

Polymerase β Simulations Suggest That Arg258 Rotation is a Slow Step Rather Than Large Subdomain Motions *Per Se*

Linjing Yang¹, William A. Beard², Samuel H. Wilson², Suse Broyde³ and Tamar Schlick^{1*}

¹Department of Chemistry and Courant Institute of Mathematical Sciences, New York University and the Howard Hughes Medical Institute, 251 Mercer Street New York, New York 10012 USA

²Laboratory of Structural Biology, National Institute of Environmental Health Sciences P.O. Box 12233, Research Triangle Park, NC 27709-2233 USA

³Department of Biology, New York University, New York NY 10003, USA

The large-scale opening motion of mammalian DNA polymerase β is followed at atomic resolution by dynamic simulations that link crystal “closed” and “open” conformations. The closing/opening conformational change is thought to be key to the ability of polymerases to choose a correct nucleotide (through “induced fit”) and hence maintain DNA repair synthesis fidelity. Corroborating available structural and kinetic measurements, our studies bridge static microscopic crystal structures with macroscopic kinetic data by delineating a specific sequence, Phe272 ring flip, large thumb movement, Arg258 rotation with release of catalytic Mg^{2+} , together with estimated time-scales, that suggest the Arg258 rearrangement as a limiting factor of large subdomain motions. If similarly slow in the closing motion, this conformational change might be restricted further when an incorrect nucleotide binds and thus play a role in pol β 's selectivity for the correct nucleotide. These results suggest new lines of experimentation in the study of polymerase mechanisms (e.g. enzyme mutants), which should provide further insights into mechanisms of error discrimination and DNA synthesis fidelity.

© 2002 Elsevier Science Ltd.

Keywords: human DNA polymerase β ; dynamics simulations; fidelity; closed state; open state

*Corresponding author

Introduction

Cells have evolved sophisticated machinery to replicate and repair DNA accurately and efficiently. DNA polymerases are crucial components of these macromolecular networks. Crystal structures of several polymerases reveal common architectural features.^{1–7} Shaped like a hand,⁸ with fingers, palm, and thumb subdomains (Figure 1), the polymerases' palm catalyzes phosphoryl transfer. The thumb and fingers subdomains of the “right-handed” DNA polymerases play a role in positioning the duplex DNA and the nascent base-pair (i.e. incoming nucleotide triphosphate and its templating base) into the polymerase active site, respectively.⁹ In contrast, these roles are played by

the fingers and thumb, correspondingly, of the “left-handed” X family DNA polymerases that include polymerase β (pol β).¹⁰ The distinct subdomain nomenclature for the left-handed X family polymerases highlights the non-homologous nature of the catalytic palm subdomain of pol β relative to the palm subdomain of the other polymerase families (A, B, Y, and reverse transcriptase; see Beard & Wilson¹⁰ for a discussion). Although the catalytic palm subdomain of members of the X family of DNA polymerases is structurally different from other families, all DNA polymerases appear to position reactive groups (i.e. metals, dNTP, and primer terminus) in a similar three-dimensional arrangement consistent with a “two-metal-ion” mechanism for nucleotidyl transfer.¹¹

During DNA replication and repair synthesis, polymerases must select the correct deoxynucleoside triphosphate (dNTP) from a pool of structurally similar molecules so as to preserve Watson-Crick hydrogen bonding rules. Kinetic analyses for several polymerases (*Escherichia coli* DNA polymerase I Klenow fragment,^{12,13} phage T7

Abbreviations used: pol β , DNA polymerase β ; TMD, targeted molecular dynamics; SD, steepest descent; ABNR, adapted basis Newton-Raphson; RMSD, root-mean-square deviation.

E-mail address of the corresponding author: schlick@nyu.edu

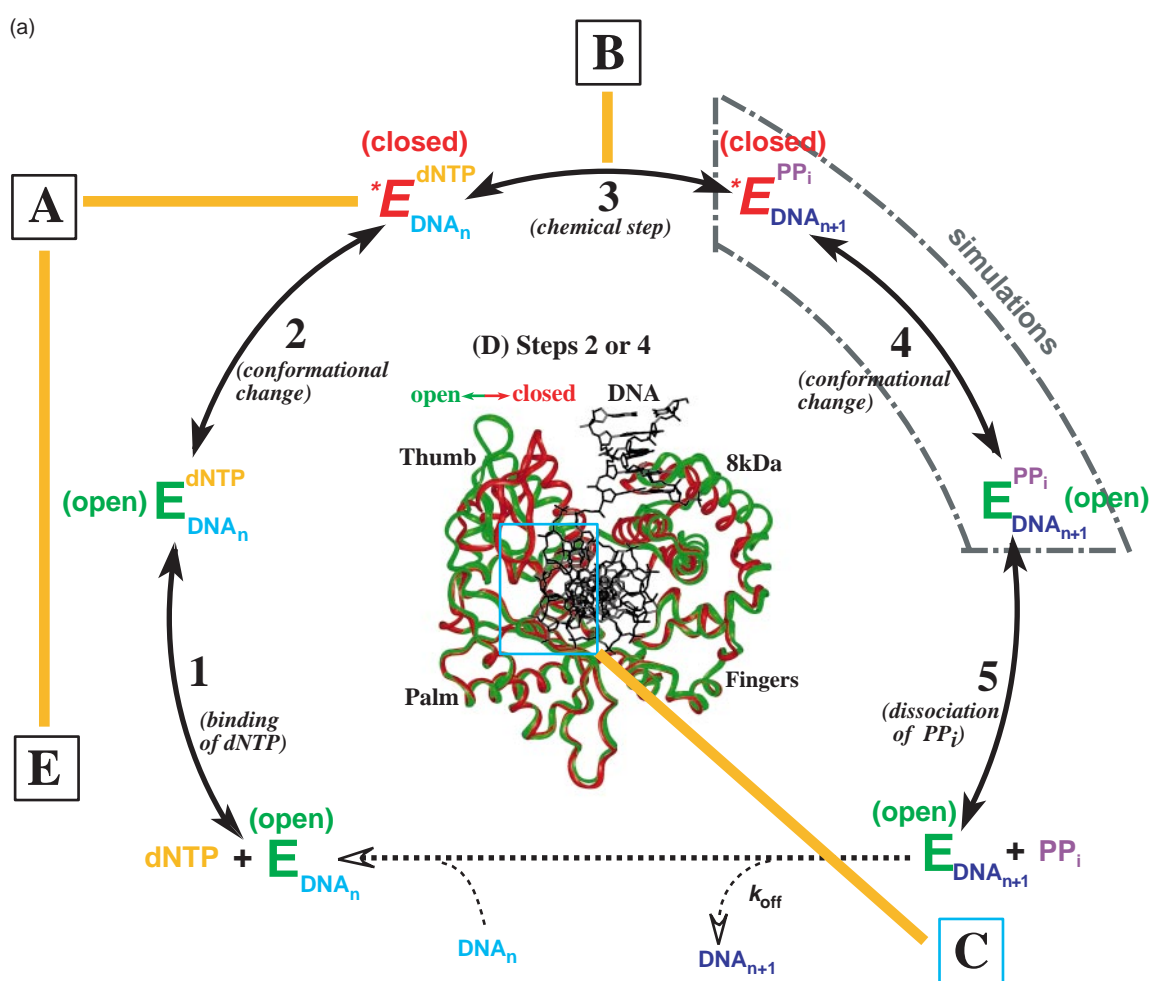


Figure 1 (legend opposite)

DNA polymerase,¹⁴ HIV-1 reverse transcriptase,¹⁵ phage T4 DNA polymerase,¹⁶ and DNA polymerase β ^{17,18} have resulted in the widely accepted scheme outlined in Figure 1 for the insertion of a nucleotide. Following DNA binding, a DNA polymerase binds dNTP to form a ternary substrate complex (step 1). This complex undergoes a conformational change to align the catalytic groups and form a productive complex (step 2). Chemistry occurs rapidly to form the product complex (step 3). This complex undergoes a second conformational change (step 4) prior to the release of the pyrophosphate moiety (PP_i).

The conformational rearrangements involved in steps 2 and 4 are believed to represent key mechanisms used by polymerases to enhance fidelity. Considerable experimental evidence suggests that binding of the correct nucleotide facilitates the first conformational change (closing, step 2), whereas binding of the incorrect nucleotide does not.¹⁹⁻²² For a proofreading exonuclease-containing DNA polymerase, such as Klenow fragment, the second conformational change (opening, step 4) may slow the kinetic cycle to allow the polymerase to remove

an incorrectly inserted nucleotide. This “induced-fit” mechanism is consistent with a large body of experimental data (static crystal structures and kinetic studies). Namely, X-ray crystal structures of various DNA polymerases indicate that the polymerase active site is “open” when bound to primer/template DNA,²⁻⁵ but “closes” around the nascent base-pair when the correct nucleotide binds.^{1-3,6,7} Thus, the enzyme/DNA complex likely alternates between open and closed states such that the binding and release of the substrates and products, as well as the linear diffusion or sliding of the DNA, occur in the open state, while the chemical reaction occurs in the closed state.

Details of the precise events involved in the large-scale opening/closing motion, and how they may regulate synthesis fidelity, are unknown. In particular, kinetic data (see Table 1) reveal slow conformational steps before and after chemistry (steps 2 and 4 of the catalytic cycle),²³ but the identity of these conformational changes is obscure. Given recent suggestions, based on enzyme kinetic measurements, that the opening/closing motions themselves are relatively fast,^{18,22,23} the question

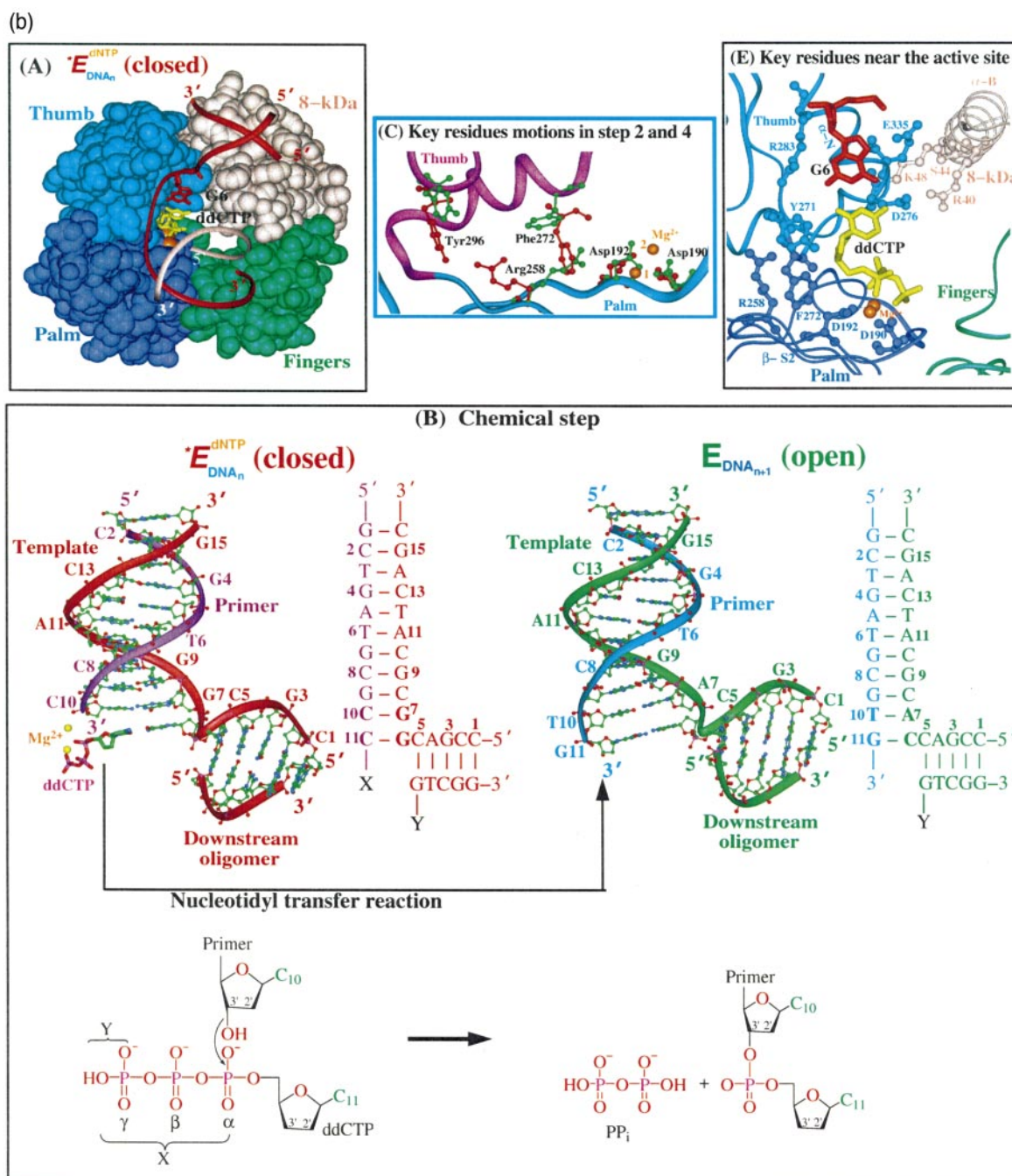


Figure 1. General pathway for nucleotide insertion by pol β and corresponding conformations of pol β /DNA complex.

arises regarding the identity of the molecular rearrangements that occur to align active-site residues so as to perform chemistry. Here, we investigate conformational changes in the polymerase opening, but do not consider the chemical step, dissociation, or translocation.

