Simulating the Effect of DNA Polymerase Mutations on Transition-State Energetics and Fidelity: Evaluating Amino Acid Group Contribution and Allosteric Coupling for Ionized Residues...

ARTICLE *in* BIOCHEMISTRY · JULY 2006

Impact Factor: 3.02 · DOI: 10.1021/bi060147o · Source: PubMed

5 AUTHORS, INCLUDING:

Yun Xiang
East China Normal University
18 PUBLICATIONS **1,061** CITATIONS
SEE PROFILE

Jan Florián
Loyola University Chicago
95 PUBLICATIONS **4,146** CITATIONS
SEE PROFILE

Arieh Warshel
University of Southern California
197 PUBLICATIONS **14,136** CITATIONS
SEE PROFILE

Simulating the Effect of DNA Polymerase Mutations on Transition-State Energetics and Fidelity: Evaluating Amino Acid Group Contribution and Allosteric Coupling for Ionized Residues in Human Pol β†

Yun Xiang,‡ Peter Oelschlaeger,‡ Jan Florián,§ Myron F. Goodman,|| and Arieh Warshel*,‡

Departments of Chemistry and Biology, University of Southern California, Seeley G. Mudd 418, 3620 McClintock Avenue, Los Angeles, California 90089-1062, and Department of Chemistry, Loyola University Chicago, 6525 North Sheridan Road, Chicago, Illinois 60626

Received January 23, 2006; Revised Manuscript Received March 29, 2006

ABSTRACT: The control of the catalytic power and fidelity of DNA polymerases involves the complex combined effect of the protein residues, the Mg^{2+} ions, and the interaction between the DNA bases. In an attempt to advance the understanding of catalytic control, we analyze the effect of the protein residues, taking human DNA polymerase β as a model system. Specifically, we examine the ability of different theoretical models to reproduce the effect of ionized residues on the transition state (TS) binding energy and the corresponding k_{pol}/K_D. We also explore the role of the Mg^{2+} ions in the binding and catalysis processes. The application of the microscopic linear response approximation (LRA) and the semimacroscopic PDLD/S-LRA methods to a benchmark of mutational studies produces a semiquantitative correlation and indicates that these methods can provide predictive power. However, pre-steady-state and steady-state kinetic studies currently available do not give a unique benchmark, owing principally to widely varying experimental conditions. We believe that a more uniform experimental benchmark is needed for further refinement of the theoretical models. The analysis of the correlation between the results obtained by a rigorous thermodynamic cycle and by simpler approximations indicates that the protein reorganization between the open, i.e., unbound, form and the closed form does not change the magnitude of the calculated mutational effects in a major way for the experimental data used in this study. The use of the PDLD/S-LRA group contributions allows us to construct energy-based correlation diagrams that can help toward understanding the coupling, i.e., transfer of information, between the base-binding and catalytic sites and to gain a deeper insight into the molecular basis of DNA replication fidelity. Our analysis suggests that the allosteric matrix obtained by subtracting the correlation matrix of the correct and incorrect base pairs should prove useful in exploring the information transfer occurring between the base-binding and catalytic sites. This type of treatment should be especially effective when coupled with structural studies of polymerase−DNA−base mispair ternary complexes and studies using polymerase double mutants. We discuss the potential of direct calculations of binding energy of the TS in a rational design of TS analogues and in drug design.

The reproduction and evolution of life depends upon the accurate replication of the genome, which is facilitated by DNA polymerases (*1*). The synthesis of new DNA molecules would be impossible without these enzymes because they increase the rate of the phosphodiester bond formation by many orders of magnitude compared to the corresponding reaction in water (*2*, *3*). DNA polymerases selectively catalyze the reaction not for a single substrate, similar to most other enzymes, but for four structurally distinct substrates. A right dNTP substrate ($\mathscr{R}$) is selected for insertion by the polymerase depending upon the identity of the templating base so that Watson−Crick base pairs are formed with high efficiency. The fidelity of DNA replication is controlled by the polymerase active site where phosphodiester bond catalysis occurs and by the binding site of the incoming nucleotide paired opposite a template base. The rate of incorporation of an incoming wrong nucleotide ($\mathscr{W}$) is drastically slower than the corresponding rate of the right nucleotide ($\mathscr{R}$), thereby ensuring high replication fidelity, reviewed in refs *4*−*6*.

Despite significant experimental advances in studies of the fidelity of DNA polymerase (e.g., refs *4*−*9*) we still do not have a clear quantitative molecular picture of the energetics of this process. Some progress in this direction has been made in recent theoretical studies, which explored the binding and chemical contributions to fidelity (*10*−*14*) as well as the possible role of conformational changes (*15*) and protein motions (*16*). Progress in these key areas is still at a relatively early stage.

† This work was supported by NIH Grants 5U19CA105010, RO1GM21422 and the HPCC of USC.

* To whom correspondence should be addressed. Telephone: (213) 740-4114. Fax: (213) 740-2701. E-mail: warshel@usc.edu.

‡ Department of Chemistry, University of Southern California.

§ Loyola University Chicago.

|| Departments of Biological Sciences and Chemistry, University of Southern California.

10.1021/bi060147o CCC: $33.50
Published on Web 05/17/2006

Relevant information about the action of DNA polymerases can be provided by mutational studies that reveal the role of specific amino acid residues in the action of DNA polymerases (e.g., refs *17*−*23*). Here, it would be useful to have theoretical models that can help to clarify the role of different residues in catalyzing the reaction for the right nucleotide and reducing the catalysis with the wrong nucleotide. Some progress in this direction has been made in modeling the free-energy contributions of different residues to the binding of correct and incorrect nucleotides (*10*, *12*). Further insight has been provided in a recent study (*16*), which simulated the structural effect of some mutations in DNA polymerase β (pol β)[1] and examined the corresponding electrostatic effects by a qualitative analysis. Additionally, theoretical studies of related problems have been reported recently (*24*−*27*).

In our view, what is clearly missing is a systematic energy-based analysis of the effect of different mutations on the fidelity of the replication process. The ability to reproduce the actual effect of the mutations is a prerequisite for a detailed analysis of the nature of the contribution of each residue. Once this ability is established, we can progress in understanding the effect of individual residues on transition-state (TS) stablization as well as on the effects of coupling between residues.

In the present work, we take a step in the above direction by evaluating the contribution of ionized residues to the energetics of the TS of pol β, which has been the subject of extensive structural (*28*−*31*) and mutational (*17*−*23*) studies. Our work involves an exploration of the ability of different simulation methods to reproduce the observed effects of mutations of ionized residues on k_{pol}/K_D of pol β. Our study indicates that proper implementation of microscopic and semimacroscopic methods can serve as a basis for an iterative process of refining theoretical approaches by feedback from further experimental studies. Last, we introduce an approach for constructing energy-based correlation diagrams that can be used to explore the coupling between protein residues, i.e., the transfer of information between the base-binding site (composed of the template base and the remaining environment of the base of the incoming dNTP) and the catalytic site.

COMPUTATIONAL PROCEDURES

Simulation Systems and Thermodynamic Cycles. Our starting point is the energy diagram presented in Figure 1, which describes the energetics corresponding to k_{pol}/K_D and $\Delta g^{\ddagger}_{enz}$ of the catalytic reaction of DNA polymerases. In the likely situation that the chemical step is rate-limiting, we expect k_{pol}/K_D to determine the overall fidelity (*14*). Our challenge is to determine the contribution of different amino acid residues to $\Delta g^{\ddagger}_{enz}$ and the corresponding TS-binding energy.

The large catalytic power of DNA polymerases is mainly due to the two Mg^{2+} ions that help in stabilizing the ionized form of the nucleophile and the charged TS (*11*, *32*). To quantify the energetics of the TS, it is important to consider

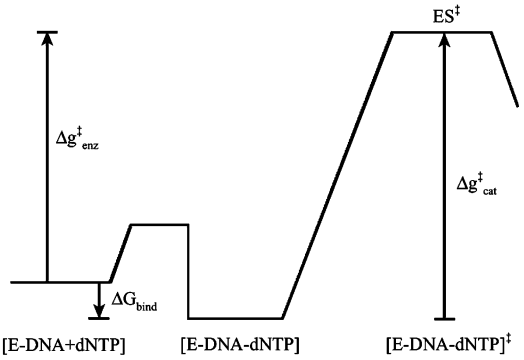

FIGURE 1: Schematic diagram describing the energetics of the reaction of DNA polymerase, for the likely case when the chemical step is rate-determining (see ref *14* for a more general diagram). ΔG_{bind} is the dNTP-binding free energy, which is related to K_D by $\Delta G_{bind} = 2.303RT \log K_D$. $\Delta g^{\ddagger}_{cat}$, the activation free energy of the reaction, is obtained from the experimentally measured catalytic rate constant k_{pol} by $\Delta g^{\ddagger}_{cat} = -2.303RT \log(hk_{pol}/k_BT)$. $\Delta g^{\ddagger}_{enz} = \Delta G_{bind} + \Delta g^{\ddagger}_{cat}$, which is directly related to k_{pol}/K_D, the catalytic efficiency of enzymatic reactions.

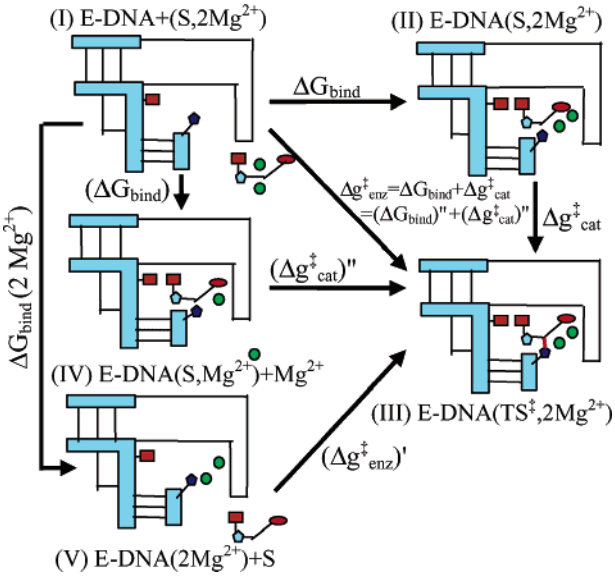

FIGURE 2: Thermodynamic cycles describing the possible ways of reaching the TS in the reaction of DNA polymerase. The cycle and detailed considerations of the binding modes of the Mg^{2+} ions allow us to relate the different approximations used in the present work. E−DNA designates the enzyme−DNA complex; S designates the dNTP substrate.

the modes of binding of these Mg^{2+} ions. Here, one should consider the thermodynamic cycle of the type described in Figure 2, where the initial and final positions of the Mg^{2+} ions are known but the intermediate states are not certain. For example, Bakhtina et al. (*33*) suggested in their recent work that the binding of the second Mg^{2+} occurs after the protein subdomain-closing conformational change but before the rate-limiting step. Thus, our simulation studies of the catalytic power of DNA polymerase will consider the state of the Mg^{2+} ions and the implication of the cycle of Figure 2.

