Mismatch Induced Conformational Distortions in Polymerase β Support an Induced-Fit Mechanism for Fidelity†

Karunesh Arora†, William A. Beard‡,
Samuel H. Wilson§, Tamar Schlick§*

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

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

Running title: DNA polymerase β mismatch simulations

Keywords (as below): DNA Polymerase β; Conformational Transitions; Fidelity; Mismatch; Molecular Dynamics, DNA repair, Induced-fit mechanism, Protein/DNA complex

*To whom correspondence should be addressed (Phone: 212-998-3116; e-mail: schlick@nyu.edu; fax: 212-995-4152)

June 22, 2005

†This work was supported by NSF grant ASC-9318159, 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).
**Abbreviations**: pol β, DNA polymerase β; dNTP, 2'-deoxyribonucleoside 5'-triphosphate; dd-CTP, 2', 3'-dideoxyribocytidine 5'-triphosphate; dCTP, 2'-deoxyribocytidine 5'-triphosphate; MD, Molecular dynamics;
Abstract:

Molecular dynamics simulations of DNA polymerase (pol) β complexed with different incorrect incoming nucleotides (template base: incoming nucleotide G:G, G:T, and T:T) at the template/primer terminus are analyzed to delineate structure/function relationships for aberrant basepairs in a polymerase active site. Comparisons, made to pol β structure and motions in the presence of a correct basepair, aim at gaining atomically detailed insights into the process of nucleotide selection and discrimination. In the presence of an incorrect incoming nucleotide, α-helix N of the thumb subdomain believed to be required for pol β’s catalytic cycling moves towards the open conformation rather than the closed conformation as observed for the correct basepair (G:C) before the chemical reaction. Correspondingly, active site residues in the microenvironment of the incoming base are in intermediate conformations for non Watson-Crick pairs. The incorrect incoming nucleotide and the corresponding template residue assume distorted conformations and do not form Watson-Crick bonds. Furthermore, the coordination number and the arrangement of ligands observed around the catalytic and nucleotide binding magnesium ions are mismatch specific. Significantly, the crucial nucleotidyl transferase reaction distance (Pα–O3′) for the mismatches between the incoming nucleotide and the primer terminus is not ideally compatible with the chemical reaction of primer extension that follows these conformational changes. Moreover, the extent of active-site distortion can be related to experimentally determined rates of nucleotide misincorporation and to the overall energy barrier associated with polymerase activity. Together, our studies provide structure/function insights into the DNA-polymerase induced constraints (i.e., α-helix N conformation, DNA basepair bonding, conformation of protein residues in the vicinity of dNTP, and magnesium ions coordination) during nucleotide discrimination and pol β-nucleotide interactions specific to each mispair and how they may regulate fidelity. They also lend further support to our recent hypothesis that additional conformational energy barriers are involved following nucleotide binding but prior to the chemical reaction.
DNA polymerases play a key role in maintaining genome integrity [1, 2, 3]. Eukaryotic DNA polymerase β (pol β) is a DNA repair enzyme that preferentially selects the correct nucleotide (e.g., C opposite G) instead of an incorrect unit (e.g., A opposite G) during DNA repair and replication synthesis. Experimental data suggest an ‘induced-fit’ mechanism [4, 5] between the DNA-bound polymerase and substrate: the correct substrate aligns catalytic groups as required for proper synthesis (‘fidelity’), while an incorrect unit distorts the geometric and electrostatic environment so that DNA synthesis is hampered. The fidelity of DNA polymerases broadly refers to their ability to incorporate correct rather than incorrect nucleotides complementary to the template DNA [6]; such fidelities span a wide range, from 1 to nearly $10^6$ errors per one million nucleotides incorporated. DNA pol β is a moderate fidelity polymerase belonging to the X-family of DNA polymerases; it fills short DNA gaps during base excision repair and has an average base-substitution error frequency of one per $10^3$ bases synthesized [7, 8].

The fidelity of DNA polymerases in selecting the correct dNTP is attributed to several structural factors: Watson-Crick hydrogen bonding [9, 10], steric fit between the incoming dNTP and the polymerase active-site pocket (including template) [11], and pattern of hydrogen bonding between polymerases and duplex DNA bases [12]. 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 are believed to play an important role in maintaining DNA synthesis fidelity [13, 10, 14].

Structurally, pol β is composed of only two domains, an amino-terminal 8-kDa region which exhibits deoxyribosephosphate lyase activity, and a 31-kDa carboxyl terminal domain which possesses nucleotidyl transfer activity. The 31-kDa domain shares an overall hand-like architecture similar to other structurally characterized DNA polymerases, containing finger, palm, and thumb subdomains [15]. Pol β is ‘left-handed’ while polymerases from other families are ‘right-handed’. Pol β lacks an intrinsic 3’ to 5’ proofreading activity.

Three crystallographic structures representing the polymerase reaction pathway (see Fig. 1) have been obtained for pol β [5]: open binary complex, containing pol β bound to a DNA substrate with a
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, pol β-nick: pol β bound to nicked DNA. Figure 1, b and c, illustrate the significant conformational difference (in the thumb subdomain about 7 Å) between these open and closed forms. Recently, DNA pol β structures with mismatches (A:C and T:C, denoting template:primer units in the nascent basepair binding pocket) were also solved [16]. The structure of the mispairs indicate that α-helix N is in an ‘intermediate’ conformation and the bases of the mispair do not form hydrogen bonds with one another.

