December 20, 2004

Fidelity Discrimination in DNA polymerase β: differing closing profiles for a mismatched (G:A) versus matched (G:C) base pair

Ravi Radhakrishnan and Tamar Schlick*

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

Abstract

Understanding fidelity — the faithful replication or repair of DNA by polymerases — requires tracking of structural and energetic changes involved, including the elusive transient intermediates, for nucleotide incorporation at the template/primer DNA junction. We report using path sampling simulations and a reaction network model strikingly different transition states in DNA pol β's conformational closing for correct d**C**TP versus incorrect d**A**TP incoming nucleotide opposite a template **G**. The cascade of transition states leads to differing active-site assembly processes toward the "two-metal-ion catalysis" geometry. We demonstrate that these context-specific pathways imply different selection processes: while active site assembly occurs more rapidly with the correct nucleotide and leads to primer extension, the enzyme remains open longer, has a more transient closed state, and forms product more slowly when an incorrect nucleotide is present.

*Corresponding author: schlick@nyu.edu

Introduction

The delicate interplay between DNA damage and repair is crucial to the integrity of our genome and has immense biomedical repercussions to various cancers, neurological aberrations, and the process of premature aging. DNA polymerases are central to these functions because of their role in replication as well as excision repair pathways [1]. The eukaryotic DNA repair enzyme, polymerase β of the X-family, with thumb, palm, and fingers subdomains, binds to DNA and fills single-stranded gaps in DNA with moderate accuracy ('*fidelity*'). X-ray crystallography has provided exquisite views of the polymerase frozen-in-action: closed (active) and open (inactive) forms of the enzyme related by a large subdomain motion ($\sim$ 6Å) of the thumb [2]. By transitioning between the inactive and active forms, the enzyme recruits a nucleotide unit (dNTP, 2′-deoxyribonucleoside 5′-triphosphate) complementary to the template base (e.g., **C** opposite **G**) about 1000 times more often than the incorrect unit (e.g., **A** opposite **G**) [3–6]; each such cycle adds a nucleotide unit to the primer strand. This '*induced-fit*' mechanism in which the correct incoming base triggers the requisite conformational change, while an incorrect unit hampers the process, is thus crucial to our understanding of pol β's activity.

The temporal bridge connecting the crystal anchors cannot be completely determined from the kinetic studies and is not easily deduced from current single-molecule experiments. This bridge could reveal mechanistic details of the pol β reaction pathway, such as key slow motions and their significance to catalytic efficiency and fidelity associated with nucleotide insertion. Prior simulations using standard dynamics techniques have: suggested that key residues in the enzyme active site (e.g., Phe272, Arg258) exhibit subtle conformational rearrangements during the thumb's subdomain motion [7–10] and incorrect nucleotide incorpo-

rations [8]; rationalized observations for pol β mutants [11]; dissected the roles of both the nucleotide-binding as well as catalytic Mg^{2+} ions [12]; and delineated key transition-state regions and the associated cooperative dynamics in DNA pol β's closing [13]. Numerous experiments on pol β mutants [14] also reveal effects of localized mutations (Y265H, Y271F, Y271H, G274P, D276V, N279A, N279L, R283A, R283K, R283L) on the enzyme's efficiency and fidelity.

Building on knowledge accumulated from these prior modeling and experimental studies, we report here detailed structural and energetic characterization of transient intermediates along the closing pathway of a correct (**G:C**) versus incorrect (**G:A**) nucleotide incorporation before the chemical reaction of primer extension. This comparison in atomic detail is made possible by an application to biomolecules [13] of the transition path sampling method [15] coupled to an efficient method to compute reaction free energy [16] and a network model for reaction rates [17]. All these components have been validated for biomolecular applications, though the well-appreciated force-field approximations are relevant to all large-scale simulations. Extensive prior modeling, however, indicates that simulations can offer important insights to link structure-function relationship. Here, our analyses reveal the disparate barriers and active-site assembly processes that guide the enzyme's conformational change and thereby serve to help discriminate between error-free and error-prone repair processes. Specifically, the active site of the closed mismatch complex is unstable with respect to the open state, and together with different metastable basins before the chemically-competent transition state, this crucial difference hampers incorporation of incorrect nucleotide units.

In bridging the conceptual gap between crystal structures and kinetic data, our computational link can contribute valuable insights into the factors that affect fidelity discrimination

in DNA synthesis and repair. The cascade of transition states navigating base-pair selection can serve to guide crystallization experiments of key intermediates along the pathway, particularly of mismatches (e.g., [18,19]), to further interpret fidelity mechanisms. Taken together, these cooperative motions of key enzyme residues suggest sequential inspection and surveillance mechanisms which trigger context-specific and substrate-sensitive conformational, energetic, and dynamic pol β pathways.

