JMB

Local Deformations Revealed by Dynamics Simulations of DNA Polymerase β with DNA Mismatches at the Primer Terminus

Linjing Yang[1], William A. Beard[2], Samuel H. Wilson[2], Benoit Roux[3] Suse Broyde[4] and Tamar Schlick[1]*

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

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

[3]*Department of Biochemistry Weill Medical College of Cornell University, 1300 York Avenue, New York, NY 10021 USA*

[4]*Department of Biology, New York University, New York, NY 10003, USA*

*Corresponding author

Nanosecond dynamics simulations for DNA polymerase β (pol β)/DNA complexes with three mismatched base-pairs, namely GG, CA, or CC (primer/template) at the DNA polymerase active site, are performed to investigate the mechanism of polymerase opening and how the mispairs may affect the DNA extension step; these trajectories are compared to the behavior of a pol β/DNA complex with the correct GC base-pair, and assessed with the aid of targeted molecular dynamics (TMD) simulations of all systems from the closed to the open enzyme state. DNA polymerase conformational changes (subdomain closing and opening) have been suggested to play a critical role in DNA synthesis fidelity, since these changes are associated with the formation of the substrate-binding pocket for the nascent base-pair. Here we observe different large C-terminal sub-domain (thumb) opening motions in the simulations of pol β with GC *versus* GG base-pairs. Whereas the conformation of pol β in the former approaches the observed open state in the crystal structures, the enzyme in the latter does not. Analyses of the motions of active-site protein/ DNA residues help explain these differences. Interestingly, rotation of Arg258 toward Asp192, which coordinates both active-site metal ions in the closed "active" complex, occurs rapidly in the GG simulation. We have previously suggested that this rotation is a key slow step in the closed to open transition. TMD simulations also point to a unique path-way for Arg258 rotation in the GG mispair complex. Simulations of the mismatched systems also reveal distorted geometries in the active site of all the mispair complexes examined. The hierarchy of the distortions (GG > CC > CA) parallels the experimentally deduced inability of pol β to extend these mispairs. Such local distortions would be expected to cause inefficient DNA extension and polymerase dissociation and thereby might lead to proofreading by an extrinsic exonuclease. Thus, our studies on the dynamics of pol β opening in mismatch systems provide structural and dynamic insights to explain experimental results regarding inefficient DNA extension following misincorporation; these details shed light on how proofreading may be invoked by the abnormal active-site geometry.

© 2002 Elsevier Science Ltd. All rights reserved

Keywords: dynamics simulations; DNA polymerase β; opening; DNA mispairs; DNA extension

Introduction

DNA polymerases are essential elements in the transmittal and maintenance of genomic integrity. The key function of polymerases during DNA replication and repair synthesis is the addition of deoxynucleotides onto the growing end of a DNA primer strand using a single-stranded DNA as a template. Base-substitution error rates for DNA
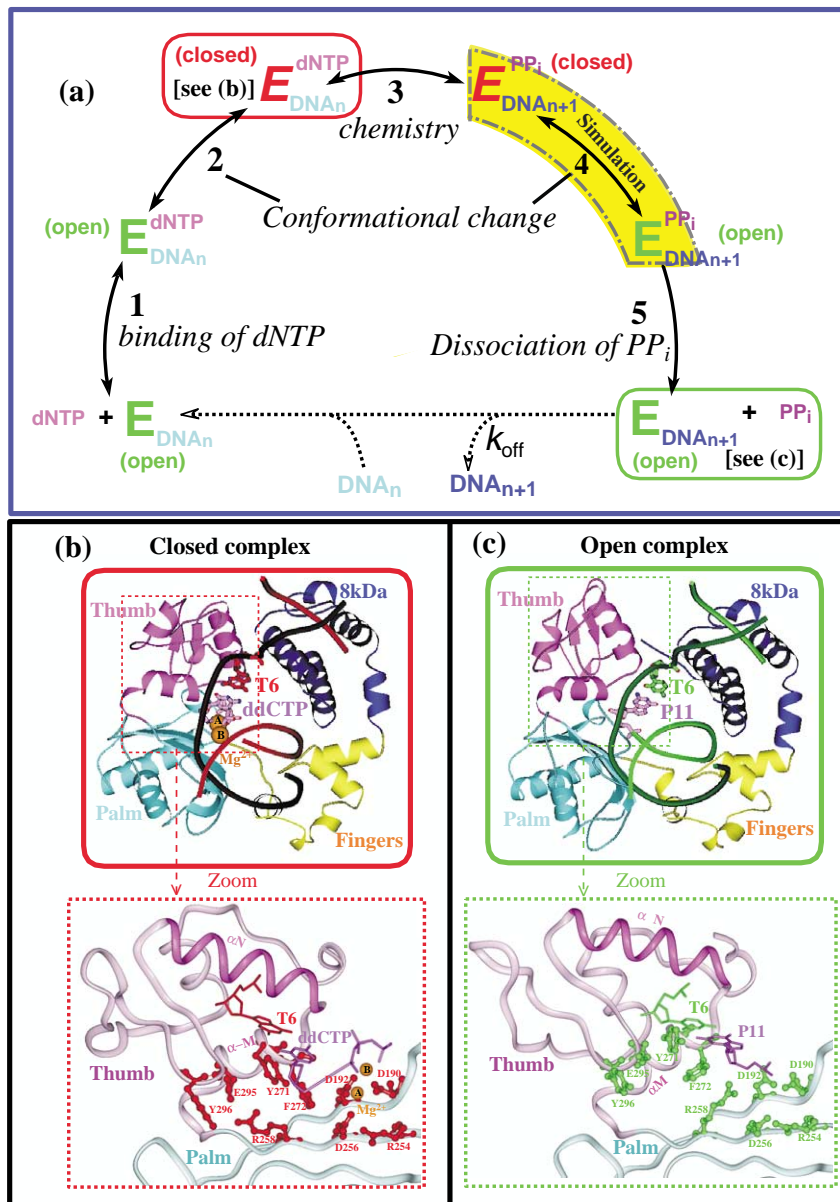
Figure 1. General pathway for (a) nucleotide insertion by DNA polymerases, with (b) corresponding crystallographic closed and (c) open conformations of pol β/DNA along the pathway. In (a), E, dNTP, PP$_i$ refer to DNA polymerase, 2′-deoxyribonucleoside 5′-triphosphate, and pyrophosphate, respectively. DNA$_n$ and DNA$_{n+1}$ represent the DNA before and after the nucleotide incorporation to the primer strand of the duplex primer/template DNA. Red (closed) and green (open) distinguish the crystal closed and open states of polymerase/DNA complexes. The highlighted yellow arc represents the pol β conformational opening (after chemistry) which we investigate here. See the text for a detailed description of the catalytic cycle. (b) The structure of the closed crystal pol β enzyme bound to DNA and ddCTP before chemistry. (c) The open crystal pol β structure bound to a nicked DNA after chemistry. Close-ups of the active site as well as the thumb subdomain (inside broken-line boxes) show the pol β movement, especially the motion of α-helix N (highlighted in dark pink). The two orange spheres labeled A and B in (b) refer to the specific nucleotide binding and catalytic magnesium ions, respectively. Note that the active-site base-pair T6-ddCTP in (b), and T6-P11 in (c), is our base-pair of focus; T and P indicate template and primer strand, respectively, with bases numbered from the 5′-end of the DNA strand. Parts of this Figure were designed with MOLSCRIPT[76] and Insight, version 2000.

synthesis vary depending on the DNA polymerase, the mispair, and the local sequence environment. The error frequency of DNA repair and replication polymerases that lack a proofreading exonuclease is approximately one in 10^3–10^6 bases synthesized.[1] DNA polymerase β (pol β) belonging to the X-family of DNA polymerases fills short DNA gaps during base excision repair and has an average base-substitution error frequency of one per 10^3 bases.[2,3] More recently, highly error-prone DNA polymerases primarily belonging to the Y-family have been observed to exhibit error frequencies that approach the theoretical maximum of 1.[4]

Since polymerase-initiated DNA synthesis errors may play a central role in human aging and diseases, it is important to understand the mechanisms utilized by DNA polymerases in selecting the

appropriate nucleotide at the molecular level. Molecular dynamics simulations, in particular, can complement experimental structural and kinetic data by probing mechanistic details associated with specific motions at the active site; in addition, long-time motions on the biological time-frames, unapproachable by standard dynamics simulations, may be examined by enhanced sampling techniques such as targeted molecular dynamics (TMD) simulations.[5–9]

Several DNA polymerases, including pol β, have been extensively characterized structurally and kinetically. As sketched in Figure 1, the initial step of DNA synthesis involves binding a primer/template duplex DNA to a DNA polymerase. After DNA binding, the binary DNA polymerase complex binds a 2′-deoxyribonucleoside 5′-triphosphate (dNTP) to form a ternary substrate

complex (step 1). This complex undergoes a conformational change of the enzyme from an open to a closed, catalytically competent state (step 2). Following the fast chemical reaction of dNTP incorporation onto the DNA primer strand (step 3), the product polymerase complex undergoes a second conformational change from a closed to an open state (step 4). The release of pyrophosphate (PP$_i$) and DNA translocation (step 5) follow, after which the reaction cycle may begin anew. The two steps involving the conformational changes of closing (before chemistry, step 2) and opening (after chemistry, step 4) have been suggested to play key roles in the faithful incorporation of nucleotides through geometric selection[10] *via* an induced fit mechanism.[11–14]

The fidelity of DNA polymerases in selecting the correct dNTP is related to several structural factors: Watson–Crick hydrogen bonding,[15] the steric fit between the incoming dNTP and the polymerase active-site pocket (including template),[10] and the pattern of hydrogen bonding between polymerases and duplex DNA bases.[16] Although Watson–Crick hydrogen bonding plays a significant role in pairing selectivity and stabilizing the existing DNA helix, steric and geometric aspects of the active site may play a more important role in maintaining DNA synthesis fidelity.[17–19]

DNA polymerase structures support a "two-metal-ion" mechanism of nucleotide addition:[20] the incoming nucleotide is accompanied by a magnesium ion (metal A) that coordinates the phosphoryl oxygen atoms from all three phosphate groups. In addition, a catalytic magnesium ion (metal B) coordinates a phosphoryl oxygen atom of the α-phosphate group. In pol β, these magnesium ions are coordinated in the active site by three conserved aspartate residues (Asp190, Asp192, and Asp256).[12] According to this two-metal-ion mechanism, the catalytic Mg^{2+} lowers the affinity of the hydrogen atom in the primer 3′-OH group, facilitating the 3′-O attack on the P^α of the incoming nucleotide. The nucleotide-binding Mg^{2+} assists the leaving of the pyrophosphate, and both ions stabilize the structure of the presumed penta-coordinated transition state.

The strategies employed by DNA polymerases to optimize their fidelity are not understood at the molecular level. For a polymerase to produce a base-substitution error, it must incorporate an incorrect nucleotide and subsequently extend the mismatch to "seal" the error. DNA polymerases pause after insertion of an incorrect nucleotide (i.e. extension is slow). This provides an opportunity for a proofreading exonuclease domain to remove the mismatched primer terminus. If the polymerase lacks an intrinsic proofreading activity as in the case for pol β, the polymerase may dissociate from DNA giving access to an extrinsic proofreading activity.

