BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

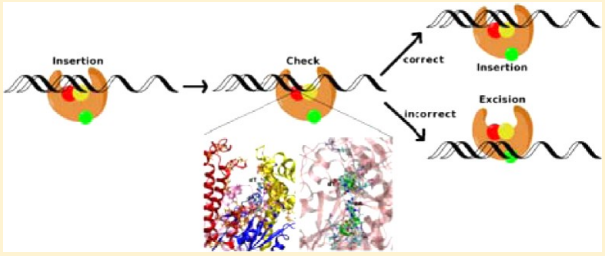
Computational Prediction of Residues Involved in Fidelity Checking for DNA Synthesis in DNA Polymerase I

Sarah E. Graham, FatimaSultana Syeda, and G. Andrés Cisneros*

Department of Chemistry, Wayne State University, Detroit, Michigan 48202, United States

Supporting Information

ABSTRACT: Recent single-molecule Förster resonance energy transfer studies of DNA polymerase I have led to the proposal of a postinsertion fidelity-checking site. This site is hypothesized to ensure proper base pairing of the newly inserted nucleotide. To help test this hypothesis, we have used energy decomposition, electrostatic free energy response, and noncovalent interaction analysis analyses to identify residues involved in this putative checking site. We have used structures of DNA polymerase I from two different organisms, the Klenow fragment from *Escherichia coli* and the Bacillus fragment from *Bacillus stearothermophilus*. Our results point to several residues that show altered interactions for three mispairs compared to the correctly paired DNA dimer. Furthermore, many of these residues are conserved among A family polymerases. The identified residues provide potential targets for mutagenesis studies for investigation of the fidelity-checking site hypothesis.



DNA polymerases play a crucial role in the cell cycle because of their involvement in genome replication and repair. Errors in replication can produce mutations that can lead to disease or even death.[1] Error rates for the replication machinery in *Escherichia coli* have been reported to be around 10^{-10}.[2] The high fidelity rate is due to at least three steps: base selection, exonucleolytic proofreading, and mismatch repair. Our understanding of the molecular mechanism for correct base pairing remains a field of intense study because it can lead to insights into the mutagenesis process.

The Klenow fragment (KF) of *E. coli* has been well characterized and widely employed as a model system to study DNA polymerases. The KF is composed of the polymerase and 3′−5′ exonuclease domains of DNA polymerase I. The minimal mechanism proposed for polymerase activity involves the binding of a DNA substrate to the polymerase domain in the "open" conformation, followed by the incoming nucleotide and a conformational change to the "closed" conformation. Subsequently, the chemical step to incorporate the nucleotide takes place. This is followed by another conformational change, release of pyrophosphate and translocation along the DNA to the next position $(n + 1)$ or DNA release.[3]

The KF belongs to the A family of polymerases, a group of highly related polymerases in both structure and function. Because of this, other polymerases in the family, including those of *Thermus aquaticus* (Klentaq fragment) and *Bacillus stearothermophilus* [Bacillus fragment (BF)], can be used to gain insight into the mechanisms of fidelity. Both the Klentaq fragment and the BF have the characteristic structure of the KF, resembling a right hand as they move along the DNA strand, and the fragments are functionally analogous to the KF. These polymerases contain both a polymerase domain and a 3′−5′ exonuclease domain. The KF and BF are 39% identical in overall sequence and 49% identical in the active site.[4] On the basis of the crystal structure of the KF, the exonuclease domain corresponds to residues 324−518 and the polymerase domain to residues 519−928. In the BF, the exonuclease domain corresponds to residues 297−468 and the polymerase domain to residues 469−876.

The KF has been shown to synthesize DNA with medium fidelity, with an average base substitution rate of approximately 2×10^{-5}.[5] The discrimination against nucleotide misinsertion in the KF is due to a lower binding affinity for the incorrect nucleotide and a slower rate for misinsertion.[3] Rate discrimination has been shown to be quantitatively stronger than dNTP discrimination, although values vary among mispairs.[6]

Loh et al. have shown that the KF can tolerate amino acid mutations that alter fidelity without a change in activity. In addition, several of the considered substitutions produce an antimutator phenotype.[7] Thompson et al. have studied nine KF derivatives with mutations in the polymerase domain. Their results suggest that DNA mismatch recognition depends on specific residues in the polymerase domain.[8] This has led to the suggestion that KF is able to recognize mismatches after a misincorporation event. Similar mutations have been shown to affect fidelity in other A family polymerases. Studies by Yoshida and co-workers as well as Tosaka et al. identify several residues important for replication fidelity in *T. aquaticus*.[9−11] It is therefore likely that studies on Klenow and related polymerases will provide insight into the mechanism of fidelity of the highly conserved A family polymerases.

Received: December 15, 2011
Revised: February 6, 2012
Published: March 8, 2012

dx.doi.org/10.1021/bi201856m | *Biochemistry* 2012, 51, 2569−2578

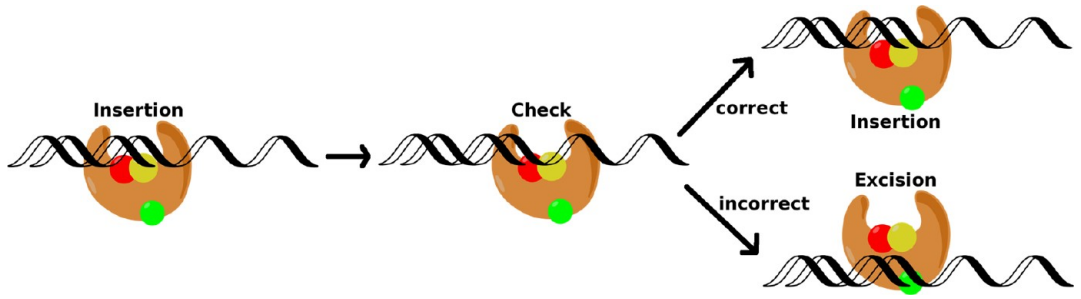


Figure 1. Schematic diagram of the proposed fidelity checking process in DNA polymerase I. Following nucleotide insertion, the DNA is translocated to the checking site. If the inserted base pair is a correct pair, the polymerase continues. If not, the DNA is shifted to the exonuclease domain and the incorrect base pair is excised.

Recent single-molecule Förster resonance energy transfer (smFRET) studies of DNA polymerase I (KF) have revealed a previously unknown step in DNA polymerization. Following each nucleotide incorporation, the KF translocated along the DNA strand in a process consistent with a two-base downstream movement (n + 2) from the active site along the template before returning to the proper location, one base downstream (n + 1), for the next base incorporation. It is suggested that this translocation is done in the open conformation of the KF. These results have led to the hypothesis of a postinsertion fidelity-checking site located at the two-base downstream position that ensures the newly added base is properly paired to the templating base, shown in Figure 1.[12] A similar proposal has been made on the basis of molecular-tweezer experiments.[13]

Here, we use energy decomposition analysis (EDA), electrostatic free energy response (EFER) analysis, and noncovalent interaction (NCI) analysis to identify residues involved in this putative fidelity-checking site. To this end, MD simulations were conducted for DNA fragments in both KF and BF. The MD simulations for KF and BF consisted of DNA dimers with a T:A pair (correctly paired) and three mispairs (C:A, G:A, and A:A). Simulations were conducted with DNA in the preinsertion (n + 1) and putative fidelity-checking (n + 2) sites. The ensembles generated from these runs have been employed for the analyses.