Computational Methodology

A. System Preparation. Models of solvated pol β/DNA/d**C**TP (correct **G:C** system) and pol β/DNA/d**A**TP (incorrect **G:A** system) complexes were prepared from 1BPX (open binary) and 1BPY (closed ternary) crystal structures [2]. Hydrogen atoms were added by CHARMM's subroutine HBUILD [20]. Also added were: a hydroxyl group to the 3′ terminus of the primer DNA strand, missing residues 1–9 of pol β, and specific water molecules coordinated to the catalytic Mg^{2+} (missing in the ternary complex). For the **G:C** system, the open complex was modified by incorporating the incoming unit d**C**TP (deoxyribocytosine 5′-triphosphate) with nucleotide-binding Mg^{2+}, producing the 1BPX ternary complex. For the **G:A** system, the open and closed complexes were built from the **G:C** system by replacing the incoming d**C**TP by d**A**TP and orienting the d**A**TP in an *anti* conformation, following the crystal structure of pol β with a mispair in the active site [18]. We note that the same mismatch (template **G**, incoming **A**) was found to be in the *anti-anti* conformation in the crystal complexes of high-fidelity Bacillus DNA polymerase I solved in the Beese group [19].

Cubic periodic domains for both initial models were constructed using Simulaid and PBCAID [21]. To neutralize the system at an ionic strength of 150 mM, water molecules

with minimal electrostatic potential at the oxygen atoms were replaced by Na^+, and those with maximal electrostatic potential were replaced with Cl^-. All Na^+ and Cl^- ions were placed more than 8 Å away from any protein or DNA atoms and from each other. The electrostatic potential for all bulk oxygen atoms was calculated with DelPhi. The resulting system has 40238 atoms (including 11249 water molecules). Consistent with a pH value of 7.0, we assume de-protonated states (i.e., -1 charge each) for Asp190, Asp192, and Asp256, as made recently [22]. Appendix A[†] provides protonation states of titratable side chains with discussion. These settings produce a net charge of $+7$ for pol β, -29 for DNA, and -4 for the dNTP. There are 42 Na^+ ions, 20 Cl^- ions, 2 Mg^{2+} ions, producing a charge of 26 and overall neutral system.

B. Minimization, Equilibration and Dynamics Protocol. Energy minimizations, equilibration, and dynamics simulations were performed using the program CHARMM [20, 23] and the all-atom version c28a3 forcefield (Chemistry Department, Harvard University, Cambridge, MA). The system was minimized using the Steepest Descent method for 10,000 steps followed by Adapted Basis Newton-Raphson [20,24] for 20,000 steps. The system was then equilibrated for 1 nanosecond at room temperature by the Verlet integrator in CHARMM prior to dynamics production runs.

C. Transition Path Sampling (TPS) and BOLAS Free Energy Protocols. Very recently, we have developed a general protocol for harvesting mechanistic pathways for macromolecular systems by transition path sampling [15,25] using a divide-and-conquer approach. Our protocol has developed strategies to [13]: (i) generate initial trajectories, (ii) identify the different transition state regions, (iii) implement transition path sampling sampling in

[†]Appendices are provided under supplementary information.

conjunction with CHARMM [20], (iv) assess convergence, and (v) compute the reaction free energy pathway [16]. In this work, we also develop and apply in this context a network model for generating reaction rates as a result of the combined transition states (see Section E below, and Fig. 4). Details of TPS and BOLAS are available in Refs. [13,16].

As described in Ref. [13] (see also Appendix B), trajectories in each transition state region are harvested using the shooting algorithm [25] to connect two metastable states via a Monte Carlo protocol in trajectory space. In each shooting run, the momentum perturbation size $dP \approx 0.002$ in units of amu$\times$Å/fs is used to yield an acceptance rate of 25 to 30%.

For the free energy calculation [16], the probability distribution $P(\chi_i)$ was calculated by dividing the range of order parameter χ_i into 10 windows. The histograms for each window are collected by harvesting 300 (accepted) trajectories per window according to the procedure outlined in Ref. [16], from which the potential of mean force $\Lambda(\chi_i)$ is calculated (Appendix B). The arbitrary constant associated with each window is adjusted to make the Λ function continuous. The standard deviation in each window of the potential of mean force calculations is estimated by dividing the set of trajectories in two blocks and collecting separate histograms. The statistical error of $\pm 3k_BT$ in the free energy is estimated from the order parameter window showing the maximum standard deviation in the potential of mean force.

D. Mixed Quantum Mechanics and Molecular Mechanics (QM/MM) Simulations of Pol β's Active-Site. The QM/MM approach we adopt is based on an existing interface between GAMESS-UK [26] (an *ab-initio* electronic structure prediction package) and CHARMM (see Appendix C for details). To focus on the active site of a solvated pol β system with correct (**G:C**) and incorrect (**G:A**) base pairs, we define the quantum region

(see circled area of Fig. 3 later) as the two Mg^{2+} ions; the conserved aspartates 190, 192, 256; incoming nucleotide; terminal primer of DNA; Ser 180; Arg 183; and water molecules within hydrogen bonding distance of the QM atoms. These 86 atoms are treated in accord with density functional theory using a B3LYP density functional and 6-311G basis set. The molecular mechanical region consists of the rest of the protein, DNA, Na^+, Cl^-, and solvent molecules up to three solvation shells (extending over 12 Å) adjoining the protein/DNA/dNTP complex. Wave function optimizations in the QM region are performed according to a density functional formalism, and geometry optimizations of the whole system were performed using the Adopted Basis Newton Raphson method implemented in CHARMM.

