and FEP levels. The FEP method has been shown to provide stable results for internal DNA mismatches [22,23] and, therefore, it can be expected to perform well for terminal mismatches of model A. The adjusted LIE empirical parameters of model A should be rigorously transferable to model B as both models contain identical base sequences.

2. Methods

2.1. BRFE and the Additivity of the Free Energy. To calculate the binding free energy of a group of structurally similar ligands to a protein (P)[24] in water (W), we divide each ligand into two regions. The first region (referred to as tail or T) is the same for all ligands. The second region (referred to as head or H) is unique for each ligand. The absolute free energy of binding ΔG_{bind} can be calculated using the thermodynamic cycle depicted in Figure 2:

$$\Delta G_{bind} = \Delta G_{wat}^{nb} - \Delta G_{prot}^{nb} + \Delta G_{bind}^{T} \qquad (1)$$

where ΔG_{wat}^{nb} and ΔG_{prot}^{nb} are the free energy differences accompanying the processes of annihilation of the head region in water and in the protein, respectively. ΔG_{bind}^{T} denotes the free-energy of binding of the tail region to the protein binding site. Like in every FEP calculation, only the free energy difference resulting from the interactions with the perturbed region (H) is taken into account. Hence, during the annihilation in water, all the free energy difference comes from head−head (HH), headwater (HW), and head−tail (HT) interactions. Supposing the additivity of the free energy contributions[19] we can write the following:

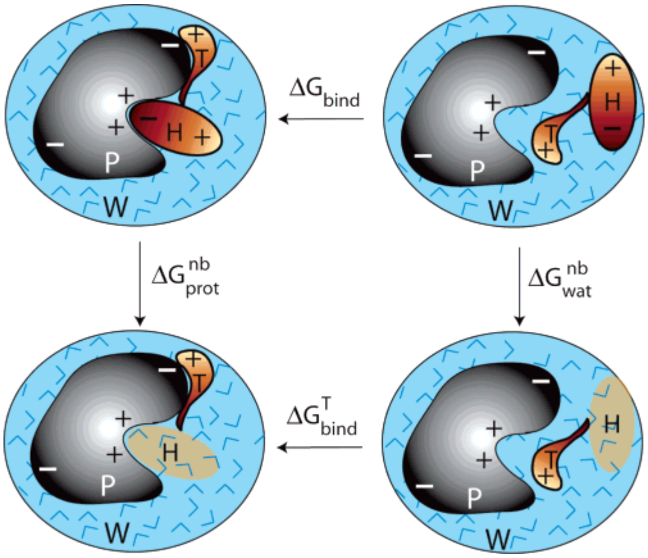

Figure 2. Thermodynamic cycle for the calculation of the binding free energy (ΔG_{bind}) from the free energies for annihilating nonbonded interactions of a rigid head moiety (H) in a protein (P) and water (W), and binding free energy of the tail (T) moiety. In the context of model A (Figure 1), P represents templating strand of DNA, H is the nucleobase which pairs with the templating base, and T is the remaining part of the primer strand. In the context of model B, P represents duplex DNA, H is the nucleobase moiety of dNTP, and T is the deoxyribose triphosphate·Mg^{2+} complex.

$$\Delta G_{wat}^{nb} = \Delta G_{wat}^{HH} + \Delta G_{wat}^{HW} + \Delta G_{wat}^{HT} \qquad (2)$$

The same procedure in the protein gives the following:

$$\Delta G_{prot}^{nb} = \Delta G_{prot}^{HH} + \Delta G_{prot}^{HW} + \Delta G_{prot}^{HP} + \Delta G_{prot}^{HT} \qquad (3)$$

If the head region is rigid, it should sample very similar configurations while being annihilated in water and in the protein, i.e., $\Delta G_{wat}^{HH} = \Delta G_{prot}^{HH}$. Using this reasonable assumption and eq 1−3 we obtain the following:

$$\Delta G_{bind} = (\Delta G_{wat}^{HW} + \Delta G_{wat}^{HT}) - (\Delta G_{prot}^{HP} + \Delta G_{prot}^{HW} + \Delta G_{prot}^{HT}) + \Delta G_{bind}^{T} = BRFE_{wat} - BRFE_{prot} + \Delta G_{bind}^{T} \qquad (4)$$

where BRFE is defined as a free energy difference originating from the interactions of the head region with its surroundings. This quantity presents a direct measure of the affinity of the perturbed region for its environment.

If we are interested in the relative free energy of binding $\Delta\Delta G_{bind}$ of two ligands consisting of the same tail and different heads (H1 and H2) to the same protein, we can write the following:

$$\Delta\Delta G_{bind} = \Delta G_{bind}^{H1} - \Delta G_{bind}^{H2} = BRFE_{wat}^{H1} + BRFE_{prot}^{H2} - BRFE_{prot}^{H1} - BRFE_{wat}^{H2} \qquad (5)$$

The binding free energy of the tail ΔG^{T}_{bind} cancels out, because the tail is the same in both ligands. Previously, this assumption has been applied in the framework of LRA and LIE methods to calculate $\Delta\Delta G_{bind}$ of dNTP substrates to the T7 DNA polymerase,[7,25] and to predict mutation effects on ΔG_{bind} for protein−protein binding.[26,27]

In a special case of non-existing tail, ΔG^{HT}_{wat}, ΔG^{HT}_{prot} and ΔG^{T}_{bind} are equal to zero, and the BRFE approach described above is reduced to

$$\Delta G_{bind} = \Delta G^{HW}_{wat} - \Delta G^{HP}_{prot} - \Delta G^{HW}_{prot} = \mathrm{BRFE}_{wat} - \mathrm{BRFE}_{prot} \quad (6)$$

This expression represents the absolute binding free energy of a rigid ligand.

Free energy is a state function. Therefore, the free energy difference depends only on the initial and final states and not on the choice of path connecting these states. On the other hand, free energy components are not state functions as only their sum is a state function. Hence, their differences are path dependent. This fact underscores the demand for finding the best possible path. In the BRFE method, we attain the annihilation of the head by simultaneously downscaling its partial atomic charges and van der Waals interactions by the same factor. This procedure ensures the natural behavior of the system by retaining a constant ratio between individual interactions of the head region with its surroundings (and also between individual interactions within the head region) along the whole annihilation path. Thus, the unique annihilation path in BRFE provides the best assessment of the role that the specific interactions play in the affinity of the unperturbed ligand to its environment.

2.2. Computational Details. *Starting Structure.* The simulations of $\mathrm{dNTP^{4-}{\cdot}Mg^{2+}{\cdot}DNA}$ complexes in the configuration depicted in Figure 1 (model B) were initiated using the crystal structure of a human DNA polymerase β ternary complex (1BPY).[20] This structure contained a nucleotide sequence $\begin{smallmatrix}5'\mathrm{C\mathbf{G}GCGCATCAGC}\\ 3'{-}\mathrm{\mathbf{N}CGCGTAGTCG}\end{smallmatrix}$ (sequence 1). Bold letters in these sequences denote the studied base pair, which consisted of the templating base (G) and the base of the dNTP substrate (N = C or T). All protein, water, sodium atoms, and the catalytic Mg^{2+} atom present in the crystal were removed. The structural Mg^{2+} ion, which was ligated by three oxygen atoms of $\alpha-$, $\beta-$, and γ-phosphates of dNTP, was retained. We also removed all the DNA atoms upstream from the templating base, except for a single overhanging nucleotide at the 5′ end of the template strand (Figure 1). To examine the sequence dependence of our results, the calculations were also carried out for the $\begin{smallmatrix}5'\mathrm{G\mathbf{G}CGCATCAGC}\\ 3'{-}\mathrm{\mathbf{N}GCGTAGTCG}\end{smallmatrix}$ sequence (sequence 2). This sequence was also used for simulations of model A (Figure 1).