Based on kinetic and structural studies of various DNA polymerases [17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30], a common nucleotide insertion pathway has been deduced where the enzyme undergoes transitions between open and closed forms (Fig. 1(a)). At step 1, following DNA binding, the DNA polymerase binds a 2'-deoxyribonucleoside 5'-triphosphate (dNTP) to form an open substrate complex; this complex is assumed to undergo conformational change to align catalytic groups and form a closed ternary complex at step 2; the nucleotidyl transferase reaction then follows. The 3'-OH of the primer strand attacks the Pα of the incoming dNTP to extend the primer strand and form the ternary product complex (step 3); this complex then undergoes conformational changes (step 4) leading to the open enzyme form. This transition is followed by dissociation of pyrophosphate (PPi) (step 5), after which the DNA synthesis/repair cycle can begin anew.

The conformational rearrangements involved in steps 2 and 4 are believed to be key for monitoring DNA synthesis fidelity [5]. Binding of the correct nucleotide is thought to induce the first conformational change (step 2) whereas binding of an incorrect nucleotide may alter or inhibit the conformational transition. This ‘induced-fit’ mechanism was thus proposed to explain the polymerase fidelity in selection of correct dNTP [5]. This mechanism suggests that the conformational changes triggered by binding of the correct nucleotide will align the catalytic groups as needed for catalysis, whereas the incorrect substrate will somehow interfere with this process. However, no direct experiments have shown this effectively. It is only recently that the structures of pol β bound to a mispair [47, 16], along with those for high-fidelity Bacillus DNA polymerase I [31] and lesion-
bypass polymerase Dpo4 [32], have begun to shed new insights into the distortions and disruptions within and around the active site of a polymerase when a mismatch is encountered.

Earlier, molecular dynamics simulations of pol β/DNA complex with the enzyme’s active site occupied by either correct incoming basepair or no substrate have delineated structural and dynamical changes that occur during conformational changes before the nucleotide transferase reaction [33]. We observed contrasting global subdomain motions as well as alignment of catalytic residues in the presence versus absence of substrate in the pol β binding site. These studies provided in silico support for the substrate-induced conformational change in pol β.

To further explore the induced-fit hypothesis and understand pol β’s nucleotide discrimination mechanism, we examine here, by several atomistic molecular dynamics simulations, the structural changes and polymerase-substrate interactions with mismatched basepairs in the polymerase active site. The present work complements prior simulations of pol β with mispairs after the nucleotide incorporation [34], which helped interpret kinetic data for DNA mispair extension. The present study investigates mispair geometric/flexibility before nucleotide incorporation and helps interpret kinetic data for pol β nucleotide misinsertion [30, 35, 19]. This investigation is warranted since limited structural information for complexes of DNA polymerases bound to mismatches exist [16, 32, 31, 48]. Modeling and simulation combined with available structural information, can help probe such structure/flexibility/function relationships systematically. Because nucleotide incorporation by pol β is a dynamic process, molecular dynamics investigations [36, 37, 38, 39] can provide important snapshots to link and extend interpretations of experimental structural anchors.

Our dynamics simulations reveal a large thumb-subdomain conformational change, closing in the presence of the correct basepair but opening in the presence of mismatched basepairs, consistent with ‘induced-fit’ hypothesis. The geometry of each mispair complex at the end of simulations indicates that incoming nucleotides and the corresponding template residue do not form Watson-Crick hydrogen bonds. The magnesium ion coordination and active-site geometries crucial for subsequent catalysis are severely distorted for all mispairs. Furthermore the coordination number and the arrangement of ligands observed around the catalytic and nucleotide binding magnesium ions are mismatch specific. Significantly, the nucleotidyl transferase reaction distance (Pα–O3') between the
incoming nucleotide and the primer terminus is not ideal for the chemical reaction following the conformational change process. The extent of distortion of the active-site geometry can be qualitatively related to the experimentally determined rates of misincorporation of different mismatches in the pol β active site [35]. These observations combined with information available from experiments suggest a direct relation between active-site structural distortions and fidelity of pol β. They also lend further credence to our recent hypothesis that additional conformational energy barriers are involved following nucleotide binding but prior to the chemical reaction.

**Computational Methodology**

**Initial Models**

All simulations of pol β/DNA complexes with incorrect incoming nucleotide were started from the intermediate ("half-closed") conformation before the nucleotidyl transferase reaction to capture pol β’s closing within 10 nanoseconds. This follows our prior experience, where the intermediate structure of pol β was successful in capturing the polymerase opening (after the chemical reaction of nucleotidyl transfer) [36] and closing [33] in the polymerase kinetic cycle. The intermediate forms used in our studies also resemble a crystal intermediate [16] and are found on the closing pathway delineated by the stochastic path approach [38].

