Supplementary Information

Extended Description of MD Simulation Results

DNA Movement in R517K and R517A Pol lambda Mutant Simulations

Although the protein remains in the ternary conformation, the DNA moves significantly toward the binary position during each mutant simulation. This DNA movement mainly involves the template strand, while the primer remains close to its initial position. Figure 3 shows the mutant-dependent range of DNA motion relative to the crystal binary and ternary positions as well as the motion observed during the wild-type simulation of the ternary state before chemistry. The corresponding RMSD plots of simulated DNA backbone atoms relative to those same atoms of the crystal binary and ternary complexes are shown in Supplementary Figure 2. From these figures, we see that the wild-type displays a small range of DNA motion and stays close to the crystal ternary position. In contrast, the Ala pol λ simulation captures much larger DNA motion, with the DNA moving over a 2-4 Å range toward the binary position. The Lys mutant simulations capture more intermediate DNA template strand movement. This trend is also evident from Supplementary Figure 2, where the RMSD for the ternary complex (red) steadily increases while the value for the binary complex (green) decreases. Using the ranges of motion in Figure 3, we can arrange the systems in increasing order of DNA motion as: WT < Lys (two ions bound) < Lys (one ion bound) < Ala.

The Lys mutant simulations reveal different ranges of DNA motion depending on the system simulated. In the Lys simulation with both ions, the DNA movement is similar to that in the wild-type enzyme. Small fluctuations toward the binary position involve only the DNA template strand backbone (Fig. 3 and Sup. Fig. 2). In the Lys simulation without the catalytic ion, more substantial DNA movement occurs and the DNA transits between the ternary and binary positions

frequently during the initial 13 ns of the simulation (Sup. Fig. 2). During this simulation, the DNA template strand backbone motion is centered between the binary and ternary crystal states as shown in Figure 3. The largest fluctuations of the template strand occur around the template base at the gap, T5, and the two following bases, T6 and T7.

Lys517's Side-chain Flexibility and Interactions with the DNA

In our prior work we showed that Arg517 can form multiple hydrogen bonds with the DNA without altering its side-chain conformation in the ternary state[1]. As observed from our Lys mutant simulations, Lys517 has more complex interactions with the DNA and can assume three different orientations; each of which produces a somewhat different DNA interaction. This side-chain flexibility results in increased DNA movement compared to wild-type pol λ.

Supplementary Figure 3 shows the three orientations of Lys517 and the single orientation of Ala517 captured during the mutant simulations. Lys517's orientations are distinguished by its side-chain dihedral angle (i.e., CG-CD-CE-NZ) ranges: 60°, 200° and 300° (Sup. Fig. 3); we term these Lys517 conformers as "extremely away", "toward", and "away" from the DNA, respectively, to indicate relative proximity to the DNA template bases T5 and T6. The 300° range or "away" conformation is similar to Lys517 in the ternary X-ray crystal structure of the mutant.

Lys517's conformation clearly affects its interactions with the DNA. As shown in Supplementary Figure 4, the "toward" orientation correlates with Lys517 hydrogen bonding with T5 and T6, while hydrogen bonding with T6:O4' can occur with all conformations of Lys517. Less stable interactions through a water molecule also form with DNA backbone atom T6:O1P in both the Lys

and Ala mutant simulations. Lys517 forms a few other transient interactions through water molecules with T5 and T6 as summarized in Supplementary Table 2. Interaction energies between the different conformers of Lys517 and T5 and T6 are described in[2], which reports additional 517 mutant simulations and is focused on identifying specific protein/DNA interactions that affect DNA template strand stability.

Active-Site Residue Rearrangements Accompanying the DNA Motion

The DNA motion in the mutant simulations destabilizes several active-site residues (i.e., Ile492, Tyr505, and Phe506) from their ternary state positions. Tyr505 moves to its binary position in all the mutant simulations, while Phe506 only flips to its binary position during the Lys simulation without the catalytic ion. Ile492 moves more variably in the mutants. In the Ala mutant[1] and Lys mutant simulation with both ions, Ile492 flips to its binary position, but this is only a brief rearrangement in the Lys simulation with both ions. In the Lys system without the catalytic ion, Ile492 flips to its binary position during the equilibration phase, but, then, moves again during production dynamics to a position opposite to its initial ternary state position. These residue changes in the Lys mutant agree with the intermediate conformations observed in the X-ray crystal structure.