Computational studies of DNA polymerase fidelity have been pioneered by Warshel and co-workers,[14−20] who have established the importance of electrostatic effects in enzyme catalysis.[21] In addition, Warshel and co-workers have developed the linear response approximation (LRA) as applied to biosimulations, which has been shown to be useful for the investigation of group contributions to ligand binding and enzyme catalysis.[22] In LRA and PDLD/LRA-S, ligand binding free energies and group contributions to ligand binding are estimated from electrostatic and nonelectrostatic contributions.[22] In addition, in LRA the dielectric screening of the Coulomb interaction is explicitly taken into account. In the case of charged ligands interacting with polar residues, a small effective dielectric screening (ε = 4) has been shown to provide a reasonable approximation. For charged residues, the solvation contribution is much larger, so a larger dielctric screening is used.[22]

EDA has been successfully used to infer catalytic roles of residues surrounding enzymatic active sites from quantum mechanical/molecular mechanical (QM/MM) catalytic studies.[23−31] The catalytic activity of some of these predictions has been subsequently confirmed experimentally.[32,33] In QM/MM calculations, the changes in nonbonded (Coulomb and van der Waals) intermolecular interaction energy can be calculated between each residue and the active site between the reactant and the transition state (TS):

$$\Delta E = \left\langle \Delta E_{i,\mathrm{QM}}^{\mathrm{nonbond}} \right\rangle_{\mathrm{MM,TS}} - \left\langle \Delta E_{i,\mathrm{QM}}^{\mathrm{nonbond}} \right\rangle_{\mathrm{MM,react}}$$

where i represents an individual residue in the MM subsystem, $E_{i,\mathrm{QM}}^{\mathrm{nonbond}}$ represents the nonbonded interaction between residue i and the QM subsystem (active site), and the broken brackets represent averages over the ensembles where the conformational space of the MM subsystem is sampled. It is important to note that this provides only a qualitative assessment and does not take into account dielectric screening effects as compared to LRA.

The EFER model applies a related strategy to calculate the electrostatic energy between two points on a restricted free energy surface. This has been previously applied successfully to biological systems to study the transition state of the DNA polymerase β reaction.[34] This approach used molecular dynamics (MD) simulations to calculate the free energy over the course of the reaction by evaluating the potential energy along the reaction coordinate. The energy difference between a probe region and the rest of the system was calculated, allowing determination of the energy contributions. In this contribution, we employ a similar approach to identify potential residues involved in the proposed fidelity checking stage.

NCI analysis[35] for selected structures of each of the considered systems has been conducted to provide additional insights. The results from these analyses are correlated with a multiple-sequence alignment of the A family polymerases to determine conservation of residues involved in the putative fidelity-checking site.

METHODS

Energy Decomposition Analysis. Two sets of energy differences have been determined to obtain the change in nonbonded interaction energy. The first corresponds to the interaction of every residue with the correctly paired bases in the preinsertion and fidelity-checking sites:

$$\Delta E_{\mathrm{ref}} = \left\langle \Delta E_{i,\mathrm{correct}}^{\mathrm{nonbond}} \right\rangle_{\mathrm{preinsertion}} - \left\langle \Delta E_{i,\mathrm{correct}}^{\mathrm{nonbond}} \right\rangle_{\mathrm{check}}$$

where i represents an individual residue, $E_{i,\mathrm{correct}}^{\mathrm{nonbond}}$ represents the nonbonded interaction (electrostatic or van der Waals) between residue i and the two correctly paired terminal bases (template and newly added correct nucleotide, T:A), and the broken brackets represent averages over the conformational

dx.doi.org/10.1021/bi201856m | Biochemistry 2012, 51, 2569−2578

space ensemble sampled from molecular dynamics (MD) simulations. The second energy difference corresponds to the interaction of every residue with the incorrectly paired bases [template and newly added incorrect terminal bases (G:A, C:A, and A:A)] also in the preinsertion and fidelity-checking sites:

$$\Delta E = \left\langle \Delta E_{i,\text{incorrect}}^{\text{nonbond}} \right\rangle_{\text{preinsertion}} - \left\langle \Delta E_{i,\text{incorrect}}^{\text{nonbond}} \right\rangle_{\text{check}}$$

The difference between the correct (reference) and incorrect interactions ($\Delta\Delta E = \Delta E_{\text{ref}} - \Delta E$) has been used to identify the residues that have the largest change in interaction between residues in the protein and correctly and incorrectly paired bases. Here, the energy difference for the T:A pair serves as reference for the G:A, C:A, and A:A mispairs. In this case, no QM/MM calculations have been performed; all results have been obtained from MD simulations only. All required energies and energy differences have been calculated by an in-house FORTRAN90 program. This program uses the snapshots obtained from the MD simulations and the force field parameter set to identify the required nonbonded interactions.

Electrostatic Free Energy Response Analysis. The free energy differences were calculated on the basis of the EFER model as described in ref 34. We have adapted the EFER free energy difference expression to calculate the free energy changes arising from correctly and incorrectly paired nucleotides in two different sites from the MD ensemble averages of electrostatic interaction energies between the DNA pair (correct or incorrect) and the rest of the system by

$$\Delta G = \frac{1}{2}\left(\left\langle V_{i,\text{correct}}^{\text{preinsertion}} - V_{i,\text{correct}}^{\text{check}} \right\rangle + \left\langle V_{i,\text{incorrect}}^{\text{preinsertion}} - V_{i,\text{incorrect}}^{\text{check}} \right\rangle\right)$$

in which broken brackets denote the ensemble average from the MD simulations, i represents an individual residue, and V_i is the electrostatic interaction energy between residue i and the DNA bases in the respective site. The free energy differences were calculated by considering the T:A pair as the correct (reference) pair and the A:A, C:A, and G:A pairs as the incorrect structures. As explained by Klvaňa et al., this formula is a special case of time-independent linear response theory.[14,34] Note that EFER and LRA are similar, although LRA also includes terms for the nonelectrostatic contributions (preorganization and van der Waals).[18,22]

MD Simulations. No structures of the KF with DNA in the active site have been determined. Therefore, we have used a structure from the highly homologous DNA polymerase I from *T. aquaticus* (Klentaq) as a guide to position the DNA in the active site of the KF. Klentaq has been crystallized in both the open and closed forms bound to a DNA primer−template sequence. The structure of Klentaq [Protein Data Bank (PDB) entry 4KTQ[36]] was superposed onto that of the KF (PDB entry 1KLN[37]) on the basis of the polymerase active sites. This provided a template for KF with DNA in the active site. The KF crystal structure includes 611 amino acids, which comprise both the exonuclease and polymerase domains. There are 16 possible DNA pairs (4 correct pairs and 12 mispairs). In this work, we have studied only one correct pair (T:A) and the three corresponding mispairs (A:A, C:A, and G:A)

For the KF, we have considered only double-stranded DNA (dsDNA) trimers with no template overhang as the DNA substrate for the polymerase. That is, the DNA substrate for KF comprises three bases each on the template and primer strands and the ends are blunt. These dsDNA trimers were docked into the active site of the KF by superposing them onto the DNA from the ternary structure of Klentaq superposed on the KF structure. Four DNA trimers were constructed corresponding to GAX opposite CTA, where X corresponds to T, C, A, or G. These trimers were built using the nucleic acid builder (NAB) utility from AMBER.[38] Eight initial structures were constructed, corresponding to the one correctly paired and three mispaired trimers, each in both the preinsertion and fidelity-checking sites.