General Simulation Conditions. The configurational ensembles for the evaluation of free energies were generated from molecular dynamics (MD) trajectories using the AMBER force field[28] implemented in the program Q.[29] Force field parameters that were used for deoxyribonucleoside triphosphates and Mg^{2+} ions were reported previously.[11,30]

The simulated solute molecules were immersed in a sphere (24 Å radius) of TIP3P water molecules subjected to the surface-constraint all-atom solvent (SCAAS) type boundary conditions.[29,31,32] These constraints were designed to mimic infinite aqueous solution. DNA atoms protruding beyond the sphere boundaries were restrained to their coordinates in the crystal structure using harmonic restraints. Nonbonding interactions of these atoms were turned off. Nonbonding interactions between atoms inside the simulation sphere were subjected to a 10 Å cutoff. The local-reaction field (LRF) method[32,33] was used to treat long-range electrostatic interactions for distances beyond a 10 Å cutoff. Non-bonding interactions of atoms forming the mutated nucleobase were explicitly evaluated for all distances. All simulated systems were equilibrated by a series of MD simulations that included gradual heating of the system to the final temperature of 298K and a total simulation time of 145 ps. All production trajectories generated constant-temperature ensembles at 298K. The SHAKE algorithm[34] was used for bonds involving hydrogen atoms. The structure and trajectory analyses were carried out using the program VMD 1.8.3.[35]

Sodium cations were added to the simulated system to achieve electroneutrality of the system. Positions of Na^{+} atoms were restrained by a flat-bottom harmonic potential (force constant of 50 kcal mol^{-1} $Å^{-2}$) that was zero for the distances less than 20 Å from the center of the simulation sphere. These potentials were applied to prevent the diffusion of the sodium ions towards the edge of the simulation sphere. The center of the simulation sphere was positioned in the center of the studied base pair. This selection placed the studied base pair plus another five base pairs downstream inside the simulation sphere. Phosphate groups that were closer than 3 Å from the sphere boundary were made electroneutral by decreasing the negative charge on the oxygen atoms as described previously.[22]

The simulations of the reference system in water in the absence of DNA were carried out for the $\mathrm{dCTP^{4-}{\cdot}Mg^{2+}}$ and $\mathrm{dTTP^{4-}{\cdot}Mg^{2+}}$ complexes with the total charge of these complexes neutralized by either another Mg^{2+} ion or by two Na^{+} ions. Simulations in water were also carried out for cytidine and thymidine nucleosides. In all simulations of dNTP or nucleosides in water, C1′ atom of the solute was constrained in the center of the simulation sphere using a harmonic potential defined by a force constant of 50 kcal mol^{-1} $Å^{-2}$.

FEP Calculations. The single-topology FEP calculations[5,36,37] were carried out for the trajectories, in which the charges and atom types (van der Waals parameters) of the thymine moiety of the substrate were slowly changed into cytosine in the forward trajectory, and vice versa in the reverse trajectory. For the model A (Figure 1), the C → T mutation was carried out for the base embedded in the duplex DNA ($\Delta G^{DNA}_{C\rightarrow T}$), and for the base moiety of a nucleoside placed in a water droplet ($\Delta G^{w}_{C\rightarrow T}$). Considering the thermodynamic cycle for the DNA melting,[22] the relative free energy

$$\Delta\Delta G_{C\rightarrow T} = \Delta G^{DNA}_{C\rightarrow T} - \Delta G^{w}_{C\rightarrow T} \quad (7)$$

expresses the destabilization of the duplex DNA by the substitution of the C base by the T base. Because C forms a Watson−Crick base pair with the templating G, this mutation should lead to DNA destabilization, i.e., the calculated $\Delta\Delta G_{C\rightarrow T}$ should be a positive number. Similarly, ΔG calculated for the C → T mutation of the base moiety of the $\mathrm{dNTP^{4-}{\cdot}Mg^{2+}}$ complex in DNA and in water in eq 7 provided the difference between the binding free energy of dCTP and dTTP (Figure 1, model B).

Several mutational pathways were tested. In order to generate these pathways, each forward (and reverse) trajectory was divided into two segments that ended and initiated in a common hybrid structure that lies approximately half-way between the thymine and cytosine structure. The four mutational pathways generated in this way (Paths I − IV) are defined in Table 1 and Figure 3.

Each segment was subdivided into 51 windows that differed in the value of the coupling parameter λ. The free energy

TABLE 1: Atomic Charges and Atom Types of the TC-hybrid Intermediate and Charges of the Cytosine and Thymine Moieties used in the FEP Calculations[a]

	TC-hybrid intermediate				cytosine	thymine
atom	path I	path II	path III	path IV	(a.u.)	(a.u.)
N1	C/C	C/C	C/C	C/C	−0.0339	−0.0239
C6	C/C	C/C	C/C	C/C	−0.0183	−0.2209
H6	C/C	C/C	C/C	C/C	0.2293	0.2607
C5	C/C	C/C	C/C	C/C	−0.5222	0.0025
C5M[b](H5)	C/C	C/C	C/C	C/C	0.1863	−0.2269
H5M[b]	0.0/T	0.0/T	0.0/T	0.0/T		0.0770
C4	C/C	C/C	C/C	C/C	0.8439	0.5194
O4(N4)	C/C	T/T	T/T	T/T	−0.9773	−0.5563
H41	0.0/C	0.0/C	0.0/C	C/C	0.4314	
H42	0.0/C	0.0/C	0.0/C	0.0/C	0.4314	
N3	C/C	C/C	T/C	C/C	−0.7748	−0.4340
H3	0.0/T	0.0/T	0.0/T	0.0/X		0.3420
C2	C/C	C/C	C/C	C/C	0.7959	0.5677
O2	C/C	C/C	T/C	T/C	−0.6548	−0.5881
total charge	−0.926	−0.505	−0.097	−0.007	−0.0631	−0.1268

[a] Atomic charges/atom types. C: the charge or atom type was taken from the corresponding atom in cytosine. T: the charge or atom type was taken from the corresponding atom in thymine. 0.0: atomic charge equals to zero. X: atom with zero van der Waals parameters. For atom numbering, see Figure 3. [b] C5M and H5M are methyl group atoms. The three H5M atoms were assigned the same charges and atoms types.

