DNA Pol lambda's Extraordinary Ability To Stabilize Misaligned DNA

ARTICLE *in* JOURNAL OF THE AMERICAN CHEMICAL SOCIETY · SEPTEMBER 2010

Impact Factor: 12.11 · DOI: 10.1021/ja1049687 · Source: PubMed

Published in final edited form as:
J Am Chem Soc. 2010 September 29; 132(38): 13403–13416. doi:10.1021/ja1049687.

The Extraordinary Ability of DNA Pol λ to Stabilize Misaligned DNA

Meredith C. Foley[*], **Victoria A. Padow**[*,†], and **Tamar Schlick**[*,‡]

[*]Department of Chemistry and Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, New York 10012

[†]Program in Neuroscience and Behavior, Barnard College, 3009 Broadway, New York, New York 10027

Abstract

DNA polymerases have the venerable task of maintaining genome stability during DNA replication and repair. Errors, nonetheless, occur with error propensities that are polymerase specific. For example, DNA polymerase λ (pol λ) generates single-base deletions through template strand slippage within short repetitive DNA regions much more readily than does the closely related polymerase *β* (pol *β*). Here we present *in silico* evidence to help interpret pol λ's greater tendency for deletion errors than pol *β* by its more favorable protein/DNA electrostatic interactions immediately around the extrahelical nucleotide on the template strand. Our molecular dynamics (MD) and free energy analyses suggest that pol λ provides greater stabilization to misaligned DNA than aligned DNA. Our study of several pol λ mutants of Lys544 (Ala, Phe, Glu) probe the interactions between the extrahelical nucleotide and the adjacent Lys544 to show that the charge of the 544 residue controls stabilization of the DNA misalignment. In addition, we identify other thumb residues (Arg538, Lys521, Arg517, and Arg514) that play coordinating roles in stabilizing pol λ's interactions with misaligned DNA. Interestingly, their aggregate stabilization effect is more important than any one component residue in contrast to aligned DNA systems, as we determined from mutations of these key residues and energetic analyses. No such comparable network of stabilizing misaligned DNA exists in pol *β*. Evolutionary needs for DNA repair on substrates with minimal base pairing, such as those encountered by pol λ in the non-homologous end-joining pathway, may have been solved by a greater tolerance to deletion errors. Other base flipping proteins share similar binding properties and motions for extrahelical nucleotides.

Keywords

DNA polymerase; pol λ; pol *β*; DNA slippage; Deletion error; Molecular Dynamics simulations; Protein/DNA complex; Conformational changes

1 Introduction

Since the genome is the master regulator of an organism, any modification can be potentially very serious. Indeed, some changes are potentially more damaging than others. In particular, insertion or deletion errors within protein-coding regions of the genome, or frameshift mutations, have the potential to alter the reading frame for protein translation and thus can

[‡] To whom correspondence should be addressed (Phone: 212-998-3116; schlick@nyu.edu; fax: 212-995-4152).

Supporting Information **Available:** Supplemental Tables S1(a–d) and S2, Figures S1–S12, and complete references 60 and 83. This material is available free of charge via the Internet at http://pubs.acs.org.

lead to changes in amino acid composition as well as truncated proteins. The human genome contains many microsatellite or simple sequence repeats that are hot spots for insertion/deletion errors. Polymorphisms within these regions increase phenotype variability and have been found to determine susceptibilities to disease and drug response. Microsatellite instability is a feature of several cancers. Numerous neurological conditions such as Alzheimer's disease, Huntington's disease, and Friedreich's ataxia are associated with specific trinucleotide repeat expansions [1]. Intrinsic properties of repetitive DNA such as non-B-DNA structure formation and base stacking contribute to its mutability [2].

Although DNA polymerases function to replicate and repair DNA, they can themselves introduce various errors. Some polymerases make more insertion/deletion errors than others [3]. Often there is a tradeoff between the ability to tackle complex breaks and lesions and the propensity to generate errors. For example, low fidelity DNA polymerase μ (pol μ) operates within the non-homologous end-joining (NHEJ) pathway to fill in gaps between two different ends of DNA without any base-pairing overlap between the strands [4]. Error-prone Y-family polymerases are capable of bypassing bulky lesions such as benzo[*a*]pyrene DNA adducts that stall high fidelity polymerases [5]. Understanding what features are responsible for creating certain errors is important both to comprehend the underlying mechanism and to assess possible pharmacological interventions to control the process and limit the generation of the most dangerous types of errors.

All DNA polymerases catalyze a nucleotidyl transfer reaction and are shaped like a hand with fingers, palm and thumb subdomains [6] (Figure 1). X-family mammalian DNA polymerases β (pol β) and λ (pol λ) have an additional 8-kDa domain with 5′-deoxyribose-5-phosphate lyase function [7, 8]. Pol λ also has an N-terminal BRCT domain, attached through a serine/proline linker region to its hand-shaped polymerase domain, for mediating interactions with other proteins [9, 10]. Both pol β and pol λ lack a proofreading domain for editing DNA replication errors. In mammalian cells, these polymerases fill short gaps in DNA within the base excision repair (BER) pathway [11, 12, 13, 14, 15]; pol λ is thought to operate in a back-up capacity to pol β within BER. Pol λ also participates in NHEJ [16, 17, 18, 4, 19], which is an error-prone pathway for double-strand break repair.

Although pol β is thought to be the predominant cause of deletion errors within the BER pathway [20], experimental data indicate that pol λ has a much higher intrinsic tendency for deletions [21]. DNA pols λ and β produce single-base deletions through classical DNA template-strand slippage [21, 22]. In the DNA slippage mechanism, deletion errors result from the extension of misaligned repetitive DNA [23]. In addition to mononucleotide repeats, repetitive DNA regions can include di-, tri-, and other nucleotide sequence repeats. Pol β has been shown to make deletion errors consistent with DNA slippage within di- and tetranucleotide microsatellite repeat sequences [24]. Insertion errors can also occur through DNA slippage, arising from primer-strand misalignments; however, these mistakes are generally less common than deletion errors in both pol λ and pol β [21, 25].

In addition to DNA slippage, experimental and structural studies suggest that deletions can occur through dNTP-stabilized misalignment and misinsertion-mediated misalignment mechanisms [26, 27], particularly in non-repetitive DNA sequences. These mechanisms differ in the stage where DNA misalignment occurs in relation to the chemical reaction. In dNTP-stabilized misalignment, the misalignment occurs prior to the chemical reaction if the nucleotide can correctly pair with the next available template base. In misinsertion-mediated misalignment, an incorrect nucleotide is first incorporated into the DNA, and then a misalignment occurs if the incorrect nucleotide can properly base pair with the next available template base.

Y-family Dpo4, known for translesion bypass, frequently causes single-base deletions using the dNTP-stabilized misalignment mechanism on both lesioned [28, 29, 30,] and nonlesioned [31, 32,] DNA; this mechanism may be preferred because Dpo4's spacious active site easily accommodates two template bases. In many DNA polymerases, DNA lesions stimulate insertion and deletion errors [3,]. Both pol λ and pol *β* make deletions when they encounter an abasic site [33, 34,]. During BER of oxidative damage within CAG repeats, pol *β* can trigger repeat expansion [35, 36, 37,], which has relevance to the progression of Huntington's disease as well as a number of other neurodegenerative disorders. Pol *μ*, another X-family polymerase, utilizes a DNA slippage mechanism to bypass several different lesions [38, 39,].

In pol *β*, like many other DNA polymerases, deletion errors increase with the length of the repetitive DNA region [3,]. Several factors are hypothesized to account for this trend such as substantial protein/DNA contacts and minor groove interactions upstream of the active site that restrict the geometry of the DNA and prevent misalignments [11, 3,]. Within long repetitive regions, the probability also increases that the misalignment will be located farther away from the polymerase active site where less-restrictive polymerase/DNA interactions may exist. Misalignments within lengthy repetitive regions are also better stabilized by an abundance of surrounding correct base pairs [11, 3,]. Within repetitive sequences, single-stranded structures such as hairpins can also form, which stabilize misalignments [1, 40,]. The greater the stability of the misalignment, the less likely it will be corrected by a proofreading domain [11, 3,]. However, for pol λ, the tendency to generate a single-base deletion error is equally as high in a two-base repeat as in a five-base repeat of identical bases [21,]. This suggests that pol λ's tracking mechanism is particularly weak when compared to some other polymerases. A structural, energetic, and functional analysis of this polymerase in comparison to other enzymes may yield important insights regarding how fidelity is maintained during replication of repetitive DNA.

Both experimental and computational studies show that, upon pol λ's binding a correct incoming nucleotide, large DNA template-strand shifting normally occurs from the inactive to active positions as a normal component of its catalytic cycle [41, 42,]. As we proposed with the Kunkel group, this DNA motion likely provides an opportunity for the DNA to slip or misalign [43, 44,]. In pol *β*, much smaller DNA motions and large-scale subdomain motions occur upon binding the correct incoming nucleotide that transition the polymerase/DNA complex from an open to closed state [45, 46, 47, 48, 49, 50, 51]. In contrast, pol λ is in a closed protein subdomain conformation both before and after correct nucleotide binding [41, 52,]. In both enzymes, a series of active-site residues, acting as “gate-keepers”, regulate the transition from inactive to active states through side-chain rearrangements [53, 42,]. Our simulations have revealed that incorrect nucleotide insertion by these polymerases is rendered less favorable by the formation of more disordered active sites [54, 55,].

Mutation studies of both pol λ and pol *β* implicate an analogous arginine residue in the thumb as an important fidelity regulator. The increased rates of single-base deletion errors found in pol λ Arg517Ala/Lys [43,] and pol *β* Arg283Ala/Lys mutants [26,] suggests that the wild-type arginine residues hamper DNA misalignments that lead to deletion errors. Indeed, our simulations of several Arg517 pol λ mutants (Ala, Lys, His, Glu, Met, Gln) bound to properly aligned DNA show increased DNA template-strand shifting between active and inactive positions compared to the wild-type enzyme [42, 43, 44,]. The extent of the motion in these mutants depends on the size and charge of the mutant residue, with the Lys mutant showing the least and the Met mutant shows the greatest amount of DNA motion. The X-ray crystal structure, obtained by the Kunkel group and co-workers, of the pol λ Arg517Ala mutant also suggests DNA template-strand flexibility since two positions of the DNA

template strand are resolved [43,]. Alterations in the hydrogen-bonding network of Arg517 were also found to increase DNA motion in simulations of pol λ bound to mismatches [55,].

Recently, pol λ was crystallized by the Kunkel group and co-workers bound to “slipped” or misaligned DNA with a repeat consisting of two identical template-strand bases both before and after correct nucleotide insertion, representing important steps in the generation of a deletion error [56,]. In these structures, one template-strand residue, located two base pairs upstream from the active-site, is extrahelical and interacts with the thumb subdomain as shown in Figure 1. Aside from this DNA template strand distortion, the rest of the polymerase/DNA complex overlays well with X-ray crystal structures of pol λ bound to normally aligned DNA and the correct incoming nucleotide. This structural similarity corresponds with kinetics data that suggest, intriguingly, that pol λ utilizes misaligned DNA as efficiently as properly aligned DNA for catalyzing the nucleotidyl transfer reaction [56,]. Such structures of deletion error intermediates also underscore pol λ's limited base-pairing requirements for chemistry since correct pairing of only the nascent and adjacent primer terminus base pairs produce a chemically-competent complex.

Understanding how pol λ stabilizes misaligned DNA and utilizes this substrate to perform correct nucleotide insertion on the atomic level is key to understanding this polymerase's tendency for making deletion errors. In addition, a detailed comparison with pol *β* can help interpret pol *β*'s reduced ability to generate deletion errors and preferred deletions on longer repetitive runs. More generally, this knowledge could aid our understanding of the interactions that occur when polymerases make deletion errors on repetitive DNA. This type of DNA is common in both coding and noncoding regions of the human genome, and mutations in these regions can lead to cancer and many neurological disorders.