Taking human pol β as a benchmark, we shall consider the effects of mutations of ionized residues in this system. The locations of most of the residues that will be considered are depicted in Figure 3.

Simulation Methods. The main objective of this paper is the evaluation of the effect of mutations of ionized residues

[1] Abbreviations: pol β, DNA polymerase β; MD, molecular dynamics; LRA, linear response approximation; LIE, linear interaction energy; EVB, empirical valence bond; FEP, free-energy perturbation; vdW, van der Waals; TS, transition state.

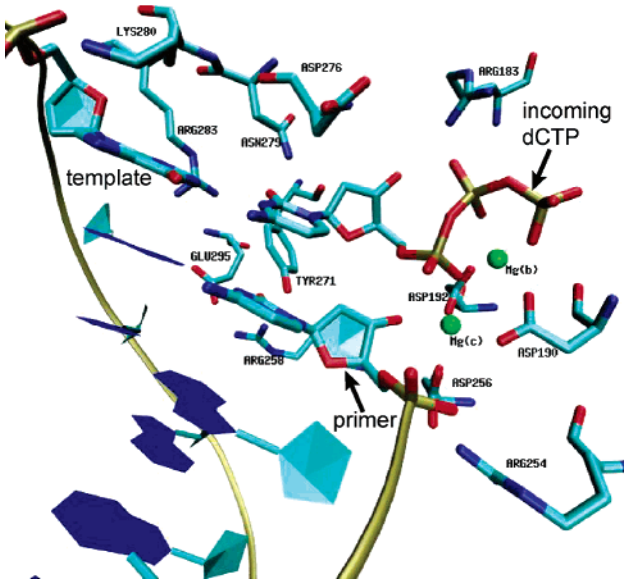

FIGURE 3: The active site of pol β including the dCTP substrate, the two Mg^{2+} ions and their coordinated three Asp residues, and the residues included in our mutational studies. Mg(b), dNTP-binding Mg^{2+} ion; Mg(c), catalytic Mg^{2+} ion. The dCTP being incorporated opposite the template nucleotide and the primer (3′ terminus) are indicated by arrows.

on the energetics of the TS in pol β. To progress toward this goal, we used several simulation approaches. These methods have been described extensively elsewhere and will be described here briefly.

As is apparent from Figure 2, it is necessary to evaluate several thermodynamic cycles. Thus, our success depends in a crucial way upon the ability to obtain reliable results for the relevant free energies. Unfortunately, doing so by statistical mechanical approaches and microscopic simulations is extremely time-consuming, owing to sampling problems. Nevertheless, it is possible in some cases to obtain meaningful results using perturbation approaches. These types of calculations are usually done by the free-energy perturbation (FEP) method (*34*, *35*) and by the related umbrella-sampling (US) method (*35*). The FEP method evaluates the free energy associated with the change of the potential surface from U_1 to U_2 by gradually changing the potential surface using the relationship

$$U_m(\lambda_m) = U_1(1 - \lambda_m) + U_2\lambda_m \quad (1)$$

where λ_m is a parameter that changes between ($0 \leq \lambda_m \leq 1$). The free-energy increment, associated with the change of U_m, can be obtained by (*35*)

$$\exp\{-\delta G(\lambda_m \rightarrow \lambda_{m'})\beta\} = \langle \exp\{-(U_{m'} - U_m)\beta\}\rangle_m \quad (2)$$

where $\langle\ \rangle_m$ indicates that the given average is evaluated by propagating trajectories over U_m and $\beta - (k_BT)^{-1}$. The overall free-energy change is now obtained by changing λ_m in n equal increments and evaluating the sum of the corresponding δG

$$\Delta G(U_1 \rightarrow U_2) = \sum_{m=0}^{n-1} \delta G(\lambda_m \rightarrow \lambda_{m+1}) \quad (3)$$

The FEP approach has been used extensively in studies of free energies of biological systems (e.g., refs *36*−*38*) and has been introduced quite early (*39*) in studies of electrostatic free energies in proteins. Unfortunately, it is often difficult to perform converging FEP calculations (e.g., binding of large ligands). In such cases, it is useful to estimate the free energy of biological processes by an equation derived by Lee et al. (*40*; see also ref *41*) and used in studies of a ligand binding to proteins. This equation expresses the free energy associated with changing the potential of the system from U_1 to U_2 by

$$\Delta G(U_1 \rightarrow U_2) = \frac{1}{2}(\langle U_2 - U_1\rangle_1 + \langle U_2 - U_1\rangle_2) \quad (4)$$

where $\langle\ \rangle_i$ designates a molecular dynamics (MD) average over trajectories obtained with $U = U_i$. The derivation of this equation was based on the assumption that the linear response approximation (LRA) (*42*) is valid. Namely, the protein and solvent environments respond linearly to the force associated with the given process. Although it is hard to accept that the LRA can provide a reliable way of describing the energetics of macromolecules or of realistic molecular systems, it was found by simulation studies that it is a reasonable approximation, in particular, for processes that depend upon electrostatic effects (*43*−*46*).

The use of the LRA approach in calculations of binding free energies involves thermodynamic cycles that consider separately the electrostatic and nonelectrostatic components of the binding process. The resulting binding free energy is expressed as

$$\Delta G_{bind} = \frac{1}{2}[\langle U^p_{elec,l}\rangle_l + \langle U^p_{elec,l}\rangle_{l'} - \langle U^w_{elec,l}\rangle_l - \langle U^w_{elec,l}\rangle_{l'}] + \Delta G^{nonelec}_{bind} \quad (5)$$

where $U^p_{elec,l}$ is the electrostatic contribution for the interaction between the ligand and its surroundings, p and w designate protein and water, respectively, and l and l′ designate the ligand in its actual charged form and the "nonpolar" ligand, where all of the residual charges are set to 0. In this expression, the terms $\langle U_{elec,l} - U_{elec,l'}\rangle$, which are required by eq 4, are replaced by $\langle U_{elec,l}\rangle$ because $U_{elec,l'} = 0$. Now, the evaluation of the nonelectrostatic contribution $\Delta G^{nonelec}_{bind}$ is still very challenging, because these contributions might not follow the LRA. A useful option, which was previously used (*40*, *47*), is to evaluate the contribution to the binding free energy from hydrophobic effects, van der Waals (vdW) forces, and water penetration. Another powerful option is the so-called linear interaction energy (LIE) approach (*45*, *48*). This approach adopts the LRA approximation for the electrostatic contribution but neglects the $\langle U_{elec,l}\rangle_{l'}$ terms. The binding energy is then expressed as

$$\Delta G_{bind} \approx \alpha[\langle U^p_{elec,l}\rangle_l - \langle U^w_{elec,l}\rangle_l] + \beta[\langle U^p_{vdw,l}\rangle_l - \langle U^w_{vdw,l}\rangle_l] \quad (6)$$

where α is a constant that is around $^1/_2$ in many cases and β is an empirical parameter that scales the vdW component of the protein−ligand interaction. A careful analysis of the relationship between the LRA and LIE approaches and the origin of the α and β parameters is given in ref *47*. This

analysis shows that β can be evaluated in a deterministic way provided that one can determine the entropic contribution and preferably the water penetration effect in a microscopic way.

Despite the formal rigor of the FEP and LRA methods, it is frequently found that such methods are subject to major convergence problems when one deals with electrostatic effects in protein interiors and that semimacroscopic models can give more reliable results. This is true in particular with regard to the PDLD/S-LRA method (*49*, *50*) that provides a direct link between the micro- and macroscopic concepts. Because this method was reviewed extensively, we review here only its main features. The PDLD/S−LRA method evaluates the change in solvation free energies upon transfer of a given group or groups to the protein by using the effective potential (*50*)

$$\Delta U_{\mathrm{sol},i}^{\mathrm{w}\rightarrow\mathrm{p}} = [-\Delta G_{\mathrm{sol},i}^{\mathrm{w}} + \Delta G_{\mathrm{sol,p}}^{\mathrm{w}}(q=q_i) - \Delta G_{\mathrm{sol,p}}^{\mathrm{w}}(q=0)]\left(\frac{1}{\epsilon_{\mathrm{p}}} - \frac{1}{\epsilon_{\mathrm{w}}}\right) + \Delta U_{q\mu}\frac{1}{\epsilon_{\mathrm{p}}} \quad (7)$$

where $\Delta G_{\mathrm{sol},i}^{\mathrm{w}}$ is the free energy of solvation of the ith ionizable group in water (the self-energy in water) and $\Delta G_{\mathrm{sol,p}}^{\mathrm{w}}(q = q_i)$ and $\Delta G_{\mathrm{sol,p}}^{\mathrm{w}}(q = 0)$ are the free energies of solvation of the entire protein in water with atomic charges present on the particular group ("charged state") and with atomic charges on the group set to 0 ("uncharged state"), respectively. The $\Delta G_{\mathrm{sol,p}}^{\mathrm{w}}(q = 0)$ term approximates the case where the ionizable group is not in the protein cavity. $\Delta U_{q\mu}$ is the vacuum interaction between the atomic charges on the ionizable group and the permanent dipoles of the protein (represented by atomic charges); ϵ_{w} is the dielectric constant of water; and ϵ_{p} is the dielectric constant of the protein, which is basically a semimacroscopic scaling factor that accounts for the interactions that are not considered explicitly. This factor is significantly different than the actual protein dielectric constant (see ref *50*).