Pol β is the smallest eukaryotic cellular DNA polymerase.[21] Its two domains consist of an N-terminal 8 kDa region which exhibits deoxyribosephosphate lyase activity and a C-terminal 31 kDa region which possesses nucleotidyl transfer activity.[22] Because of its relatively small size and the large amount of available experimental structural and kinetic data, pol β is an ideal model for studying fidelity by computational techniques. Most DNA polymerases share a common overall architecture which can be likened to a right-hand consisting of "thumb", "palm", and "fingers" subdomains.[23] In contrast, members of the X-family of DNA polymerases, such as pol β, are left-handed so that the fingers and thumb subdomain nomenclature is opposite to that of other DNA polymerases.[22]

For many DNA polymerases, slow conformational changes both before and after the incorporation of a correct nucleotide into the DNA primer strand have been suggested on the basis of presteady-state kinetic data[24]). Kinetic and structural data have recently suggested that the large subdomain movement *per se* is relatively rapid.[25–28] However, details of the polymerase mechanism in stabilizing an incoming dNTP complementary to the template to maintain DNA synthesis fidelity remain unknown, especially regarding the precise events involved in the large-scale subdomain motions. In our previous work,[29] consisting of dynamics simulations (both at room and high temperatures, and TMD simulations) for the pol β/DNA complex, a specific sequence of events in the polymerase opening has been proposed: (1) Phe272 ring flip; (2) large thumb motion; and (3) Arg258 rotation toward Asp192. Our data also suggest that the local motions might limit the conformational change in closing when an incorrect nucleotide binds and thus help maintain pol β's ability to discriminate between correct and incorrect nucleotides.

To further explore polymerase structural transitions that may enhance DNA synthesis fidelity, we examine here, by atomistic molecular dynamics simulations over 6 ns, the behavior of pol β/DNA complexes with the mismatched base-pairs GG, CA, and CC (primer/template nucleotides), at the polymerase active site. Another 6 ns simulation trajectory of the pol β/DNA complex with the correct nascent GC base-pair was also performed for comparison. Our simulations are experimentally anchored to the polymerase reaction pathway through available crystallographic structures for pol β with primer/template DNA.[12] Namely, available are crystal structures of two binary complexes of pol β·gap and pol β·nick in the open state, and one ternary complex of pol β·gap·ddCTP in the closed state.

Pol β binds a single-nucleotide gap DNA in pol β·gap before chemistry, and is bound to a nicked DNA in pol β·nick after chemistry. In the pol β·gap·ddCTP complex, pol β binds both a single-nucleotide gap DNA substrate and a 2′,3′-dideoxy-ribocytidine 5′-triphosphate (ddCTP). The catalytic reaction cycle proceeds from the open
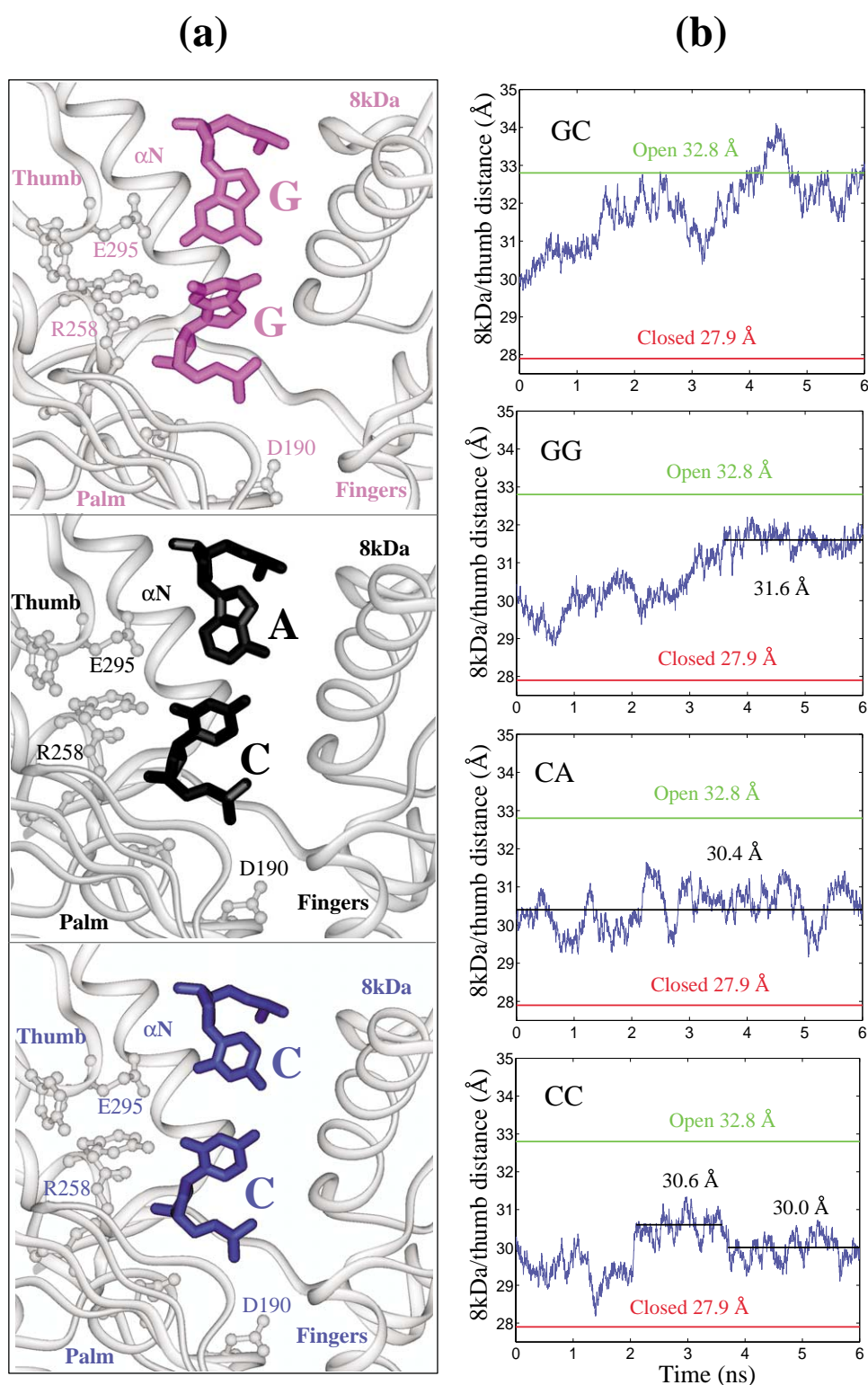
(a)　　　　　　　　　　　**(b)**

Figure 2. (a) The starting structures for 6 ns simulations of the three half-open mismatches GG (pink), CA (black), and CC (blue) in the active site. (b) The evolutions of center-to-center distance (d_f in Å) between 8 kDa domain and thumb subdomain as a function of time for the GC, GG, CA, and CC trajectories. The reference d_f values in both the closed (red) and open (green) crystal structures are also presented for comparison.

gapped DNA binary complex to the closed ternary complex to the nicked DNA binary complex. The crystal structures of the closed ternary complex and the open binary nick complex

were used to construct our initial models for the simulations.

Our dynamics simulations reveal a large C-terminal subdomain (thumb) motion in the
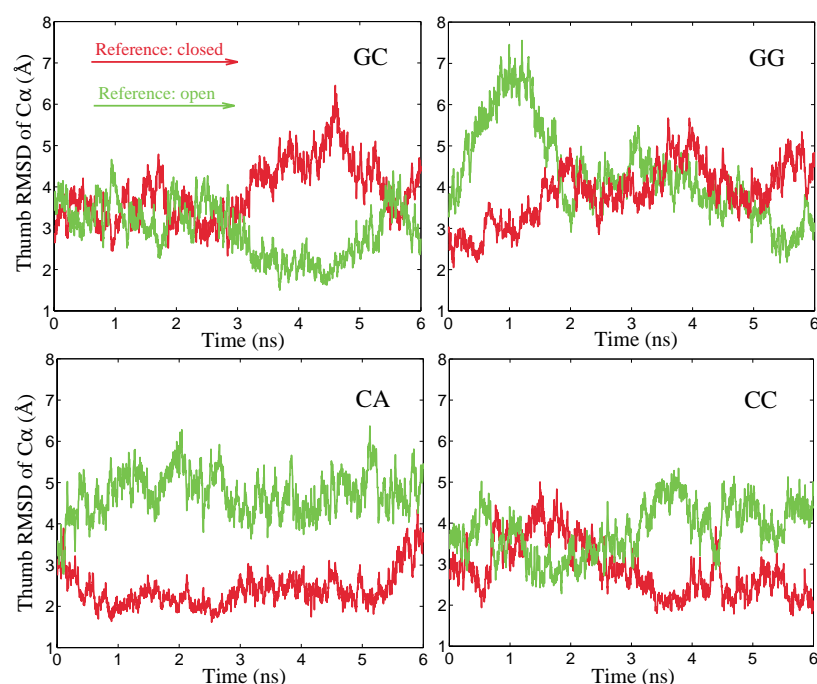
Figure 3. The evolution of the RMSD of the thumb C^α atoms in all trajectories relative to the crystal closed (red) and open (green) structures. Superimpositioning is performed according to the palm C^α atoms.

half-open GG and GC trajectories, but the thumb in the former does not approach that in the crystal open form (as it does for the GC system). Among the mispair complexes, Arg258 rotation toward Asp192 is only observed in the GG simulation. Analyses of the motions of active-site residues provide structural insights for the differing behavior regarding polymerase opening. Simulations also reveal structural distortions of the nascent mispairs GG, CC, and CA, with GG displaying the largest distortion. These results are consistent with the experimentally determined inefficiency of correct nucleotide incorporation extending different mispairs. Since pol β does not possess an intrinsic exonuclease function, inefficient mismatched DNA extension would lead to polymerase dissociation; this in turn would provide access to an extrinsic proofreading exonuclease. TMD simulations for four pol β/DNA complexes with GC, GG, CA, and CC nascent base-pair starting from a closed state suggest a different and aberrant pathway for Arg258 rotation in the opening of the GG complex. We also suggest, on the basis of the analysis of protein residues in close contact with nascent base-pairs, that protein residues Tyr271, Glu295, Asp276, and Lys280 may play an important role in pol β opening. Further studies show that the CA mispair resembles the correct GC base-pair in terms of water and sodium ion coordination with active-site residues, whereas the GG and CC mispairs do not. Our simulation results thus provide some structural and dynamic clues into the inefficient extension of mismatched base-pairs, especially for the GG mispair.

Results

Initial models

In our previous studies for pol β/DNA complex,[29] the simulations starting from a closed state failed to capture the C-terminal (thumb) subdomain opening. To accelerate conformational changes associated with subdomain opening, we constructed a half-open structure, which did approach an open state in the simulations and provided insightful results especially in relation to experimental observations. For the same reason, all simulations for the mispair complexes were also started from the half-open state, which in addition facilitates comparison. The half-open model for the pol β/DNA system was built as an average of the crystallographic coordinates of the closed ternary (1BPY) and open binary nick (1BPZ) complexes[29] (see Computational Methodology). Specifically, the model's thumb subdomain is in a partially open state and the nascent DNA base-pair is GC. On the basis of this half-open structure, three initial models of pol β/DNA complex with GG, CA, or CC mispairs in the active site were built by replacing the correct GC base-pair with GG, CA, or CC. The INSIGHT II package, version 2000, was used for this construction, which left the other protein residues and DNA sequences in place. The generated steric clashes in all systems were removed by subsequent energy minimization and equilibration (see Computational Methodology). These half-open pol β/DNA models with mismatched base-pairs were similarly solvated in a face-centered cube for the periodic domain, and were neutralized with Na^+ and Cl^- counterions at
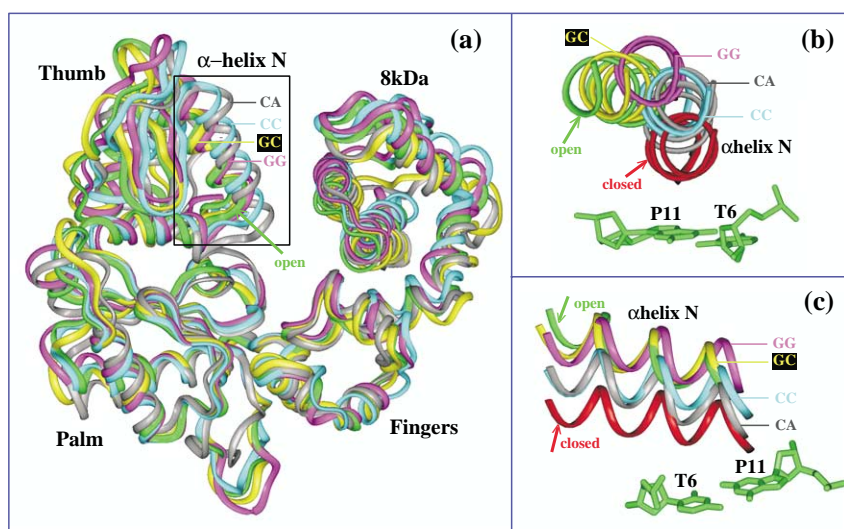
Figure 4. Comparison of simulated systems with the open crystal structure for (a) pol β and the α-helix N in the thumb subdomain ((b) and (c)). The α-helix N positions in all simulated systems are also compared with that in the closed crystal structure. All systems are superimposed according to the palm C^α atoms. The colored ribbons red, green, yellow, pink, gray, and cyan indicate the closed crystal, the open crystal structure, and the simulated half-open GC, GG, CA, and CC, respectively. The position of α-helix N is shown from two points of view.

an ionic strength of 150 mM.[29] This modeling produced three solvated models representing a half-open pol β/DNA complex with GG (41,976 total atoms), CA (41,971 atoms), or CC (41,970 atoms) mispairs. Full details are given Yang *et al.*[29] and under Computational Methodology. We call these three mismatch models half-open GG, CA, and CC herein. These mismatch half-open models, as well as the half-open pol β/DNA with the correct base-pair GC (termed half-open GC, for comparison), were subjected to molecular dynamics simulations over 6 ns. The starting structures of the three mispairs in the active site are shown in Figure 2(a).