Structures for BF were created using a crystal structure determined with DNA in the active site (PDB entry 1U4B[39]). This crystal structure contains 23 bp of DNA, with a one-base template overhang. The BF, analogous to the KF, is composed of both polymerase and exonuclease domains totaling 603 amino acids. Eight structures were created with mispairs A:A, C:A, and G:A and correct pair T:A, both in the preinsertion sites and in the fidelity-checking sites, as shown in Figure 2.

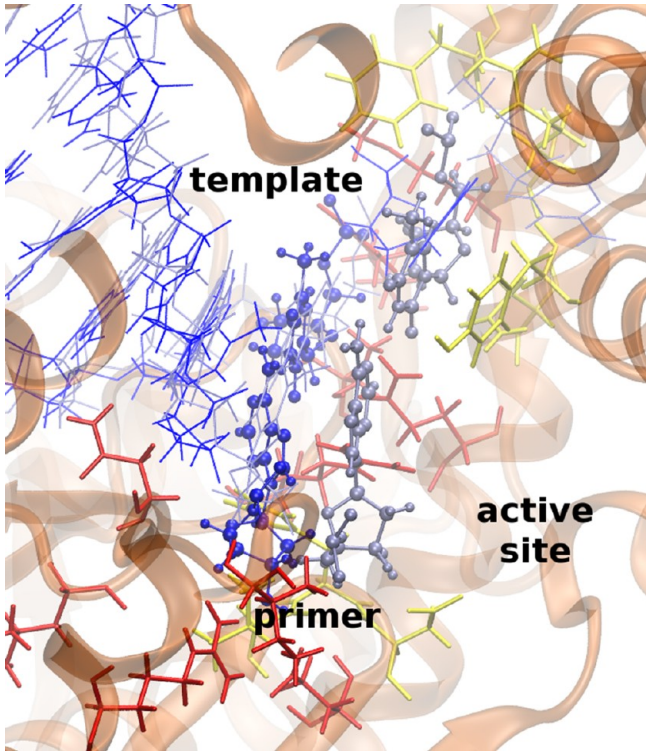


Figure 2. DNA positioned in the BF. DNA with a T:A pair in the preinsertion site is shown as light blue balls and sticks (the rest of the DNA shown as light blue sticks). DNA with a T:A pair in the checking site is shown as dark blue balls and sticks (the rest of the DNA shown as dark blue sticks). Residues within 3 Å of the T:A pair in the preinsertion and checking sites are colored yellow and red, respectively.

This structure served as a comparison to the KF, both as a control to ensure that the superposition of DNA from Klentaq into Klenow did not prejudice the results obtained and to determine if the DNA substrate in the KF shifts position in the DNA binding groove because of its small size and lack of a template overhang.

These structures were subjected to minimization and single-run MD with the PMEMD program from the AMBER11 suite of programs.[38] Initially, all structures were minimized with 1000 steps of conjugate gradient minimization, followed by solvation,

dx.doi.org/10.1021/bi201856m | *Biochemistry* 2012, 51, 2569−2578

and NPT simulations to set the density to 1 g/mL. The KF was solvated in a 103 Å × 127 Å × 108 Å box of TIP3P[40] water, and the BF was solvated in a 100 Å × 100 Å × 110 Å box of TIP3P water. Once the systems achieved the target temperature, the constraints were gradually removed over 100 ps until no constraints were left, all with the *NVT* ensemble. In all cases, the smooth particle mesh Ewald method was used for the calculation of long-range electrostatics.[41,42] The total time of the simulations for the KF was 22 ns. Simulations for the BF were conducted for 20 ns. Snapshots of the simulations were created every 1 ps. Root-mean-square deviations (rmsds) for the systems throughout the simulations are around 3−4 Å. This can be compared with the rmsd of 3.0 Å for the free Klenow structure under the same simulation conditions. The rmsds for BF are <3.0 Å for all the systems except for the A:A (check) structure, for which the rmsd increases to 3.5 Å (see the Supporting Information).

Snapshots corresponding to 10 ns from each of the 24 single-run MD simulations were employed for the energy decomposition and EFER analyses. That is, 10000 structures for each system were employed for both EDA and EFER. The nonbonded intermolecular interaction energies between each residue in the enzyme and the two nucleotides in the preinsertion or checking site were determined for all systems to obtain the energy differences for decomposition analysis. On the basis of our previous studies, we have selected a cutoff of ±1 kcal/mol for the interactions. Thus, only residues that have interactions that change more than ±1 kcal/mol are considered as having a significant contribution.

NCI Calculations. NCI is a visualization tool recently proposed by Johnson et al.[35] NCI provides a way to identify noncovalent interactions between molecules (e.g., van der Waals, hydrogen bond, etc.). It is a visualization index that relies on the molecular density and its derivatives. NCI is based on the peaks that appear in the reduced density gradient at low densities.[35] This method provides a way to investigate the noncovalent intra- and intermolecular interactions. In this way, it is possible to analyze interactions between a ligand and a protein as has been previously shown.[35] The results obtained from the NCI analysis consist of surfaces between the interacting molecules. These surfaces are assigned specific colors to denote the strength and characteristic of the interactions: green surfaces denote weak interactions, e.g., van der Waals, blue surfaces strong attractive interactions, e.g., hydrogen bonds, and red surfaces strong repulsive interactions.[35]

For this study, three snapshots for each of the 16 systems have been subjected to NCI analysis (see the Supporting Information). The final templating and primer bases have been considered as ligands interacting with a spherical region of 10 Å around the binding site. All calculations were obtained with a step size of 0.2 for the cube and a cutoff of 5 Å for the calculation of the interactions between the nucleotides and the active site.

RESULTS AND DISCUSSION

Sixteen structures total from the KF and the BF combined have been considered in this study. Several residues show a change in interaction with the DNA at the mispaired bases compared to the correct pair based on the EDA.