To capture the physics of the reorganization of the protein dipoles in the charging process, it is essential to relax the protein structure in the relevant charged and uncharged states (note that eq 7 corresponds to a single protein structure). Moreover, for accurate free-energy differences, several protein configurations should be averaged. The configurational space can be adequately sampled by utilizing Monte Carlo or MD techniques (*51*). In this study, we use an MD approach in the LRA framework described above. This approach approximates the free energy associated with a transformation between two charged states by averaging the potential difference between the initial and final states over trajectories propagated on these two states, respectively. Using the PDLD/S effective potential energy of eq 7, the free energy of solvation is given by

$$\Delta G_{\mathrm{sol},i}^{\mathrm{w}\rightarrow\mathrm{p}} = \frac{1}{2}[\langle \Delta U_{\mathrm{sol},i}^{\mathrm{w}\rightarrow\mathrm{p}} \rangle_{q=q_i} + \langle \Delta U_{\mathrm{sol},i}^{\mathrm{w}\rightarrow\mathrm{p}} \rangle_{q=0}] \quad (8)$$

where $\Delta U_{\mathrm{sol},i}^{\mathrm{w}\rightarrow\mathrm{p}}$ is the PDLD/S potential of eq 7, which is considered as the effective potential for the LRA approach, and the $\langle\ \rangle_{q=q_i}$ and $\langle\ \rangle_{q=0}$ terms designate an average over protein configurations generated in the charged and uncharged state of the given group, respectively. Although this approach takes into account the reorganization of the environment explicitly, it may not fully account for some effects such as the complete water penetration and protein reorganization. These factors and the effect of induced dipoles are implicitly included in the model, which leads to the use of ϵ_{p} in the PDLD/S model.

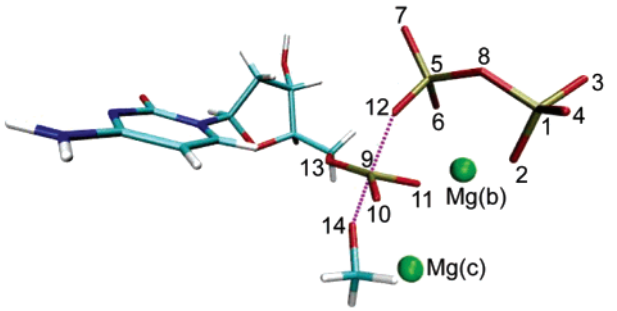

FIGURE 4: The TS model used in the present calculations. The charges of the labeled atoms are given in Table 1. The standard Amber charges are used for all other atoms. The two distances, $R_{9,14}$ and $R_{9,12}$, which define the TS geometry, are 1.7 and 2.2 Å, respectively.

Table 1: Residual Charges Used for the TS Complex[a]

atom number	name	charge	atom number	name	charge
1	PG	1.200	8	O3B	−0.500
2	O1G	−0.915	9	PA	1.155
3	O2G	−0.915	10	O1A	−0.830
4	O3G	−0.915	11	O2A	−0.830
5	PB	1.195	12	O3A	−0.785
6	O1B	−0.905	13	O5′	−0.500
7	O2B	−0.905	14	O3′	−0.585

[a] The atom number is in accordance with the notation of Figure 4.

The present study focuses on the direct binding of the TS. For this purpose, we have generated an empirical valence bond (EVB) type TS, whose nature is described in Figure 4. This "construct" represents the TS of the associative pathway found in our early studies (*11*) (probably with a proton transfer to the bulk solvent). The charge distribution of the TS was obtained by the Gaussian03 package (*52*) DFT (B3LYP/6-31G*) calculations of a $[(CH_3O)_2(PO_3)_3]^{5-}$ system using the PCM solvation model and is given in Table 1. The structure of this $[(CH_3O)_2(PO_3)_3]^{5-}$ TS was partially optimized in the gas phase and used to derive the relevant charges. The parts that were considered as the "probe" or region I in the MOLARIS terminology (*49*) include all of the atoms of the incoming dNTP, as described in Figure 4. The attacking ribose of the DNA primer strand, including the O3′ atom that forms a covalent bond with the α-phosphate of the dNTP substrate in the TS structure was included formally in region II (see below), as well as the two Mg^{2+} ions. A moderate position constraint (K = 10 kcal mol^{-1} Å^{-2}) on the Mg^{2+} ions was used to avoid instabilities in the LRA uncharging calculations. As required by EVB treatments, we excluded the 1−3 interaction between O3′ and the atoms of the α-phosphate. Placing the O3′ nucleophile in region II is a convenient treatment that is, in fact, quite rigorous (see also the Results and Discussion). That is, although U in eq 6 does not include the O3′−P_α bond energy, the free-energy contribution from the changes in this bond are included in the reorganization energy of region II, which is automatically obtained by FEP and LRA treatment. Of course, we could consider C3′ rather than O3′ as a "link atom", but this would give exactly the same result as long

Table 2: Calculated and Observed Effects of Mutations of Pol β on the TS Binding Free Energy[a]

enzyme	origin	$(\Delta G^{PDLD/S}_{bind,TS})'$ (kcal/mol)	$(\Delta G^{LRA}_{bind,TS})'$ (kcal/mol)	$(k_{pol}/K_D)^{obs}$ (s^{-1} M^{-1})	$\Delta g^{\ddagger}_{enz}$ (kcal/mol)	reference	method	system
	rat			6 600 000	8.5	*19*	pss	gap
wild type	rat	−103.6	−394.7	1 100 000	9.6	*65*	pss	nongap
	human			643 000	9.9	*70*	ss	gap
R183A	rat	−85.2	−363.7	440 000	10.1	*19*	pss	gap
R254A	rat	−95.0	−353.3	6700	12.7	*71*	ss (A/T)	nongap
R258A	rat	−98.6	−373.6	26 100	11.9	*71*	ss (A/T)	nongap
Y271A	rat	−114.5	−385.9	415 000	10.2	*20*	pss	nongap
D276V	human	−125.9	−407.0	10 500 000	8.2	*17*	pss	gap
K280G	rat	−96.0	−371.2	41 700	11.0	*18*	ss	gap
R283A	rat	−92.7	−366.3	4900	12.9	*65*	pss	nongap
E295A	rat	−109.5	−379.1	200 000	10.6	*19*	pss	gap

[a] The calculated $(\Delta G^{PDLD/S}_{bind,TS})'$ and $(\Delta G^{LRA}_{bind,TS})'$ correspond to the TS binding in closed proteins that already include 2 Mg^{2+} ions. This corresponds to the $(\Delta g^{\ddagger}_{enz})'$ of Figure 2. The relationship between the calculated $(\Delta G_{bind,TS})'$ and $\Delta g^{\ddagger}_{enz}$ is established in eq 9−11. Note that $(\Delta g^{\ddagger}_{enz})_{calc}$ is $(\Delta g^{\ddagger}_{enz})'_{calc} + 60$ in the LRA treatment and $(\Delta g^{\ddagger}_{enz})_{calc} + 20$ in the PDLD/S approach. The calculated values of the observed $\Delta g^{\ddagger}_{enz}$ are obtained from $(k_{pol}/K_D)^{obs}$ with the help of the TS theory. The sources of the different experimental values and the type of experiment (pss and ss designate pre-steady state and steady state, respectively) are also given. All systems were done with G/C, except when indicated.

as one is using an EVB-type treatment. Note in this respect that any treatment that includes a covalent bonding to the protein must divide some bond between regions I and II (unless one treats the entire protein quantum mechanically). This point is illustrated in Figure 1 of ref *53* for pK_a calculations.

The FEP simulation was performed using ENZYMIX of the MOLARIS simulation package (*49*). The nucleic acid bases were represented by AMBER charges (*54*) and ENZYMIX vdW parameters (after validation that this gives reliable solvation energies and reliable base-pairing energies and structures). A single center model with the vdW parameters $r^* = 1.30$ Å and $\epsilon = 0.06$ kcal/mol (*11*) was used for the Mg^{2+} ions. Somewhat better structural results were obtained with the six-center model of Åqvist and Warshel (*62*), but the use of this more sophisticated model is left to further studies and more careful parametrization.

The FEP simulation involved 21 windows of 500 ps each at 310 K. Only the charges of the mutated residue were gradually changed. For example, for R283A, part of the atom charges were changed from those of Arg to those of Ala, while part of the residual charges were changed to 0. The vdW parameters were not mutated, both because of the inherent instability of such a process and because the corresponding contributions are small. The all-atom LRA simulations were performed with a 500 ps simulation time for each of the two states in eq 4. The radius of the surface-constrained all-atom solvent (SCAAS) (*55*) simulation center was taken as 22 Å, and the center of the sphere was placed at the geometric center of region I. In the case of the LRA treatment, we performed the mutational calculations while changing both the atoms and the charges of the mutated residues (in this case, we used the cycle of Figure 2 while doing separate calculations for the wild-type and mutant enzymes).

The PDLD/S-LRA calculations were performed with the POLARIS module of MOLARIS on the automatically generated MD configurations for the uncharged and charged states needed for the LRA treatment of eq 4. The calculations used $\epsilon_p = 4$ in eq 7, keeping in mind that it is only because of the LRA treatment that we can consistently use such a low value and noting that $\epsilon_p = 4$ has very little to do with the actual dielectric of the protein (*50*, *56*). The $\Delta G^{w\rightarrow p}_{sol,i}$ of eq 8 was taken as the estimated binding free energy of the group [a more sophisticated treatment is described elsewhere (*47*), but it is not justified in the case of binding of ionized groups]. The LRA-binding energy was evaluated using eq 5 and ignoring the nonelectrostatic term that was found to be relatively small. The PDLD/S-LRA treatment is similar in many respects to the MM/PBSA approach (*57*), except that the earlier PDLD/S-LRA uses a more consistent LRA averaging as well as a more consistent dielectric treatment (see the Discussion in ref *38*).

The open and closed protein structures used as starting points for the simulations were constructed from the pol β/DNA binary complex (PDB ID 1BPX) (*29*) and the pol β/DNA/ddCTP ternary complex (PDB ID 1BPY) (*29*), respectively. We also added to the terminal nucleotide of the primer DNA strand and to the ddCTP substrate the 3′ hydroxyl groups. The mutants were constructed by manually mutating the corresponding residues in the wild-type enzyme.

RESULTS AND DISCUSSION

Calculating the Contributions of Charged Residues in Pol β. This principal goal of this study is to begin to analyze the effects of mutations in DNA polymerases by focusing on the change in k_{pol}/K_D (and the corresponding $\Delta g^{\ddagger}_{enz}$) caused by mutations of ionized residues. The selection of the experiments for analysis was complicated, owing to an understandable lack of similar experimental conditions and kinetics protocols in the literature. Therefore, we have focused mainly, although not exclusively, on analyzing pre-steady-state experiments that employed a "G/C single-base gap", in which a dCTP substrate is present for insertion opposite a template G residue located in a one-base gap. Notably, human pol β is used mainly to fill in one-nucleotide gaps during short-patch base-excision repair (*58*). The chosen mutations and their origin as well as the experimental approach used are listed in Table 2, which also lists the relevant experimental references.