Toward these ends, we utilize a molecular modeling and dynamics simulation approach to sample the motions and energetics of several polymerase/DNA complexes over a nanosecond time frame. Such biomolecular simulations have yielded important insights into the behavior of a wide-range of systems [57, 58,]. Nonetheless, biomolecular simulations are subject to many, well-recognized inherent limitations such as force-field uncertainties, solvent approximations, limited sampling, and finite size effects. Like other standard and widely-accepted force fields, CHARMM [59, 60,] is subject to these general approximations. In particular, the force field's treatment of divalent ions, which are modeled based on van der Waals (described by the phenomenological Lennard-Jones potential) and Coulombic interactions [61, 62, 63, 64,], may translate into shorter ligand/ion distances than those observed in high resolution X-ray crystal structures. Inclusion of divalent ions like Mg^{2+} in polymerase models is necessary since they play important roles in both the preparation and evolution of the “two-metal-ion” nucleotidyl transfer mechanism [65,]. Since our study focuses on general trends in Mg^{2+} ion coordination and the possible long-range effects of the ions on DNA and protein rearrangements, and involves systematic comparisons of the trends among closely-related systems, these divalent ion force field limitations are acceptable. When analyzing biomolecular simulation data, comparison of results for closely-related systems is useful to detect any simulation-dependent artifacts. General trends are also typically more informative than detailed quantitative analyses. Ultimately, the best validation of modeling and simulation work is derived from the predictive power of its findings as supported by subsequent experimental data.

In this work, we describe analyses based on several simulations of wild-type and mutant forms of pol λ bound to misaligned DNA to elucidate the factors that determine this enzyme's propensity for generating deletion errors. We find that a significant factor is the number of points of contact between pol λ's thumb and the DNA closest to the extrahelical nucleotide that stabilize misaligned DNA (see positions of residues in Figure 2).

Interestingly, these residues (Lys544, Arg538, Lys521, Arg517) function as a group, almost like a moveable clamp, and replacement of one residue with alanine does not destroy the overall function of the complex. It also appears that Arg514, which stacks with the templating base, may provide further DNA stabilization in the active site. Moreover, energetic analyses of both misaligned and aligned DNA complexes reveal that this cooperative residue effect on DNA binding does not occur when the DNA is aligned. The significance of this proposed effect is further supported by a separate analysis we conducted showing a greater binding free energy for pol λ bound to misaligned DNA compared to aligned DNA. For comparison, we also analyze simulations of pol *β* bound to misaligned DNA modeled from the pol λ X-ray crystal structure. Together, our studies reveal that thumb/DNA interactions present in pol λ, but absent in pol *β*, allow pol λ to stabilize misalignments close to its active site to help explain the different substrate specificities of these enzymes. Remarkably, pol λ stabilizes misaligned DNA better than aligned DNA. We also raise interesting mechanistic similarities in binding misaligned DNA for base-flipping in pol λ with other types of proteins.

2 Computational Methods

2.1 Initial Models (Wild-type Pol λ and Pol *β* plus Pol λ Lys544Ala, Lys544Phe, Lys544Glu, Lys521Ala, Arg517Ala, and Arg538Ala Mutants

As shown in Table 1, seven initial pol λ models are prepared based on the X-ray crystal pol λ/DNA/dTTP ternary complex in which the primer-template is misaligned and poised for a deletion error (PDB entry 2BCV). To assess the roles of size and charge of Lys544, located in the thumb loop containing *β*-strand 8, on the stabilization of the misaligned DNA structure, we mutated Lys544 to alanine (Ala), phenylalanine (Phe), or glutamate (Glu), respectively, in three models. The effects of other nearby positively-charged residues (i.e., Lys521, Arg517, and Arg538) are also probed by mutating these residues to alanine in three other models. The last pol λ model is left unchanged to serve as our control.

In all pol λ models (see the wild-type system shown in Figure 1), the sodium ion in the catalytic ion position is changed to magnesium, the hypothesized species of the divalent ion required for DNA polymerase activity. Protein residues 1–12 and all other atoms not resolved in the X-ray crystal structure are also added. To correct the primer terminus, which contained the unreactive 3′-deoxycytidine 5′-monophosphate nucleotide, an oxygen atom is added to the 3′ carbon and an oxygen atom is removed from the 2′ carbon so that the natural 2′-deoxycytidine 5′-monophosphate (dC) nucleotide is formed. The overall geometry of our modeled wild-type pol λ active site remains very similar to the crystal structure geometry following 20 ns of simulation (Figure S1 in the Supporting Information).

To interpret the lower deletion error rates of pol *β* on an atomic-level, we modeled pol *β* bound to misaligned DNA. This model is built from the X-ray crystal structure of the pol *β*/DNA/dUMPNPP ternary complex (PDB entry 2FMS). The DNA and dUMPNPP in the pol *β* complex are replaced with the misaligned DNA and dTTP in the pol λ models after superimposing the pol *β* protein C_α atoms onto those of pol λ. Missing pol *β* residues 1–9 and all other atoms not resolved in the X-ray crystal structure are built.

In all pol λ complexes, optimized periodic boundary conditions in a cubic cell are introduced using the PBCAID program [66,]. The smallest image distance between the solute, the protein complex, and the faces of the periodic cubic cell is 10 Å. The pol *β* model was solvated in a water box using the VMD program [67,]. We modeled all systems in a neutral form with an 150 mM ionic strength by using the Delphi package [68,] to calculate the electrostatic potential of all bulk water (TIP3 model) oxygen atoms. Those water oxygen atoms with minimal electrostatic potential are replaced with Na^+ and those with maximal electrostatic potential

are replaced with Cl^-. In placing the ions, a separation of at least 8 Å was maintained between the Na^+ and Cl^- ions and between the ions and protein or DNA atoms.

As shown by the model of the wild-type pol λ system in Figure 1, all pol λ models contain approximately 37,837 atoms, 330 crystallographically resolved water molecules, 10,260 bulk water molecules, two Mg^{2+} ions, an incoming nucleotide, and 34 Na^+ and 23 Cl^- counterions. The final dimensions of the box are: 68.96 Å × 73.02 Å × 69.12 Å.

For pol *β*, the model is slightly larger in size with 42,780 atoms, 292 crystallographically resolved water molecules, 11,897 bulk water molecules, two Mg^{2+} ions, an incoming nucleotide, and 40 Na+ and 27 Cl^- counterions. The simulation box dimensions are: 71.41 Å × 74.67 Å × 73.59 Å.

2.2 Minimization, Equilibration and Dynamics Protocol

All pol λ and pol *β* model systems are energy minimized and equilibrated using the CHARMM program [61,] with the all-atom CHARMM protein and nucleic acid force field [59, 60,]. First, each system is minimized with fixed positions for all protein and nucleic heavy atoms except those from the added residues using SD for 5,000 steps followed by ABNR for 10,000 steps. Three cycles of further minimization are carried out for 10,000 steps using SD followed by 20,000 steps of ABNR. During these minimizations, the Cl^-, Na^+ and water relax around the protein/DNA complex. The equilibration process is started with a 30 ps simulation at 300 K using single-timestep Langevin dynamics and keeping the constraints used in the previous minimization step. The SHAKE algorithm is employed to constrain the bonds involving hydrogen atoms. This is followed by unconstrained minimization using 10,000 steps of SD followed by 20,000 steps of ABNR and a further 30 ps of equilibration at 300 K.

Production dynamics are performed using the NAMD program [69,] with the CHARMM force field [59, 60,]. First, the energy in each system is minimized using the Powell conjugate gradient algorithm. Systems are then equilibrated for 100 ps at constant pressure and temperature. Pressure is maintained at 1 atm using the Langevin piston method [70,], with a piston period of 100 fs, a damping time constant of 50 fs, and piston temperature of 300 K. Temperature coupling was enforced by velocity reassignment every 2 ps. The water and ions are further energy minimized and equilibrated at constant volume and temperature for 30 ps at 300 K while holding all protein and DNA heavy atoms fixed. This is followed by minimization and 30 ps of dynamics at 300 K on the entire system. Then, production dynamics are performed at constant temperature and volume. The temperature is maintained at 300 K using weakly coupled Langevin dynamics of non-hydrogen atoms with damping coefficient $\gamma = 10\ ps^{-1}$ used for all simulations performed; bonds to all hydrogen atoms are kept rigid using SHAKE [71,], permitting a time step of 2 fs. The system is simulated in periodic boundary conditions, with full electrostatics computed using the PME method [72,] with grid spacing on the order of 1 Å or less. Short-range nonbonded terms are evaluated every step using a 12 Å cutoff for van der Waals interactions and a smooth switching function. The total simulation length for all pol λ systems and the pol *β* system is 20 ns and 10 ns, respectively.

Simulations using the NAMD package were run on local and NCSA SGI Altix 3700 Intel Itanium 2 processor shared-memory systems running the Linux operating system.

2.3 Electrostatic Potential Surface Calculations

The equilibrated models of wild-type pol λ and pol *β* bound to misaligned DNA are used for calculations of the thumb subdomain electrostatic potential with the QNIFFT program [73, 74,]. The parameters used include a macromolecule dielectric constant of 2, a solvent dielectric

constant of 78.357, a 1.4 Å water probe, a 2 Å ion exclusion radius, and a 150 mM salt concentration. The CHARMM force field was used to generate radius and charge data. To calculate the potentials, we used a focusing procedure with our polymerase/DNA/dTTP systems positioned in a 65 Å × 65 Å × 65 Å cubic lattice. In each initial calculation, our pol β or pol λ system occupied 19–23% of the lattice, corresponding to 0.2 grids per Å. Following two or three rounds of focusing, each system occupied ~94–99% of the lattice, corresponding to 0.86 or 0.96 grids per Å. These calculations could help elucidate environmental differences in the polymerase thumb binding region for the extrahelical nucleotide in the misaligned DNA substrate.

2.4 MM-PBSA Calculations

We use the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) method [75, 76, 77, 78,] to compute the relative binding free energy ($\Delta\Delta G$) of aligned and misaligned DNA in pol λ. This method for computing free energies is essentially a post-processing technique for our 20 ns MD simulations of pol λ aligned DNA and misaligned DNA trajectories that requires removal of waters, ions, and the incoming nucleotide. All calculations are performed using the c35b2 version of the CHARMM program [61,] with the CHARMM force field [59, 60,]. Full details for the pol λ/aligned DNA simulation can be found in [42,].

In the MM-PBSA approach, the binding free energy is calculated from the free energy difference between the polymerase/DNA complex and the two unbound components of the complex. The free energy of the complex, protein, and DNA is calculated separately according to:

$$G = E_{MM} + G_{solv} - TS \quad (1)$$

where E_{MM} is the molecular mechanical energy, G_{solv} is the solvation energy, and $-TS$ is the solute entropic contribution. E_{MM} is the sum of the internal (e.g., bonds, angles, dihedrals), van der Waals, and electrostatic components. G_{solv} is the sum of the hydrophobic energy (G_{np}) and the electrostatic solvation energy (G_{PB}). The nonpolar contribution to the solvation free energy is calculated from: $G_{np} = \gamma\text{SASA} + \beta$, where $\gamma = 0.00542$ kcal/Å^2, $\beta = 0.92$ kcal/mol, and SASA is the solvent accessible surface area, which is determined using a water probe radius equal to 1.4 Å. The polar contribution to the solvation free energy (G_{PB}) is determined by solving the finite-difference Poisson-Boltzmann equation using the PBEQ module in CHARMM [79,]; these calculations were run at 300 K with a solute dielectric constant of 1.0, solvent dielectric constant of 80, reference gas phase dielectric constant of 1.0, water probe radius of 1.4 Å, monovalent salt concentration of 0.15 M, and ion exclusion radius of 2 Å. A two step focusing procedure was used with an initial grid spacing of 0.4 Å and a final grid spacing of 0.1 Å. A maximum of 1000 iterations were used for each calculation. The entropy was approximated from quasiharmonic analysis of the MD trajectories using the VIBRAN module in CHARMM [80, 81,].