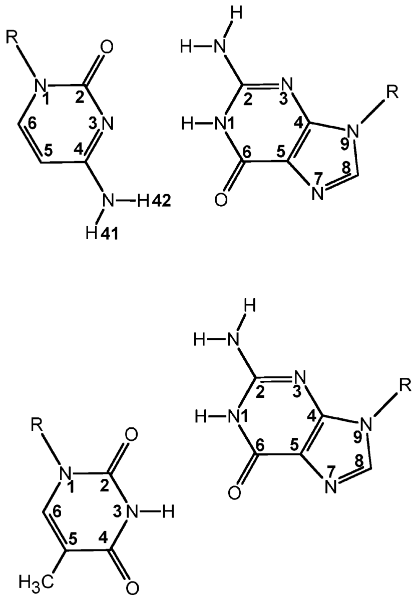

Figure 3. Atom numbering of the CG (top) and TG (bottom) base pairs. Hydrogen atoms bonded to the sp^2 carbon atoms are not shown.

difference for the *i*th window, $\Delta G_{i\to i+1,}$ was evaluated as[38]

$$\Delta G_{i\to i+1} = -\beta^{-1}\ln\langle \exp(-\beta\Delta U)\rangle_i \qquad (8)$$

where $\Delta U = U_{i+1} - U_i$ represents the potential energy difference between the states characterized by coupling parameters λ_{i+1} and λ_i. $\beta^{-1} = k_B T$, where k_B stands for the Boltzmann constant, and T is the thermodynamic temperature. Notation $\langle\cdots\rangle_i$ indicates averaging over the ensemble of configurations generated by a MD simulation on the potential energy surface of the state i. The total free energy change, ΔG, was calculated as a sum of the free energy differences over the 51 windows. The average ΔG was evaluated by averaging over the results calculated for the forward and reverse trajectories. The error in the calculated ΔG was determined as one half of the difference between the free energies calculated from the forward and reverse trajectories. Integration step size of 1 fs was used for all trajectories. The total simulation length of each segment varied, depending on the studied system, between 0.5 and 2 ns.

The energy of the system was sampled every 10th step. The first 10 energies were discarded for each window before the free energy change was calculated. In addition, for each window the electrostatic($\Delta G^{ES}_{i\to i+1}$) and van der Waals ($\Delta G^{vdW}_{i\to i+1}$) contributions to the free energy difference were evaluated by inserting only the electrostatic (ΔU^{ES}) and van der Waals (ΔU^{vdW}) portions of the potential energy difference in eq 8, respectively. The electrostatic *(*ΔG^{ES}) and van der Waals *(*ΔG^{vdW}) components of the total free energy change were then obtained as a sum over the 51 windows.

BRFE Calculations. The BRFE calculations were performed only for duplex DNA (Figure 1, model A). The single-topology FEP calculations [5,36,37] were carried out for the trajectories, in which the charges and the depth of the van der Waals potential function of the atoms forming the thymine or cytosine moiety at the 3′ terminus of the primer strand were simultaneously changed to zero. Note that the van der Waals radii and the bonded parameters of all mutated atoms were retained at their original values through the whole simulation. That is, each base was mutated to “nothing” by allowing it to become gradually “transparent” rather than by shrinking its size. The free energies for the mutation of a base to “nothing” in the DNA environment will be denoted BRFE^{B}_{DNA}. Identical calculations were carried out for the base moiety of deoxycytidine and thymidine in a water droplet, yielding a free energy BRFE^{B}_{w}, where the subscript B denotes either C or T. These free energy differences present a direct measure of the affinity of the nucleobase moiety towards its respective DNA or water environment. Therefore, using the thermodynamic cycle of Figure 2 (with the vertical free energy differences equal to BRFE^{B}_{DNA} and BRFE^{B}_{w}), the DNA destabilization due to the creation of an abasic site can be calculated as

$$\Delta G_{B\to 0} = \mathrm{BRFE}^{B}_{DNA} - \mathrm{BRFE}^{B}_{w} \qquad (9)$$

where B can be either C or T. The results of BRFE calculations can be compared to the results obtained by the FEP mutation of T to C (eq. 7) using the expression

$$\Delta\Delta G_{C\to T} = \Delta G_{C\to 0} - \Delta G_{T\to 0} \qquad (10)$$

The trajectories used to evaluate BRFE free energies consisted of 101 windows. The BRFE was calculated for each window by inserting in eq 8 only ΔU originating from interactions of the annihilated nucleobase with its surroundings. These free energy differences were then summarized over all 101 windows. The total length of each trajectory was 2 ns. The integration step was 1 fs. In order to prevent the dissociation of the uncharged substrate from the binding pocket, we applied a weak flat bottom distance restraint between the template and substrate bases. The magnitude of the distance for which this restraint started to appear was varied in a series of separate calculations. The reported free energies and standard deviations reflect an average over various constraint distances.

LIE Calculations. The LIE[8] calculations were carried out for molecular dynamics trajectories, the length of which varied between 0.5 and 2 ns depending on the studied system. A 2 fs integration step was used. Energies were sampled every 10 steps. The “probe” region of the substrate, i.e., the collection of atoms for which the average electrostatic ($\langle U_{ES}\rangle$) and van der Waals ($\langle U_{vdW}\rangle$) interaction energies with the rest of the system were explicitly evaluated, represents the LIE analog of the head region

TABLE 2: FEP Path Dependence of the Change in the Free Energy (kcal/mol) of a DNA Duplex and a dCTP·DNA Complex Due to the Formation of the Terminal TG Mispair[a]

	model A			model B[c]		
FEP pathway[b]	$\Delta\Delta G^{ES}_{C\to T}$	$\Delta\Delta G^{vdW}_{C\to T}$	$\Delta\Delta G_{C\to T}$	$\Delta\Delta G^{ES}_{C\to T}$	$\Delta\Delta G^{vdW}_{C\to T}$	$\Delta\Delta G_{C\to T}$
path I	1.0 ± 2.3, *M* = *2*	−0.5 ± 0.1, *M* = *2*	0.5±2.2, *M* = *2*	0.4 ± 0.2, *M* = *2*	0.3 ± 0.3, *M* = *2*	0.7 ± 0.4, *M* = *2*
path II				−0.1 ± 1.2, *M* = *4*	−0.2 ± 0.1, *M* = *4*	−0.3 ± 1.0, *M* = *4*
path III	2.4 ± 0.5, *M* = *2*	−0.6 ± 0.3, *M* = *2*	1.8 ± 0.6, *M* = *2*	−0.4 ± 1.2, *M* = *4*	−0.3 ± 0.1, *M* = *4*	−0.7 ± 1.2, *M* = *4*
path IV				1.5 ± 1.2, *M* = *6*	−0.4 ± 0.4, *M* = *6*	1.0 ± 1.1, *M* = *6*
all paths	1.7 ± 1.5, *M* = *4*	−0.6 ± 0.1, *M* = *4*	1.1± 1.5, *M* = *4*	0.5 ± 1.3, *M* = *16*	−0.3± 0.3, *M* = *16*	0.2 ±1.3, *M* = *16*

[a] *M* denotes a number of independent simulations used to determine the average free energy and its standard deviation. Simulation time of each independent simulation was 1 and 2 ns for models A and B, respectively. For detailed description of model systems see Figure 1 and Methods section. [b] Table 1 [c] 5′CGG sequence of the "template" strand.