To explore the effects of the mutations on $\Delta g^{\ddagger}_{enz}$, we used a thermodynamic cycle that considered the attack of the ribose on the dNTP in the protein (with a proton transfer to the bulk water) and the reference reaction of a nucleophilic

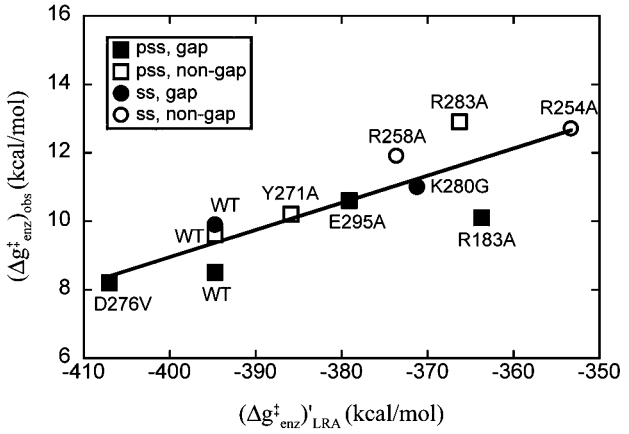

FIGURE 5: Calculated LRA TS-binding free energies for the pol β wild type and mutants and the corresponding experimental results. The experimental approaches used and the specific systems studied are labeled according to the notation of Table 2. The relationship between $(\Delta g^{\ddagger}_{enz})'$ and $(\Delta g^{\ddagger}_{enz})$ is described in the text.

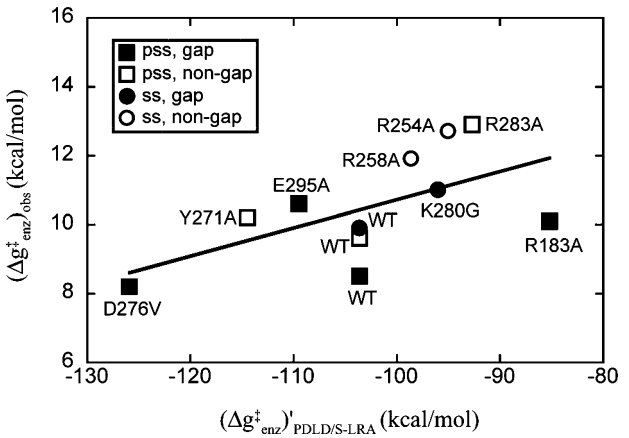

FIGURE 6: Calculated PDLD/S-LRA TS-binding free energies for the pol β wild type and mutants and the corresponding experimental results. The experimental approaches used and the specific systems studied are labeled according to the notation of Table 2. The relationship between $(\Delta g^{\ddagger}_{enz})'$ and $(\Delta g^{\ddagger}_{enz})$ is described in the text.

attack of a ribose on dNTP in water. This gave the relationship

$$\Delta g^{\neq}_{enz} = \Delta g^{\neq}_{w} + \Delta G_{bind}(TS) + \Delta G^{replace}_{bond} \qquad (9)$$

where $\Delta G^{replace}_{bond}$ is the free energy associated with breaking the bond between the ribose and TS in solution and assembling this bond in the protein, without considering the change in environment (which is a part of ΔG_{bind}). Because $\Delta G^{replace}_{bond}$ is close to 0 and because we are only interested in the effects of mutations, we can write

$$\Delta\Delta g^{\neq}_{enz} = \Delta\Delta G_{bind}(TS) \qquad (10)$$

With this relationship in mind, we evaluated the $\Delta\Delta G_{bind}$(TS) for the mutations listed in Table 2, using both the microscopic LRA and the PDLD/S-LRA method. The corresponding calculated results are correlated with the corresponding observed results in Figures 5 and 6, respectively. Before discussing the correlation, we want to emphasize that these calculations correspond to $(\Delta g^{\ddagger}_{enz})'$ of Figure 2, where the two Mg^{2+} ions are already at the active site in the unbound state. Of course, the correct $\Delta g^{\ddagger}_{enz}$ (see

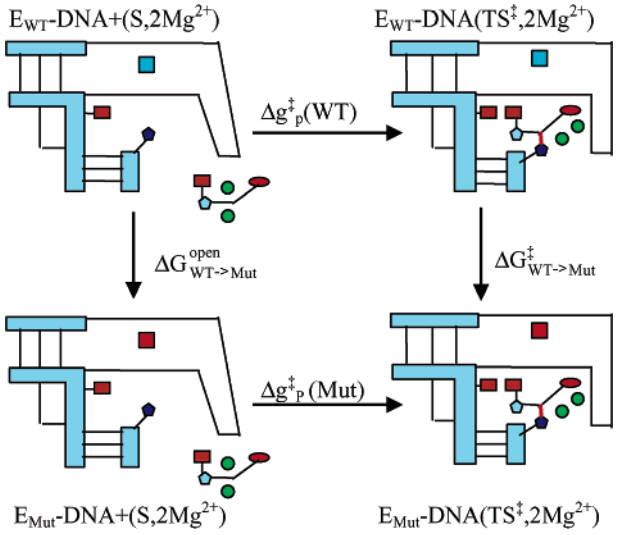

FIGURE 7: Thermodynamic cycle used in the FEP calculations. WT designates wild type; Mut designates mutant; E_{WT}−DNA designates wild-type enzyme−DNA complex; E_{Mut}−DNA designates mutant enzyme−DNA complex; S designates dNTP substrate.

Figure 2) should involve an initial state, where both the Mg^{2+} ions and the incoming dNTP are bound together in water. However, the binding of the two Mg^{2+} ions to the active site of the enzyme−DNA (E−DNA) system gave similar results in our LRA calculations of the different mutants (about +60 kcal/mol), and thus, we used the relationship

$$\Delta g^{\neq}_{enz} = \Delta G_{bind}(2Mg^{2+}) + (\Delta g^{\neq}_{enz})' \qquad (11)$$

and assumed that $\Delta G_{bind}(2Mg^{2+})$ (according to the notation of Figure 2) is constant. We also neglected the effect of the mutation on reorganization of the protein upon moving from the open to the closed form because the calculations started with the closed protein (this issue will be discussed below). The reason for the very large overestimate of the absolute $\Delta g^{\ddagger}_{enz}$ will also be discussed below.

An inspection of the correlation between the calculated and observed results reveals a reasonable but not quantitative correlation. The microscopic LRA (Figure 5) gives a better correlation than the PDLD/S-LRA method (Figure 6), but there are several outliers in both methods. Thus, we can consider the present agreement as semiquantitative, where the uncertainties in the experimental results leave room for improvement both on the theoretical and experimental fronts. Nevertheless, there are several important points that we can deduce from the calculations and from the additional studies reported below.

In trying to assess possible ways to improve our simulation approach, we start by examining the most rigorous available approach, one that involves the use of the FEP method and the mutation cycle of Figure 7. This cycle takes into account, at least formally, all of the relevant effects including the binding of the Mg^{2+} ions and the movement between the open and closed forms of the protein. Here, we can attempt to address two questions. First, are we able to obtain accurate estimates of the change in TS energetics using the full FEP cycle? Second, what can we learn about the validity of the LRA cycle from the more complete FEP cycle? The first question is explored in Table 3, where we compare the FEP calculations to the corresponding experimental results. As

Table 3: Calculated FEP Mutational Effects and the Corresponding Observed Values[a]

	calculated			observed	
enzyme	$\Delta\Delta G_{open}^{WT\rightarrow Mut}$ (kcal/mol)	$\Delta\Delta G_{TS}^{WT\rightarrow Mut}$ (kcal/mol)	$\Delta\Delta g_{enz}^{\ddagger}$ [b] (kcal/mol)	$\Delta g_{enz}^{\ddagger}$ (kcal/mol)	$\Delta\Delta g_{enz}^{\ddagger}$ (kcal/mol)
wild type	0.0	0.0	0.0	8.5−9.9	0.0
D276V	100.3	97.5	−2.8	8.2	−1.4
K280G	50.9	53.8	2.9	11.0	2.0
R283A	52.6	54.2	1.6	12.9	3.3

[a] The calculation used the cycle of Figure 7. [b] Note that $\Delta\Delta g_{enz}^{\ddagger} = \Delta\Delta G_{TS} - \Delta\Delta G_{open}$.

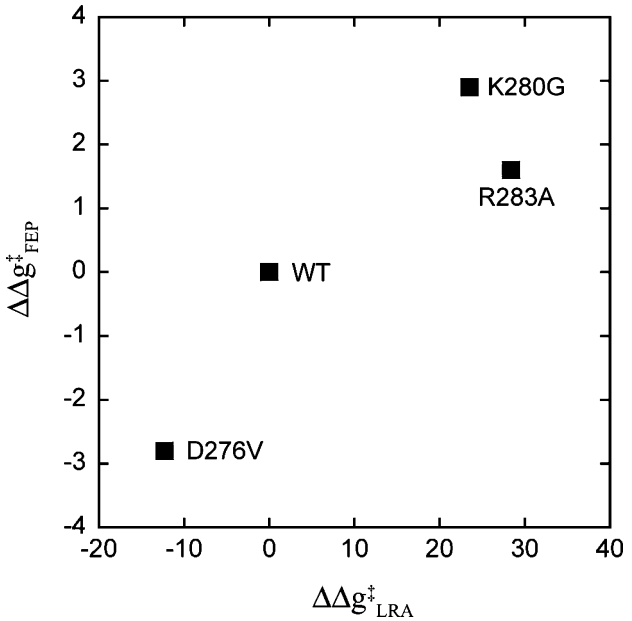

FIGURE 8: Correlation between the calculated LRA and the calculated FEP values of $\Delta\Delta g^{\ddagger}$ (kcal/mol) for the wild-type enzyme and selected mutants of pol β. Note that the difference in magnitude of the two quantities reflects the use of different cycles.

seen from the table, the correlation is reasonable but far from perfect, reflecting probably significant convergence problems or perhaps some problems with the experimental benchmark (see also the discussion below).