Once the free energy of each species is calculated, we compute the binding free energy (ΔG) from:

$$\Delta G_{bind} = G(complex) - G(protein) - G(DNA), \quad (2)$$

and the relative binding free energy ($\Delta\Delta G$) for the misaligned DNA complex versus the aligned DNA complex from:

$$\Delta\Delta G_{bind} = \Delta G_{bind}\,[misaligned\ DNA\ complex] - \Delta G_{bind}\,[aligned\ DNA\ complex]. \quad (3)$$

Since these calculations are based on our MD trajectories, averages are obtained for each free energy term. To monitor the convergence of the free energy calculations, each half of the 20 ns simulation was analyzed by 625 frames (every 16 ps). As shown in Supplementary Tables S1a–S1d in the Supporting Information, the computed free energies from both halves are similar and within the error ranges. Data in Table 2 are from the second half of each simulation. For the quasiharmonic analysis, a total of 18,000 frames were used from each trajectory, spanning 3–20 ns with 1 ps frame spacing.

3 Results

3.1 Charge of Pol λ's 544 Residue Is Important for Stabilizing the Extrahelical Template Nucleotide

Our examination of the wild-type pol λ complex reveals that it contains a conserved nucleotide binding pocket that shows great similarity to base-flipping enzymes [82,]. In this binding motif for extrahelical nucleotides, a positively charged residue interacts with the DNA phosphate backbone and a hydrophobic residue lies adjacent to the nucleotide base. As shown in Figure 1d, pol λ's Lys544 in the thumb satisfies the electrostatic requirements of the positively-charged residue by forming several hydrogen bonds to the DNA phosphate backbone near the extrahelical nucleotide (T7) and Pro547 is the base-interacting hydrophobic residue.

Since Lys544 appears to be a key component of the pol λ/misaligned DNA complex, we elucidate its role in the stability of the complex by comparing wild-type pol λ to three Lys544 mutants (Ala, Phe, Glu) bound to misaligned DNA. We chose these three mutants to examine the effect of size and charge at this location: Ala for its much smaller size and neutral charge, Phe for its bulky side chain and neutral charge, and Glu for its opposite charge to lysine.

For example, a comparison shows that the charge of residue 544 is more important than its size in explaining the resulting complex's stability. While the wild-type system does not deviate significantly from the crystal, the Lys544Glu mutant exhibits significant rearrangements in both the DNA and the active site. Some rearrangements also occur in the Ala and Phe mutants.

In all systems, hydrogen bonding interactions between the residue at position 544 and T7 form when favorable electrostatic interactions occur between them. A qualitative comparison of the strength of these interactions also helps assess how well the misaligned DNA structure is stabilized. In the wild-type system, interactions between Lys544 (positively charged, hydrogen bond donor at its terminal ∈-amino group) and the thymine base of T7 are always favorable, and become even more so when hydrogen bonds form between Lys544's terminal ∈-amino group and the T7:O4′ and T7:O5′ atoms (Figure 3a,b) in addition to Lys544's interactions with the T7:O1P/O2P and T7:H3 atoms (Figure S2 in the Supporting Information); Lys544's side chain rotates to facilitate all these interactions with the DNA (Figure 3c). Similarly, in the Lys544Ala and Lys544Phe mutants, electrostatic interactions become much more favorable when a hydrogen bond forms between T7 and the backbone oxygen of the residue at position 544 (Figure 4). However, these interactions are more limited and less frequent in relation to the wild-type pol λ system. In the Lys544Glu mutant, the presence of a negative charge at position 544 leads to unfavorable electrostatic interactions between T7 and Glu544, preventing any hydrogen bond formation (Figure 4).

The range of motion of T7 and residue 544 in each of these systems is depicted in Figure 5. Even the wild-type pol λ system shows oscillations in the position of T7. The largest T7 rearrangement occurs in the Lys544Glu mutant system, in which the extrahelical residue moves significantly (~14 Å, double the wild-type range) toward reentering the helix; this is accompanied by a widening of the spacing between the surrounding T6 and T8 template bases to accommodate T7 within the helix (Figure S3 in the Supporting Information). In addition, the T7 base rotates to the *syn* conformation as shown by a change in its glycosidic torsion angle (χ) (Figure S4 in the Supporting Information). The next largest motion occurs in the Lys544Ala mutant, but T7 tends to shift in the opposite direction. In the Lys544Phe system, the DNA kinks more around the extrahelical nucleotide as revealed by a tightening of the distance between T7 and T9 backbone atoms (Figure S5 in the Supporting Information) and T7 oscillates over a small range. Changes in the hydration pattern around T7 may affect DNA motion in some of the mutant systems. In particular, replacement of lysine with alanine creates a more open pocket around T7 that is more accommodating to water molecules; these additional intervening water molecules weaken protein/T7 interactions and may account for the larger T7 range of motion in the Ala mutant compared to the Phe mutant. Despite these local changes around T7, the DNA remains in the active position in all systems (Figure S6 in the Supporting Information).

The mutations of Lys544 and DNA motions discussed above lead to changes in several active-site residues (see overlaid residues in Figure S7a in the Supporting Information). In all Lys544 mutant systems, Ile492 and Tyr505 flip to their inactive positions. Phe506 also rearranges to its inactive position in the Phe and Ala mutants. Even in the wild-type system, the positions of some active-site residues fluctuate as seen by transient rearrangements in Ile492 and Asn513 (see torsion changes in Figure S8 in the Supporting Information).

Further transitions toward an inactive state occur in the Lys544Glu and Lys544Ala mutants for active-site atoms involved in the chemical reaction and magnesium ion coordination. In these mutants, the crucial O3′–P_α distance elongates, Asp490 rotates, and more water molecules enter the active site compared to the wild-type and Lys544Phe systems (Figure 6). The additional water molecules, in particular, serve to distort the active-site from the ideal two-metal-ion arrangement by blocking the coordination of the primer terminus' O3′ atom to the catalytic ion, which allows the primer terminus to drift away from the incoming nucleotide. These combined changes hamper the chemical reaction.

Taken together, we suggest that a single mutation in the thumb at position 544 can reduce the likelihood of deletion errors by changing specific protein/DNA interactions as well as altering the active-site geometry of the complex in such a way that the process of extending a DNA misalignment is hampered. The wild-type Lys544 residue thus appears to have been evolutionarily adapted to stabilize misaligned DNA well in pol λ's active site.

3.2 The Active Pol λ Complex is More Stable When the Bound DNA is Misaligned

Intriguingly, we find that the active pol λ/misaligned DNA/dTTP complex is more stable than when the bound DNA is properly aligned. This results from pol λ's tighter grip on misaligned DNA through extra points of contact with the extrahelical nucleotide, mediated by Lys544, and other synergistic thumb/DNA interactions as shown in Figure 2. We deduced this effect from additional mutant simulations of other positively-charged residues in the thumb (Arg517, Lys521, Arg538), which interact with the DNA, and from energetic analyses. Additionally, Arg514, which stacks with the templating base at the gap, contributes to this combined effect by forming important stabilizing interactions with the active-site DNA. (See Table S5 in the Supporting Information for a summary of hydrogen bonding interactions of these five residues with the DNA).

From a structural standpoint, we found that replacement of only one of these residues with alanine produces some alteration in the complex, but does not destabilize the complex because of compensatory interactions from the other residues that we discuss in terms of thumb/DNA interaction energy below. The stronger binding of misaligned DNA compared to normally aligned DNA is further emphasized by an examination of Arg517Ala pol λ with both of these DNA types. In the misaligned DNA system, the DNA does not shift and remains in the chemically-active position (Figure S6 in the Supporting Information) while, in the aligned DNA system, frequent DNA shifting occurs between active and inactive positions, which we previously reported [42,]. Similarly, the misaligned DNA remains in the active position in the Lys521Ala and Arg538Ala mutants (Figure S6 in the Supporting Information), a commonality in all our pol λ/misaligned DNA complexes.

Nevertheless, smaller-scale DNA rearrangements within the active position occur. In the Lys521Ala and Arg517Ala systems, the increase in DNA spacing around the extrahelical T7 residue is similar to the Lys544Glu mutant (Figure S3 in the Supporting Information). Moreover, in the Arg538Ala system, a combination of fluctuations occurs in the templating base at the gap (T5) and the adjacent template base (T6) that pairs with the primer terminus (Figure S9a in the Supporting Information) in addition to intermittent movement in the thumb loop containing *β*-strand 8 toward the inactive loop position (Figure S9b,c in the Supporting Information).

The combined large motions for Arg538Ala lead to more active-site distortions compared to the other systems (see positions of residues in Figure S7b in the Supporting Information). This includes Tyr505 shifting toward the inactive position, twisting in the Phe506 side chain, and fluctuations in Ile492's position; these changes help to elongate the O3′–P_α distance and alter the coordination pattern of the catalytic ion such that the O3′ atom is too distant to coordinate with the ion (Figure 6). The chemical reaction in this mutant is thus significantly hampered. The active-site geometries in the Arg517Ala and Lys521Ala mutants more closely resemble the wild-type system in that only fluctuations in the position of Ile492 occur.

From our energetic analysis of all wild-type and mutant pol λ/misaligned DNA complexes together with wild-type and Arg517Ala pol λ/aligned DNA complexes [42,], important underlying differences in pol λ's handling of misaligned vs. aligned DNA emerge as well as suggestive trends for relative contributions of specific residues to misaligned DNA binding. These energetic analyses consist of sampling the interaction energy between the thumb (residues at positions 544, 538, 521, 517, and 514), and the adjacent DNA (DNA template-strand nucleotides T5–T8), for all systems (Figure 7). We see that electrostatics are the dominant component of the interaction energy. Van der Waals interactions are smaller and similar in all systems, except for the Arg517Ala/aligned DNA system, which has less favorable interaction energy than the others. Interestingly, the wild-type, Lys521Ala, and Arg538Ala pol λ/misaligned DNA systems have more negative electrostatic interaction energy than the wild-type/aligned DNA complex. In addition, the Arg517Ala/misaligned DNA system has similar average electrostatic interaction energy, but smaller fluctuations (standard deviations) than the wild-type/aligned DNA system. In fact, electrostatic energy fluctuations among the misaligned DNA complexes are similar and smaller than the aligned DNA complexes analyzed; these similar fluctuations suggest that the thumb residues in the misaligned DNA complexes energetically compensate for one another to effectively prevent DNA motion to the inactive position. This is in contrast to the much larger fluctuations occurring in the Arg517Ala/aligned DNA complex, which exhibits 5–7 Å in DNA template-strand motion from the active to inactive DNA position [42, 43, 44,]. Importantly, whole system energy fluctuations in total energy, kinetic energy, potential energy are very similar for all

systems and constitute between 0.2–0.5% of the average energies, reinforcing the significance of the above differences.

Based on the relative magnitudes of the electrostatic component of the interaction energy, the contributions of individual residues to the binding of misaligned DNA can be ranked as follows: Lys544 > Arg517 > Lys521 > Arg538. Mutations in Lys544 can be interpreted as having charge-dependent effects on misaligned DNA binding since negatively-charged glutamate at position 544 has poorer electrostatic interactions than systems with an alanine or phenylalanine residue at this position. The large DNA change in the Lys544Glu system contributes to reducing this electrostatic interaction.

Comparison of the Arg517Ala systems reveals why deletion errors are increased in this mutant. In the aligned DNA system, less favorable energetic interactions between the DNA and the thumb and larger fluctuations compared to the wild-type/aligned DNA system provide a supportive framework for the DNA to slip into the misaligned DNA position, which is energetically much more stable and similar to wild-type/aligned DNA complexes.

Binding free energies lend further evidence for the stronger binding of misaligned compared to aligned DNA. Using the MM-PBSA approach as described in Computational Methods to compute the binding free energies (Table 2), the binding difference $\Delta\Delta G_{bind} = -162.2$ kcal/mol indicates stronger binding of the misaligned DNA compared to the aligned DNA complex. Interestingly, favorable electrostatic interactions (ΔE_{elec}) within the protein/DNA complex more than compensate for the energetic penalty paid by the electrostatics of solvation (ΔG_{PB}) in both complexes. However, the greater ΔE_{elec} in the misaligned DNA complex compared to the aligned DNA complex, as also inferred from our energetic analysis above (Figures 7, 8), appears to be the main source of the misaligned DNA complex's stronger binding, further substantiating our prior results.