To supplement the standard dynamics simulations for the half-open systems, we also constructed four separate pol β/DNA complexes with GC, GG, CA, or CC base-pairs in the active site on the basis of the modified, crystal ternary closed structure[12] (see Figure 1(b)). In these closed complexes, the incoming dNTP has reacted with the primer $3'$ terminus, as in the above half-open models. These closed GC, GG, CA, and CC systems are subjected to TMD simulations (see Computational Methodology) to explore possible pathways of polymerase opening. The target state for the TMD simulations is the crystal open binary nick complex of pol β/DNA (see Figure 1(c)).

Large thumb motion in the GC and GG simulations

We first use the center-to-center distance (d_f) between the 8 kDa domain and the thumb subdomain (heavy atoms only) to assess the thumb opening motion. Figure 2(b) presents the time evolution of d_f in the four half-open simulations as compared to the reference d_f distances in the crystal closed and open structures. We see that the GC system has already approached an open state resembling that of the crystal open structure at the

end of the simulation, with d_f larger than the open-form reference of 32.8 Å. The simulation of the GG system reveals opening, although with final d_f that is 1.2 Å smaller than the open reference value. The CA and CC mismatch models remain far away from the open state at the end of the 6 ns simulation.

To further explore the degree of opening, we compare the thumb motions in all half-open systems, as reflected by the evolution of the root-mean-square deviations (RMSD) of the thumb C^α atoms relative to those of both the closed and open crystal structures (Figure 3). We observe a large thumb motion in both the GC and GG systems. However, the minimum RMSD values relative to the open crystal structure in the GC and GG trajectories are different: 1.5 Å in the former and 2.2 Å in the latter. This difference, coupled to a simple inspection of the structures, indicates that while the thumb in the GG has approached an open state, this state differs from that approached in the GC system. The thumb subdomain is relatively stable in both simulations of the half-open CA and CC.

Figure 4 also shows for the half-open GC, GG, CA, and CC complexes the positions of α-helix N (which defines the closed and open states for pol β) at the end of the simulations, as well as for the crystal closed and open structure. The six systems are superimposed according to the C^α atoms in the palm subdomain. We see that α-helix N in the simulated open state of the GC system is close to that in the crystal open structure (RMSD = 0.8 Å). The position of α-helix N in the simulated GG system is also fairly near that in the crystal open structure (RMSD = 1.6 Å). However, the α-helix N positions in the simulated CA and CC systems are far away from the corresponding open crystal locations. Therefore, the degree of opening for the four half-open systems can be described by the relation GC > GG > CC ≈ CA.
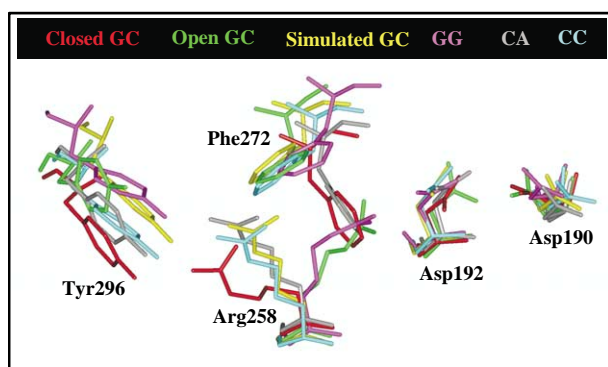
Figure 5. Conformational comparisons for active-site protein residues in the closed (red) and open (green) crystal structures, and in the simulated half-open GC (yellow), GG (pink), CA (gray), and CC (cyan) complexes. Note that the Arg258 rotation occurs only in the GG simulation.

Arg258 rotation occurs in the GG simulation

Crystallographic structures of pol β/DNA complexes[12] have indicated that many local structural transitions accompany the pol β closing and opening motions. Our earlier simulations[29] for the pol β/DNA complex with GC in the active site

(dynamics at room and high temperatures, as well as TMD simulations) have proposed a sequence of events during pol β opening:[29] first, the Phe272 phenyl ring flips; second, a large thumb movement occurs; and third, Arg258 rotates toward Asp192. The Arg258 rotation did not occur in the GC system at room temperature following the thumb motion, which suggested that the Arg258 rotation toward Asp192 is a slow step in the pol β opening process while thumb movement is relatively rapid. Indeed, our high-temperature simulations for the pol β/DNA complex[29] have suggested a relatively high energy barrier for the Arg258 rotation toward Asp192 following the breakage of the Arg258/Tyr296 hydrogen bond. Our hypothesis ties well with a considerable body of experimental data.[26–28] This sequence of events is consistent with our present DNA-mismatch simulations. The most interesting observation is that Arg258 rotates toward Asp192 in the GG simulation.

The conformations for the active-site residues of Asp190, Asp192, Arg258, Phe272, and Tyr296 in our four half-open structures after 6 ns simulation are displayed in Figure 5, along with those in the crystal closed and open structures. We see that the Phe272 phenyl ring flips in the simulations of GC,
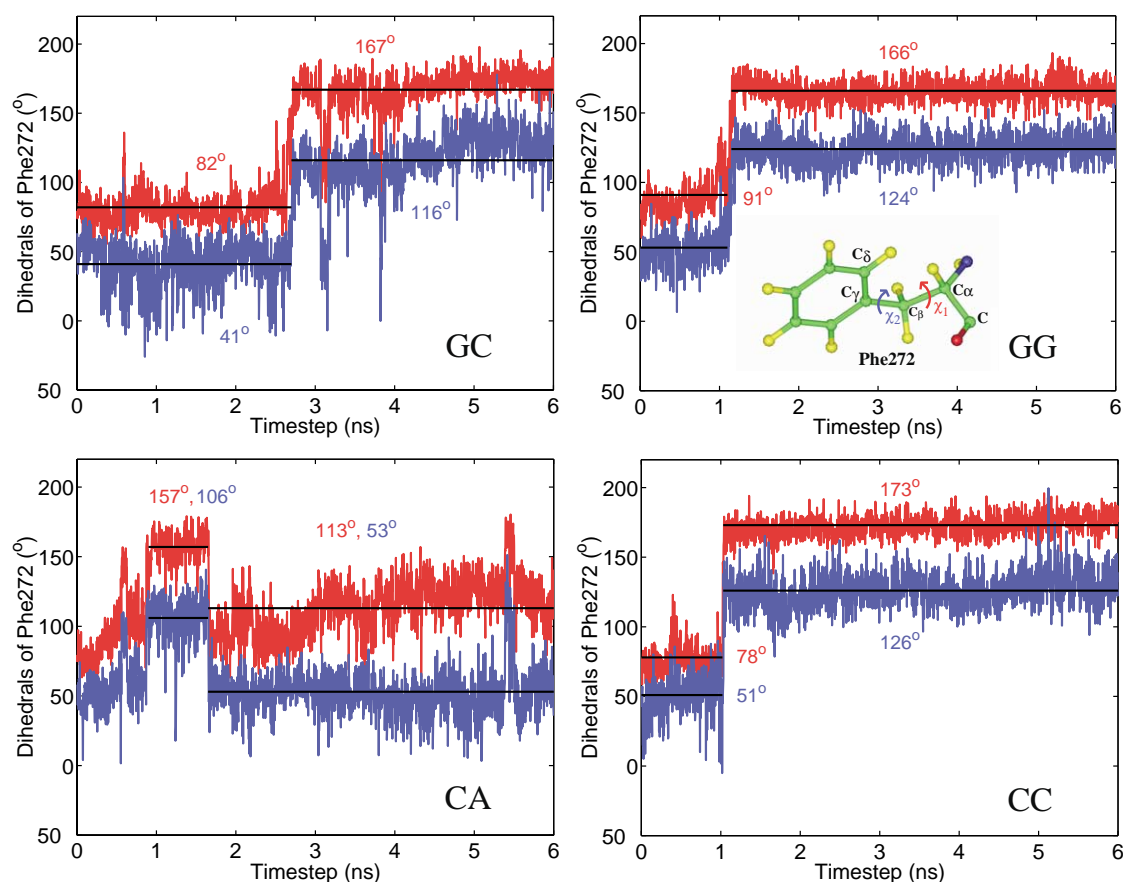