Although the FEP approach might give better results than the LRA simulation, we can use the correlation between the two approaches (Figure 8) as an instructive tool. That is, in correlating $\Delta g_{calc,LRA}^{\ddagger}$ and $\Delta g_{calc,FEP}^{\ddagger}$, we can assess the validity of some of our assumptions without being subjected to possible experimental uncertainties. For example, structural studies indicated that Arg283 significantly changes its position and environment in the open and closed forms of the protein. However, the reasonable correlation between $\Delta g_{calc,LRA}^{\ddagger}$ (that does not reflect a conformational change) and $\Delta g_{cal,FEP}^{\ddagger}$ (that reflects the reorganization of the environment of Arg283) indicates that the catalytic effect of Arg283 does not depend upon the open−closed transition. Now, although we do not expect this to be a completely general case, the correlation in Figure 8 indicates that the LRA cycle represents a useful approximation to the FEP results.

One of the goals of this study is to estimate the effective dielectric constant ϵ_{eff} for the interaction between the protein ionized groups and the TS. Establishing the magnitude of this parameter can provide significant insight into the protein reorganization as well as the ability to use simple models in studies of TS analogues. Defining the effective dielectric constant by

$$\Delta\Delta g_{obs}^{\neq} = \frac{\Delta\Delta g_{calc}^{\neq}}{\epsilon_{eff}} \quad (12)$$

we find, from the calculations described above, ϵ_{eff} of 1.5 ± 0.5, 10, and 12 for the FEP, PDLD/S-LRA, and LRA approaches, respectively. The finding that ϵ_{eff} is somewhat larger than 1 for the FEP calculations may reflect in part the use of a nonpolarizable force field and perhaps difficulties in obtaining full convergence because of incomplete sampling (see the Concluding Remarks). It is also possible that a part of the reason for having an effective dielectric of 1.5, rather than 1, in the FEP calculations is associated with the use of a mutational procedure that only mutated the charges while leaving the atoms of the wild-type protein unchanged. However, using the cycle of Figure 2 in the LRA calculations and comparing the results obtained by charge mutations and complete mutations gave very similar results. Thus, clarifying this issue will require further studies.

The large values of ϵ_{eff} of the PDLD/S-LRA and LRA methods reflect the fact that the calculations do not represent a full response of the protein/solvent system to the charging process and the use of a simplified cycle. The trend of the ϵ_{eff} for the PDLD/S-LRA and LRA approaches is common to charge−charge interactions in proteins (e.g., see refs *50* and *59*) and in particular to cases with metal ions. It is important to realize at this point that the microscopic LRA and the FEP method usually give similar dielectric constants and lead to a similar effective dielectric constant. Thus, the difference between the ϵ_{eff} of the LRA and FEP methods reflects basically the difference between the cycles used. In the FEP case, we only charge the mutated group (changing a charge from 0 to ±1 while considering explicitly the transition from the open to the closed form). In the LRA case, on the other hand, we consider the charging of a highly charged substrate. At any rate, the knowledge of the effective dielectric for the different treatments should be quite helpful in the semiquantitative prediction of mutational effects. Moreover, the ϵ_{eff} of the PDLD/S-LRA method will be used in the construction of the interaction matrices that will be described in the next section.

Energy-Based Correlation Diagrams and the Allosteric Coupling between the Base-Binding and Catalytic Sites. The results described in the previous section suggest that the PDLD/S-LRA method can be used to obtain qualitative insights that reflect the contributions of different residues to the energetics of the TS. Using this method, we can move beyond the generic proposal that polymerase fidelity results from an "induced fit" mechanism (*60*), where it is stated that the binding to the correct substrate moves the enzyme

to a more favorable catalytic configuration. That is, although the induced fit observation is in principle valid, it does not tell us what configurational change was induced and, more importantly, it does not tell us why it provides better catalysis. The "why" and "how" are the questions that must be resolved to understand the molecular basis of polymerase fidelity. Our working hypothesis is that the interaction between the base-binding and catalytic sites operates in an allosteric way, where the binding of an incorrect base leads to changes in the base-binding site, which are then propagated to the catalytic site and lead to structural changes that reduce the stabilization of the TS. To explore this proposal, it is important to try to identify the role of individual residues that transfer the information between the base-binding and catalytic sites.

The origin of the structural control of fidelity can be quantified by considering the interplay between the binding site of the incoming base and the stabilization of the TS in the catalytic site. Our previous studies (*10*, *11*, *14*) have indicated that the binding of the incoming base is determined by the preorganization energy provided by the base-binding site that includes the template base. Now, as shown in a schematic way in Figure 9, the binding of an incorrect nucleotide results in a situation where the wrong base pair of the incoming dNTP does not bind optimally with the template. The incoming base can relax to a configuration with more favorable binding energy, but this will be at the expense of having a suboptimal interaction between the TS and its binding site. In other words, the reorganization of the environment at the base-binding site, i.e., the shift of the base, forces a reorganization in the catalytic site. One could attempt to consider this correlation by examining the correlation between the motion of different residues as was done in studies of hydride transfer in dihydrofolate reductase (e.g., ref *61*). However, although these correlations are clearly instructive, they do not tell us about the energetics that control the catalytic process. To attain a more quantitative picture, it is important to asses the interaction of free energy between the protein groups. Now, the use of energy diagrams for the interaction between protein parts is not new (see ref *59*, *63*) and here we use a related treatment by introducing the following approach. We start by evaluating the PDLD/S-LRA group contribution (as defined by eq 8) to the binding energy of selected ionizable residues proximal to the base-binding and catalytic sites, at the TS configuration. These contributions constitute the diagonal elements (the $\Delta\Delta G_{ii}$) of the interaction matrix. Note that $\Delta\Delta G_{ii}$ is not equal to the contribution of the *i*th residue to TS binding but is equal to the energetics of uncharging the *i*th group at the TS. This is similar to $\Delta G^{\ddagger}_{WT\rightarrow Mut}$ in Figure 7. Next, we define the off-diagonal elements of the correlation matrix following our previous studies (*59*)

$$\Delta\Delta G_{ij} = \Delta G(q_i = 0 \rightarrow q_i = \overline{q_i})^{w\rightarrow p}_{q_j=\overline{q}_j} - \Delta G(q_i = 0 \rightarrow q_i = \overline{q_i})^{w\rightarrow p}_{q_j=0} \quad (13)$$

where $\Delta G(q_i = 0 \rightarrow q_i = \overline{q_i})^{w\rightarrow p}_{q_j=\overline{q}_j}$ is evaluated with all of the q values (except q_i and q_j) set to 0 and $q_j = \overline{q_j}$. Here, $\overline{q_i}$ designates the charge distribution of the *i*th residue, and $q_i = 0$ means that all of the residual charges at the *i*th group are set to 0. In this way, we evaluate the change in the group

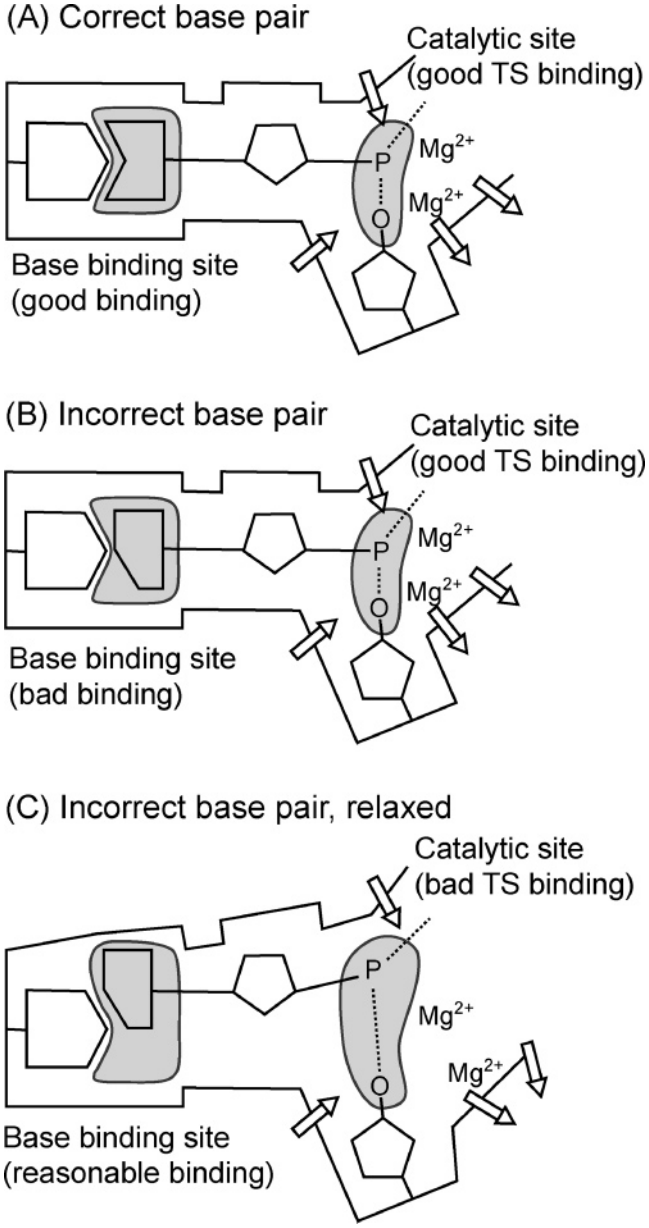

FIGURE 9: Schematic description of the protein rearrangement upon binding of $\mathscr{W}$. The protein provides perfect sites for both the chemical and base-pairing parts of $\mathscr{R}$. On the other hand, in the case of $\mathscr{W}$, the protein has to relax in the base-pairing site, and this relaxation destroys the preorganization in the chemical part. Arrows indicate the protein dipoles.

contribution of residue *i* upon uncharging residue *j*, as can be done by double-mutation experiments. The off-diagonal elements can be evaluated as was done in ref *59* by systematic PDLD/S-LRA calculations, where $\Delta G(q_i = 0 \rightarrow q_i = \overline{q_i})$ is evaluated by using eq 8. This is, however, an expensive procedure that is left to subsequent studies. Here, we use a simplified approach based on the simplified group contribution treatment of the PDLD/S method (*63*, *64*). Using this approximation, we obtain from eq 13 the following result:

$$\Delta\Delta G_{ij} \cong \langle \Delta U^{ij}_{q\mu}/\epsilon^{ij}_{eff} \rangle \quad (14)$$

where $\Delta U^{ij}_{q\mu}$ is the contribution to the $\Delta U^{ij}_{q\mu}$ of eq 7 from the gas phase interaction between the ith and jth residues and ϵ^{ij}_{eff} is taken as 4 for interactions between nonionized residues

