Conformational Transition Pathway of Polymerase β/DNA upon Binding Correct Incoming Substrate

Karunesh Arora and Tamar Schlick*

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

Received: November 24, 2004; In Final Form: January 13, 2005

The closing conformational transition of wild-type polymerase β bound to DNA template/primer before the chemical step (nucleotidyl transfer reaction) is simulated using the stochastic difference equation (in length version, "SDEL") algorithm that approximates long-time dynamics. The order of the events and the intermediate states during polβ's closing pathway are identified and compared to a separate study of polβ using transition path sampling (TPS) (Radhakrishnan, R.; Schlick, T. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 5970−5975). Results highlight the cooperative and subtle conformational changes in the polβ active site upon binding the correct substrate that may help explain DNA replication and repair fidelity. These changes involve key residues that differentiate the open from the closed conformation (Asp192, Arg258, Phe272), as well as residues contacting the DNA template/primer strand near the active site (Tyr271, Arg283, Thr292, Tyr296) and residues contacting the β and γ phosphates of the incoming nucleotide (Ser180, Arg183, Gly189). This study compliments experimental observations by providing detailed atomistic views of the intermediates along the polymerase closing pathway and by suggesting additional key residues that regulate events prior to or during the chemical reaction. We also show general agreement between two sampling methods (the stochastic difference equation and transition path sampling) and identify methodological challenges involved in the former method relevant to large-scale biomolecular applications. Specifically, SDEL is very quick relative to TPS for obtaining an approximate path of medium resolution and providing qualitative information on the sequence of events; however, associated free energies are likely very costly to obtain because this will require both successful further refinement of the path segments close to the bottlenecks and large computational time.

Introduction

Capturing enzyme motions on the time scale of milliseconds to microseconds is one of the greatest challenges in computational biology.[1,2] Although molecular dynamics (MD) and Brownian dynamics (BD) simulations can reveal detailed insights into structure and function of biomolecular systems,[3−5] routine applications of standard dynamics simulations are limited to the time scales of a few nanoseconds. This limitation has spurred interest in developing a rich menu of alternative methodologies that can tackle biomolecular motions on longer time scales.

If the starting and final states are known in advance, biomolecular systems can be treated with microscopic, coupled quantum mechanical/molecular mechanical (QM/MM) methods[6] (e.g., EVB[7,8]), or "steered" (see ref 3 and references therein)—subject to time-dependent external forces along certain degrees of freedom or along local free energy gradient—to study chemical reactions and study folding/unfolding events, or generate insights into disallowable configurational states and common pathways. Activated and long-time processes can also be studied by using high-temperature simulations,[9] aggregate dynamics,[10] enhanced sampling by MC/MD,[11−14] replica dynamics,[15,16] free-energy calculations,[17−19] path sampling[20−22] or the stochastic path approach.[23] All these methodologies are based on different approximations and address varied problems (see, for example, ref 24).

The stochastic path algorithm in length (SDEL for stochastic difference equation) of Elber and co-workers provides a framework for obtaining numerically stable solutions at arbitrary time steps and approximate dynamics trajectories on extended time scales.[23] SDEL is based on boundary value formulation and is useful in investigating processes where reactant and product states are known. An advantage of SDEL is that determination of "order parameters" is not required. SDEL and its variants have been used to follow various biological processes on time scales ranging from nanoseconds to microseconds; examples include cytochrome C's folding kinetics, DNA's sugar repuckering pathway,[25] protein A's folding, cyclodextrin glycosyltransferase's mechanism, and hemoglobin's R to T transitions (see citations in ref 24). In this work, we use SDEL to study a biologically important system of a DNA polymerase, namely the conformational transition pathway of mammalian DNA polymerase β (polβ) complexed with DNA template/primer and correct incoming substrate (dCTP).

DNA polymerases are important enzymes associated with DNA replication and repair. The genome template in human cells frequently becomes damaged by exposure to harmful radiation (e.g., UV), oxidation, etc.[26,27] and can be repaired by the cellular machinery that involves many enzymes, including DNA polymerases. Mammalian DNA polβ functions primarily in Base Excision Repair,[28] which involves restoring the damaged

* To whom correspondence should be addressed. Phone: 212-998-3116. E-mail: schlick@nyu.edu. Fax: 212-995-4152.

10.1021/jp0446377 CCC: $30.25
Published on Web 03/01/2005

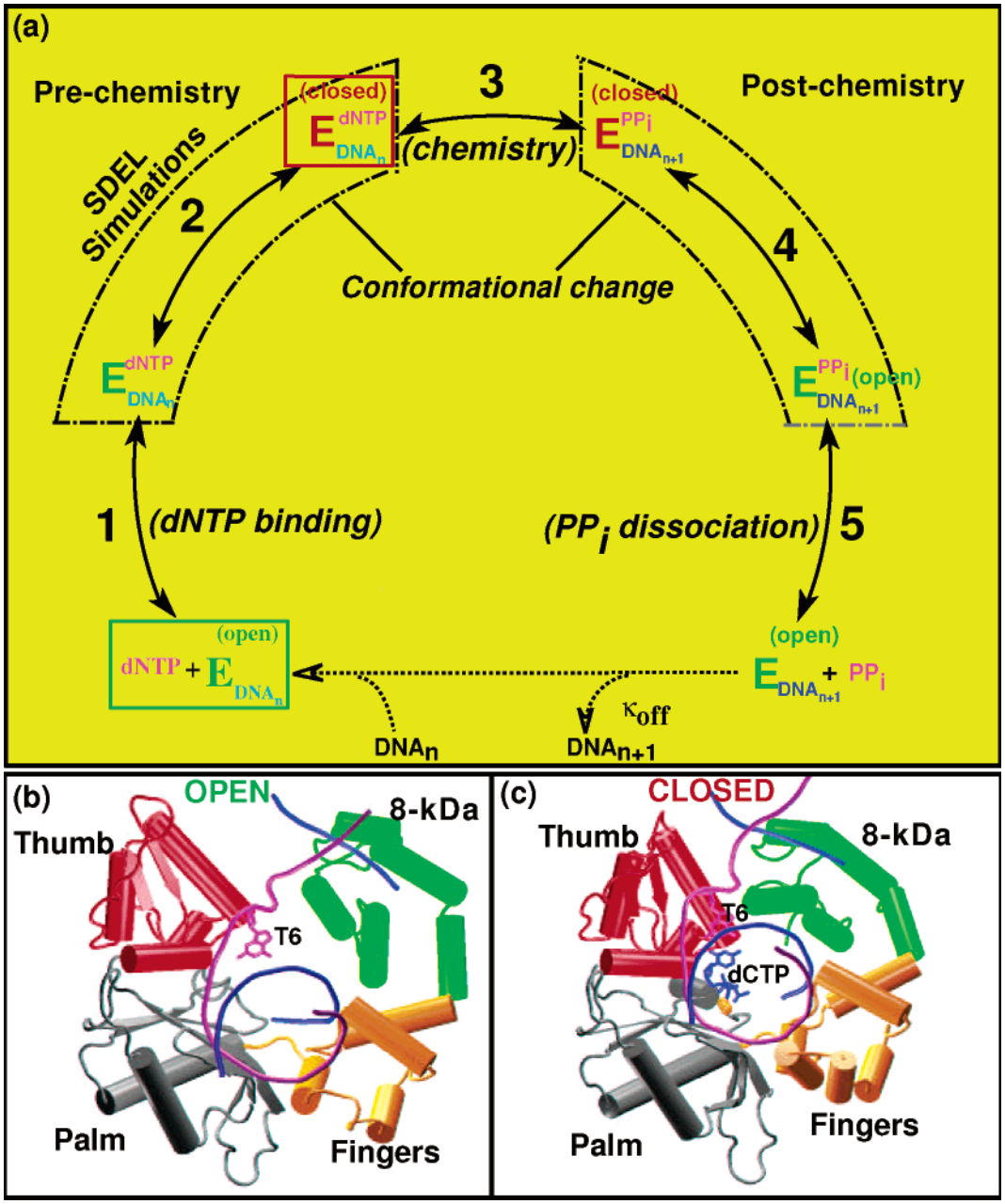

Figure 1. General pathway for nucleotide insertion by DNA polβ (a) and corresponding (b) crystal open (bottom left) and closed (bottom right) conformations of polβ/DNA complex. E: DNA polymerase; dNTP:2′-deoxyribonucleoside 5′-triphosphate. PP_i: pyrophosphate. DNA_n/DNA_{n+1}: DNAbefore/after nucleotide incorporation to DNA primer. T6 is the template residue (G) corresponding to the incoming nucleotide (dCTP).

base in a DNA strand that sustained damage with a correct nucleotide using the undamaged strand's base as a Watson−Crick template.[29]

Structurally, polβ is composed of only two domains, an N-terminal 8 kDa region that exhibits deoxyribose phosphate lyase activity, and a 31 kDa C-terminal domain that possesses nucleotidyl transfer activity. The 31 kDa domain resembles structurally characterized polymerases to date, containing finger, palm, and thumb subdomains.[30] Studies on polβ, therefore, can serve as a model for other DNA polymerases. The relatively small size of polβ (335 protein residues and 16 DNA base pairs) also renders it attractive for computational studies.