3.3 Pol *β* Has Fewer Residues to Stabilize the Misaligned DNA than in Pol λ

Unlike pol λ, where several positively-charged residues are present, only one positively-charged protein residue, Arg299, interacts with the DNA backbone near the extrahelical thymine (T7) in pol *β*. The electrostatic potential surfaces of the thumb subdomains of pol λ and pol *β* reveal striking differences in the misaligned DNA/thumb interaction landscapes (Figure 8). Whereas pol λ has a substantial positively-charged surface for interactions with the DNA near the extrahelical nucleotide, pol *β* has a much smaller region in the vicinity of Arg299, which is adjacent to T7. Thus, differences in electrostatic interactions indicate that pol λ has a much stronger grip on misaligned DNA compared to pol *β*.

Still, the pol *β*/misaligned DNA/dTTP complex appears to be relatively stable with the active site exhibiting a good geometry for the chemical reaction (Figure 6). Some rearrangements, however, occur in the thumb and DNA. The loop in the thumb containing *β*-strand 7 twists to open up the area around the extrahelical nucleotide (T7) (Figure S10a,d in the Supporting Information), and the DNA template strand backbone rearranges near T7 to allow T7's thymine base to stack with Arg299 (Figure 9 and Figure S10b in the Supporting Information). This DNA template-strand rearrangement is accompanied by intermittent shifts in the T5–T8 template strand bases slightly upstream (Figure S10c,e in the Supporting Information). In addition to Arg299, nearby Asn294 and Ala307 sometimes form hydrogen bonds to T7 (Figure S11a–c in the Supporting Information). Together, these changes suggest that pol *β* forms a less stable complex with misaligned DNA than pol λ. The relative stability of pol *β*'s active site may result from Arg283's strong interactions with the DNA template strand bases in the active site (Figure S11d–f in the Supporting Information), which are comparable to its interactions with normally aligned DNA.

4 Discussion

DNA replication on a misaligned template-primer results in either a deletion or insertion error, unless corrected by a proofreading domain or the mismatch repair pathway [3,]. Our atomic-level modeling of pol λ and pol *β* bound to deletion error intermediates reveals distinct differences in how these enzymes bind misaligned DNA containing a two-base repeat, which helps interpret their respective tendencies for deletion error generation. Our simulated pol λ/misaligned DNA structures and dynamics reveal that pol λ binds very tightly to the region of the DNA template-strand between the nascent base pair and the extrahelical nucleotide, located two bases upstream from the nascent base pair. These strong interactions stabilize the extrahelical nucleotide and anchor the DNA in the active position for the chemical reaction; this in turn accounts for pol λ's high catalytic efficiency on misaligned substrates and high propensity for deletion error generation [21, 56,]. The primary residues involved in the binding (Lys544, Arg538, Lys521, and Arg517) are all located in pol λ's thumb subdomain. Arg514, also in the thumb, may provide some secondary support through its stacking interactions with the templating base at the gap. Our examination of the electrostatic potential landscape of the thumb near the extrahelical nucleotide suggests that the positive charge of these residues explains the strong interaction between pol λ and the misaligned DNA. In contrast, pol *β* bound to the same DNA substrate has a much less positively-charged electrostatic potential surface in the vicinity of the extrahelical nucleotide and thus binds the misaligned DNA more weakly. This weaker interaction helps explain its lower tendency for deletion errors, especially within the short repetitive DNA regions that we have modeled.

Our studies of three Ala mutants in place of Arg517, Lys521, Arg538 as well as three Lys544 (Ala, Phe, Glu) pol λ mutant systems reveal that loss of one of these residue's interactions with the DNA is not critical for the stability of the complex, with the single exception of Lys544Glu. Since Lys544 is closest to T7, it is crucial for stabilizing the misaligned DNA. While the Ala and Phe mutants show small DNA changes, the Glu substitution leads to a rotation in the extrahelical nucleotide towards reentering the helix. Interestingly, in the Lys544Glu, Arg517Ala, and Lys421Ala systems, a gap is created within the helix disrupting stacking interactions between the template strand bases flanking T7; this effect further underscores pol λ's tolerance of DNA rearrangement upstream of the active site.

Together, our mutation studies show that alterations in the electrostatic environment around the extrahelical nucleotide can lead to protein and DNA rearrangements that affect the progression of the chemical reaction. Our energetic analysis of thumb/DNA interactions allows us to predict the deletion error tendencies of the various mutants. Generally, systems that interact with the misaligned DNA most strongly would be expected to generate the greatest amounts of deletion errors. This accounts for the higher experimentally determined deletion error rate of the Arg517Ala mutant [43,] and the lower error rate for the Lys544Ala mutant [56,]. The Lys544Phe mutant, which has similar energetic properties, but better active-site geometry than the Lys544Ala mutant, would have a similar or slightly higher deletion error tendency than the Lys544Ala mutant while the Lys544Glu mutant would have a much lower deletion error rate because of its poorer thumb/misaligned DNA interactions. The exception to this trend is Arg538. Although this residue contributes the least energetically to misaligned DNA binding, the Arg538Ala system exhibited motions in both the protein and DNA toward the inactive state, which indicates that an alanine mutation at this position may reduce the efficiency of nucleotide insertion on both aligned and misaligned DNA substrates.

The absence of a DNA shift toward the inactive position in the Arg517Ala pol λ mutant system is surprising given that this same mutation destabilizes the template strand when pol λ is bound to normally aligned DNA, as our prior modeling has shown [42, 43]. Another examination of six different Arg517 pol λ mutant systems bound to properly aligned DNA revealed that both the size and charge of the side chain at position 517 are important for stabilizing properly aligned DNA [44]. Clearly, the stability of the misaligned DNA when bound to the Arg517Ala mutant indicates that compensating interactions are present. Our free energy calculations support the stronger binding of misaligned compared to aligned DNA and suggest that the difference lies in stronger electrostatic interactions within the misaligned DNA complex. These interactions are mediated by Arg514, Lys521, Arg538, and Lys544, which render Arg517 a less critical residue for DNA stabilization. These stronger interactions with misaligned DNA would also explain why crystal structures with pol λ bound to misaligned DNA have been obtained at a higher resolution than those bound to properly aligned DNA [56, 41].

These results highlight pol λ's remarkable robustness, which allows it to be flexible as needed to handle different types of substrates with high efficiency. Together with this versatility, however, comes an inherent drawback. Since pol λ only tracks the base pairing geometry of the nascent and primer terminus base pairs through Arg517/DNA template strand base interactions, it has a restricted capacity to eliminate deletion errors. The normal motion of the DNA during pol λ's catalytic cycle further contributes to its difficulty in monitoring the geometry of the DNA. Its ability to handle unusual substrates, however, allows pol λ to repair double-strand breaks containing jagged DNA ends with minimal base pairing within the NHEJ pathway [21]. Interestingly, a recent study suggests that defects in NHEJ repair are a likely source of deletions and other mutations in breast cancers [83].

Our simulations of the pol *β*/misaligned DNA complex reveals differences from pol λ in the arrangement of the DNA template strand near the extrahelical nucleotide. Pol *β*'s extrahelical T7, located two bases upstream from the nascent base pair, moves to stack with Arg299 in the thumb, causing alterations in both the DNA and the adjacent region of the thumb. The few stabilizing factors for misaligned DNA we identified for pol *β* agree with the experimental observation that misalignments involving the template strand decrease catalytic efficiency [25]. Interestingly, pol *β*'s efficiency can be restored when the extrahelical nucleotide is shifted at least five nucleotides upstream from the active site [25]. Thus, the position of the extrahelical nucleotide plays an important role in pol *β*'s deletion error tendency. Since pol *β* readily makes slippage-mediated deletion errors on five-nucleotide and longer repeat regions [21, 27], the extrahelical nucleotide may be positioned five nucleotides or more upstream of the active site where it is sufficiently stabilized by the surrounding correct base pairs instead of by thumb interactions. Like pol λ, pol *β* is limited to monitoring the geometry of the nascent and primer terminus base pairs primarily through Arg283. However, as compared with pol λ, pol *β*'s catalytic cycle demonstrates a lack of large-scale DNA motion as a result of a more positively-charged template strand binding groove upstream of the active site [9], and this increases pol *β*'s fidelity by preventing large template-strand rearrangements.

When binding repetitive DNA, error rate and structural data [21, 56] suggest that pol λ prefers positioning an extrahelical nucleotide two bases upstream from the active site. The formation of misalignments close to the polymerase active site requires strong protein/DNA interactions with the active-site DNA as well as around the extrahelical nucleotides, a feature we find in pol λ. Pol *β*'s weaker interactions around the extrahelical residue adapt it better for misalignments within longer nucleotide runs that can be alternatively stabilized by multiple correct base pairs surrounding the misalignment. When combined with the different amounts of DNA motion induced upon binding the correct incoming nucleotide, these

factors account for the significantly lower rates of deletions by pol β compared to pol λ. Interestingly, related pol μ easily makes slippage-mediated deletion errors on two and three nucleotide repeats [27, 84,]. Thus, it may have interactions more like pol λ than pol β. Indeed, pol μ conserves pol λ's Arg514 and Arg517 residues with Arg444 and Arg447; our examination of pol μ's structure [85,] also shows another thumb residue, Arg451, to be located in the same vicinity as pol λ's Lys521 and Arg538 residues (Figure S12 in the Supporting Information). Y-family DNA polymerase Dbh, which also makes deletions on repetitive DNA through DNA template-strand slippage, likewise prefers positioning the extrahelical nucleotide three bases upstream from the active site [86,]. As expected, X-ray crystal structures reveal that this position is well-stabilized by the C-terminal domain of Dbh [86,]. These similarities reinforce our hypotheses regarding pol λ's versatile ability to stabilize misaligned DNA and hence generate deletion errors. Of course, these factors are part of a complex picture that involves other long-range interactions within the enzyme and with other biomolecules associated with repair. In fact, other studies have found that residues more distal to the active site can affect deletion error rates for pol β and pol λ [87, 88, 89,].

Interesting connections between pol λ's slippage mechanism and DNA interactions in other base-flipping enzymes can also be drawn. For example, the position of T7 in the pol λ Lys544Glu mutant shows remarkable similarity to a crystal intermediate in the base flipping of thymine by uracil DNA glycosylase (UNG) [90,]. In the UNG structure, the extrahelical thymine is in the *syn* conformation and located in a position that resembles T7 in the Lys544Glu mutant. The paucity of protein/nucleotide interactions in both systems permits rotation of the glycosidic torsion angle from *anti* to *syn*. Interestingly, the *syn* conformation of an extrahelical uracil has also been observed in crystal structures of an archaeal B-family DNA polymerase, which stalls when encountering this base [91,]. MutM, a bacterial DNA glycosylase, is also known to bind its target, 8-oxoguanine, in its active site in the *syn* orientation [92,]. In addition to UNG, extrahelical intermediates on base flipping pathways have been identified for non-target DNA bases in O^6-alkylguanine-DNA alkyltransferase (AGT) from transition path sampling of base flipping [93,] and both 8-oxoguanine DNA glycosylase I (hOGG1) [94,] and MutM [95, 96,] from X-ray crystallographic studies. In these DNA glycosylases and AGT, contacts with *anti vs. syn* bases and transition of the base to an extrahelical intermediate position may aid discrimination of target from non-target substrates for transition to the active site. In pol λ, base recognition is less specific, and pol λ's partially solvent-exposed binding pocket for extrahelical nucleotides allows it to accommodate all four standard bases. Our proposed conserved extrahelical nucleotide binding motif in pol λ in Figure 1d highlights the fundamental nature of these interactions and motions in a cell's functioning.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Research described in this article was supported in part by Philip Morris USA Inc. and Philip Morris International and by NSF grant MCB-0316771, NIH grant R01 ES012692, and the American Chemical Society's Petroleum Research Fund award (PRF #39115-AC4) to T. Schlick. The computations are made possible by support for the SGI Altix 3700 by the National Center for Supercomputing Applications (NCSA) under grant MCA99S021, and by the NYU Chemistry Department resources under grant CHE-0420870. M. F. acknowledges funding from NYU's Kramer fellowship for the 2008-2009 academic year. Molecular images were generated using the INSIGHTII package (Accelrys Inc., San Diego, CA),VMD [67,] , and the PyMOL Molecular Graphics System (DeLano Scientific LLC, San Carlos, CA).