E. Analysis of Reaction Kinetics Using the Gillespie Algorithm. We use Gillespie's method [17] to simulate the time evolution of our system in terms of a network of elementary chemical reactions. In contrast to the traditional approach of treating a network of reactions as a deterministic system and solving a coupled set of differential equations to obtain the temporal evolution of concentrations of species, Gillespie's method considers reaction kinetics profile as a random walk governed by the master equation; thus, reaction events occur with specified probabilities, and each event alters the probabilities of subsequent events. The stochastic nature of the calculation is crucial if the absolute number of reactant or intermediate species in the system is not large.

As detailed in Appendix D, the elements in the reaction network consist of the identified metastable states between the open and closed conformations of the enzyme and, in addition, the chemical reaction step. The rate constants for hopping between adjacent metastable states are obtained from the BOLAS free energy computations for the **G:C** and **G:A** systems and that for the chemical step derived from experimentally measured $k_{\rm pol}$ values [3–5,27–35].

A summary of the free energy barriers, rate constants, and $k_{\rm pol}$ values for the matched (**G:C**) system and the mismatch (**G:A**) system is provided in Table I later. These network simulations are performed using the STOCKS simulator [36], with a timestep corresponding to one hundredth of the timescale of the fastest transition. One hundred independent trajectories of evolution were simulated to account for the stochastic nature inherent in the kinetic model.

Results

Transition State Identification. Our analyses of detailed closing pathways before the chemical reaction identify five transition states for the matched **G:C** system and four transition states for the mismatched **G:A** system (Fig. 1). Fig. 1

The following order parameters (χ_1–χ_5) characterize the reaction profiles [13]: χ_1 = measuring the root mean square deviation (RMSD) of heavy atoms in amino acid residues 275–295 in pol β that form the thumb's helix N [2] with respect to the same atoms in the enzyme's closed state (1BPY); χ_1 varies from $\sim$ 6 to 1.5 Å between the open and closed states. χ_2 is the dihedral angle defined with respect to the quadruplet of atoms C_γ–C_β–C_α–C in Asp192; χ_2 varies from $\sim$ 90 to 180° between the unflipped and the flipped states of Asp192. χ_3 is dihedral angle C_γ–C_δ–N_ϵ–C_ζ, describing the rotation state of Arg258 (dihedral angle C_β–C_γ–C_δ–N_ϵ is an alternative); χ_3 varies between $\sim$100 and 260° between the unrotated and fully-rotated states of Arg258. (A partially-rotated state ($\chi_3 \approx 180°$) is also observed as a metastable state). χ_4 is dihedral angle $C_{\delta 1}$–C_γ–C_β–C_α describing Phe272; χ_4 varies from ~ -50 to 50° between the unflipped and the flipped states of Phe272. χ_5 is the distance between the nucleotide binding Mg^{2+} ion and the oxygen atom $O1_\alpha$ of d**C**TP; χ_5 varies between 3.5 Å and 1.5 Å in a subtle ion-rearrangement in the catalytic region.

These five order parameters characterize all transition states. For **G:C**, TS 1 is the closing of the thumb, TS 2 is the Asp192 flip, TS 3 is the partial rotation of Arg258, TS 4 is the Phe272 flip, and TS 5 defines a subtle ion-rearrangement in the catalytic region involving the Mg^{2+} ions and is also associated with the stabilization of Arg258 in its fully rotated state (TS 3 is the partial rotation) which accompanies the ion-rearrangement. For the mismatch, TS 1 is the thumb closing, TS 2 is Asp192's flip, TS 3 is the Phe272 flip, and TS 4 is the complete rotation of Arg258. Significantly, the ion-rearrangement step is absent in the mismatched system, resulting in a disordered active-site that is further away (in space) from the reaction competent state (Fig. 1).

The sequence of the transition states along the closing pathways for both the **G:C** and **G:A** systems is determined by ranking the values of the order parameters in the metastable states using a histogram analysis [13]; for example, if χ_2 changes from unflipped (in open) to flipped (in closed) values while χ_3, χ_4, χ_5 remain at values of the open structure, we know that TS 2 precedes TS 3–5. For the **G:C** pathway, the order of events was: thumb closing, Asp192 flip, partial rotation of Arg258, Phe272 flip, and ion-rearrangement in the catalytic region. For the **G:A** pathway, the thumb closing was followed by Asp192 flip, Phe272 flip, and rotation of Arg258. Thus, the Phe272 flip occurs before rather than after the Arg258 rotation in the mismatch system, and the crucial ion-rearrangement is lacking for **G:A** compared to **G:C**.

Pathway Analysis. The striking differences in the sequence of events in these pathways are evident by analyzing the conformational states visited in the dynamics trajectories (Fig. 2). Here, we describe the conformational landscapes of the Arg258 rotation and Phe272 flip for each system as contour plots of the function $-\ln P(\chi_3, \chi_4)$, where χ_3, χ_4 are the dihedral

Fig. 2

angles characterizing Arg258 rotation, Phe272 flip, and $P(\chi_3, \chi_4)$ is the two-dimensional (2D) probability distribution; $P(\chi_3, \chi_4)$ is calculated by accumulating a 2D histogram (of χ_3 and χ_4) using the harvested trajectories. The metastable states correspond to the blue basins (high probability states), while the transition states correspond to the saddles. Red regions correspond to low probability (high free energy) states. Physical pathways capturing the Arg258's rotation and Phe272's flip are those that connect basins (A1,P1) and (A3,P2) and pass through saddle regions. In both the **G:C** and **G:A** cases, the Arg258 rotation occurs in two steps (through the intermediate state A2). The Phe272 flip follows the partial rotation of Arg258 for the **G:C** case, while the order is reversed for the **G:A** case. Even more interesting, the **G:C** landscape exhibits a unique pathway with most of the local minima lying along the physical pathway, while the **G:A** landscape reveals multiple paths as well as other local minima that do not lie along any R258, F272 rotation/flip pathway.