Polβ has been crystallized[31] in complexes representing three intermediates of the opening/closing transition: *open binary complex*, containing polβ bound to a DNA substrate with single nucleotide gap; *closed ternary complex*, containing pol β·gap·ddCTP (pol β bound to gapped DNA as well as 2′,3′-dideoxyribocytidine 5′-triphosphate (ddCTP)); and *open binary product complex*, polβ·nick (polβ bound to nicked DNA). Recently, DNA polβ structures with DNA mismatches (A:C and T:C) were also solved and found to be in partially open conformations.[32] Figures1b,c illustrate the significant conformational difference between open and closed forms of the polβ with the matched base pair at the template/primer terminus (G:C).

Based on kinetic and structural studies of several polymerases[33−45] complexed with primer/template DNA, a common nucleotide insertion pathway has been characterized for some polymerases (e.g., HIV-1 reverse transcriptase, phage T4 DNA polymerase, *Escherichia coli* DNA polymerase I Klenow fragment) that undergo transitions between open and closed forms, like polβ (Figure 1a). At step 1, following DNA binding, the DNA polymerase incorporates a 2′-deoxyribonucleoside 5′-triphosphate (dNTP) to form an open substrate complex; this

complex undergoes a conformational change to align catalytic groups and form a closed ternary complex at step 2; the nucleotidyl transfer reaction then follows: 3′-OH of the primer strand attacks the P_α of the dNTP to extend the primer strand and form the ternary product complex (step 3); this complex then undergoes a reverse conformational change (step 4) back to the open enzyme form. This transition is followed by dissociation of pyrophosphate (PP_i) (step 5), after which the DNA synthesis/repair cycle can begin anew.

The conformational rearrangements involved in steps 2 and 4 (see Figure 1a) are believed to be key for monitoring DNA synthesis fidelity.[31] That is, *binding of the correct nucleotide induces the first conformational change (step 2) whereas binding of an incorrect nucleotide (e.g., A rather than C opposite a G) may alter or inhibit the conformational transition*. This "induced-fit" mechanism[46,47] was thus proposed to explain the polymerase fidelity in selecting the correct dNTP.[31] This mechanism suggests that the conformational changes triggered by the binding of the correct nucleotide will align the catalytic groups as needed for catalysis, whereas the incorrect substrate will somehow interfere with this process. Our recent studies using standard molecular dynamics also provide *in silico* support for the presence of substrate-induced conformational change in polβ's closing.[48]

Nevertheless, whether the conformational change or the chemistry step is rate limiting is not known for polβ. Recent studies in our lab (L. Yang, K. Arora., R. Radhakrishnan, and T. Schlick, unpublished) suggest the latter but this still does not distract from the importance of conformational changes which likely steer the enzyme correctly to the chemical reaction. Indeed, Post and Ray[49] argue that an induced-fit mechanism can alter enzyme specificity even if conformational changes are not kinetically dominant if the correct and incorrect nucleotide incorporation involve nonidentical interactions with the active site in the substrate-dependent transition state. Ongoing experimental and theoretical studies will undoubtedly resolve these important questions.

Given the role of polβ in maintaining genome integrity and the relation between polβ's malfunction and various diseases (e.g., cancer, premature aging),[50−52] polβ has attracted many experimental (see ref 53 and references therein) and theoretical investigations.[8,54,55−57] Our prior modeling work of the polβ dynamics before and after the chemical step of nucleotidyl transfer reaction using MD[48,58−61] and transition path sampling (TPS)[22] employed the CHARMM force field.[62] Importantly, these works suggested that subdomain motions per se are rapid but that subtle side chain residue motions that involve key protein residues such as 272, 258, 271, 283, and 192 and the movements of the magnesium ions are slow and possibly rate limiting the conformational changes; distortions at the active sites for mismatches[59] and mutant studies[60] and dissected compensatory molecular interactions also provided atomic level mechanisms to interpret kinetic data. For example, Tyr271 of polβ is observed to hydrogen bond to the minor groove edge of the primer terminus in the closed active ternary complex.[31] Removing this hydrogen bond through site-directed mutagenesis resulted in an increase in the binding affinity for the incoming nucleotide;[63] this result was not expected or easily interpreted with available crystallographic structures of polβ. However, our modeling revealed a clear explanation to the puzzling observation: compensatory backbone motions allow the adjacent Phe272 to interact with the incoming nucleotide.[60] As another example, our modeling of the R283A mutant[60] confirms the suggestion that the equilibrium of the open-closed subdomain transition is shifted toward the open conformation and argues against the idea that a larger (nascent base pair) binding pocket is provided by the alanine substitution in the closed conformation. Because correct nucleotide insertion is specifically affected with this mutant (decreased 30 000-fold) and incorrect insertion is not,[64] our modeling also suggests that incorrect insertion may occur from an open conformation.

Because of inherent time scale limitations, initial studies[48,58−61] were performed on constructed intermediate states of polβ based on available crystal structures in open and closed conformations. Here, like our TPS study,[22] we generate approximate paths connecting the crystallographic anchors without resorting to constructed intermediates. Specifically, we apply SDEL based on the AMBER force field,[65] to investigate in atomic detail the closing conformational transition pathway of polβ in the presence of correct incoming substrate before the nucleotidyl transfer chemical reaction. Solvent effects in SDEL simulations are considered implicitly using the Gibbs−Born model.[66−68] Our use of AMBER instead of CHARMM as in other studies validates the existence of the dominant pathway identified in ref 22. Already, our implementation of SDEL with the AMBER nucleic acid force field was tested and validated for an application to deoxyadenosine, which generated the pseudorotation pathway[25] between C2′-endo and C3′-endo sugar conformations. In this work, the order of events along polβ's closing conformational pathway is delineated, and the slow reaction coordinates along the path are determined. The minima associated with each transition state region are computed and compared to those obtained by transition path sampling (TPS).[22] The results obtained from SDEL simulations agree well qualitatively with those from TPS, despite each method's very different approximations, and reveal important intermediates and bottlenecks along the closing pathway associated with motions of key active site residues.

Methods

SDEL. SDEL is based on the formulation of molecular dynamics as a boundary value problem.[69] From specified initial (Y_1) and final (Y_2) configurations, mass-weighted coordinates (Y) are obtained as a function of the trajectory length l. This is accomplished by constructing trajectory that renders the action (S) stationary, where S is

$$S = \int_{Y_1}^{Y_2} \sqrt{2(E - U)}\, \mathrm{d}l \qquad (1)$$

Here, E is the total energy, U is the potential energy, and dl is a length element. The prescribed endpoints Y_1 and Y_2 in our application are coordinates of polβ's open and closed conformations, respectively. SDEL uses a discretized version of the action $S \cong \sum_i \sqrt{2(E-U)}\ \Delta l_{i,i+1}$ to define the classical trajectory. According to *the principle of least action*, a stationary point of S exists (though S need not be a minimum or maximum); a trajectory that makes S stationary also solves the classical equations of motion. Such a trajectory is computed by optimizing the gradient norm given by the following target function T_G:

$$T_G = \sum_i \left(\frac{\partial S/\partial Y_i}{\Delta l_{i,i+1}} \right)^2 \Delta l_{i,i+1} + \lambda \sum_i (\Delta l_{i,i+1} - \langle \Delta l \rangle)^2 \qquad (2)$$

where the variables Y_i are all configurations along the path and the parameter λ defines the coupling strength of a restraint that maintains the structures equally distributed along the trajectory