References

1. Mirkin SM. Nature. 2007; 447:932–940. [PubMed: 17581576]
2. Bacolla A, Wells RD. Mol Carcinog. 2009; 48:273–285. [PubMed: 19306308]
3. Garcia-Diaz M, Kunkel TA. Trends Biochem Sci. 2006; 31:206–214. [PubMed: 16545956]
4. Nick McElhinny SA, Havener JM, Garcia-Diaz M, Juarez R, Bebenek K, Kee BL, Blanco L, Kunkel TA, Ramsden DA. Mol Cell. 2005; 19:357–366. [PubMed: 16061182]
5. Broyde S, Wang L, Rechkoblit O, Geacintov NE, Patel DJ. Trends Biochem Sci. 2008; 33:209–219. [PubMed: 18407502]
6. Steitz TA. J Biol Chem. 1999; 274:17395–17398. [PubMed: 10364165]
7. Matsumoto Y, Kim K. Science. 1995; 269:699–702. [PubMed: 7624801]
8. Garcia-Diaz M, Bebenek K, Kunkel TA, Blanco L. J Biol Chem. 2001; 276:34659–34663. [PubMed: 11457865]
9. Garcia-Diaz M, Bebenek K, Krahn JM, Blanco L, Kunkel TA, Pedersen LC. Mol Cell. 2004; 13:561–572. [PubMed: 14992725]
10. Moon AF, Garcia-Diaz M, Batra VK, Beard WA, Bebenek K, Kunkel TA, Wilson SH, Pedersen LC. DNA Repair. 2007; 6:1709–1725. [PubMed: 17631059]
11. Kunkel TA. J Biol Chem. 2004; 279:16895–16898. [PubMed: 14988392]
12. Braithwaite EK, Prasad R, Shock DD, Hou EW, Beard WA, Wilson SH. J Biol Chem. 2005; 280:18469–18475. [PubMed: 15749700]
13. Braithwaite EK, Kedar PS, Lan L, Polosina YY, Asagoshi K, Poltoratsky VP, Horton JK, Miller H, Teebor GW, Yasui A, Wilson SH. J Biol Chem. 2005; 280:31641–31647. [PubMed: 16002405]
14. Lebedeva NA, Rechkunova NI, Dezhurov SV, Khodyreva SN, Favre A, Blanco L, Lavrik OI. Biochim Biophys Acta. 2005; 1751:150–158. [PubMed: 15979954]
15. Tano K, Nakamura J, Asagoshi K, Arakawa H, Sonoda E, Braithwaite EK, Prasad R, Buerstedde JM, Takeda S, Watanabe M, Wilson SH. DNA Repair. 2007; 6:869–875. [PubMed: 17363341]
16. Lee JW, Blanco L, Zhou T, Garcia-Diaz M, Bebenek K, Kunkel TA, Wang Z, Povirk LF. J Biol Chem. 2004; 279:805–811. [PubMed: 14561766]
17. Fan W, Wu X. Biochem Biophys Res Commun. 2004; 323:1328–1333. [PubMed: 15451442]
18. Ma Y, Lu H, Tippin B, Goodman MF, Shimazaki N, Koiwai O, Hsieh CL, Schwarz K, Lieber MR. Mol Cell. 2004; 16:701–713. [PubMed: 15574326]
19. Capp JP, Boudsocq F, Bertrand P, Laroche-Clary A, Pourquier P, Lopez BS, Cazaux C, Hoffmann JS, Canitrot Y. Nucleic Acids Res. 2006; 34:2998–3007. [PubMed: 16738138]
20. Chan K, Houlbrook S, Zhang QM, Harrison M, Hickson ID, Dianov GL. Mutage-nesis. 2007; 22:183–188.
21. Bebenek K, Garcia-Diaz M, Blanco L, Kunkel TA. J Biol Chem. 2003; 278:34685–34690. [PubMed: 12829698]
22. Kunkel TA. J Biol Chem. 1985; 260:5787–5796. [PubMed: 3988773]
23. Streisinger G, Okada Y, Emrich J, Newton J, Tsugita A, Terzaghi E, Inouye M. Cold Spring Harb Symp Quant Biol. 1966; 31:77–84. [PubMed: 5237214]
24. Eckert KA, Mowery A, Hile SE. Biochem. 2002; 41:10490–10498. [PubMed: 12173936]
25. Beard WA, Shock DD, Wilson SH. J Biol Chem. 2004; 279:31921–31929. [PubMed: 15145936]
26. Osheroff WP, Beard WA, Yin S, Wilson SH, Kunkel TA. J Biol Chem. 2000; 275:28033–28038. [PubMed: 10851238]
27. Tippin B, Kobayashi S, Bertram JG, Goodman MF. J Biol Chem. 2004; 279:45360–45368. [PubMed: 15339923]
28. Zang H, Goodenough AK, Choi JY, Irimia A, Loukachevitch LV, Kozekov ID, Angel KC, Rizzo CJ, Egli M, Guengerich FP. J Biol Chem. 2005; 280:29750–29764. [PubMed: 15965231]
29. Zang H, Irimia A, Choi JY, Angel KC, Loukachevitch LV, Egli M, Guengerich FP. J Biol Chem. 2006; 281:2358–2372. [PubMed: 16306039]
30. Eoff RL, Stafford JB, Szekely J, Rizzo CJ, Egli M, Guengerich FP. Biochemistry. 2009; 48:7079–7088. [PubMed: 19492857]

31. Ling H, Boudsocq F, Woodgate R, Yang W. Cell. 2001; 107:91–102. [PubMed: 11595188]
32. Kokoska RJ, Bebenek K, Boudsocq F, Woodgate R, Kunkel TA. J Biol Chem. 2002; 277:19633–19638. [PubMed: 11919199]
33. Efrati E, Tocco G, Eritja R, Wilson SH, Goodman MF. J Biol Chem. 1997; 272:2559–2569. [PubMed: 8999973]
34. Blanca G, Villani G, Shevelev I, Ramadan K, Spadari S, Hubscher U, Maga G. Biochem. 2004; 43:11605–11615. [PubMed: 15350147]
35. Kovtun IV, Liu Y, Bjoras M, Klungland A, Wilson SH, McMurray CT. Nature. 2007; 447:447–452. [PubMed: 17450122]
36. Liu Y, Prasad R, Beard WA, Hou EW, Horton JK, McMurray CT. J Biol Chem. 2009; 284:28352–28366. [PubMed: 19674974]
37. Hartenstine MJ, Goodman MF, Petruska J. J Biol Chem. 2002; 277:41379–41389. [PubMed: 12196536]
38. Duvauchelle JB, Blanco L, Fuchs RPP, Cordonnier AM. Nuc Acids Res. 2002; 30:2061–2067.
39. Zhang Y, Wu X, Guo D, Rechkoblit O, Taylor JS, Geacintov NE, Wang Z. J Biol Chem. 2002; 15:44582–44587. [PubMed: 12228225]
40. Wells RD. Trends Biochem Sci. 2007; 32:271–278. [PubMed: 17493823]
41. Garcia-Diaz M, Bebenek K, Krahn JM, Kunkel TA, Pedersen LC. Nat Struct Mol Biol. 2005; 12:97–98. [PubMed: 15608652]
42. Foley MC, Arora K, Schlick T. Biophys J. 2006; 91:3182–3195. [PubMed: 16920835]
43. Bebenek K, Garcia-Diaz M, Foley MC, Pedersen LC, Schlick T, Kunkel TA. EMBO Reports. 2008; 9:459–464. [PubMed: 18369368]
44. Foley MC, Schlick T. J Am Chem Soc. 2008; 130:3967–3977. [PubMed: 18307346]
45. Ahn J, Kraynov VS, Zhong X, Werneburg BG, Tsai MD. Biochem J. 1998; 331:79–87. [PubMed: 9512464]
46. Zhong X, Patel SS, Werneburg BG, Tsai MD. Biochemistry. 1997; 36:11891–11900. [PubMed: 9305982]
47. Beard WA, Wilson SH. Chem Rev. 2006; 106:361–382. [PubMed: 16464010]
48. Yang L, Beard WA, Wilson SH, Broyde S, Schlick T. J Mol Biol. 2002; 317:651–671. [PubMed: 11955015]
49. Arora K, Schlick T. Biophys J. 2004; 87:3088–3099. [PubMed: 15507687]
50. Arora K, Schlick T. J Phys Chem B. 2005; 109:5358–5367. [PubMed: 16863202]
51. Radhakrishnan R, Schlick T. Proc Natl Acad Sci USA. 2004; 101:5970–5975. [PubMed: 15069184]
52. Fowler JD, Brown JA, Kvaratskhelia M, Suo Z. J Mol Biol. 2009; 390:368–379. [PubMed: 19467241]
53. Radhakrishnan R, Arora K, Wang Y, Beard WA, Wilson SH, Schlick T. Biochemistry. 2006; 45:15142–15156. [PubMed: 17176036]
54. Arora K, Beard WA, Wilson SH, Schlick T. Biochemistry. 2005; 44:13328–13341. [PubMed: 16201758]
55. Foley MC, Schlick T. J Phys Chem B. 2009; 113:13035–13047. [PubMed: 19572669]
56. Garcia-Diaz M, Bebenek K, Krahn JM, Pedersen LC, Kunkel TA. Cell. 2006; 124:331–342. [PubMed: 16439207]
57. Lee EH, Hsin J, Sotomayor M, Comellas G, Schulten K. Structure. 2009; 17:1295–1306. [PubMed: 19836330]
58. Schlick, T. Molecular Modeling and Simulation: An Interdisciplinary Guide. Springer-Verlag; New York: 2002.
59. MacKerell A Jr, Banavali NK. J Comput Chem. 2000; 21:105–120.
60. MacKerell AD Jr, et al. J Phys Chem B. 1998; 102:3586–3616.
61. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. J Comput Chem. 1983; 4:187–217.
62. Stote RH, Karplus M. Proteins. 1995; 23:12–31. [PubMed: 8539245]