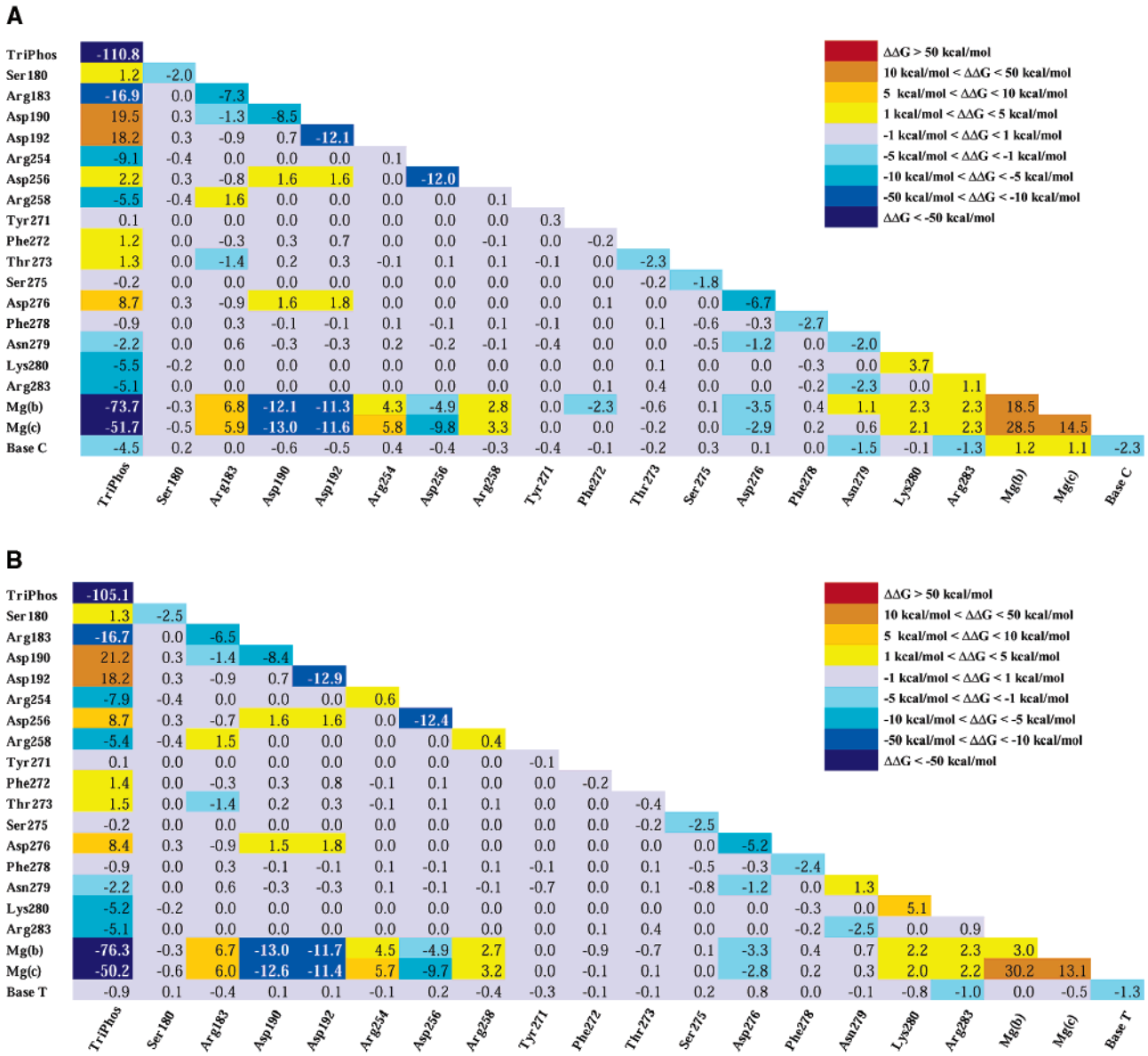

FIGURE 10: Interaction matrices for the TS of the incorporation of the dNTP substrate in the active site of pol β calculated with the PDLD/S-LRA method. The diagonal elements describe the electrostatic contributions ($\Delta\Delta G_{ii}$, in kcal/mol) of the indicated residues to TS binding, while the off-diagonal elements describe the effect of the indicated *j*th residue (in the given column) on the TS binding by the *i*th residue ($\Delta\Delta G_{ij}$) (see eq 14). The intensity of colors corresponds to the strength of the interaction (e.g., the red color shows the strongest interaction; the light gray color shows interactions close to 0; and the blue color shows the negative interaction). (A) Template guanine with incoming dCTP ($\mathscr{R}$), (B) Template guanine with incoming dTTP ($\mathscr{W}$). TriPhos denotes the triphosphate part of the incoming dNTP; Mg(b) denotes the binding magnesium ion, and Mg(c) denotes the catalytic magnesium ion. The bases are cytosine in A and thymine in B, respectively. The interaction between the incoming base and the template base is −5.2 kcal/mol and −2.9 kcal/mol for the correct and incorrect pairs, respectively.

and as the ϵ_{eff} obtained from eq 12 for interactions between ionized residues.

The above matrix describes the complex interactions between protein groups at the TS configuration, and allows one to probe the coupling between the template base and incoming nucleotide through the protein residues and Mg^{2+} ions. The corresponding correlation diagrams are presented in Figure 10A for the enzyme with the correct base pair (G/C) and in Figure 10B for an incorrect base pair (G/T). We also evaluated the interaction between the incoming base and the template for the correct and incorrect base pairs and listed the results in the caption of Figure 10.

It should be noted that the group contributions are not additive (in the same way that alanine-scanning experiments may not be additive) and that our interaction matrix should be viewed as a qualitative tool for probing the residues involved in the coupling between the base-binding and catalytic sites.

Using the difference between the correlation diagrams for incorporation of wrong and right nucleotides, depicted in Figure 11, provides insight into the allosteric information transfer governing replication fidelity. That is, the difference matrix (referred to here as the "allosteric matrix") helps to identify the residues that are involved in coupling between the base-binding and catalytic sites. The present analysis is preliminary in its nature because the structure of the incorrect base pair was generated by starting from the same protein structure (PDB ID 1BPY) (*29*) with the correct base pair

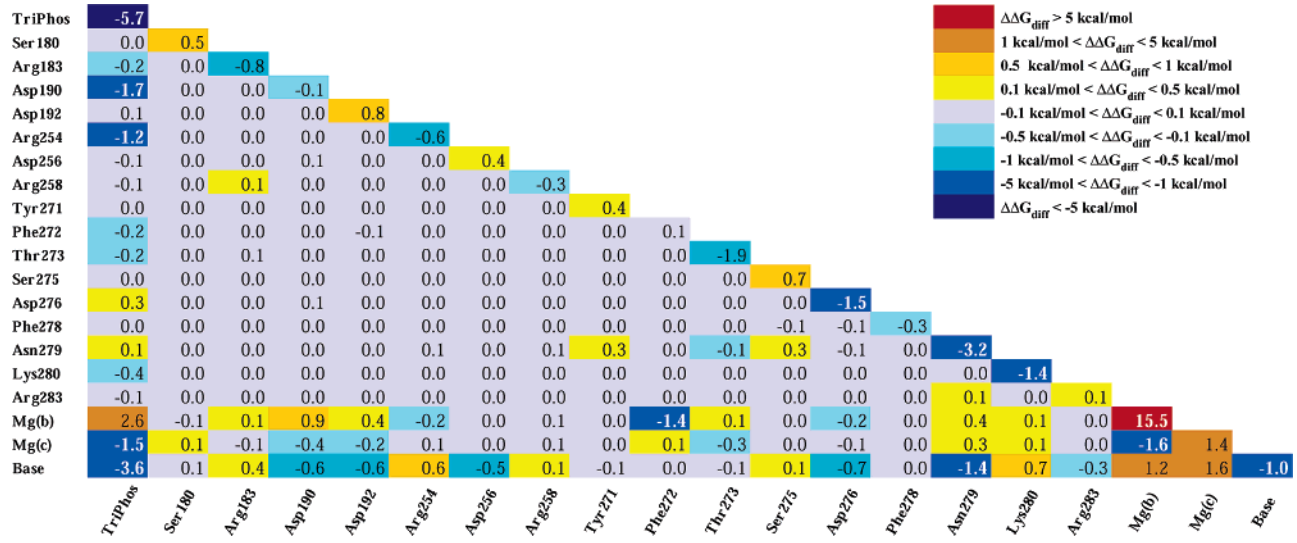

FIGURE 11: Allosteric matrix constructed from the difference between the interaction matrices of G/C and G/T, which are given in Figure 10. The notation is the same as in Figure 10. Note that the scale of the color code is different than in Figure 10.

rather than from the actual structure containing a mismatched base pair. Nevertheless, inspection of the first column and the last row provides an insight about residues that interact in an allosteric way with the TS and the base, respectively. Thus, Asp190, Arg254, and the Mg^{2+} ions are coupled to the TS, while Arg254, Asp256, Asp276, Asn279, Lys280, and the Mg^{2+} ions are coupled to the base. This coupling can be understood by noticing that, if the structure of the catalytic site was kept constant after changing the base to an incorrect base, we would obtain the same interaction between the TS and Arg254 in the interaction matrices of the correct and the "unrelaxed" interaction matrix of the incorrect base (not shown). Now, once the structure of the mismatch system is allowed to relax, we obtain a 1.2 kcal/mol reduction in the interaction between Arg254 and the TS. This reduction in TS stabilization is the value given in the allosteric matrix. The quantitative validity of this analysis requires further experimental and theoretical studies. In particular, it will be useful to obtain structural information for incorrect base pairs and to examine the effect of pol β double mutants.

It is encouraging that the calculated fidelity obtained from the first element of the allosteric matrix (the change in the TS binding for G/C and G/T base pairs) is around 5.7 kcal/mol. This value is a reasonable approximation to the observed fidelity. That is, the observed fidelity for a G/T pair is 3800 for a nongapped DNA (*65*) and 12 000 for a gapped DNA (*66*) substrate. The calculated energy difference in Figure 11 results in a fidelity $[(k_{pol}/K_D)_{G/C}/(k_{pol}/K_D)_{G/T}]$ of 11 000. However, it is important to emphasize that we do not presently view the allosteric matrix as a quantitative tool for estimating polymerase fidelity. That is, although the overall fidelity is reproduced in a reasonable way by the allosteric matrix, the off-diagonal coupling terms seem too small, reflecting perhaps insufficient calculated structural rearrangement upon misincorporation (note that we do not have structural information for the mismatch case). Thus, we still need to refine a proper effective dielectric constant that is relevant to the structural change upon misincorporation, which is much lower than the dielectric constant for charge−charge interactions. To arrive at more quantitative conclusions will require comparisons of the calculated and observed effect of base substitutions similar to the analysis of the effects of polymerase mutations on kinetics and fidelity.