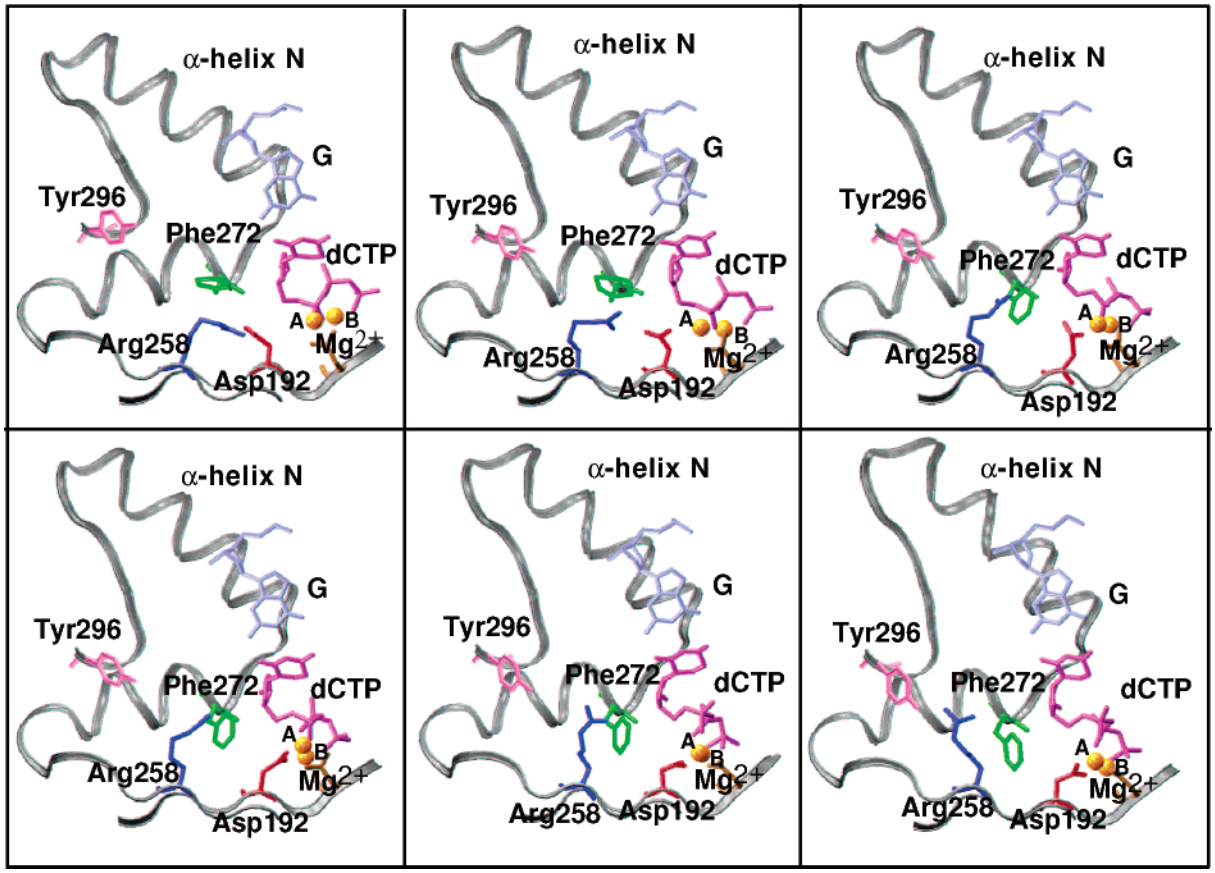

Figure 2. (Left to right) snapshots of the conformational transition path of the polβ/DNA complex upon binding the correct incoming nucleotide (dCTP) opposite a template G as generated by SDEL. Shown in the figure are α-helix N in gray (in the thumb subdomain), residues in the microenvironment of the incoming nucleotide (dCTP), magnesium ions (catalytic (Mg^{2+}(A)), and nucleotide binding (Mg^{2+}(B))), and the enzymes active site. A dynamic view of complete path is available on our website http://monod.biomath.nyu.edu/~arora.

($\langle \Delta l \rangle = 1/(N+1)\sum \Delta l_{i,i+1}$). For sufficiently small steps, $\Delta l_{i,i+1}$, the exact classical trajectory is recovered. The target function T_G is minimized subject to additional constraints to keep rigid-body translations and rotations constant.[70] Two major assumptions in the application of the SDEL method as implemented in the MOIL package[71] are the use of the implicit Born solvation model[66−68] and the filtering of high-frequency motions.[69] The former can be alleviated using greater computer time, but the filtering makes SDEL applicable to long-time, albeit approximate, trajectories.

Note that the trajectories are parametrized as a function of length and not time according to recent formulations.[69] The obtained SDEL trajectories cannot be described clearly as a function of time, but only the temporal order in which configurations are valid. An algorithm to determine the time scales of complex processes following predetermined milestones along the reaction coordinate was recently developed but has not been tested yet on complex biomolecular processes.[72] For a comparison of the formulations and their evaluation, see ref 23. An overview of SDEL and its computational efficiency are provided elsewhere.[73]

Model Building for Simulations. Our endpoints are derived from the open and closed conformations for matched base pair on the basis of the crystal open binary complex (PDB ID: 1BPX) and closed ternary complex (PDB ID: 1BPY). Hydrogen atoms were added to the crystallographic heavy atoms. The open complex was modified by positioning the incoming substrate dCTP and both ions (nucleotide and catalytic magnesium) in the active site by superimposing the palm subdomain of the open binary and ternary closed complex. The INSIGHTII package, version 2000 (Accelrys Inc., San Diego, CA), was used to retain the coordinates of protein residues and DNA sequences in place during these preparations. The hydroxyl group was also added to the 3′ terminus of the primer DNA strand.

The steric clashes in generated models were removed by subsequent energy minimization and equilibration. Solvation effects were included on the basis of the Gibbs−Born (GB) solvation model.[66−68] The GB model implemented in MOIL has been successfully applied to study various biological systems.[24,25,69,74] Simulations employed the parm99.dat version of the AMBER nucleic acid force field.[65] Our implementation of this force field in MOIL was tested previously on a nucleoside unit.[25] The MOIL package is available at *http://cbsu.tc.cornell.software/moil/moil.html*.

Generation of SDEL Trajectories. From the coordinates of the polβ/DNA/dCTP complex in the open and closed conformations, the minimum energy path was calculated and used as the initial guess for the SDEL trajectory. The minimum energy path was generated using the self-penalty walk (SPW) functional[70] with 100 grid points. SPW minimizes the energies in the structures along the path generated by a simple linear interpolation of Cartesian coordinates and provides a reasonable initial approximate path close to steepest descent path. The path was optimized for 2000 steps. At the end of these minimum energy path calculations, the RMS distance between sequential structures was on the order of 0.1−0.2 Å.

SPW calculations were further refined by the SDEL formalism. The total energy of the system, E, was estimated from standard initial-value molecular dynamics simulations that were equilibrated at room temperature. Both open and closed

conformations of polβ/DNA/dCTP complexes were used to estimate the energy E. The target function T_G (or the path) was optimized by using five cycles of 2000 simulated annealing steps. The value of the gradient of the target function, T_G, normalized to the number of degrees of freedom, was 30 (amu·kcal)/(mol/Å).

The annealed trajectory deviates from the minimum energy path and includes more oscillations in the minima. Furthermore, because the number of configurations is kept constant, at the end of the simulated annealing process, the distance between sequential structures increases to about 0.4−0.6 Å. Increasing the number of grid points from 100 to 200 did not reduce the sequential distance between structures significantly and made annealing difficult. Therefore, results discussed here are based on the trajectory with 100 grid points with intermediate step size between the steepest descent path and exact classical trajectories.[24] The resulting trajectory has kinetic energy included explicitly and approximates dynamical features of the real system compared to the steepest descent path with no inertial terms. The total root-mean-square deviation of the intermediate structures in the final minimized trajectory compared to the initial guess trajectory is of the order 0.45−0.7 Å. This is a significant change from the starting path and suggests that the path was not trapped in a nearby local minimum.

