Theoretical Investigation of the Binding Free Energies and Key Substrate-Recognition Components of the Replication Fidelity of Human DNA Polymerase β

Jan Florián,*,† Myron F. Goodman,†,‡ and Arieh Warshel†

Department of Chemistry and Department of Biological Sciences—Hedco Molecular Biology Laboratories, University of Southern California, Los Angeles, California 90089-1062

Received: March 22, 2002

We present a theoretical study of the selection of right/wrong dNTP substrates by DNA polymerases at the initial binding step, a major component of the DNA replication fidelity. Linear-response analysis (LRA) and molecular dynamics simulations are performed starting from the X-ray crystal structure of a ternary DNA polymerase β·DNA · ddCTP complex. These simulations provide converged structures of ternary complexes containing all four Watson−Crick (W−C) pairs as well as 11 neutral mismatched dNTP-template base pairs in the anti−anti conformations. The signs and overall magnitudes of the calculated relative free energies for binding of each dNTP to pol β·DNA complexes, which contain either correct or incorrect templating bases, agree with the observed universal preference of DNA polymerases for W−C base pairs. Overall, the binding free-energy differences of each dNTP to right versus wrong templates are found to be dominated by electrostatic interactions between templating and dNTP bases. However, about half of the electrostatic contribution can be attributed to the steric preorganization of the polymerase active site that was generated by the protein folding process. The preorganized site maintains optimal W−C pairing for matched bases while forcing mismatched pairs into configurations far from their ideal gas-phase geometries. Consequently, the preorganized site is responsible for large template contributions to fidelity. Individual additive contributions to fidelity are determined for active site residues. Interactions between incoming dNTPs and Asn279 and Tyr262 protein residues contribute significantly to the binding component of fidelity, with the Asn279 residue most effective in destabilizing each of the 15 nucleotide mispairs in neutral anti−anti conformations. Active site amino acids can also exert deleterious effects on fidelity. Tyr262 enhances base substitution fidelity via mispair destabilization, but it also stabilizes slipped mispaired primer·template structures that are potential precursors for +1 frameshift mutations. The inclusion of an extra water molecule in the active cleft was found to stabilize several wobble base pairs. Calculations performed at this level of "fine" detail are used to predict the effects of amino acid substitutions on the fidelity for mutant forms of pol β, thus providing a deeper understanding of the role of the polymerase active site in ensuring replication accuracy.

1. Introduction

The genomic stability of living organisms relies on highly accurate template-directed synthesis of chromosomal DNA by DNA polymerases. High replication fidelity is enforced in three steps, which include the preferential insertion of correct deoxyribonucleotide triphosphates (dNTPs) at the 3′-end of the growing DNA strand (primer), "proofreading" in the exonuclease site, and postreplicative editing.[1] In the first step, DNA polymerases are remarkably efficient in favoring the incorporation of Watson−Crick (WC) base pairs in the newly synthesized DNA. Nucleotide misinsertion frequencies measured for a variety of exonuclease-deficient DNA polymerases are in the range of 10^{-3}−10^{-6}, depending on sequence context and specific polymerase properties.[2] Such highly selective outcome of the nucleotide transfer reaction in the DNA polymerase active site cannot be explained by the relatively small differences in stability of WC versus non-WC base pairs in DNA in aqueous solution.[3,4]

The selection of the right nucleotide for incorporation in the growing DNA strand may occur during several reaction steps, including initial binding of dNTP and divalent metal cations, a conformational transition in the finger domain of the polymerase, and the phosphodiester bond formation. It has been often assumed that the conformational change is the rate-limiting step in the polymerization process, and the interpretation of kinetic experiments aimed at discerning the rate-limiting step from the binding step[5,6] seemed to confirm this notion. However, recent kinetic and crystallographic evidence excluded the conformational change as a rate-limiting step in the polymerization reaction.[7] Thus, it is reasonable to assume that the initial discrimination step reflects to a large extent differences in equilibrium binding constants for competing right and wrong dNTP substrates, where the insertion rate is proportional to the dNTP residence time in the polymerase active cleft.[8,9] This initial binding step includes the relaxation of the conformations of both the polymerase·DNA complex and the dNTP substrate. Correct base-pairing substrates remain bound longer than poorly paired substrates and are consequently favored for insertion. Of course, the total discrimination should also reflect the contribution of the chemical step.

In this paper, we calculate the selectivity of the right versus wrong dNTP binding in the active site of human polymerase β (pol β);[10−12] an analysis of the chemical step will be presented separately (J. Florian, A. Warshel, M. F. Goodman, unpublished; see also refs 13 and 14). Pol β plays important role in the repair

* Corresponding author. E-mail: florian@usc.edu.
† Department of Chemistry.
‡ Department of Biological Sciences.

10.1021/jp020790u CCC: $22.00
Published on Web 05/14/2002

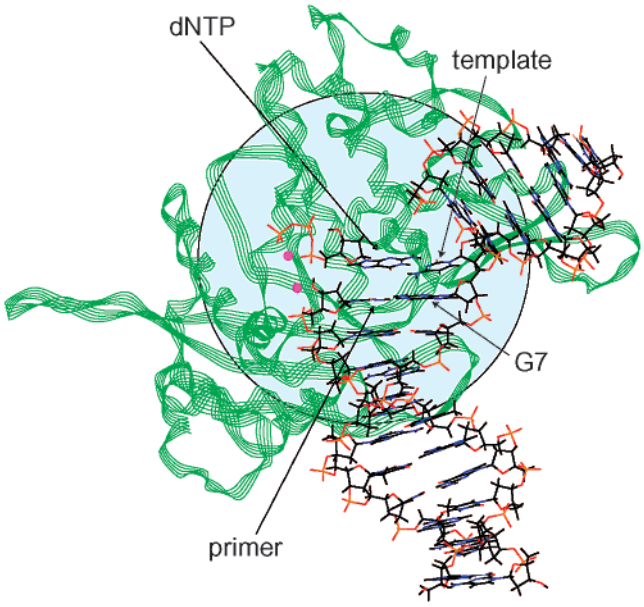


Figure 1. Ternary complex of human DNA polymerase β (green), DNA and dNTP substrates (atom-based colors). The size of the simulation sphere is indicated by the light blue color. Positions of Mg^{2+} ions are indicated by purple spheres. Throughout this paper, the "template" and "primer" denote the template base opposite to dNTP, and the 3′-terminal base of the newly synthesized DNA strand.

of damaged and chemically modified bases in mammalian DNA.[15] Using the crystal structure of the ternary pol β·DNA·dNTP complex[16] as the starting point for analysis, we examine nanosecond atomic-scale dynamics of the 18 Å sphere containing WC or mismatched dNTP·template base pairs and their environment (Figure 1). The average energies sampled during these simulations are used to predict relative stabilities of the base pairs in the pol β active site, and to calculate the contributions of each active site amino acid and primer/template bases to dNTP-binding discrimination. The accompanying paper addresses dNTP substrate binding contribution to the accuracy of proofreading-defective bacteriophage T7 DNA polymerase, providing a basis for understanding why T7 polymerase is able to synthesize DNA more accurately than pol β.

2. Computational Methods

The accurate calculation of the binding free energies of the protein−ligand complexes represents one of the most important and also most difficult tasks faced by computational chemists. Approaches that are formally rigorous within the limits of molecular mechanics potential energy surfaces, for example, thermodynamic integration (TI) or free-energy perturbation (FEP) calculations,[17−20] require extensive sampling of the protein−substrate configurations. Therefore, these methods are currently practical only for the evaluation of the relative binding free energy for two structurally very similar ligands. Thus, while FEP studies of transition mutations (pyrimidine → pyrimidine or purine−purine) are feasible,[21,22] energetics of the transversion mutations (purine ↔ pyrimidine) are currently beyond the reach of the FEP methodology. Even more importantly, binding free energies obtained by TI or FEP methods cannot be uniquely decomposed into a sum of contributions from the individual protein residues. These practical deficiencies can be alleviated when average electrostatic and van der Waals interaction energies between the ligand and the protein are used to determine binding free energies. The rationalization for the ability of average interaction energies to represent binding free energies with reasonable accuracy is formulated below in the framework of the linear response approximation (LRA). A more general discussion of the use of the LRA approach in binding calculations has been presented elsewhere.[23,24]

2.1. Selection of the Reaction Coordinate. Let us consider a chemical reaction in solution, in which the charge distribution Q of the solute changes from the reactant (Q_r) to product (Q_p) state. Usually, a single variable that changes monotonically during the course of the reaction is selected as a reaction coordinate. The magnitude of the reaction coordinate then characterizes the reaction progress. For example, the bond length to the leaving group would be a reasonable choice for the reaction coordinate of a bond dissociation reaction in the gas phase. However, because of the solvent/protein participation in the reaction, the selection of an appropriate reaction coordinate on such a geometric basis becomes more difficult for the reactions occurring in proteins and solutions. This problem becomes most obvious for charge-transfer reactions that are associated with small intramolecular energy changes, but large solvation changes. For such reactions, it is advantageous[18] to define the reaction coordinate x in terms of the difference in the solvation energy U for an instantaneous change in the solute charge distribution from Q_p to Q_r,

$$x = \Delta U = U_r - U_p \tag{1}$$

Equation 1 enables us to associate every snapshot on the molecular dynamics (MD) trajectory with its position on the reaction coordinate. Clearly, the value of the reaction coordinate defined by eq 1 monotonically increases upon going from the reactant-like solute−solvent configurations to the product-like configurations.

2.2. Formulation of the Linear Response Approximation (LRA). By running MD or Monte Carlo (MC) simulations, the probability density, $p_r(x)$, of finding the system at certain point of the reaction coordinate can be in principle evaluated as a configurational average ($\langle\ \rangle$) of the delta function (δ) on the potential surface U_r,[28]

$$p_r(x) = \int \delta(\Delta U - x) \exp[-\beta U_r]\, d\Gamma / \int \exp[-\beta U_r]\, d\Gamma = \langle \delta(\Delta U - x) \rangle_r \tag{2}$$

where $\beta = 1/k_B T$, k_B is the Boltzman constant, T is the absolute temperature, and $d\Gamma$ represents an element in the $3N$-dimensional configurational space, where N is the number of particles in the system. The δ function is defined by an integral equation

$$\int \delta(y - y_0) F(y)\, dy = F(y_0) \tag{3}$$

The probability density $p_r(x)$ determines the free-energy function $g_r(x)$ as

$$g_r(x) = -\beta^{-1} \ln p_r(x) \tag{4}$$

The typical profile of this free-energy function as well as the corresponding free-energy function for the simulation on the product surface (i.e., with the charge distribution Q_p) is illustrated in Figure 2. Because the probability of finding solute−solvent system in the configuration interacting less optimally with the charge distribution Q_r is small, the diabatic free energy $g_r(x)$ of the system increases for x more distant from x_r. Thus, functions $g_r(x)$ and $g_p(x)$ are schematically depicted in Figure 2 as two parabolic functions of equal curvature. The vertical distance from the bottom of parabola g_r to the value attained by this parabola at $x = x_p$ is defined as the

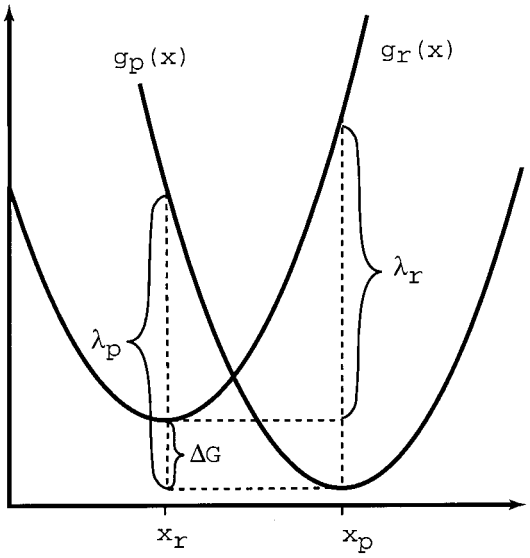


Figure 2. Diabatic free-energy surfaces corresponding to reactants and products of a charge-transfer reaction.

solvent reorganization energy λ_r,[28]

$$\lambda_r = g_r(x_p) - g_r(x_r) \tag{5}$$

Analogously, the solvent reorganization energy λ_p for the product state is defined as

$$\lambda_p = g_p(x_r) - g_p(x_p) \tag{6}$$

If we denote the reaction free energy as $\Delta G_{r\to p}$, we can write

$$\lambda_r = \Delta G_{r\to p} + g_r(x_p) - g_p(x_p) \tag{7}$$

and

$$\lambda_p = g_p(x_r) - g_r(x_r) - \Delta G_{r\to p} \tag{8}$$

The assumption that functions g_r and g_p have the same curvature leads to the condition $\lambda_r = \lambda_p$. Note that this assumption is very reasonable for electrostatic free-energy surfaces for charge-transfer reactions in condensed phase.[25,26] Using the condition $\lambda_r = \lambda_p$ along with eqs 7 and 8, the reaction free energy can be expressed as

$$\Delta G_{r\to p} = [g_p(x_r) - g_r(x_r) + g_p(x_p) - g_r(x_p)]/2 \tag{9}$$

Equation 9 can be transformed into formula involving average potential energy and the change of the internal energy of the solute (E^{gas}) by using the identity

$$x + \Delta E^{gas}{}_{p\to r} = U_r - U_p + \Delta E^{gas}{}_{p\to r} = g_r(x) - g_p(x) \tag{10}$$

for the derivation of which we refer the reader to the literature (eq 3.8 in ref 27 or eq 24 in ref 28).