in BRFE. We used a reference system consisting of deoxycytidine and thymidine in aqueous solution and defined their nucleobase portion as the probe region for model A. The sum of $\langle U_{ES}\rangle$ scaled by an empirical factor β and $\langle U_{vdW}\rangle$ scaled by an empirical factor α presents a direct measure of affinity of the probe region for its surroundings [8]. Therefore, using a thermodynamic cycle depicted in Figure 2 for model A, the DNA destabilization due to the creation of an abasic site can be calculated as

$$\Delta G_{B\to 0} = -\alpha(\langle U^{p}_{vdW}\rangle - \langle U^{w}_{vdW}\rangle) - \beta(\langle U^{p}_{ES}\rangle - \langle U^{w}_{ES}\rangle) \qquad (11)$$

where superscripts p and w denote the DNA(dN) + water + counterions and the reference system in aqueous solution, respectively. This equation presents a direct LIE analog of eq 9. Therefore, we determined the coefficients α and β by the least squares fit using the BRFE (and FEP) results for model A. The fitted data contained five independent parameters for two different DNA sequences. These adjusted empirical parameters were then reused for the model B where we defined the probe region as the nucleoside portion of the corresponding dNTP. The dividing line between the nucleoside and non-nucleoside parts coincided with the center of the C4′−C5′ bond. The two reference systems for model B consisted of $dCTP^{4-}\cdot Mg^{2+}$ and $dTTP^{4-}\cdot Mg^{2+}$ complexes in aqueous solution containing two Na^{+} ions. By combining eq 10 and 11, the results of LIE calculations can be compared to the results obtained by the FEP mutation of T to C (eq 7).

3. RESULTS

FEP calculations generally require long simulation times for the convergence of the calculated ΔG.[39] As a state function, the magnitude of ΔG should be independent of the mutation path that connects the two states, for which this ΔG is calculated. However, the convergence may occur faster for some pathways than for others. Furthermore, ΔG components may depend on the chosen pathway. In order to quantify these issues, we evaluated several different mutational pathways for the T → C mutation.

Our previous FEP simulations of the DNA destabilization by internal mismatches[22] involved a simple perturbation path in which the charges and van der Waals radii were changed concomitantly from one base to the other. However, this perturbation pathway allowed a non-zero charge to be present on atoms with nearly zero van der Waals radii. For example, this situation may occur in the initial part of the T → C mutation when the amino group hydrogens start to appear. Such hydrogens could get in a very close distance to the hydrogen atoms of TIP3P water molecules that have also zero van der Waals radii. These interactions occasionally resulted in large instabilities if standard 1 fs integration step was used.

To avoid these difficulties, we investigated several variants of a two-stage perturbation protocol (Table 1). In the first step of this protocol, thymine was mutated to a TC hybrid in a single FEP simulation. The presence of the TC hybrid on the mutation path assured that the van der Waals radii of all atoms were fully created before the charges on these atoms started to grow. This hybrid was mutated in a separate FEP simulation to cytosine, and the free energies calculated from these two FEP calculations were added. The reverse pathway led from cytosine to TC to thymine.

In the path I, all thymine charges were mutated to their cytosine counterparts in the first step except for the hydrogen atoms of the amino group that were grown in the uncharged state. However, this pathway allowed a large negative charge on the N4 atom (Figure 3) to fully develop in the TC hybrid without being compensated by positive charges on the amino group hydrogens. Consequently, the amino group of the TC hybrid showed strong interactions with a sodium cation present in solution. These interactions made the calculated results in DNA duplex unstable (Table 2).

Although strong interactions of the negatively charged NH_2 group with Na^{+} could have been averaged out using very long simulation trajectories, we decided to avoid these instabilities by modifying the mutation pathway. This modification was accomplished by simultaneously growing the negative charge from −0.56 in O4 to −0.98 in N4 and positive charges on the amino group hydrogens in the TC → C segment of the path II (Table 1). However, results calculated for pathway II still showed a significant dependence on starting conditions (Table 2). We attributed this instability to the large variation of the total charge of the mutated base along the perturbation path: Namely, whereas initial (thymine) and final (cytosine) states had a zero total charge, the charge of the TC hybrid had overall charge of −0.5 in path II.

To avoid creating this unnatural net negative charge along the pathway, we delayed the mutation of the charges of N3 and O2 atoms to the TC → C segment (Table I, path III). This pathway provided stable results for the match−mismatch free-energy difference in model A, but in model B we were faced with the dissociation of dNTP from DNA in the simulation stage near the TC hybrid (Table 2). We believe that this instability was due to an absence of hydrogen bonding interactions between the templating base and the TC hybrid. Therefore, in path IV we maintained at least one hydrogen bond during the entire trajectory leading from dCTP to dTTP (see Table 1 and Supporting Information Movie 1). However, an improved visual appearance of the intermediate structures sampled along path IV did not translate into significantly smaller statistical fluctuation of the ΔG values calculated using this path.

Overall, none of the four pathways examined here significantly outperformed the other pathways. This means that in our

TABLE 3: Nucleoside Contributions to the DNA Stabilization Free Energy (kcal/mol) Calculated Using the BRFE Method for Model A

	3′−TG 5′GGC	3′−CG 5′GGC
$\Delta G_{B\to 0}$[a]	8.5 ± 1.1	10.3 ± 1.9
$\Delta G_{B\to 0}^{ES}$	2.2 ± 0.4	5.5 ± 1.1
$\Delta G_{B\to 0}^{vdW}$	6.3 ± 0.7	4.8 ± 1.0
ΔG_{nad}	0.0	0.0

[a] Equation 9: the base that was mutated to "nothing" is positioned at the 3′ end of the top (primer) strand (see also Figure 1). Free energies and their components were calculated as an average over 7 separate 2 ns calculations (i.e., 14 ns total simulation time). [b] Nonaditivity of the free energy for a single trajectory, $\Delta G_{nad} = \Delta G_{B\to 0} - (\Delta G_{B\to 0}^{ES} + \Delta G_{B\to 0}^{vdW})$. The electrostatic and van der Waals components were calculated by inserting the corresponding potential energy term in eq 8.

case a simple prescription for obtaining a reliable free energy difference is to sample as many configurations as possible rather than to search for a perfect path. In our case, it appears that sufficient sampling requires at least 20 ns of the simulation time. Averaging over the results of all paths, which totaled 32 ns for both models, yielded ΔG of 1.1 ± 1.5 and 0.2 ± 1.3 kcal/mol for the DNA duplex and dNTP·DNA systems, respectively. Although the standard deviations of these energies are large, they provide the measure of energy fluctuations among different trajectories rather than the accuracy of the average free energies. This interpretation is consistent with the fact that the standard deviation increases after including more independent trajectories. On the other hand, the accuracy of the overall ΔG should improve when more independent trajectories are included because the added trajectories improve sampling of the configuration space. Thus, a realistic estimate of the actual error of the average ΔG values presented in the bottom row of Table 2 is ±0.7 kcal/mol.

In accordance with our results for the additivity of the solvation free energies of nucleoside phosphates [19], the van der Waals and electrostatic components are additive within each unique FEP pathway. There are variations in the relative magnitudes of the contribution of the van der Waals and electrostatic components among different pathways that appear to indicate that these contributions are non-additive. However, due to the significant statistical uncertainty of the total free-energy differences and their components, it is difficult to quantify the exact magnitude of the non-additivity effects related to small changes in the FEP path. At any rate, this non-additivity is small enough to allow us to conclude that the C·G → T·G transition is favored by the van der Waals component, but it is disfavored by electrostatic interactions that dominate the total free energy difference.