The calculation of SDEL trajectories was performed in parallel mode with the Message Passing Interface Library. Trajectories were computed on the LINUX cluster at Cornell University, using 20 CPU nodes of 600 MHz processing speed, for roughly 1 month for a single trajectory of 100 grid points.

Results

Orchestrated Order of Events During Polβ Closing. Snapshots of polβ's closing conformational transition path when complexed with template-matched incoming nucleotide are depicted in Figure 2. Shown in the figure is the evolution of polβ's:α-helixN (in the thumb subdomain), residues in the microenvironment of the incoming nucleotide (dCTP), and the enzymes active site. Note the significant change in the conformation of α-helixN and other residues along the pathway. Figure 3 illustrates the evolution of order parameters $\{\chi_i\}$ associated with each significant conformational event as a function of the path configuration 1−100 and indicates the sequence of these changes: the thumb partially closes, Asp192's side chain flip, Phe272 flips, and Arg258 rotates fully with complete thumb closure, resulting in formation of stable closed state. The choice of representing polβ's conformational change in terms of these variables was based on the analysis of available crystal structures of polβ and our prior investigations of polβ/DNA complex using dynamics simulations,[48,58] which suggest marked changes in the conformation of chosen variables during transition from open to closed state.

Thus we see that, upon binding the correct incoming nucleotide, the conformational changes that follow are cooperative, subtle, and organized. The partial thumb closure corresponds to the motion of α-helixN that repositions itself, with concomitant decrease in RMSD from 6.7 to 3.2 Å (order parameter χ_1 in Figure 3) with respect to the crystal closed conformation. This large thumb movement triggers the conformational change of the local side chains in the microenvironment of the incoming dCTP.

When the thumb is partially closed (thumb RMSD ≈3.2 Å), residues Asp192, Arg258 and Phe272 depart from their initial conformations and pass through various metastable states before reaching the final closed conformation. First, Asp192 completes its flip with change in dihedral ($C^{\gamma}-C^{\beta}-C^{\alpha}-C$) from ≈60° to 180° (order parameter χ_2 in Figure 3) and coordinates with both Mg^{2+} ions. In the open conformation, Asp192 is in an inactive state and engaged in salt bridge with Arg258. Following Asp192's flip, Phe272 flips (χ_3 in Figure 3), disrupting the salt bridge between Asp192 and Arg258 and vacating space for the complete Arg258 rotation (χ_4 in Figure 3). In the partially rotated state, Arg258 ($\chi_4 = 130°$) sterically clashes with the Phe272 ($\chi_3 = 50°$).

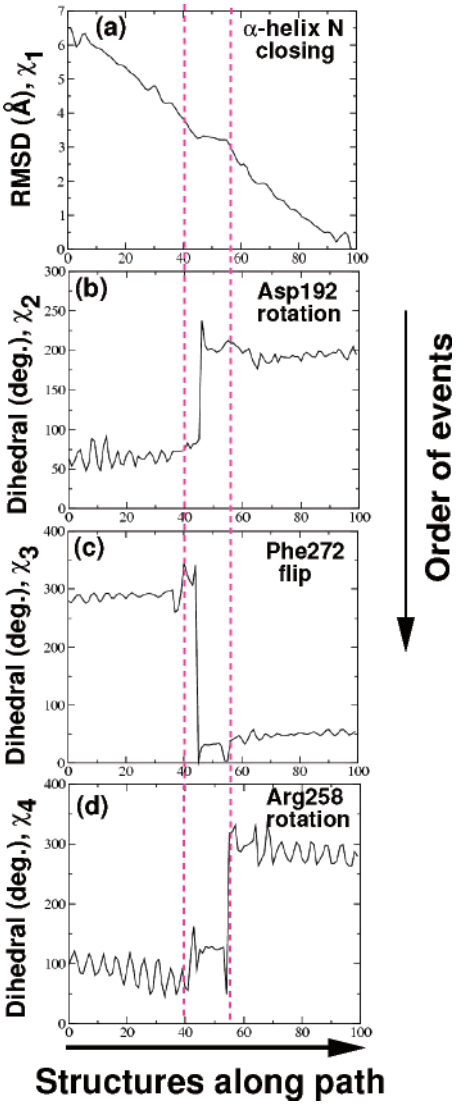

Figure 3. Order of events (top to bottom) of polβ's closing pathway depicted as a function of different order parameters evolving as the progress along the path of 100 steps (left to right): (a) RMSD of the α-helix N (thumb residues 275 to 295) with respect to the closed state (χ_1), (b) dihedral angle ($C^{\gamma}-C^{\beta}-C^{\alpha}-C$) characterizing the flip of Asp192 (χ_2), (c) dihedral angle ($C^{\delta 1}-C^{\gamma}-C^{\beta}-C^{\alpha}$) characterizing the flip of Phe272 (χ_3), and (d) dihedral angle ($C^{\gamma}-C^{\delta}-N^{\epsilon}-C^{\zeta}$) characterizing the rotation of Arg258 (χ_4).

Only following these crucial side chain motions in the microenvironment of the incoming nucleotide does the thumb close completely to stabilize the closed state and prepare it for the chemical step. This network of conformational changes provides several checks for polβ to select the correct dNTP preceding the primer elongation step. We hypothesize that complete thumb closing may not occur for an incorrect incoming nucleotide (dNTP), and side chain motions (e.g., Arg258 rotation) may follow alternative high-energy path, resulting in

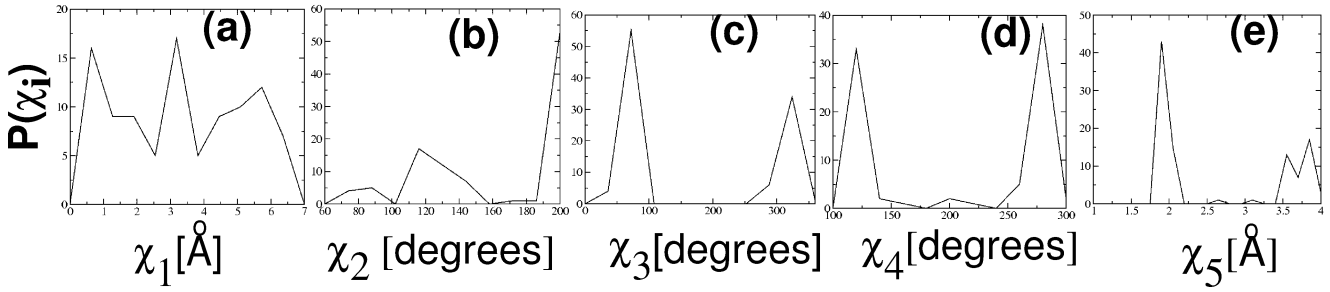

Figure 4. Probability distribution, $P(\chi_i)$, along the reaction coordinate of different order parameters (χ_i, $i = 1-5$) associated with conformational events along the pathway. (a) χ_1 is the RMSD of α-helix N (thumb residues 275−295) with respect to the closed state, (b) χ_2 is the dihedral angle characterizing the flip of Asp192, (c) χ_3 is the dihedral angle characterizing the flip to Phe272, (d) χ_4 is the dihedral characterizing the rotation of Arg258, and (e) χ_5 is the distance between the nucleotide binding Mg^{2+} ion (labeled as Mg^{2+}(B) in Figure 2) and the oxygen atom $O1_\alpha$ of dCTP. The multimodality of these distributions indicate the transitions involved.

reduced efficiency of incorrect nucleotide incorporation. Such evidence has already emerged from TPS.[75]

Overall, our observed sequence of events in polβ's closing agrees with our previous investigations and lends credence to the hypothesis that side chain rotations (including that of Arg258) may be rate limiting in the conformation sequence, not the overall pathway (Figure 1). These subtle residue motions are crucial for stabilizing the closed state before the nucleotidyl transfer reaction step. The simulations also highlight the multistep aspects of this conformational change rather than a simple subdomain motion, as suggested by the crystal coordinates alone.