CONCLUDING REMARKS

This study examines the ability of simulation methods to reproduce the effect of mutation of ionized residues on the catalytic reaction of pol β. The observed trend was reproduced in a semiquantitative way by both the PDLD/S-LRA and the all-atom LRA methods, indicating that these methods can provide a useful tool for studies of DNA polymerases. Furthermore, our study provides an insight into the effective dielectric constants for the charge−charge interaction in DNA polymerases. These effective dielectric constants can be used as a powerful tool in the qualitative analysis of mutational effects.

Our study introduces an energy-based correlation diagram that should help in identifying the residues responsible for information transfer between the base-binding and catalytic sites. The allosteric matrix obtained by computing the difference between the correlation matrices of the correct and incorrect base pairs can provide an important bridge between experimental and theoretical studies. That is, the calculations of the allosteric matrix can best be accomplished starting with X-ray structures of right and wrong base pairs. The predictions of the correlation diagrams can be verified by performing experiments with polymerase double mutants, which can lead hopefully to an eventual understanding of the control of information transfer governing the fidelity process.

The present study focuses on the direct calculation of the binding of the TS rather than the stepwise calculation of ΔG_{bind} and $\Delta g^{\ddagger}_{cat}$. This approach provides significantly better results and correlations than those obtained by separate calculations of the binding and catalytic steps, which we have not reported. The utility of this approach may offer insight into the logical design of TS analogues and thus may provide a route toward drug design. The success of the PDLD/S-LRA approach in computing the energy of the TS offers the possibility to perform a rapid mutational screening of TS analogues based on their predicted ability to bind tightly to the polymerase active cleft.

Obtaining quantitative results for the effects of mutations in DNA polymerases presents a major challenge for computer simulation approaches. We believe that we use the state of the art approaches and note that our earlier studies (e.g., refs *11*−*14*) are to the best of our knowledge the only studies that reproduce the trend in fidelity in DNA polymerases. Nevertheless, we must still consider the present results as qualitative results. One example of the problems faced with studies of mutational effects in the highly charged active site is provided by our attempt to explore the R258A mutation by a FEP treatment. In this case, we found that we have major problems in reproducing the experimental trend. These problems were traced back to the difficulties of obtaining sufficient solvation energy for Arg258 in the protein active site. The origin of this problem seems to be related to the fundamental problem of reproducing the pK_a of the V66E mutant of staphylococcal nuclease (SNase) by microscopic and semimacroscopic calculations (*67*; Kato and Warshel, *J. Phys. Chem. B*, in press), because it reflects significant changes in water penetration and/or local folding. Fortunately, we developed recently an overcharging FEP approach (Kato and Warshel, *J. Phys. Chem. B*, in press) that can be used to overcome this sampling problem. However, we feel that such treatment should be left to subsequent studies.

It might be useful to comment at this point on our new allosteric matrix as a general tool for analysis of allosteric effects, in view of the current tendency to focus on the effect of correlated motions on enzyme catalysis (e.g. *61*, *68*). As discussed elsewhere (*69*; Warshel et al., manuscript submitted), the fact that changes far from the chemical site can influence catalysis is important, but this is unlikely to be due to specific motions per se. The actual effect reflects most likely the changes in the preorganization of the active site. Even in chemical reactions in solution, we have a very strong coupling between the solvent modes along the reaction coordinate, but this is simply the nature of the reaction coordinate in condensed phases and not a feature that determines the activation barrier. The control of fidelity or other allosteric effects is due to the transfer of energy or more precisely to the change in energy balance, where changing the binding in the effector site changes the binding of the TS (see Figure 9). Thus, it seems to us that focusing on energy correlation is more effective than focusing on the correlation between structural changes (that must anyhow be converted to energy).

We view the present study as a "work in progress" along a path toward deriving a quantitative understanding of the origin of DNA replication fidelity. We are not yet at a stage where the computational analysis can be regarded as definitive, nor have we provided a complete analysis of the fidelity problem. Nevertheless, the present study provides compelling evidence that, even for the challenging task of elucidating the origins of DNA synthesis fidelity, there is much to be learned by performing computer-aided structure−function correlation analysis. Combining computer simulations with structural and kinetic approaches should provide a more complete description of protein function, including the extent to which complex protein conformational changes contribute to the enzyme mechanism. For the example of DNA polymerases, we believe that this type of synergistic approach may eventually lead to a quantitative picture of the molecular origin of replication fidelity.

REFERENCES

1. Kornberg, A., and Baker, T. A. (1992) *DNA Replication*, W. H. Freeman, New York.
2. Wolfenden, R., and Snider, M. J. (2001) The depth of chemical time and the power of enzymes as catalysts, *Acc. Chem. Res. 34*, 938−945.
3. Schroeder, G. K., Lad, C., Wyman, P., Williams, N. H., Wolfenden, R. (2006) The time required for water attack at the phosphorus atom of simple photodiesters and of DNA, *Proc. Natl. Acad. Sci. U.S.A. 103*, 4052−4055.
4. Echols, H., and Goodman, M. F. (1991) Fidelity mechanisms in DNA replication, *Annu. Rev. Biochem. 60*, 477−511.
5. Johnson, K. A. (1993) Conformational coupling in DNA polymerase fidelity, *Annu. Rev. Biochem. 62*, 685−713.
6. Joyce, C. M., and Benkovic, S. J. (2004) DNA polymerase fidelity: Kinetics, structure, and checkpoints, *Biochemistry 43*, 14317−14324.
7. Goodman, M. F., Creighton, S., Bloom, L. B., and Petruska, J. (1993) Biochemical basis of DNA replication fidelity, *Crit. Rev. Biochem. Mol. Biol. 28*, 83−126.
8. Kunkel, T. A., and Bebenek, K. (2000) DNA replication fidelity, *Annu. Rev. Biochem. 69*, 497−529.
9. Goodman, M. F. (2002) Error-prone repair DNA polymerases in prokaryotes and eukaryotes, *Annu. Rev. Biochem. 71*, 17−50.
10. Florian, J., Goodman, M. F., and Warshel, A. (2002) Theoretical investigation of the binding free energies and key substrate-recognition components of the replication fidelity of human DNA polymerase β, *J. Phys. Chem. B 106*, 5739−5753.
11. Florian, J., Goodman, M. F., and Warshel, A. (2003) Computer simulation of the chemical catalysis of DNA polymerases: Discriminating between alternative nucleotide insertion mechanisms for T7 DNA polymerase, *J. Am. Chem. Soc. 125*, 8163−8177.
12. Florian, J., Warshel, A., and Goodman, M. F. (2002) Molecular dynamics free-energy simulations of the binding contribution to the fidelity of T7 DNA polymerase, *J. Phys. Chem. B 106*, 5754−5760.
13. Florian, J., Goodman, M. F., and Warshel, A. (2003) Computer simulation studies of the fidelity of DNA polymerases, *Biopolymers 68*, 286−299.
14. Florian, J., Goodman, M. F., and Warshel, A. (2005) Computer simulations of protein functions: Searching for the molecular origin of the replication fidelity of DNA polymerases, *Proc. Natl. Acad. Sci. U.S.A. 102*, 6819−6824.
15. Radhakrishnan, R., and Schlick, T. (2004) Orchestration of cooperative events in DNA synthesis and repair mechanism unraveled by transition path sampling of DNA polymerase β's closing, *Proc. Natl. Acad. Sci. U.S.A. 101*, 5970−5975.
16. Yang, L. J., Beard, W. A., Wilson, S. H., Broyde, S., and Schlick, T. (2004) Highly organized but pliant active site of DNA polymerase β: Compensatory mechanisms in mutant enzymes revealed by dynamics simulations and energy analyses, *Biophys. J. 86*, 3392−3408.
17. Berg, B. J. V., Beard, W. A., and Wilson, S. H. (2001) DNA structure and aspartate 276 influence nucleotide binding to human DNA polymerase β—Implication for the identity of the rate-limiting conformational change, *J. Biol. Chem. 276*, 3408−3416.
18. Beard, W. A., Shock, D. D., Yang, X. P., DeLauder, S. F., and Wilson, S. H. (2002) Loss of DNA polymerase β stacking interactions with templating purines, but not pyrimidines, alters catalytic efficiency and fidelity, *J. Biol. Chem. 277*, 8235−8242.
19. Kraynov, V. S., Showalter, A. K., Liu, J., Zhong, X. J., and Tsai, M. D. (2000) DNA polymerase β: Contributions of template-positioning and dNTP triphosphate-binding residues to catalysis and fidelity, *Biochemistry 39*, 16008−16015.
20. Kraynov, V. S., Werneburg, B. G., Zhong, X. J., Lee, H., Ahn, J. W., and Tsai, M. D. (1997) DNA polymerase β: Analysis of the contributions of tyrosine-271 and asparagine-279 to substrate specificity and fidelity of DNA replication by pre-steady-state kinetics, *Biochem. J. 323*, 103−111.
21. Werneburg, B. G., Ahn, J., Zhong, X. J., Hondal, R. J., Kraynov, V. S., and Tsai, M. D. (1996) DNA polymerase β: Pre-steady-state kinetic analysis and roles of arginine-283 in catalysis and fidelity, *Biochemistry 35*, 7041−7050.
22. Kim, S. J., Beard, W. A., Harvey, J., Shock, D. D., Knutson, J. R., and Wilson, S. H. (2003) Rapid segmental and subdomain motions of DNA polymerase β, *J. Biol. Chem. 278*, 5072−5081.