The catalytic aspartates (i.e., Asp427, Asp429, and Asp490) are less affected by the DNA motion and are able to coordinate with the active-site Mg^{2+} ions. As we observed in our prior wild-type simulations[1], the nucleotide-binding ion is coordinated to one water molecule, Asp427, Asp429, and dTTP. The catalytic ion, present in the Lys simulation with both ions, is coordinated to two water molecules and all three aspartate residues.

Supplementary Methods

Molecular Dynamics (MD) Simulation

Initial Models. Two initial models were prepared based on the pol λ ternary complex with the correct incoming nucleotide bound (PDB entry 1XSN). In both models, Arg517 is mutated to lysine. The catalytic ion was added to the active site in one model from another pol λ ternary complex (PDB entry 2BCV) that contains a Na^+ in the catalytic ion position. The ion was then changed to Mg^{2+}. All other missing atoms and residues were added. The protein/DNA/dTTP complex was solvated in a cubic cell and a neutral system at an ionic strength of 150 mM was obtained by replacing water oxygen atoms with minimal electrostatic potential with Na^+ and those with maximal electrostatic potential with Cl^-. The Na^+ and Cl^- ions were placed at least 8 Å away from each other and both protein and DNA atoms. Both initial models contain around 38,323 atoms, including waters, Mg^{2+} ions, dTTP, and 41 Na^+ and 29 Cl^- counterions. The final dimensions of the box are: 72.61 Å x 75.97 Å x 70.91 Å.

Minimization, Equilibration and Dynamics Protocol. Two simulations were performed. System 1 is the ternary Arg517Lys system with the nucleotide-binding ion only, whereas system 2 has both magnesium ions. Initial energy minimizations and equilibration simulations were performed using the CHARMM program[3] with the all-atom C26a2 force field[4]. First, the positions of added residues were optimized, while keeping all other heavy atoms fixed, using SD for 5,000 steps followed by ABNR for 10,000 steps. This minimization cycle was repeated twice using 10,000 steps of SD followed by 20,000 steps of ABNR. During minimizations, Cl^-, Na^+, and water molecules were allowed to relax around the protein/DNA complex. Using the constraints

imposed during the minimizations, the systems were equilibrated with a 30 ps simulation at 300 K using single-timestep Langevin dynamics and SHAKE[5] to constrain the bonds involving hydrogen atoms. Unconstrained minimization was then performed using 10,000 steps of SD followed by 20,000 steps of ABNR. Another 30 ps simulation at 300 K followed by minimization consisting of 2,000 steps of SD followed by 4,000 steps of ABNR was then performed. This was followed by a longer 130 ps dynamics simulation at 300 K. The equilibration procedure was continued using the NAMD[6] program, which was used for production dynamics simulations. The energy in each system was minimized using the Powell algorithm with fixed positions for all protein and DNA heavy atoms. Systems were further equilibrated for 100 ps at constant pressure and temperature. Pressure was maintained at 1 atm using the Langevin piston method[7], 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.

Production dynamics were performed at constant temperature and volume using NAMD[6] with the CHARMM force field[4]. The temperature was 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. SHAKE[5] was employed with a 2 fs time step. The system was simulated in periodic boundary conditions, with full electrostatics computed using the PME method[8]. Short-range nonbonded terms were evaluated at every step using a 12 Å cutoff for van der Waals interactions and a smooth switching function. The total simulation length for all systems was 20 ns. All NAMD simulations were run on local and NCSA SGI Altix 3700 Intel Itanium 2 processor shared-memory systems running the Linux operating system.

Supplementary References:

1 Foley, M. C., Arora, K., and Schlick, T., Sequential side-chain residue motions transform the binary into the ternary state of DNA polymerase λ. *Biophys. J.* **91**, 3182-3195 (2006).

2 Foley, M. C. and Schlick, T., Simulations of DNA pol λ R517 mutants indicate 517's crucial role in ternary complex stability and suggest DNA slippage origin. *J. Am. Chem. Soc.* (2008) in press.