To gain insights into this key conformational arrangement and its biological ramifications, we explore this process by dynamic simulations of pol β , the smallest mammalian polymerase (335 amino acid residues).^{24,25} Its 31 kDa C-terminal domain

performs phosphoryl transfer common to polymerases,²⁶ while the 8 kDa N-terminal domain of pol β has lyase activity, that is, it catalyzes the release of the resultant 5'-deoxyribose phosphate after an endonuclease cleaves an apurinic/apyrimidinic site during base excision repair.^{27,28} Though computational approaches are vulnerable to many basic approximations in force fields and technique limitations,²⁹ state-of-the-art simulations that are well anchored to experimental structures and other measurements can help consolidate observations

Table 1. Experimental kinetics data

Polymerase	Rate constants (s ⁻¹)		References
	Before chemistry (k_{pol})	After chemistry	
DNA polymerase β	3-90	70	18-22, 48, 49
DNA polymerase I (Klenow)	50	15	13, 40
Phage T7 DNA polymerase	300	1200	14, 88
Bacteriophage T4	400		16, 89
HIV-1 reverse transcriptase	30 ^{DNA} 70 ^{RNA}		15 90, 91

Kinetics of slow conformational changes prior to and following the chemical bond-forming step (reaction of the dNTP with the 3'-end of the primer strand). The column k_{pol} refers to the rate constant describing nucleotide insertion. This step is believed to be limited by a conformational change (step 2 in Figure 1).

and propose new hypotheses, which can in turn be tested, for example, by measurements with rationally designed mutant enzymes.

Pol β has been crystallized in complexes representing three intermediates of the opening/closing transition:² the open binary complex of pol β ·gap, containing pol β bound to a DNA substrate with a single nucleotide gap; the closed ternary complex containing pol β ·gap·ddCTP (i.e. pol β bound to the gapped DNA as well as a 2',3'-dideoxyribocytidine 5'-triphosphate (ddCTP)); and the open binary product complex, pol β ·nick: pol β bound to nicked DNA. The large motion between the open and closed forms is illustrated in Figure 1(D); the DNA sequences in the closed ternary and open binary product crystal complexes are shown. Note that the template residues C6 and A7 in the pol β ·nick structure are different from G6 and G7 in the pol β ·gap·ddCTP structure, as are their complementary residues in the primer strand.

The three crystal structures provide static information relating to different catalytic states of pol β . They indicate that pol β 's thumb (residues 262-335) repositions itself after binding the correct dNTP (by rotating about the α -helix M axis) to position α -helix N so that several side-chains can interact with the nascent base-pair in the closed conformation. As mentioned, though the structural changes that occur in conjunction with the kinetic conformational changes are unknown, recent evidence suggests that the large subdomain motions, inferred from comparison of structures of binary polymerase/DNA complexes with those bound with an incoming dNTP, are relatively rapid.¹⁸ Our reported simulations suggest collectively that rather than the closing/opening motion *per se*, a slow step for alignment of the catalytic residues is the rotational rearrangement of Arg258 coupled to Mg²⁺ release/binding. If similarly slow in the closing motion, this conformational change might be further restricted when an incorrect nucleotide binds and thus play a role in pol β 's selectivity for the correct nucleotide.

Results

Models setup

Two models of solvated pol β /DNA complexes were prepared for dynamic simulations based on the 1BPY (closed ternary) and 1BPZ (open binary product) crystal structures,² as detailed in Computational Methodology. The first model of a closed complex (post-chemistry) was constructed by reacting the α -phosphate group of the ddCTP with the 3'-hydroxyl group of the DNA primer strand in the crystal closed ternary complex, leaving PP_i and two specific magnesium ions in place. We call this model the modified closed form. The second model (post-chemistry) was an intermediate between the open and closed states. It was constructed as an average of the coordinates of the modified closed ternary complex (the first model) and the open binary product crystal structure, to represent a complex with the thumb in a partially open state. We call this model half-open. The preparation of both models involved adding hydrogen atoms and missing residues in the crystals. Our constructed half-open structure (ensured to be free of steric clashes by small modeling adjustments using INSIGHT II version 2000, followed by minimization and equilibration) resembles well (Figure 2(B)) a crystal structure of a half-open intermediate complex (J. M. Krahn, W.A.B. & S.H.W., unpublished results).

Our solvated models were prepared in face-centered cubes with water using the programs Simulaid³⁰ and PBCAID.³¹ Counterions (Na⁺ and Cl⁻) were used to neutralize the biomolecular systems (using the Delphi package^{32,33}) at an ionic strength of 150 mM. This produces a closed solvated model of 43,751 atoms, and a half-open solvated model of 41,973 atoms. Cartesian-coordinate energy minimizations, equilibrations, and dynamics simulations were performed using the program CHARMM^{34,35} with the all-atom force field, version 27. Production dynamics of 6 ns duration for the modified closed model and 4.8 ns for the half-open model were performed at room temperature 300 K using our efficient stochastic LN approach.³⁶⁻³⁸ In addition, several dynamics runs for the modified closed model at high temperature

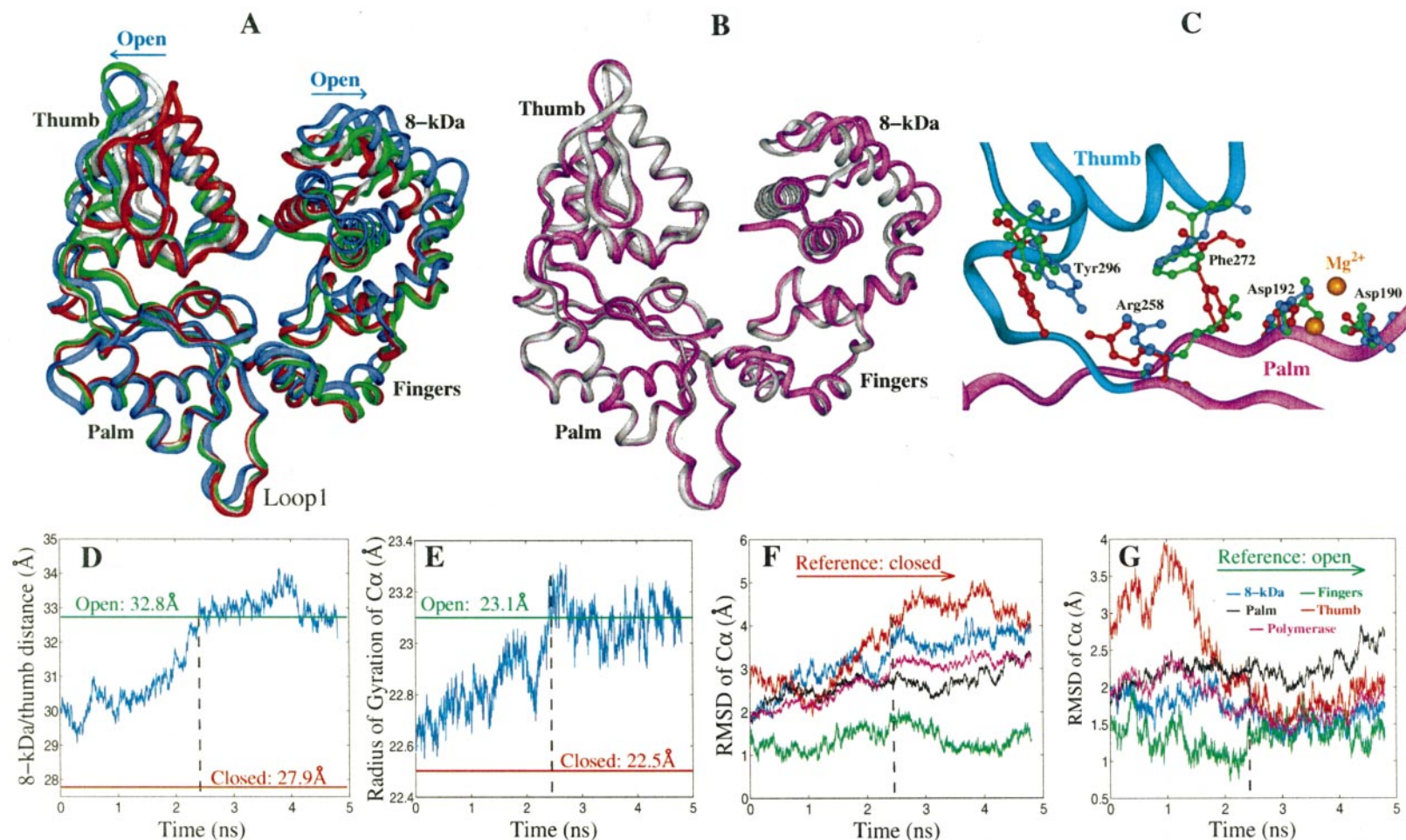


Figure 2. Comparison of crystallographic and simulated structures. (A) Superimposition of the averaged trajectory structure of pol β in the open form (blue) to the crystallographic closed (red) and open binary product (green) complex as well as the initial constructed half-open (grey) according to the palm C α atoms. (B) Comparison of initial constructed half-open (grey) and crystal mismatch half-open (pink) according to the palm subdomain (RMSD of 1.3 Å obtained for the thumb). (C) Conformational comparison of key residues in the simulated open, crystallographic closed and open binary product structures, with the same blue/red/green color definitions. Calculated (D) center to center distance (d_i) between the 8 kDa and thumb subdomains; (E) radius of gyration (R_g) for all C α atoms; (f) and (g) RMS C α atom deviation of instantaneous simulation structure with respect to the crystal closed ternary and open binary product complexes, superimposed according to all C α atoms except for those of residues 1-9, whose coordinates are missing from the crystal ternary complex.

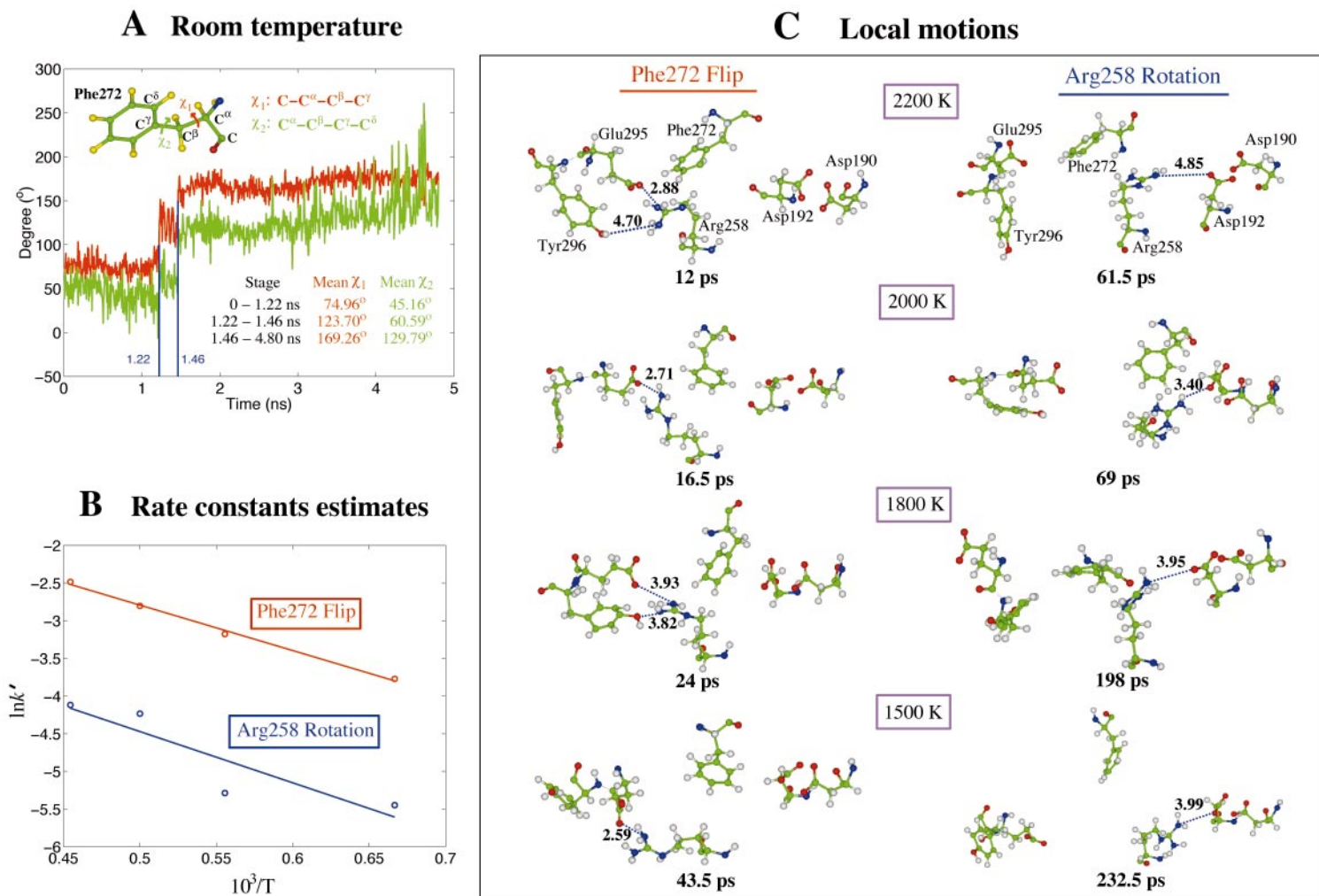


Figure 3. Phe272 and Arg258 conformational transitions during opening. (A) Time evolution of two dihedral angles around Phe272 over the entire simulation: χ_1 (red) and χ_2 (green). Values for three intervals in the trajectory for the two dihedral angles are given. (B) Arrhenius equation plots for the temperature-dependence of specific rate constants for the Phe272 ring flip (red) and the Arg258 rotation toward Asp192 (blue). (C) Conformations in the active site and time-scales for Phe272 ring flip (left) and Arg258 rotation (right) at temperatures 2200 K, 2000 K, 1800 K, and 1500 K. The green, red, blue, and white spheres represent carbon, oxygen, nitrogen, and hydrogen atoms, respectively. The atom-atom distances between Arg258 NH1 and Glu295 O^{δ2}, Arg258 NH2 and Tyr296 OH, Arg258 NH2 and Asp192 O^{δ2} are given.

as well as several targeted molecular dynamics (TMD) simulations, were performed from a closed to an open state, as detailed in Computational Methodology.