Potential of Mean Force Calculations. The free energy changes associated with each transition event (Table I) are derived from the potential of mean force calculations (see methodology, and Figs. B2 and B3 of Appendix B), which are used to construct the overall reaction kinetics profiles below.

Tab. I

The conformational profile (Fig. 1) reveals key differences in thermodynamic quantities for **G:C** vs. **G:A**. Despite similar overall activation free-energy barriers for conformational change ($19 \pm 3\ k_BT$), the closed state is thermodynamically stable for the correct **G:C** system, but only metastable (i.e., higher in free energy by $9\ k_BT$ than the open state) for the **G:A** mispair. Significantly, these barriers are of the same order of magnitude as that for the overall process (including chemistry): 25–28 k_BT for **G:C**, 36 k_BT for **G:A**, obtained from the experimentally measured rate-constants $k_{\rm pol} = 3$–$100\ {\rm s}^{-1}$ and $0.002\ {\rm s}^{-1}$, respectively

[3–6], where $k_{\rm pol}$ is related to the free energy barriers ΔF via $k_{\rm pol} = (k_B T/h) \times \exp(-\beta \Delta F)$. These values suggest that even if the chemical step and not the conformational step is rate limiting, the conformational rearrangements prior to the actual chemical reaction direct the system to the reaction-competent state in a substrate-sensitive manner and cause the mismatched closed state to be unstable. This relative instability, together with the disordered catalytic active site for the mispair (Fig. 3), produce the different free energy barriers and hence overall fidelity discrimination. Fig. 3

That substrate-induced conformational changes directly affect fidelity of some moderate and high-fidelity polymerases is consistent with the notion of geometric selection criteria, as recently shown for T7 polymerase encountering a lesion, which blocks the conformational change [37]. Exquisite crystal structures of mismatches for the Bacillus fragment also reflect the relationship between geometric distortion and fidelity [19].

Discussion

The highly cooperative dynamics associated with the conformational transition for pol β trigger systematic differences in the evolution of active-site geometries near the closed state (Figure 3). According to the two-metal-ion-catalyzed phosphoryl transfer mechanism [38], for which functional, kinetic evidence was provided by Bolton et. al. [39], the conserved protein residues of pol β (Asp190, Asp192, and Asp256) strategically orient the Mg^{2+} ions with respect to the P_α of the d**C**TP and the O3′ hydroxy terminal of last residue of the DNA primer (see Fig. 3). Our computed geometry of the catalytic region in the closed conformation for **G:C** is consistent with the two-metal-ion catalytic mechanism *with the exception of O3′-P_α distance*, which on average is 1.2 Å larger than the ideal distance of 3.2 Å for phos-

phoryl transfer by a dissociative mechanism [40,41]. The corresponding distance for the **G:A** mispair is significantly higher: 5.5 Å; the crucial distance between the catalytic Mg^{2+} and the nucleophilic O3′ oxyanion approaches different values (higher by about 1.8 Å for **G:A** vs. **G:C**). Our combined quantum/classical optimizations (Appendix C) have shown that these differences are not mere force-field artifacts. Although subtle differences in the geometries resulting from CHARMM27 and the QM/MM were observed (Fig. C1), the essential differences in the active-site assembly between the **G:C** and **G:A** systems were preserved, suggesting that they indeed likely contribute to the discrimination of the incorrect substrate at the active-site. Moreover, the observed geometry for the correct substrate (**G:C** system) suggests a likely pathway for the initial proton abstraction by Asp256 — a network of hydrogen bonds through two mediating water molecules separating the Oδ2 atom of Asp256 and the O3′ atom — strongly implicating a concerted proton transfer mechanism for the deprotonation of the O3′ group. The possible occurrence of this proton transfer as a first step of the nucleotide incorporation reaction is consistent with the tour-de-force calculation by Warshel et. al. of the energy profiles for nucleotide incorporation in T7 polymerase [22], as well as the recent QM/MM study of the phosphoryl transfer catalyzed by β–phosphoglocomutase by Webster [42].

Thus, our data concerning the existence and impact of different coordination networks and their evolution toward the closed state compatible for the chemical reaction support early suggestions based on NMR studies by the Mildvan group; namely, that error verification or prevention steps following substrate binding but prior to primer elongation involve coordination of the enzyme-bound metal by the α and β-phosphoryl groups [43,44].

In addition, we find that the intimate interaction of the charged residue Arg283 (Fig. 3)

with the nucleotide binding pocket (particularly the template residue) is different in the **G:C** vs. **G:A** cases, suggesting an energetic discrimination (causing a destabilization of the mismatch) in the environment of the nucleotide binding pocket. These cumulative differences likely produce the higher barrier for **G:A** in the chemical incorporation step and also explain why Arg283 mutant experiments reveal crucial effects on fidelity [6,45].