Figure 6. Time evolution of two dihedral angles associated with the Phe272 phenyl ring over the entire trajectory of the GC, GG, CA, and CC systems. The dihedral angles are defined by the following atomic sequence: χ_1 (red) for $C-C_\alpha-C_\beta-C_\gamma$ and χ_2 (blue) for $C_\alpha-C_\beta-C_\gamma-C_\delta$. Average values for the two dihedral angles during characteristic ranges are also given in all trajectories.
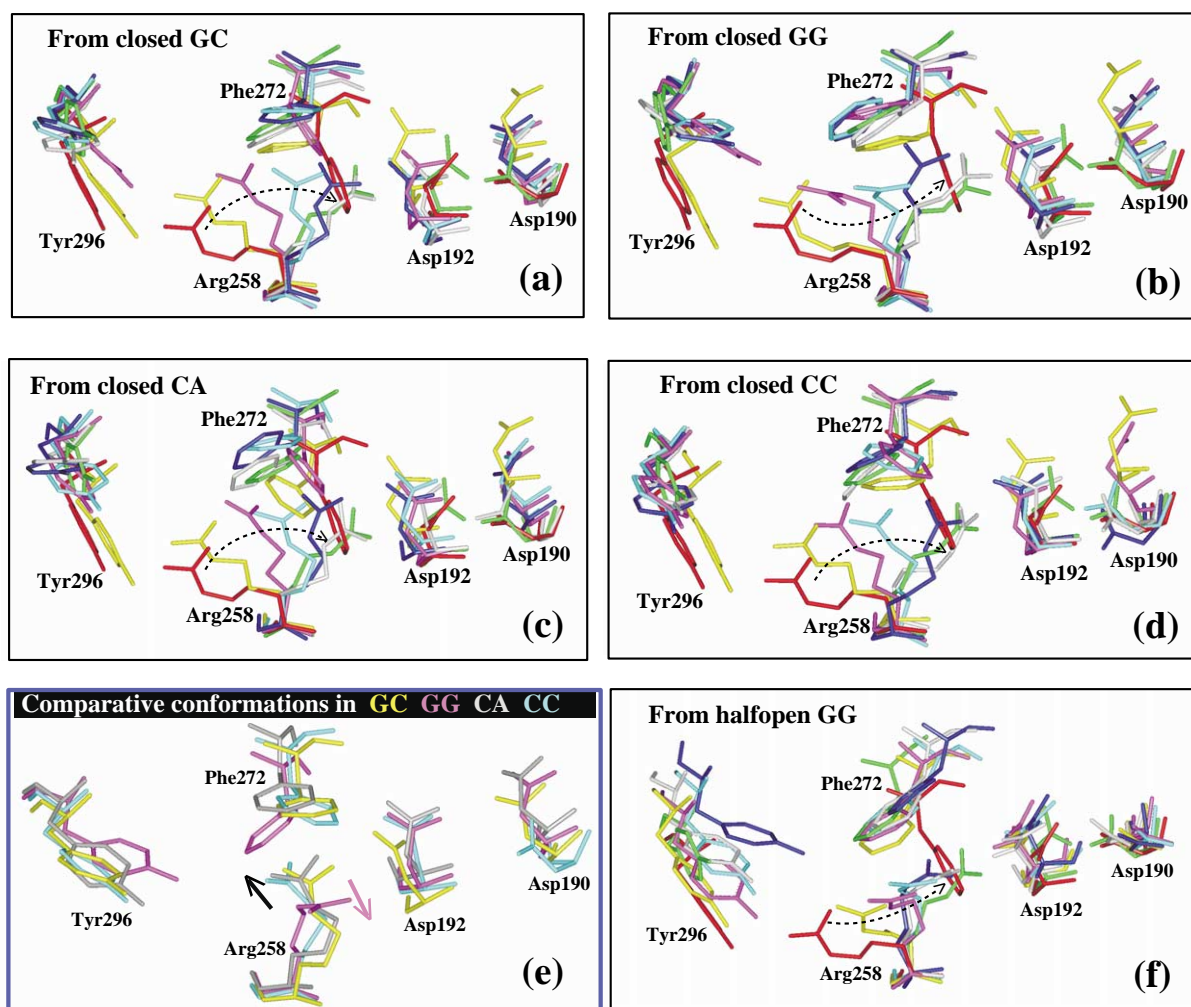
Figure 7. Possible pathways of Arg258 rotation in the TMD simulations of pol β/DNA complexes with nascent base-pairs (a) GC, (b) GG, (c) CA, and (d) CC from a closed to an open state, and (e) conformations of key active-site residues midway during the Arg258 rotation. (f) The possible pathway of Arg258 rotation in the standard (untargeted) molecular dynamics simulation of the GG is also presented. Note that the pink arrow in (e) indicates the amino groups' direction along the Arg258 rotation pathway in the GG complex, while the black arrow indicates their directions in the GC, CA, and CC complexes. The DNA minor groove would be situated below Arg258 and Phe272 in all Figures.

GG, and CC. However, in the CA trajectory, where the Phe272 ring does at first flip away from Asp192, it then flips back toward it again. This can be seen from the evolutions of the two dihedral angles in Phe272 shown in Figure 6: χ_1 for C–C_α–C_β–C_γ, indicating the rotation of the entire phenyl ring, and χ_2 for C_α–C_β–C_γ–C_δ, indicating the inclination of the ring plane. The two dihedral angles occupy two similar characteristic ranges as a function of time for the simulations of the GC, GG, and CC, but not for the CA simulation. We also see that Phe272 flips prior to the large thumb motion in both the GC and GG systems. However, Arg258 rotates toward Asp192 only in the GG simulation, and this is accompanied by rotation of Asp192 to form a salt bridge with Arg258. The resulting distance between Arg258:NH1 and Asp192:O$^{\delta 1}$ is 2.67 Å by the end of the dynamics trajectory. The Arg258 rotation follows, or couples with, the thumb movement in the GG trajectory.

TMD simulations suggest a unique Arg258 rotation pathway in the GG complex

Aside from the free dynamics simulations described above, several 100 ps TMD simulations[5–9,30] were performed to investigate possible pathways from a closed (rather than half-open as in the standard simulations) to an open state for pol β/DNA complexes, with our three mispairs at the polymerase active site (see Computational Methodology for TMD simulation algorithm). For comparison, TMD simulations for the pol β/DNA complex with the correct GC base-pair were also performed. Closed enzyme systems were constructed on the basis of a modified crystal closed, ternary complex, in which the incoming dGTP was reacted with the primer 3'-end and the product pyrophosphate was left in place. Then, the correct GC nascent base-pair was replaced by GG, CA, and CC mispairs to represent
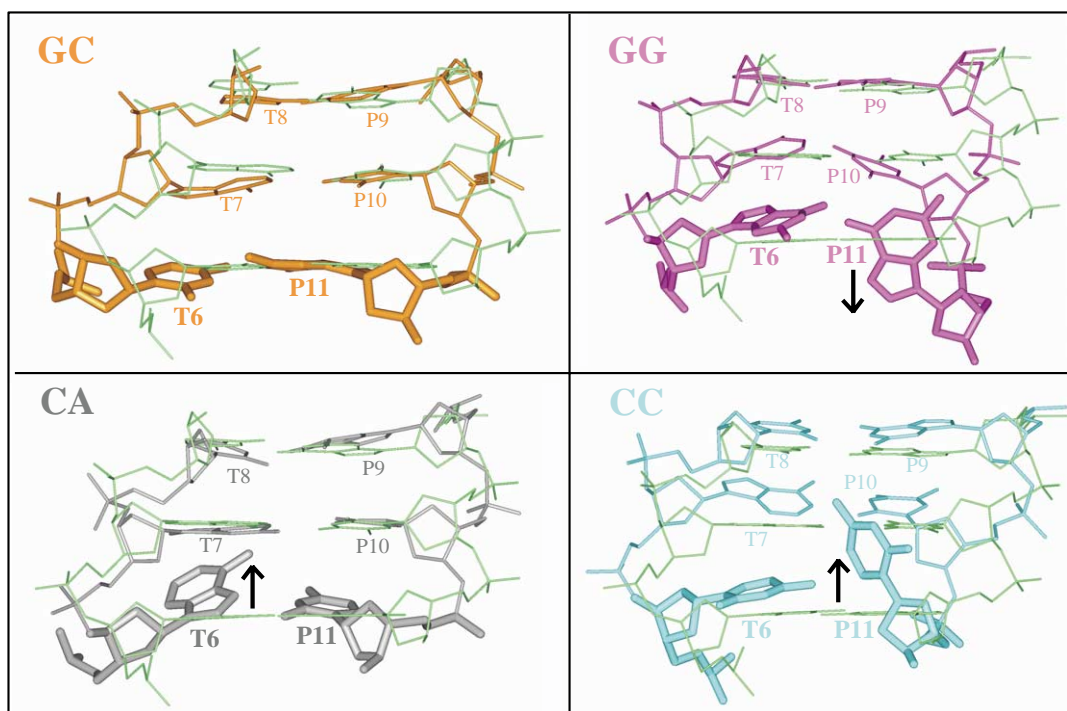
Figure 8. Conformational comparison of active-site base-pairs and their two neighboring base-pairs in the open crystal structure (green) and the simulated GC (orange), GG (pink), CA (gray), and CC (cyan) complexes of pol β/DNA. The nascent base-pair (P11-T6) plane in the open crystal complex is used for reference. The arrows (black) indicate the flipping directions of P11 in both GG and CC and of T6 in CA, with respect to the reference base-pair plane in the open crystal structure. The DNA minor groove is nearest the viewer, whereas the major groove is farthest from the viewer.

comparable systems. The target structure for the TMD simulations is the open binary nick pol β/DNA complex.

The TMD pathways for Arg258 rotation from closed to open states in the four pol β/DNA complexes are illustrated in Figure 7. In all systems, the Phe272 ring has flipped away from Asp192 before the thumb subdomain movement and Arg258 rotation, consistent with our previous TMD simulation results for the pol β/DNA complex.[29] We observed that the Arg258 rotation always follows the large thumb motion, occurring at the end of the TMD simulations, associated with a high potential energy barrier. TMD simulations of the GG system suggest a somewhat different pathway for Arg258 rotation: Arg258 in the GG complex rotates from its hydrogen bonding with Tyr296 to form a salt-bridge with Asp192 in such a way that the Arg258 amino groups point toward the primer/template DNA minor groove; in the other TMD simulations, the Arg258 rotates toward Asp192 with the Arg258 amino groups directed away from the DNA minor groove (see Figure 7(a)–(e)).

The TMD simulations indicate that the Arg258 rotation is easier in the GG complex than in the other complexes; this is on the basis of the smaller force constant required for the motion. Furthermore, the TMD simulation pathway of Arg258 rotation in the GG system starting from a closed state is similar to that obtained from the standard (untargeted) molecular dynamics simulation for the half-open GG (see Figure 7(b) and (f)). This different pathway of Arg258 rotation in the GG complex may change the geometry of the binding pocket in the active site, which would further affect the binding of the incoming nucleotide in the next cycle of DNA incorporation, namely, the DNA extension step. See Discussion for further details.

GG mispair geometry is most distorted, followed by CC and CA

In Figure 8, we superimpose the nascent base-pairs and their two neighboring primer/template DNA base-pairs in the simulated half-open GC, GG, CA, and CC systems with the corresponding base-pairs in the open crystal complex with nicked DNA,[12] according to the palm C^α atoms. The DNA base-pairs in the open crystal pol β/DNA complex are chosen as a reference to check the distortion of nascent base-pairs in the simulated half-open structures. The nascent base-pair in the simulated GC system is similar to that in the crystal open structure, with the two bases of the GC pair coplanar, maintaining Watson–Crick hydrogen bonding throughout the 6 ns simulation. The primer 3′-end base (P11) of G in the GG pair and of C in the CC pair both have flipped ∼90° with respect to the reference crystal structure base-pair

plane, and the two bases of GG or CC mispairs are almost perpendicular to each other. However, the flipping directions of P11 in GG and CC are somewhat different: the G has flipped toward the primer $3'$ terminus with its five-membered ring below the crystal structure reference base-pair plane and its mismatch template partner G (T6), while the C ring has flipped away from the primer $3'$ terminus with the P11 base above the crystal reference plane and its mismatch template partner C (T6) (flipping directions shown by arrows in Figure 8). The template base A (T6) of the nascent CA mispair also flips somewhat with respect to the crystal structure reference plane. The dihedral angle between the two bases of the CA mispair is $155°$ after the simulation, indicating that the CA bases are closer to coplanar than the two bases in the GG and CC mispairs.

The nascent base-pairs in all four half-open systems have slid toward the DNA major groove to varying degrees, relative to the corresponding base-pair in the crystal open complex. For the primer $3'$-terminal residues, the hierarchy of the magnitude of the movement is GG > CC > CA > GC, with associated P–P distances of 7.37 Å, 6.80 Å, 3.51 Å, and 2.80 Å, respectively, between the primer $3'$-terminal residue P11 in the half-open system and the crystal structure of the open complex. For the partner template residues, the ranking of motions toward the major groove is CA > GG > GC > CC, with corresponding P–P distances of 4.80 Å, 3.86 Å, 3.51 Å, and 1.77 Å, respectively, between the template residue in each system and in the crystal structure. These measurements suggest that the primer $3'$-terminal residues are generally more flexible than their partner template residues. The $C1'$–$C1'$ distances between the two residues of the nascent base-pairs are 10.79 Å, 12.18 Å, 9.45 Å, and 11.39 Å in the GC, GG, CA, and CC systems at the simulation's end, respectively. For reference, the $C1'$–$C1'$ distance of P11-T6 in the crystal open structure is 10.56 Å. It is clear that the $C1'$–$C1'$ distance in the GG complex deviates significantly from that in the crystal structure of the open complex; this further indicates a large distortion of the GG mispair.