The free energies for the annihilation of the cytosine and thymine bases using the BRFE variant of FEP calculations are presented in Table 3. The decomposition of these energies into the electrostatic and van der Waals components was found to be additive (within the accuracy of 0.1 kcal/mol). By subtracting the free energies in the two columns of Table 3 (eq 10) we obtain $\Delta\Delta G_{bind}$ for the C·G → T·G transition of 1.8 kcal/mol. The electrostatic and van der Waals components of this free-energy are 3.3 and −1.5 kcal/mol, respectively (data not shown in Table 3). Each of these components is significantly larger than its FEP counterpart (Table 2, model A). This result is not surprising in view of the large difference between FEP pathways used for the calculations presented in Tables 2 and 3. However, some qualitative results, such as signs of the electrostatic and van der Waals components or the dominance of the electrostatic term, remain path independent even for very large path variations.

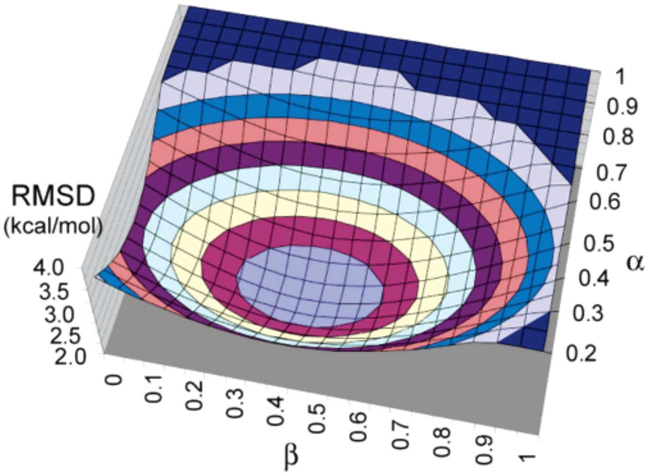

Figure 4. Root-mean-square deviation surface used to find optimal values of the α and β parameters for LIE calculations of nucleotide binding in duplex DNA.

The BRFE method has two features that make it an ideal source of benchmark data for fitting parameters of approximate binding methods such as LIE or LRA. First, the ratio of the average electrostatic and van der Waals energies of the "head" is kept constant along the whole perturbation pathway. Second, BRFE calculations exclude the contribution of head−head interactions from the computed free energies and focus only on the interactions of the head region with its surroundings. These features ensure that the BRFE decomposition of the calculated free energies into electrostatic and van der Waals components is consistent with the decomposition by the LRA and/or LIE methods. To take advantage of the compatibility between the BRFE and LIE decompositions, we generated a RMSD function that contained differences between the LIE energies expressed as a function of parameters α and β (eq 11) and five target energies. The target energies included four BRFE electrostatic and van der Waals free energies (Table 3), and $\Delta\Delta G_{C\to T}$ of 1.1 kcal/mol calculated using the FEP method (Table 2, model A). The resulting function shows a minimum for $\alpha = 0.45$ and $\beta = 0.43$ (Figure 4). These α and β parameters were reused in LIE calculations of model B (Table 4). Since the BRFE calculations for this system are impractical and the FEP results are rather unstable (Table 2), the most accurate $\Delta\Delta G_{C\to T}$ for model B should be obtainable by properly parametrized LIE method. To examine the stability and DNA sequence dependence of the LIE results for model B, the calculations were carried out for two different sequences of the "primer" strand and for two different starting conditions. The comparison of the resulting energies (Table 4) shows that the LIE results are stable. The van der Waals contributions to ΔG_{bind} for dCTP or dTTP demonstrate significant sequence dependency (Table 4) but the magnitudes of $\Delta\Delta G_{C\to T}$ of 0.4 kcal/mol are sequence independent (Table 5).

The energetic preference for the formation of the GC base pair increases upon going from model B to model A, i.e., when the triphosphate substituent that is not covalently connected to DNA is replaced by the DNA backbone. Using different computational methods, the formation of the Watson−Crick base pair in this arrangement is favored by 1.1 to 1.8 kcal/mol (Table 5).

Structurally, the dangling 5′-dC or 5′-dG in the simulations of models A and B are found to stack with the templating base

TABLE 4: Nucleoside Contributions to the dNTP Binding and DNA Stabilization Free Energy (kcal/mol) Calculated Using the LIE Method[a]

	model B				model A	
	TG 5′GGC	**C**G 5′GGC	**T**C 5′CGG	**C**C 5′CGG	3′−TG 5′GGC	3′−CG 5′GGC
$\langle U^{p}_{ES}\rangle - \langle U^{w}_{ES}\rangle$	−0.2 ± 0.2	−2.8 ± 2.1	−0.4 ± 0.8	−2.8	−4.2	−9.1
$\langle U^{p}_{vdW}\rangle - \langle U^{w}_{vdW}\rangle$	−11.8 ± 0.3	−10.1 ± 0.5	−10.8 ± 0.7	−9.0	−12.8	−11.9
$\Delta G_{B\rightarrow 0}$[c]	5.4 ± 0.1[b]	5.8 ± 0.7[b]	4.8 ± 0.1[b]	5.2	7.6	9.2

[a] Bold letters in the DNA sequences are used to denote nucleosides that are a part of the $dNTP^{4-}\cdot Mg^{2+}$ complex. Base-paired nucleotides are indicated by vertical alignment of the DNA sequences. [b] Energies reported for this system were obtained as an average of two independent 2 ns simulations that differed in their starting geometries. [c] Equation 11: $\alpha = 0.45$, $\beta = 0.43$

TABLE 5: Change in the Free Energy of a dCTP·DNA Complex and a DNA Duplex Due to the Formation of the Terminal TG Mispair[a]

	$\Delta\Delta G_{C\rightarrow T}$(kcal/mol)[b]		
	dNTP·DNA		DNA duplex
method	**N**G 5′GG	**N**C 5′CGG	3′**N**G 5′GGC
FEP[c]		0.2 ± 1.3	1.1 ± 1.5
BRFE[d]			1.8 ± 1.5
LIE[e]	0.4 ± 0.8	0.4 ± 0.6	1.7 ± 0.7
exp[f]			1.4

[a] The $dNTP^{4-}\cdot Mg^{2+}$ complex (model B) and C or T (model A) are denoted by a letter **N**. See also captions to Table 2. [b] Calculated using eq 7 and data from Table 2 (FEP), eq 10 and data from Table 4 (LIE), and eq 10 and data in Table 3 (BRFE). [c] Mean values and standard deviations were determined for data corresponding to 32 ns of total simulation time (Table 2). [d] Mean value and its standard deviation was determined from seven separate simulations that correspond to 14 ns of total simulation time (Table 3). [e] Mean value and its standard deviation was determined from four separate simulations that correspond to 2 ns of total simulation time (Table 4). [f] Free energy difference at 25 °C determined from the nearest neighbor parameters ΔG (5′GC/3′CG) = −2.24 kcal/mol, ΔH (5′GC/3′CG) = −9.8 kcal/mol, ΔG (5′GC/3′TG) = −0.92 kcal/mol, ΔH (5′GC/3′TG) = −5.9 kcal/mol determined for the correct and incorrect terminal DNA pairs at 37°C and 1M NaCl solution.[40,41] Conversion of the ΔΔG value from 37 to 25 °C was done by assuming that its ΔΔH and ΔΔS components are temperature independent in the 25−37 °C range.

for most of the time. This dominant configuration is shown in Figure 5. The stability of the stacked complex with 5′-dG nucleotide appears to be higher than with 5′-dC because we did not observe flipping of the base of 5′-dG out of the stacked geometry. The cytosine moiety of the dangling 5′-dC was observed to become unstacked during the simulation, which was started from the stacked complex, but the stacked configuration was reformed within 1 ns. This folding process was facilitated by the formation of a T-shaped complex between the two terminal bases (see Supporting Information Movie 2).