Subdomain motions

The simulation from the modified closed structure at room temperature failed to capture the pol β opening over several nanoseconds. However, opening was captured in our two 4.8 ns simulations started from the half-open structure. Analysis of a representative simulation for the half-open structure at room temperature (Figure 2) reveals this opening, as reflected by the evolution of the center-to-center distance (d_i) between the 8 kDa and thumb domains (heavy-atoms only) as a function of time, and the radius of gyration over all C α atoms (R_g). We analyze this simulation throughout the work; the other trajectory (differing by initial velocities from the same Gaussian distribution) yields similar results.

The subdomain movement analysis shows that the thumb moves more than other subdomains (largest RMSD of 5 Å compared to <4 Å for others), approaching a stable open state midway through the simulation, at \sim 2.4 ns. In overall flexibility, the 8 kDa domain is second to the thumb (largest RMSD is about 4 Å), with fingers and palm relatively less flexible (nearly superimposable when comparing the crystal closed, the crystal open binary product, and the simulation's open structures). The very close thumb overlay between our simulated open and the crystal open product structures indicates that a stable model for open pol β has been approached. Indeed, crystallographic data suggest that large thumb movements for the main chain occur during the opening/closing motion.^{2,25}

Since our half-open structure was designed to accelerate the opening process, our room-temperature simulation from this structure probably does not follow a natural reaction pathway. However, given the partial opening captured in a nanosecond time-frame, it is plausible that the full-scale opening, though certainly slower, is not slower by many orders of magnitude. Taken together with the much longer time-scales estimated for the conformational change of the ternary product complex (step 4 in the catalytic cycle, see Table 1), the simulations suggested to us that other specific rearrangements must be involved.

Arg258 rotation is slow

Sawaya *et al.*² deduced the following characteristic conformational changes accompanying the thumb movement from the closed to the open state from crystal structures (Figure 2(b)). In the closed state, Arg258 hydrogen bonds with residues Glu295 and Tyr296, and the phenyl ring of Phe272 disrupts the salt-bridge between Asp192 and Arg258, freeing Asp192 to ligand the catalytic and

nucleotide-binding Mg²⁺. In the open state, the phenyl ring of Phe272 has moved away from Asp192, and Arg258 has rotated toward Asp192, poised to engage in a salt-bridge. Note that, in the open binary product crystal structure,² the Arg258 rotation readies it to form a salt-bridge with Asp192, but that Asp192 has not moved to complete this interaction; the actual salt-bridge is observed, however, in the crystal open binary gapped complex.

Our simulations at room temperature from the half-open structure capture the Phe272 ring movement away from Asp192, allowing Arg258 to rotate so as to form the salt-bridge with Asp192. We did not capture the Arg258 rotation toward Asp192, as observed in the open crystal structure (Arg258 remains hydrogen bonded to Tyr296). This rotation, along with the breakage of the Arg258/Tyr296 hydrogen bond, likely requires additional energy. Indeed, high-temperature molecular dynamics simulations capture the Arg258 rotation fully, as do the targeted molecular dynamics simulations (described below).

The flipping process of the Phe272 phenyl ring can be described by two dihedral angles (Figure 3(a)): χ_1 for C-C α -C β -C γ , measuring the rotation of the entire phenyl ring, and χ_2 for C α -C β -C γ -C δ , measuring the inclination of the ring plane. Figure 3(a) shows that each dihedral angle occupies characteristic ranges as a function of time for the room-temperature simulation. Larger fluctuations for χ_2 at the end suggest greater flexibility of the ring's inclination than its overall rotation. The second transition in Figure 3(a) corresponds to the motion of the Phe272 ring away from Asp192, which occurs at \sim 1.5 ns, that is, before the thumb subdomain opens. Together with the high-temperature simulations (see more below), we propose the following sequence of events for pol β 's motion: (1) flip of Phe272's phenyl ring, vacating room for Arg258 rotation; (2) opening of pol β through a large thumb movement; and (3) Arg258 rotation toward Asp192. This sequence suggests that the rearrangement of certain key residues might be much slower than pol β 's thumb subdomain opening.

Exploratory simulations at temperatures of 1500, 1800, 2000, and 2200 K were started from the modified closed structure. All coordinates for the C α atoms in the fingers and palm subdomains were fixed, as suggested by the stable positions of these atoms in open and closed crystal forms.² Though the polymerase secondary structure is not maintained at high temperature, the local movements of Phe272 and Arg258 are reasonable. The conformations of six key residues, involved in the flip of the Phe272 phenyl ring and the rotation of Arg258, as well as the time-scales for both motions and certain distances, are displayed in Figure 3(c) at differing temperatures.

The suggested sequence of events in pol β opening is in accord with estimated activation barriers E_{a1} and E_{a2} , and rates associated with the Phe272

ring flip and Arg258 rotation, respectively. From an Arrhenius plot of $\ln k'$ versus $1/T$, where k' is the rate constant and T is absolute temperature, we estimate the k' as the reciprocal of time required for these two motions. A least-squares fit of the natural logarithms of these constants k' plotted against $1/T$ (Figure 3(B)) shows a fairly linear dependence in this broad temperature range, with activation barriers of ~ 12.0 kcal/mol (1 cal = 4.284 J) for Phe272 flip and ~ 13.5 kcal/mol for Arg258 rotation. This linearity reveals approximate temperature-independence for the activation barriers over the temperature range examined.

The former estimate is close to that determined experimentally for a phenylalanine ring flip in basic pancreatic trypsin inhibitor.³⁹ The latter is comparable to the experimental Gibbs free energy barrier of 12-16 kcal/mol for the conformational step following chemistry in DNA polymerase I (Klenow fragment)^{13,40} and T7 DNA polymerase.¹⁴ The corresponding extrapolated times for the Phe272 flip and Arg258 rotation at room temperature (300 K) are approximately 4×10^{-4} second and 2×10^{-2} second, respectively. Thus, completing the conformational change from the closed to the fully open complex, including Arg258 rotation toward Asp192, would require roughly 10 ms. We

note that the rate constant of thumb movement itself cannot be estimated from the Arrhenius plot, since the thumb's secondary structure is not maintained at high temperature.

In addition, several 100 ps TMD simulations (see Computational Methodology)⁴¹⁻⁴⁶ for the modified closed structure from the closed to open state provide further support to our main conclusion that Arg258 rotation is slow. The TMD simulations, in which the full Arg258 rotation occurs only when a large force is applied and only following pol β opening, also suggest a high energy barrier. The constrained potential energy versus time in one of the TMD simulations is plotted in Figure 4, with a series of snapshots illustrating a possible pathway for Arg258 rotation and Phe272 flip. Note that the high peak for the potential energy appears at the last segment of the simulation, accompanying Arg258 rotation. Although the TMD simulations do not provide a natural pathway for pol β opening, they support our suggestion that a high energy barrier is involved for Arg258 full rotation.

Thus, the room-temperature simulations from the half-open model and the TMD simulations from the modified closed model show the following sequence of events: (1) Phe272 flip; (2) large thumb movement; (3) Arg258 rotation toward Asp192; the high-temperature simulations from the

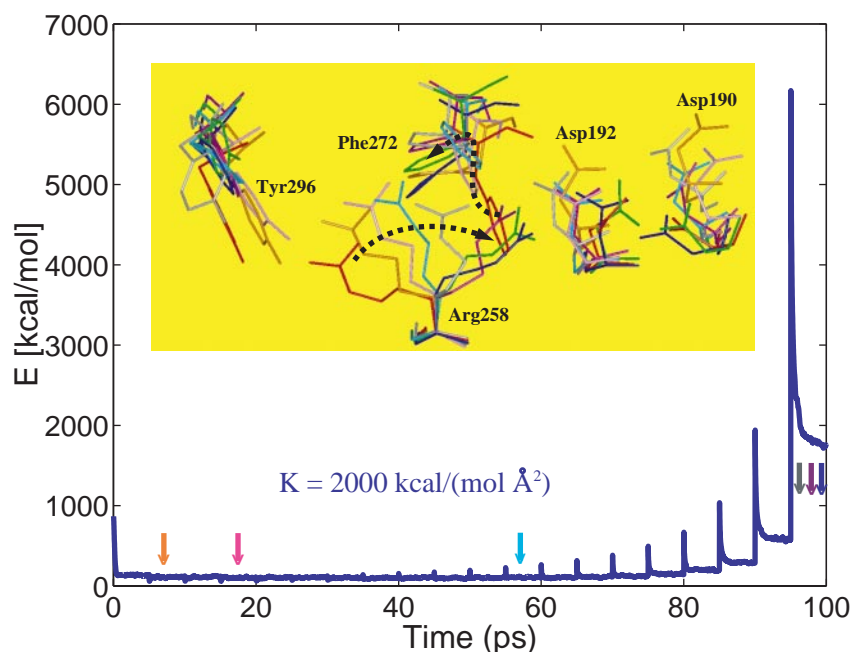


Figure 4. Targeted molecular dynamics simulation from the closed to open form. A representative pathway for Arg258 rotation and Phe272 flip is shown along with restrained potential energy versus time. The active-site residues in the crystal closed ternary and the open binary product complexes are shown in red and green, respectively; those in simulated structures are shown sequentially in time along the TMD pathway in orange, magenta, light blue, grey, purple, and deep blue, with corresponding arrows along the x -axis. The functional form of the RMS restraint energy is: $E_{\text{RMS}} = K[D_{\text{RMS}}(X(t), X^{\text{target}}) - d_0]^2$, where K is a force constant (in kcal mol $^{-1}$ Å $^{-2}$), D_{RMS} represents the relative RMS distance for a selected set of atoms between the instantaneous conformation $X(t)$ and the reference X^{target} , and d_0 is an offset constant (in Å).

modified closed model reveal that Arg258 rotation follows the Phe272 flip, but do not provide evidence regarding thumb movement. The question then arises as to why the Phe272 flip and thumb movement occur in the nanosecond time regime in the room-temperature simulations from the half-open model. A probable answer is that the energy barriers for these motions have been crossed in this half-open model.

The time estimate for Arg258 rotation can now be related to experimental kinetic data regarding slow conformational steps (Table 1), before and after the chemical step, for several polymerases, to place the slow conformational rearrangement in the larger reaction context. Recall that the first slow conformational change before chemistry produces a tight, closed complex upon binding the correct dNTP; the fast chemical step that follows produces a tightly bound, closed enzyme/product complex, which undergoes a second conformational change that converts the closed complex into the open loose form.^{13,14,40,47} For pol β , slow conformational changes occurring both before and after chemistry have been identified by fluorescence and kinetic analyses.^{18-21,48,49} These conformational changes are of comparable magnitudes (Table 1).

Catalytic Mg^{2+} departure may be slow during the pol β opening process

In the closed crystal structure complex, two magnesium ions lie in the active site. The first, the nucleotide-binding ion, is coordinated with the α , β , and γ -phosphate groups of ddCTP (Figure 1), two conserved aspartate residues Asp190 and Asp192, and one water molecule. The second Mg^{2+} is a catalytic ion coordinated with the α -phosphate group of ddCTP, the 3'-OH group of the primer, Asp190, Asp192, Asp256, and one water molecule. These two magnesium ions play an essential role in assisting the addition of a correct dNTP to the 3' terminus of the DNA primer strand and stabilizing

the positioning of the incoming dNTP, and are key to the polymerase's fidelity.^{1-3,6,11,23,50-52} In the open binary product crystal complex, in contrast, there are no crystallized Mg^{2+} nor PP_i in the active site.² Hence, in this initial study of the polymerase opening motion, we do not include the two magnesium ions or the product PP_i in our half-open structure. Sodium counterions are included in our solvated model; their role is to achieve physiological salt concentration rather than to model the functional role of the Mg^{2+} .

In the simulation, we observe one Na^+ counterion coordinated with the three active-site aspartate residues Asp190, Asp192, and Asp256 (see Figure 5). These residues are normally bound to the catalytic Mg^{2+} in the closed complex. During the first two-thirds of the trajectory, this sodium ion coordinates with Asp190, the incoming nucleotide, and a number of dynamic water molecules. During the last third, it moves to coordinate with Asp192 and Asp256, as well as two stable water molecules. Since our Na^+ position coincides with that occupied by the catalytic Mg^{2+} in the closed ternary crystal structure,² it may be acting as a surrogate electrostatically, but not catalytically, for the missing (and crystallographically absent from the open form) Mg^{2+} in the half-open structure. We infer that the catalytic Mg^{2+} may remain in the active site throughout the opening process. Dynamic simulations of the pol β /DNA complexes containing the catalytic Mg^{2+} and/or the nucleotide-binding Mg^{2+} (accompanied by pyrophosphate) are now being performed, to assess their effects on the pol β conformational opening.