The number of residues is different for each mispair, suggesting a different selectivity depending on the mismatch in agreement with experimental results.[6] In particular, for the G:A, C:A, and A:A mispairs in the KF, there are 137, 157, and 168 residues that show a significant change in interaction, respectively. Similar results were obtained for the BF. For the A:A, C:A, and G:A mispairs, 76, 72, and 70 residues, respectively, had significant $\Delta\Delta E$ compared to the correctly paired TA. Most of the residues obtained from the energy decomposition analysis (EDA) are located in the polymerase. Standard deviation (SD) analysis for the EDA shows that the deviation for the calculated interactions is between 10 and 20%.

Comparison of the EDA results shows that approximately 30 residues have altered interactions for all mispairs in either the KF or the BF. These residues are located on all three polymerase domains (palm, fingers, and thumb) with a few in the exonuclease domain. Several residues have a $\Delta\Delta E$ of ≥10 kcal/mol for at least one of the mismatches. These have also been selected as contributing to the checking mechanism because the EDA interaction is considerably larger than the cutoff of ±1 kcal/mol to consider the interaction as significant. The majority of the residues obtained from the EDA show changes only in electrostatic interactions, although a few residues showed changes only in the van der Waals interactions.

Correlation of a sequence alignment with the EDA results of the KF and the BF reveals 14 residues that have interactions for the T:A mispairs in both polymerases (see Figure 3).

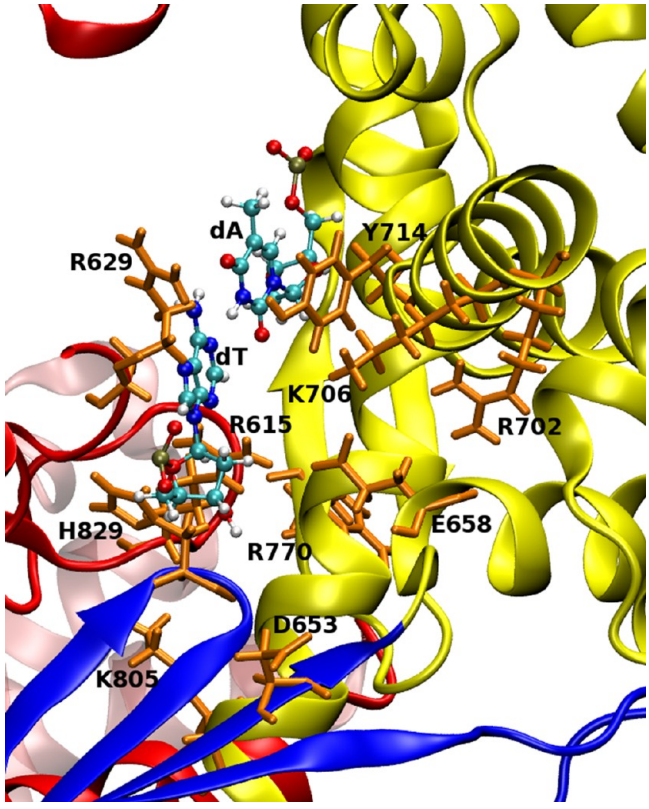


Figure 3. Positions of the residues found to have an interaction with the mispairs in both the KF and the BF (orange sticks). The structure and numbering correspond to those of the BF. Polymerase subdomains are colored red, blue, yellow, and mauve for the palm, fingers, thumb, and exonuclease, respectively. DNA bases (T:A) in the preinsertion site are shown as balls and sticks.

Interestingly, these residues are located primarily near what would be the newly inserted nucleotide. This is somewhat intuitive, as these residues are proposed to participate in the

Table 1. NCI Results for the BF (left) and the KF (right)[a]

residue	A:A	C:A	G:A	T:A
T611	(+)			(+)
Q612	(+)	(+)	(+)	(+)
T613	(+)		(+)	(+)
R615	+(+)	(+)	+(+)	+(+)
Q624	(+)	(+)	(+)	(+)
N625	(+)	(+)	(+)	(+)
I626		(+)	(+)	(+)
P627	(+)	(+)	(+)	(+)
I628	+(+)	+(+)	+(+)	+(+)
R629	(+)	(+)	+(+)	+(+)
R637	(+)	+(+)	(+)	
S655	+			
L659	+			
F710	+(+)	+	+	+
Y714	+(+)	+	+	+(+)
Q723	+			
R771	(+)	(+)	(+)	(+)
F786		+(+)	+	+
R789	+	+	+	+
M790	(+)	+(+)	+	+(+)
N793	(+)	+	+	+
T794	(+)			
Q797	+(+)	+	+	+(+)
V828	+(+)	+(+)		+
H829	+(+)	+(+)	+(+)	+(+)
D830	+		+(+)	+

residue	A:A	C:A	G:A	T:A
A606				+
S608				+
T609				+
E611	+			
E612	+			
V613	+			
L614	+			
T664			(+)	(+)
A665	(+)			(+)
T666		+(+)		(+)
R668	+	+(+)		
S670			(+)	
Q677	(+)			(+)
N678	(+)		(+)	(+)
I679	(+)		(+)	(+)
P680	(+)	(+)	(+)	(+)
V681	(+)	(+)	+(+)	+(+)
R682	(+)	+(+)	+(+)	+(+)
N683	(+)		(+)	
R690			+(+)	
S707			+(+)	
D710	+			+
K758		+		
F762	+	+(+)	+(+)	+(+)
Y766	+(+)	+(+)	+(+)	+(+)
M768	+(+)	(+)	+	+
Q776	+			
R835	+			+
A838	+	+	+	+
R841	+	+	+	+(+)
A842	+	+(+)	+	+(+)
N845	+(+)	+(+)	+	+(+)
A846		(+)		(+)
Q849	+	+(+)	+(+)	+(+)
V880	(+)			+(+)
H881	+(+)	+(+)	+(+)	+(+)
D882	+	+		+

[a]Residues conserved among A family polymerases are shown in bold. Residues in checking site in parenthesis.

fidelity checking mechanism that would determine whether this was the correct nucleotide inserted and the polymerization continues or the incorrect nucleotide resulting in movement to the exonuclease site.

Selected structures from the MD simulations were subjected to NCI. This method has been recently proposed by Johnson et al.[35,43] and allows the visualization of interactions such as hydrogen bonds, van der Waals interactions, etc. Table 1 lists the residues that have been identified by NCI to interact with the correctly paired and mispaired nucleotides in the preinsertion and fidelity-checking sites. Several of these residues coincide with the EDA.

NCI analysis of the KF shows interactions for 37 residues among all considered nucleotide pairs and 26 residues in the BF. Some of these residues show interactions for only one pair or a few of the pairs. A group of residues, corresponding to Q677−P680 in the KF and homologous residues Q624−P627 in the BF, show interactions exclusively when the nucleotides are in the checking site. Several other residues such as R682, Y766, and H881 in the KF appear to have interactions with the terminal nucleotides regardless of which site they occupy. In particular, Y766 (Y714 in BF) shows interesting results of EDA and NCI analysis. This residue is only one of two (A838 in the KF is the other one) that have altered interactions in the van der Waals component. It also shows stacking interactions with the properly matched templating base when it is in the preinsertion site. However, Y766 appears to stack preferentially with the primer base when there is a mismatch (e.g., see Figure 4A and Figure S4E of the Supporting Information).