In sum, the nascent base-pairs in the GG and CC systems are severely distorted, especially GG. Such distortions would be expected to significantly affect the efficiency of nucleotide insertion in the next cycle of DNA extension.

Motions of active-site residues during pol β opening

We now analyze motions of active-site residues during pol β's opening to help interpret the differences in the sequence of events for the various mismatch systems. The positions of active-site protein/DNA residues in both the crystal closed and open pol β/DNA complex are shown in Figure 1(b) and (c). For reference, in the GC complex, hydrogen bonds form early between Glu295

and P11/Gua (Glu295:O^ϵ/Gua:N2H) and between Glu295 and Tyr271 (Glu295:O^ϵ/Tyr271:OH). The hydrogen bond between Arg258:NH2 and Tyr296:OH forms prior to the Phe272 flip (see Figure 6) and is maintained thereafter. Watson–Crick hydrogen bonds between the nascent GC base-pair are maintained throughout the simulation.

In the GG simulation, we observe early hydrogen bonds between Glu295 (both $O^{\epsilon 1}$ and $O^{\epsilon 2}$ atoms) and the primer Gua:N1H as well as template Gua:N2H; the former hydrogen bond breaks after 3.3 ns but the latter is maintained throughout the trajectory. The hydrogen bond between Arg258:NH2 and Tyr296:OH forms early but breaks soon thereafter. Prior to the Phe272 flip (Figure 6), Tyr271:OH moves to form a hydrogen bond with Glu295:O, which is maintained during the remainder of the simulation. Following the Phe272 flip, for GG only we observe interactions between the N2H group of the P11/Gua with both the Phe272 ring and the Arg258:NH2 groups (distances of around 3.5 Å). By the end of the GG simulation, Arg258 has rotated toward Asp192 and formed an Arg258/Asp192 salt-bridge.

In the CA simulation, Tyr271:OH moves to form a hydrogen bond with Glu295:O^ϵ early in the simulation, Phe272 flips temporarily, and subsequently Tyr271:OH moves to hydrogen bond with primer Cyt:O2. The hydrogen bond between Arg258 and Tyr296 does not form during the entire simulation. Arg258 moves to interact with Tyr271 during certain intervals. We do not observe interactions between Glu295 and the nascent base-pair throughout the CA simulation. The absence of such interactions may provide the freedom for the glycosydic torsion angle of dA in the nascent CA base-pair to flip from the *anti* to the *syn* conformation, as we observe at the end of the simulation (at ~5.4 ns) (see Figure S1, Supplementary Material). A movement of this type in purine residues is possible, since purine bases can adopt the *syn* domain.[31] In the GC, GG, and CC simulations, the template and primer nucleotides preserve the *anti* conformation ($\chi \approx -104°$).

Finally, in the CC system, a hydrogen bond between Arg258:NH2 and Tyr296:OH is maintained throughout the simulation. Tyr271 hydrogen bonds with Glu295 (Tyr271:OH/Glu295:O^ϵ) but breaks after about 2.8 ns. Tyr271 also moves to make van der Waals contacts with the distorted P11/Cyt. After Tyr271 flips, Phe272 interacts with Arg258. By the end of the CC simulation, P11/Cyt flips about 90° relative to its template partner, as seen in Figure 8, and it somewhat approaches Arg258.

In addition, a hydrogen bond between Tyr271:O and P11:O3'H is formed in all simulations; this interaction may help stabilize the Tyr271 position prior to the Phe272 flip. Following the Phe272 flip, this hydrogen bond is not necessarily required to stabilize Tyr271; indeed it breaks at the end of the GG simulation because the bulky P11/Gua

becomes distorted to approach Arg258. During the GC and CA simulations, Arg254 rotates very early to change the hydrogen bonding with Asp256 to hydrogen bond with Asp190; however, in the GG and CC simulations, Arg254 hydrogen bonds with both Asp256 and Asp190 simultaneously.

We also analyze the interactions between the key protein residue Arg258 and the newly incorporated primer nucleotide P11, and between Arg258 and its contacting protein residues (Figure S2). The electrostatic and van der Waals interaction energy for Arg258 and P11 is the highest in GG and lowest in CC; in GC and CA, the values are around zero. The high interaction energy in GG indicates a stronger repulsive interaction between P11 and Arg258, which possibly triggers the thumb's opening and the Arg258 rotation for the GG complex. The lowest interaction energies between Arg258 and P11 and between Arg258 and its adjacent protein residues in the CC system may explain in part the stability of the thumb in the CC simulation, but movement to an open state in the GC simulation.

Tyr271, Glu295, Asp276, and Lys280 contact the DNA nascent base-pair during pol β opening

The interaction between polymerase and DNA bases, especially the hydrogen bonding between certain protein residues and the nascent base-pair, can be important for the polymerase to maintain DNA synthesis fidelity.[16] Crystal structures of the closed pol β/DNA complex[12,32] indicate that several protein residues are in contact with the DNA bases in the minor groove: Lys234 and Tyr271 hydrogen bond to the O2 of a template cytidine and the O2 of a primer cytidine, respectively; Arg283 of α-helix N hydrogen bonds to the sugar O4' in template T7; Asn279 hydrogen bonds to the minor groove edge (N3 of purine bases and O2 of pyrimidine bases) of the incoming nucleoside triphosphate. These protein residues may play a key role in disrupting the spine of well-ordered water molecules normally found in the *B*-DNA minor groove, thereby stabilizing the somewhat *A*-like DNA conformation at the pol β active site.[32] Asp276 and Lys280 are observed to make van der Waals contact with the bases of the incoming nucleotide and templating base, respectively, in the crystal structures of the closed ternary complex of pol β.[12,26,32,33] In addition, Asp276 is observed to form a salt-bridge with Arg40 of the N-terminal lyase domain.[26] Site-directed mutagenesis of Asp276 has indicated that there is an increase in the binding affinity for the incoming nucleotide with valine substitution.[26] In contrast, mutagenesis of Lys280 indicates that there is a loss in binding affinity for the incoming nucleotide with smaller residue 280 side-chains. Interestingly, this loss was more pronounced for templating purine bases than pyrimidine bases.[33] Substitutions

of protein residues Tyr271, Asn279, and Arg283 influence the catalytic efficiency and fidelity of pol β.[3,34–37]

To determine protein residues which may be involved in the pol β opening motion, and possibly also in pol β closing, we analyze in Figure S3 all protein residues contacting (within 3.5 Å) the nascent base-pair and its neighboring pair during the simulations of all four half-open systems. We also monitor the time evolution of the corresponding distances between protein residues and DNA base-pairs.

We note that in all simulations a greater number of protein residues contact the newly incorporated primer base (P11) relative to the partner template base (T6). Tyr271 remains in close contact with P11, forming the hydrogen bond Tyr271:O/P11:O3'H throughout all trajectories, except that for GG Tyr271 moves away from P11 at the end of the simulation. Furthermore, we observe other hydrogen bonds between Tyr271 and P11, namely Tyr271:OH/P11:N2 in the GC system, Tyr271:OH/P11:O2 in the CA system, and Tyr271:OH/P11:N1 in the CC simulation. A hydrogen bond between Asn279:H$^\delta$ and P11:O3' also forms in the CA and CC simulations, while Asn279 moves away from P11 in the GC and GG simulations. Lys280 contacts P11 and/or T6 at certain periods of the GC, CA, and CC simulations and then moves away by the end of the trajectories. Glu295 contacts P11 in the GC and T6 in the GG simulations while it interacts with P11 only initially in the CA and CC trajectories and subsequently moves away. Asp276 contacts P11 throughout the CA simulation but moves away from P11 in the GC simulation; this movement is associated with the thumb opening. We also find that Arg283, which contacts the template base in the closed crystal structure,[12,32] is out of the hydrogen bond range with the nascent base-pairs in most simulations; it only contacts the template in the nascent base-pair at the end of the CC trajectory. This trend suggests that Arg283 may not be involved in the pol β opening even though it is most likely important in closing; in the closed conformation, Arg283 hydrogen bonds with the T7 template sugar and possibly Glu295.[12]

Furthermore, residues Phe272, Thr273, Gly274, and Ser275 near the α-helix N remain close to P11 in the CC and CA simulations, especially for CC; these residues move away from P11 in the GC and GG simulations. These differences are significant because the relatively large motions of α-helix N captured in the GC and GG simulations are correlated to the thumb opening.

Finally, we find that the primer nucleotide P10, adjacent to the newly incorporated P11, forms a hydrogen bond with protein residue Arg254 throughout the GC and CA simulations. In the GG and CC simulations, hydrogen bond Arg254:NH2/P10:O2P breaks early. The breakage of Arg254/P10 hydrogen bond in the GG (Pur:Pur) and CC (Pyr:Pyr) systems and not GC (Pur:Pyr) nor CA (Pyr:Pur) systems suggests a

discriminatory role for Arg254 with regards to the size of the nascent base-pair through interactions with the residues adjacent to the incoming nucleotide.

In sum, protein residues Tyr271, Asp276, Asn279, Lys280, and Glu295 make close contact with the DNA nascent base-pair. They may thus play an important role in the pol β opening process.

A sodium ion coordinated with active-site residues may be a surrogate for the catalytic Mg²⁺ in the GC and CA simulations

In the GC simulation, a neutralizing Na^+ counterion was observed to coordinate with the active-site carboxylate groups Asp190, Asp192, and Asp256, which is similar to our previous observation in another GC trajectory.[29] This sodium ion coordinates Asp190 at the start of the simulation and then coordinates both Asp192 and Asp256; in the second half of the simulation, it only coordinates Asp256. In addition, three water molecules are also coordinated with this sodium ion throughout the simulation, as shown in Figure 9.

A sodium ion coordinates both Asp192 and Asp256, as well as two to four water molecules, throughout the CA simulation, which is similar to the sodium ion coordination pattern in the GC simulation. However, in both the GG and CC trajectories, a sodium ion coordinates only Asp190, as well as two to four water molecules in GG and two to five water molecules in CC systems. The water molecules coordinating the sodium ions are non-specific and exchange rapidly with other water molecules in the simulation.

The sodium ion position in the GC and CA simulations is similar to that occupied by the specific catalytic Mg^{2+} (metal B), which coordinates with the active-site residues of Asp190, Asp192, and Asp256 in the closed, ternary pol β/DNA complex.[12] The catalytic Mg^{2+} plays an important role in facilitating nucleotidyl transfer.[12,13,20,27,32,38–41] The effects of the catalytic Mg^{2+} and the nucleotide-binding Mg^{2+} on the pol β opening are under study using molecular dynamics simulations for the pol β/DNA complex.

Active-site water-bridge patterns in GC and CA are different from those in GG and CC

A network of water molecules hydrogen bonded with the key protein residues Asp190, Asp192, Arg254, and Asp256 throughout all the simulations is shown in Figure 10. However, the water networks are somewhat different in the four systems. For GC, a network of three to four water molecules connect the four protein residues like a chain, and there is no water-bridge between Asp256 and Arg254. For GG, seven to ten water molecules form a hydrogen bond network with the above four residues and the rotated Arg258. In the CA trajectory, the water network only involves two

water molecules with sequential hydrogen bonding with Asp190, Asp192, and Asp256; this is similar to the water network we observe for GC. The pattern of the water network for CC is similar to that seen in the GG system, with both having one water molecule hydrogen bonded to Asp192, Arg254, and Asp256. However, there are no water-bridges between Arg258 and Asp192 or between Arg258 and Asp256, since Arg258 has not rotated toward Asp192 in the CC trajectory. The most stable water molecule is the one that is hydrogen bonded to both Asp190 and Asp192 (by three hydrogen bonds), maintained throughout all simulations of the four half-open systems. The other water molecules in the network are not specific and exchange with solution water molecules.