63. MacKerell AD Jr. J Phys Chem B. 1997; 101:646–650.
64. Yang L, Arora K, Beard WA, Wilson SH, Schlick T. J Amer Chem Soc. 2004; 126:8441–8453. [PubMed: 15238001]
65. Steitz TA. Nature. 1998; 391:231–232. [PubMed: 9440683]
66. Qian X, Strahs D, Schlick T. J Comput Chem. 2001; 22:1843–1850. [PubMed: 12116415]
67. Humphrey W, Dalke A, Schulten K. J Mol Graph. 1996; 14:33–38. [PubMed: 8744570]
68. Klapper I, Hagstrom R, Fine R, Sharp K, Honig B. Proteins. 1986; 1:47–59. [PubMed: 3449851]
69. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. J Comput Chem. 2005; 26:1781–1802. [PubMed: 16222654]
70. Feller SE, Zhang Y, Pastor RW, Brooks BR. J Chem Phys. 1995; 103:4613–4621.
71. Ryckaert JP, Ciccotti G, Berendsen HJC. J Comput Phys. 1977; 23:327–341.
72. Darden TA, York DM, Pedersen LG. J Chem Phys. 1993; 98:10089–10092.
73. Sharp KA, Honig B, Harvey SC. Biochemistry. 1990; 29:340–346. [PubMed: 2405900]
74. Chin K, Sharp KA, Honig B, Pyle AM. Nature Struct Biol. 1999; 6:1055–1061. [PubMed: 10542099]
75. Chong LT, Duan Y, Wang L, Massova I, Kollman PA. Proc Natl Acad Sci USA. 1999; 96:14330–14335. [PubMed: 10588705]
76. Massova I, Kollman PA. J Am Chem Soc. 1999; 121:8133–8143.
77. Gorfe AA, Jelesarov I. Biochemistry. 2003; 42:11568–11576. [PubMed: 14529266]
78. Adcock ST, McCammon JA. Chem Rev. 2006; 106:1589–1615. [PubMed: 16683746]
79. Im W, Beglov D, Roux B. Comput Phys Comm. 1998; 111:59–75.
80. Janezic D, Venerable RM, Brooks BR. J Comput Chem. 1995; 16:1554–1566.
81. Andricioaei I, Karplus M. J Chem Phys. 2001; 115:6289–6292.
82. Estabrook RA, Lipson R, Hopkins B, Reich N. J Biol Chem. 2004; 279:31419–31428. [PubMed: 15143064]
83. Stephens PJ, et al. Nature. 2009; 462:1005–1010. [PubMed: 20033038]
84. Zhang Y, Wu X, Yuan F, Xie Z, Wang Z. Mol Cell Biol. 2001; 21:7995–8006. [PubMed: 11689691]
85. Moon AF, Garcia-Diaz M, Bebenek K, Davis BJ, Zhong X, Ramsden DA, Kunkel TA, Pedersen LC. Nat Struct Mol Biol. 2007; 14:45–53. [PubMed: 17159995]
86. Wilson RC, Pata JD. Mol Cell. 2008; 29:767–779. [PubMed: 18374650]
87. Clairmont CA, Narayanan L, Sun KW, Glazer PM, Sweasy JB. Proc Natl Acad Sci USA. 1999; 96:9580–9585. [PubMed: 10449735]
88. Li SX, Vaccaro JA, Sweasy JB. Biochemistry. 1999; 38:4800–4808. [PubMed: 10200168]
89. Garcia-Diaz M, Bebenek K, Larrea AA, Havener JM, Perera L, Krahn JM, Pedersen LC, Ramsden DA, Kunkel TA. Nat Struct Mol Biol. 2009; 16:967–972. [PubMed: 19701199]
90. Parker JB, Bianchet MA, Krosky DJ, Friedman JI, Mario Amzel L, Stivers JT. Nature. 2007; 449:433–437. [PubMed: 17704764]
91. Firbank SJ, Wardle J, Heslop P, Lewis RJ, Connolly BA. J Mol Biol. 2008; 381:529–539. [PubMed: 18614176]
92. Fromme JC, Verdine GL. J Biol Chem. 2003; 278:51543–51548. [PubMed: 14525999]
93. Hu J, Ma A, Dinner AR. Proc Natl Acad Sci USA. 2008; 105:4615–4620. [PubMed: 18353991]
94. Banarjee A, Yang W, Karplus M, Verdine GL. Nature. 2005; 434:612–618. [PubMed: 15800616]
95. Qi Y, Spong MC, Nam K, Karplus M, Verdine GL. J Biol Chem. 2009; 285:1468–1478. [PubMed: 19889642]
96. Qi Y, Spong MC, Nam K, Banerjee A, Jiralerspong S, Karplus M, Verdine GL. Nature. 2009; 462:762–766. [PubMed: 20010681]

List of Abbreviations

pol λ	DNA polymerase λ
pol *β*	DNA polymerase *β*
dNTP	2′-deoxyribonucleoside 5′-triphosphate
T5	templating base at the gap
T6	adjacent template base that pairs with the primer terminus
T7	extrahelical template base
P6	primer terminus

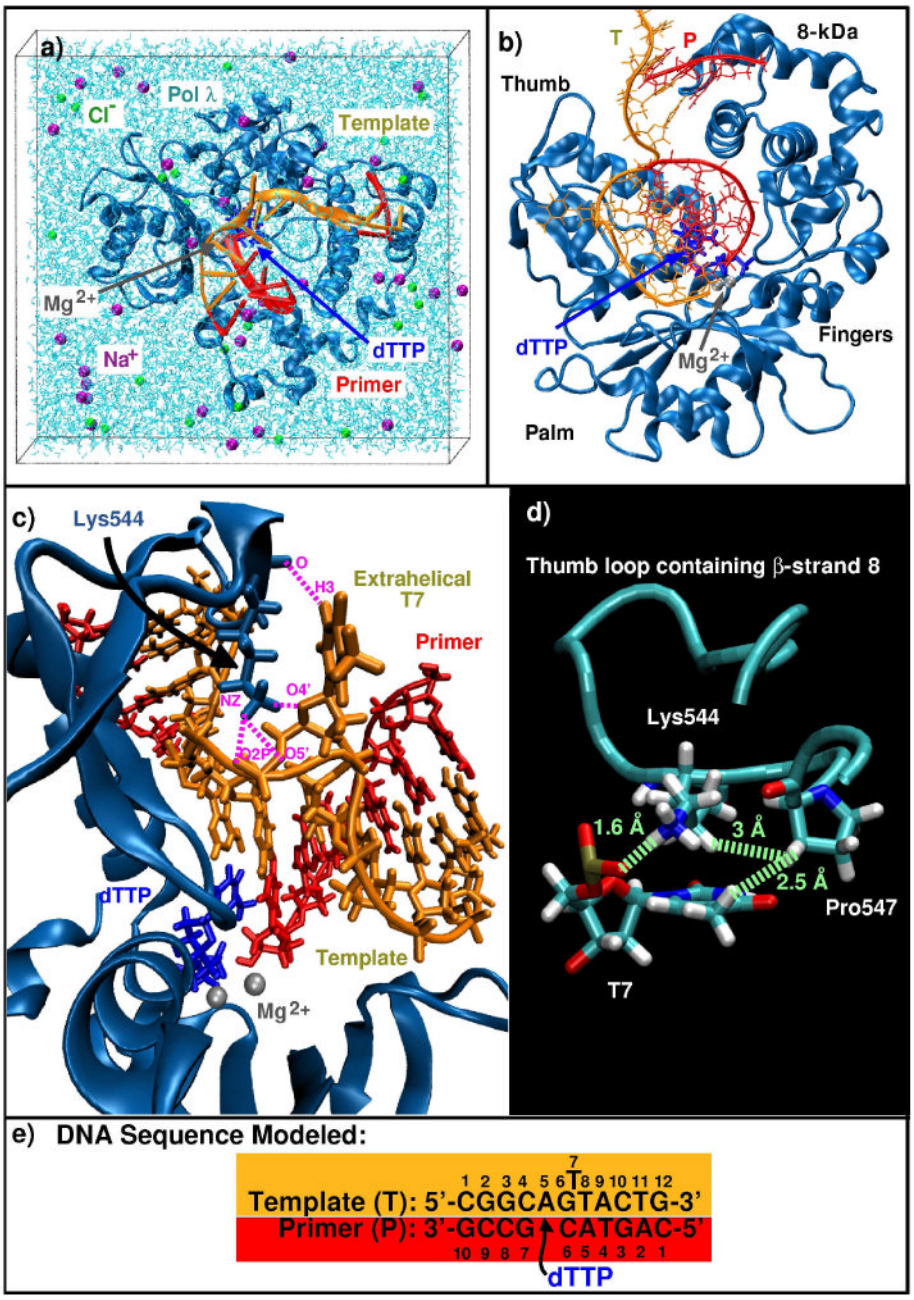

Figure 1.

The pol λ/misaligned DNA/dTTP ternary system poised for a deletion error. (a) Solvated and (b) unsolvated views of the complex. (c) stereo image of DNA pairing in the active site and Lys544/DNA interactions. T: DNA template strand (red); P: DNA primer strand (orange). (d) Pol λ residues directly interacting with the extrahelical nucleotide. The positions and interactions of these residues resemble a conserved motif found in base-flipping enzymes [82,] . See text for more details. (e) DNA sequence modeled in all systems. Numbers shown are used for referencing the DNA bases in the text.

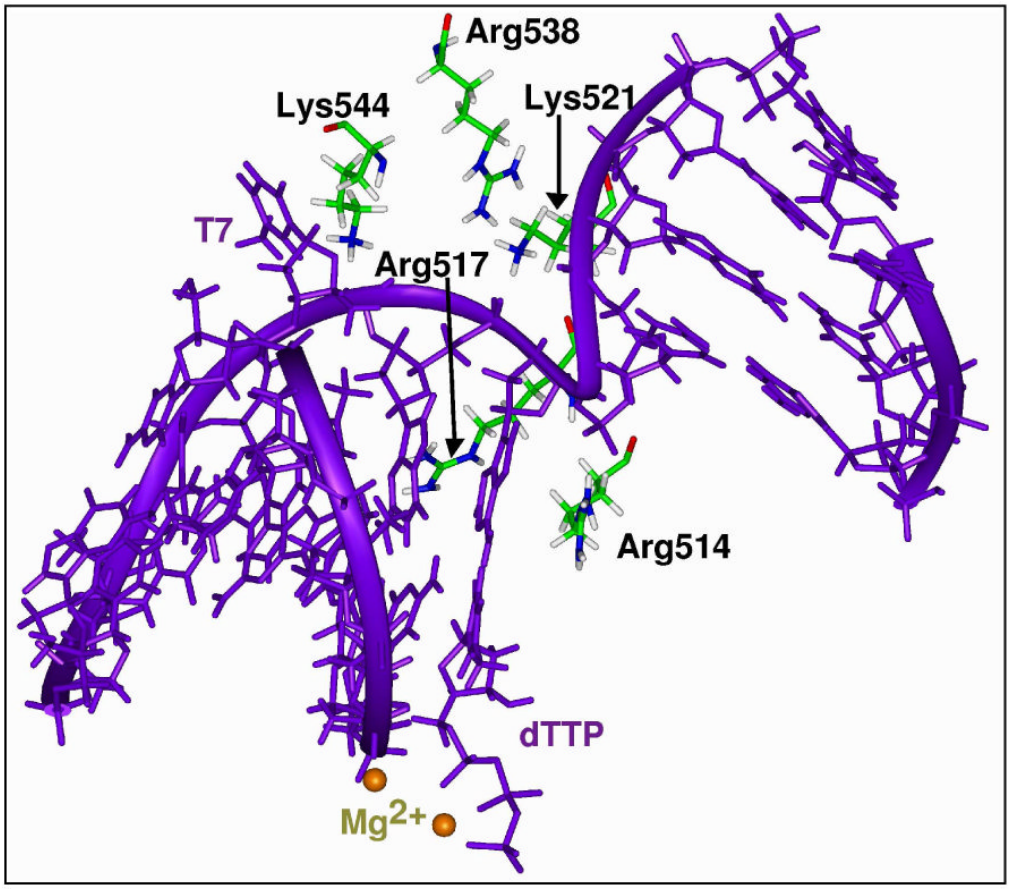

Figure 2.
Positively-charged residues in pol λ's thumb (Lys544, Lys521, Arg538, Arg517, and Arg514) that together allow tight binding of misaligned DNA.