To locate the different minima (metastable states) associated with order parameters, all the structures on the optimized path were independently minimized using Powell's conjugate gradient algorithm. The minimized structures were clustered together in the sequential order obtained from the trajectory. This trajectory, from which the kinetic energy has been removed, can yield significant insights into the system's dynamics and kinetic properties. This strategy of classifying potential energy minima resembles that used by Stillinger and Weber to study inherent structures in liquids.[76]

Figure 4 shows the probability distribution of different order parameters along the reaction coordinate obtained from the quenched trajectory. Qualitatively, these probability distribution plots can be compared to a coarse-grained potential of mean force along the reaction coordinate as reported independently by transition path sampling.[22] The minima associated with each order parameter in Figure 4 coincide approximately with the maxima in ref 22 (see Figure 10 (Appendix 1)). This agreement suggests that SDEL can capture approximate long-time dynamics of the system. The TPS studies employed a Monte Carlo sampling of thousands of short symplectic molecular dynamics trajectories to connect metastable basins;[20] free energies based on TPS were computed by the BOLAS algorithm.[77]

Thumb/Nucleic-Acid Contacts. Structural and kinetic analyses of polβ mutants involving residues at the active site suggest that selection of incoming dCTP is directed by the templating residue by an induced-fit mechanism.[31,78] However, the templating residue (G) is not properly aligned to base pair with the incoming nucleotide (dCTP) when the thumb is open. In Figure 5, we monitor the evolution of distances between residues in the thumb that interact with the DNA template residue and the two nucleic acid residues that follow on the template strand.

Initially, when the thumb is open, all distances are in the range >6 Å but then gradually decrease to reach the range of van der Waals radii when the thumb is closed. The decrease in distances corresponds to the movement of α-helix N in the thumb subdomain, with some distances reaching their final state when the thumb is in the intermediate half-closed conformation (RMSD ≈ 3.2 Å) (Figure 5). Other distances such as those involving Tyr271, Arg283, and Thr292 attain the values of the final conformation only when α-helix N is completely closed. The conformational change of the thumb aligns the templating residues into the active conformation, where it can base pair with incoming residue and anchor it to the polymerase. Arg283 forms part of the active site pocket and is particularly important for nucleotide discrimination, as suggested by mutagenesis experiments.[63] Namely, the site-specific mutation of Arg283 with sterically less demanding residue alanine results in marked decrease in the polymerase fidelity.[33,63] Thus, the increased interaction between Arg283 and the templating residue might signal the presence of a correct nucleotide and trigger a cascade of conformational changes in the enzyme's active site pocket. Tyr271 is another particularly interesting signaling residue; it interacts with the partner template base (see Figure 5, top right) in the open thumb conformation, and with the primer base in the closed thumb conformation (see Figure 5, top left). Only after the transition state is stabilized by the partial thumb closing and the active site assembly does Tyr271 move, breaking the bond with template and forming new bond with the primer 3′-end base. This bond pegs the last primer base in place for nucleophilic attack by P_α of the incoming nucleotide. In this scenario, it is likely that, in the absence of Tyr271 interaction, the primer will be relatively free to move away from the P_α of incoming nucleotide, directly influencing the chemical reaction distance ($P_\alpha-O3'$) and hence the rate of nucleotide incorporation. This observation provides structural rationalization for the observed decrease in k_{pol} on mutating Tyr271 with alanine.[38,63]

Role of Incoming Nucleotide Triphosphate Binding Residues in Catalysis. The nucleotidyl transfer reaction that follows the conformational closing of polβ involves the nucleophilic attack of the 3′-OH of the primer on P_α of the incoming substrate. Because this chemical reaction depends on the precise location of P_α within a certain distance of 3′-OH, residues interacting directly with the incoming nucleotide might affect the positioning of P_α. On the basis of crystal structures, residues Arg183, Ser180, and Gly189 located in the palm subdomain of polβ interact with the β and γ phosphates (see Figure 6) of the incoming nucleotide and might thus play important roles in catalysis or preparation for the chemical reaction.[31] In Figure 6 we follow the evolution of distances between these residues and the phosphate atoms of the incoming nucleotide. As shown in Figure 6a, the distances decrease by 1−2 Åwith the thumb's closing motion. Arg183, which interacts with the β phosphate of the dCTP, is most likely involved in the transition state stabilization by positioning P_α in the optimal orientation for the chemical reaction. Kinetic studies of the Arg183 mutant suggest

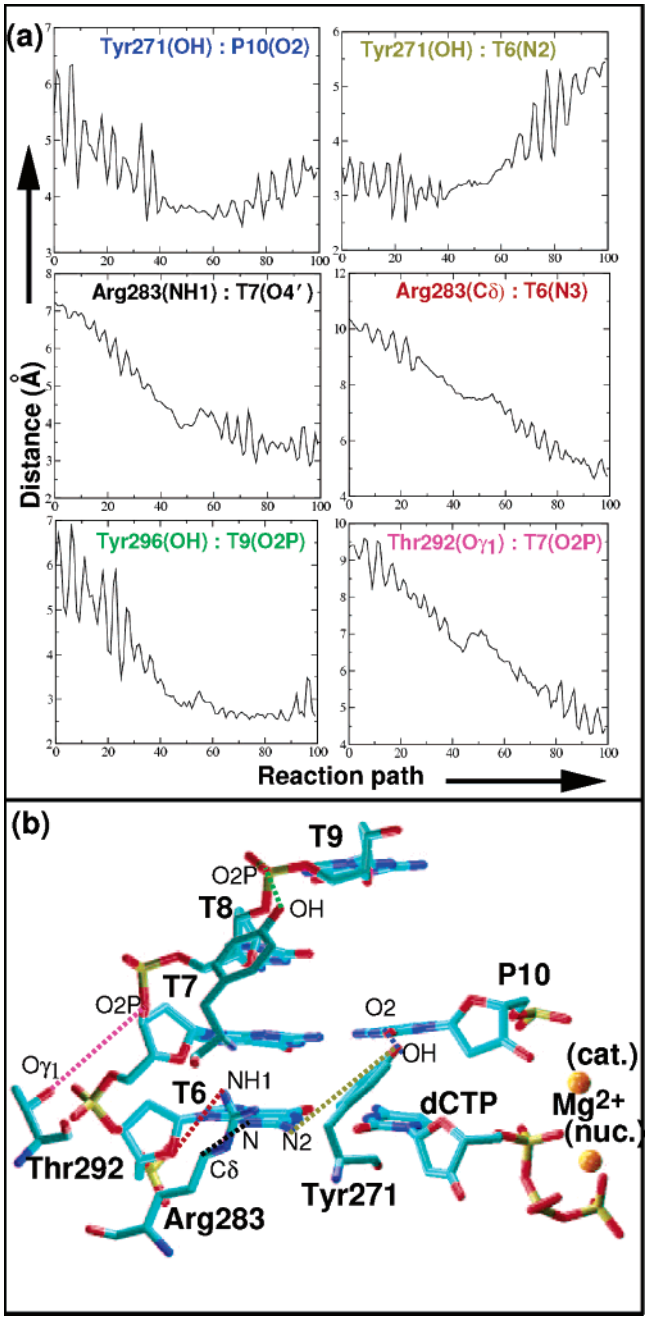

Figure 5. Evolution of distances of protein residues in the α-helix N of thumb domain interacting with DNA template/primer strand residues and incoming nucleotide (dCTP) along the pathway. Distances are diagrammed (below) (b) with arrows color coded corresponding to the color of legends in each panel window (top) (a).

significant decrease in k_{pol} for both correct and incorrect base pairs.[79] The importance of the β-phosphate coordination in the assembly of polymerase active site for catalysis was also noted by Mildvan and co-workers.[80]

Magnesium Ions Rearrangement. DNA polymerases across the different families have common two-metal ion coordination with few conserved acidic residues (e.g., two highly conserved aspartate residues and possibly either another aspartate or glutamate) in the enzyme active site.[55,56,80−82] In polβ's active site, three aspartate residues (residues 190, 192, and 256) coordinate the Mg^{2+} ions, resulting in a tight binding pocket for the metal ions. The magnesium ions have been implicated with assembling the acidic residues and catalyzing the nucleotide insertion via a "two-metal-ion" mechanism,[83] but mechanistic details are unknown. The critical role of magnesium ions in polβ's closing and active site assembly was also underscored in our recent studies,[61] where we showed that polβ's closing before chemistry requires both divalent metal ions in the active site, and opening after chemistry is triggered by release of the catalytic metal ion. We further suggested that subtle but slow adjustments of the catalytic and nucleotide binding magnesium ions may help guide polymerase selection for the correct nucleotide.