A sequence alignment of A family DNA polymerases was performed with TCoffee. This alignment shows that 26 of the residues obtained from the EDA of the KF are totally or partially conserved across the family and across species (see Figure 5). These residues correspond to E541, K593, R631, R654, T664, R668, R673, P680, R682, R690, D705, E710, R712, D732, R754, R755, K758, Y766, E783, R821, R822, D854, K857, H881, D882, and E883. Of these residues, only R654 and R673 (K591 and E620, respectively, in the BF) are not conserved in the BF. Residues D705 and D882 in the KF (D653 and D830, respectively, in the BF) correspond to the

dx.doi.org/10.1021/bi201856m | Biochemistry 2012, 51, 2569−2578

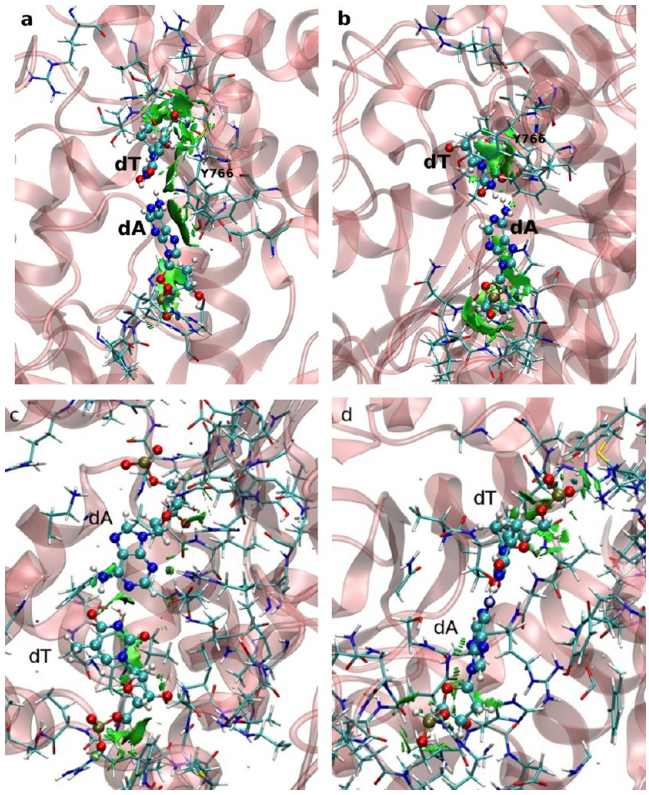


Figure 4. NCI analysis of the TA nucleotide pair (a) in the preinsertion site and (b) in the fidelity-checking site shown of the KF and (c) in the preinsertion site and (d) in the fidelity-checking site of the BF. Colored surfaces denote noncovalent interactions: green surfaces denote weak interactions, e.g., van der Waals, and blue surfaces strong attractive interactions, e.g., hydrogen bonds.[35]

active site residues that form bonds to the catalytic metals for the polymerase reaction. The remaining residues are totally conserved except for R690 and R712, where K and W substitute for a third of the sequences. In the case of R755, this residue is only partially conserved. All these residues show altered interactions on the basis of the Coulomb component except for Y766, which shows altered van der Waals interactions.

There were 14 residues that were found to have interactions shown in the EDA for all the mispairs in both the KF and the BF. Ten of these residues were found to be totally conserved across the A family of polymerases. These residues are R668, R682, D705, E710, R754, R758, Y766, R822, K857, and H881 in the KF, which correspond to R615, R629, D653, E658, R702, K706, Y714, R770, K805, and H829 in the BF. Correlating this sequence alignment to the NCI analysis reveals that 11 of the residues identified for the KF are conserved, eight of which were also found in the NCI analysis of BF. These residues are highlighted in bold in Table 1. This suggests that, in addition to electrostatic interactions, these residues have other types of interactions such as $\pi-\pi$ stacking or form hydrogen bonds with the terminal nucleotides. Of the conserved residues, only P680 (P627) in the KF (BF) shows noncovalent interactions for all the terminal nucleotide pairs only when they are in the checking site. Comparing the results of the analysis with those from NCI and the sequence alignment reveals two residues, R682 and H881 in the KF and R629 and H829 in the BF, that are totally conserved across the A family and that are also found in EDA and NCI.

The EFER analysis yields similar but not identical results, as shown in Figure 6 (see the Supporting Information for full tables of interactions). Many of the same residues were found in both the EDA and EFER analyses. The EFER analysis for BF showed 150 residues that had an interaction energy of >1 for all three mispairs. The EFER analysis for KF showed 58 residues that showed a change in interaction energy for all T:A mispairs. Comparing these residues with the sequence alignment yields 39 residues that show interactions in all mispairs for both the KF and the BF. Several of these residues were found to have an interaction in both EDA and EFER for all mispairs (A:A, C:A, and G:A in both the KF and the BF). These are residues 600, 612, 668, 682, 687, 705, 710, 754, 758, 822, and 857 in the KF and corresponding homologous residues 548, 559, 615, 629, 634, 653, 658, 702, 706, 770, and 805 in the BF. Of these, only residues 600, 612, and 687 (KF numbering) are not conserved across the A family of polymerases.

Mutagenesis Studies Linking Fidelity with Specific Residues. Thompson et al. have used time-resolved fluorescence anisotropy to determine the partitioning of the DNA primer terminus between the polymerase and the exonuclease.[8] This methodology allowed the calculation of free energy differences for correctly paired and mispaired G:X DNAs in the KF. Their results indicate that R668 and N845 function to recognize matched base pairs. These residues are homologous to R615 and N793 in the BF. On the other hand, Q849 shows negligible energetic change, and its mutation did not impair discrimination of G:G mispairs.

Other experimental studies have shown that mutation of R668, R682, E710, and N845 causes decreased fidelity in the KF. In addition, the mispair discrimination changes as a function of the mispair.[44] For example, the most frequent error for the E710A mutant is the A:dCTP mispair. This correlates with the our results for KF in which the change in interaction for the A:C mispair is negative whereas $\Delta\Delta E$ for the G:A or A:A mispair is positive. In the BF, the change in interaction was positive for the A:C mispair but negative for the A:A and G:A mispairs. The R668A mutant has been reported to favor errors involving insertion of dGTP. Our results show that the interaction energy change for all mispairs in both the BF and the KF is 1 order of magnitude larger than the 1 kcal/mol cutoff for R668 and the homologous residue in BF.

The R682A and N845A mutants give an increase in the frequency of T:dCTP errors. For the latter, all $\Delta\Delta E$ values involving dCTP misinsertion are positive, in both the KF and the BF. Misinsertions of A:dATP mispairs are also favored by the N845A mutant. EDA in this case shows only the interaction energy change above the cutoff for the C:A mispair for the KF, although all three mispairs in BF have significant interaction energy. When dCTP is in the primer position, all three mispairs show a positive interaction energy change.