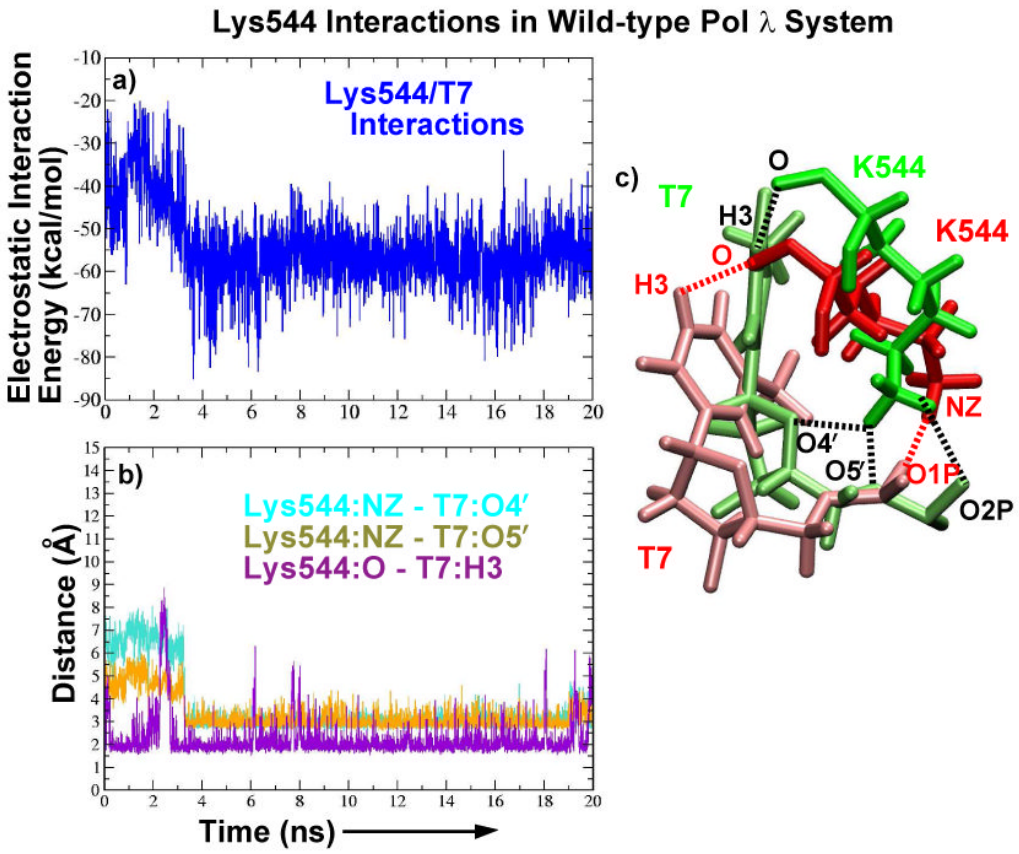

Figure 3.

Lys544/DNA interactions in pol λ. (a) Electrostatic interaction energy changes between Lys544 and T7. (b) Time-evolution of distances between Lys544 and T7:O4′, O5′, and H3 atoms. (c) Change in Lys544's side chain from the experimentally-derived orientation (red) to a new orientation (green), obtained during the dynamics simulation, that facilitates formation of two additional hydrogen bonds to T7.

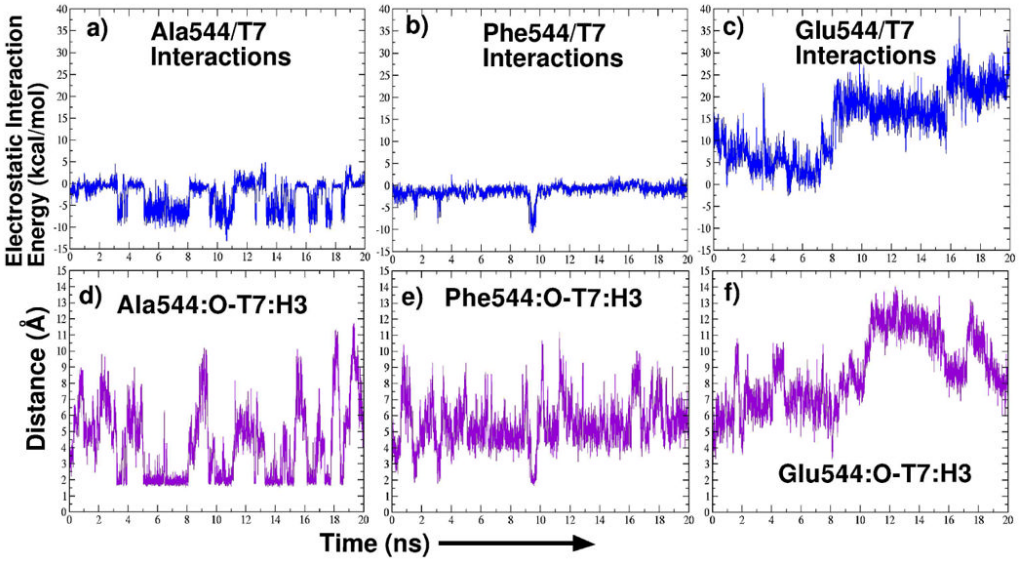

Figure 4.

Time evolution of electrostatic interaction energy changes between residue 544 and T7 (a–c) and corresponding distance changes (d–f) in Lys544 pol λ mutant systems.

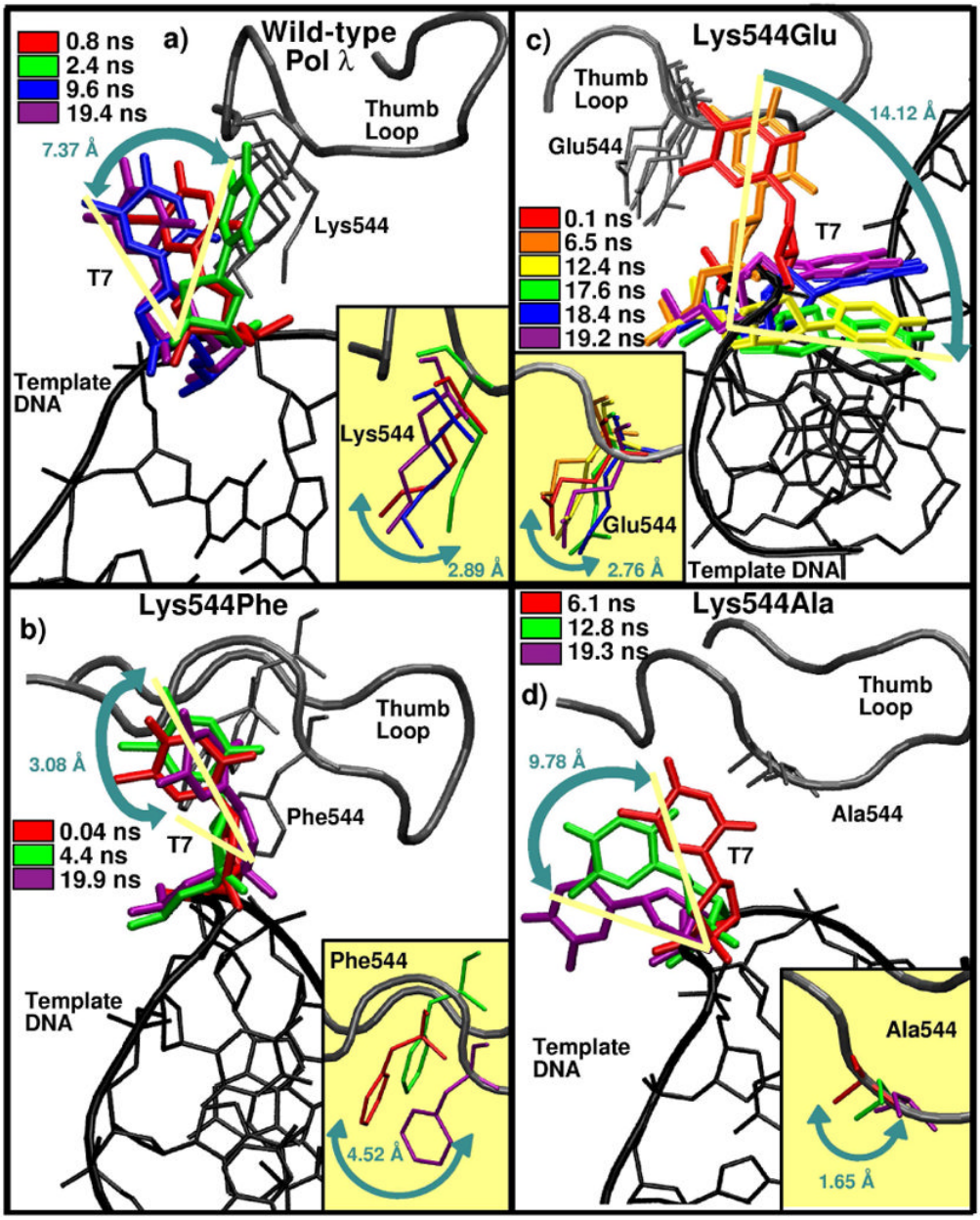

Figure 5.

Ranges of motion of extrahelical thymine and residues at position 544 in pol λ wild-type and Lys544 mutant (i.e., Glu, Phe, and Ala) systems, the latter in the corner insets.

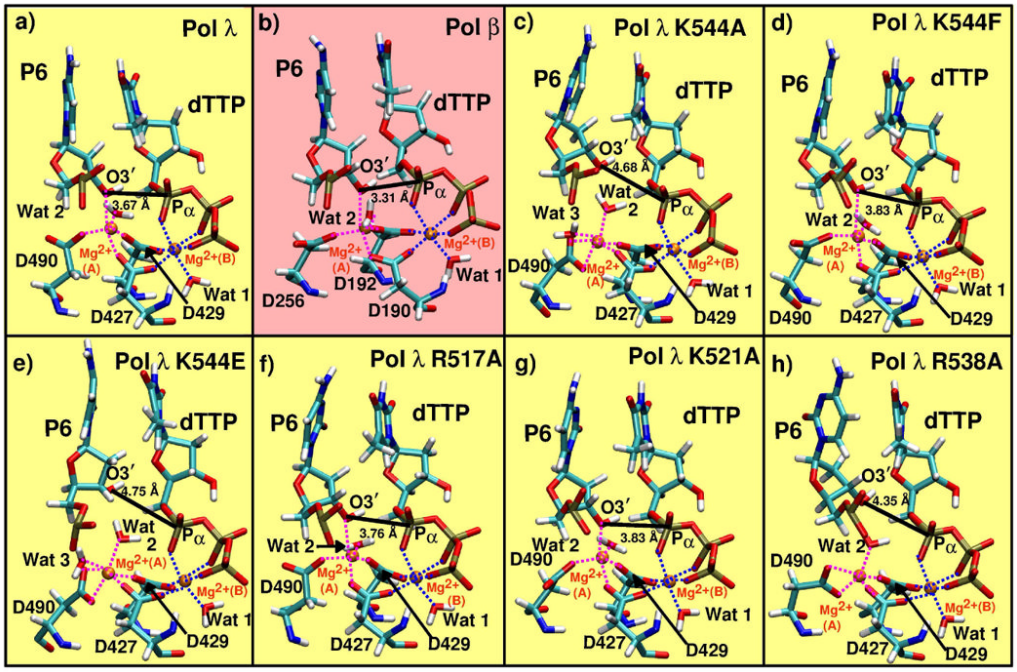

Figure 6.

Active-site geometries of pol λ, pol *β*, and all pol λ mutant systems as seen in the final snapshot of each system trajectory. Abbreviations used: Wat, a water molecule; Mg^{2+} (A), catalytic ion; Mg^{2+} (B), nucleotide-binding ion.