The difference in thermodynamic stabilities of the closed conformation between **G:C** and **G:A** systems (of 9 k_BT, see Figure 1) is almost quantitatively rationalized by the different active-site geometries (Figure 3). The reduced electrostatic interaction (for the mismatch **G:A** system) between the O3′ of the terminal base of primer DNA and catalytic Mg^{2+} and that between C_ζ of Arg283 and N_3 of template base opposite the incomer each account for a lowering of 4 k_BT energy. These simple estimates are based on differences in distances (r) translated into energies (E) by the application of Coulomb's law of electrostatics ($E \propto 1/r$). Correspondingly, we expect the catalytic Mg^{2+} ion and Arg283 to play significant roles in stabilizing the closed conformation for the correct substrate (**G:C** system). *In silico* evidence for the former comes from dynamics studies [12] which suggest that *closing* before the chemical incorporation requires both divalent metal ions in the active site while *opening* after chemical incorporation is triggered by release of the catalytic metal ion. The importance of Arg283 in pol β's activity is already appreciated from the studies of the enzyme mutant R283A, which exhibits a 5000-fold decrease in catalytic efficiency and a 160-fold decrease in fidelity, in comparison to wildtype [6]. These combined studies lend additional support to the group-contribution view (of template stabilized discrimination [45]) in rationalizing the stability differences between pol β complexes with correct and incorrect substrates.

We further demonstrate the significance of the cascade of subtle events orchestrating the

active-site assembly for the correct vs. mismatch systems prior to the chemical incorporation and subsequent catalysis by solving a network model of elementary chemical reactions (inset in Figure 4 and Appendix D) to produce the overall rates for the combined process. The Fig. 4 elementary reactions and associated rate constants used to construct the entire reaction evolution are gleaned from the transition state and free energy estimates (Table I, also see Appendix D). The resulting profiles in Fig. 4 reveal a striking difference in the evolution of reactants, products, and reaction intermediates between the **G:C** and **G:A** systems. The (blue) curve corresponding to the open enzyme state for the matched system rapidly disappears, with the closed state (red band, MS 7) quickly emerging and transitioning into product (black band, MS 8), where d**C**TP has been incorporated into the primer strand opposite the template guanine residue. Around 0.1 s, which corresponds to the experimental $k_{\rm pol}$ of 10 $\rm s^{-1}$ [3–5,27], the product curve sharply rises, until all species are product ($\sim$ 1 s).

For the mismatched system, in contrast, the open enzyme state (blue band) disappears very slowly, and the closed enzyme state (red) disappears sharply due to its instability. The product forms much more slowly (black band), in significant amount (>67 %, the time at which the percentage of product species reaches this level corresponds to $k_{\rm pol}^{-1}$) only around 500 s, corresponding to $k_{\rm pol} = 0.002$ $\rm s^{-1}$ [3–5,27] for the mismatched **G:A** system.

Taken together, these network solutions show that the overall rate is sensitive to the relative stability of the closed state with respect to the open state; a loss by a factor of 10 in rate corresponds to a 3 k_BT decrease in stability. Since the transition events follow in sequence, the individual k_{ij}s can affect this relative stability in numerous ways and all corroborate to produce the overall difference in fidelity discrimination.

In conclusion, the unraveled cascade of transition states during the closing pathways of

correct (**G:C**) system versus incorrect (**G:A**) pol β systems suggests crucial differences in the evolution of the active-site assembly towards the two-metal-ion transition state geometry. The more highly distorted and notably less stable active-site for the mismatch (by 9 k_BT) establishes a source of discrimination and hence selection criteria for the incoming nucleotide unit. Subject to the force-field approximations and statistical inaccuracies [13], we identify similar barriers to closing prior to chemical incorporation for both systems (19 k_BT). These, taken together with overall barriers inferred from experimentally measured k_{pol} values (27 k_BT for **G:C** and 36 k_BT for **G:A** [3–5,27]), point to a rate-limiting chemical incorporation step for both systems. However, work in progress suggests that this rate-limiting step occurs subsequent to nucleotide alignment in the active-site and just prior to the actual (phosphoryl transfer) reaction, in a "pre-chemistry avenue" that slowly adjusts metal/phosphoryl coordination at the active-site [Schlick et. al., unpublished results]. Recent NMR studies [46] that indicate localized motions near methionine residues (of which Met191 and Met155 are proximal to metal binding ligands Asp190 and Asp192) along with pioneering work by the Mildvan group [44] corroborate this intriguing possibility that error verification or prevention steps — following substrate binding but prior to primer elongation — involve coordination of enzyme-bound metal-ions by the alpha and beta-phosphoryl groups. Like Alice in Wonderland searching for the golden key to unlock the mysterious doors, the search for unraveling fidelity mechanisms is revealing a sequence of gates — "paths" in conformational, pre-chemistry, and chemistry — through which the passage has crucial biological ramifications. Further experimental and modeling studies are underway to explore the existence of the pre-chemistry avenue and detail the chemical reaction pathway.