3 Brook, B. R. et al., Charmm: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187-217 (1983).

4 MacKerell, A. D., Jr. and Banavali, N. K., All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comput. Chem.* **21**, 105-120 (2000).

5 Ryckaert, J. -P., Ciccotti, G., and Berendsen, H. J. C., Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327-341 (1977).

6 Phillips, J. C. et al., Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781-1802 (2005).

7 Feller, S. E., Zhang, Y., Pastor, R. W., and Brooks, B. R., Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **103**, 4613-4621 (1995).

8 Darden, T. A., York, D. M., and Pedersen, L. G., Particle mesh Ewald: An $N\cdot\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys*. **98**, 10089-10092 (1993).

9 Arora, K. and Schlick, T., In silico evidence for DNA polymerase-beta's substrate-induced conformational change. *Biophys. J.* **87**, 3088-3099 (2004).

Supplementary Figure Legends

Supplementary Figure 1. Conformational changes in the template strand and in pol λ resulting from dNTP binding. A, Overlay of binary (PDB 1XSL; brown) and ternary (PDB 1XSN; green) pol λ complexes[7]. As a consequence of dNTP binding, the template strand undergoes a conformational change that is accompanied by repositioning of the β-hairpin formed by β-strand 3-4 and four active site side chains, Tyr505, Phe506, Arg514 (not shown) and Arg517. B, In the binary configuration, Arg517 stacks

with the templating base (black arrow) and prevents the template strand from adopting the ternary catalytic conformation

Supplementary Figure 2. Time evolution of the RMSD of DNA backbone atoms (P, O3', C3', C4', C5', O5', O5T) in all mutant simulations and wild-type (WT) simulation started from the ternary state relative to the crystal binary (PDB entry 1XSL, green) and ternary (PDB entry 1XSN, red) structures. Superimposition is performed with respect to protein Cα atoms. RMSD is calculated using a previously developed method designed to capture the complex motions of enzymes that is detailed in the Appendix. R517K (nuc), Lys mutant simulation without the catalytic ion; R517K (nuc + cat), Lys mutant simulation with both the nucleotide-binding and catalytic ions.

Supplementary Figure 3. Time evolution of the Lys517 CG-CD-CE-NZ dihedral angle and corresponding three main conformational states: "extremely away" (red), "toward" (green), and "away" (orange) orientations, as discussed in the text. Average distances between Lys517:NZ and T6:N3 are provided. For comparative purposes, Ala517 is shown and the average distance between Ala517:CB and T6:N3 is given. R517K (nuc), Lys mutant simulation without the catalytic ion; R517K (nuc + cat), Lys mutant simulation with both the nucleotide-binding and catalytic ions.

Supplementary Figure 4. Time evolution of key distances between Lys517 and nearby DNA bases, T5 and T6. Superimposed onto plots a, b, c, and f are the Lys517 dihedral angle plots of CG-CD-CE-NZ (brown) from Supplementary Figure 3 (panels e and f) to reveal correlations in the data. Panels a, c, e are from the Lys mutant simulation without the catalytic ion [R517K (nuc)];