Water density is low around the new base-pair in the closed state but increases during the pol β opening

Analysis of the water environment within 6 Å of the new base-pair (T6-P11) for the simulated closed, half-open, and open structures shows that local water oxygen densities increase systemati-

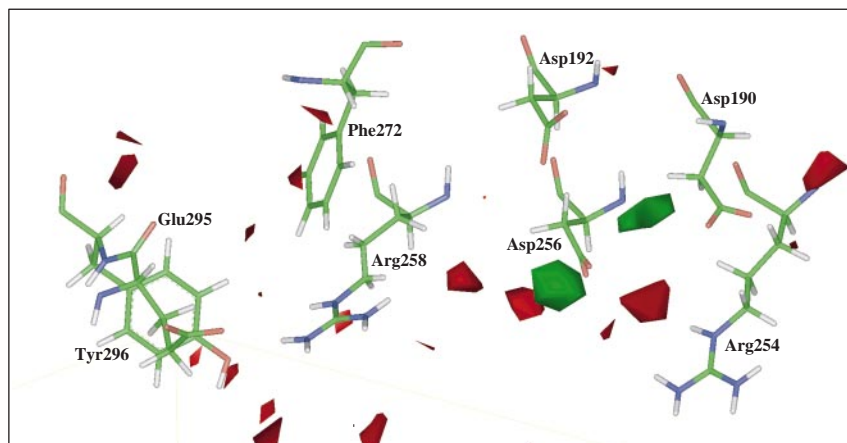


Figure 5. Na^+ probability density of >6% (green) and water probability density of >35% (red) in the active site of open pol β /DNA complex. The density was accumulated on a 1 Å cubic lattice over 200 snapshots sampled at a frequency of 3 ps.

cally during the pol β opening, due to the α -helix N motion away from the new base-pair (see Figure 6). The tight packing in the closed structure excludes water molecules near the new base-pair and thereby permits the polymerase to assemble active-site components so they are poised for the chemical reaction.^{1,2,47}

Dynamics of key hydrogen bonds reflect thumb motion

Key hydrogen bonds between subdomains or between polymerase and DNA help describe the pol β opening. Hydrogen bond hb1 between Arg40 NH2 in the 8 kDa domain and Asp276 O δ^2 in the thumb forms during the initial part of the room-temperature simulation and breaks as the polymerase opens (Figures 7 and 8). Indeed, hb1 is present in the closed crystal complex but not in the open form and enhances dNTP binding affinity.¹⁸ Hydrogen bond hb2 between Lys48 NH2 (8 kDa) and Glu335 O ϵ^2 (thumb), unlike hb1, is present in

the open crystal complexes but not in the closed form, where its Glu335 O ϵ^2 atom is hydrogen bonded to the nearby Ser44 OH (8 kDa) (hb3). This alternate arrangement can be explained as follows: as the thumb moves away from the 8 kDa domain during the opening, Glu335 leaves Ser44 and approaches Lys48, rupturing hb3 and forming hb2. The hydrogen-bond between the thumb's Glu316 O ϵ^1 and palm's Arg182 NH2 found in both closed and open crystal structures persists throughout the simulated opening. In addition, a salt-bridge between Asp256 and Arg254, found in both closed and open crystal complexes,² was present in our initial half-open model and persisted for ~ 2 ns in the simulation; subsequently, however, the amino groups of Arg254 rotated somewhat so that Arg254 was no longer within hydrogen bonding range with Asp256.

Our simulations indicate that Tyr271, which may affect polymerase closing by hydrogen bonding with the 3'-end primer base,^{20,53} participates in

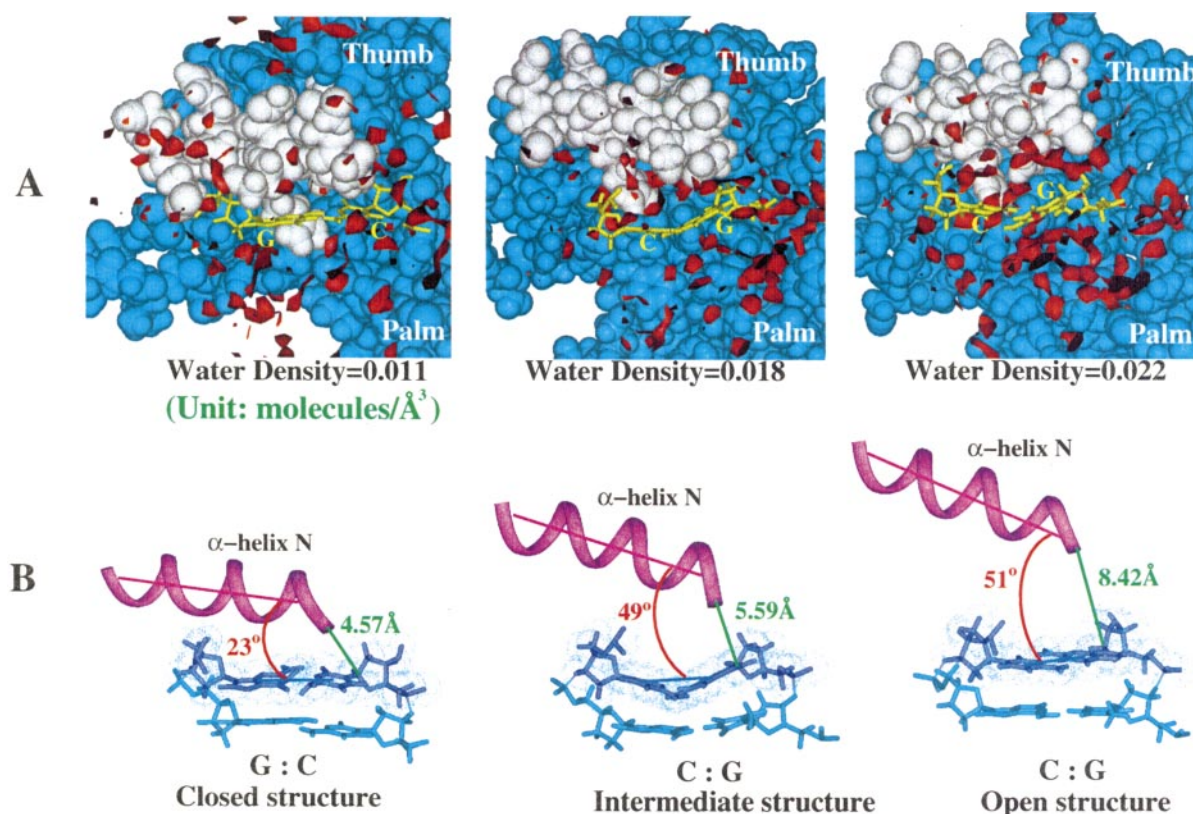


Figure 6. Water density during opening. Top: Water probability density of $>20\%$ (red) within 6 \AA around the incoming nucleotide and its complementary template base (yellow) in the simulated closed, half-open, and open structures, respectively. The water density is 0.011 molecules/ \AA^3 , 0.018 molecules/ \AA^3 , and 0.022 molecules/ \AA^3 for the three structures, respectively. For reference, the standard (bulk) water density is 0.033 molecules/ \AA^3 . Bottom: Relative position of the polymerase α -helix N (purple) with respect to the new base-pair (dark blue) and its neighbor base-pair (light blue). The van der Waals surface of the new base-pair is illustrated (blue dots). The angles (red) between the α -helix N axis and the line through the two center points of individual base planes in the new base pair, and the distances (green) of Asp276 C $^\alpha$ with the Cyt N1 in the closed structure and with Gua N9 in the half-open and open structures are given.

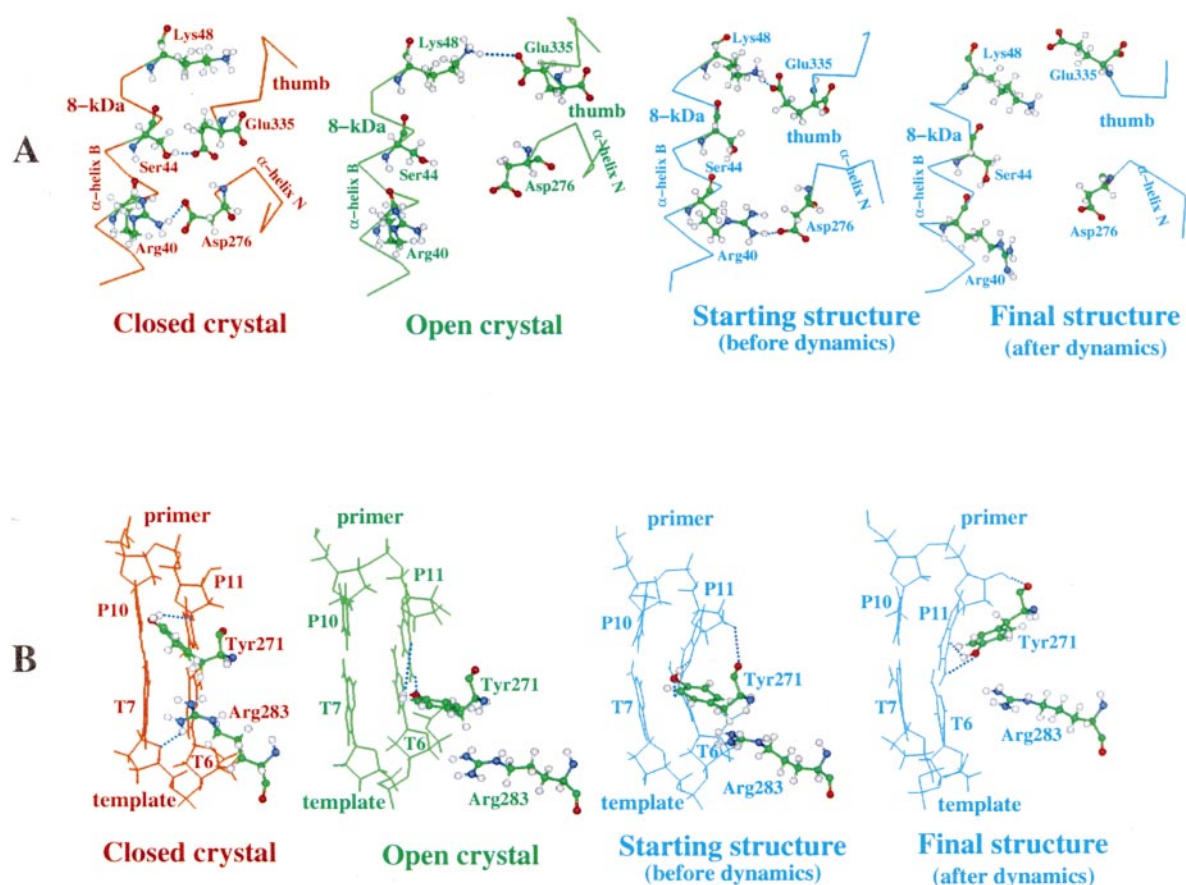


Figure 7. Important hydrogen bonds in the closed ternary and open binary product crystal structures and in the simulated half-open structure before and after MD simulation: (a) between thumb and 8 kDa domain; (b) between protein residues, Tyr271 and Arg283, and the nascent DNA base-pair.

three hydrogen bonds with the incoming dNTP (Figures 7 and 8) throughout the simulation: Tyr271: O/P11: NH₂, Tyr271: OH/P11: N3, and Tyr271: O/P11O3H. Tyr271 exhibits small movements away from P10 toward the 3'-end residue P11 by breaking the hydrogen bonds with P10 as the polymerase opens. These trends suggest that Tyr271 plays an important role in the pol β opening.