Acknowledgments

We thank Samuel Wilson and William Beard for helpful comments on this work, and Linjing Yang and Karunesh Arora for many stimulating discussions throughout this work. We thank Martin Karplus for use of the CHARMM program and Bernard Brooks, Martin Guest, and Paul Sherwood for help with GAMESS-UK. Acknowledgment is made to NIH grant R01 GM55164, NSF grant MCB-0239689, and the donors of the American Chemical Society Petroleum Research Fund for support of this research. Computational resources were partly provided by the Advanced Biomedical Computing Center.

Literature cited

1. Friedberg, E. C. (2003) *Nature* **421**, 436–439.

2. Sawaya, M. R, Prasad, R, Wilson, S. H, Kraut, J, & Pelletier, H. (1997) *Biochemistry* **36**, 11205–11215.

3. Ahn, J, Kraynov, V. S, Zhong, X, Werneburg, B. G, & Tsai, M.-D. (1998) *Biochem. J.* **331**, 79–87.

4. Vande Berg, B. J, Beard, W. A, & Wilson, S. H. (2001) *J. Biol. Chem.* **276**, 3408–3416.

5. Shah, A. M, Li, S.-X, Anderson, K. S, & Sweasy, J. B. (2001) *J. Biol. Chem.* **276**, 10824–10831.

6. Beard, W. A, Osheroff, W. P, Prasad, R, Sawaya, M. R, Jaju, M, Wood, T. G, Kraut, J, Kunkel, T. A, & Wilson, S. H. (1996) *J. Biol. Chem.* **271**, 12141–12144.

7. Yang, L, Beard, W. A, Wilson, S. H, Broyde, S, & Schlick, T. (2002) *J. Mol. Biol.* **317**, 651–671.

8. Yang, L, Beard, W. A, Wilson, S. H, Roux, B, Broyde, S, & Schlick, T. (2002) *J. Mol. Biol.* **321**, 459–478.

9. Arora, K & Schlick, T. (2003) *Chem. Phys. Lett.* **378**, 1–8.

10. Rittenhouse, R. C, Apostoluk, W. K, Miller, J. H, & Straatsma, T. P. (2003) *Phys. Rev. Lett.* **53**, 667–682.

11. Yang, L, Beard, W. A, Wilson, S. H, Broyde, S, & Schlick, T. (2004) *Biophysical J.* **86**, 3392–3408.

12. Yang, L, Arora, K, Beard, W. A, Wilson, S. H, & Schlick, T. (2004) *J. Am. Chem. Soc.* **126**, 8441–8453.

13. Radhakrishnan, R & Schlick, T. (2004) *Proc. Nat. Acad. Sci.* **101**, 5970–5975.

14. Kunkel, T. A & Bebenek, K. (2000) *Annu. Rev. Biochem.* **69**, 497–529.

15. Dellago, C, Bolhuis, P. G, & Geissler, P. L. (2002) *Adv. Chem. Phys* **123**, 1–81.

16. Radhakrishnan, R & Schlick, T. (2004) *J. Chem. Phys.* **121**, 2436–2444.

17. Gillespie, D. T. (1977) *J. Phys. Chem.* **81**, 2340–2361.

18. Krahn, J. M, Beard, W. A, & Wilson, S. H. (2004) *Structure* **12**, 1823–1832.

19. Johnson, S. J & Beese, L. S. (2004) *Cell* **116**, 803–816.

20. Brooks, B. R, Bruccoleri, R. E, Olafson, B. D, States, D. J, Swaminathan, S, & Karplus, M. (1983) *J. Comp. Chem.* **4**, 187–217.

21. Qian, X, Strahs, D, & Schlick, T. (2001) *J. Comp. Chem.* **22**, 1843–1850.

22. Florián, J, Goodman, M. F, & Warshel, A. (2003) *J. Am. Chem. Soc.* **125**, 8163–8177.

23. MacKerell, Jr., A. D & Banavali, N. K. (2000) *J. Comp. Chem.* **21**, 105–120.

24. Schlick, T. (1992) in *Reviews in Computational Chemistry*, eds. Lipkowitz, K. B & Boyd, D. B. (VCH Publishers, New York, NY) Vol. III, pp. 1–71.

25. Bolhuis, P. G, Dellago, C, & Chandler, D. (1998) *Faraday Discuss.* **110**, 421–436.

26. Schmidt, M. W, Baldridge, K. K, Boatz, J. A, Elbert, S. T, Gordon, M. S, Jensen, J. J, Koseki, S, Matsunaga, N, Nguyen, K. A, Su, S, Windus, T. L, Dupuis, M, & Montgomery, J. A. (1993) *J. Comput. Chem.* **14**, 1347–1363.

27. Suo, Z & Johnson, K. A. (1998) *J. Biol. Chem.* **273**, 27250–27258.

28. Kraynov, V. S, Werneburg, B. G, Zhong, X, Lee, H, Ahn, J, & Tsai, M.-D. (1997) *Biochem. J.* **323**, 103–111.

29. Zhong, X, Patel, S. S, Werneburg, B. G, & Tsai, M.-D. (1997) *Biochemistry* **36**, 11891–11900.

30. Dahlberg, M. E & Benkovic, S. J. (1991) *Biochemistry* **30**, 4835–4843.

31. Kuchta, R. D, Mizrahi, V, Benkovic, P. A, Johnson, K. A, & Benkovic, S. J. (1987) *Biochemistry* **26**, 8410–8417.

32. Wong, I, Patel, S. S, & Johnson, K. A. (1991) *Biochemistry* **30**, 526–537.

33. Patel, S. S, Wong, I, & Johnson, K. A. (1991) *Biochemistry* **30**, 511–525.

34. Frey, M. W, Sowers, L. C, Millar, D. P, & Benkovic, S. J. (1995) *Biochemistry* **34**, 9185–9192.

35. Capson, T. L, Peliska, J. A, Kaboord, B. F, Frey, M. W, Lively, C, Dahlberg, M, & Benkovic, S. J. (1992) *Biochemistry* **31**, 10984–10994.