Because simulations on the potential energy U_r (i.e., with the reactant's charges) will sample the configurations close to x_r, we can write

$$\langle U_r(x) - U_p(x)\rangle_r = U_r(x_r) - U_p(x_r) \tag{11a}$$

and similarly for averaging on the product potential surface we obtain

$$\langle U_r(x) - U_p(x)\rangle_p = U_r(x_p) - U_p(x_p) \tag{11b}$$

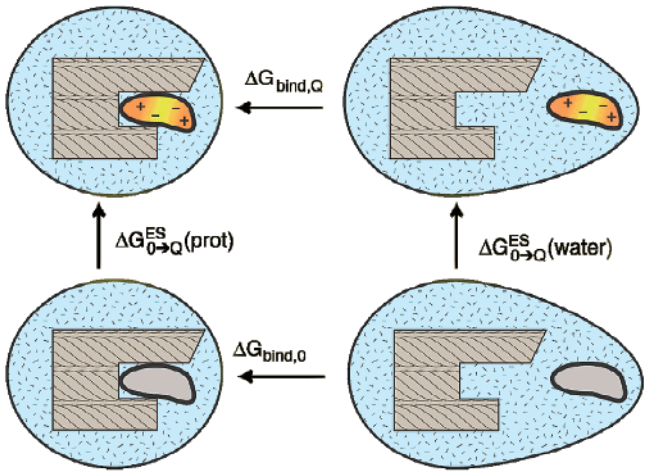


Figure 3. Thermodynamic cycle for the calculation of the binding free energy (ΔG_{bind}) of a ligand to a macromolecule in water. Subscripts Q and 0 indicate a fully charged ligand and the ligand with zero atomic charges. Note that the vertical free-energy changes, i.e., the free energies for charging a ligand in water and protein environment, are reduced in LRA framework (eq 13) to changes in the corresponding electrostatic components, which is indicated here by superscript ES. Also note that the van der Waals potential does not change in our cycle.

Using eqs 10 and 11, eq 9 can be rewritten as

$$\Delta G_{r\to p} = [\langle U_p - U_r\rangle_r + \langle U_p - U_r\rangle_p]/2 + \Delta E^{gas}{}_{r\to p} \tag{12}$$

This equation approximates the reaction free energy by the sum of differences in the solvation potential energy sampled on both the reactant and product potential surfaces, augmented by the change in the internal energy of the solute. In general, we can express U as a sum of electrostatic (U^{ES}) and van der Waals (U^{vdW}) energies. Now, if we consider processes in which U^{vdW} does not change (i.e., $\langle U^{vdW}{}_p - U^{vdW}{}_r\rangle_r = \langle U^{vdW}{}_p - U^{vdW}{}_r\rangle_p = 0$), eq 12a can be rewritten in terms of the electrostatic part of the potential energy as

$$\Delta G_{r\to p} = [\langle U^{ES}{}_p - U^{ES}{}_r\rangle_r + \langle U^{ES}{}_p - U^{ES}{}_r\rangle_p]/2 + \Delta E^{gas}{}_{r\to p} \tag{13}$$

2.3. Application of LRA to the Calculations of the Electrostatic Component of the Binding Free Energy. In this paper, we express the binding free energy ($\Delta G_{bind} = \Delta G_{bind,Q}$) of a ligand to a macromolecule in water as (Figure 3)

$$\Delta G_{bind} = \Delta G^{ES}{}_{bind} + \Delta G_{bind,0} \tag{14a}$$

where the electrostatic part of the binding free energy ($\Delta G^{ES}{}_{bind}$) is defined as a difference in free energies of charging a ligand in water and in the protein active site,

$$\Delta G^{ES}{}_{bind} = \Delta G^{ES}{}_{0\to Q}(\text{prot}) - \Delta G^{ES}{}_{0\to Q}(\text{water}) \tag{14b}$$

The free energies for these charging processes are described here by LRA of eq 13 (see also below). Unfortunately, the free energy for the binding of the uncharged ligand ($\Delta G_{bind,0}$) does not fall into the realm of LRA. As pointed out in ref 24, this contribution could be rigorously evaluated as a difference in free energies for shrinking the uncharged ligand into "nothing" in water and in the protein active site. However, the relevant computer simulations of this process converge extremely slowly and therefore they were not considered as a viable option for the present study. The $\Delta G_{bind,0}$ term includes van der Waals ($\Delta G^{vdW}{}_{bind}$), entropic ($T\Delta S_{bind,0}$), and other contributions ($\Delta G_{rest,0}$).

The $\Delta G_{rest,0}$ includes, for example, the change in the solvation of the protein−ligand system due to the binding of uncharged ligand (for more detailed discussion and explicit calculation of this term see ref 24). Thus, we can write

$$\Delta G_{bind,0} = \Delta G^{vdW}{}_{bind} - T\Delta S_{bind,0} + \Delta G_{rest,0} \quad (14c)$$

Now we are ready to evaluate the free enrgy terms of eq 14. We start by evaluating the free energies for charging ligand (solute) in the protein binding site and in water. This is done by expressing eq 12 in terms of the statistical averages over the differences in the potential energies calculated for the charged (U_Q) and uncharged (U_0) ligand,

$$\Delta G^{ES}{}_{0\rightarrow Q} = [\langle U_Q - U_0\rangle_0 + \langle U_Q - U_0\rangle_Q]/2 + \Delta E^{gas}{}_{0\rightarrow Q} \quad (15)$$

The potential energy U is defined as the sum of the solute−solvent interaction energy (U_{Ss}) and solvent−solvent interaction energy (U_{ss}),

$$U = U_{Ss} + U_{ss} \quad (16)$$

where the protein atoms (if present) are considered as a part of the solvent. If the potential energies U_Q and U_0 are sampled on the same potential energy surface, the U_{ss} terms in U_Q and U_0 will mutually cancel because the charges of the solvent atoms are identical in U_Q and U_0. Analogously, the nonelectrostatic part of U_{Ss} will contribute equally to U_Q and U_0. Consequently, eq 15 will be reduced to

$$\Delta G^{ES}{}_{0\rightarrow Q} = [\langle U^{ES}{}_{Ss,Q} - U^{ES}{}_{Ss,0}\rangle_0 + \langle U^{ES}{}_{Ss,Q} - U^{ES}{}_{Ss,0}\rangle_Q]/2 + \Delta E^{gas}{}_{0\rightarrow Q} \quad (17)$$

Equation 17 can be further simplified to

$$\Delta G^{ES}{}_{0\rightarrow Q}(prot) = [\langle U^{ES}{}_{Ss,Q}(prot)\rangle_0 + \langle U^{ES}{}_{Ss,Q}(prot)\rangle_Q]/2 + \Delta E^{gas}{}_{0\rightarrow Q}(prot) \quad (18a)$$

$$\Delta G^{ES}{}_{0\rightarrow Q}(water) = \langle U^{ES}{}_{Ss,Q}(water)\rangle_Q/2 + \Delta E^{gas}{}_{0\rightarrow Q}(water) \quad (18b)$$

due to the fact that $U^{ES}{}_{Ss,0} = 0$ for zero solute charges, and $\langle U^{ES}{}_{Ss,Q}\rangle_0 = 0$ in water. Finally, making a reasonable assumption that the change of the intramolecular energy of the solute is the same in the protein and in water, i.e.,

$$\Delta E^{gas}{}_{0\rightarrow Q}(prot) - \Delta E^{gas}{}_{0\rightarrow Q}(water) = 0 \quad (19)$$

we obtain

$$\Delta G^{ES}{}_{bind} = (\langle U^{ES}{}_{Ss,Q}(protein)\rangle_Q - \langle U^{ES}{}_{Ss,Q}(water)\rangle_Q + \langle U^{ES}{}_{Ss,Q}(protein)\rangle_0)/2 \quad (20)$$

The electrostatic solute−solvent interactions in eq 20 are determined by Coulomb's law

$$U^{ES}{}_{Ss,Q}(kcal/mol) = 332\Sigma q_i Q_j / r_{ij} \quad (21)$$

where q_i and Q_j denote charges (in atomic units) on the solute and solvent atoms, respectively, and r_{ij} is the distance (angstroms) between the ith solute atom and jth solvent atom.

Equation 20 is based on the LRA of eq 12 and thus it provides a rigorous theoretical basis for the simple evaluation of $\Delta G^{ES}{}_{bind}$ from statistical ensembles generated by MD or MC simulations. Unfortunately, such a rigorous yet simple prescription is not available for the $\Delta G^{vdW}{}_{bind}$, $T\Delta S_{bind}$, and $\Delta G_{rest,0}$ terms. However, a reasonable estimate of these terms can be made if we attempt to evaluate the relative binding free energy,

$$\Delta\Delta G_{bind} = \Delta G_{bind}(B) - \Delta G_{bind}(A) \quad (22)$$

where $\Delta G_{bind}(B)$ and $\Delta G_{bind}(A)$ are the binding free energies of the same ligand to the protein active sites B and A, respectively. In our case, A and B correspond to the two different templates in the polymerase active site. If the sites A and B are structurally similar, we can expect the entropic contribution to ΔG_{bind} to partially cancel out, since we are dealing with uncharged ligands that can move without hydrogen bonding restrictions. Similarly, we expect that the change in solvation of the sites A and B upon binding of the same ligand will partially cancel out. The errors due to these assumptions will be proportional to the difference in the empty space that is not filled by the ligand. Consequently, these errors, as well as the whole $\Delta\Delta G_{bind,0}$ term, will be proportional to the difference in the average van der Waals solute−solvent interaction energies, that is

$$\Delta\Delta G_{bind,0} = \alpha[\langle U_{Ss}{}^{vdW}(B)\rangle_Q - \langle U_{Ss}{}^{vdW}(A)\rangle_Q] \quad (23)$$

Although our choice of the statistical averaging over the trajectory of the fully charged ligand ($\langle\ \rangle_Q$) in eq 23 is somewhat arbitrary, this procedure is justified by the considerations mentioned above. At any rate, one usually obtains similar results for $\langle\ \rangle_0$. In the Amber force field used in this study, $U_{Ss}{}^{vdW}$ energies consist of the exchange repulsion term, which varies with the interatomic distance as $1/r_{ij}{}^{12}$, and the attractive London dispersion term, which is proportional to $1/r_{ii}{}^{6}$. The attractive part of $U_{Ss}{}^{vdW}$ overcomes the repulsive term for interatomic separations that are larger than the sum of the atomic vdW radii of atoms i and j.

Using the LRA (eq 20, or in a more general form, eq 12), the electrostatic part of the relative binding free energy is evaluated as

$$\Delta\Delta G^{ES} = \Delta\Delta G_Q{}^{ES} + \Delta\Delta G_0{}^{ES} \quad (24a)$$

where

$$\Delta\Delta G_Q{}^{ES} = 0.5\,[\langle U^{ES}{}_{Ss,Q}(B)\rangle_Q - \langle U^{ES}{}_{Ss,Q}(A)\rangle_Q] \quad (24b)$$

and

$$\Delta\Delta G_0{}^{ES} = 0.5[\langle U^{ES}{}_{Ss,Q}(B)\rangle_0 - \langle U^{ES}{}_{Ss,Q}(A)\rangle_0] \quad (24c)$$

Note that the uncharged state (denoted by the subscript 0 in eq 24a,c) is our "reactant", and thus, in accord with the general formulation of the LRA approximation (eq 12), the correct electrostatic component of binding must include $\Delta\Delta G_0{}^{ES}$ and not just $\Delta\Delta G_Q{}^{ES}$.