23. Beard, W. A., Shock, D. D., Vande Berg, B. J., and Wilson, S. H. (2002) Efficiency of correct nucleotide insertion governs DNA polymerase fidelity, *J. Biol. Chem. 277*, 47393−47398.
24. Feig, M., Zacharias, M., and Pettitt, B. M. (2001) Conformations of an adenine bulge in a DNA octamer and its influence on DNA structure from molecular dynamics simulations, *Biophys. J. 81*, 352−370.
25. Orozco, M., Perez, A., Noy, A., and Luque, F. J. (2003) Theoretical methods for the simulation of nucleic acids, *Chem. Soc. Rev. 32*, 350−364.
26. Seibert, E., Ross, J. B., and Osman, R. (2003) Contribution of opening and bending dynamics to specific recognition of DNA damage, *J. Mol. Biol. 330*, 687−703.
27. Oostenbrink, C., and van Gunsteren, W. F. (2005) Efficient calculation of many stacking and pairing free energies in DNA from a few molecular dynamics simulations, *Chemistry 11*, 4340−4348.
28. Sawaya, M. R., Pelletier, H., Kumar, A., Wilson, S. H., and Kraut, J. (1994) Crystal structure of rat DNA polymerase β: Evidence for a common polymerase mechanism, *Science 264*, 1930.
29. Sawaya, M. R., Prasad, R., Wilson, S. H., Kraut, J., and Pelletier, H. (1997) Crystal structures of human DNA polymerase β complexed with gapped and nicked DNA: Evidence for an induced fit mechanism, *Biochemistry 36*, 11205−11215.
30. Krahn, J. M., Beard, W. A., and Wilson, S. H. (2004) Structural insights into DNA polymerase β deterrents for misincorporation support an induced-fit mechanism for fidelity, *Structure 12*, 1823−1832.
31. Batra, V. K., Beard, W. A., Shock, D. D., Pedersen, L. C., and Wilson, S. H. (2005) Nucleotide-induced DNA polymerase active site motions accommodating a mutagenic DNA intermediate, *Structure 13*, 1225−1233.
32. Fothergill, M., Goodman, M. F., Petruska, J., and Warshel, A. (1995) Structure-energy analysis of the role of metal ions in phosphodiester bond hydrolysis by DNA polymerase I, *J. Am. Chem. Soc. 117*, 11619−11627.
33. Bakhtina, M., Lee, S., Wang, Y., Dunlap, C., Lamarche, B., and Tsai, M. D. (2005) Use of viscogens, dNTP α S, and rhodium(III) as probes in stopped-flow, experiments to obtain ne evidence for the mechanism of catalysis by DNA polymerase β, *Biochemistry 44*, 5177−5187.
34. Zwanzig, R. W. (1954) High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys. 22*, 1420.
35. Valleau, J. P., and Torrie, G. M. (1977) *Modern Theoretical Chemistry*, Vol. 5, Plenum Press, New York.
36. Kollman, P. (1993) Free energy calculations: Applications to chemical and biochemical phenomena, *Chem. Rev. 93*, 2395−2417.
37. Warshel, A., and Åqvist, J. (1991) Electrostatic energy and macromolecular function, *Annu. Rev. Biophys. Chem. 20*, 267−298.
38. Shurki, A., and Warshel, A. (2003) Structure/function correlations of protreins using MM, QM/MM and related approaches; methods, concepts, pitfalls and current progress, *Adv. Protein Chem. 66*, 249−312.
39. Warshel, A., Sussman, F., and King, G. (1986) Free energy of charges in solvated proteins: Microscopic calculations using a reversible charging process, *Biochemistry 25*, 8368−8372.
40. Lee, F. S., Chu, Z. T., Bolger, M. B., and Warshel, A. (1992) Calculations of antibody−antigen interactions: Microscopic and semi-microscopic evaluation of the free energies of binding of phosphorylcholine analogs to McPC603, *Protein Eng. 5*, 215−228.
41. Hummer, G., and Szabo, A. (1996) Calculation of free energy differences from computer simulations of initial and final states, *J. Chem. Phys. 105*, 2004−2010.
42. Kubo, R., Toda, M., and Hashitsume, N. (1985) *Statistical Physics II: Nonequilibrium Statistical Mechanics*, Springer-Verlag, Berlin, Germany.
43. Hwang, J.-K., and Warshel, A. (1987) Microscopic examination of free energy relationships for electron transfer in polar solvents, *J. Am. Chem. Soc. 109*, 715−720.
44. Kuharski, R. A., Bader, J. S., Chandler, D., Sprik, M., Klein, M. L., and Impey, R. W. (1988) Molecular model for aqueous ferrous ferric electron transfer, *J. Chem. Phys. 89*, 3248−3257.
45. Åqvist, J., and Hansson, T. (1996) On the validity of electrostatic linear response in polar solvents, *J. Phys. Chem. 100*, 9512−9521.
46. Morreale, A., de la Cruz, X., Meyer, T., Gelpi, J. L., Luque, F. J., and Orozco, M. (2005) Partition of protein solvation into group contributions from molecular dynamics simulations, *Proteins 58*, 101−109.
47. Sham, Y. Y., Chu, Z. T., Tao, H., and Warshel, A. (2000) Examining methods for calculations of binding free energies: LRA, LIE, PDLD-LRA, and PDLD/S-LRA calculations of ligands binding to an HIV protease, *Proteins 39*, 393−407.
48. Åqvist, J., Luzhkov, V. B., and Brandsdal, B. O. (2002) Ligand binding affinities from MD simulations, *Acc. Chem. Res. 35*, 358−365.
49. Lee, F. S., Chu, Z. T., and Warshel, A. (1993) Microscopic and semimicroscopic calculations of electrostatic energies in proteins by the POLARIS and ENZYMIX programs, *J. Comput. Chem. 14*, 161−185.
50. Schutz, C. N., and Warshel, A. (2001) What are the dielectric "constants" of proteins and how to validate electrostatic models, *Proteins 44*, 400−417.
51. Allen, M. P., and Tildesley, D. J. (1987) *Computer Simulation of Liquids*, Oxford University Press, New York.
52. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Montgomery, J. A., Kudin, K. N., Burant, J. C., Millam, J. M., Iyengar, S. S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G. A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, N., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J. E., Hratchian, H. P., Cross, J. B., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Ayala, P. Y., Morokuma, K., Voth, G. A., Salvador, P., Dannenberg, J. J., Zakrezewski, V. G., Dapprich, S., Daniels, A. D., Strain, M. C., Farkas, O., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Ortiz, J. V., Cui, Q., Baboul, A. G., Clifford, S., Cioslowski, J., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Gonzalez, C., Pople, J. A. (2003) *Gaussian03*, Gaussian, Inc., Pittsburgh, PA.
53. Sham, Y. Y., Chu, Z. T., and Warshel, A. (1997) Consistent calculations of pK_a's of ionizable residues in proteins: Semi-microscopic and macroscopic approaches, *J. Phys. Chem. B 101*, 4458−4472.
54. 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., and Kollman, P. A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *J. Am. Chem. Soc. 117*, 5179−5197.
55. King, G., and Warshel, A. (1989) A surface constrained all-atom solvent model for effective simulations of polar solutions, *J. Chem. Phys. 91*, 3647−3661.
56. Warshel, A., and Papazyan, A. (1998) Electrostatic effects in macromolecules: Fundamental concepts and practical modeling, *Curr. Opin. Struct. Biol. 8*, 211−217.
57. Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., and Case, D. A. (1998) Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate−DNA helices, *J. Am. Chem. Soc. 120*, 9401−9409.
58. Wilson, S. H., and Kunkel, T. A. (2000) Passing the baton in base excision repair, *Nat. Struct. Biol. 7*, 176−178.
59. Sham, Y. Y., Muegge, I., and Warshel, A. (1998) The effect of protein relaxation on charge−charge interactions and dielectric constants of proteins, *Biophys. J. 74*, 1744−1753.
60. Koshland, D. E. (1958) Application of a theory of enzyme specificity to protein synthesis, *Proc. Natl. Acad. Sci. U.S.A. 44*, 98−104.
61. Wong, K. F., Selzer, T., Benkovic, S. J., and Hammes-Schiffer, S. (2005) Impact of distal mutations on the network of coupled motions correlated to hydride transfer in dihydrofolate reductase, *Proc. Natl. Acad. Sci. U.S.A. 102*, 6807−6812.
62. Åqvist, J., and Warshel, A. (1990) Free-energy relationships in metalloenzyme-catalyzed reactions—Calculations of the effects of metal-ion substitutions in staphylococcal nuclease, *J. Am. Chem. Soc. 112*, 2860−2868.
63. Muegge, I., Schweins, T., and Warshel, A. (1998) Electrostatic contributions to protein−protein binding affinities: Application to Rap/Raf interaction, *Proteins 30*, 407−423.

64. Muegge, I., Tao, H., and Warshel, A. (1998) A fast estimate of electrostatic group contributions to the free energy of protein inhibitor binding, *Protein Eng. 10*, 1363−1372.

65. Ahn, J., Werneburg, B. G., and Tsai, M. D. (1997) DNA polymerase β: Structure−fidelity relationship from pre-steady-state kinetic analyses of all possible correct and incorrect base pairs for wild type and R283A mutant, *Biochemistry 36*, 1100−1107.

66. Ahn, J. W., Kraynov, V. S., Zhong, X. J., Werneburg, B. G., and Tsai, M. D. (1998) DNA polymerase β: Effects of gapped DNA substrates on dNTP specificity, fidelity, processivity, and conformational changes, *Biochem. J. 331*, 79−87.

67. Garcia-Moreno, B., Dwyer, J. J., Gittis, A. G., Lattman, E. E., Spencer, D. S., and Stites, W. E. (1997) Experimental measurement of the effective dielectric in the hydrophobic core of a protein, *Biophys. Chem. 64*, 211−224.

68. Thorpe, I. F., and Brooks, C. L., III (2005) Conformational substates modulate hydride transfer in dihydrofolate reductase, *J. Am. Chem. Soc. 127*, 12997−13006.

69. Olsson, M. H. M., Parson, W. W., and Warshel, A. Dynamical contributions to enzyme catalysis: Critical tests of a popular hypothesis, *Chem. Rev.*, in press.

70. Miller, H., Prasad, R., Wilson, S. H., Johnson, F., and Grollman, A. P. (2000) 8-oxodGTP incorporation by DNA polymerase β is modified by active-site residue Asn279, *Biochemistry 39*, 1029−1033.

71. Menge, K. L., Hostomsky, Z., Nodes, B. R., Hudson, G. O., Rahmati, S., Moomaw, E. W., Almassy, R. J., and Hostomska, Z. (1995) Structure−function analysis of the mammalian DNA polymerase β active site: Role of aspartic acid 256, arginine 254, and arginine 258 in nucleotidyl transfer, *Biochemistry 34*, 15934−15942.

BI060147O