In Figure 7, we follow the evolution of distances of ligands coordinating the catalytic and nucleotide binding Mg^{2+} ions along the polβ closing pathway. We note here, concomitant to the thumb's motion, significant changes in the coordination of the magnesium ions to the conserved aspartate residues. Figure 7 also follows the crucial distance for the chemical reaction, namely the interaction between O3′ of the last primer residue and the P_α of the dCTP. Values of P_α−O3′ distance less than 4 Å are sampled for intermediate states in the second half of the polβ closing pathway. Of course, a proper study of the active site geometry requires a quantum/classical mechanics treatment; however, our preliminary optimization suggests that additional high-energy barriers must be overcome to reach the ideal geometry appropriate for the chemical reaction and that a separate "pre-chemistry" phase likely follows closing but precedes the chemical reaction itself.[84] In fact, we speculate that these motions involves adjustments of residues that contact the α, β, and γ phosphates of the incoming nucleotide.

Concluding Remarks

Determining pathways in biomolecular systems is an extremely challenging problem, in large part due to enormous sampling of phase space required to locate all the transition states. Earlier, the successful application of transition path sampling to determine polβ's pathway in explicit solvent conditions was made possible due a divide and conquer strategy for enumerating the multiple transition states[22] and an efficient new free energy method termed "BOLAS";[77] the original TPS formalism did not address these aspects. These TPS applications required about 8 months (24 CPUs of SGI Origin R10000 machines) of computing time to determine the complete polβ pathway with the free energy profiles and yet more time for analysis. In comparison, our application of SDEL to polβ's pathway with 100 intermediate structures and implicit treatment of solvent took only about 1 month (20 CPUs of 600 MHz LINUX machines) of computing time, but prior simulations may have helped understand the systems behavior. Still, the final path is likely a representative dominant trajectory, as revealed by the agreement to TPS. Efforts to increase resolution of the path further by increasing the number of intermediate structures to 200 did not succeed because of the inefficient optimization available in the package for a system size of polβ. Moreover, the final path obtained provides only qualitative information on the sequence of events. Determining thermodynamic free energy along the pathway from SDEL trajectory is theoretically possible, either by computing many independent paths from the ensemble of room-temperature structures[25,85] or by methods such as umbrella sampling.[86] This, however, would require successful refinement of the SDEL segments near the transition states and large computational efforts for complex systems such as polβ.

A comparison of polβ's conformational closing pathway generated using SDEL and an independent transition path sampling study[20] reveals excellent agreement, despite the very different approaches and different force fields. Although the TPS application was performed in explicit solvent using CHARMM,

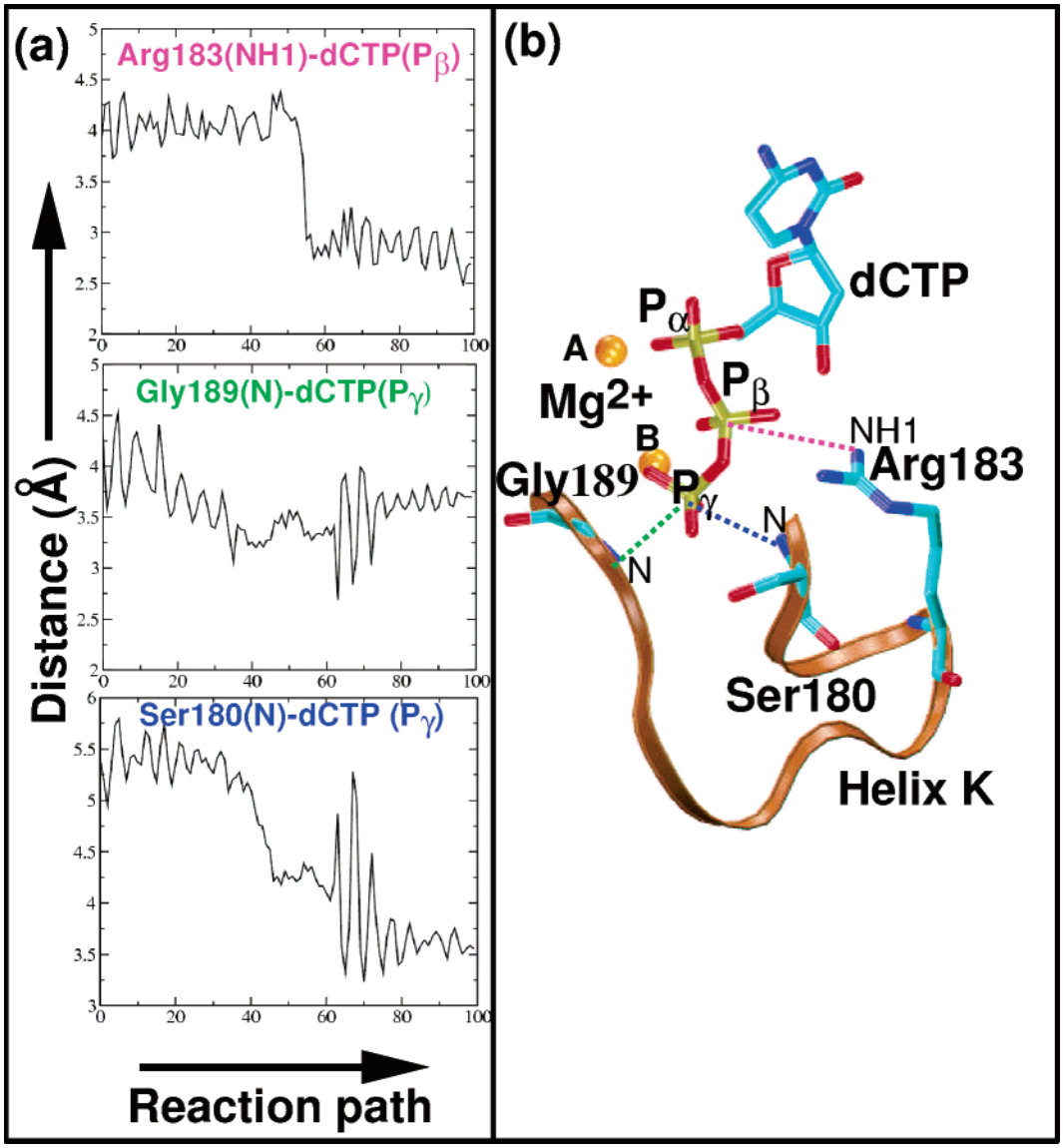

Figure 6. Evolution of distances of residues between triphosphate-bound residues Arg183, Ser180 and Gly189 and the incoming nucleotide (dCTP). Distances are diagrammed (right) (b) with arrows color coded corresponding to the color of legends in each panel window (left) (a). Catalytic and nucleotide binding magnesium ions are labeled as (A) and (B), respectively.

solvent effects were only considered here implicitly, and the AMBER force field was employed instead. The very good qualitative agreement between the two polβ closing pathways points to a dominant sequence of events during polβ's closing that is likely biologically significant. Namely, the enzyme's active site has evolved so as to trigger a sequence of subtle conformational rearrangements that is sensitive to the active site content (e.g., no substrate, right or wrong substrate complementary to the template) and, thereby, regulates replication or repair fidelity.

The conformational transition pathway of polβ was investigated using the stochastic difference equation algorithm that approximates long-time dynamics. The order of events during polβ closing was identified, and the evolution of key protein residues in the microenvironment of the incoming substrate interacting with template/primer DNA was followed along the conformational transition pathway. Simulations provided a detailed atomistic view of intermediate metastable conformations of polβ connecting crystal states, not available experimentally. We find the general sequence during polβ's closing to be: thumb's partial subdomain motion, Asp192 flip, Phe272 flip, and Arg258 rotation coupled to ion rearrangements (Figures 2−4). We also highlight cooperative motions of residues interacting with DNA template/primer (Tyr271, Arg283, Thr292, Tyr296) and those coordinating the β and γ phosphates (Ser180, Arg183, Gly189). The final state of the active site also suggests additional conformational rearrangements, which we term the "pre-chemistry avenue" prior to the chemical reaction.