EDA shows a striking result for residue R682 in the KF in that the $\Delta\Delta E$ for the T:C mispair is almost 10-fold higher than for the other mispairs involving dCTP insertion. In addition, the T:C mispair energy is positive compared to all other energies, which are negative; this suggests an unfavorable interaction between the T:C mispair and R682. An interaction of this magnitude was not found in the BF, although the NCI results indicate an interaction for all mispairs. The NCI results for the KF show that R682 (R629 in the BF) interacts with all dCTP mispairs in both the preinsertion and fidelity-checking sites.

dx.doi.org/10.1021/bi201856m | *Biochemistry* 2012, 51, 2569–2578

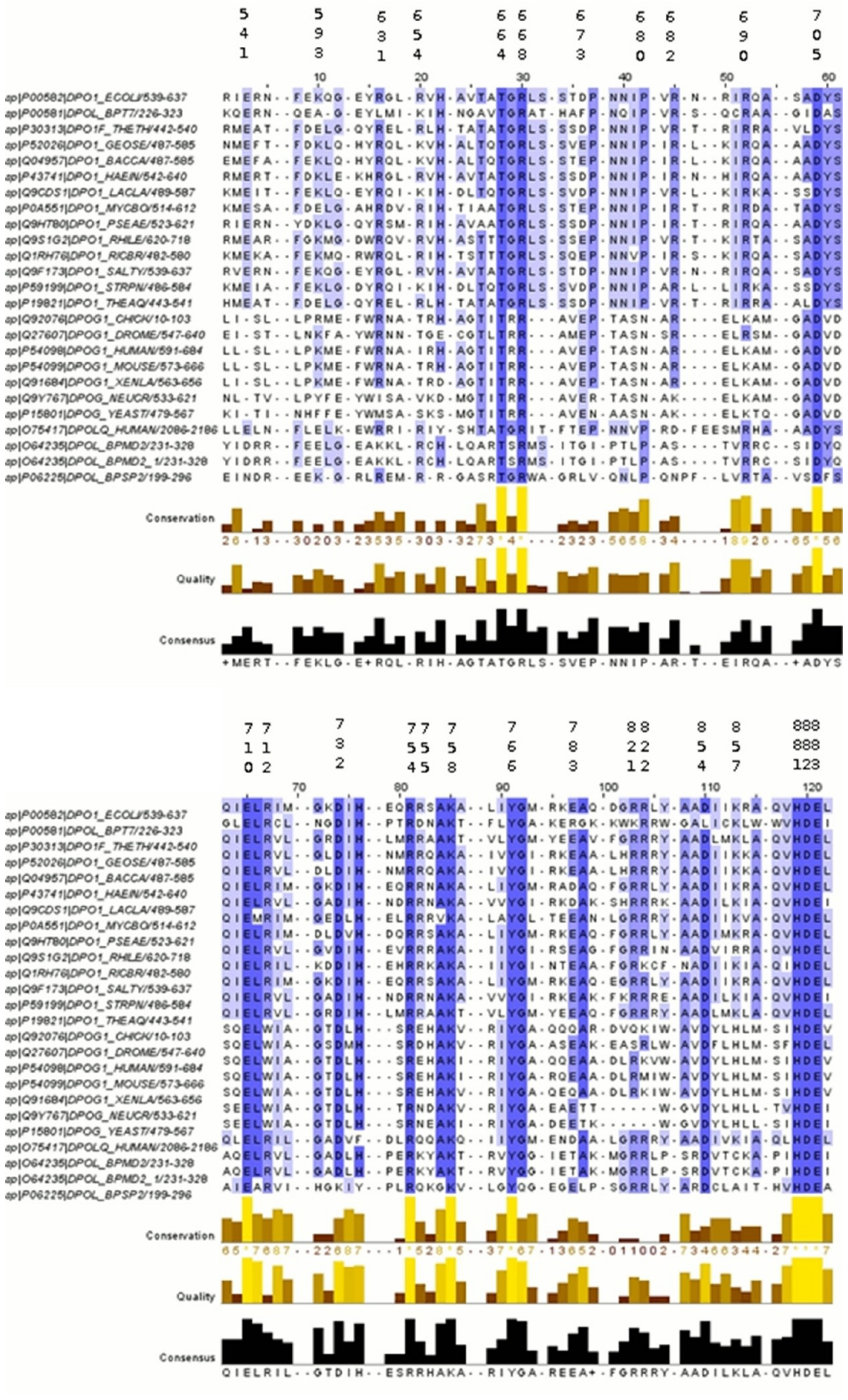


Figure 5. Condensed alignment for family A DNA polymerases. Labeled residues correspond to the KF (UniProtKB entry P00582).

The change of Q849 to an Ala in the KF has been shown to weaken DNA binding while maintaining dNTP binding.[45] This was also tested by Singh and Modak, who reported the presence of a hydrogen bond track that includes N845, Q849, R668, H881, and Q677.[46] All of the equivalent residues in the BF (793, 797, 615, 829, and 624) were shown to have an interaction with all three mispairs. The Q849A mutant and the R754A and H881A mutants have been shown to have an increase in replication accuracy against extension of a T:G mispair.[44] Residues R754 and H881 (702 and 829, respectively,