In earlier work, we investigated conformational changes in pol β/DNA complex in the presence of a correct basepair (G:C) [33]. There the intermediate model was constructed as an average of the crystallographic open, binary gapped complex (1BPX) and closed, ternary complex (1BPY) from the PDB/RCSB resource [40]. Specifically, the model’s thumb subdomain is partially closed, with the correct basepair (G:C) in the active site. On the basis of this model, complexes of pol β with mismatches G:T, G:G, and T:T (template:incoming dNTP) were built by replacing the correct incoming nucleotide (dCTP) with incorrect nucleotides dTTP, dGTP and dTTP, respectively. For the T:T mispair, the template G was replaced by T. The protein residues and other DNA base sequences remain unchanged. Both the incoming nucleotide and template are in anti conformation for all mis-
pairs, following the crystal structure of pol β with a mispair [16]. The starting conformations of all three mispairs in the active site are shown in Fig. 2(a). Mispairs G:T and T:T are hydrogen bonded with one another in the initial models. The thumb subdomain is in the intermediate conformation between crystal “open” and “closed” complex (see Fig 2(b)).

All mismatch models were similarly solvated in a cubic periodic domains using our program PBCAID [41]. To neutralize the system at an ionic strength of 150 mM, water molecules (TIP3 model) with minimal electrostatic potential at the oxygen atoms were replaced by Na\(^{+}\), and those with maximal electrostatic potential were replaced with Cl\(^{-}\). All Na\(^{+}\) and Cl\(^{-}\) ions were placed more than 8 Å away from any protein or DNA atom and from each other. The electrostatic potential for all bulk oxygen atoms was calculated using the DelPhi package [42]. All solvated systems with mismatch incoming nucleotides contain about 40,000 atoms including water of solvation (11,249 water molecules).

**Minimization, Equilibration and Dynamics Protocol**

Energy minimizations, equilibrations, and dynamics simulations for all systems were performed using the program CHARMM (Chemistry Department, Harvard University, Cambridge, MA) [43] with the all-atom version C26a2 force field [44]. First, each system was minimized with fixed positions of all protein and DNA heavy atoms using Steepest Descent (SD) for 10,000 steps; this was followed by Adapted Basis Newton-Raphson minimization [43, 45] for 20,000 steps. The 30 ps of equilibration at 300 K was followed by further minimization using SD and ABNR until the gradient of root-mean-square deviations was less than 10\(^{-6}\)kcal/mol-Å. Finally, the protein/DNA/counterions/water coordinates were re-equilibrated for 30 ps at 300 K by the Langevin multiple timestep method LN (See Ref. [36] for a thorough examination of the stability and reliability of the integrator for large macromolecular systems in terms of thermodynamic, structural and dynamic properties compared to single-timestep Langevin as well as Newtonian (Velocity Verlet) propagators [36]). Our triple-timestep protocol uses an inner timestep \(\Delta \tau = 1\) fs for updating local bonded interactions; medium timestep \(\Delta t_m = 2\) fs to update nonbonded interactions within 7 Å; and outer timestep of \(\Delta t = 150\) fs for calculating the remaining terms; this yields a speedup factor of
4 over single-timestep Langevin dynamics at $\Delta \tau = 1$ fs. The SHAKE algorithm was employed to constrain the bonds involving hydrogen atoms. Electrostatic and van der Waals interactions were smoothed to zero at 12 Å with a shift function and a switch function, respectively. The Langevin collision parameter of $\gamma = 10$ ps$^{-1}$ was chosen to couple the system to a 300°C heat bath. The total simulation length for each system is between 10 to 15 ns.

**Results and Discussions**

**Subdomain Conformation and Catalytic Activation**

For all mismatch trajectories, started from the intermediate position of $\alpha$-helix N of the thumb subdomain, we capture movements towards the open conformation observed in the binary pol $\beta$ structure with gapped DNA (Fig. 3). The extent of $\alpha$-helix N opening is mispair dependent (see Fig. 4). Specifically, the RMSD of $\alpha$-helix N with respect to the open conformation can be ranked as G:G < T:T < G:T < G:C. Thus, the purine-purine mismatch (G:G) favors the inactive conformation most strongly while the purine-pyrimidine and pyrimidine-pyrimidine mismatches (G:T and T:T) tend to intermediate conformations between closed and open states as recently observed in the solved crystal mismatch structures (A:C and T:C) of pol $\beta$ [16]. A distorted open conformation that differs from that seen in the presence of a cognate basepair was also observed in high-fidelity *Bacillus* DNA polymerase I [31]; note that there mispairs are at different part of the catalytic cycle than those studied here.

In contrast to the open form of the thumb preferred for all mismatch complexes, thumb *closing* is observed in the presence of the correct nucleotide (G:C). Prior works using dynamics and advanced sampling techniques have shown that thumb closing in the presence of a correct base pair occurs in multiple steps rather than a single concerted manner [38, 37]. That is, a metastable conformation exists for the thumb corresponding to a “half-rotated” state of Arg258 on the pathway between crystal open and closed conformations; this thumb movement triggers several side chain motions in the microenvironment of the incoming substrate. In concert, coupled motions produce unique interactions between the polymerase and the substrate present neither in the open or closed
crystal forms and provide specific suitability checks for pol \( \beta \) to differentiate correct from incorrect substrates. We speculate that a disruption of the thumb closing in the presence of a mismatched incoming nucleotide might signal a stalling of such a reaction pathway. The observed global thumb conformational changes in response to mismatched incoming substrates are consistent with the proposed induced-fit mechanism for DNA polymerase fidelity.