By adding the electrostatic (eq 24) and the scaled vdW (eq 23) contributions, we obtain an LRA expression for the relative binding free energy of a solute to proteins A and B,

$$\Delta\Delta G_{bind} = \Delta\Delta G_Q{}^{ES} + \Delta\Delta G_0{}^{ES} + \Delta\Delta G_{bind,0} \quad (25)$$

which uses a single adjustable parameter α. Thus, the approach of eqs 23−25 will be referred to as LRA/α. The solute−protein A and solute−protein B interaction energies, which are needed

TABLE 1: Comparison of Calculated[a] and Observed Hydration Free Energies of Relevant Model Systems

				ΔG_{hydr} (kcal/mol)	
solute	ES_{Ss}[b]	vdW_{Ss}[c]	$-T\Delta S_{hydr}$[d]	calcd[a]	exptl
pyridine	−12.2	−8.0	6.6[e]	−4.0	−4.7[e]
acetamide	−27.6	−3.9	5.1[f]	−10.9	−9.7[g]
difluorotoluene	−4.2	−11.5	8.2[f]	−0.3	−0.3[h]

[a] Calculated using eq 28, with the parameters $\alpha = 0.56$ and $\beta = 0.50$. The parameter α was determined by minimizing the RMS deviation between the calculated and observed ΔG_{hydr}. β=0.5 was obtained by assuming the validity of LRA for the electrostatic part of ΔG_{hydr}. The solute−water interaction energies were calculated by averaging over a 500 ps trajectory generated by molecular dynamics simulation (for further details see section 2.5). [b] Average of the solute−solvent electrostatic energy (kcal/mol) sampled on the potential surface of the fully charged solute in water ($\langle U^{ES}_{sS,Q}(water)\rangle_Q$ in eq 28). [c] Average of the solute−solvent van der Waals energy (kcal/mol) sampled on the potential surface of the fully charged solute in water ($\langle U^{vdW}_{sS}(water)\rangle_Q$ in eq 28). [d] Entropic contribution to ΔG_{hydr} (kcal/mol, $T = 298$ K). [e] Reference 68. [f] Calculated by the Langevin dipoles solvation model.[29] [g] Reference 69. [h] Reference 21.

for the evaluation of eq 25, are calculated in this paper by the MD averaging described in section 2.5.

2.4. Adjustment of the Parameter α. Whereas the parameter 0.5 that appears in eq 24 is determined by the LRA, coefficient α for the van der Waals (vdW) interactions depends on the system studied, and thus it has to be adjusted empirically. This is done here by comparing the observed and calculated hydration free energies, ΔG_{hydr}. ΔG_{hydr} is defined as the free energy of transfer of the solute from the 1 M gas-phase state to 1 M aqueous solution. This free energy corresponds to "binding" of a gas-phase molecule by water. Hence we can use the same considerations that led to eq 14 and write

$$\Delta G_{hydr} = \Delta G^{ES}_{hydr} + \Delta G^{vdW}_{hydr} - T\Delta S_{hydr} \qquad (26)$$

Although ΔG^{ES}_{hydr} may in principle include some entropic contribution, the explicit $T\Delta S_{hydr}$ term is needed to account for the positive observed ΔG_{hydr} for hydrophobic solutes, such as alkanes. The algorithm based on eq 26 and the dipolar model of water has been successful in reproducing hydration free energies of charged and uncharged medium-sized solutes.[29] Using LRA (eq 13), ΔG^{ES}_{hydr} can be written as

$$\Delta G^{ES}_{hydr} = \Delta G^{ES}_{0\rightarrow Q}(water) - \Delta G^{ES}_{0\rightarrow Q}(gas) = \beta\langle U^{ES}_{sS,Q}(water)\rangle_Q \qquad (27)$$

and

$$\Delta G_{hydr} = \beta\langle U^{ES}_{Ss,Q}(water)\rangle_Q + \alpha\langle U^{vdW}_{Ss}(water)\rangle_Q - T\Delta S_{hydr} \qquad (28)$$

where $\beta = 0.5$. To select a reasonable value of parameter α, we assume that the magnitude of this parameter for the solute-water interactions is the same as for solute−protein interactions. We also exploit the fact that the LRA is expected to be valid for the electrostatic part of ΔG_{hydr}, and that experimental ΔG_{hydr} and in some cases also experimental $T\Delta S_{hydr}$ are available. More specifically, $\alpha = 0.56$ is obtained here by comparing the calculated and observed solvation free energies for acetamide, pyridine, and difluorotoluene (Table 1). The acetamide and pyridine molecules were selected as benchmark systems because they contain structural features similar to those present in nucleic acid bases. (Note that the nucleic acid bases themselves could not be used as benchmarks because their experimental absolute hydration free energies are unknown.) However, both acetamide and pyridine molecules have a disadvantage in that the electrostatic and vdW components of ΔG_{hydr} are strongly correlated. That is, even small deviations of the β parameter for pyridine or acetamide, from its theoretical LRA value of $^1/_2$, would greatly affect the magnitude of the coefficient α. Therefore, we augmented our set of benchmark molecules by difluorotoluene, which is a nonpolar analogue for thymine.

2.5. Molecular Dynamics Simulations. The configurational ensembles for the evaluation of average solute−solvent electrostatic and van der Waals energies were generated from the unconstrained molecular dynamics (MD) trajectories using the AMBER force field[30] implemented in the program Q, version 3.77.[31] The AMBER force field was selected for its ability to provide stable duplex DNA conformations (see, e.g., refs 32−35) as well as reasonable structural properties of proteins and protein-nucleic acid complexes.[36] In addition, our program MOLARIS,[37] which allowed us to use induced dipoles on the DNA and protein atoms, was used to examine the force-field dependence of the calculated results.

Starting structures for MD simulations were generated from the crystal structure of human DNA polymerase β, complexed with the DNA and ddCTP338 [16] (1bpy, resolution 2.2 Å) as described below. The 2′,3′-dideoxyribose moieties of DCT338 (incoming nucleotide) and DC10 (3′-terminal nucleotide of the primer strand) in the crystal structure were modified by adding 3′-OH groups. With the exception of the histidine residues, which were kept in their neutral form, the protonation state of ionizable residues was determined from their pK_a constants in water. That is, all Glu and Asp residues and DNA phosphate groups were negatively charged (−1), whereas all Lys and Arg residues carried charge +1. This selection of charged groups resulted in overall electroneutral systems within the 7 and 15 Å spheres centered on the base of dCTP338 (substrate). The base of the incoming nucleotide and the opposite template were "mutated" to form all 16 possible base pairs in the pol β active site, while keeping glycosidic bonds in their anti conformations corresponding to standard Watson−Crick structures. These structural modifications were done using 'mutate residue' command in the SYBYL 6.6 program,[38] and were followed by manual adjustment of selected torsional angles to remove large steric clashes. Because steric problems could not be satisfactorily removed for the G(anti)·G(anti) base pair, this pair was not included in the structures studied by MD simulations. The water molecules were added into the simulated system by immersing the simulation sphere into the sphere of bulk water molecules. Those water molecules that were not sterically overlapping with the atoms present in the crystal structure were retained in the starting structure for the MD simulations. In certain cases (e.g., for some pyridine−pyridine mismatches) an extra water molecule was manually inserted (in addition to the above mentioned water molecules) in the binding pocket to stabilize the given mispair by the formation of hydrogen bonds with some template and/or dNTP atoms.

The simulated DNA−protein−substrate complex was immersed in 18 Å sphere of TIP3P water molecules subject to the surface-constraint all-atom solvent (SCAAS) type boundary conditions.[39,40] The substrate base was positioned near the center of this sphere. Positions of DNA atoms protruding beyond the 18 Å sphere were fixed at their crystallographic positions, and their nonbonded interactions with the atoms within the simulation sphere were turned off. Except for the early stages of the equilibration process, no positional constraints were used for

atoms inside the simulation sphere. Limiting the fully simulated part of the system to 18 Å sphere was necessary to achieve simulation times that were long enough to allow for the proper reorganization of the protein environment perturbed by the presence of mismatched base pairs. For each base pair, the structure of the pol β/DNA/substrate/water system was equilibrated by the following protocol. First, the solvent and substrate were relaxed by a series of three short-step (0.05−0.5 fs) simulations at 10 K (total 3 ps simulation time). Subsequently, the constraints on the protein atoms were gradually removed in a series of six MD simulations (total 22ps simulation time) at the temperature 10−30 K with the integration step 1fs. These calculations were followed by the gradual heating of the simulated system from 30 to 298 K in a series of 10 unconstrained MD simulations with 1 fs step and 85 ps total simulation time. The equilibration protocol was concluded with 1.66 ns simulation with the simulation parameters identical with the actual production calculations described below. We choose such a long equilibration period in order to provide sufficient time for the protein and solvent environment to adapt to the presence of mismatched base pairs that were not part of the original crystal structure. The final geometry of equilibration protocol described above was used as a starting point of the production calculations with fully charged ligand (leading to the averages denoted $\langle\rangle_Q$ in section 2.3), and also as a starting point for the equilibration protocol for the simulations with the uncharged substrate. This equilibration included 80 ps simulation with the 1 fs step at 298 K followed by 660 ps equilibration with the same conditions as the production run.

The electrostatic and vdW interaction energies between the nucleobase part of the substrate and the solvent in the pol β simulations were sampled at 20 fs intervals along 200 ps constant-temperature (298K) trajectories generated with the 2 fs integration step. The "shake" algorithm was applied to both solute and solvent molecules. The nonbonded interactions were evaluated explicitly for distances shorter than 10 Å. The local-reaction field (LRF) method [40,41] was used to treat long-range electrostatic interactions for distances beyond 10 Å cutoff.

The electrostatic contributions of a protein or DNA residue *(R)* to $\Delta\Delta G_{\text{bind}}$ of eq 25 were estimated as

$$\Delta\Delta G^{\text{R}}_{\text{bind}} = 0.5\,[\langle U_{\text{SR}}{}^{\text{ES}}(\text{B})\rangle_Q - \langle U_{\text{SR}}{}^{\text{ES}}(\text{A})\rangle_Q + \langle U_{\text{SR}}{}^{\text{ES}}(\text{B})\rangle_0 - \langle U_{\text{SR}}{}^{\text{ES}}(\text{A})\rangle_0] \quad (29)$$

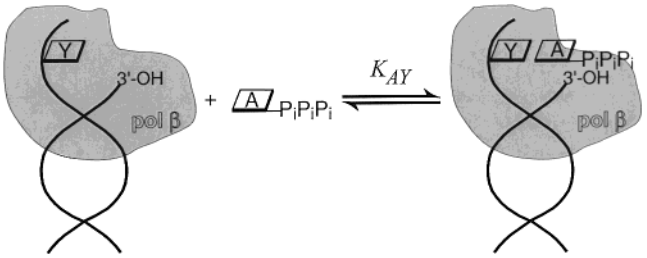
where subscript SR denotes that only the interactions between the residue R and the nucleobase part of dNTP were included. As stated above, A and B correspond to different templates in the enzyme active site. In these calculations, $U_{\text{SR}}{}^{\text{ES}}$ energies were averaged over 100 structures sampled at 2 ps intervals along the 200ps production trajectories. For R corresponding to a template or primer residue, only the contribution from its nucleobase moiety (i.e., the base + C1′H atoms of the sugar) was included in the $U_{\text{SR}}{}^{\text{ES}}$ energy.