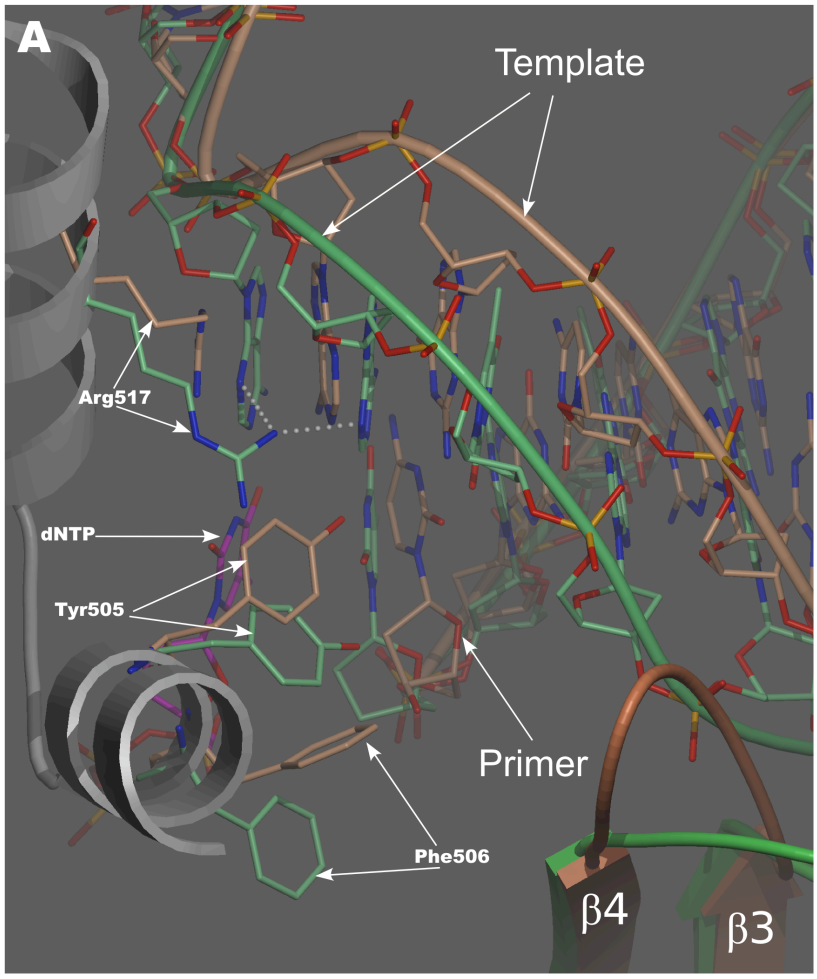
panels b, d, f are from the Lys mutant simulation with both the nucleotide-binding and catalytic ions [R517K (nuc + cat)].

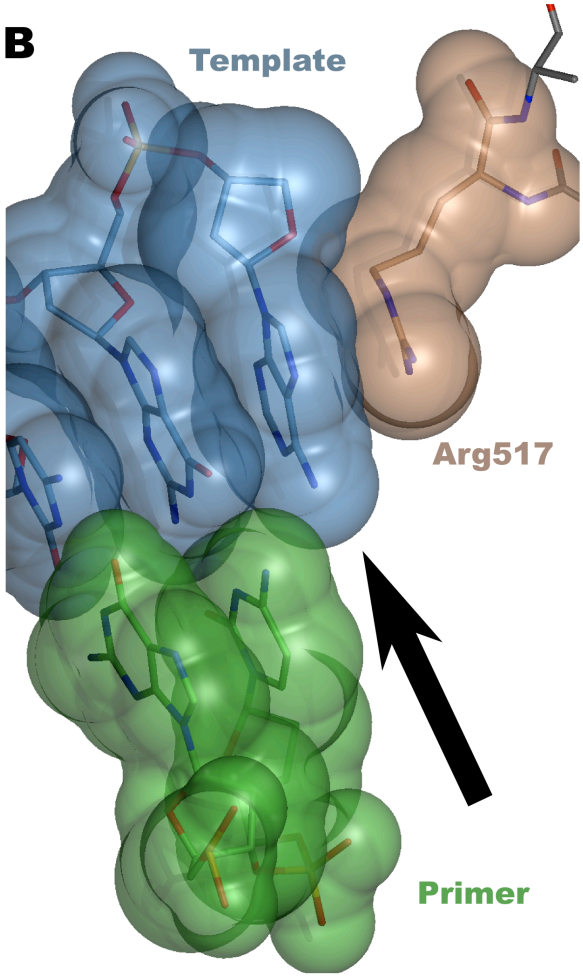
Supplementary Figure 5. Diagram showing the hypothetical relationship between the position of the simulated structure and the crystal states. The sides of the triangle are labeled as a, b and c. The shift distance, h, represents the displacement of the simulated structure in the direction perpendicular to the line joining the geometric centers of the crystal structures, c. The variable lengths L1 and L2 describe the RMSD of the simulated structure projected onto c with respect to the crystal binary and ternary states, respectively.