In addition, there are transient water molecules hydrogen bonding with protein residues in the simulations. In the GC simulation, one water molecule occasionally hydrogen bonds with both Asp276 and Lys280 before thumb movement. For GG, one water molecule is hydrogen bonded with both Tyr271 and Glu295 after the Tyr271 ring flip, but before Arg258 rotation. A network of two to three water molecules connecting key protein residues Arg258, Tyr271, and Glu295 was observed in the CA simulation; Arg283 is directly hydrogen bonded with Glu295 throughout this simulation as shown in Figure 10(e). The CC trajectory reveals a network of two water molecules hydrogen bonding with the protein residues of Arg283, Tyr271, and Glu295 (see Figure 10(f)). This network stabilizes the structure of these three residues following the Tyr271 flip.

We also examined the water-bridges between protein residues and DNA bases. In the GC simulation, several protein residues such as Asp192, Glu295, and Asp276 hydrogen bond with primer bases P11/Gua and P10/Thy through several water molecules. During the GG simulation, Arg258 hydrogen bonds with P10/Thy through one water molecule before Arg258 rotation and then soon breaks; Glu295 interacts with the two template bases T6/Gua and T7/Ade, and with one primer base P11/Gua through three water-bridges. In the CC simulation, a water-bridge connecting Asp276 and P11/Cyt was observed, and the primer P11/Cyt base partially stacks with the Tyr271 ring due to the flip of the base. These transient water-bridges between the polymerase and DNA may contribute to local structural stability at certain points along the catalytic cycle.

We also find some short-lived water-bridges between DNA bases in the simulations. For example, primer P11/Gua interacts with template T7/Ade through a water-bridge early in the GG simulation; however, the water-bridge breaks soon after the P11/Gua base flips. Some temporary water-bridges forming hydrogen bonds with two primer DNA residues, P11 and P10, were observed in all simulations. However, there are no stable water-bridges between DNA bases in the active site of the pol β/DNA complex.
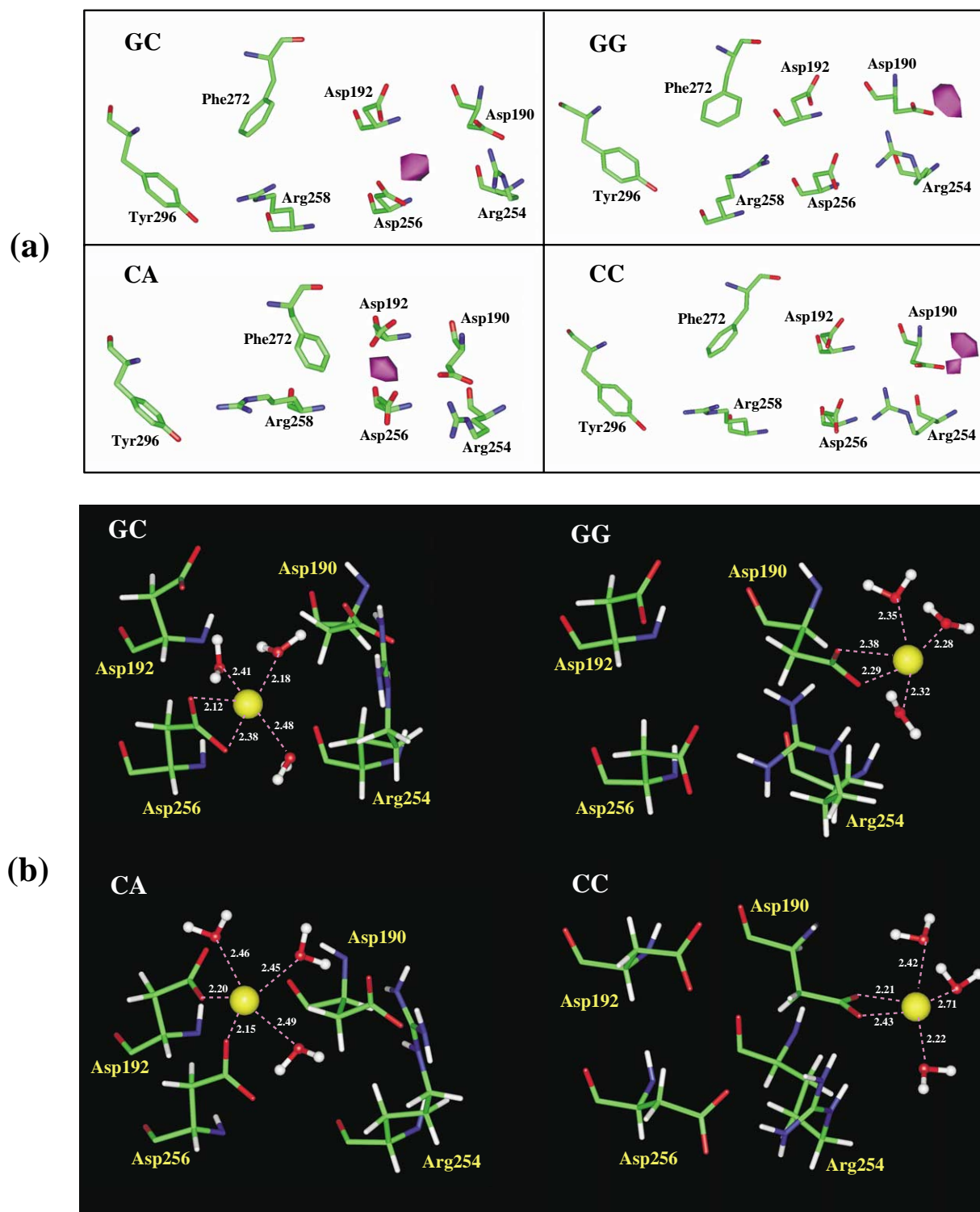
Figure 9. (a) Na⁺ probability density of >6% in the active site of the GC, GG, CA, and CC systems, respectively. The densities were accumulated on a 1 Å cubic lattice over 200 snapshots sampled at a frequency of 3 ps for each trajectory at the end of last 600 ps. (b) Sodium ion coordination with key protein residues Asp190, Asp192, Asp256 and several water molecules in the active site during each of the four half-open trajectories.
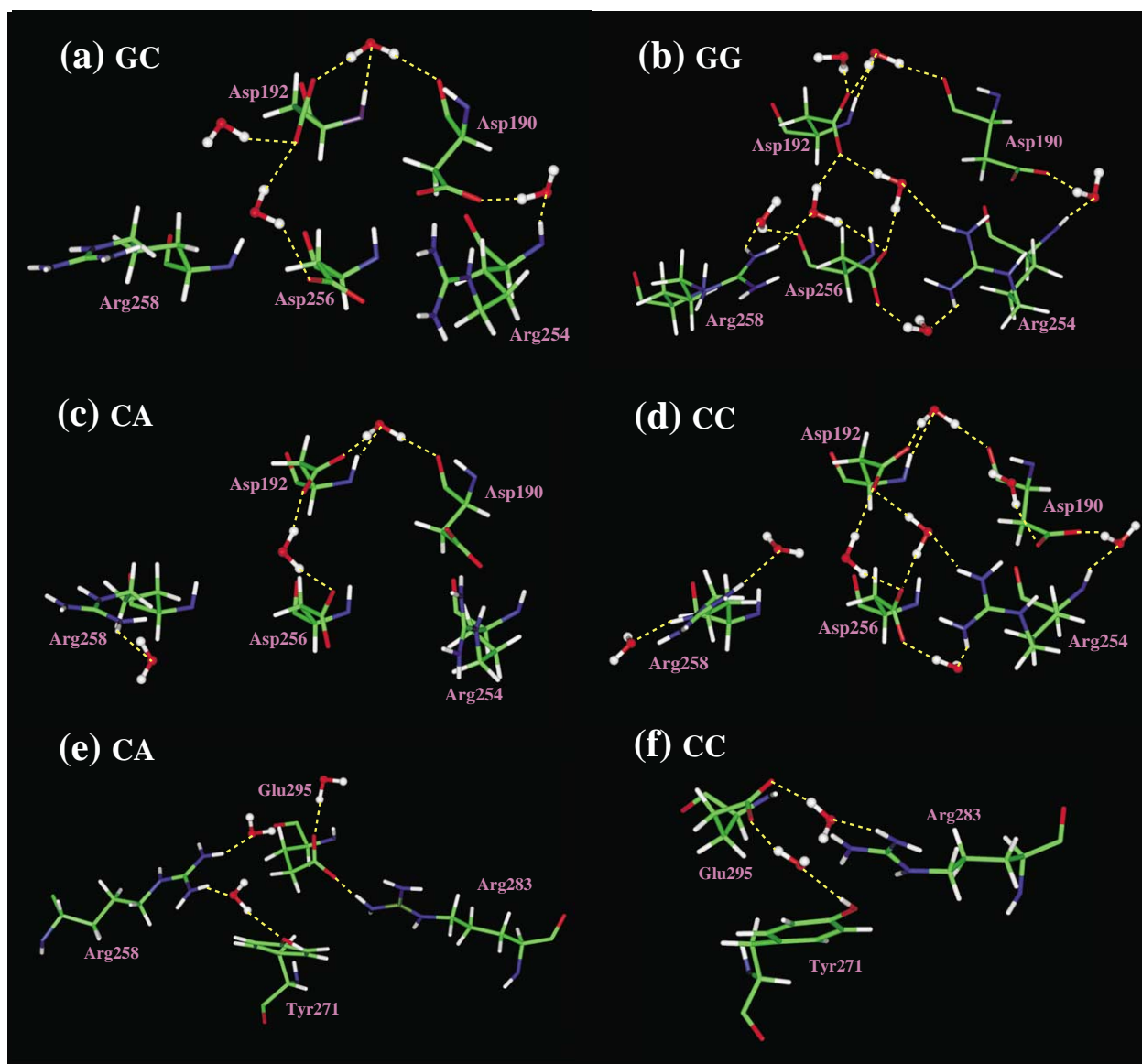
Figure 10. A network of water molecules that are hydrogen bonded with key protein residues Asp190, Asp192, Arg254, Asp256, and Arg258 during the (a) GC, (b) GG, (c) CA, and (d) CC trajectories. Another water network connecting with Arg258, Tyr271, Arg283, and Glu295 in the simulations of the (e) CA and (f) CC trajectories is also presented.

In sum, the long-lived water bridges connecting polymerase residues might help stabilize the active-site structure during the thumb opening.

Discussion

Different active-site motions help explain how DNA mismatches influence polymerase opening

The strategies utilized by DNA polymerases to select the correct nucleotide from a pool of structurally similar molecules are poorly understood. It is generally believed that geometric constraints imposed by the polymerase active site are critically important.[10] X-ray crystal structures of DNA polymerases in various liganded states[12,13,23,32,42–49] reveal several protein and nucleic acid structural transitions during the binding of the correct nucleotide to a polymerase/DNA complex. Kinetic approaches to study nucleotide insertion and fidelity[25,26,28,29,34,35,50–58] have indicated that rate-limiting conformational changes occur before and after chemistry. This suggests an induced-fit mechanism for nucleotide insertion and fidelity. The identity of the proposed conformational changes is unknown, but recent kinetic,[26] structural,[27] and modeling[29] studies with pol β indicate that the source is not the large subdomain movements inferred from structural studies but rather subtle side-chain movements (e.g. Arg258)[29] and/or DNA conformational transitions that could limit nucleotide insertion. Furthermore, we emphasize that, although

rate-limiting conformational changes are usually suggested to limit correct and incorrect nucleotide insertion, the identity of these structural transitions need not be the same for these matched *versus* mismatched insertions.