Because the standard Amber parameter library does not contain parameters for the triphosphate part of dNTP, reasonable values of these parameters had to be determined. Our choice of atomic charges on the triphosphate moiety of the substrate was based on the ab initio charges of hydrogentriphosphate methylester ($CH_3O(PO_3)_3H^{3-}$) complexed with Mg^{2+}. The geometry of this complex was optimized by the gas-phase Hartree−Fock ab initio calculations using the 6-31G(d) set of atomic orbitals. During this optimization, the torsional angles involving backbone atoms of the triphosphate unit (C−O_1−P_α−O_2−P_β−O_3−P_γ−OH, where O_1, O_2, and O_3 denotes the bridging oxygen) were fixed at their values corresponding to the conformation of the substrate observed in the pol β crystal. The atomic charges were obtained by fitting to the HF/6-31G(d) electrostatic potential of the complex calculated self-consistently in the presence of dielectric continuum with the dielectric constant 80 by the PCM method. Standard Pauling atomic radii multiplied by 1.2 were used for the PCM calculations. Ab initio calculations were carried out using the Gaussian 94 program.[42] The calculated charge for the Mg^{2+} ion (1.66) was used for both Mg^{2+} atoms present in the active site. The charges calculated for the triphosphate part of the complex were manually redistributed subject to the following restraints: (i) zero total charge for the $dNTP^{4-}\cdot 2Mg^{2+}$ complex (i.e., the γ-phosphate is fully deprotonated), (ii) standard Amber charges are retained for the nucleoside part of the $dNTP^{4-}$, (iii) charges of all nonbridging oxygens attached to Pγ (denoted Oγ) are equal, (iv) charges on two nonbridging oxygens attached to P_β (denoted O_β) are equal, (v) charges on two nonbridging oxygens attached to P_α (denoted O_α) are equal. The charge transfer required for satisfying the first two conditions was accomplished by decreasing the charges calculated for the P atoms, because these atoms are significantly more polarizable than oxygen atoms. This procedure resulted in charges −0.8128, 1.2, −0.77, −0.5, 1.19, −0.55, 1.18, −0.7761, −0.4954, and −0.0069 for the O_γ, P_γ, O_β, O_3, P_β, O_2, P_a, O_α, O_1, and $C_{5'}$ atoms, respectively. The Amber atom types used were O3 for O_γ atoms, O2 for O_α and O_β atoms, and OS for bridging oxygens. The atom type O3 was newly added in the Amber parameter set. The following parameters involving this atom type were used: $R^* = 1.70$ Å, $\epsilon = 0.21$ kcal/mol, $k_{\text{P-O3}} = 950$ kcal/mol/Å^2, $r_{\text{P-O3}} = 1.52$ Å, $k_{\text{O3-P-O3}} = 250$ kcal/mol/rad^2, $\phi_{\text{O3-P-O3}} = 109.9°$, $k_{\text{O3-P-O3}} = 200$ kcal/mol/rad^2, and $\phi_{\text{O3-P-OS}} = 108.5°$, where R^* is the vdW radius, ϵ is the bonding parameter, k is the force constant, r is the equilibrium bond distance, and ϕ is the equilibrium bond angle. These parameters differ only marginally from the standard Amber parameters for the O2 atom type.

Separate LRA calculations with the MOLARIS 8.10 program and the polarizable ENZYMIX force field [37,43] were carried out to estimate the contribution of the atomic polarizabilities on DNA and protein atoms to the relative binding energetics. For this purpose, we compared the results of two 50 ps simulations (1 fs step) at 298 K that were initiated from the same geometry of the pol β ternary complex, and were carried out using the ENZYMIX force field with the induced dipoles turned off and on.

3. Results

The computer simulations presented in this paper were carried out with the goal of achieving maximal accuracy of the calculated energies using current state-of-the-art computer resources. This requirement led us to the setup of a computer experiment that differs from in vitro polymerase fidelity measurements currently available in the literature. Misinsertion frequencies (and their binding and k_{cat} contributions) are typically reported for variable dNTP opposite to a fixed template base.[44−46] We have calculated the relative binding of a single substrate with protein−DNA complexes differing in the identity of the template base (Figure 4). This approach ensures that the nonelectrostatic $\Delta G_{\text{bind},0}$ term (Figure 3) will be small, and consequently that approximations used in eq 23 will not detrimentally affect the calculated $\Delta\Delta G_{\text{bind}}$. This approach also significantly limits inaccuracies due to the limited radius of the simulation sphere. Our definition of $\Delta\Delta G_{\text{bind}}$ (Figure 4) means that our results reveal the source of template fidelity rather than



$$\log K_{AY} - \log K_{AT} = -\Delta\Delta G_{bind}/1.36,\ Y = A, C, T, G$$

Figure 4. Setup of the computer experiment addressing the template fidelity for the dATP substrate. The templating base is denoted as Y. The relationship between the calculated relative binding free energies ($\Delta\Delta G_{bind}$) and the corresponding equilibrium binding constants (K) is given for the temperature 298 K and dATP substrate.

substrate fidelity of pol β. The existence of dual fidelity (selectivity) of pol β is due to the fact that this protein works with two classes of substrates: DNA and dNTP. If one of these classes is restricted to contain only a single member, pol β preferentially selects such a member from the other class that allows it to form a Watson−Crick base pair. It is important to note that the "substrate fidelity" and "template fidelity" for the formation of the same mispairs may in principle differ because the template and dNTP environments in the protein active site are different. Still, in both cases, the active site is epected to discriminate against base-pair mismatches. This discrimination implies positive values of $\Delta\Delta G_{bind}$ for all dNTP substrates. (Note that although $\Delta\Delta G_{bind}$ is explicitly defined in Figure 4 only for the specific case of dATP substrate, the generalization of this definition to other substrates is straightforward.) A relevant experiment addressing "template fidelities" and their binding/catalytic components would involve the study of the extension of the 3′primer end of DNAs containing various template bases competing for a single dNTP substrate. In addition, it should be emphasized that the calculated $\Delta\Delta G_{bind}$ corresponds to neutral mispairs, in which both the templating and the substrate nucleotides are in the anti conformation. Obviously, some mispairs containing protonated, deprotonated, or flipped (syn conformations) nucleosides may be in principle more stable then their neutral anti−anti counterparts.[47−49] Moreover, a participation of rare tautomeric forms of the bases cannot be completely disregarded. Therefore, the observed misinsertion frequencies (when they become available for the "template fidelities") and their binding components may be larger than probabilities corresponding to the calculated free-energy differences for neutral anti−anti base pairs.[50]

The magnitudes of $\Delta\Delta G_{bind}$ calculated by the LRA method (eq 25) and their electrostatic and vdW components are compared in Table 2. Here, the positive numbers in the $\Delta\Delta G_{bind}$ column indicate the destabilization (in kcal/mol) of the mispair relative to the corresponding Watson−Crick pair in the pol β active site (Figure 5). Because all the calculated $\Delta\Delta G_{bind}$ are positive, the computer experiment predicts that the substrate binding is always responsible for a part of the "template fidelity" of polymerase β. The mispairs calculated to be the most stable are the dATP·G, dTTP·C, dTTP·G, and dTTP·T.

The analysis of Table 2 also reveals that the electrostatic energies are the largest contributors to the calculated $\Delta\Delta G_{bind}$. The additivity of the LRA formalism used in our simulations (eq 25) enabled us to provide more detailed information about the makeup of these interactions by expressing the calculated $\Delta\Delta G_{bind}$, in terms of electrostatic contributions of nearby protein and DNA residues (eq 29). The residues that were chosen for

TABLE 2: The Calculated Relative Binding Free Energies ($\Delta\Delta G_{bind}$) for the Pol β-DNA−Substrate Complexes Involving the Formation of Mispairs[a]

dNTP	template	$\Delta\Delta G^{ES}{}_Q$	$\Delta\Delta G^{ES}{}_0$	$\Delta\Delta G_{bind,0}$	$\Delta\Delta G_{bind}$
dATP	T	0	0	0	0
	A	5.0	5.1	1.9	11.9
	C	5.3	7.0	0.3	12.7
	G	0.9	0.0	0.8	1.6
dTTP	A[b]	0	0	0	0
	C[b]	−0.5	1.4	0.7	1.6
	G	0.2	1.3	−0.6	0.9
	T[b]	1.9	−0.1	0.8	2.6
dGTP	C	0	0	0	0
	A	1.0	6.4	0.1	7.4
	T[b]	1.5	9.3	−0.4	10.4
dCTP	G	0	0	0	0
	C[c]	−0.6	4.1	0.0	3.5
	A[c]	3.6	5.1	2.4	11.1
	T[b]	4.9	2.3	−1.1	6.1

[a] The results (in kcal/mol) are given relative to the Watson−Crick, which involves the same dNTP. $\Delta\Delta G^{ES}{}_Q$ and $\Delta\Delta G^{ES}{}_0$ denote the electrostatic components of $\Delta\Delta G_{bind}$ calculated by averaging over trajectories evaluated for the charged and uncharged nucleobase part of dNTP, respectively (eq 24). $\Delta\Delta G_{bind,0}$ denotes the contribution from the binding of an uncharged ligand calculated using eq 23. [b] Average of the results presented in Table 4. [c] Initial structure for MD simulations included an extra water molecule added manually in the binding pocket in order to stabilize the mispair.

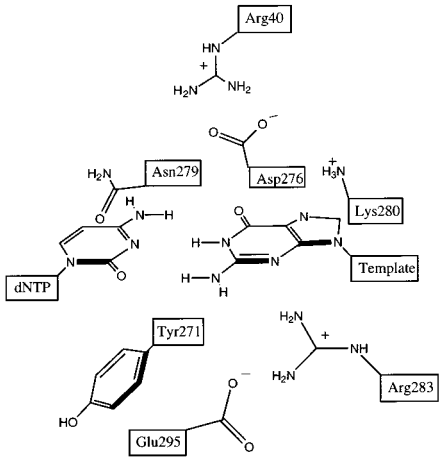


Figure 5. Schematic representation of the protein residues interacting with the base moieties of the template and dNTP substrate in the active site of pol β. The numbering of the protein residues used in the original paper of Sawaya et al.[16] was retained.

the group contribution analysis include Arg283, Glu285, and Tyr271 residues located in minor groove, and Asn279, Arg40, Asp276, and Lys280 residues that are stacked above the dNTP·template base pair (Figure 5). In addition, we included the templating base, the 3′-terminal base of the primer (dC352), and the guanosine base pairing with dC352 (dG7). Note that the pol β active site does not contain any amino acid side chains interacting directly with the major groove edge of the dNTP·template base pair, and that this region of the active site is accessible to bulk water molecules. A detailed structural and energetic account is given below.

3.1. dATP as Incoming dNTP. The calculated structure of the reference dATP·T base pair (Figure 6) shows a regular planar Watson−Crick pair with 2.9 Å average length of N−H···O and N···H−N hydrogen bonds. The deoxyribose of the substrate and primer nucleosides adopt the C3′-endo conformation that is typically found in A-DNA and RNA duplexes. In contrast, the template strand, including the template nucleoside, is found

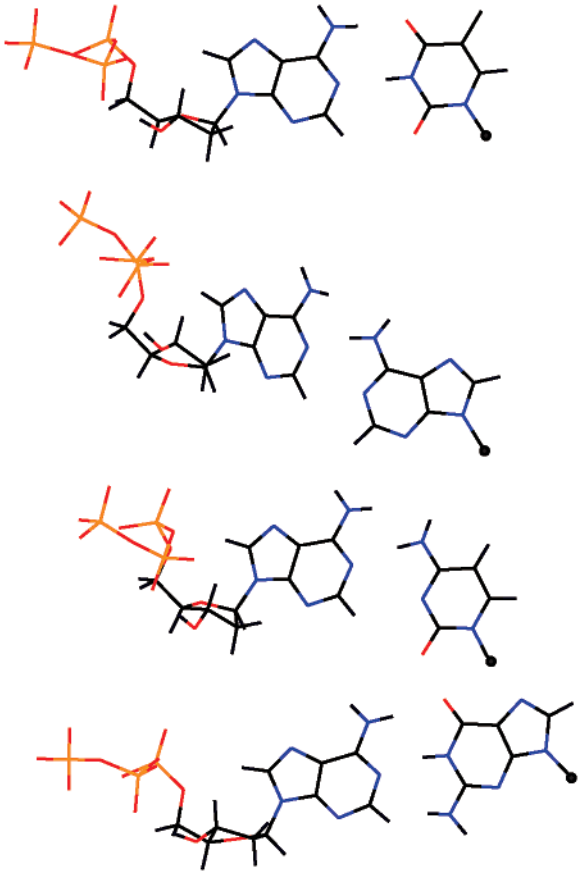


Figure 6. Average structures of the Watson−Crick and mismatched base pairs in the polymerase β active site. dATP binding opposite to T, A, C, and G (from top to bottom, respectively)

to have C2′-endo (B-DNA) sugars. A switch from the B- to A-DNA conformation in a polymerase active site has been observed in the crystal structure of DNA polymerase I (pol I) determined at 1.8 Å resolution.[51] However, the conformation of the primer nucleoside in the pol I crystal structure was C2′-endo. At 2.2 Å resolution, the available crystal structure of pol β is not accurate enough to distinguish puckering modes of nucleoside sugars. In addition, the substrate and primer nucleosides are present in the pol β crystal structure in the 2′,3′-dideoxy forms. The missing 3′-OH groups may affect conformations of the respective sugars.