We propose that a mismatched incoming unit to the template base will produce a very different trajectory and likely not a closed state. This hypothesis is supported by our standard molecular dynamics simulations with various mismatches[87] in the active site that show that the open thumb conformation is favored over the closed state and that the bases are in staggered conformations rather than forming hydrogen bonds. The recent A:C and T:C mismatch crystal structures of polβ corroborates these observations.[32] A similar line of evidence is provided by the transition path sampling study of a G:A mispair which found multiple pathways in the mispair compared to one dominant pathway for the correct pair.[75] An extensive experimental study on mismatches for the high-fidelity *Bacillus* DNA polymerase I also reveals that the interactions of the polymerase with the mismatched bases may be unique to each mispair.[88] Furthermore, DNA polymerases differ in their response in relation to

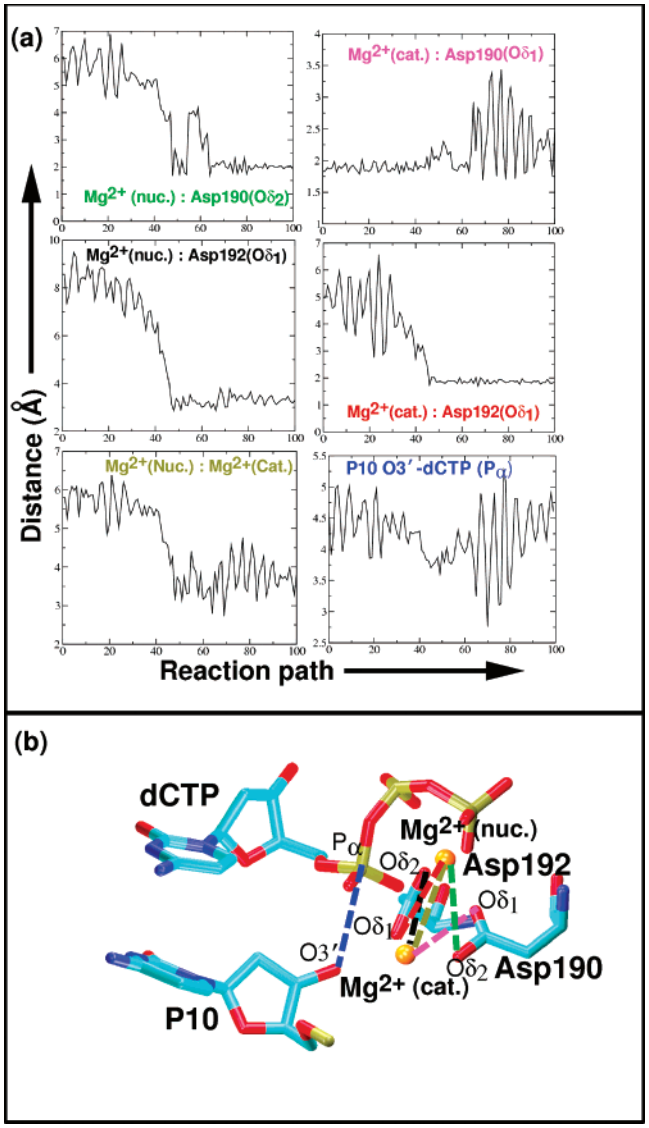

Figure 7. Evolution of distances of conserved aspartates coordinating catalytic and nucleotide binding Mg^{2+} ions along the pathway. These plots underscore the subtle rearrangement of magnesium ions in the enzymes active site concomitant with the thumb's subdomain motion. The crucial distance for the nucleotidyl transfer chemical reaction (O3′ of last primer residue to P_α of dCTP) is also provided (lower right) (a). Distances are diagrammed (below) (b) with arrows color coded corresponding to the color of legends in each panel window (top) (a).

a mispair. For example, the G:T mismatch which is readily extended by *Bacillus* DNA polymerase I was shown to form a reverse wobble in low-fidelity polymerase Dpo4[89] and these wobbles may hamper extension.

Acknowledgment. We thank Prof. Ron Elber (Cornell University) for providing the SDEL code and CPU time on the LINUX cluster at Cornell. We thank Prof. Alfredo Càrdenas (University of South Florida) for assisting with implementation of SDEL code. We thank Prof. Ravi Radhakrishnan (University of Pennsylvania) and Dr. Linjing Yang for stimulating discussions throughout this work. We are grateful to Drs. S. H. Wilson and W. A. Beard (NIEHS) for many insightful comments. Molecular images were generated using VMD[90] and INSIGHTII (Accelrys Inc., San Diego, CA). This work was supported by NSF grant ASC-9318159, and NIH grant R01 GM55164. Acknowledgment is also made to the donors of the American Chemical Society Petroleum Research Fund for support (or partial support) of this research (Award PRF39115-AC4 to T. Schlick).

Supporting Information Available: Cartesian coordinates of the incoming nucleotide (dCTP) and magnesium ions from the starting (open) and final (closed) conformations of polβ are available as PDB files. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Schlick, T.; Skeel, R. D.; Brünger, A. T.; Kale, L. V.; Board, J. A.; Hermans, J.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 9−48.
(2) Schlick, T.; Beard, D.; Huang, J.; Strahs, D.; Qian, X. *IEEE Comput. Sci. Eng.* **2000**, *2*, 38−51.
(3) Schlick, T. *Molecular Modeling and Simulation: An Interdisciplinary Guide;* Springer-Verlag: New York, 2002.
(4) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646−652.
(5) McCammon, J. A.; Harvey, S. C. *Computer modeling of chemical reactions in enzymes and solution;* Cambridge Univeristy Press: Cambridge, 1987.
(6) Benkovic, S. J.; Hammes-Schiffer, S. *Science* **2004**, *301*, 1196−1202.
(7) Warshel, A. *Computer modeling of chemical reactions in enzymes and solution;* John Wiley and Sons: New York, 1989.
(8) Florián, J.; Goodman, M. F.; Warshel, A. *J. Am. Chem. Soc.* **2003**, *125*, 8163−8177.
(9) Daggett, V. *Acc. Chem. Res.* **2002**, *35*, 422−429.
(10) Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. *Nature* **2002**, *420*, 102−106.
(11) Zhou, R.; Berne, B. J.; Germain, R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14931−14936.
(12) Zhou, R.; Huang, X.; Margulis, C.; Berne, B. J. *Science* **2004**, *305*, 1605−1609.
(13) Hummer, G.; Schotte, F.; Anfinrud, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15330−15334.
(14) Hamelberg, D.; Mongan, J.; McCammon, J. A. *J. Chem. Phys.* **2004**, *120*, 11919−11929.
(15) Voter, A. F. *Phys. Rev. B* **1998**, *57*, 13985−13988.
(16) Rhee, Y. M.; Pande, V. S. *Biophys. J.* **2003**, *84*, 775−786.
(17) Brooks, C. L. *Acc. Chem. Res.* **2002**, *35*, 447−454.
(18) Becker, O. M., MacKerell, A. D., Jr., Roux, B., Wantanabe, M., Eds. *Computational Biochemistry and Biophysics*; Marcel Drekker, Inc.: New York, 2001; pp 169−197.
(19) Lu, N.; Wu, D.; Woolf, T. B.; Kofke, D. A. *Phys. Rev. E* **2004**, *69*, 057702−057706.
(20) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291−318.
(21) Singhal, N.; Snow, C. D.; Pande, V. S. *J. Chem. Phys.* **2004**, *121*, 415−425.
(22) Radhakrishnan, R.; Schlick, T. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 5970−5975.
(23) Elber, R.; Cárdenas, A.; Ghosh, A.; Stern, H. *Adv. Chem. Phys.* **2003**, *126*, 93−129.
(24) Cárdenas, A.; Elber, R. *Biophys. J.* **2003**, *85*, 2919−2939.
(25) Arora, K.; Schlick, T. *Chem. Phys. Lett.* **2003**, *378*, 1−8.
(26) Ames, B. N.; Shigenaga, M. K.; Hagen, T. M. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *90*, 7915−7922.
(27) Lindhal, T.; Wood, R. D. *Science* **1999**, *286*, 1897−1905.
(28) Seeberg, E.; Eide, L.; Bjoras, M. *Trends Biochem. Sci.* **1995**, *20*, 391−397.
(29) Wilson, S. H. *Mutat. Res.* **1998**, *407*, 203−215.
(30) Joyce, C. M.; Steitz, T. A. *Annu. Rev. Biochem.* **1994**, *63*, 777−822.
(31) Sawaya, M. R.; Parsad, R.; Wilson, S. H.; Kraut, J.; Pelletier, H. *Biochemistry* **1997**, *36*, 11205−11215.
(32) Krahn, J. M.; Beard, W. A.; Wilson, S. H. *Structure* **2004**, *12*, 1823−1832.
(33) Ahn, J.; Werneburg, B. G.; Tsai, M.-D. *Biochemistry* **1997**, *36*, 1100−1107.
(34) Ahn, J.; Kraynov, V. S.; Zhong, X.; Werneburg, B. G.; Tsai, M.-D. *Biochem. J.* **1998**, *331*, 79−87.
(35) Vande Berg, B. J.; Beard, W. A.; Wilson, S. H. *J. Biol. Chem.* **2001**, *276*, 3408−3416.
(36) Shah, A. M.; Li, S. X.; Anderson, K. S.; Sweasy, J. B. *J. Biol. Chem.* **2001**, *276*, 10824−10831.
(37) Suo, Z.; Johnson, K. A. *J. Biol. Chem.* **1998**, *273*, 27250−27258.
(38) Kraynov, V. S.; Werneburg, B. G.; Zhong, X.; Lee, H.; Ahn, J.; Tsai, M.-D. *Biochem. J.* **1997**, *323*, 103−111.