DNA in the active site exhibits some A-like characteristics

Lu *et al.*⁵⁴ have noted partial A-like character in the DNA steps adjacent to the site of nucleotide incorporation for pol β /DNA complexes in their analysis of these crystal structures. Our analysis of key geometric features of the DNA demonstrates such partial A-DNA-like features near the active site (see Figure 9).^{54,55} A-like characteristics in the active site, as well as B-DNA structural features in the duplex regions away from the active site, are both present in the starting half-open model and are preserved during the simulation. This A-like tendency^{1,3,4,7,56–58} has been considered biologically

significant, since a widened minor groove permits the primer/template DNA to contact better the active-site residues such as Lys234, Tyr271, and Arg283, thereby assisting in positioning the 3'-terminus of the primer for the nucleotidyl transfer reaction, as noted previously.^{1,52,57,58}

Discussion

Our dynamic studies bridge kinetic and structural approaches for studying pol β mechanisms by suggesting the sequence and time-frames of structural transitions associated with the rate of correct nucleotide insertion.⁴⁷ These data corroborate and further elaborate upon experimental measurements (Table 1) and structural data. Our simulations suggest that the thumb movement *per se* is rapid and may not account for the slower conformational changes that govern the kinetics of correct nucleotide insertion. On the other hand, slower, subtle structural transitions involving a network of active-site groups (e.g., Arg258 and others, see below) might occur subsequent to such subdomain movements. Though we have not modeled the magnesium ions in our simulations from the half-

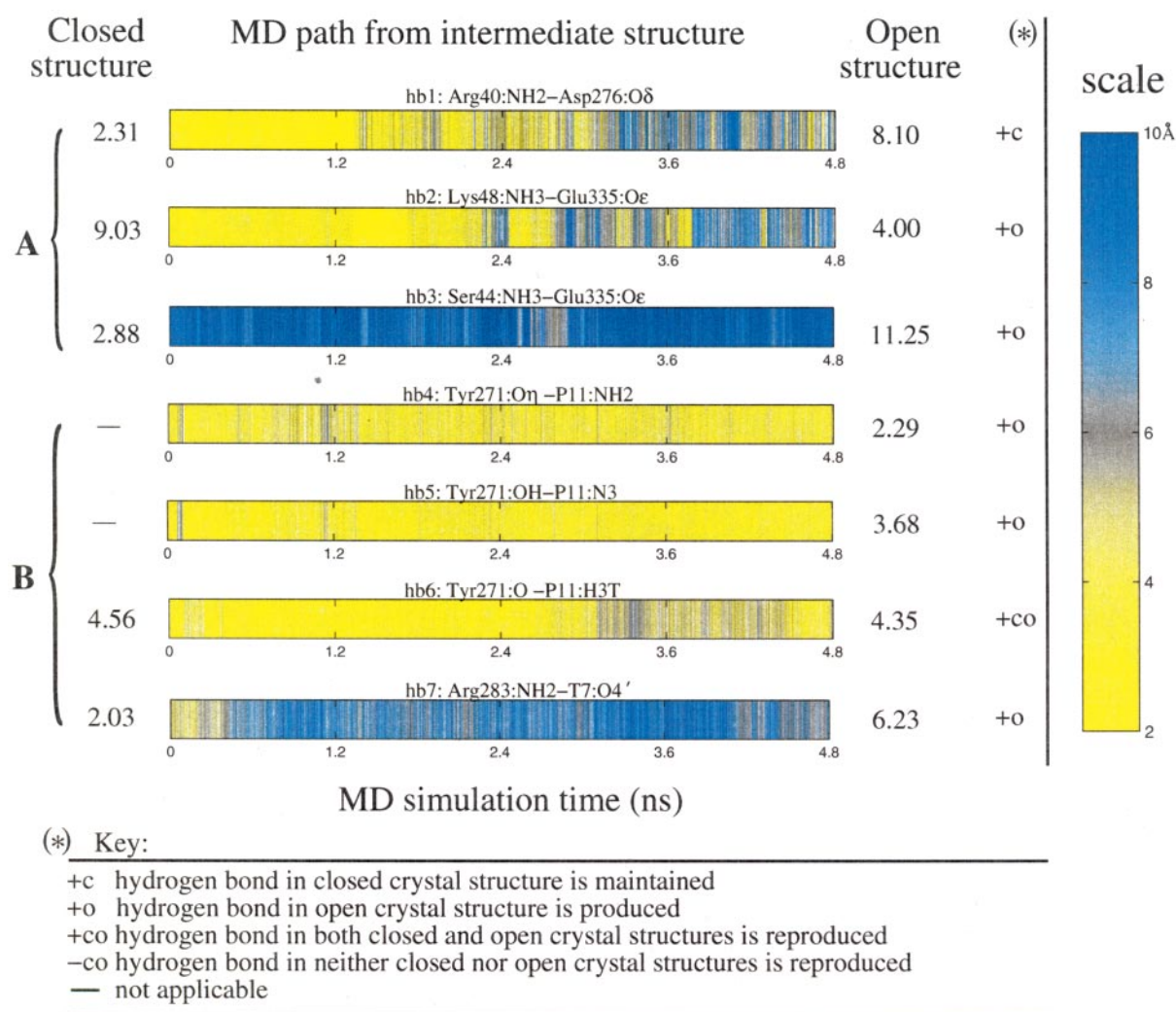


Figure 8. Distance variations (in Å) between the two heavy-atoms in the hydrogen bonds shown in Figure 7. The distances are compared with the corresponding values in the closed and open crystal structures. Hydrogen bonds in groups A and B are classified as in Figure 7. The color scale from yellow to blue indicates increasing length values.

open form (since they are absent from the open binary product crystal structure), our sodium ions serve as surrogates electrostatically. The simulations suggest that the catalytic ion in the active site (Figure 5) is important, since, by coordinating Asp192, it prevents formation of the Asp192/Arg258 salt-bridge and possibly hinders the Arg258 rotation in pol β opening. This scenario suggests that the release of the catalytic Mg^{2+} and the Arg258 rotation may occur in concert, following pol β 's opening. Kinetic data⁴⁸ have indicated that the conformational transitions that occur before and after chemical reaction, as monitored by fluorescence changes of the template base analogue 2-aminopurine, have very similar half-lives. Thus, the conformational change identified here may be a limiting factor both before and after the chemical step. Similarly, we hypothesize that the binding of the catalytic Mg^{2+} and the rearrangement of

Arg258 may be coupled and represent a slow step in pol β 's closing before chemistry.

Our suggestion for the slow, local structural motion of Arg258 in the active site ties well with the Vande Berge *et al.* experiments,¹⁸ which recently demonstrated that template-directed ground-state dNTP binding is altered by mutagenesis of the thumb subdomain. Since this ground-state binding occurs prior to the rate-limiting conformational change, and dNTP interactions with the thumb subdomain occur in the closed conformation of pol β , these results indicate that closing of the thumb is rapid (i.e. not rate-limiting). The open and closed conformations were noted to be dramatically different in these two structures, specifically in residues Asp192, Arg258, Phe272, and Glu295. Additionally, Arg283 of helix N in the thumb subdomain was seen to approach Glu295 when it closed. The importance of these residues was demonstrated also by amino acid substitutions

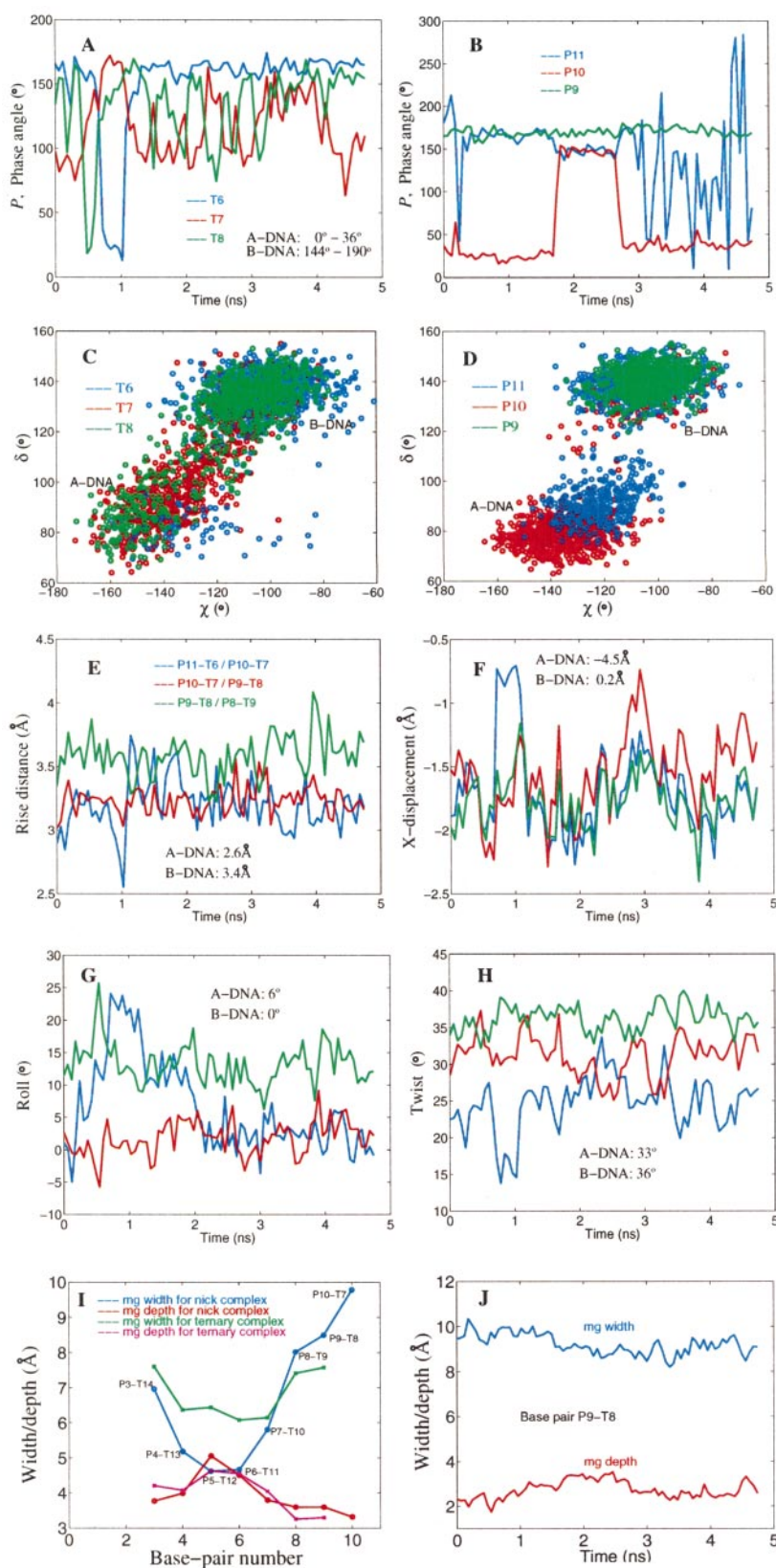


Figure 9. Time-dependence of conformational parameters for three base-pairs closest to the primer 3' terminus. (A)-(D) sugar pseudorotation angle P and scattergrams of χ/δ for the six base-pair residues, respectively; (E)-(H) are the local base-pair parameters rise, x -displacement, roll, and twist for the three base-pairs. The corresponding values in the canonical fiber models of *A*-DNA and *B*-DNA⁸⁴ are presented for comparison. The torsion angles χ and δ correspond to $O_4-C_1-N_9-C_4$ and $C_5-C_4-C_3-O_3$.⁸⁵ The sugar pseudorotation parameter is defined elsewhere⁸⁶ and the rise, x -displacement, roll, and twist are the nucleic acid structure parameters as defined.⁸⁷ (i) Minor groove (mg) width and depth of primer/template DNA in crystal structures: closed ternary complex and open binary product complex. (j) Minor groove width/depth of base-pair P9-T8 as a function of time from the half-open structure. P and T indicate primer and template strand, respectively; the subscripts indicate the residue numbers, 5' to 3', in each strand. The minor groove width/depth for canonical X-ray fiber diffraction models of *A*-DNA and *B*-DNA⁸⁴ is 11.0/2.8 Å and 5.7/7.5 Å, respectively.

for Asp192,⁵⁹ Phe272,⁶⁰ Arg283,⁵³ Glu295,⁶¹ or Tyr265,²² which resulted in a significant decrease in catalytic efficiency and/or fidelity. Thus, these

side-chain rotations are required for catalytic metal-binding/dissociation; the loss of this network can result in a mutator polymerase.^{53,62} Our

work points to one of these residues, Arg258, as a potential major player in the reaction.

That domain movement may be faster than side-chain motion has been suggested by work reported by Li *et al.*^{63,64} In studies of ddNTP-trapped ternary complexes of the large fragment of DNA polymerase I from *Thermus aquaticus* (klentaq1),^{63,64} it was found that an arginine residue at position 660 (outside the nucleotide-binding site) selectively affects ddGTP incorporation, but not dGTP incorporation. Note that the polymerase can be locked in the closed conformation by the ddGTP. This suggests that the side-chain motion of Arg660 to interact with the G base is slower than domain movement, a deduction that agrees well with our observation that the Arg258 motion is slower than the thumb subdomain movement in the pol β /DNA complex.

Available crystallographic evidence for DNA polymerases does not contradict our hypotheses above. For example, Pelletier *et al.*⁶⁵ have reported structures of pol β in the open conformation that have metal ions (Mn^{2+}) bound to the catalytic and nucleotide metal site subsequent to chemistry. Our simulations at high temperature have shown that the catalytic Mg^{2+} remains very stable in the active site by coordinating the key conserved aspartate residues. Still, recent crystallographic structures²³ of pol β -DNA-Cr(III) \cdot dTMPPCP (before chemistry) and pol β -DNA-Cr(III) \cdot PCP (after chemistry) have suggested that the binding of the catalytic ion follows the pol β closing before chemistry, and the release of the catalytic Mg^{2+} is prior to the pol β opening after chemistry. However, the absence of electron density for the catalytic Mn^{2+} in that work²³ may simply be due to the "excess" charge on Cr^{3+} . Stopped-flow measurements examining changes in 2-aminopurine fluorescence have observed rapid conformational transitions,^{22,66} but apparently this spectroscopic probe does not respond to slower conformational changes such as catalytic metal ion binding. Further approaches to isolate and measure the dynamics of the catalytic metal ion are needed.