When dATP is bound next to A in the template, it is found to be displaced into the major groove (Figure 6). This displacement enables the formation of a wobble base pair, which appears in the projection of Figure 6 to be stabilized by two hydrogen bonds. However, the dATP·A pair is significantly buckled, probably due to steric constraint imposed by the active site residues. In addition, the departure from the base pair planarity is driven by the formation of a N···HN hydrogen bond between the N1 atom of dATP and N1H group of dG7 in the template strand (Figure 7). Because dG7 lies opposite to C on the 3′ end of the primer strand (Figure 1), the O3′H group of which serves as an attacking nucleophile in the polymerization reaction, the interaction between dATP and dG7 may affect the rate of the formation of the new PO bond. Although the interactions of the triphosphate moieties in the pol β active site represent a rich platform for the elaboration of their possible kinetic consequences, we feel that understanding of the selectivity due to the chemical step of the DNA replication should be based on calculated activation barriers rather than on the analysis of the ground state structures. Therefore, in this paper we limit

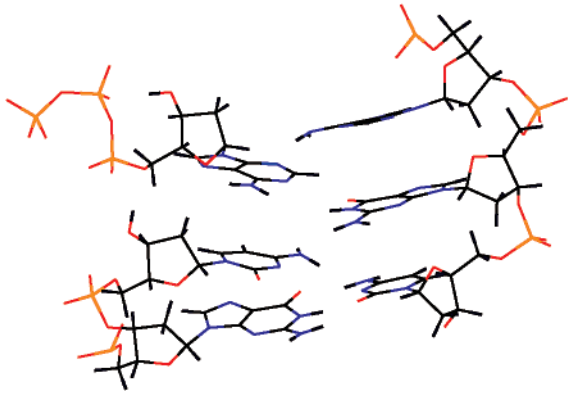


Figure 7. The calculated structure of the neutral dATP·A mismatch in the active site of human polymerase β.

TABLE 3: Group Contributions to the Electrostatic Part of the Relative Binding Free Energy of DATP[a]

residue	dATP·T	dATP·A	dATP·C	dATP·G
Arg283+Glu295[b]	0	1.3	0.3	0.8
Asn279	0	3.2	2.4	2.8
Lys280	0	−0.2	−0.2	−0.6
Asp276+Arg40[b]	0	−0.9	−0.3	0.1
Tyr271	0	1	0.2	0.4
primer	0	−0.5	−0.2	−1
G7	0	−2.2	0.8	−0.1
template	0	7.2	10.3	3.2
Σ[d]	0	8.9	13.3	5.5
pol$\beta - \Sigma$[e]	0	1.2	−1.0	−4.6

[a] Relative energies are given in kcal/mol. The reference state corresponds to dATP opposite to T in the template. [b] Average of the results calculated with and without extra water molecule inserted in the binding pocket. [c] Joint effect of the two listed amino acid residues. [d] Sum of the contributions from the groups listed above. [e] The contribution of the remaining part of the protein + DNA.

ourselves to discussing geometries of base and deoxyribose moieties, while leaving the analysis of triphosphate conformations and interactions for future studies. This is because the base and sugar interactions are directly associated with the main subject of the present study: the energetics of dNTP binding opposite to the "right" and "wrong" template nucleosides.

Binding of dATP opposite to A in the template is disfavored by 11.9 kcal/mol with respect to binding opposite to T (Table 2). At 298 K, this free-energy difference translates into a large ratio of binding constants ($K_{AT}/K_{AA} = 10^9$). This result indicates that the incorporation of A·A mispairs by pol β does not involve the formation of the neutral anti−anti base pair. The most probable alternative is the syn−anti conformer of the neutral A·A base. The involvement of the syn−anti A·A base pairs is supported by the observation of these structures in the crystal structure of RNA double helix.[52]

The displacement of the base moiety of dATP into the major groove also occurs when this substrate is inserted opposite to C in the template strand. Although the resulting wobble base pair is somewhat propeller twisted, the average distance between the closest H atoms of the amino groups on C and dATP is only 2.05 Å. Therefore, it is not surprising that the steric and electrostatic interactions between the substrate and template were found to be the most important factors that determine the lower stability of the neutral dATP·C mismatch in the pol β active site (Table 3).

The A·G mispair may in principle adopt three different forms: A(anti)·G(anti), A(syn)·G(anti), and the protonated A(anti)·G(syn), all of which have been observed in DNA in crystals [53−55] and DNA in solution,[48,56] and as such they can be

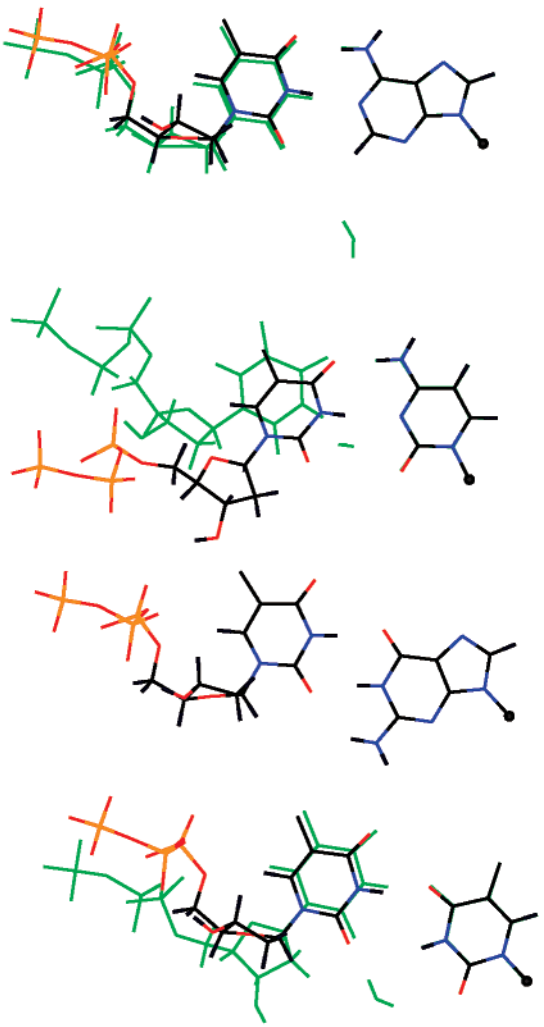


Figure 8. Average structures of the Watson−Crick and mismatched base pairs in the polymerase β active site. dTTP binding opposite to A, C, G, and T (from top to bottom, respectively). The structures obtained for the simulations in the presence and absence of an extra water molecule in the minor groove are drawn using green and atom-type based colors, respectively.

expected to be relatively stable also in the pol β active site. Indeed, the calculated $\Delta\Delta G_{bind}$ for the dATP(anti)·G(anti) base pair (Table 2) is only 1.6 kcal/mol, which corresponds to $K_{AT}/K_{AG} = 15$ binding contribution to the template fidelity of pol β.

3.2. dTTP as Incoming dNTP. The geometry of the dTTP·A base pair, including the deoxyribose conformations, is very similar to the dATP·T base pair discussed above. We found the structure and energetics of the dTTP·A pair to be little perturbed by placing an additional water molecule in the active site (Figure 8, Table 4). This water diffuses away from the substrate during the simulation, and settles instead in the vicinity of the N3 atom of dG7.

A rather small calculated $\Delta\Delta G_{bind}$ (1.6 kcal/mol, Table 2) for the binding of dTTP in the active site containing template C is in part due to the formation of the hydrogen bond between the O2 carbonyl of dTTP and the N1−H group of dG7. As can be seen from Table 5, the interaction between thymine of the substrate and dG7 stabilizes the complex. This interaction is facilitated by a large propeller twist of the dTTP·C base pair. The propeller twist also relieves the electrostatic repulsion between O2 oxygens of dTTP and C bases, while retaining the N−H···N and N−H···O hydrogen bonds. Alternatively, the

TABLE 4: The Comparison of $\Delta\Delta G_{bind}$ and Its Components[a] Calculated from Trajectories Generated in the Presence/Absence of a Water Molecule in the Initial Structure of dNTP Binding Site of Pol β

dNTP	template[b]	$\Delta\Delta G^{ES}{}_Q$	$\Delta\Delta G^{ES}{}_0$	$\Delta\Delta G_{bind,0}$	$\Delta\Delta G_{bind}$
dTTP	A	0.4	−0.1	0.0	0.4
	A(w)	−0.4	0.1	−0.1	−0.4
dTTP	C	−2.0	0.9	1.6	0.4
	Cw	1.1	1.9	−0.2	2.7
dTTP	T	3.2	−0.4	−0.7	2.1
	Tw	0.6	0.2	2.3	3.1
dGTP	T	−0.9	10.8	−1.1	8.8
	Tw	3.8	7.7	0.4	12.0
dCTP	T	2.8	5.0	−1.4	6.4
	Tw	6.9	−0.5	−0.7	5.7

[a] The results (in kcal/mol) are given relative to the Watson−Crick base pair that involves the same dNTP. $\Delta\Delta G^{ES}{}_Q$ and $\Delta\Delta G^{ES}{}_0$ denote the electrostatic components of $\Delta\Delta G_{bind}$ calculated by averaging over trajectories evaluated for the charged and uncharged nucleobase part of dNTP, respectively (eq 24). $\Delta\Delta G_{bind,0}$ denotes the contribution from the binding of an uncharged ligand (eq 23). [b] Symbols A(w), C(w), and T(w) denote structures in which an extra water molecule was present in the minor groove of the dNTP·A, dNTP·C, and dNTP·T base pairs.

TABLE 5: Group Contributions to the Electrostatic Part of the Relative Binding Free Energy of dTTP[a]

residue	dTTP·A[b]	dTTP·C[b]	dTTP·G	dTTP·T[b]
Arg283+Glu295[b]	0	0.6	0.2	−0.4
Asn279	0	1.7	3.8	2.2
Lys280	0	−0.6	0.1	−2
Asp276+Arg40[b]	0	−1.1	−4.8	−0.4
Tyr271	0	0.1	0.0	0.6
primer	0	0.8	0	0.5
G7	0	−2.6	−1.3	−0.6
template	0	1.3	1.8	4.3
Σ^d	0	0	0	4.0
pol $\beta - \Sigma^e$	0	0.9	1.5	−2.2

[a] Relative energies are given in kcal/mol. The reference state corresponds to dTTP opposite to A in the template. [b] Average of the results calculated with and without extra water molecule inserted in the binding pocket. [c] Joint effect of the two listed amino acid residues. [d] Sum of the contributions from the groups listed above. [e] The contribution of the remaining part of the protein + DNA.

O2···O2 repulsion is relieved by widening of the dTTP·C base pair (Figure 8) observed in the simulation with a water molecule inserted in the minor groove, However, this widening prevents the formation of the stabilizing hydrogen bond between dTTP and dG7 bases.

The wobble dTTP·G base pair is the most stable mismatch obtained in this study. This finding is consistent with the stability of the T·G wobble base pair in DNA duplexes.[57,58] The calculated $\Delta\Delta G_{bind}$ of 0.9 kcal/mol includes a significant negative contribution from the combined electrostatic effect of the Asp276 and Arg40 residues (Table 5, Figure 5). We present here the overall contribution of the ion-pair rather than separate contributions from each charged residue because contributions of charged groups forming an ion pair have always large magnitudes and the opposite signs.

A wobble base pairing associated with a small displacement of dTTP base into the major groove is also characteristic for the dTTP·T mispair. Although the extra water molecule in the minor groove does not affect the geometry and binding energy of this mismatch significantly (Figure 8, Table 4), it induces a change of the conformation of deoxyribose of dTTP from C3′-endo to C2′-endo.