More specifically, from structural analyses of pol \( \beta \) in the crystal open and closed conformations, several key residues (Asp192, Phe272, Arg258) have been implicated to act in concert with the thumb’s movement [5]. In the open thumb conformation, Asp192 is engaged in a salt bridge with Arg258 while in the closed thumb conformation, the phenyl ring of Phe272 disrupts the salt bridge between Asp192 and Arg258, freeing Asp192 to ligand the nucleotide-binding metal ion. The closed conformation also puts the side chains of thumb residue Tyr296 in position to hydrogen bond with Arg258, ensuring that Arg258 will not interfere with Asp192’s ability to ligand the metal ion. According to the current induced-fit hypothesis, the conformational closing (in presence of a correct nucleotide) prepares these residues in the binding site for the subsequent step of nucleotidyl transfer along the polymerase kinetic pathway (Fig. 1).

Fig. 5 shows the conformation of these key active site residues in the presence of correct versus incorrect incoming nucleotides as observed during the simulations. For the correct incoming nucleotide (dCTP) Asp192 coordinates both magnesium ions, and Phe272 insulates Asp192 from Arg258 by intervening between these side chains. In the G:T mismatch conformations, these residues are positioned similar to those in the presence of the correct incoming nucleotide. For G:G and T:T, Asp192 coordinates the magnesium ions but Phe272 is not positioned between Asp192 and Arg258 and instead resembles the conformation of Phe272 in the open thumb conformation. Fig. 6 depicts the conformation of Phe272 in all trajectories as characterized by two dihedral angles. To determine the position of Arg258 during the simulations we follow in Fig. 7 the interatomic distance evolutions of Arg258:NH1 to both Asp192:OD1 and Tyr296:OH in the presence of correct and incorrect nucleotides. The figure shows how Arg258 rotates and forms hydrogen bond with Tyr296 in the matched (G:C) and mismatched (G:T) trajectories. However, for the G:G and T:T mismatch simulations, distances between Arg258 and Tyr296 remain constant throughout the tra-
DNA polymerase β mismatch simulations

jectory as in the starting intermediate conformation. Thus, results indicate that the motion of key active site residues (Arg258, Asp192 and Tyr296) is correlated with the thumb subdomain movement (see Fig. 4), which in turn is related to the presence and the type of the incoming nucleotide in the basepair binding pocket.

The thumb subdomain movement towards open or closed states can also be interpreted by following the motion of protein residues Asp276, Arg283, Asn279 and Lys280 that interact with the nascent basepair only in the closed conformation.

Structural and kinetic analyses of pol β mutants involving residues at the active site suggest that the selection of the incoming nucleotide is guided by the templating residue via an induced-fit mechanism [46]. The templating residue is not properly aligned to basepair with the incoming nucleotide in the open thumb conformation, but thumb closure favors the active conformation of template by directly forming van der Waals contacts with the template, anchoring it to the polymerase [5].

In Figure 8, we follow the evolution of distances between the nascent basepair binding residues (Asp276, Arg283, Asn279 and Lys280) and the DNA template base in the presence of correct and incorrect nucleotides. Figure 8 (top left) shows that template binding residues move to interact with the DNA template G in the presence of the correct incoming nucleotide. In the G:T mismatch simulation as illustrated in Fig. 8 (top right), the templating residues first move away from the DNA template but then move towards it as the trajectory evolves. For the G:G and T:T mismatch simulations where the thumb assumes the open conformation, the templating residues evolve away from the template (Fig. 8, bottom). The observed motion of the templating residues is also directly correlated with the thumb’s conformational change.

Nascent Basepair Conformation in the DNA Pol β Active Site

In Figure 9, we superimpose the nascent basepair and two neighboring primer/template DNA basepairs in the simulated G:C, G:T, G:G and T:T systems with corresponding basepairs in the closed crystal ternary pol β DNA complex with correct incoming nucleotide according to the protein Cα atoms. The DNA basepairs in the closed crystal pol β DNA complex serve as a reference to determine the distortion of the nascent basepairs in the matched and mismatch structures. The sim-
ulated correct basepair deviates slightly from the perfect Watson-Crick pairing, but the deviations are much greater for incorrect incoming nucleotides. Interestingly, for all mispair geometries, either the primer or template strand slides towards the major groove (see Fig. 9). The template bases G and T are displaced toward the major groove in the G:T and T:T mispairs, respectively. For the G:G and G:T mispairs, the incoming nucleotide bases are displaced toward major groove. The template and primer bases slide toward the major groove to varying degrees relative to the corresponding basepair in the crystal closed complex. Displacement of the incoming nucleotide distances the \( \alpha \)-phosphorous from the primer terminus results in a larger \( P_\alpha-O3' \) distance and discourages incorrect nucleotide insertion.

We also note non-Watson-Crick interactions within all mispairs; in particular, the basepairs stack against one another in the G:G mispair. This staggered conformation of nascent basepairs in the confines of pol \( \beta \) active site is consistent with the structure of a purine-purine mismatch in the pol \( \beta \) active site [47].

According to the geometric selection model steric size and shape complementarity of nascent basepairs is more crucial than the ability to form Watson-Crick hydrogen bonds [9, 10]. The ability of several DNA polymerases to insert non-polar nucleoside isoteres that do not form Watson-Crick bonds is consistent with this model [48, 49]. Still, the significance of Watson-Crick hydrogen bonding in determining DNA polymerase synthesis and high fidelity appears to be polymerase dependent [50]. For example, pol \( \beta \) fails to replicate non-polar isoteres, indicating that minor groove hydrogen bonds with the template base and incoming nucleotide may be crucial in nucleotide selection process.