Further analyses of the MD trajectories show that the dCTP·G pair forms a regular Watson−Crick base pair whereas the dTTP·G mispair forms a wobble base pair (Figure 5). A small shift in the average position of the dTTP base, which is required to accommodate wobble base pairing, causes an increase of the average distance between the phosphorus atom of α-phosphate of dNTP and the oxygen atom of the 3′OH group of the preceding nucleotide ("primer") from 3.9 Å (in the case of dCTP) to 4.3 Å. Both the time-dependent variation (Figure 6) and the histogram (Figure 7) of the O3′−P_{α} distances show that the O3′−P_{α} distance oscillates between two stable values of about 3.8 and 5.2 Å. The distances around 5.2 Å are more abundant for the mispair. In these configurations, the 3′OH group of the primer nucleotide is hydrogen bonded with a nearby water molecule, and its ribose moiety has the O4′-endo conformation. In dCTP·G pair, the O3′ atom is the most frequently found 3.68 Å from the P_{α} atom. These configurations are characterized by the presence of the hydrogen bond between the 3′OH group and the bridging oxygen (O5′) of dCTP (Figure 5, top), and the C2′-endo ribose conformation. In contrast, the most populated configurations in DNA containing the dTTP·dG mispair have O3′−P_{α} distances around 3.86 Å (Figure 7). These configurations are accompanied by hydrogen bonding of the 3′OH group with one of the non-bridging oxygen atoms of the α-phosphate of dTTP (Figure 5, bottom), and the C2′-endo ribose conformation.

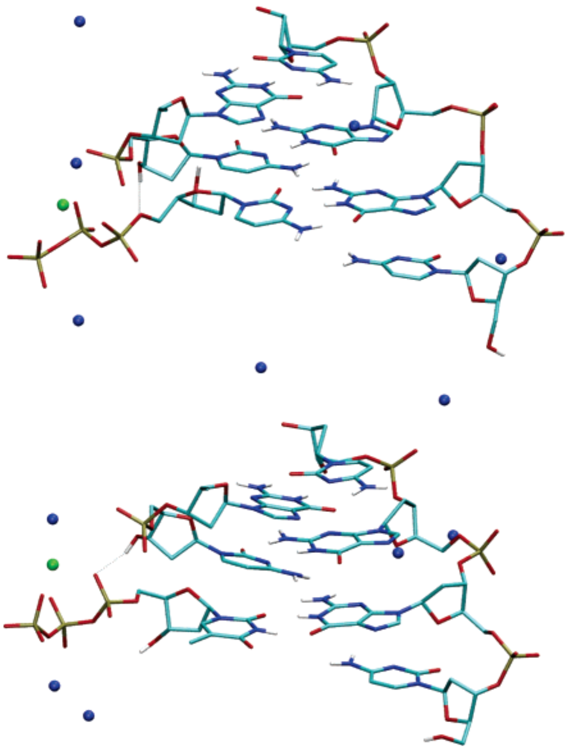

Figure 5. Representative snapshots of the dCTP·G-DNA (top) and dTTP·G-DNA (bottom) structures in aqueous solution (model B). Na^{+} and Mg^{2+} ions are rendered as blue and green spheres, respectively. Hydrogen bonds formed by the terminal 3′-OH group of the primer DNA strand are indicated by dotted lines. DNA helices beyond the three terminal base pairs, water molecules, and hydrogen atoms attached to carbon are not shown.

4. DISCUSSION

The accuracy of DNA replication is characterized by misinsertion frequency, $f_i = n(I)/n(C)$, where n(I) and n(C) represent number of incorrect and correct dNTP insertions during the DNA polymerization process.[42,43] The logarithm of this frequency is proportional to the difference between the free energies of the rate-limiting activated complexes containing the incorrect and correct dNTP ($\Delta\Delta g^{\ddagger}$). [44] If the protein confor-

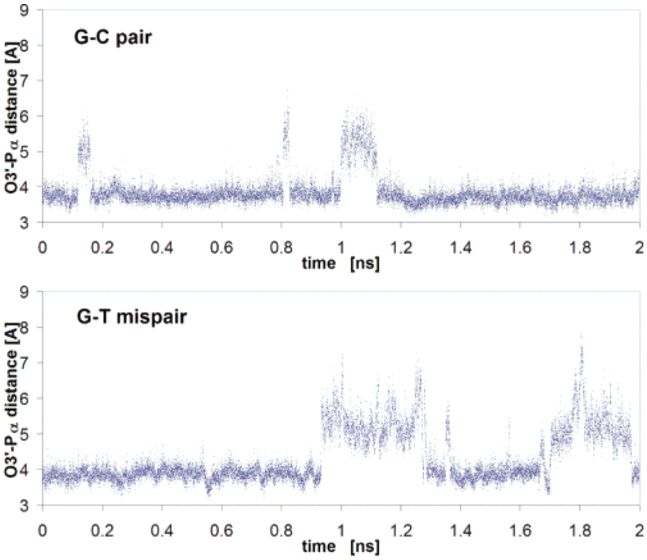

Figure 6. Distances between the terminal 3′O atom of the "primer" strand of DNA and the P_α atom of the bound dCTP (top) and dTTP (bottom) sampled during a 2 ns MD trajectory.

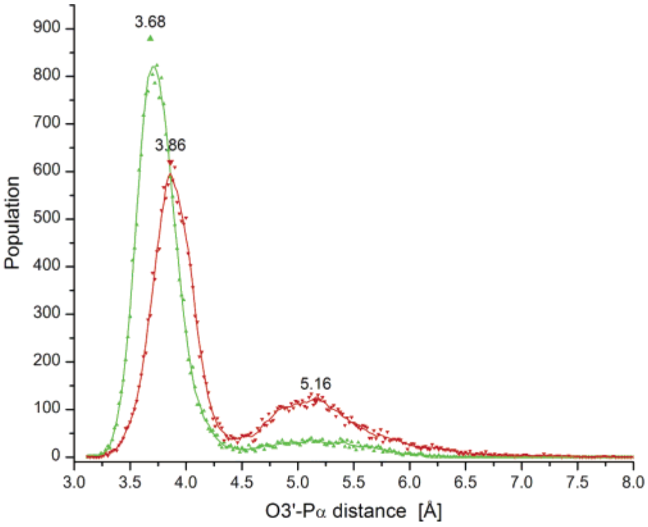

Figure 7. Distribution of the distances between the terminal 3′O atom of the "primer" strand of DNA and the P_α atom of the bound dCTP (green) and dTTP (red) that were sampled during a 2 ns MD trajectory.

mational change is not rate-limiting for dNTP insertion this free-energy difference can be divided into differences due to binding ($\Delta\Delta G_{bind}°$) and the chemical step ($\Delta\Delta g_{cat}^{\ddagger}$).[7] Experiments or calculations that would provide further partitioning into contributions from different amino acid residues of the DNA polymerase, the catalytic metal, and DNA would equate to full understanding of the DNA replication fidelity.