3.3. dGTP as Incoming dNTP. Our analysis for dGTP substrate is limited to the dGTP·C, dGTP·A and dGTP·T pairing

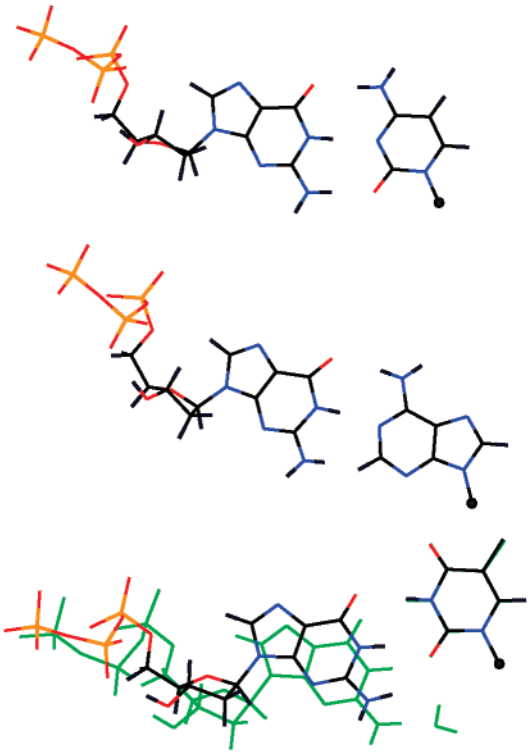


Figure 9. Average structures of the Watson−Crick and mismatched base pairs in the polymerase β active site. dGTP binding opposite to C, A, and T (from top to bottom, respectively). The structures of the dGTP·T base pair were obtained with simulations in the presence (green) or absence (atom-type based colors) of an extra water molecule in the binding pocket.

(Figure 9) because our simulations of the dGTP(anti)·G(anti) mispair in the pol β active site repeatedly failed due to large steric clashes between the bases. Thus, the dGTP·G base pair is likely to contribute to the polymerase misinsertion frequency only in its dGTP(anti)·G(syn) or dGTP(syn)·G(anti) forms that exceed the scope of the present paper.

The reference dGTP·C base pair is nearly planar, with the lengths of the three hydrogen bonds in the 2.8−2.9 Å range. The puckering of the deoxyribose moieties (i.e., C3′-endo for the dNTP and primer, and C2′-endo for the template) parallels puckering states of all other Watson−Crick base pairs studied in the present work.

The dGTP·A base pair, which is somewhat buckled and propeller twisted, interacts favorably with the Arg283·Glu285 ion pair (Table 6). In addition, the $-NH3^+$ group of Lys280, which forms ion-pair with the template PO_2^- group in other binding complexes studied here, undergoes a large conformation change that allows it to come to 3.5 Å from the carbonyl oxygen of dGTP, and thus to stabilize the dGTP·A mismatch. Despite this stabilization by Lys280, $\Delta\Delta G_{bind}$ energy of the dGTP·A mismatch is 7.4 kcal/mol (Table 2). This result characterizes the neutral dGTP(anti)·A(anti) mismatch as an improbable contributor to the polymerase misinsertion frequency. A more stable dGTP·A mispairs in the pol β active site may involve A(anti)·G(anti), A(syn)·G(anti), or the protonated A(anti)·G(syn) mismatches, whose binding energetics has been left for future studies.

The dGTP·T mismatch forms a wobble base pair, in which the dGTP base is displaced into the minor groove. The insertion of a water molecule into the minor groove increases this displacement and slightly stabilizes the mispair. However, the dGTP·T mismatch is greatly destabilized with respect to the reference dGTP·C base pair due to the interactions between dNTP and template bases (Table 6).

TABLE 6: Group Contributions to the Electrostatic Part of the Relative Binding Free Energy of dGTP[a]

residue	dGTP·C	dGTP·A	dGTP·T[b]
Arg283+Glu295[b]	0	−6.2	−2.5
Asn279	0	0.6	0
Lys280	0	−5.4	0.9
Asp276+Arg40[b]	0	0.9	−2.2
Tyr271	0	0.0	0.7
primer	0	−0.5	−0.3
G7	0	−0.1	−1
template	0	14.6	17.5
Σ[d]	0	4	13
pol $\beta - \Sigma$[e]	0	3.4	−2.2

[a] Relative energies are given in kcal/mol. The reference state corresponds to dGTP opposite to C in the template. [b] Average of the results calculated with and without extra water molecule inserted in the binding pocket. [c] Joint effect of the two listed amino acid residues. [d] Sum of the contributions from the groups listed above. [e] The contribution of the remaining part of the protein + DNA.

3.4. dCTP as Incoming dNTP. The reference dCTP·G base pair is slightly propeller-twisted (∼10°) while maintaining short hydrogen bonds (2.8−2.9 Å). Because all the neutral anti−anti mispairs involving the dCTP substrate show no stabilizing dCTP·template contacts (Figure 10), the direct template·dCTP interactions contribute about 20 kcal/mol to the relative stability of the dCTP·G base pair. This contribution is largely offset by the stabilization of mismatches by the water molecule inserted in the active site. In addition, strong hydrogen bonding between dCTP and G7 plays a significant role for the dCTP·C and dCTP·T mispairs (Table 7), and thus a somewhat surprising stability of the dCTP·C mispair ($\Delta\Delta G_{bind} = 3.5$ kcal/mol, Table 2) may be a sequence-dependent rather than a general feature.

The presence of a minor groove water molecule is essential for the stability of the MD trajectories of the ternary complex of pol β containing C in the template strand opposite to dCTP. However, the minor groove water molecule does not serve as a stabilization factor for the formation of the dCTP·C base pair, but rather hydrates the Watson−Crick edge of the template base after the substrate base becomes hydrogen bonded with the G7 nucleotide (Figure 11). Unlike the interactions between dATP and G7, or interaction of dTTP with G7, which involved a twisted substrate and the formation of a single hydrogen bond with G7 (Figure 7), the dCTP inserted opposite to C forms a regular Watson−Crick base pair with G7. This base pair is slightly buckled, and the lengths of its three hydrogen bonds fall in the 2.9−3.2 Å range. The formation of the dCTP·G7 base pair is accompanied with the loss of the hydrogen bonding interactions between the terminal base of the primer and G7. This loss is partly compensated by hydrogen bonding between the primer and the OH group of Tyr271, enabled by twisting of the 3-terminal primer base into the minor groove. The C3′-endo conformation of the primer, which is typical for a correct dNTP binding (see above), is retained despite this twist. On the other hand, the dCTP sugar switches its conformation from C3′-endo, calculated for the binding of dCTP opposite to G, to C2′-endo. A reasonable alternative to the neutral dCTP(anti)·C(anti) mismatch might be the hemiprotonated dCTP(anti)·C(anti) wobble base pair with the proton shared between the O2 and N3 atoms of dCTP and C, respectively. The calculations of the protonated CC mismatch are left for future studies.

The neutral dCTP·A mismatch involves the dCTP base fully displaced into the major groove. The N1 atom of the template base is hydrated by a water molecule. The deoxyribose of dCTP

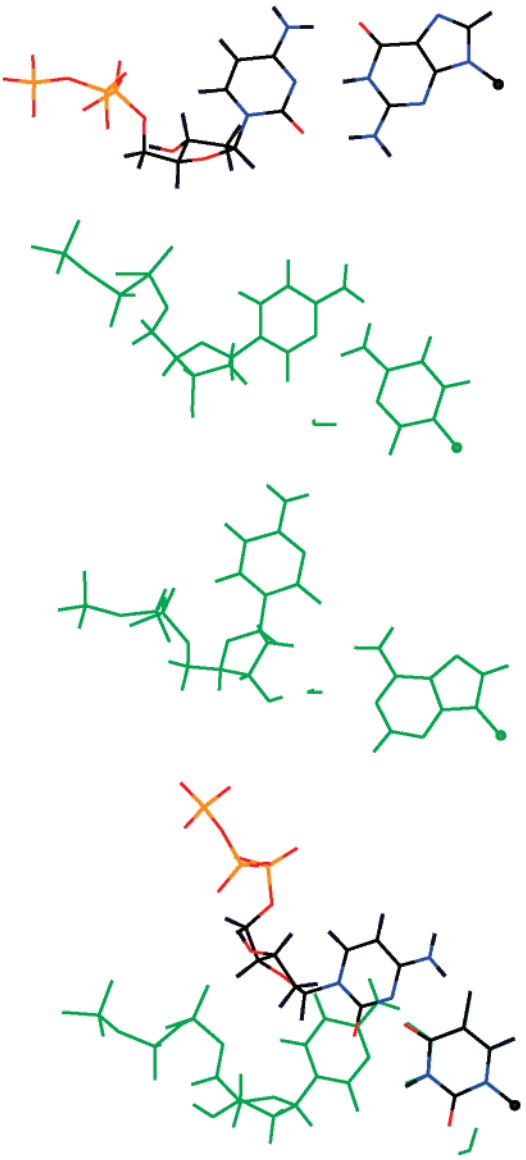


Figure 10. Average structures of the Watson−Crick and mismatched base pairs in the polymerase β active site. dCTP binding opposite to G, C, A, and T (from top to bottom, respectively). The structures obtained for the simulations in the presence and absence of a water molecule in the minor groove are drawn using green and atom-type-based colors, respectively.

attains the C2′-conformation, which is probably induced by the presence of a hydrogen bond between the O3′H hydroxyl group and the oxygen atom of the water molecule. The template and primer have C2′-endo and C3′endo sugars, respectively, which are most common throughout this study. Given the large calculated $\Delta\Delta G_{bind}$ energy, we expect the alternative binding of dCTP opposite to the N1-protonated adenosine as a more viable structure contributing to the polymerase misinsertion frequency for C·A mismatches.

The structure of the dCTP·T mispair is similar to the dCTP·C mispair in that the dCTP forms hydrogen bonds with G7. The resulting Watson−Crick pair is nearly planar, with three hydrogen bonds about 2.9 Å long. The puckering modes for the deoxyriboses are C2′-endo for the primer and the template, and C3′-endo for dCTP. The primer base loses all the hydrogen bonding interactions with G7. These interactions are replaced by hydrogen bonding with Tyr271 and a water molecule. The template base is stabilized by the interaction of its O4 oxygen

TABLE 7: Group Contributions to the Electrostatic Part of the Relative Binding Free Energy of dCTP[a]

residue	dCTP·G	dCTP·C[f]	dCTP·A[f]	dCTP·T[b]
Arg283+Glu295[b]	0	3.4	−1.4	2.9
Asn279	0	2.2	3.6	1.5
Lys280	0	0.9	−3.3	1.2
Asp276+Arg40[b]	0	−1.5	−0.9	−1.3
Tyr271	0	0.5	1.8	1.8
primer	0	−1.9	−3.3	−2.8
G7	0	−13	2.6	−8.9
template	0	24.5	20.6	19.9
Σ[d]	0	15.1	19.7	14.3
pol β − Σ[e]	0	−11.6	−11	−7.1

[a] Relative energies are given in kcal/mol. The reference state corresponds to dCTP opposite to G in the template. [b] Average of the results calculated with and without extra water molecule inserted in the binding pocket. [c] Joint effect of the two listed amino acid residues. [d] Sum of the contributions from the groups listed above. [e] The contribution of the remaining part of the protein + DNA. [f] Calculated with a water molecule inserted in the binding pocket in order to stabilize the mispair.

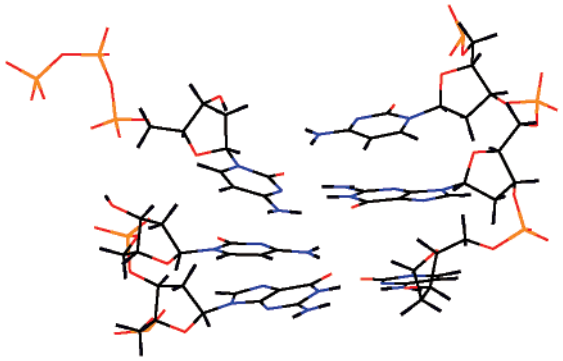


Figure 11. The calculated structure of the neutral dCTP·C mismatch in the active site of human polymerase β.

with the amino group of Asn279, and its N3−H group with a water molecule positioned in the minor groove. However, in contrast with the dCTP·C mismatch, for which the water molecule in the minor groove was inserted manually, the water molecule interacting with T in the dCTP·T mispair came to the vicinity of T during the 1.6 ns equilibration simulation. The diffusion of this water molecule occurred along a distance of about 5 Å defined as a length of a straight line connecting the initial and final position of the molecule. Surprisingly, the separate simulation of the dCTP·T mispair that was started with a water molecule inserted in the minor groove close to the template did not lead to the interaction between dCTP and G7, but rather to the formation of the wobble dCTP·T base pair. The $\Delta\Delta G_{bind}$ energies calculated with and without an extra water molecule are similar (5.7 and 6.4 kcal/mol, respectively), whereas their components differ significantly (Table 4).