Our prior work investigated three mismatched basepairs, namely G:G, A:C and C:C (template:primer) at the DNA polymerase active site after chemistry (step 4 in Fig. 1) [34]. It appeared that the hierarchy of distortions for these basepairs (G:G > C:C > A:C) paralleled the experimentally determined ability of pol \( \beta \) to extend these mispairs. The work here investigates mispairs, namely G:G, G:T, T:T (template/primer) at the polymerase active before the nucleotide incorporation (step 2 in Fig. 1). The observed mispair distortions in pol \( \beta \) in either case, before and after nucleotidyl transferase reaction, are quite similar. That is, Watson-Crick pairing is not observed
for the mispairs, and in general, purine:purine (e.g., G:G) mismatches are most severely distorted and difficult to incorporate or extend. Kinetic analyses of the incorporation and extension of all 12 mispairs by pol β agree with the trends observed in our simulations [35].

Mismatch-Induced Disruption of the Active-site Geometry

DNA polymerase structures are consistent with a “two-metal-ion” mechanism [51, 52, 53] of nucleotide incorporation: the incoming nucleotide is accompanied by a magnesium ion (the nucleotide binding ion) that coordinates the phosphoryl oxygen atoms with all three phosphate groups. In addition, a catalytic magnesium ion coordinates a phosphoryl oxygen atom of the α-phosphate group. In pol β, these magnesium ions are coordinated in the active site by oxygen atoms of three conserved aspartate residues (Asp190, Asp192, and Asp256). According to this two-metal-ion mechanism, the catalytic Mg\(^{2+}\) lowers the affinity of the proton in the primer 3′-OH group, facilitating the 3′-O attack on the P\(_α\) of the incoming nucleotide [54, 55]. The nucleotide-binding Mg\(^{2+}\) assists the pyrophosphate’s departure, and both ions stabilize the structure of the presumed penta-coordinated transition state.

However, in the available crystal ternary pol β/DNA complex with a correct incoming nucleotide, the primer terminus lacks the 3′-OH group. The recently reported crystal structures of DNA mismatches (template:primer, A:C and T:C) in the active site were solved by using nicked product DNA that traps the mispair in the nascent basepair binding pocket [16]. Thus, the affect of the missing 3′-OH on the magnesium ion coordination and the pol β active-site geometry cannot be inferred from the available structural information.

In our simulations, we modeled the missing 3′-OH group at the primer terminus in all starting structures. Figure 10 shows the coordination of both catalytic and nucleotide binding magnesium ions as observed during the simulations for matched versus mismatched incoming nucleotides at the template/primer terminus. Clearly, the ligand arrangement around both magnesium ions is mispair specific. Already, the magnesium ion coordination sphere for the correct basepair differs from that observed in the ternary closed crystal complex of pol β; in the crystal structure (PDB entry 1BPY) [5], the catalytic magnesium ion is found coordinated with the cation bound to the
OD2 oxygen of Asp256, the non-bridging O1A oxygen of the substrate (dCTP), the OD1 oxygen of Asp192, and the carboxylate oxygen atoms of Asp190. In addition, a water molecule was postulated to coordinate with catalytic magnesium, though not observed in the electron density map [5]. Interestingly, this predicted water molecule becomes bound to the catalytic ion during the simulation in the presence of the correct incoming nucleotide [33] and, is observed in a high resolution (1.65 Å) structure of pol β ternary complex recently solved in the Wilson lab (unpublished results). Some differences in ligand coordination are listed in Table 1 and have been discussed previously [33], where we attribute non-ideal geometries as indication that additional high-energy barrier must be surmounted before chemical reaction (see further discussion at end of subsection). Most strikingly, we note larger than expected distances between the catalytic magnesium ion and the modeled O3', not present in the crystal structure; the O3' atom is expected to coordinate the catalytic ion to complete the coordination sphere. In addition, the O1A oxygen atom on the α-phosphorus of the incoming nucleotide coordinates only one of the magnesium ion and not both ions, as also suggested by the closed ternary crystal complex (1BPY).

For the mismatched incoming substrates in the active site, magnesium ion coordination departs further from the perfect ligand arrangement as assumed to be required for the subsequent nucleotidyl transferase reaction (see Fig. 10 and Table 1). Though the magnesium ions are hexa-coordinated, ligands that complete the coordination are distinct for each mispair and also depart from that for the correct pair. This implies loss of other interactions with protein side chains and the incoming nucleotide triphosphate, leading to more water molecules coordinating the catalytic magnesium ion so as to complete the ion coordination sphere in all mispair active site geometries. According to Petruska et al. [56], enzymes amplify free energy differences between correct and incorrect basepair by partially excluding water from the active site, thus increasing enthalpy differences and reducing entropy differences and leading to improved fidelity. Water exclusion and enthalpy-entropy compensation has been put forward as likely explanation for higher fidelity of polymerases than can be explained by free energy differences between correct and incorrect basepairs in aqueous solution.