The contribution of DNA to the replication fidelity (in the absence of DNA polymerase) is pursued in this study using two methods from the arsenal of classical statistical mechanics, FEP and LIE.[39] These calculations focus on $\Delta\Delta G_{bind}°$ for dTTP versus dCTP insertion opposite dG in the template. It might be preferable to establish this binding selectivity experimentally, but dNTP binding to a predetermined DNA site would be extremely difficult to detect. Thus, free energy calculations represent the only practical option.

The main issue with calculations of free energy differences for complex systems, such as DNA in aqueous solution, is their limited accuracy. This issue is complicated by the difficulty in estimating the actual error of these results. Besides using two independent computational methods, we tried to diminish these issues by averaging over multiple perturbation pathways in the FEP calculations and over multiple starting conditions in the LIE calculations. Last but not least we "anchored" the calculated results by performing calculations for a structurally related system for which an experimental free energy estimate was available. By achieving a very good agreement between the theoretical and experimental free energy differences for model A (DNA duplex column in Table 5), we eliminated the possibility that free-energies calculated for system B would be affected by systematic errors due to imperfect force field parameters. This anchoring procedure left the configurational sampling as the most serious potential source of computational errors.

Of the two methods used to calculate relative dNTP binding free energy, the LIE method is less prone to be affected by sampling problems. For example, our LIE calculations were able to take into account all the configurations corresponding to folding and unfolding of the dangling end of DNA (Supporting Information Movie 2). However, the weakness of the LIE method is that it uses two empirical parameters that may depend on the studied system. We developed a version of the FEP protocol called BRFE that is especially suitable to provide data for determining LIE parameters. The LIE parameter for scaling electrostatic interaction ($\beta = 0.43$), which was derived using data from BRFE calculations for model A, is identical to the parameter recommended by Åqvist and coworkers for the use in binding calculations with the Gromos87 force field.[9] Although our parameter $\alpha = 0.45$ is significantly larger than the $\alpha = 0.18$ recommended by Åqvist et al,[9] this difference affects relative binding free energies much less than absolute binding free energies. That is, if the generic $\alpha = 0.18$ and $\beta = 0.43$ parameters were used, the LIE free energies presented in Table 5 would increase by 0.4 and 0.3 kcal/mol for models A and B, respectively, which is below the reported error range. This outcome along with the reported small dependence of LIE on the force field used,[45] appears to lessen the importance of the custom fit carried out in this work. However, our extra parametrization effort is warranted by significant structural differences between the protein−small ligand systems, for which the generic LIE parameters were developed,[9] and DNA systems.

Besides being an invaluable source for LIE parameters, the BRFE method has additional qualities that make it a promising new tool for binding simulations. First, BRFE is more useful than FEP for calculations of free energy changes due to large structural modification of a rigid moiety of the ligand. For instance, the BRFE method can facilitate a change from pyrimidine to purine nucleobase without the use of problematic dual topology. An additional feature of the BRFE method is that it eliminates errors originating from subtracting two large numbers of similar magnitudes by focusing only on the head region of the ligand and neglecting head-head interactions in a free and bound state. Thus, BRFE has the potential to provide differences in DNA melting free energies due to the replacement of one of the DNA nucleotides by the abasic site.[46] Note that complete utilizing this BRFE capability in the present study would require the use of a different reference system in water calculations of model A (for example, a dinucleotide). In order to fully focus on the replication fidelity issue we opted to use a nucleoside as our reference system in water because this reference system does not require such extensive sampling as dinucleotides, and thus, it is expected to provide more stable relative free-energies.

Second, the BRFE method allows us to decompose the free-energy into additive residue or force field contributions that are consistent with the decomposition obtained by LIE approximation and the decomposition observed by protein mutagenesis studies. Thus, this method has the potential to provide direct link to additive properties studied experimentally. Although we did not calculate the decomposition of our free energy differences into group contributions, we confirmed the additivity of the decomposition of BRFE free-energies into electrostatic and van der Waals components (Table 3).

The calculated binding contribution to the selectivity of DNA replication in the absence of the enzyme is 0.4 kcal/mol (Table 5, model B). This $\Delta\Delta G_{bind}^\circ$ agrees with the free energy preference of 0.2 to 0.4 kcal/mol observed for binding of the terminal AT base pair versus terminal GT, CT, and TT mispairs.[47] This agreement is somewhat coincidental given the large structural differences between the calculated and experimental systems. $\Delta\Delta G_{bind}^\circ$ of 0.4 kcal/mol corresponds to a two-fold difference between the equilibrium binding constants of the correct and incorrect dNTPs.

A complete determination of DNA replication fidelity by computer simulations must involve calculations of both the $\Delta\Delta G_{bind}^\circ$ and $\Delta\Delta g_{cat}^\ddagger$. Because our results were obtained using ground state molecular dynamics, they cannot be used to directly evaluate $\Delta\Delta g_{cat}^\ddagger$. However, the calculated ground state trajectories do indicate that the insertion of the correct dNTP may also be favored by the chemical step. In particular, the 3′-teminal oxygen, which attacks the P_α phosphorus of dNTP in the polymerization reaction, is found closer on average to P_α of the correct than of the incorrect dNTP (Figure 7). In a related class of ester hydrolysis reactions in solution, configurations that bring the nucleophilic oxygen closer than 3.2 Å from the attacked carbon were classified as "near attack conformations" (NACs) by Lightstone and Bruice.[48] The increased probability of the ground state dynamics to generate NACs was found to be linearly correlated with the rate constant for the corresponding uncatalyzed reaction.[48] By assuming that the same correlation is valid for our reaction, we estimate that the rate constant for the nucleophilic attack on the correct dNTP is about three times larger than that for the incorrect dNTP. This ratio corresponds to $\Delta\Delta g_{cat}^\ddagger$ of 0.6 kcal/mol. Our estimate is independent of the choice of the NAC distance for the new PO bond provided that this distance is chosen anywhere from 3.3 to 3.6 Å. Our use of a slightly larger NAC distance for the attack on phosphorus (than 3.2 Å suggested by Lightstone and Bruice) is consistent with a larger van der Waals radius of the phosphorus atom compared to carbon.