3.5. Stability of the Calculated Energetics. The computational protocol used in this study involved a long equilibration period of 1.66 ns to ensure that the protein structure was given sufficient time to relax after the original crystal structure (containing the dCTP·G base pair) was modified by incorporating another dNTP·template base pair. The convergence of positions of active site residues during the equilibration period was studied by monitoring their electrostatic contributions to $\Delta\Delta G_{bind}$ for different simulation windows (Table 8). The results are reasonably converged for mispairs containing dATP or dTTP. However, the convergence is slower for mispairs containing dCTP or dGTP as incoming dNTP. Here, the largest change occurs for the dGTP·A mispair upon going from the

TABLE 8: Variation of the Sum of Contributions of Active Site Residues[a] to the Electrostatic Part of Relative Binding Free Energy ($\Delta\Delta G^{ES}_{Q}$, kcal/mol) along the 1.97 ns MD Trajectory of the Ternary Complex of DNA Polymerase β[a]

		$\Delta\Delta G^{ES}_{Q}$ (kcal/mol)[b]			
substrate	template	110−270 ps[c]	270−770 ps[c]	770−1770 ps[c]	1770−1970 ps[c]
dATP	A	4.7	4.5	4.8	6.2
	C	4.8	6.2	5.3	5.4
	G	2.3	0.5	2.3	1.9
dTTP	C	2.6	2.0	1.5	0.8
	G	0.5	0.4	1.4	−0.7
	T	4.6	5.0	4.7	3.5
dGTP	A	4.7	5.0	−0.4	−1.1
	T	5.2	5.1	4.6	3.1
dCTP	C	10.7	12.0	10.7	7.8
	A	15.9	15.1	13.6	12.5
	T	8.4	7.5	5.5	5.8

[a] Arg283 + Glu295 + Asn279 + Lys280 + Asp276 + Arg40 + Tyr271 + dC352(primer) + dG7 + template (Figure 5). [b] The results (see eq 24b) are given relative to the Watson−Crick base pair, which involves the same dNTP. [c] The electrostatic energies were sampled in the given time period along a continuous 1970 ps MD trajectory.

270−770 ps to 770−1770 ps window. This drop in $\Delta\Delta G^{ES}_{Q}$ is associated with global reorientation of the side chains of the Arg40 and Lys280 residues. In general, the longer simulation times seem to lead to more stable mispairs containing dCTP or dGTP substrates. Because the calculated magnitudes of $\Delta\Delta G_{bind}$ (around 10 kcal/mol, Table 2) for these mispairs are clearly too large, the extension of the simulation time beyond 2 ns would yield more realistic values of binding contributions to the polymerase fidelity. Unfortunately, the required total simulation times of about 10 ns are practically unattainable with the present-time supercomputer resources.

Another important technical issue involves the question of whether the standard, nonpolarizable, molecular-mechanics force fields are sufficiently robust for realistic studies of protein−ligand binding. Because of the presence of multiple charged groups (e.g., two magnesium ions and phosphate groups) in the polymerase active site the importance of including atomic polarizabilities on the DNA and protein atoms might turn out to be significant. To address this issue, we carried out LRA simulations of the pol β ternary complexes containing dGTP·C and dGTP·T base pairs with the ENZYMIX force field that allows one to turn the polarizable component of the force field (induced dipoles) on and off. The effect of the presence of induced dipoles was a change in the $\Delta\Delta G_{bind}$ by −1.2, 2.5, and 1.1 kcal/mol in 30, 60, and 100 ps trajectories, respectively. Given that the overall $\Delta\Delta G_{bind}$ for the dGTP·T mispair is 10.4 kcal/mol (Table 2), the calculated changes in this energy due to induced dipoles are quite small. Thus, at present the benefits of the use of more rigorous nonaditive (polarizable) force fields seem to be partially offset by the sampling limitations caused by larger CPU-time demands (about a factor of two) of these force fields.

4. Discussion

The preference of the DNA polymerase for synthesizing DNA containing W−C base pairs is determined by the energetics of binding and bond-formation steps. The selectivity of both the binding and chemical steps are known to make significant contributions to the polymerase fidelity[4,6,59] These two steps are formally separable as long as the dNTP·DNA·pol β ternary complex corresponds to a minimum on the free-energy surface. The existence of such a minimum is evidenced by the stability of the ternary complex in the crystalline state. Thus, it is reasonable to present calculations that address only the binding contribution to the polymerase fidelity, while leaving the study of the catalytic step for future studies (J. Florian, M. F., Goodman, and A. Warshel, unpublished, see also refs 13 and 14. Note that the polymerase insertion fidelity is given by the ratio of k_{cat}/K_m values for right and wrong substrates.[60] The binding energy is related in part to K_m, and its contribution to the overall fidelity stems from the greater affinity of right compared to wrong dNTP substrates for the polymerase−primer/template DNA complex.[8,9]

We have investigated the binding contribution to the polymerase fidelity at an atomic level of detail determined by the crystal structure of the pol β ternary complex. Starting from this structure, we considered explicitly interatomic forces between all the pairs of atoms in the 18 Å sphere centered on the dNTP base, and the evolution of atomic positions and the corresponding forces in time. These calculations were carried out with the goal to collect sufficient information to determine how the free energy of the entire ternary complex changes upon replacing the W−C pair between dNTP and template with a mismatched pair.

Unfortunately, the calculations of the free-energy change by sampling the potential energy of the whole system are unstable in the sense that they require averaging over very long time periods for reaching a reasonable convergence. This inherent instability makes global algorithms of free-energy calculations, among them the free-energy perturbation (FEP) or thermodynamic integration (TI) methods, impractical for the study of the polymerase fidelity. An instructive example of the performance and limitations of the FEP method is our calculation of the stability of the neutral GT and AC mispairs in duplex DNA.[21] Therefore, we resorted to an approximation that determines the free energy by monitoring the potential energy only for the nucleobase moiety of dNTP substrate. This part of the substrate becomes a "probe", whose average interaction with the surrounding atoms is proportional to the free energy of the whole ternary complex.

The viability of this reduction in the number of sampled interactions depends on the validity of the linear response approximation (LRA), as formulated and rationalized in the Methods section. In brief, the LRA approximation assumes that the average energetic cost of the reorganization of the environment, induced by turning on the charges on the probe, is one-half of the change in the average electrostatic interaction energy between the probe and its environment. If the environment is randomly oriented relative to the probe then $\Delta\Delta G^{ES}_{0} = 0$, and the $\Delta\Delta G^{ES}_{Q}$ of Tables 2−7 produces the entire free energy change. However, the environment may be already preorganized (for example, by specific protein folding, or in the case of DNA polymerase, due to the presence of the templating base. In this case, the LRA approximates the contribution of the preorganized environment to the total binding free energy using the $\Delta\Delta G^{ES}_{0}$ term (eq 24c). This term is evaluated by calculating atomic positions and interatomic forces of the ternary complex with zero charges on the probe atoms. In this simulation, the protein and DNA atoms do not adapt their geometry to the electrostatic presence of the probe. If the geometries from such a simulation are used to calculate electrostatic interaction between the fully charged probe and its environment, which is exactly how the $\Delta\Delta G^{ES}_{0}$ term is calculated, the corresponding interaction energy will reflect the stabilization or destabilization of the probe by the preorganization of the environment. The $\Delta\Delta G^{ES}_{0}$ term is formally an electrostatic energy. However, because the orienta-

tion of the probe in the active site during the calculation of $\Delta\Delta G^{ES}{}_0$ is determined only by its van der Waals interactions with the active site, the geometry (steric) of the active site is implicitly taken into account in the magnitude of the $\Delta\Delta G^{ES}{}_0$ term.

Inspection of the results presented in Table 2 reveals that the calculated $\Delta\Delta G_{bind}$ is dominated by the electrostatic terms. Also the nonelectrostatic contribution ($\Delta\Delta G_{bind,0}$), which is approximated here by the scaled van der Waals interaction energies, usually stabilizes the W−C base pairs, but this contribution is always smaller than the contribution from the electrostatic interactions. In fact, there are some mismatches (dTTP·G, dGTP·T, and dCTP·T) that are slightly stabilized by $\Delta\Delta G_{bind,0}$ term. The simple electrostatic term ($\Delta\Delta G^{ES}{}_Q$) contributes more to the overall $\Delta\Delta G_{bind}$ than $\Delta\Delta G_{bind,\phi}$, but taking the magnitude of $\Delta\Delta G^{ES}{}_Q$ as the only determinant of the binding fidelity would lead to preferential incorporation of the dTTP·C and dCTP·C mismatches in the replicated DNA. Similarly, the contribution of the protein and template preorganization alone ($\Delta\Delta G^{ES}{}_0$) would give a preference to dTTP·T mismatch over the corresponding W−C pair, and predicts the dATP·G mispair to be equally stable as the dATP·T pair. Thus, it is essential to consider all three components to the binding free energy (eq 25) in order to reproduce the binding fidelity of pol β. It should be noted that many LRA models for calculations of binding free energies in proteins[61−64] neglect or overlook the preorganization contribution ($\Delta\Delta G^{ES}{}_0$) (see also discussion in ref 24). Although this simplification might sometimes be justified, in the present case the preorganization contribution is very significant and cannot be neglected.

Considering only anti−anti conformers, the mispairs most likely to be bound in the polymerase β active site are dATP·G, dTTP·C, and dTTP·G. Of these, dTTP·C may lead to a frameshift mutation rather than to substitution mutation because of a strong interaction between dTTP and G7 (Table 5). However, a much stronger tendency for frameshifting is predicted for the dCTP·C (Figure 11) and dCTP·T mismatches if the template base is preceded by G. This is because the formation of a displaced mispair illustrated in Figures 7 and 11 seems to represent a plausible structural basis for the generation of +1 frameshift mutations by pol β. Although a few +1 frameshift mutations are indeed observed during in vitro studies of pol β fidelity, −1 frameshifts occur much more frequently.[65] Obviously, predictions based solely on the binding calculations must be viewed as tentative until the chemical step and the product structure are also taken into account. Therefore, additional calculations of changes in the activation barrier due to the formation of mispairs are required for elucidation of the determinants of the overall fidelity of pol β.

One of the virtues of the LRA calculations is that the calculated $\Delta\Delta G_{bind}$ can be decomposed into the contributions of individual protein or DNA residues. An overview of the decomposition for nearby protein and DNA groups is presented in Figure 12. Here we note that among protein residues, Asn279 contributes most significantly to pol β fidelity. The ability of this group to destabilize all 15 mispairs is truly remarkable. In connection with a possible use of our results to guide protein engineering efforts, it should be noted that our group contribution analysis reflects only direct electrostatic effects of a given group and not its structural effects. Nevertheless, we suggest the mutation of Asn279 to Gln as a good candidate for a mutation that decreases the fidelity of pol β, especially for transversion and transition mutations involving A·T and T·A pairs. An important role in increasing fidelity of the binding

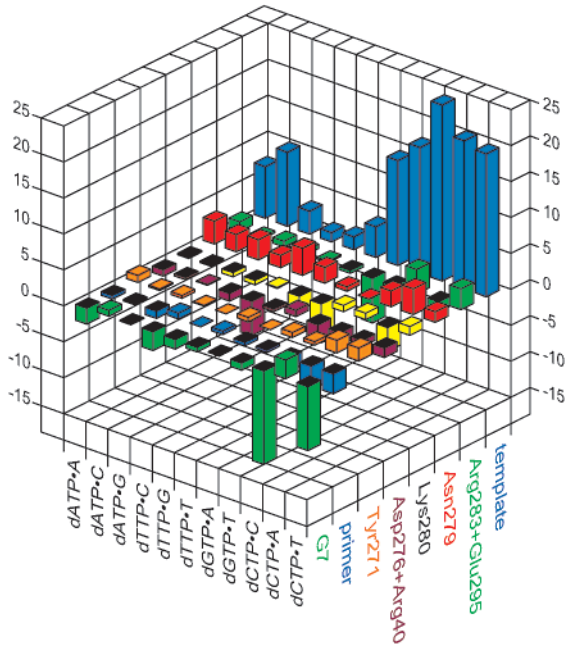


Figure 12. Group contributions to $\Delta\Delta G_{bind}$. For approximate positions of the studied groups in the active site of pol β, see Figures 1 and 5.