While we cannot establish that Arg258 rearrangement is a slow, limiting factor by computer simulations alone, specific experiments can be designed to test our suggested order of events. Since Arg258 is involved in a network of interactions with Asp192, which itself coordinates both catalytic and nucleotide-binding ions, mutating Arg258 and examining the effects on this network can help analyze the importance of these residues, the role of the ions, and how conformational change influences nucleotide selection. Indeed, alanine and lysine mutants of Arg258 are being examined kinetically and computationally for effects on metal binding and fidelity. A previous study demonstrated that alanine substitution for Arg258 did not have a dramatic affect on activity,⁶⁷ but fidelity measurements are lacking. Additional approaches, such as pulsed electron paramagnetic resonance (EPR) and time-resolved fluorescence techniques, should provide insights into the

dynamics of active-site metals, the nascent base-pair, subdomains, and specific protein side-chains for incorporation of correct/incorrect nucleotides. Together, such kinetic, structural, and simulation results should provide further insights into the mechanisms involved in error discrimination. In particular, studies are needed to suggest how recently discovered DNA polymerases, such as DNA polymerase η or ι , replicate lesion-containing and undamaged DNA templates with low fidelity.⁶⁸⁻⁷¹

Computational Methodology

Systems preparation

PDB/RCSB⁷² coordinates of human DNA polymerase β entries 1BPY (closed, ternary complex) and 1BPZ (open, binary product complex) were used to construct initial solvated models. All hydrogen atoms were added to the crystallographic heavy-atoms by the CHARMM subroutine HBUILD.⁷³

The conformation of the closed complex, in which reaction of primer DNA with the incoming nucleotide has not yet taken place, was modified by connecting the ddCTP to the 3'-end of the DNA primer strand and breaking the bond between P^α and the oxygen atom that bonds with both P^α and P^β , leaving the pyrophosphate moiety in place. A hydroxyl group was added to the 3' terminus of the primer in the modified dC residue. Residues 1-9 of the polymerase and a specific water molecule coordinated to the catalytic Mg^{2+} were added, as they were missing from the ternary crystal structure. Residues 1-4 in the open binary product structure were added.

A half-open model was then constructed as an average of the crystallographic coordinates of the modified closed and open complexes, and the PP_i unit and the magnesium ions are absent from this half-open model. For consistency, the template residues G6 and G7 in the closed structure were replaced by the C6 and A7 residues, respectively, present in the open binary product structure, as were their complementary residues in the primer strand. The half-open model strongly resembles a crystal intermediate structure of pol β /DNA (nicked) complex with a mismatched A:C base-pair at the active site (J. M. Krahn, W.A.B., & S.H.W., unpublished results). The mismatch complex is essentially thought to stabilize the intermediate state. Our constructed half-open structure was expected to accelerate the conformational changes to the open state.

Face-centered cubes for the periodic domain of both initial models were constructed using the program Simulaid³⁰ and PBCAID.³¹ The smallest image distance is 16 Å for the half-open structure and 20 Å for the closed structure. Solvent molecules that were closer than 1.8 Å to the nearest protein, DNA, crystallographic water atoms or ions were removed. To neutralize the system at an ionic strength of 150 mM, water molecules with minimal electrostatic potential at the oxygen atoms were replaced by Na^+ and those with maximal electrostatic potential at the oxygen atoms were replaced by Cl^- . All Na^+ and Cl^- were placed more than 8 Å away from any protein or DNA atoms and from each other. The electrostatic potential of all bulk water oxygen atoms were calculated with the DelPhi package.^{32,33} The initial half-open structure of pol β complexed with the primer/template DNA contains 41,973 atoms, consisting

of 11,830 bulk water molecules, 44 Na⁺ and 21 Cl⁻ counterions. The initial closed structure of pol β complexed with the primer/template DNA and the leaving pyrophosphate contains 43,751 atoms, consisting of 322 crystallographically resolved water molecules as well as 12,096 bulk water molecules, two Mg²⁺ and two Na⁺ obtained from crystallographic coordinates, together with 43 Na⁺ and 23 Cl⁻ counterions (see Figure 10).

Minimization, equilibration and dynamics protocol

Cartesian coordinate energy minimizations and dynamics simulations were performed using the program CHARMM^{34,35} and the all-atom version 27 force-field (Chemistry Department, Harvard University, Cambridge, MA). The partial charges for the pyrophosphate were evaluated on the basis of the partial charges of two reference structures already parameterized in CHARMM; namely, the inorganic phosphate and the anionic methylphosphate. Relevant atomic partial charges were averaged and normalized⁷⁴ so that the total partial charge equals -3 . (The pyrophosphate structure and the atomic partial charges, as well as the atom labels in CHARMM, are shown in Figure S1 of the Supplementary Material).

Each solvated system was equilibrated by energy minimization with fixed positions of all protein and DNA heavy-atoms using the method of steepest descent (SD) for 10,000 steps; this was followed by 20,000 steps minimization by the adapted basis Newton-Raphson (ABNR)^{34,75} scheme with same constraints. To ensure that all counterions were located at the minima or maxima of electrostatic potential around the protein/DNA complex, 30 ps of equilibration was run at 1000 K by the stochastic LN approach^{36-38,76} with all heavy-atoms of protein and DNA fixed, thus allowing the counterions to relax and move freely. The entire system was again minimized by SD for 10,000 steps followed by ABNR for 20,000 steps while still fixing all heavy-atoms of protein and DNA. Another 45 ps of equilibration at 300 K was followed by further minimization using SD and ABNR until the gradient of root-mean-square deviation was less than 10^{-6} kcal/mol Å. Finally, the protein, DNA, counterions and water coordinates were re-equilibrated for 45 ps at 300 K by LN.

The current LN protocol was applied to both the half-open and the closed structures using three timesteps $\Delta\tau$, Δt_m , and Δt . The inner timestep of 1 fs is used for updating all local bonded interactions. The medium timestep of 2 fs is for updating the non-bonded interactions within a spherical distance of 7 Å associated with healing and buffer lengths of 4 Å each. The outer timestep of 150 fs is employed to update the remaining non-bonded interactions up to the global non-bond interaction cutoff (14 Å here) using a heuristic procedure if any atom moves more than 1 Å. SHAKE was employed to constrain all bonds involving hydrogen atoms. The Langevin damping constant γ was 10 ps^{-1} . A dielectric constant of 1.0 was used. Electrostatic and van der Waals interactions were smoothed to zero at 12 Å with a shift function and a switch function, respectively.

LN integrator and performance

LN is an efficient method for propagating biomolecular dynamics according to the Langevin equation.³⁶⁻³⁸ The alternative stochastic formalism can be used for numerical stability considerations as well as to represent

random fluctuations in the cellular environment. The addition of stochastic terms alters the Newtonian description, but is suitable for thermodynamic sampling, as used here, and to deduce configurational patterns. Though time-scales of specific events can change due to the stochastic coupling, with a relatively small damping constant, as used here, the perturbations to Newtonian dynamics are expected to be small, with motions similar and time-scales of the same order.³⁷ For example, we have confirmed that barrier crossing events for butane are very similar at room temperature for Verlet and LN (data not shown). Detailed comparisons for pol β between LN and Velocity Verlet demonstrate this as well. The computational advantages are significant (speedup factors of around 10 and around 8)^{77,78} with respect to inner timesteps of 0.5 fs (around 5 with respect to 1 fs) and important for the required sampling used in this work.

Since LN is being used here for a complex biomolecular system, we present here aspects of LN's stability and reliability in terms of thermodynamic, structural and dynamic properties compared to single-timestep Langevin as well as Newtonian (Velocity Verlet) propagators.

The three LN timesteps are related by integers k_1 and k_2 as follows:

$$\Delta t_m = k_1 \Delta\tau, \Delta t = k_2 \Delta t_m$$

Thus, the combined force at the reference point X_r , $-\nabla E(X_r)$, is evaluated only once per outer timestep Δt ; other force terms are updated more frequently, as described above. The splitting of the non-bonded interactions into medium-range and long-range parts is accomplished using a smooth switching function. The LN algorithm has been presented in detail elsewhere.^{37,38,77}

We evaluate LN trajectories through comparison with the traditional small-timestep methods of Langevin (BBK) and Velocity Verlet (VV) in terms of thermodynamic, structural and dynamic properties as a function of time. For this purpose, we fix $\Delta\tau$ to 1 fs, Δt_m to 2 fs, and vary the outer timestep Δt to 6, 20, 40, 80, 100, 120, and 150 fs. All bonds involving hydrogen atoms employed SHAKE both in LN and the reference Langevin method.

Energy components calculated over 12 ps of simulation from the equilibrated state are compared with the corresponding energies calculated by the single-timestep Langevin method at $\Delta t = 1$ fs in Figure 10(A), in terms of errors of the energy components and speedup factors relative to the reference Langevin method at $\Delta t = 1$ fs. With varied outer timesteps, bond-length, bond-angle, and torsion energies have the smallest relative errors, around 0.2%. The largest relative error of 2.5% occurs in the van der Waals energy for $\Delta t = 50$ fs, but the relative error of the total energy is still only around 1.2%. For small outer timesteps, relative errors are less than 1%. The higher energy errors for $\Delta t = 20$ fs protocol are likely due to resonance artifacts.^{38,77} The above results demonstrate the stability of the LN protocol at $\Delta\tau/\Delta t_m/\Delta t = 1/2/150$ fs and a speedup factor of approximately 4.5 with respect to the reference Langevin at $\Delta t = 1$ fs.

The spectral analysis in Figure 10(B) for the protein, DNA, and water atoms obtained from the three algorithms, LN as above, and BBK, at $\gamma = 10 \text{ ps}^{-1}$ and Velocity Verlet ($\gamma = 0$, $\Delta\tau = 1$ fs), further supports the reliability of this protocol for thermodynamic sampling. Spectral densities are sensitive measures of the algorithmic protocol. The data presented resulted from sampling every 2 fs from the above three 12 ps trajectories. The

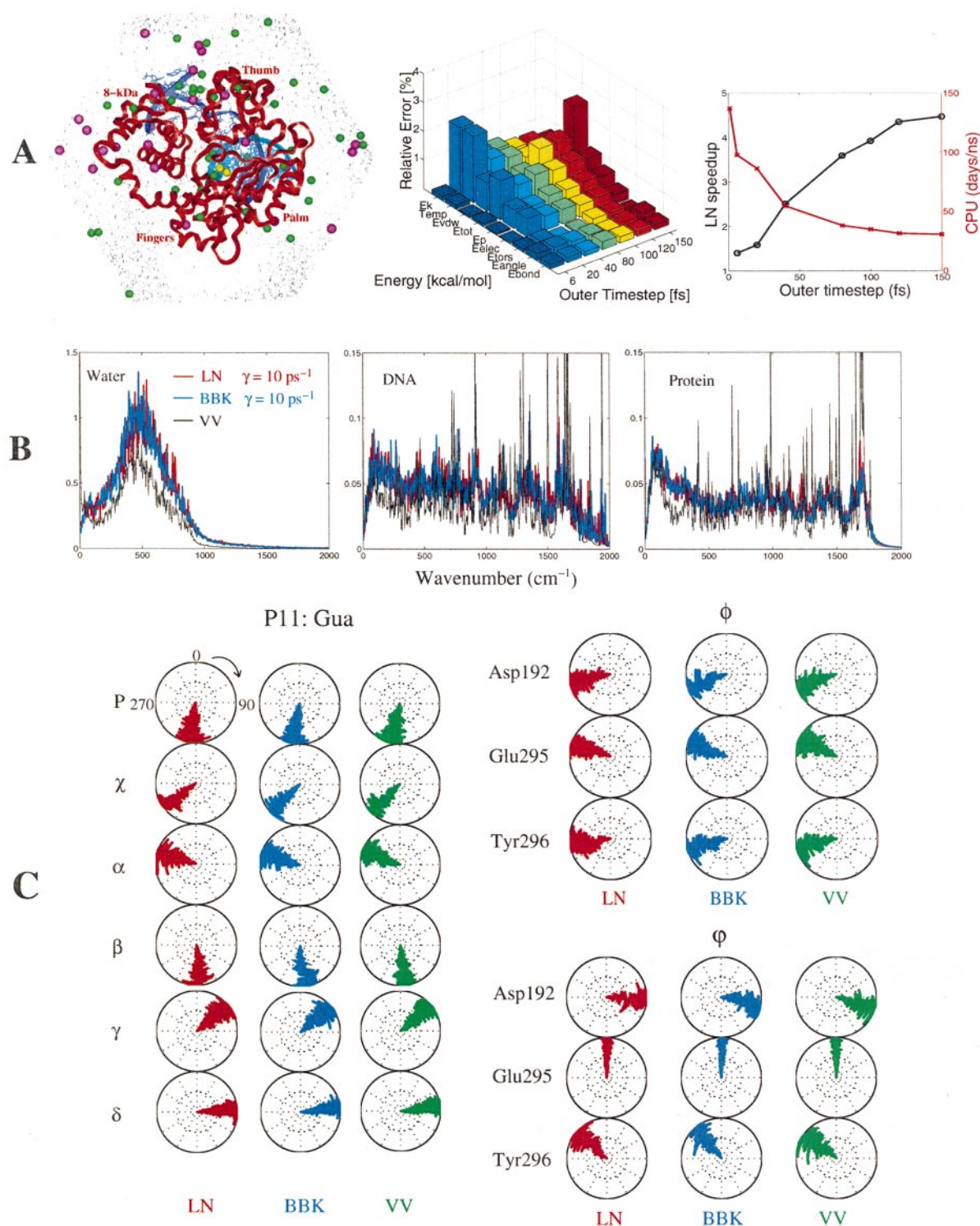


Figure 10. LN performance. (A) Initial model of ternary complex with two bound Mg^{2+} (yellow), 45 Na^+ (green) and 23 Cl^- (purple), as well as 12,418 solvated water molecules (left), error in LN energy components (middle) and LN speedup (right) for DNA pol β complex simulations relative to the explicit Langevin integrator at the single timestep $\Delta t = 1$ fs. LN error is shown for different energy components propagated over 12 ps. (B) Water, DNA, and protein spectra calculated over 12 ps runs and sampled every 2 fs. LN spectra are compared to Langevin spectra as well as spectra from Velocity Verlet. (C) Time-evolution of DNA sugar pseudorotation phase angle P and backbone torsion angles of primer residue 11, and time-evolutions of ϕ and ψ torsion angles of protein residues Asp192, Glu295, and Tyr296 over 12 ps, respectively. SHAKE was employed for all bonds involving hydrogen atoms in the simulation.

spectral analysis procedure⁷⁷ involves computing the velocity autocorrelation time series for all atoms of interest and obtaining a power spectrum by fast Fourier transformation. These spectra are averaged over the water, DNA, and protein atoms separately for the global characterization of the motion.