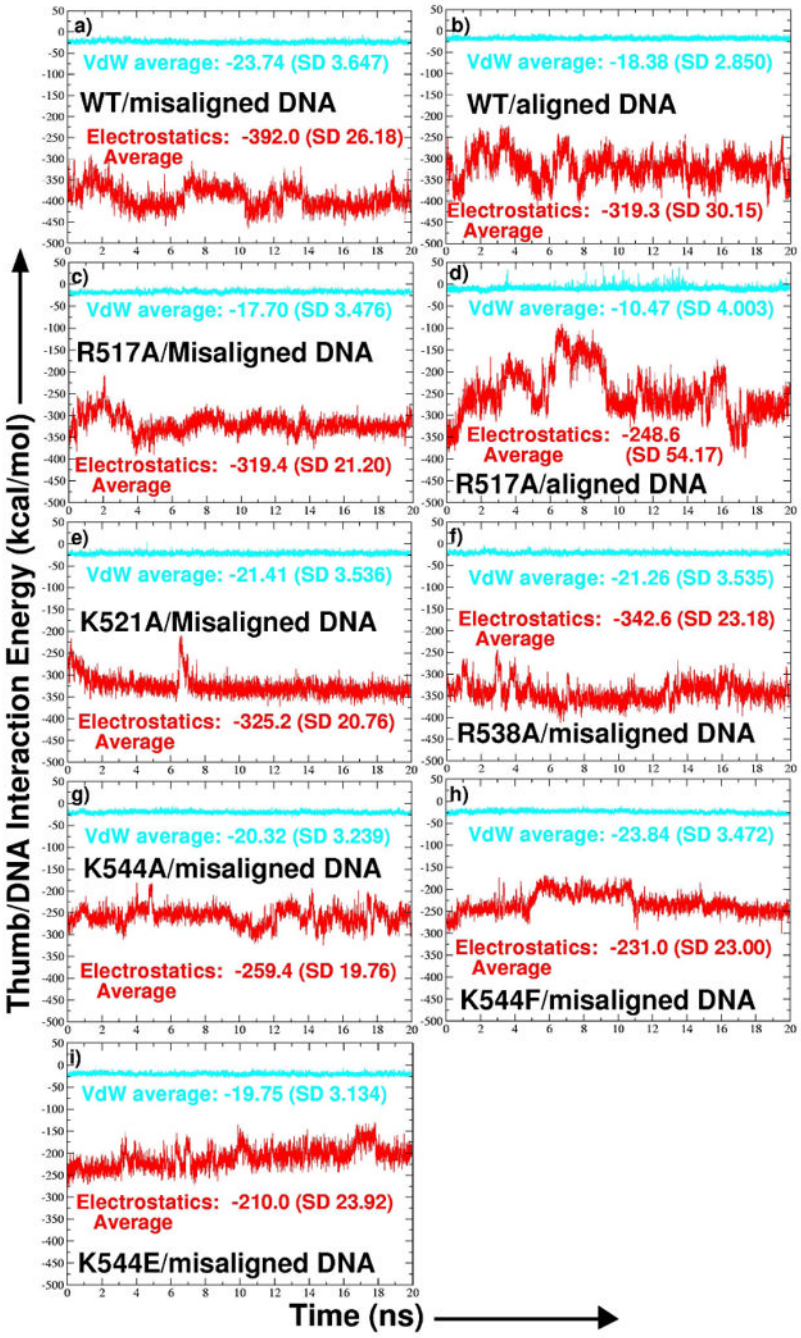

Figure 7.

Time evolution of thumb (residues 544, 538, 521, 517, and 514)/DNA (T5–T8) interaction energy divided into van der Waals (VdW) and electrostatics energy components. Energy averages and standard deviations (SD) are provided.

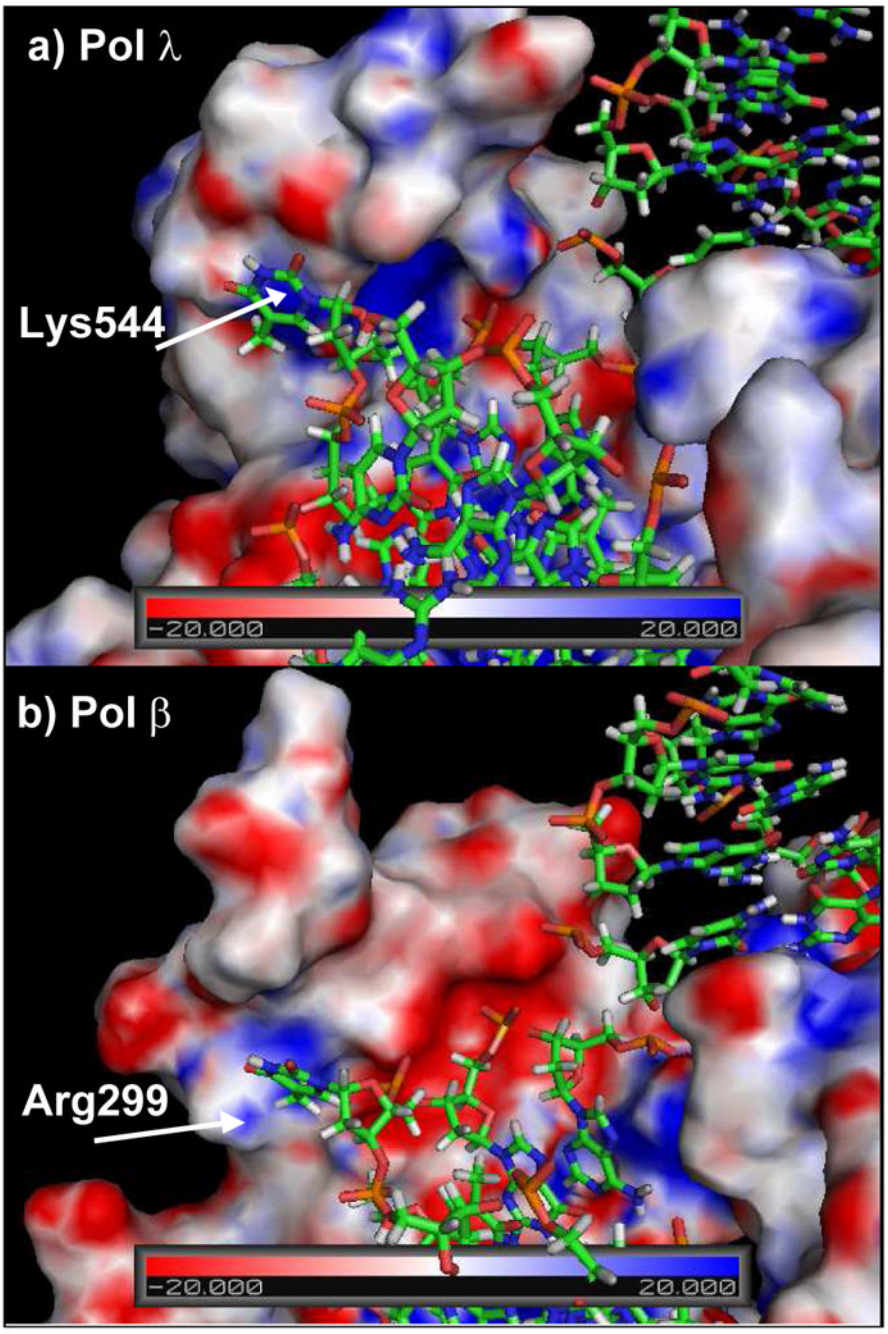

Figure 8.
Electrostatic potential landscapes of thumb/DNA interactions in (a) pol λ and (b) pol *β* misaligned DNA systems following equilibration.

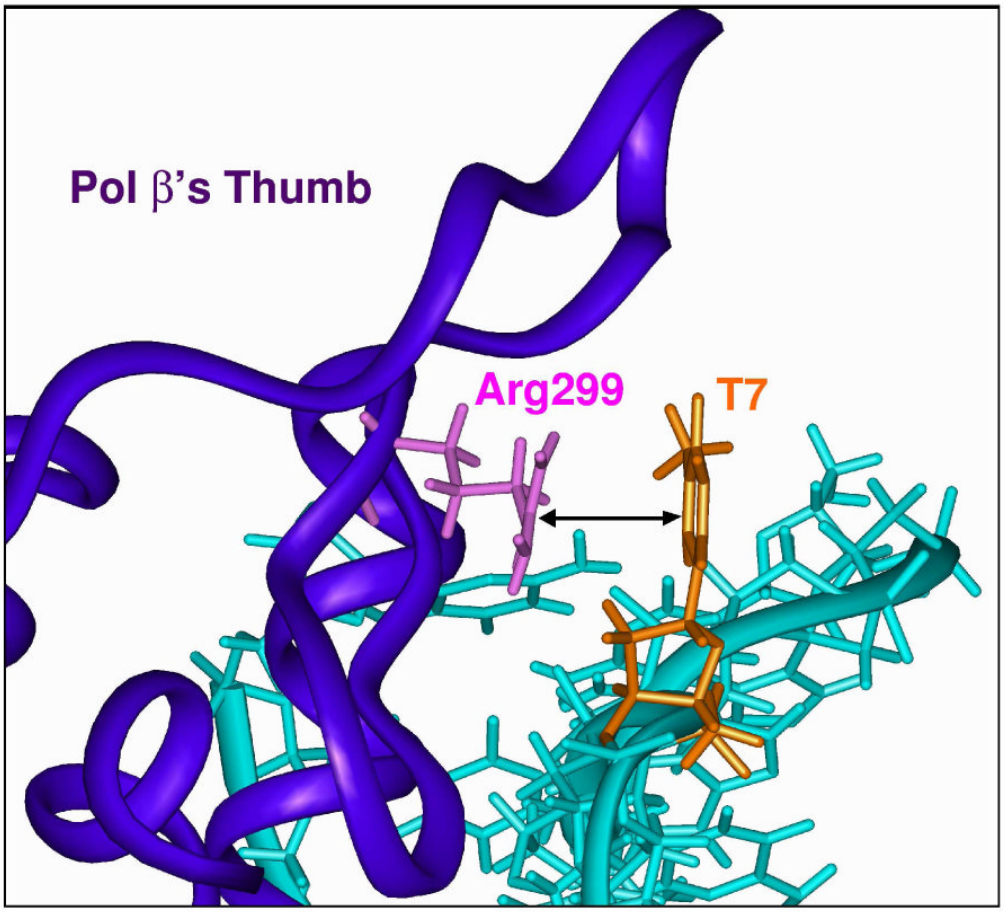

Figure 9.
Pol *β* interactions with the misaligned DNA are mediated by stacking interactions between thumb-residue Arg299 and T7.

Table 1
Summary of Ternary Polymerase/Misaligned DNA/dTTP Complexes Analyzed

Polymerase System	Length of Simulation	Initial DNA Position[a]	DNA Change
WT[b] Pol λ	20 ns	active	no major motion, but small fluctuations in T7
K544A Pol λ	20 ns	active	no major motion, but larger T7 fluctuations than WT Pol λ
K544F Pol λ	20 ns	active	small kink in DNA backbone around T7
K544E Pol λ	20 ns	active	T7 rotates to intrahelical position, loss of stacking interactions between T6 and T8
R517A Pol λ	20 ns	active	loss of stacking interactions between T6 and T8
K521A Pol λ	20 ns	active	loss of stacking interactions between T6 and T8
R538A Pol λ	20 ns	active	small fluctuations toward inactive DNA position[c]
WT Pol β	10 ns	active	small rotation in DNA backbone near T7

[a] For pol λ, active DNA position corresponds to that in PDB entry 1XSN For pol β, active DNA position corresponds to that in PDB entry 1BPY

[b] WT, wild-type

[c] For pol λ, inactive DNA position corresponds to that in PDB entry 1XSL

Table 2

Energetic Analysis of the Formation of Pol λ Aligned DNA and Misaligned DNA Complexes

	Misaligned DNA	Aligned DNA
ΔE_{elec}	-705.4 (35.1)	-599.3 (36.1)
ΔE_{vdw}	-163.3 (7.1)	-146.7 (6.3)
ΔE_{MM}	-868.7 (34.5)	-746.0 (36.3)
ΔG_{np}	-22.8 (0.4)	-20.2 (0.5)
ΔG_{PB}	13.5 (23.7)	61.0 (36.5)
ΔG_{Solv}	-9.3 (23.8)	40.8 (36.5)
$\Delta G_{elec,tot}$	-691.9 (42.9)	-538.3 (48.9)
ΔG_{tot}	-878.0 (42.6)	-705.2 (49.3)
$-T\Delta S$	91.2	80.6
ΔG_{Bind}	-786.8	-624.6

All results are in kcal/mol. Standard deviations are the values in parentheses. Definitions include: ΔE_{elec} = electrostatic MM energy, ΔE_{vdw} = van der Waals MM energy, ΔE_{MM} = total MM energy ($\Delta E_{elec} + \Delta E_{vdw} + \Delta E_{int}$), ΔG_{np} = nonpolar contribution to the solvation energy, ΔG_{PB} = electrostatic contribution to the solvation energy, ΔG_{Solv} = total solvation energy ($\Delta G_{PB} + \Delta G_{np}$), $\Delta G_{elec,tot}$ = total electrostatic energy ($\Delta E_{elec} + \Delta G_{PB}$), ΔG_{tot} = total energy without solute entropic contribution ($\Delta E_{MM} + \Delta G_{Solv}$), $-T\Delta S$ = solute entropic contribution (T = 300 K), ΔG_{bind} = total energy with solute entropic contribution ($\Delta G_{tot} - T\Delta S$).