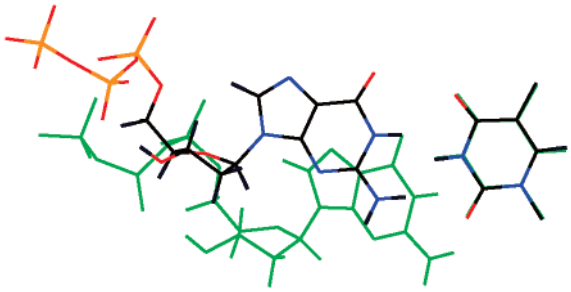


Figure 13. The structure of the dGTP·T base pair obtained by averaging over the trajectory calculated with uncharged dGTP base in the active site of pol β. The results of the simulations with or without extra water molecule in the minor groove are drawn with green and atom-based colors, respectively.

step is played also by the Tyr271 residue. However, for the DNA sequence studied here, Tyr271 has also a structural role in stabilizing unpaired primer and thus in facilitating the interactions of dNTP (especially dCTP) with G7, where G7 is the 3′-neighbor of the templating base (Figure 1). These interactions might promote the generation of +1 frameshift mutations discussed in the preceding paragraph.

Our calculations clearly associate the binding contribution to the pol β fidelity with the interactions between the incoming dNTP substrate and the template. This template contribution is significantly larger for mutation of the G·C or C·G than the A·T or T·A base pairs. This is not surprising in view of much larger electrostatic attraction between the bases in G·C compared to the A·T base pairs. The magnitude of the electrostatic attraction between bases in the gas-phase can be rigorously determined from the basis-set corrected Hartree−Fock interaction energy obtained by quantum mechanical calculations.[66] These calculations indicate that W−C G·C base pairs are electrostatically about 14 kcal/mol more stable than A·T pairs.

Turning attention to the magnitudes of the destabilization of various mismatches due to dNTP-template interactions, it is instructive to compare contributions presented in Figure 12 and Tables 3−7, with the gas-phase interaction energies calculated using the AMBER force field.[67] Here we find that typically the

energies in Figure 12 are larger than then their gas-phase counterparts. For example, the calculated template contribution to the destabilization of the dGTP·T mismatch relative to the dGTP·C pair is 17.5 kcal/mol (Table 6) whereas the corresponding gas-phase destabilization is only 12.4 kcal/mol (cf. A41 column in Table 1 of ref 67). Similarly, the template-driven destabilization energies for the dGTP·A, dTTP·C, and dTTP·T mismatches (14.6, 1.3 and 4.3 kcal/mol) are larger than their gas-phase counterparts (13.3, 0.7 and 0.7 kcal/mol,[66d] respectively). The explanation for such large template contributions is that the preorganized pol β active site keeps the template base optimally positioned for Watson−Crick pairing, while forcing mismatched pairs to attain configurations that are far from their ideal gas-phase geometries. This is especially true of the template contributions calculated using simulations with uncharged nucleobase. These template contributions result in large magnitudes of the $\Delta\Delta G^{ES}_0$ term (Table 2). For example, the geometry of the dGTP·T base pair in the pol β active site, which was obtained by averaging over the configurations of the uncharged substrate trajectory, shows a structure that is close to the shape of the Watson−Crick base pair (Figure 13). Notably, the introduction of an extra water molecule in the active site helps to maintain a wobble geometry of the dGTP·T pair by perturbing the steric environment of the pol β active site (Figure 13). The LRA method with its explicit preorganization term is essential theoretical tool for capturing this preorganization effect and for transforming the geometric concepts into free-energy contributions.

Acknowledgment. This work was supported by the NIH grants GM21422 (to M.F.G.) and GM24492 (to A.W.).

References and Notes

(1) Kornberg, A.; Baker, T. A. *DNA Replication*; W. H. Freeman: New York, 1992.

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

(3) 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.

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

(5) Zhong, X. J.; Patel, S. S.; Werneburg, B. G.; Tsai, M. D. *Biochemistry* **1997**, *1997*, 11891.

(6) Johnson, K. A. *Annu. Rev. Biochem.* **1993**, *32*, 685.

(7) Arndt, J. W.; Gong, W.; Zhong, X.; Showalter, A. K.; Liu, J.; Dunlap, C. A.; Lin, Z.; Paxson, C.; Tsai, M.-D.; Chan, M. K. *Biochemistry* **2001**, *40*, 5368.

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

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

(10) Davies, J. F.; Almassy, R. J.; Hostomska, Z.; Ferre, R. A.; Hostomsky, Z. *Cell* **1994**, *76*, 1123.

(11) Pelletier, H.; Sawaya, M. R.; Kumar, A.; Wilson, S. H.; Kraut, J. *Science* **1994**, *264*, 1891.

(12) Sawaya, M. R.; Pelletier, H.; Kumar, A.; Wilson, S. H.; Kraut, J. *Science* **1994**, *264*, 1930.

(13) Abashkin, Y. G.; Erickson, J. W.; Burt, S. K. *J. Phys. Chem. B* **2001**, *105*, 287.

(14) Fothergill, M.; Goodman, M. F.; Petruska, J.; Warshel, A. *J. Am. Chem. Soc.* **1995**, *117*, 11619.

(15) Srivastava, D. K.; Vande Berg, B. J.; Prasad, R.; Molina, J. T.; Beard, W. A.; Tomkinson, A. E.; Wilson, S. H. *J. Biol. Chem.* **1998**, *273*, 21203.

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

(17) Kollman, P. A. *Chem. Rev.* **1993**, *93*, 2395.

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

(19) Straatsma, T. P.; McCammon, J. A. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407.

(20) Warshel, A.; Sussman, F.; Hwang, J.-K. *J. Mol. Biol.* **1988**, *201*, 139.

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

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

(23) Lee, F. S.; Chu, Z. T.; Bolger, M. B.; Warshel, A. *Prot. Eng.* **1992**, *5*, 215.

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

(25) Marcus, R. A.; Sutin, N. *Biochim. Biophys. Acta* **1985**, *811*, 265.

(26) Warshel, A.; Parson, W. W. *Q. Rev. Biophy.* **2001**, *34*, 563.

(27) Zhou, H. X.; Szabo, A. *J. Chem. Phys.* **1995**, *103*, 3481.

(28) King, G.; Warshel, A. *J. Chem. Phys.* **1990**, *93*, 8682.

(29) Florián, J.; Warshel, A. *J. Phys. Chem. B* **1999**, *103*, 10282.

(30) 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.

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

(32) Cheatham, T. E., III; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4193.

(33) Cubero, E.; Sherer, E. C.; Luque, F. J.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **1999**, *121*, 8653.

(34) Jayaram, B.; Sprous, D.; Young, M. A.; Beveridge, D. L. *J. Am. Chem. Soc.* **1998**, *120*, 10629.

(35) Spackova, N.; Berger, I.; Egli, M.; Sponer, J. *J. Am. Chem. Soc.* **1998**, *120*, 6147.

(36) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. *Acc. Chem. Res.* **2000**, *33*, 889.

(37) Warshel, A.; Creighton, S., In *Computer Simulation of Biomolecular Systems*; van Gunsteren, W. F., Wiener, P. K., Eds.; ESCOM: Leiden, 1989; p 120.

(38) *SYBYL 6.6*; Tripos, Inc.: 1699 South Hanley Rd., St. Louis, MO, 63144, 1999.

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

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

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

(42) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian94, revision D.2*; Gaussian, Inc.: Pittsburgh, 1995.

(43) Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, *14*, 161.

(44) Boosalis, M. S.; Mosbaugh, D. W.; Hamatake, R.; Sugino, A.; Kunkel, T. A.; Goodman, M. F. *J. Biol. Chem.* **1989**, *264*, 11360.

(45) Chagovetz, A. M.; Sweasy, J. B.; Preston, B. D. *J. Biol. Chem.* **1997**, *272*, 27501.

(46) Werneburg, B. G.; Ahn, J.; Zhong, X.; Hondal, R. J.; Kraynov, V. S.; Tsai, M.-D. *Biochemistry* **1996**, *35*, 7041.

(47) Sowers, L. C.; Shaw, B. R.; Veigl, M. L.; Sedwick, W. D. *Mutat. Res.* **1987**, *177*, 201.

(48) Gao, X.; Patel, D. *J. Am. Chem. Soc.* **1988**, *110*, 8.

(49) Hunter, W. N.; Brown, T.; Kennard, O. *Nature* **1986**, *320*, 552.

(50) It should be noted that because at present the most theoretical calculations of complex biological systems do not give the right answers, the choice of the appropriate studied system is as important as the choice of the experimental conditions in the laboratory (for example, the choice of the right conditions for the crystallization of proteins). Some conditions give stable results whereas others do not. In our case, the conditions for facile experimental studies (substrate fidelity) and computer simulations (template fidelity, neutral anti−anti base pairs) differ. This does not mean, however, that either the experimental or theoretical results should be considered incapable of addresing important fidelity issues just because they are not complete enough.

(51) Kiefer, J. R.; Mao, C.; Braman, J. C.; Beese, L. S. *Nature* **1998**, *391*, 304.

(52) Baeyens, K. J.; Bondt, H. L. D.; Pardi, A.; Holbroo, S. R. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 12851.

(53) Prive, G. G.; Heinemann, U.; Chandrasegaran, S.; Kan, L. S.; Kopka, M. L.; Dickerson, R. E. *Science* **1987**, *238*, 498.

(54) Brown, T.; Hunter, W. N.; Kneale, G.; Kennard, O. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 2402.

(55) Brown, T.; Leonard, G. A.; Booth, E. D.; Chambers, J. *J. Mol. Struct.* **1989**, *207*, 445.

(56) Fazakerley, G. V.; Quignard, E.; Woisard, A.; Guschlbauer, W.; van der Marel, G. A.; van Boom, J. H.; Jones, M.; Radman, M. *EMBO J.* **1986**, *5*, 3697.

(57) Allawi, H. T.; SantaLucia, J., Jr. *Biochemistry* **1997**, *36*, 10581.

(58) Kennard, O. *J. Biomol. Struct. Dyn.* **1985**, *3*, 205.
(59) Carroll, S. S.; Benkovic, S. J. *Chem. Rev.* **1990**, *90*, 1291.
(60) Fersht, A. R. *Enzyme Structure and Mechanism*; W. H. Freeman and Company: New York, 1985.
(61) Åqvist, J.; Medina, C.; Samuelson, J. E. *Protein Eng.* **1994**, *7*, 385.
(62) Hansson, T.; Marelius, J.; Åqvist, J. *J. Comput.−Aided Mol. Des.* **1998**, *12*, 27.
(63) Paulsen, M. D.; Ornstein, R. L. *Protein Eng.* **1996**, *9*, 567.
(64) Jones-Hertzog, D. K.; Jorgensen, W. L. *J. Med. Chem.* **1997**, *41*, 55272.
(65) Kunkel, T. A. *J. Biol. Chem.* **1985**, *260*, 5787.
(66) (a) Hobza, P.; Sandorfy, C. *J. Am. Chem. Soc.* **1987**, *109*, 1302. (b) Hrouda, V.; Florian, J.; Hobza, P. *J. Phys. Chem.* **1993**, *97*, 1542. (c) Florián, J.; Leszczynski, J.; Scheiner, S. *Mol. Phys.* **1995**, *84*, 469. (d) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965.
(67) Hobza, P.; Kabelac, M.; Sponer, J.; Mejzlik, P.; Vondrasek, J. *J. Comput. Chem.* **1997**, *18*, 1136.
(68) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Solution Chem.* **1981**, *10*, 563.
(69) Wolfenden, R. *Science* **1983**, *222*, 1087.