The coordination of the magnesium ions directly influences the nucleotidyl transferase reaction that follows subdomain motion. Presumably, this chemical reaction occurs either by means of a
metal-assisted (Mg$^{2+}$ ion) dissociative mechanism or associative in-line nucleophilic attack on the P$_\alpha$ of the incoming nucleotide by 3'-OH of the DNA primer terminus, producing an extended DNA primer. The ideal P$_\alpha$–O3' distance for the phosphoryl transferase reaction to proceed via a dissociative mechanism should be around 3.3 Å [57]. In our simulations we observe that the crucial P$_\alpha$–O3' distance is larger than than ideal distance in the presence of both correct and incorrect incoming nucleotides in the pol $\beta$ active site. For incorrect compared to correct incoming nucleotides, the P$_\alpha$–O3' distance is greater by 1–2 Å. These notable deviations from the ideal active site geometries compatible with nucleotidyl transferase reactions imply a hampering of the incorrect nucleotide incorporation permitting the incorrect dNTP to diffuse away.

Based on the above discussion, we propose that the extent of distortion of pol $\beta$'s active-site geometry can be broadly described by three crucial distances (P$_\alpha$–O3', Catalytic Mg$^{2+}$–O3', and Catalytic Mg$^{2+}$–dNTP (O1A)) required to be within ideal distance for the chemical reaction of nucleotidyl transfer. With these distance criteria, the hierarchy of the mismatch-induced distortions of the active-site geometry, in increasing order, is: G:T < G:G < T:T (see Table 2).

Based on the experimentally determined kinetic parameters, $k_{pol}$ and $K_d$ listed in Table 2, the fidelity of different mismatches in pol $\beta$ can be arranged in the following increasing order: G:T < G:G < T:T. Since fidelity is inversely related to the error frequency, we can infer that G:T is the most favorable mispair for pol $\beta$ followed by G:G and T:T; T:T is the least favorable mispair to be misincorporated [35]. Thus, the degree of active-site geometry distortion determined from our simulations approximately parallels the experimentally determined efficiencies of misincorporation for different mismatches. This suggests that a relationship between active-site structural distortion and fidelity may be appropriate for understanding pol $\beta$. Moreover, given that most DNA polymerases insert incorrect nucleotides with similar efficiencies [58], strategies utilized by pol $\beta$ to discriminate against incorrect nucleotide may be generally applicable.

This comparative analysis of mispair conformations provides new insights on polymerase mechanisms for the following reasons. Systematic QM/MM investigations of pol $\beta$ in which active-site atoms treated with the ab-initio approach also resulted in larger than expected active site distances [59]. Larger distances also arose in dynamics simulations of Dpo4 from the Broyde lab [60]:
2.5 ns simulations from the ternary complex of Dpo4 with modified/unmodified template and different incoming substrates produced in all but one simulation $P_{\alpha-O3'}$ distances far from ideal and larger for mismatches ([60], supplementary figure S5); though these findings are not discussed in their work, perturbed active-site geometries by another group employing AMBER rather than CHARMM lend credence to our interpretation that these distortions are not mere force-field artifacts and that requisite rearrangements before chemistry may have important biological ramifications.

Specifically, this scenario suggests that additional active site rearrangements must follow the thumb’s conformational change to evolve the system to a state compatible for the chemical reaction. Such conformational changes will likely involve slow and stochastic adjustments of Mg$^{2+}$ and phosphoryl coordination at the pol $\beta$ active-site in a part of the reaction profile we call “pre-chemistry avenue” (after subdomain closing but before chemistry). From emerging studies, we speculate that the chemical reaction might proceed from partially open states or suboptimal ‘activated’ complexes for mismatched incoming nucleotides, after these “pre-chemistry avenue” changes. The existence of such pre-chemistry changes is supported by the recently solved A:C and T:C mismatch crystal structures of pol $\beta$ where the thumb is in an intermediate conformation [16]. The pre-chemistry avenue can be viewed as another fidelity check point on en route to the chemical reaction that promotes selectivity and hence fidelity. The existence of pre-chemistry changes does not diminish the role of large subdomain rearrangements, still required, but not rate limiting, for incoming nucleotide recognition [19, 36, 61]. NMR data from the London/Wilson [62] and Mildvan [63] groups also suggest localized motions proximal to metal-binding ligands and near the $\beta/\gamma$ phosphates of the incoming nucleotide prior to chemistry but following dNTP binding.

**Concluding Remarks**

Our dynamics simulations of pol $\beta$/DNA complexes with nascent G:G, T:T and G:T mispairs before chemistry. Simulations suggest several structural and dynamic features for all mispairs which may represent mechanisms employed by pol $\beta$ to evade error initiation. The closing motion of the $\alpha$-helix N of the thumb subdomain required for catalytic cycling of pol $\beta$ is disrupted in the presence of mispairs in the pol $\beta$ active site. The extent of thumb opening in trajectories started from
an intermediate conformation is mismatch specific. These global conformational changes are consistent with the induced-fit mechanism for substrate discrimination enhancing polymerase fidelity. The induced-fit mechanism can alter enzyme specificity even if conformational changes are not kinetically rate-limiting if nucleotide incorporation involves non-identical interactions with the active site in a substrate-dependent transition state [64].