36. Kierzek, A. M. (2002) *Bioinformatics* **18**, 470–481.

37. Li, Y, Dutta, S, Doubli, S, Bdour, H. M, Taylor, J, & Ellenberger, T. (2004) *Nat. Struc. Biol.* **11**, 784–790.

38. Steitz, T. A & Steitz, J. A. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 6498–6502.

39. Bolton, E. C, Mildvan, A. S, & Boeke, J. D. (2002) *Molec. Cell* **9**, 879–889.

40. Mildvan, A. S. (1997) *Proteins: Stru. Fun. Gen.* **29**, 401–416.

41. Lahiri, S. D, Zhang, G, Dunaway-Mariano, D, & Allen, K. N. (2003) *Science* **299**, 2067–2071.

42. Webster, C. E. (2004) *J. Am. Chem. Soc.* **126**, 6480.

43. Ferrin, L. J & Mildvan, A. S. (1986) *Biochemistry* **25**, 5131–5145.

44. Ferrin, L. J & Mildvan, A. S. (1987) *DNA replication and recombination.* (Alan R. Liss, Inc.), pp. 75–87.

45. Beard, W. A & Wilson, S. H. (1998) *Chem. Biol.* **5**, R7–R13.

46. Bose-Basu, B, DeRose, E. W, Kirby, T. W, Mueller, G. A, Beard, W. A, Wilson, S. H, & London, R. E. (2004) *Biochemistry* **43**, 8911–8922.

Figure 1: Overall captured reaction kinetics profile for pol β's closing transition followed by chemical incorporation of dNTP for **G:C** and **G:A** systems. The barriers to chemistry (dashed peaks) are derived from experimentally measured $k_{\rm pol}$ values [3–5,27]. The profiles were constructed by employing reaction coordinate characterizing order parameters (χ_1–χ_5) in conjunction with transition path sampling (Appendix B). The order parameters χ_1–χ_5 serve as reaction coordinates to characterize the transition states TS 1–TS 5 in the matched **G:C** system, as well as TS 1–TS 4 in the mismatched **G:A** system. The potential of mean force along each reaction coordinate is computed for each conformational event (Appendix B, Figs. B2 and B3). The relative free energies of the metastable states and the free-energy barrier characterizing each transition state are calculated with BOLAS [16].

Figure 2: Conformational landscapes for the rotation and flipping of Arg258 and Phe272, in the conformational closing pathway of pol β for **G:C** versus **G:A** systems.

Figure 3: Evolution of average distances of ligands coordinating the catalytic and nucleotide-binding Mg^{2+} ions along the reaction coordinate for **G:C** and **G:A**. Metastable states 1 to 7 evolve the system in the closing pathway. The extent of thumb closing (χ_1 at top), and a crucial distance for the chemical reaction (O3$'$ of last primer (guanine) residue to P_α of d**C**TP in bottom plot) are also provided. Coordination and distances are diagrammed on the *right*: catalytic site ready for the phosphoryl transfer reaction. Circled area represents the QM region

Figure 4: Comparison of reaction kinetics for **G:C** and **G:A** systems. The temporal evolution of open (blue band), closed (red band) and product (black band) species are derived based on 100 evolution trajectories from binary (open) complexes. Inset describes the reaction networks according to profiles in Fig. 1. The networks are solved with the stochastic algorithm of Gillespie [17] (Appendix D). The spread in the kinetics (thickness of bands shown) represents the inherent stochasticity of the system.