The LN spectra for water molecules closely resemble the spectra obtained from BBK and VV. For the protein and DNA atoms, although the intensities of both LN and BBK spectra ($\gamma = 10 \text{ ps}^{-1}$) are weaker than VV spectra intensities ($\gamma = 0$), as expected from Langevin damping, most peak positions calculated from these three algorithms are in good agreement with each other, especially the LN and BBK spectra. Thus, as γ decreases the spectra should become more similar.^{37,38}

We also analyzed the conformational changes based on the time evolution of ϕ and ψ torsion angles for three protein residues: Asp192, Glu295, and Tyr296, and the DNA backbone torsion angles (χ , α , β , γ , δ) as well as the sugar pseudorotation phase angles, P , for the 3'-primer terminus guanine residue as shown in Figure 10(C). The time evolutions of these variances using the three algorithms over 12 ps are very similar. In most cases, the residues exhibit stable oscillations of widths of around 30° . Observed differences in the time evolutions obtained from the different approaches are reasonable and expected.

Finally, we ascertain the reliability of the simulation comparing the position fluctuations of protein atoms during the production trajectory with values obtained from the experimental temperature factors using the relation $\langle \Delta r_i^2 \rangle = 3B_i / (8\pi^2)$, where Δr_i is the atomic displacement of atom i from its average geometric position and B_i is the corresponding X-ray temperature B -factor.⁷⁹⁻⁸¹ The atomic fluctuations for the backbone atoms of the polymerase are calculated over the entire simulation range as well as the first and second half of the simulation as averages per residue basis. (Figure S2 of the Supplementary Material compares these fluctuations to those obtained from the crystal B -factors of the open binary product complex.²)

The computed atomic fluctuation patterns show fairly good agreement with the experimental data: the peaks and valleys coincide in general. The regions of higher fluctuation peaks are loop and turn segments positioned between regions of stable secondary structures of α -helices and β -sheets. In particular, the three regions of residues 1-9, 243-249 and 299-305 exhibit relatively large mobility. The intrinsic flexibility (or disorder) of residues 1-9 might explain their absence from the crystallographic structures. Residues 243-249 define the palm's loop1 as shown in Figure 2(A). Residues 299-305 are part of a loop connecting a β -strand and an α -helix in the thumb. Furthermore, the peak heights for thumb residues 260-335 are larger than those for palm, fingers, and 8 kDa residues, which is consistent with the fact that the thumb subdomain has the largest movement in the simulation. Generally, the peak heights for all residues over the second half of the simulation are lower than those over the entire simulation and over the first half. The experimental fluctuations are larger except in three regions (discussed above), which are related to the enzyme's large motion. Differences are expected also given that the simulated protein is free in aqueous solution and the experimental protein is relatively restricted in a crystal-line lattice.

TMD simulation

The TMD simulations are performed by introducing an additional constraint force in a normal MD simulation.^{42,44-46} Our TMD simulations are carried out using academic version c28a2 of the program CHARMM (Chemistry at Harvard Molecular Mechanics),³⁴ modified to incorporate an energy restraint based on the RMS distance with respect to a reference target conformation. The functional form of the RMS restraint energy is as follows:

$$E_{\text{RMS}} = K[D_{\text{RMS}}(X(t), X^{\text{target}}) - d_0]^2$$

where K is a force constant (in $\text{kcal mol}^{-1} \text{\AA}^{-2}$), D_{RMS} represents the relative RMS distance for a selected set of atoms between the instantaneous conformation $X(t)$ and the reference X^{target} , and d_0 is an offset constant (in \AA). In the targeted dynamics, the selected set of atoms are free to rearrange spontaneously without experiencing any spurious forces while the RMSD relative to the target conformation remains constant. With the decreasing of d_0 (RMSD offset) as a function of the simulation time, the conformational change is driven from the initial to the final targeted conformation.

In our TMD simulations, the restrained trajectories are initiated from an equilibrated closed system and the target is the crystal open binary product structure. The conformational transition from closed to open state is driven by applying RMSD restraints with a total force constant of $2000 \text{ kcal mol}^{-1} \text{\AA}^{-2}$ to all heavy-atoms except those in residues 1-9, which are missing in the crystal closed structure. The offset parameter of d_0 is decreased from 3.6\AA by 0.18\AA every 5 ps of the simulation until it reaches zero deviation. The TMD simulations indicate that the pol β has approached the open state before Arg258 rotates.

Geometric characteristics

The radius of gyration and RMSD of C^α atomic positions in the dynamic structures with respect to the crystal, closed ternary complex and open binary product complex were monitored as a function of time. The radius of gyration is defined as:

$$R_g(t) = \left[\frac{1}{N_{C^\alpha}} \sum_{i=1}^{N_{C^\alpha}} (r_i(t) - r_{CG}(t))^2 \right]^{1/2}$$

where $r_i(t)$ and $r_{CG}(t)$ are the coordinates of atom i and of the center of geometry of the polymerase, respectively, at time t . N_{C^α} is the total number of the C^α atoms involved. The RMS deviation is defined as:

$$d_{\text{rms}}(t) = \left[\frac{1}{N_{C^\alpha}} \sum_{i=1}^{N_{C^\alpha}} (r_i(t) - r_i^{\text{CRY}})^2 \right]^{1/2}$$

where r_i^{CRY} is the coordinate of atom i in the crystal structures, after a least-squares fit superposition with dynamic structure at time t .

Arrhenius equation

The Arrhenius equation is used to estimate the activation barriers and time-frames for the Phe272 ring flip and Arg258 rotation at 300 K, based on four high-temperature molecular dynamics simulations. According to

the Arrhenius equation,^{82,83} the temperature-dependence of a chemical reaction's specific rate constant, k' , is expressed as:

$$\ln k' = \frac{E_a}{RT} + \ln A$$

where E_a is the activation energy, and A is the frequency factor or pre-exponential factor. A linear Arrhenius plot of $\ln k'$ versus $1/T$ is obtained when E_a is temperature-independent. Because different vibrational modes are activated differently as a function of the temperature, activation barriers in complex biomolecular processes are not temperature-independent. However, for the purpose of making rough estimates (i.e. an order of magnitude), and comparing them with experiment, we employ Arrhenius plots, and assess whether E_a is relatively temperature-independent from the plot linearity.

Acknowledgments

We thank Dr Daniel Strahs for providing the program for analyzing the water and counterion density, Mr Xiaoliang Qian for providing the program for building the periodic boundary conditions of the face-centered cube,³¹ and both for helpful discussions. We thank Dr Joseph Krahn for helpful advice in the initial stages of the work. The work was supported by NSF grant ASC-9318159 and NIH grant R01 GM55164 to T.S., and NIH grants CA75449, CA28038 and DOE grant DE-FG0290ER60931 to S.B. Computations were supported by National Computational Science Alliance under MCA99S021N and utilized the NCSA SGI/CRAY Origin2000. T.S. is an investigator of the Howard Hughes Medical Institute.

References

- Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. & Kraut, J. (1994). Structures of ternary complexes of rat DNA polymerase β , a DNA template-primer, and ddCTP. *Science*, **264**, 1891-1903.
- Sawaya, M. R., Prasad, R., Wilson, S. H., Kraut, J. & Pelletier, H. (1997). Crystal structures of human DNA polymerase β complexed with gapped and nicked DNA: evidence for an induced fit mechanism. *Biochemistry*, **36**, 11205-11215.
- Li, Y., Korolev, S. & 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**, 7514-7525.
- Kiefer, J. R., Mao, C., Braman, J. C. & Beese, L. S. (1998). Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal. *Nature*, **391**, 302-305.
- Ding, J., Das, K., Hsiou, Y., Sarafianos, S. G., Clark, A. D., Jr, Jacobo-Molina, A. *et al.* (1998). Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 Å resolution. *J. Mol. Biol.* **284**, 1095-1111.
- Doublé, S. & Ellenberger, T. (1998). The mechanism of action of T7 DNA polymerase. *Curr. Opin. Struct. Biol.* **8**, 704-712.
- Huang, H., Chopra, R., Verdine, G. L. & Harrison, S. C. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science*, **282**, 1669-1675.
- Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. & Steitz, T. A. (1985). Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature*, **313**, 762-766.
- Steitz, T. A. (1999). DNA polymerases: structural diversity and common mechanisms. *J. Biol. Chem.* **274**, 17395-17398.
- Beard, W. A. & Wilson, S. H. (2000). Structural design of a eukaryotic DNA repair polymerase: DNA polymerase β . *Mutat. Res.* **460**, 231-244.
- Steitz, T. A., Smerdon, S. J., Jäger, J. & Joyce, C. M. (1994). A unified polymerase mechanism for non-homologous DNA and RNA polymerases. *Science*, **266**, 2022-2025.
- Kuchta, R. D., Benkovic, P. & Benkovic, S. J. (1988). Kinetic mechanism whereby DNA polymerase I (Klenow) replicates DNA with high fidelity. *Biochemistry*, **27**, 6716-6725.
- Dahlberg, M. E. & Benkovic, S. J. (1991). Kinetic mechanism of DNA polymerase I (Klenow fragment): identification of a second conformational change and evaluation of the internal equilibrium constant. *Biochemistry*, **30**, 4835-4843.
- Patel, S. S., Wong, I. & Johnson, K. A. (1991). Pre-steady-state kinetic analysis of processive DNA replication inducing complete characterization of an exonuclease-deficient mutant. *Biochemistry*, **30**, 511-525.
- Kati, W. M., Johnson, K. A., Jerva, L. F. & Anderson, K. S. (1992). Mechanism and fidelity of HIV reverse transcriptase. *J. Biol. Chem.* **267**, 25988-25997.
- Frey, M. W., Sowers, L. C., Millar, D. P. & Benkovic, S. J. (1995). The nucleotide analog 2-aminopurine as a spectroscopic probe of nucleotide incorporation by the Klenow fragment of *Escherichia coli* polymerase I and bacteriophage T4 DNA polymerase. *Biochemistry*, **34**, 9185-9192.
- Werneburg, B. G., Ahn, J., Zhong, X., Hondal, R. J., Kraynov, V. S. & Tsai, M.-D. (1996). DNA polymerase β : pre-steady-state kinetic analysis and roles of arginine-283 in catalysis and fidelity. *Biochemistry*, **35**, 7041-7050.
- Vande Berg, B. J., Beard, W. A. & Wilson, S. H. (2001). DNA structure and aspartate 276 influence nucleotide binding to human DNA polymerase β . *J. Biol. Chem.* **276**, 3408-3416.
- Ahn, J., Werneburg, B. G. & Tsai, M.-D. (1997). DNA polymerase β : structure-fidelity relationship from pre-steady-state kinetic analyses of all possible correct and incorrect base pairs for wild type and R283A mutant. *Biochemistry*, **36**, 1100-1107.
- Kraynov, V. S., Werneburg, B. G., Zhong, X., Lee, H., Ahn, J. & Tsai, M.-D. (1997). DNA polymerase β : analysis of the contributions of tyrosine-271 and asparagine-279 to substrate specificity and fidelity of DNA replication by pre-steady-state kinetics. *Biochem. J.* **323**, 103-111.
- Ahn, J., Kraynov, V. S., Zhong, X., Werneburg, B. G. & Tsai, M.-D. (1998). DNA polymerase β : effects of gapped DNA substrates on dNTP specificity, fidelity, processivity and conformational changes. *Biochem. J.* **331**, 79-87.