(39) Zhong, X.; Patel, S. S.; Werneburg, B. G.; Tsai, M.-D. *Biochemistry* **1997**, *36*, 11891−11900.
(40) Dahlberg, M. E.; Benkovic, S. J. *Biochemistry* **1991**, *30*, 4835−4843.
(41) Kuchta, R. D.; Mizrahi, V.; Benkovic, P. A.; Johnson, K. A.; Benkovic, S. J. *Biochemistry* **1987**, *26*, 8410−8417.
(42) Wong, I.; Patel, S. S.; Johnson, K. A. *Biochemistry* **1991**, *30*, 526−537.
(43) Patel, S. S.; Wong, I.; Johnson, K. A. *Biochemistry* **1991**, *30*, 511−525.
(44) Frey, M. W.; Sowers, L. C.; Millar, D. P.; Benkovic, S. J. *Biochemistry* **1995**, *34*, 9185−9192.
(45) Capson, T. L.; Peliska, J. A.; Kaboord, B. F.; Frey, M. W.; Lively, C.; Dahlberg, M.; Kovic, S. J. *Biochemistry* **1992**, *31*, 10984−10994.
(46) Koshland, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1958**, *44*, 98−104.
(47) Koshland, D. E. *Angew Chem., Int. Ed. Engl.* **1994**, *33*, 2375−2378.
(48) Arora, K.; Schlick, T. *Biophys. J.* **2004**, *87*, 3088−3099.
(49) Post, C. B.; Ray, W. J., Jr. *Biochem.* **1995**, *34*, 15881−15885.
(50) Iwanaga, A.; Ouchida, M.; Miyazaki, K.; Hori, K.; Mukai, T. *Mutat. Res.* **1999**, *435*, 121−128.
(51) Hoeijmakers, J. H. J. *Nature* **2001**, *411*, 366−374.
(52) Starcevic, D.; Dalal, S.; Sweasy, J. B. *Cell Cycle* **2004**, *3*, 998−1001.
(53) Beard, W. A.; Shock, D. D.; Wilson, S. H. *J. Biol. Chem.* **2004**, *279*, 31921−31929.
(54) Andricioaei, I.; Goel, A.; Herschbach, D.; Karplus, M. *Biophys. J.* **2004**, *87*, 1478−1497.
(55) Abashkin, Y. G.; Erickson, J. W.; Burt, S. K. *J. Phys. Chem. B* **2001**, *105*, 287−292.
(56) Rittenhouse, R. C.; Apostoluk, W. K.; Miller, J. H.; Straatsma, T. P. *Proteins: Struct. Funct. Genet.* **2003**, *53*, 667−682.
(57) Florián, J.; Warshel, A.; Goodman, M. F. *J. Phys. Chem. B* **2002**, *106*, 5754−5760.
(58) Yang, L.; Beard, W. A.; Wilson, S. H.; Broyde, S.; Schlick, T. *J. Mol. Biol.* **2002**, *317*, 651−671.
(59) Yang, L.; Beard, W. A.; Wilson, S. H.; Roux, B.; Broyde, S.; Schlick, T. *J. Mol. Biol.* **2002**, *321*, 459−478.
(60) Yang, L.; Beard, W. A.; Wilson, S. H.; Broyde, S.; Schlick, T. *Biophys. J.* **2004**, *86*, 3392−3408.
(61) Yang, L.; Arora, K.; Beard, W. A.; Wilson, S. H.; Schlick, T. *J. Am. Chem. Soc.* **2004**, *126*, 8441−8453.
(62) MacKerell, A. D., Jr.; Banavali, N. K. *J. Comput. Chem.* **2000**, *21*, 105−120.
(63) Beard, W. A.; Osheroff, W. P.; Prasad, R.; Sawaya, M. R.; Jaju, M.; Wood, T. G.; Kraut, J.; Kunkel, T. A.; Wilson, S. H. *J. Biol. Chem.* **1996**, *271*, 12141−12144.
(64) Beard, W. A.; Shock, D. D.; VandeBerg, B. J.; Wilson, S. H. *J. Biol. Chem.* **2002**, *277*, 47393−47398.
(65) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049−1074.
(66) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122−129.
(67) Tsui, V.; Case, D. A. *Biopolymers* **2000**, *56*, 275−291.
(68) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(69) Elber, R.; Ghosh, A.; Cárdenas, A. *Acc. Chem. Res.* **2002**, *35*, 396−403.
(70) Czerminski, R.; Elber, R. *Int. J. Quantum Chem.* **1990**, *24*, 167−185.
(71) Czerminski, R.; Elber, R.; Roiterberg, A.; Simmerling, C.; Goldstein, R.; Li, H.; Verkhivker, G.; Kesar, C.; Zhang, J.; Ulitsky, A. *Comput. Phys. Commun.* **1995**, *91*, 159−189.
(72) Faradjian, T.; Elber, R. *J. Chem. Phys.* **2004**, *120*, 10880−10889.
(73) Zaloj, V.; Elber, R. *Comput. Phys. Commun.* **2000**, *128*, 118−127.
(74) Cárdenas, A.; Elber, R. *Proteins* **2003**, *51*, 245−257.
(75) Radhakrishnan, R.; Schlick, T. Submitted for publication.
(76) Stillinger, F. H.; Weber, T. A. *Science* **1984**, *225*, 983−989.
(77) Radhakrishnan, R.; Schlick, T. *J. Chem. Phys.* **2004**, *121*, 2436−2444.
(78) Beard, W. A.; Wilson, S. H. *Chem. Biol.* **1998**, *5*, R7−R13.
(79) Kraynov, V. S.; Showalter, A. K.; Liu, J.; Zhong, X.; Tsai, M.-D. *Biochemistry* **2000**, *39*, 16008−16015.
(80) Ferrin, L. J.; Mildvan, A. S. *Biochemistry* **1986**, *25*, 5131−5145.
(81) Steitz, T. A. *Nature* **1998**, *391*, 231−232.
(82) Mildvan, A. S. *Proteins: Struct. Funct. Genet.* **1997**, *29*, 401−416.
(83) Beese, L. S.; Steitz, T. A. *EMBO J.* **1991**, *9*, 25−33.
(84) Lahiri, S. D.; Zhang, G.; Dunaway-Mariano, D.; Allen, K. N. *Science* **2003**, *299*, 2067−2071.
(85) Ghosh, A.; Elber, R.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10394−10398.
(86) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187−199.
(87) Arora, K.; Schlick, T. Manuscript in preparation.
(88) Johnson, S. J.; Beese, L. S. *Cell* **2004**, *116*, 803−816.
(89) Trincao, J.; Johnson, R. E.; Wolfle, W. T.; Escalante, C. R.; Prakash, S.; Prakash, L.; Aggarwal, A. K. *Nat. Struct. Biol.* **2004**, *11*, 457−462.
(90) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33−38.