Supplementary Table 2: Percentage of Each R517K Pol λ Mutant Simulation During Which Water-Mediated Interactions Occur Between K517's ε-Amino Group and the DNA

R517K Pol λ Simulation	DNA Atoms		
	T5:N3	**T6:N3**	**T6:O4′**
1 (R517K pol λ/DNA/dTTP/nuc Mg^{2+})	3.4 %	17.3 %	4.2 %
2 (R517K pol λ/DNA/dTTP/nuc+cat Mg^{2+})	0 %	18 %	0.2 %

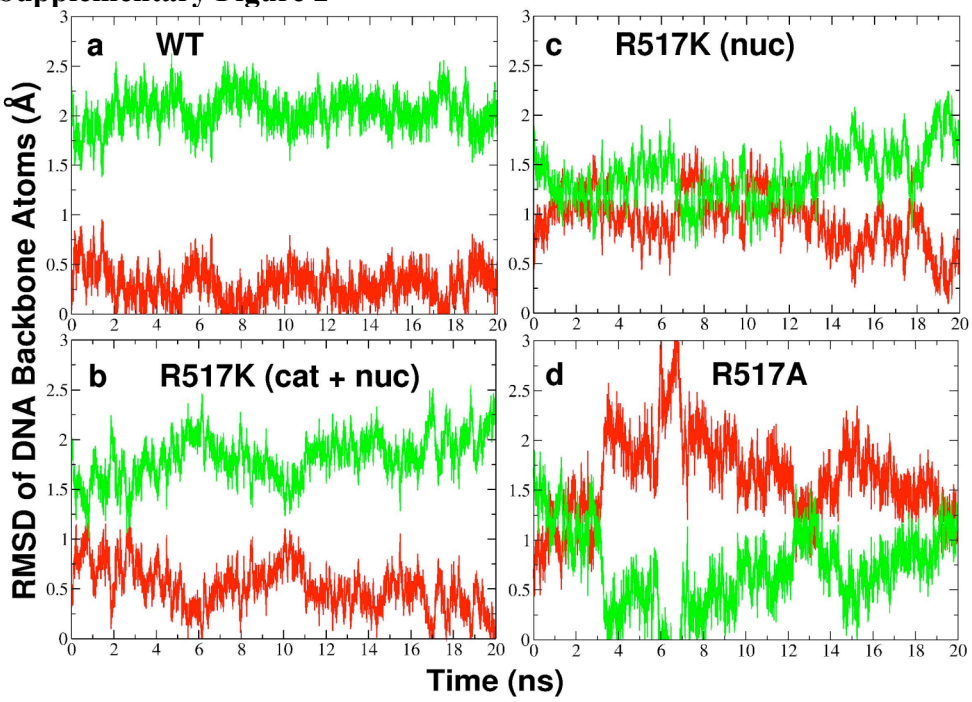
T5, template base at gap; T6, template base of primer terminal base pair; nuc, nucleotide-binding ion; cat, catalytic ion