22. Shah, A. M., Li, S.-X., Anderson, K. S. & Sweasy, J. B. (2001). Y265H mutator mutant of DNA polymerase β . *J. Biol. Chem.* **276**, 10824-10831.
23. Arndt, J. W., Gong, W., Zhong, X., Showalter, A. K., Liu, J., Dunlap, C. A. *et al.* (2001). Insight into the catalytic mechanism of DNA polymerase β : structures of intermediate complexes. *Biochemistry*, **40**, 5368-5375.
24. Kumar, A., Widen, S. G., Williams, K. R., Kedar, P., Karpel, R. L. & Wilson, S. H. (1990). Studies of the domain structure of mammalian DNA polymerase β : identification of a discrete template binding domain. *J. Biol. Chem.* **265**, 2124-2131.
25. Beard, W. A. & Wilson, S. H. (1998). Structural insights into DNA polymerase β fidelity: hold tight if you want it right. *Chem. Biol.* **5**, R7-R13.
26. Kumar, A., Abbotts, J., Karawya, E. M. & Wilson, S. H. (1990). Identification and properties of the catalytic domain of mammalian DNA polymerase β . *Biochemistry*, **29**, 7156-7159.
27. Matsumoto, Y. & Kim, K. (1995). Excision of deoxyribose phosphate residues by DNA polymerase β during DNA repair. *Science*, **269**, 699-702.
28. Prasad, R., Beard, W. A., Strauss, P. R. & Wilson, S. H. (1998). Human DNA polymerase β deoxyribose phosphate lyase. Substrate specificity and catalytic mechanism. *J. Biol. Chem.* **273**, 15263-15270.
29. Schlick, T. (1999). Computational molecular biophysics today: a confluence of methodological advances and complex biomolecular applications. *J. Comput. Phys.* **151**, 1-8.
30. Mezei, M. (1997). Optimal position of solute for simulations. *J. Comput. Chem.* **18**, 812-815.
31. Qian, X., Strahs, D. & Schlick, T. (2001). A new program for optimizing periodic boundary models of solvated biomolecules (PBCAID). *J. Comput. Chem.* **22**, 1843-1850.
32. Klapper, I., Hagstrom, R., Fine, R., Sharp, K. & Honig, B. (1986). Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: effects of ion strength and amino-acid modification. *Proteins: Struct. Funct. Genet.* **1**, 47-59.
33. Gilson, M. K., Sharp, K. & Honig, B. H. (1987). Calculating the electrostatic potential of molecules in solution: method and error assessment. *J. Comput. Chem.* **9**, 327-335.
34. Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187-217.
35. MacKerell, A. D., Jr & Banavali, N. K. (2000). 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.
36. Schlick, T., Barth, E. & Mandziuk, M. (1997). Biomolecular dynamics at long timesteps: bridging the timescale gap between simulation and experimentation. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 181-222.
37. Barth, E. & Schlick, T. (1998). Overcoming stability limitation in biomolecular dynamics. I. Combining force splitting via extrapolation with Langevin dynamics in LN. *J. Chem. Phys.* **109**, 1617-1632.
38. Barth, E. & Schlick, T. (1998). Extrapolation versus impulse in multiple-timestepping schemes. II. linear analysis and applications to Newtonian and Langevin dynamics. *J. Chem. Phys.* **109**, 1633-1642.
39. Wagner, G., Brühwiler, D. & Wüthrich, K. (1987). Reinvestigation of the aromatic side-chains in the basic pancreatic trypsin inhibitor by heteronuclear two-dimensional nuclear magnetic resonance. *J. Mol. Biol.* **196**, 227-231.
40. Kuchta, R. D., Mizrahi, V., Benkovic, P. A., Johnson, K. A. & Benkovic, S. J. (1987). Kinetic mechanism of DNA polymerase I (Klenow). *Biochemistry*, **26**, 8410-8417.
41. Elber, R. (1990). Calculation of the potential of mean force using molecular dynamics with linear constraints: an application to a conformational transition in a solvated dipeptide. *J. Chem. Phys.* **93**, 4312-4321.
42. Schlitter, J., Engels, M., Krüger, P., Jacoby, E. & Wollmer, A. (1993). Targeted molecular dynamics simulations of conformational change - application to the T \leftrightarrow R transition in insulin. *Mol. Sim.* **10**, 291-309.
43. Schlitter, J., Engels, M. & Krüger, P. (1994). Targeted molecular dynamics: a new approach for searching pathways of conformational transitions. *J. Mol. Graph.* **12**, 84-89.
44. Ferrara, P., Apostolakis, J. & Caflisch, A. (2000). Computer simulations of protein folding by targeted molecular dynamics. *Proteins: Struct. Funct. Genet.* **39**, 252-260.
45. Krüger, P., Verheyden, S., Declerck, P. J. & Engelborghs, Y. (2001). Extending the capabilities of targeted molecular dynamics: simulation of a large conformational transition in plasminogen activator inhibitor 1. *Protein Sci.* **10**, 798-808.
46. Young, M. A., Gonfloni, S., Superti-Furga, G., Roux, B. & Kuriyan, J. (2001). Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Cell*, **105**, 115-126.
47. Johnson, K. A. (1993). Conformational coupling in DNA polymerase fidelity. *Annu. Rev. Biochem.* **62**, 685-713.
48. Zhong, X., Patel, S. S., Werneburg, B. G. & Tsai, M.-D. (1997). DNA polymerase β : multiple conformational changes in the mechanism of catalysis. *Biochemistry*, **36**, 11891-11900.
49. Suo, Z. & Johnson, K. A. (1998). Selective inhibition of HIV-1 reverse transcriptase by an antiviral inhibitor, (R)-9-(2-phosphonylmethoxypropyl)adenine. *J. Biol. Chem.* **273**, 27250-27258.
50. Beese, L. S. & Steitz, T. A. (1991). Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.* **9**, 25-33.
51. Steitz, T. A. (1993). DNA- and RNA-dependent DNA polymerases. *Curr. Opin. Struct. Biol.* **3**, 31-38.
52. Steitz, T. A. (1998). A mechanism for all polymerases. *Nature*, **391**, 231-232.
53. Beard, W. A., Osheroff, W. P., Prasad, R., Sawaya, M. R., Jaju, M., Wood, T. G. *et al.* (1996). Enzyme-DNA interactions required for efficient nucleotide incorporation and discrimination in human DNA polymerase β . *J. Biol. Chem.* **271**, 12141-12144.
54. Lu, X. J., Shakked, Z. & Olson, W. K. (2000). A-form conformational motifs in ligand-bound DNA structures. *J. Mol. Biol.* **300**, 819-840.
55. Schneider, B., Neidle, S. & Berman, H. M. (1997). Conformations of the sugar-phosphate backbone in helical DNA crystal structures. *Biopolymers*, **42**, 113-124.

56. Eom, S. H., Wang, J. & Steitz, T. A. (1996). Structure of Taq polymerase with DNA at the polymerase active site. *Nature*, **382**, 278-281.
57. Doublé, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature*, **391**, 251-258.
58. Doublé, S., Sawaya, M. R. & Ellenberger, T. (1999). An open and closed case for all polymerases. *Structure*, **7**, R31-R35.
59. Date, T., Yamamoto, S., Tanihara, K., Nishimoto, Y. & Matsukage, A. (1991). Aspartic acid residues at positions 190 and 192 of rat DNA polymerase β are involved in primer binding. *Biochemistry*, **30**, 5286-5292.
60. Li, S. X., Vaccaro, J. A. & Sweasy, J. B. (1999). Involvement of phenylalanine 272 of DNA polymerase beta in discriminating between correct and incorrect deoxynucleoside triphosphates. *Biochemistry*, **38**, 4800-4808.
61. Kraynov, V. S., Showalter, A. K., Liu, J., Zhong, X. & Tsai, M.-D. (2000). DNA polymerase β : contributions of template-positioning and dNTP triphosphate-binding residues to catalysis and fidelity. *Biochemistry*, **39**, 16008-16015.
62. Osheroff, W. P., Beard, W. A., Yin, S., Wilson, S. H. & Kunkel, T. A. (2000). Minor groove interactions at the DNA polymerase β active site modulate single-base deletion error rates. *J. Biol. Chem.* **275**, 28033-28038.
63. Li, Y., Mitaxov, V. & Waksman, G. (1999). Structure-based design of Taq DNA polymerases with improved properties of dideoxynucleotide incorporation. *Proc. Natl Acad. Sci. USA*, **96**, 9491-9496.
64. Li, Y. & Waksman, G. (2001). Crystal structures of a ddATP-, ddTTP-, ddCTP-, and ddGTP-trapped ternary complex of KlenTaq1: insights into nucleotide incorporation and selectivity. *Protein Sci.* **10**, 1225-1233.
65. Pelletier, H., Sawaya, M. R., Wolfle, W., Wilson, S. H. & Kraut, J. (1996). A structural basis for metal ion mutagenicity and nucleotide selectivity in human DNA polymerase β . *Biochemistry*, **35**, 12762-12777.
66. Zhong, X., Patel, S. S. & Tsai, M.-D. (1998). DNA polymerase β . 5. Dissecting the functional roles of the two metal ions with Cr(III)dTTP. *J. Am. Chem. Soc.* **120**, 235-236.
67. Menge, K. L., Hostomsky, Z., Nodes, B. R., Hudson, G. O., Rahmati, S., Moomaw, E. W. *et al.* (1995). Structure-function analysis of the mammalian DNA polymerase β active site: role of aspartic acid 256, arginine 254 and arginine 258 in nucleotidyl transfer. *Biochemistry*, **34**, 15934-15942.
68. Matsuda, T., Bebenek, K., Masutani, C., Hanaoka, F. & Kunkel, T. A. (2000). Low fidelity DNA synthesis by human DNA polymerase- η . *Nature*, **404**, 1011-1013.
69. Johnson, R. E., Washington, M. T., Haracska, L., Prakash, S. & Prakash, L. (2000). Eukaryotic polymerase ι and ζ act sequentially to bypass DNA lesions. *Nature*, **406**, 1015-1019.
70. Tissier, A., McDonald, J. P., Frank, E. G. & Woodgate, R. (2000). Pol ι , a remarkably error-prone human DNA polymerase. *Genes Dev.* **14**, 1642-1650.
71. Johnson, R. E., Washington, M. T., Prakash, S. & Prakash, L. (2000). Fidelity of human DNA polymerase η^* . *J. Biol. Chem.* **275**, 7447-7450.
72. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al.* (2000). The protein data bank. *Nucleic Acids Res.* **28**, 235-242.
73. Brünger, A. T. & Karplus, M. (1988). Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison. *Proteins: Struct. Funct. Genet.* **4**, 148-156.
74. Wu, X., Shapiro, R. & Broyde, S. (1999). Conformational analysis of the major DNA adduct derived from the food mutagen 2-amino-3-methylimidazo[4,5-f]quinoline. *Chem. Res. Toxicol.* **12**, 895-905.
75. Schlick, T. (1992). Optimization methods in computational chemistry. In *Reviews in Computational Chemistry* (Lipkowitz, K. B. & Boyd, D. B., eds), vol. 3, pp. 1-71, VCH Publishers, New York, NY.
76. Schlick, T. (2001). Time-trimming tricks for dynamic simulations: splitting force updates to reduce computational work. *Structure*, **9**, R45-R53.
77. Sandu, A. & Schlick, T. (1999). Masking resonance artifacts in force-splitting methods for biomolecular simulations by extrapolative langevin dynamics. *J. Comput. Phys.* **151**, 74-113.
78. Strahs, D. & Schlick, T. (2000). A-tract bending: insights into experimental structures by computational models. *J. Mol. Biol.* **301**, 643-663.
79. van Gunsteren, W. F., Berendsen, H. J., Hermans, J., Hol, W. G. & Postma, J. P. (1983). Computer simulation of the dynamics of hydrated protein crystals and its comparison with X-ray data. *Proc. Natl. Acad. Sci. USA*, **80**, 4315-4319.
80. Levy, R. M., Sheridan, R. P., Keepers, J. W., Dubey, G. S., Swaminathan, S. & Karplus, M. (1985). Molecular dynamics of myoglobin at 298 K. Results from a 300-ps computer simulation. *Biophys. J.* **48**, 509-518.
81. Arnold, G. E. & Ornstein, R. L. (1997). Protein hinge binding as seen in molecular dynamics simulations of native and M61 mutant T4 lysozymes. *Biopolymers*, **41**, 533-544.
82. Moore, W. J. (1972). *Physical Chemistry*, Prentice-Hall, Inc., Englewood Cliffs, NJ.
83. Makinen, M. W. & Fink, A. L. (1977). Reactivity and cryoenzymology of enzymes in the crystalline state. *Annu. Rev. Biophys. Bioeng.* **6**, 301-343.
84. Leslie, A. G., Arnott, S., Chandrasekaran, R. & Ratliff, R. L. (1980). Polymorphism of DNA double helices. *J. Mol. Biol.* **143**, 49-72.
85. Saenger, W. (1984). *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
86. Altona, C. & Sundaralingam, M. (1972). Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation. *J. Am. Chem. Soc.* **94**, 8205-8212.
87. Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kitzing, E. V. *et al.* (1989). Definitions and nomenclature of nucleic acid structure parameters. *EMBO J.* **8**, 1-4.
88. Wong, I., Patel, S. S. & Johnson, K. A. (1991). An induced-fit mechanism for DNA replication fidelity: direct measurement by single-turnover kinetics. *Biochemistry*, **30**, 526-537.
89. Capson, T. L., Peliska, J. A., Kaboord, B. F., Frey, M. W., Lively, C., Dahlberg, M. & Benkovic, S. J. (1992). Kinetic characterization of the polymerase and exonuclease activities of the gene 43 protein of bacteriophage T4. *Biochemistry*, **31**, 10984-10994.
90. Sou, Z. C. & Johnson, K. A. (1997). Effect of RNA secondary structure on the kinetics of DNA syn-

thesis catalyzed by HIV-1 reverse transcriptase. *Biochemistry*, **36**, 12459-12467.

91. Kerr, S. G. & Anderson, K. S. (1997). RNA dependent DNA replication fidelity of HIV-1 reverse transcriptase: evidence of discrimination between DNA and RNA substrates. *Biochemistry*, **36**, 14056-14063.

Edited by B. Honig

(Received 13 August 2001; received in revised form 18 December 2001; accepted 22 January 2002)



<http://www.academicpress.com/jmb>

Supplementary Material is available on IDEAL