Our dynamics simulations for pol β/DNA complexes with mismatched base-pairs at the polymerase active-site detail structural transitions that may occur during subdomain opening induced by misincorporations. The simulations for the pol β/DNA complex with the correct nascent GC base-pair suggested the following sequence of events during pol β's opening:[29] Phe272 ring flip, a large thumb movement, and Arg258 rotation toward Asp192. During the GC, GG, and CC simulations, the flip of the Phe272 phenyl ring indeed occurs prior to thumb movement. In the CA simulation, the Phe272 flips early but then flips back toward Asp192. A large thumb motion following the Phe272 ring flip only occurs in the GC and GG simulations, while the thumb remains stable in the CA and CC trajectories. Although a large thumb motion occurs in the GG simulation, the simulated open pol β in the GG system is different from the open crystal structure; in the GC system, the simulated open state is close to the open crystal structure. The TMD simulations from a closed to an open state suggest that the pathway for Arg258 rotation in the GG complex is different from that in the GC, CA, and CC complexes.

Analyses of the motions of the protein and DNA residues in the active site for the matched and mismatched simulations help explain differing sequences of events during pol β's opening. In the crystal closed pol β/DNA complex,[12] both residues Glu295 and Tyr296 hydrogen bond with Arg258, and Phe272 is positioned between Arg258 and Asp192 (see Figure 1(b)). Tyr271 hydrogen bonds with P10, the adjacent base to the incoming nucleotide. In the crystal open nick structure, Phe272 has flipped away from Asp192, and the Arg258 has rotated and is poised to form a salt-bridge with Asp192 (see Figure 1(c)). Asp192 does not rotate toward Arg258 in the crystal open nick complex, whereas it does in the crystal open gapped complex.[12] In addition, the Glu295 carboxyl group moves away from Tyr296 and toward the DNA template, and Tyr271 moves to hydrogen bond with the newly incorporated nucleotide P11. In both the closed and open states, Arg254 and Asp256 form a salt-bridge. These motions of the active-site protein and DNA residues play an important role in the polymerase opening.

Our simulation analyses reveal the following structural features that usually occur prior to the Phe272 flip: (1) the hydrogen bond between Arg258 and Tyr296 forms; (2) the hydrogen bond between Arg258 and Glu295 breaks; and (3) the Tyr271 moves to interact with the newly incorporated primer base P11 and/or with Glu295. Most important is the movement of Tyr271, since it provides room for the Phe272 flip. This flip occurs in the GC, GG, and CC simulations, but only temporarily in the CA simulation. In the CA system, the hydrogen bond between Arg258 and Tyr296 does not form, and this leaves Arg258 free to interact with Tyr271. In addition, Tyr271 does not move sufficiently to form a stable hydrogen bond with P11 early in the simulation. These interactions of Tyr271 with Arg258 and Glu295 may lock the Tyr271 to prevent a complete Phe272 flip. Since we suggested a sequence of events for pol β's opening where the Phe272 flip occurs first, the absence of this motion for the CA system may impede the thumb opening and Arg258 rotation in that mismatched simulation. In the CC system, the strong, favorable interaction energy between Arg258 and the binding pocket residues (Figure S2) might stabilize the thumb subdomain and prevent its opening. In the GG system, the interactions of P11 with both Arg258 and Phe272 may activate the binding pocket to accelerate the polymerase opening as well as Arg258 rotation. However, the distortion we observe in the GG base-pair and the unique Arg258 rotation pathway, both in the standard and TMD simulations, explain the incomplete opening in GG.

Finally, we emphasize recent findings[26,29] suggesting that the movement of the thumb to an open or closed structure is likely rapid and not rate limiting. The observation here that the thumb subdomain adopts an open form for the GG and GC systems but a stable intermediate conformation in the CA and CC simulations might imply that the free energy difference and barrier between the open and closed states could be small, consistent with structural characterizations of the molecular contacts observed in the open and closed forms.[44]

Distorted mismatched DNA base-pairs are hard to extend

Our simulations for the three GG, CA, and CC mispair systems reveal that the geometry of the GG is the most distorted, with CC second, and CA being the least distorted. Such geometric distortions of the nascent mispairs likely result from the steric rearrangements undertaken to accommodate the mispair in the active-site binding pocket. (The binding pocket includes the polymerase, the corresponding template base, and the incoming nucleotides.) For Pur:Pur base-pairs (Pur and Pyr refer to purine and pyrimidine bases, respectively), the incorporated purine base is too large to be accommodated by the binding pocket, which thus leads to a base flip of the primer. For Pyr:Pyr base-pairs, the primer pyrimidine base cannot tightly approach the template pyrimidine base, which also results in the flipping of the primer base due to its freedom in the binding pocket. For Pur:Pyr (or Pyr:Pur) base-pairs, the primer bases can be accommodated by the binding pocket, since the geometries of the nascent mispairs are similar to those of the normal Watson–Crick base-pairs. However, the geometries of the Pur:Pyr (or Pyr:Pur) mispairs can become distorted due to the

absence of Watson–Crick hydrogen bonds; this could cause the flipping of the template base in the CA mispair system. In general, the mismatched Pur:Pyr (or Pyr:Pur) base-pairs are less geometrically distorted than the Pyr:Pyr and Pur:Pur base-pairs compared to the orientation of Watson–Crick base-pairs. The structural analyses for the nascent base-pairs is consistent with the geometric selection model proposed by Echols & Goodman.[10] In this model, shape complementarity of the nascent base-pair is more important than the ability to form Watson–Crick hydrogen bonds. The ability of several DNA polymerases to efficiently insert non-polar isosteres is consistent with this model.[59]

Kinetic analyses of the extension of the 12 possible mispairs by pol β indicate that, in general, Pur:Pyr mispairs (e.g. CA) are the easiest to extend, whereas Pur:Pur (e.g. GG) the most difficult (Shock *et al.*, unpublished results). These extension efficiencies parallel the degree of distortion found in the polymerase active site during the simulations. Steady-state kinetic data for avian myeloblastosis virus reverse transcriptase,[60] which is considered to have a fidelity similar to that of pol β, also indicate a similar hierarchy for the extension of mismatched DNA primer termini: Pur:Pyr and Pyr:Pur > Pyr:Pyr > Pur:Pur; in particular, the GG mispair is the most difficult to extend and has a significantly lower relative efficiency of extension.[60] Combined with our structural analyses, we suggest that the greater distortions impede extension. Thus, the dramatic distortion observed in the GG and CC mismatches (flipped bases and displacement into the major groove) correlates with the weak ability of pol β to extend these mispairs. The coplanar nature of the CA mispair and its smaller displacement into the major groove (relative to the GG and CC mispairs) indicate that CA might be extended more easily than GG and CC, but not as well as a correctly matched terminus (e.g. GC here).

Distorted DNA mismatches would cause inefficient extension leading to extrinsic proofreading

DNA polymerases must synthesize DNA with high fidelity so that deleterious mutations are not introduced into the genome. Many DNA polymerases have an associated, intrinsic or extrinsic, proofreading exonuclease that excises misincorporated nucleotides. In some instances, the polymerase may continue DNA synthesis (i.e. extend the mispair) resulting in a "buried" mismatch in the DNA. For mismatches that escape proofreading, the cell has evolved a post-replication mismatch repair system to correct replication errors.

For a base-substitution error to occur, a polymerase insertion error must also be extended. This kinetically challenging event for the polymerase provides an additional opportunity for an intrinsic proofreading exonuclease to edit the error and enhance overall fidelity. When this occurs, the mispaired primer terminus is transferred to the exo-

nuclease active site. The molecular events that manage the partitioning of the primer terminus between the polymerase and exonuclease active sites are unclear but are believed to involve melting of the primer terminus. Following exonuclease cleavage, the primer terminus returns to the polymerase active site for further extension. In some cases for proofreading to occur, the unpaired primer terminus (DNA product) must completely dissociate from the polymerase to rebind to the intrinsic exonuclease active site, to a separate sub-unit, or to an accessory protein. During pol β-dependent base excision repair, recent evidence suggests that polymerase errors are proofread by apurinic/apyrimidinic (AP) endonuclease.[61]

Since pol β does not have an intrinsic proofreading activity, the distorted mispaired terminus could have an effect on proofreading by hastening DNA dissociation after misinsertion; this would allow an extrinsic enzyme to edit the mismatch, and/or to inhibit further DNA synthesis (i.e. extension). It is believed that the latter mechanism (extension inhibition) is more important, since polymerases bind paired or mispaired primer termini DNA with similar affinities.[62] This implies that the DNA dissociation rate constant is not affected by structural abnormalities at the polymerase active site. Thus, the positions of the 3′-OH of the deoxyribose of the primer terminus (P11) and the catalytic metal are crucial determinants of whether continued synthesis would occur.

Indeed, DNA pol β extends the mispairs inefficiently (i.e. correct insertion onto a mismatch is inefficient).[63] Kinetic analyses of the extension of all 12 possible mispairs by pol β indicate that the rate of incorporation of the correct dNTP onto the mismatch is dramatically reduced in all cases. According to our mismatch simulations, these inefficient mispair extensions result from the distortion of the active site, including the abnormal position of the primer 3′-OH terminus. The degrees of active-site distortion in the mismatched systems agree well with the corresponding extension inefficiencies. This distortion effect is relatively localized, with small structural perturbations occurring over two to three base-pairs from the primer terminus. For a DNA polymerase that possesses intrinsic exonuclease activity, such local deformations associated with a mismatch at the polymerase active site may facilitate melting of several duplex nucleotides and thereby hasten intramolecular transfer to the exonuclease active site.

Therefore, our modeling efforts shed light on the structural (geometric, steric, and electrostatic) consequences of the presence of a DNA mispair rather than a Watson–Crick base-pair in the pol β active site. For continued DNA synthesis to occur, the primer terminus (matched or mismatched) and critical polymerase side-chains must assume an appropriate geometric arrangement. If this cannot be achieved within some critical time interval, the product DNA would likely dissociate from the polymerase. A newly incorporated, mismatched

base-pair could be proofread by an apurinic/apyrimidinic endonuclease, whereas a matched base-pair (nicked DNA) could be ligated by a DNA ligase.

Specific polymerase–DNA interactions produce unique active-site geometry

The observation that key protein residues, such as Tyr271, Glu295, Asp276, and Lys280, interact with the newly formed mismatched base-pairs highlights the dynamic nature of the polymerase active site. It suggests that the geometric characteristics of each mispair will depend not only on the local DNA structure, but more importantly on its local protein environment. In this context, it is therefore not surprising that the opening motion of the thumb subdomain is influenced by the nature of the mispair situated in the active site. Since formation of a mispair will depend on geometric, steric, and electrostatic characteristics of each mispair, it is expected that the insertion of an incorrect nucleotide will also depend on local protein–DNA interactions that will be specific for each mispair. This, in turn, would be predicted to alter the specific nature of the closing motion of the thumb subdomain of pol β. These suggestions are supported by the observation that the spectrum of errors observed for DNA polymerases (misinsertions and extensions) are highly dependent on the identity of the polymerase where the local protein environment will be unique in each case. Even for a given DNA polymerase, the energetic contribution of a specific interaction at the active site is expected to be unique for formation of a base-pair (correct or incorrect). This idea is supported by the observation that removal of the Lys280 side-chain of pol β results in a loss of binding affinity for a correct incoming pyrimidine nucleotide, but not a purine triphosphate.[33]

Thus, geometric selection plays an important role in both selection of the correct dNTP as well as extension from the primer terminus. The observation that base-pair geometry of the primer terminus is critically important is not surprising, since it forms one face of the binding pocket for the nascent base-pair.[11] It remains to be determined what polymerase–DNA interactions (e.g. steric and/or electrostatic) with the primer terminal base-pair also influences insertion of the incoming nucleotide. More broadly, additional experimental and computational work, including behavior of mutant enzymes (L.Y. *et al.*, unpublished results), should provide further insight into these fundamental processes of DNA synthesis and repair fidelity.

Computational Methodology

Systems setup

The initial, solvated half-open model for human pol β complexed with primer/template DNA was constructed by using the PDB/RCSB[64] coordinates entries 1BPY (closed, ternary complex) and 1BPZ (open, binary product complex). The missing protein residues in the crystal structures were added by using the INSIGHT II package, version 2000. CHARMM subroutine HBUILD[65] was employed to add all hydrogen atoms to the crystallographic heavy atoms.