Besides large-scale conformational motions, there are several subtle local conformational changes in pol β’s active site in response to an incoming nucleotide. The specific structural changes are dependent of the identity of the incoming nucleotide. The magnesium ions which are crucial for stabilizing pol β’s active site (in the presumed penta-coordinated transition state for phosphoryl transferase reaction) have distorted ligand arrangements. The extent of distortion of pol β’s active-site geometry can be broadly described by three crucial distances (Pα–O3′, Catalytic Mg$^{2+}$–O3′, and Catalytic Mg$^{2+}$–dNTP (O1A)) required to be within ideal distance for the of nucleotide transferase reaction. The degree of active-site geometry distortion determined from our simulations roughly parallels the experimentally determined rates of misincorporation for different mismatches suggesting a direct relation between active-site structural distortions and fidelity of pol β.

The related chemical reaction distance (Pα–O3′) between the matched or mismatched incoming nucleotide and the primer terminus is larger than expected for phosphoryl transfer by a dissociative mechanism [54, 65, 66]. The base of the incorrect nucleotide and its template partner stack against each other rather than forming Watson-Crick hydrogens bonds. Overall, our observed active site geometries for both the Watson-Crick G:C pair and the mispairs, do not exhibit the ideal geometry for the chemical reaction. This suggests that additional subtle adjustments in the pol β active site are required to evolve the system to the transition state poised for chemistry. The nature of these adjustments, their energetic costs, and relation to the overall reaction barriers will be the subject of future investigations.

Acknowledgement

Molecular images were generated using VMD [67] and the INSIGHTII package (Accelrys Inc., San Diego, CA).
References


DNA polymerase β mismatch simulations


Table 1: Average active site interatomic distances for the simulated matched (G:C) and mismatched (G:T, G:G, T:T) simulations.

<table>
<thead>
<tr>
<th>Nucleotidyl Transfer distance</th>
<th>Distance (Å)</th>
<th>G:C&lt;sup&gt;a&lt;/sup&gt;</th>
<th>G:T</th>
<th>G:G</th>
<th>T:T</th>
<th>G:C&lt;sup&gt;b&lt;/sup&gt; crystal</th>
</tr>
</thead>
<tbody>
<tr>
<td>dNTP (P&lt;sub&gt;α&lt;/sub&gt;) : P10 (O3')</td>
<td>4.19</td>
<td>4.11</td>
<td>4.97</td>
<td>5.53</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Catalytic Magnesium ion Coordination</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mg&lt;sup&gt;2+&lt;/sup&gt;(A) : P10 (O3')</td>
<td>4.98</td>
<td>4.51</td>
<td>4.36</td>
<td>4.41</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Mg&lt;sup&gt;2+&lt;/sup&gt;(A) : dNTP (O1A)</td>
<td>1.81</td>
<td>1.89</td>
<td>4.26</td>
<td>4.26</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>Mg&lt;sup&gt;2+&lt;/sup&gt;(A) : 256 (Oδ2)</td>
<td>1.81</td>
<td>1.87</td>
<td>1.83</td>
<td>1.83</td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td>Mg&lt;sup&gt;2+&lt;/sup&gt;(A) : 192 (Oδ1)</td>
<td>1.83</td>
<td>1.84</td>
<td>1.86</td>
<td>1.85</td>
<td>2.2</td>
<td></td>
</tr>
<tr>
<td>Nucleotide Magnesium ion Coordination</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mg&lt;sup&gt;2+&lt;/sup&gt;(B) : dNTP (O1A)</td>
<td>4.06</td>
<td>3.78</td>
<td>1.93</td>
<td>1.93</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Mg&lt;sup&gt;2+&lt;/sup&gt;(B) : 192 (Oδ2)</td>
<td>1.86</td>
<td>1.84</td>
<td>1.85</td>
<td>1.87</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Mg&lt;sup&gt;2+&lt;/sup&gt;(B) : dNTP (O2γ)</td>
<td>1.81</td>
<td>1.89</td>
<td>1.83</td>
<td>1.83</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Mg&lt;sup&gt;2+&lt;/sup&gt;(A) : Mg&lt;sup&gt;2+&lt;/sup&gt;(B)</td>
<td>4.07</td>
<td>4.26</td>
<td>4.38</td>
<td>4.37</td>
<td>n/a</td>
<td></td>
</tr>
</tbody>
</table>

Mg<sup>2+</sup>(A), catalytic magnesium; Mg<sup>2+</sup>(B), nucleotide binding magnesium; dNTP, 2'-deoxynucleoside 5'-triphosphate; P10, primer terminus; 
<sup>a</sup>Template : Incoming nucleotide; 
<sup>b</sup>PDB accession code 1BPY
Table 2: Base substitution fidelity of pol β taken from refs. [19,35]

<table>
<thead>
<tr>
<th>Enzyme Complex</th>
<th>Rate Constants(^+) (k_{pol})</th>
<th>Fidelity(^*)</th>
<th>Catalytic Efficiency (k_{pol} / K_d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G:dCTP</td>
<td>(s^{-1})</td>
<td>n/a</td>
<td>750,000</td>
</tr>
<tr>
<td>G:dTTP</td>
<td>0.0058</td>
<td>63,560</td>
<td>12</td>
</tr>
<tr>
<td>G:dGTP</td>
<td>0.00029</td>
<td>2,885,000</td>
<td>0.3</td>
</tr>
<tr>
<td>T:dTTP</td>
<td>0.00032</td>
<td>1,533,000</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\(^*\) Fidelity = error frequency\(^{-1}\) = [(\(k_{pol} / K_d\)_c + (\(k_{pol} / K_d\)_i)] / (\(k_{pol} / K_d\)_c), where c and i denote correct and incorrect nucleotide incorporation, and \(K_d [\mu M]\) is the apparent equilibrium dissociation constant of dNTP.