dx.doi.org/10.1021/bi201856m | *Biochemistry* 2012, 51, 2569−2578

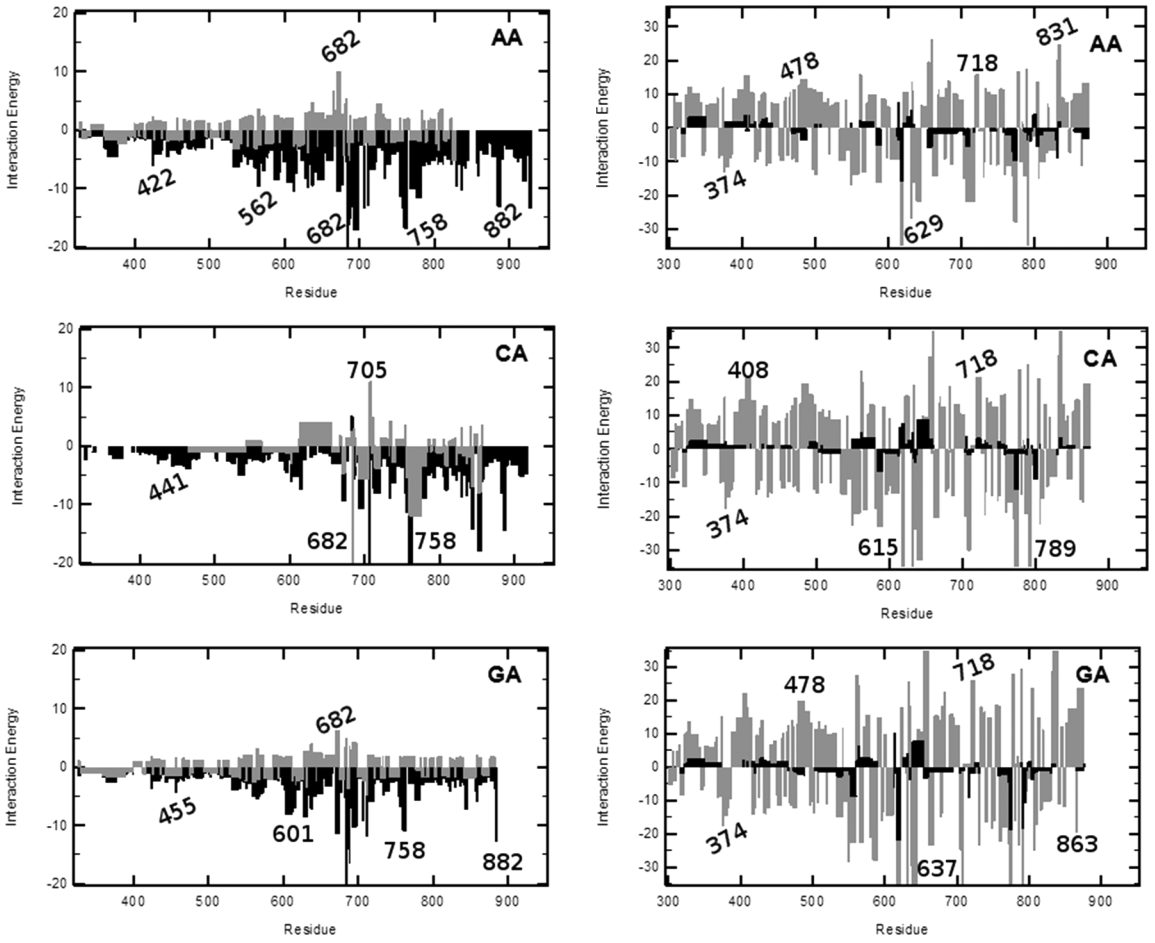


Figure 6. EDA and EFER results for the KF (left) and the BF (right). The EFER results are colored gray and the EDA results black. The interaction energy on the ordinate axis refers to intermolecular interactions ($\Delta\Delta E$ for EDA or ΔG for EFER) in kilocalories per mole as described in Methods.

in the BF) were found to have an interaction for all three mispairs in both the KF and the BF, and Q849 had significant interaction energy for all three mispairs in the BF and for the C:A mispair in the KF.

Loh et al. have reported several tolerated amino acid substitutions that alter fidelity without a loss of activity in the KF.[7] Their study also pointed out that many mutations that result in an antimutator phenotype mapped to residues in helix M (residues 720−728 in the KF). Two of the residues, D720 and K721, are denoted as being significant by EDA of the KF but not in the BF. Their study also highlights D854 as being highly conserved across eukaryotic, prokaryotic, and viral evolution, and D854 is thought to be likely to be critical for polymerase function.[7]

These experimental studies correlate KF fidelity to several of the residues obtained from our analyses for both the KF and the BF. There have been no mutagenesis studies examining fidelity in the BF, although we would expect similar results on the basis of sequence homology, as multiple mutagenesis studies on other A family polymerases, particularly the Klentaq fragment, have found residues that affect replication fidelity. Our computational results suggest that (at least part of) the discrimination arises from altered nonbonded interactions between the DNA and certain protein residues. On the basis of the single-molecule results of Christian et al.[12] and our results, one possibility could be that these residues are involved in the proposed fidelity-checking site that discriminates proper base pairs from mispairs after nucleotide incorporation.

■ CONCLUSIONS

Potential residues involved in fidelity checking that could also be involved in the putative postinsertion fidelity-checking site of KF and BF have been computationally studied using EDA, EFER, and NCI analyses. Eight systems including one properly paired and three mispaired nucleotide bases have been employed for both the KF and the BF. The EDA and EFER results show that ∼30 residues show a change in interaction energy difference for all mispairs. Of these, most of the amino acids are in the polymerase domain, and only a few correspond to the exonuclease domain. In general, the majority of the residues show altered electrostatic interactions with the terminal primer−template pair. NCI analysis gives 37 residues for the KF and 26 for the BF with noncovalent interactions. A sequence alignment of nonredundant A family polymerases reveals that 26 of the residues obtained from EDA of the KF are conserved, including 10 of the residues identified from the NCI results. For the BF, eight of the residues that are conserved across the family were found in the NCI analysis. Comparison with experimental studies shows that mutation of some of the identified residues affects the fidelity of DNA synthesis in the KF. Our results also agree with the selectivity of misinsertions reported by experimental studies. The identified residues

dx.doi.org/10.1021/bi201856m | *Biochemistry* 2012, 51, 2569−2578

are likely to play a role in the fidelity of DNA polymerase I and provide possible targets for mutagenesis experiments to further investigate the fidelity-checking mechanism hypothesis.

■ ASSOCIATED CONTENT

Supporting Information

Intermolecular interaction energy tables for all residues with significant interactions, NCI figures for A:A, C:A, G:A, and T:A pairs, and rmsd plots for apo-KF and all systems considered. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: andres@chem.wayne.edu. Phone: (313) 577-2571.

Funding

Financial support and computing time from Wayne State University (WSU) and WSU computing are gratefully acknowledged.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Drs. L. J. Romano and D. Rueda for stimulating discussion and Dr. L. Perera for the initial version of the EDA program and critical reading of the manuscript.

■ REFERENCES

(1) Thomas, A. K. (2003) Considering the cancer consequences of altered DNA polymerase function. *Cancer Cell 3* (2), 105−110.

(2) Schaaper, R. M. (1993) Base Selection, Proofreading, and Mismatch Repair during DNA Replication in *Escherichia coli*. *J. Biol. Chem. 268* (32), 23762−23765.

(3) Kuchta, R. D., Benkovic, P., and Benkovic, S. J. (1988) Kinetic mechanism whereby DNA polymerase I (Klenow) replicates DNA with high fidelity. *Biochemistry 27* (18), 6716−6725.

(4) Kiefer, J. R., et al. (1998) Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature 391* (6664), 304−307.

(5) Bebenek, K., et al. (1990) The fidelity of DNA synthesis catalyzed by derivatives of *Escherichia coli* DNA polymerase I. *J. Biol. Chem. 265* (23), 13878−13887.

(6) Carroll, S. S., Cowart, M., and Benkovic, S. J. (1991) A mutant of DNA polymerase I (Klenow fragment) with reduced fidelity. *Biochemistry 30* (3), 804−813.

(7) Loh, E., Choe, J., and Loeb, L. A. (2007) Highly Tolerated Amino Acid Substitutions Increase the Fidelity of *Escherichia coli* DNA Polymerase I. *J. Biol. Chem. 282* (16), 12201−12209.

(8) Thompson, E. H. Z., et al. (2002) Determinants of DNA Mismatch Recognition within the Polymerase Domain of the Klenow Fragment. *Biochemistry 41* (3), 713−722.

(9) Suzuki, M., et al. (2000) *Thermus aquaticus* DNA Polymerase I Mutants with Altered Fidelity. *J. Biol. Chem. 275* (42), 32728−32735.