On the basis of the half-open model constructed in our initial studies for pol β/DNA complex after dNTP had been incorporated,[29] three complexes of pol β with mispairs GG, CA, and CC at the DNA extension site were built by replacing the correct nascent base-pair GC with the above three mismatched base-pairs, respectively, leaving protein residues and other DNA base sequences unchanged. The chemical product pyrophosphate (PP$_i$) and two specific magnesium ions in the active site were not considered in this work, since they are absent from the crystal open structures. The three mismatch half-open models were similarly solvated in a periodic domain of face-centered cube with smallest image distance of 16 Å. Then, the systems were neutralized at an ionic strength of 150 mM by replacing water molecules with minimal electrostatic potential at the oxygen atoms with Na$^+$ and those with maximal electrostatic potential at the oxygen atoms with Cl$^-$. The electrostatic potentials of all bulk water oxygen atoms were calculated with the DelPhi package.[66,67] The initial models of half-open pol β with GG, CA, and CC mispairs contain the same number of counterions (44 Na$^+$ and 21 Cl$^-$) and bulk water molecules (11,830).

Minimization, equilibration and dynamics protocol

Energy minimizations, equilibrations, and dynamics simulations for all systems were performed using the program CHARMM[68,69] and the all-atom version 26a2 force field (Chemistry Department, Harvard University, Cambridge, MA). Each system was minimized using the method of steepest descent for 10,000 steps followed by that of adapted basis Newton–Raphson[68,70] for 20,000 steps. Then, each system was equilibrated for 50 ps at room temperature by using the stochastic LN integrator[71–74] before the dynamics production began.

The multiple timestep LN integrator[71–73] for Langevin dynamics is so called for its origin in a Langevin/normal modes scheme. The current LN protocol applied to the simulations of large biological systems uses three timesteps $\Delta\tau$, Δt_m, and Δt. The short timestep cycle updates the bond, angle, and dihedral energy terms every $\Delta\tau$ interval. The medium timestep cycle is used for updating the non-bonded interactions within a spherical distance (here 7 Å and associated with heating and buffer lengths of 4 Å each) every Δt_m interval. The long timestep is for updating the remaining non-bonded interactions up to the global non-bonded interaction cutoff distance (14 Å here); this updating utilizes a

heuristic procedure if any atom moves more than 1 Å. Here we use the LN protocol of $\Delta\tau/\Delta t_m/\Delta t = 1/2/150$ fs. The stability and reliability of this LN protocol for large macromolecular systems have been demonstrated in term of thermodynamic, structural, and dynamic properties compared to single-timestep Langevin as well as Newtonian (velocity Verlet) integrators.[29] The SHAKE algorithm was employed in all runs to constrain the bonds involving hydrogen atoms. Electrostatic and van der Waals interactions are smoothed to zero at 12 Å with a shift function and a switch function, respectively. The Langevin collision parameter of $\gamma = 10\,\text{ps}^{-1}$ is chosen to couple the system to a 300 °C heat bath. The computational speedup for our large pol β/DNA systems is around fivefold with respect to the inner timesteps of 1 fs.[29] The NCSA Origin 2000 further provides a speedup factor of 8.7 on 16 processors relative to single processor runs.

RMSD definition

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

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

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.

TMD simulations algorithm

The TMD simulations are performed by introducing an additional constraint force in a normal MD simulation.[6,8,9,30,75] Our TMD simulations are performed using the academic version c28a2 of the CHARMM program (Chemistry at Harvard Molecular Mechanics),[68] modified to incorporate an energy restraint on the basis of the RMS distance with respect to a reference target conformation. The functional form of the RMS restraint energy is as follows:

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

where K is a force constant (in kcal/mol per Å²), 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). 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 structure. The conformational transition from a closed to an open state is driven by applying RMSD restraints to all heavy atoms except those in residues 1–9, which are missing from the crystal closed structure. The offset parameter of d_0 is decreased from 3.6 Å by 0.18 Å every 5 ps of the simulation until it reaches zero deviation.

Note that while clearly non-physical, TMD simulations can help analyze certain large-scale conformational dynamics by suggesting conformations and energy trends along the pathway.

References

1. Kunkel, T. A. & Bebenek, K. (2000). DNA replication fidelity. *Annu. Rev. Biochem.* **69**, 497–529.
2. Kunkel, T. A. & Alexander, P. S. (1986). The base substitution fidelity of eucaryotic DNA polymerases. *J. Biol. Chem.* **261**, 160–166.
3. Osheroff, W. P., Beard, W. A., Wilson, S. H. & Kunkel, T. A. (1999). Base substitution specificity of DNA polymerase β depends on interactions in the DNA minor groove. *J. Biol. Chem.* **274**, 20749–20752.
4. Friedberg, E. C., Fischhaber, P. L. & Kisker, C. (2001). Error-prone DNA polymerases: novel structures and the benefits of infidelity. *Cell*, **107**, 9–12.
5. 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.
6. Schlitter, J., Engels, M., Krüger, P., Jacoby, E. & Wollmer, A. (1993). Targeted molecular dynamics simulations of conformational change—application to the T ↔ R transition in insulin. *Mol. Sim.* **10**, 291–309.
7. 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.
8. 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.
9. 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.
10. Echols, H. & Goodman, M. F. (1991). Fidelity mechanism in DNA replication. *Annu. Rev. Biochem.* **60**, 477–511.

11. 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.

12. 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.

13. 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.

14. Doublié, S., Sawaya, M. R. & Ellenberger, T. (1999). An open and closed case for all polymerases. *Structure*, **7**, R31–R35.

15. Watson, J. D. & Crick, F. H. C. (1953). Genetical implications of the structure of deoxyribonucleic acid. *Nature*, **171**, 964–967.

16. Kunkel, T. A. & Wilson, S. H. (1998). DNA polymerases on the move. *Nature Struct. Biol.* **5**, 95–99.

17. Kool, E. T., Morales, J. C. & Guckian, K. M. (2000). Mimicking the structure and function of DNA: insight into DNA stability and replication. *Angew. Chem. Int. Ed.* **39**, 990–1009.

18. Kool, E. T. (1998). Replication of non-hydrogen bonded bases by DNA polymerases: a mechanism for steric matching. *Biopolymers*, **48**, 3–17.

19. Goodman, M. F. (1997). Hydrogen bonding revisited: geometric selection as a principal determinant of DNA replication fidelity. *Proc. Natl Acad. Sci. USA*, **94**, 10493–10495.

20. Steitz, T. A. (1998). A mechanism for all polymerases. *Nature*, **391**, 231–232.

21. Wilson, S. H. (1998). Mammalian base excision repair and DNA polymerase β. *Mutat. Res.* **407**, 203–215.

22. Beard, W. A. & Wilson, S. H. (2000). Structural design of a eukaryotic DNA repair polymerase: DNA polymerase β. *Mutat. Res.* **460**, 231–244.

23. 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.

24. Johnson, K. A. (1993). Conformational coupling in DNA polymerase fidelity. *Annu. Rev. Biochem.* **62**, 685–713.

25. 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.

26. 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.

27. 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.

28. 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.

29. Yang, L., Beard, W. A., Wilson, S. H., Broyde, S. & Schlick, T. (2002). Polymerase β simulations suggest that Arg258 rotation is a slow step rather than large subdomain motion *per se. J. Mol. Biol.* **317**, 651–671.

30. Ferrara, P., Apostolakis, J. & Caflisch, A. (2000). Computer simulations of protein folding by targeted molecular dynamics. *Proteins: Struct. Funct. Genet.* **39**, 252–260.

31. Saenger, W. (1984). *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.

32. 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.

33. Beard, W. A., Shock, D. D., Yang, X.-P., DeLauder, S. F. & Wilson, S. H. (2002). Loss of DNA polymerase β stacking interactions with templating purines, but not pyrimidines, alters catalytic efficiency and fidelity. *J. Biol. Chem.* **277**, 8235–8242.

34. 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.

35. 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.

36. 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.

37. 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.

38. Doublié, S. & Ellenberger, T. (1998). The mechanism of action of T7 DNA polymerase. *Curr. Opin. Struct. Biol.* **8**, 704–712.

39. 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.

40. Steitz, T. A. (1993). DNA- and RNA-dependent DNA polymerases. *Curr. Opin. Struct. Biol.* **3**, 31–38.

41. 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.

42. Beese, L. S., Derbyshire, V. & Steitz, T. A. (1993). Structure of DNA polymerase I Klenow fragment bound to duplex DNA. *Science*, **260**, 352–355.

43. Kim, Y., Eom, S. H., Wang, J., Lee, D. S., Suh, S. W. & Steitz, T. A. (1995). Crystal structure of *Thermus aquaticus* DNA polymerase. *Nature*, **376**, 612–616.

44. Pelletier, H., Sawaya, M. R., Wolfle, W., Wilson, S. H. & Kraut, J. (1996). Crystal structures of human DNA polymerase β complexed with DNA: implications for catalytic mechanism, processivity, and fidelity. *Biochemistry*, **35**, 12742–12761.

45. Eom, S. H., Wang, J. & Steitz, T. A. (1996). Structure of Taq polymerase with DNA at the polymerase active site. *Nature*, **382**, 278–281.

46. Kiefer, J. R., Mao, C., Hansen, C. J., Basehore, S. L., Hogrefe, H. H., Braman, J. C. & Beese, L. S. (1997). Crystal structure of a thermostable Bacillus DNA polymerase I large fragment at 2.1 Å resolution. *Structure*, **5**, 95–108.

47. Doublié, 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.

48. 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.

49. 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.

50. 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.

51. 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.

52. 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.

53. 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.

54. Patel, S. S., Wong, I. & Johnson, K. A. (1991). Presteady-state kinetic analysis of processive DNA replication inducing complete characterization of an exonuclease-deficient mutant. *Biochemistry*, **30**, 511–525.

55. Wong, I., Patel, S. S. & Johnson, K. A. (1991). An induced-fit kinetic mechanism for DNA replication fidelity: direct measurement by single-turnover kinetics. *Biochemistry*, **30**, 526–537.

56. 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.

57. 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.

58. 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.

59. Kool, E. T. (2001). Hydrogen bonding, base stacking, and steric effects in DNA replication. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 1–22.

60. Mendelman, L. V., Petruska, J. & Goodman, M. F. (1990). Base mispair extension kinetics. Comparison of DNA polymerase α and reverse transcriptase. *J. Biol. Chem.* **265**, 2338–2346.

61. Chou, K. M. & Chou, Y. C. (2002). An exonucleolytic activity of human apurinic/apyrimidinic endonuclease on 3′ mispaired DNA. *Nature*, **415**, 655–659.

62. Goodman, M. F., Creighton, S., Bloom, L. B. & Petruska, J. (1993). Biochemical basis of DNA replication fidelity. *Crit. Rev. Biochem. Mol. Biol.* **28**, 83–126.

63. Kosa, J. L. & Sweasy, J. B. (1999). The E249K mutator mutant of DNA polymerase β extends mispaired termini. *J. Biol. Chem.* **274**, 35866–35872.

64. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al.* (2000). The Protein Data Bank. *Nucl. Acids Res.* **28**, 235–242.

65. 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.

66. 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.

67. 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.

68. Brooks, B. R., Bruccoleri, 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.

69. MacKarell, 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.

70. 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.

71. 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.

72. 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.

73. 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.

74. Schlick, T. (2001). Time-trimming tricks for dynamic simulations: splitting force updates to reduce computational work. *Structure*, **9**, R45–R53.

75. Ma, J. & Karplus, M. (1997). Molecular switch in signal transduction: reaction paths of the conformational changes in *ras* p21. *Proc. Natl Acad. Sci. USA*, **94**, 11905–11910.

76. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.