\(^+\) Rate Constant \(k_{pol}\): rate of nucleotide incorporation for first-enzyme turnover.
List of Figures

1 (a) General pathway for nucleotide insertion by pol β and (b, c) corresponding crystal open (bottom left) and closed (bottom right) conformations of pol β/DNA complex. E: DNA polymerase; dNTP: 2'-deoxyribonucleoside 5'-triphosphate; PPi: pyrophosphate; DNA\textsubscript{n}/DNA\textsubscript{n+1}: DNA before/after nucleotide incorporation to DNA primer. T6 is a template residue (G) for incoming nucleotide (dCTP).

2 (a) The starting structures of three mismatches G:T (purple), G:G (pink) and T:T (orange) in the pol β active site. The incoming nucleotides are in anti conformation in all mismatches (b) Comparison of one of the initial constructed intermediate model (gray) with respect to the crystallographic binary open (green) and ternary closed structures (red). Conformation of protein in all initial mismatch models is similar. Structures are superimposed with respect to the palm C\textalpha atoms.

3 Time evolution of the RMSD of α-helix N C\textalpha atoms in all trajectories (G:C, G:T, T:T, and G:G) relative to those in the crystal open (green) and closed (red) structures. All systems are superimposed by the palm subdomain (residue 152 to 258) C\textalpha atoms.

4 Comparison of the simulated systems with the closed crystal structure for (a) entire pol β and (b, c) α-helix N in the thumb subdomain in two orientations. All systems are superimposed with respect to palm C\textalpha atoms. The colored ribbons identify the closed (red) crystal structure (1BPY) and the simulated G:T (yellow), G:G (gray), T:T (turquoise), and G:C (blue), respectively.

5 The conformation of key active site residues in the simulated pol β/DNA complexes with different incoming nucleotides as observed during last 1 ns of G:C, G:T, G:G, and T:T trajectories, respectively.
DNA polymerase β mismatch simulations

6 Time evolution of dihedral angles associated with the Phe272 phenyl ring over the trajectories. The two dihedral angles of Phe272 are defined by the following atomic sequence: \( \chi_1 \) for \( \text{C-C}_\alpha\text{-C}_\beta\text{-C}_\gamma \) and \( \chi_2 \) for \( \text{C}_\alpha\text{-C}_\beta\text{-C}_\gamma\text{-C}_\delta \). In the crystal open conformation (1BPX) dihedral values \( \chi_1 \) and \( \chi_2 \) are 146° and 167°, respectively; in the crystal closed conformation (1BPY) values \( \chi_1 \) and \( \chi_2 \) correspond to 96° and 21°, respectively.

7 Time evolution of atom-atom distances for Arg283:NH1 with both Asp192:OD1 and Tyr296:OH in the mismtatched trajectories.

8 Time evolution of distances between the nascent basepair and protein residues Asp276, Asn279, and Arg283 in all mismatched trajectories. Distances for the correct basepair (G:C) are diagrammed (below) with dotted lines.

9 Conformational comparisons of active-site basepairs and two neighboring basepairs in the crystal closed (red) versus simulated G:C (blue), G:T (yellow), G:G (gray), and T:T (cyan) complexes of pol β/DNA. The nascent basepair plane in the closed crystal complex (red) is used for reference. Structures are superimposed using \( \text{C}_\alpha \) atoms of protein residues 10 to 335 of pol β. The DNA minor groove is nearest the viewer, whereas the major groove is farthest from the viewer. The arrows (green) depict the deviation from the Watson-Crick bonding. The labels on DNA bases correspond to the simulated structures.

10 Catalytic (\( \text{Mg}^{2+}(A) \)) and nucleotide (\( \text{Mg}^{2+}(B) \)) binding magnesium ion coordination environment reflected at the end of G:C, G:T, G:G, and T:T trajectories. Black lines are drawn to the coordinated oxygen atoms. Average distances over last one nanosecond for atoms coordinating the magnesium ions are listed in Table 1.
Figure 1
Figure 2
Figure 3
Figure 4
Figure 5
DNA polymerase β mismatch simulations

Figure 6

![Graph showing dihedrals of Phe272 for different base pairs (G:C, G:T, G:G, T:T) over time (ns).]
Figure 7
Figure 8
<table>
<thead>
<tr>
<th>Template</th>
<th>Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>G :C</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>G</td>
</tr>
<tr>
<td>C</td>
<td>G</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>dCTP</td>
</tr>
</tbody>
</table>

| G :T     | G      |
|          | C      |
|          | G      |
| C        | G      |
| G        | C      |
|          | dTTP   |

| G :G     | G      |
|          | C      |
|          | G      |
| C        | G      |
| G        | C      |
|          | dGTP   |

| T :T     | T      |
|          | C      |
|          | G      |
| C        | G      |
| G        | C      |
|          | dTTP   |

**Figure 9**
Figure 10
Mismatch Induced Conformational Distortions in Polymerase β Support an Induced-Fit Mechanism for Fidelity

Karunesh Arora, William A. Beard, Samuel H. Wilson, Tamar Schlick