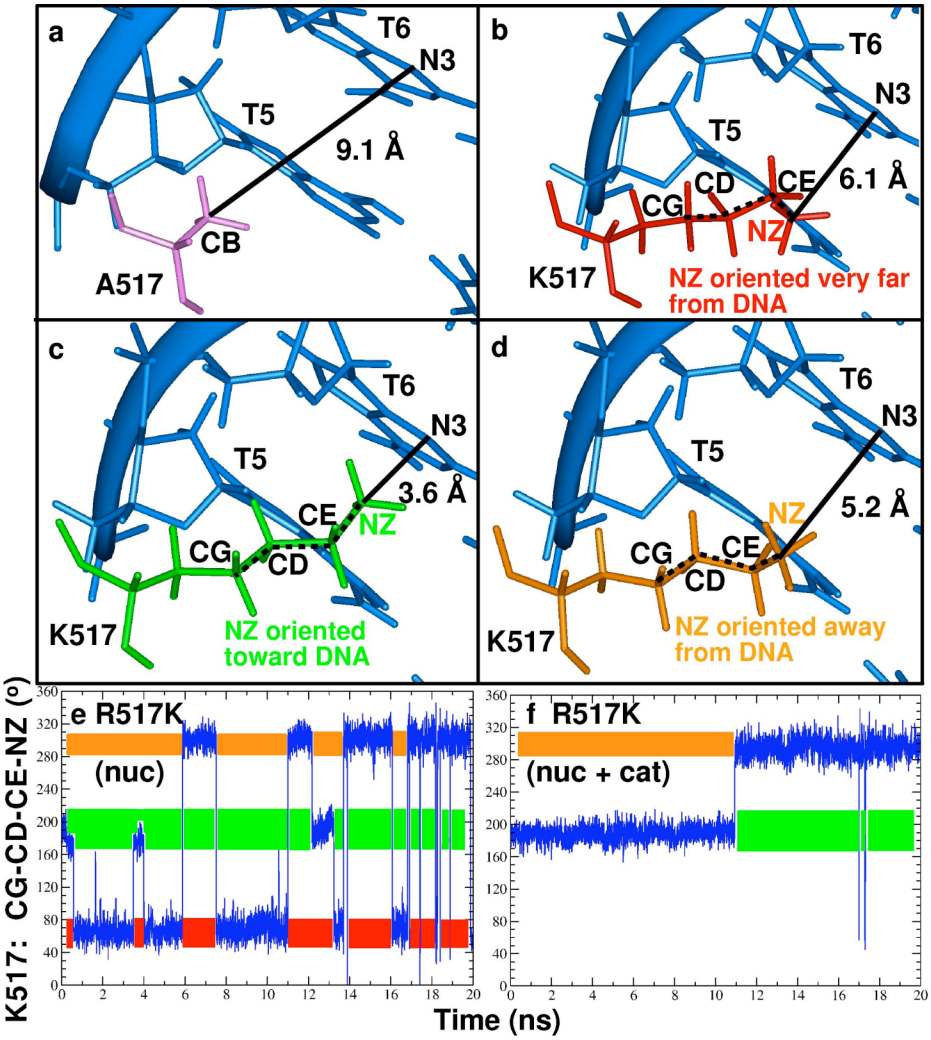


Supplementary Fig. 1

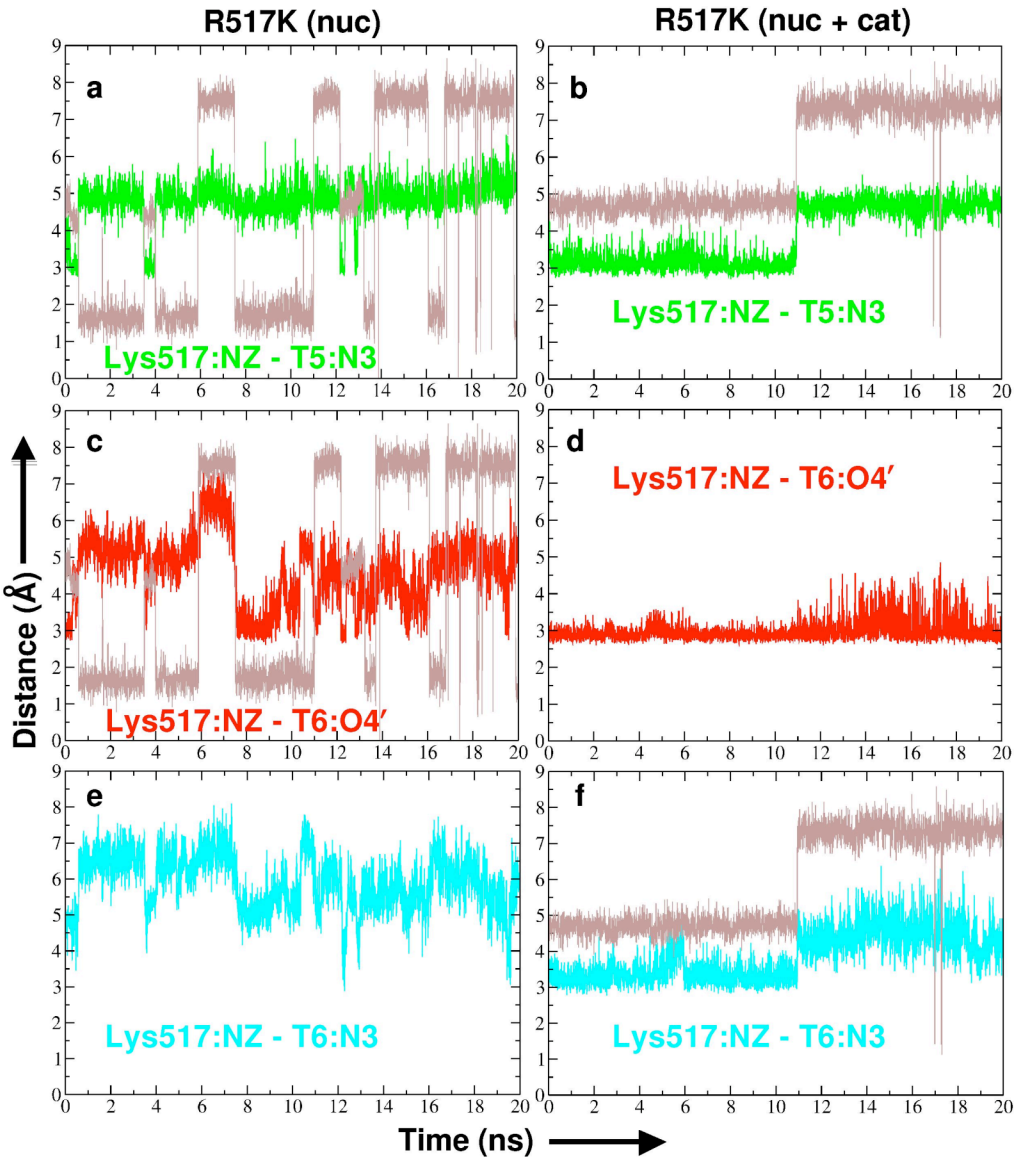
Supplementary Figure 2



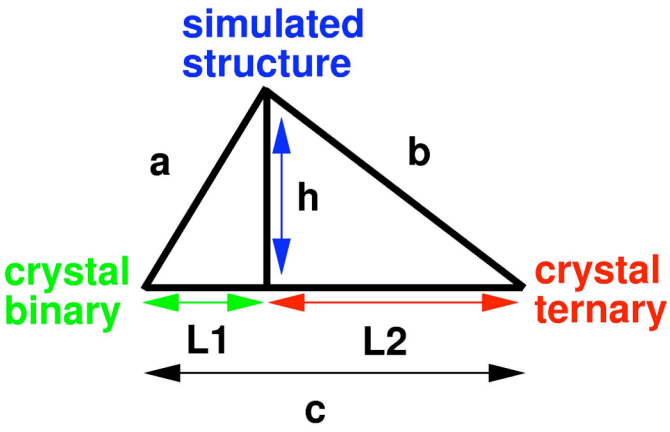
Supplementary Figure 3



Supplementary Figure 4



Supplementary Figure 5



Appendix

A previously developed approach to represent the RMSD data more accurately in enzymes was employed to capture the motion of the DNA and subdomains in our pol λ simulations[9]. In this alternative procedure, we project the RMSD of the simulated structure on the line joining the geometric centers of the structure in the crystal binary and ternary conformations. As shown in Supplementary Figure 5, this can be represented by a triangle where the crystal binary and ternary conformations form two vertices of the triangle and the simulated structure forms the third. The lengths of side a and b of the triangle are given by the RMSD of the simulated structure relative to the crystal binary conformation and crystal ternary conformation, respectively, when the simulated structure and the corresponding crystal structure are superimposed with respect to all protein Cα atoms. The RMSD between the crystal binary and ternary conformations forms side c of the triangle and is held fixed. The shift distance, h, describes the displacement of the simulated structure in the direction perpendicular to the line joining the geometric centers of the structure in the crystal binary and ternary conformations. When h is constant, RMSD alone is a good measure of the motion. When h is not constant, this is an appropriate methodology to use to represent motion toward the crystal states. The variable lengths L1 and L2 correspond to the projected

RMSD of the simulated structure with respect to the crystal binary and ternary conformations, respectively.