(10) Tosaka, A., et al. (2001) O-helix Mutant T664P of *Thermus aquaticus* DNA Polymerase I. *J. Biol. Chem. 276* (29), 27562−27567.

(11) Yoshida, K., et al. (2001) Arg660Ser mutation in *Thermus aquaticus* DNA polymerase I suppresses T → C transitions: Implication of wobble base pair formation at the nucleotide incorporation step. *Nucleic Acids Res. 29* (20), 4206−4214.

(12) Christian, T. D., Romano, L. J., and Rueda, D. (2009) Single-molecule measurements of synthesis by DNA polymerase with base-pair resolution. *Proc. Natl. Acad. Sci. U.S.A. 106* (50), 21109−21114.

(13) Ibarra, B., et al. (2009) Proofreading dynamics of a processive DNA polymerase. *EMBO J. 28* (18), 2794−2802.

(14) 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* (22), 5739−5753.

(15) 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* (27), 8163−8177.

(16) 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* (19), 6819−6824.

(17) Xiang, Y., et al. (2008) Exploring the role of large conformational changes in the fidelity of DNA polymerase β. *Proteins: Struct., Funct., Bioinf. 70* (1), 231−247.

(18) Ishikita, H., and Warshel, A. (2008) Predicting Drug-Resistant Mutations of HIV Protease. *Angew. Chem., Int. Ed. 47* (4), 697−700.

(19) Rucker, R., Oelschlaeger, P., and Warshel, A. (2010) A binding free energy decomposition approach for accurate calculations of the fidelity of DNA polymerases. *Proteins: Struct., Funct., Bioinf. 78* (3), 671−680.

(20) Ram Prasad, B., and Warshel, A. (2011) Prechemistry versus preorganization in DNA replication fidelity. *Proteins: Struct., Funct., Bioinf. 79* (10), 2900−2919.

(21) Warshel, A., et al. (2006) Electrostatic Basis for Enzyme Catalysis. *Chem. Rev. 106* (8), 3210−3235.

(22) Muegge, I., Tao, H., and Warshel, A. (1997) A fast estimate of electrostatic group contributions to the free energy of protein-inhibitor binding. *Protein Eng. 10* (12), 1363−1372.

(23) Senn, H. M., O'Hagan, D., and Thiel, W. (2005) Insight into Enzymatic C−F Bond Formation from QM and QM/MM Calculations. *J. Am. Chem. Soc. 127* (39), 13643−13655.

(24) Cisneros, G. A., et al. (2003) Ab Initio QM/MM Study Shows There Is No General Acid in the Reaction Catalyzed by 4-Oxalocrotonate Tautomerase. *J. Am. Chem. Soc. 125* (34), 10384−10393.

(25) Cisneros, G. A., et al. (2004) The Protein Backbone Makes Important Contributions to 4-Oxalocrotonate Tautomerase Enzyme Catalysis: Understanding from Theory and Experiment. *Biochemistry 43* (22), 6885−6892.

(26) Cisneros, G. A., et al. (2008) Catalytic mechanism of human DNA polymerase λ with Mg^{2+} and Mn^{2+} from ab initio quantum mechanical/molecular mechanical studies. *DNA Repair 7* (11), 1824−1834.

(27) Cisneros, G. A., et al. (2009) Reaction Mechanism of the ε Subunit of *E. coli* DNA Polymerase III: Insights into Active Site Metal Coordination and Catalytically Significant Residues. *J. Am. Chem. Soc. 131* (4), 1550−1556.

(28) Martí, S., et al. (2003) Preorganization and Reorganization as Related Factors in Enzyme Catalysis: The Chorismate Mutase Case. *Chem.—Eur. J. 9* (4), 984−991.

(29) Szefczyk, B., et al. (2007) Quantum chemical analysis of reaction paths in chorismate mutase: Conformational effects and electrostatic stabilization. *Int. J. Quantum Chem. 107* (12), 2274−2285.

(30) Li, G., and Cui, Q. (2003) What Is So Special about Arg 55 in the Catalysis of Cyclophilin A? Insights from Hybrid QM/MM Simulations. *J. Am. Chem. Soc. 125* (49), 15028−15038.

(31) Cui, Q. (2003) Catalysis and specificity in enzymes: A study of triosephophate isomerase and comparison with methyl glyoxal synthase. In *Advances in Protein Chemistry*, pp 315−372, Academic Press, New York.

(32) Metanis, N., et al. (2004) Electrostatic Interactions Dominate the Catalytic Contribution of Arg39 in 4-Oxalocrotonate Tautomerase. *J. Am. Chem. Soc. 126* (40), 12726−12727.

(33) Bebenek, K., et al. (2010) Loop 1 modulates the fidelity of DNA polymerase λ. *Nucleic Acids Res. 38* (16), 5419−5431.

(34) Klvaňa, M., et al. (2011) An Abridged Transition State Model To Derive Structure, Dynamics, and Energy Components of DNA Polymerase β Fidelity. *Biochemistry 50* (32), 7023−7032.

(35) Johnson, E. R., et al. (2010) Revealing Noncovalent Interactions. *J. Am. Chem. Soc. 132* (18), 6498−6506.
(36) Li, Y., Korolev, S., and Waksman, G. (1998) Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: Structural basis for nucleotide incorporation. *EMBO J. 17* (24), 7514−7525.
(37) Beese, L. S., Friedman, J. M., and Steitz, T. A. (1993) Crystal structures of the Klenow fragment of DNA polymerase I complexed with deoxynucleoside triphosphate and pyrophosphate. *Biochemistry 32* (51), 14095−14101.
(38) Case, D. A., et al. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem. 26* (16), 1668−1688.
(39) Hsu, G. W., et al. (2004) Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase. *Nature 431* (7005), 217−221.
(40) Jorgensen, W. L., et al. (1983) *Comparison of simple potential functions for simulating liquid water*, Vol. *79*, pp 926−935, American Institute of Physics, College Park, MD.
(41) Essmann, U. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys. 103* (19), 8577.
(42) Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys. 98* (12), 10089−10092.
(43) Contreras-García, J., et al. (2011) NCIPLOT: A Program for Plotting Noncovalent Interaction Regions. *J. Chem. Theory Comput. 7* (3), 625−632.
(44) Minnick, D. T., et al. (1999) Side Chains That Influence Fidelity at the Polymerase Active Site of *Escherichia coli* DNA Polymerase I (Klenow Fragment). *J. Biol. Chem. 274* (5), 3067−3075.
(45) Loh, E., and Loeb, L. A. (2005) Mutability of DNA polymerase I: Implications for the creation of mutant DNA polymerases. *DNA Repair 4* (12), 1390−1398.
(46) Singh, K., and Modak, M. J. (2003) Presence of 18-Å Long Hydrogen Bond Track in the Active Site of *Escherichia coli* DNA Polymerase I (Klenow Fragment). *J. Biol. Chem. 278* (13), 11289−11302.