By combining the estimated magnitude of $\Delta\Delta g_{cat}^\ddagger$ with the calculated $\Delta\Delta G_{bind}^\circ$, we arrive at a 1 kcal/mol free energy preference for the insertion of dCTP opposite G in the template compared to the insertion of dTTP. This prediction is experimentally testable by standard steady state experiments[49] in the absence of polymerase if these experiments are able to monitor the formation of the product of the uncatalyzed insertion of dNTP at the DNA primer terminus for several weeks without DNA or dNTP degradation. Thus, the lability of dNTP substrates with respect to the hydrolysis of their terminal phosphate would have to be eliminated, for example, by using dNTP analogs with β-γ bridging oxygen replaced by a NH, CH_2 or CF_2 group. The addition of Mg^{2+} ions in the same concentration as dNTP would ensure experimental conditions commensurate with the simulated system. Adding Mg^{2+} in a concentration higher than equimolar is expected to increase the reaction rate[50] and perhaps also the insertion fidelity.

5. Conclusions

Our LIE and FEP simulations of a model system for the uncatalyzed DNA replication fidelity predict a 6-fold preference for the incorporation of the correct nucleotide at the end of the DNA primer strand. This preference is nearly equally distributed between the contributions originating from a stronger binding of the correct dNTP, and a larger rate constant for the formation of the new PO bond to the correct dNTP.

Acknowledgment. This work was supported by the NIH grant 1U19CA105010. We thank John SantaLucia, Jr. for providing us with the nearest-neighbor parameters for the terminal GT mismatches and Janez Mavri for fruitful discussions.

Supporting Information Available: Movies of the molecular dynamics trajectories for the alchemistic mutation of dTTP into dCTP, and for an elemental DNA folding event in which the dangling base spontaneously forms a stacked conformation via an intermediate T-shaped configuration.

References and Notes

(1) Wolfenden, R. *Acc. Chem. Res.* **1972**, *5*, 10.

(2) Amidon, G. L.; Pearlman, R. S.; Anik, S. T. *J. Theor. Biol.* **1979**, *77*, 161.

(3) Warshel, A. *Acc. Chem. Res.* **1981**, *14*, 284.

(4) Hadzi, D.; Kidric, J.; Koller, J.; Mavri, J. *J. Mol. Struct.* **1990**, *237*, 139.

(5) Pearlman, D. A. *J. Phys. Chem.* **1994**, *98*, 1487.

(6) Echols, H.; Goodman, M. F. *Annu. Rev. Biochem.* **1991**, *60*, 477.

(7) Florián, J.; Goodman, M. F.; Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6819.

(8) Åqvist, J.; Medina, C.; Samuelson, J. E. *Protein Eng.* **1994**, *7*, 385.

(9) Hansson, T.; Marelius, J.; Åqvist, J. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27.

(10) Sham, Y. Y.; Chu, Z. T.; Tao, H.; Warshel, A. *Proteins Struct. Funct. Genet.* **2000**, *39*, 393.

(11) Florián, J.; Goodman, M. F.; Warshel, A. *J. Phys. Chem. B* **2002**, *106*, 5739.

(12) Åqvist, J.; Hansson, T. *J. Phys. Chem.* **1996**, *100*, 9512.

(13) Simonson, T.; Brunger, A. T. *Biochemistry* **1992**, *31*, 8661.

(14) Smith, P. E.; van Gunsteren, W. F. *J. Phys. Chem.* **1994**, *98*, 13735.

(15) Mark, A. E.; van Gunsteren, W. F. *J. Mol. Biol.* **1994**, *240*, 167.

(16) Boresch, S.; Karplus, M. *J. Mol. Biol.* **1995**, *254*, 801.

(17) Boresch, S.; Archontis, G.; Karplus, M. *Proteins* **1994**, *20*, 25.

(18) Brady, G. P.; Sharp, K. A. *J. Mol. Biol.* **1995**, *254*, 77.

(19) Bren, U.; Martinek, V.; Florian, J. *J. Phys. Chem. B* **2006**, submitted.

(20) Sawaya, M. R.; Prasad, R.; Wilson, S. H.; Kraut, J.; Pelletier, H. *Biochemistry* **1997**, *36*, 11205.

(21) Goodman, M. F. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10493.

(22) Florián, J.; Goodman, M. F.; Warshel, A. *J. Phys. Chem. B* **2000**, *104*, 10092.

(23) Cubero, E.; Laughton, C. A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2000**, *122*, 6891.

(24) In the context of this paper, the term "protein" or "protein environment" means DNA and DNA environment, respectively.

(25) Florián, J.; Warshel, A.; Goodman, M. F. *J. Phys. Chem. B* **2002**, *106*, 5754.

(26) Brandsdal, B. O.; Åqvist, J.; Smalås, A. O. *Protein Sci.* **2001**, *10*, 1584.

(27) Almlof, M.; Åqvist, J.; Smalås, A. O.; Brandsdal, B. O. *Biophys. J.* **2006**, *90*, 433.

(28) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(29) Marelius, J.; Kolmodin, K.; Feierberg, I.; Åqvist, J. *J. Mol. Graphics Modell.* **1999**, *16*, 213.

(30) Florián, J.; Goodman, M. F.; Warshel, A. *J. Am. Chem. Soc.* **2003**, *125*, 8163.

(31) King, G.; Warshel, A. *J. Chem. Phys.* **1989**, *91*, 3647.

(32) Sham, Y. Y.; Warshel, A. *J. Chem. Phys.* **1998**, *109*, 7940.

(33) Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 3100.

(34) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Physics* **1977**, *23*, 327.

(35) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14.1*, 33.

(36) Valleau, J. P.; Torrie, G. M. A Guide to Monte Carlo for Statistical Mechanics. 2. Byways. In *Modern Theoretical Chemistry;* Berne, B. J., Ed.; Plenum: New York, 1977; Vol. 5; pp 169.

(37) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions;* John Wiley & Sons: New York, 1991.

(38) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420.

(39) Leach, A. R. *Molecular Modelling. Principles and Applications;* Prentice Hall: Harlow, UK, 2001.

(40) SantaLucia, J., Jr.; Allawi, H. T.; Senevirante, P. A. *Biochemistry* **1996**, *35*, 3555.

(41) SantaLucia, J., Jr., Personal communication.

(42) Clayton, L. K.; Goodman, M. F.; Branscomb, E. W.; Galas, D. J. *J. Biol. Chem.* **1979**, *254*, 1902.

(43) Galas, D. J.; Branscomb, E. W. *J. Mol. Biol.* **1978**, *88*, 653.

(44) Florián, J.; Goodman, M. F.; Warshel, A. *Biopolymers* **2003**, *68*, 286.

(45) Almlof, M.; Bransdal, B. O.; Åqvist, J. *J. Comput. Chem.* **2004**, *25*, 1242.

(46) Gelfand, C. A.; Plum, G. E.; Grollman, A. P.; Johnson, F.; Breslauer, K. J. *Biochemistry* **1998**, *37*, 7321.

(47) Petruska, J.; Goodman, M. F.; Boosalis, M. S.; Sowers, L. C.; Cheong, C.; I. Tinoco, J. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6252.

(48) Lightstone, F. C.; Bruice, T. C. *J. Am. Chem. Soc.* **1996**, *118*, 2595.

(49) Goodman, M. F.; Creighton, S.; Bloom, L. B.; Petruska, J. *Critical Rev. Biochem. Mol. Biol.* **1993**, *28*, 83.

(50) Blasko, A.; Bruice, T. C. *Acc. Chem. Res.* **1999**, *32*, 475.