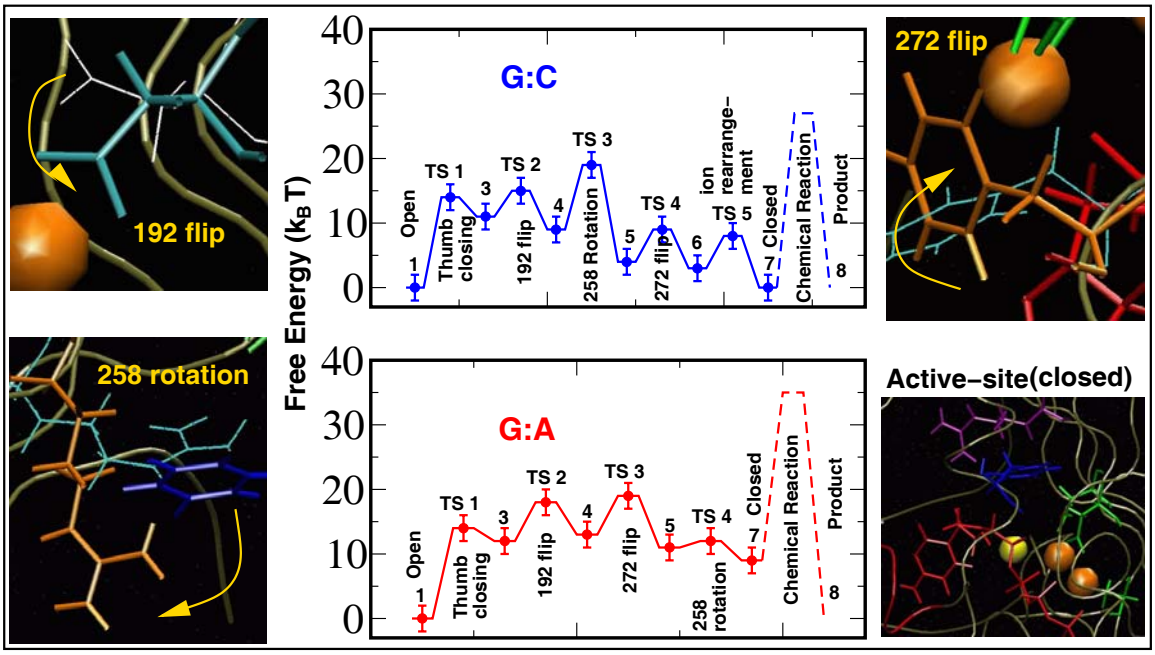
192 flip
258 rotation
272 flip
Active−site(closed)
G:C
G:A
Free Energy (k_B T)
40
30
20
10
0
Open
Thumb closing
192 flip
258 Rotation
272 flip
ion rearrange−ment
Closed
Chemical Reaction
Product
TS 1
TS 2
TS 3
TS 4
TS 5
258 rotation
1
2
3
4
5
6
7
8

Figure 1

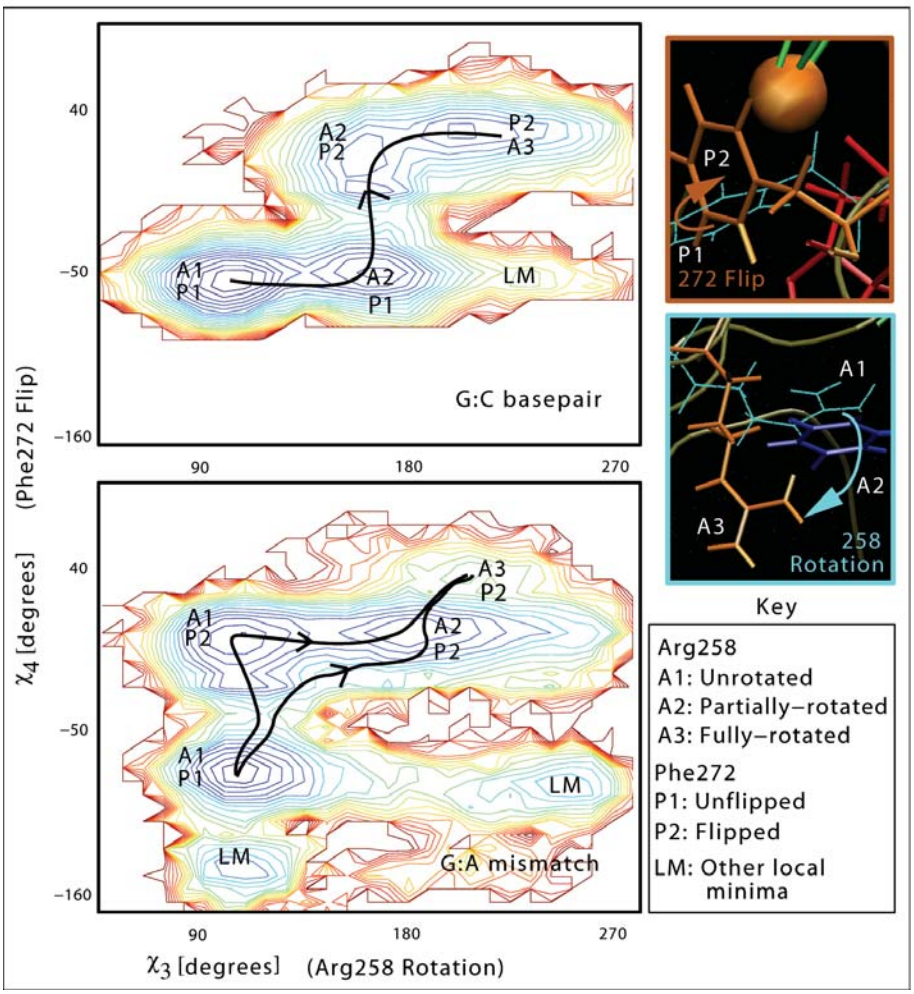
40
−50
−160
90
180
270
A2
P2
P2
A3
A1
P1
A2
P1
LM
G:C basepair
χ4 [degrees] (Phe272 Flip)
A3
P2
A1
P2
A2
P2
A1
P1
LM
LM
G:A mismatch
χ3 [degrees] (Arg258 Rotation)
P2
P1
272 Flip
A1
A2
A3
258
Rotation
Key
Arg258
A1: Unrotated
A2: Partially−rotated
A3: Fully−rotated
Phe272
P1: Unflipped
P2: Flipped
LM: Other local minima

Figure 2

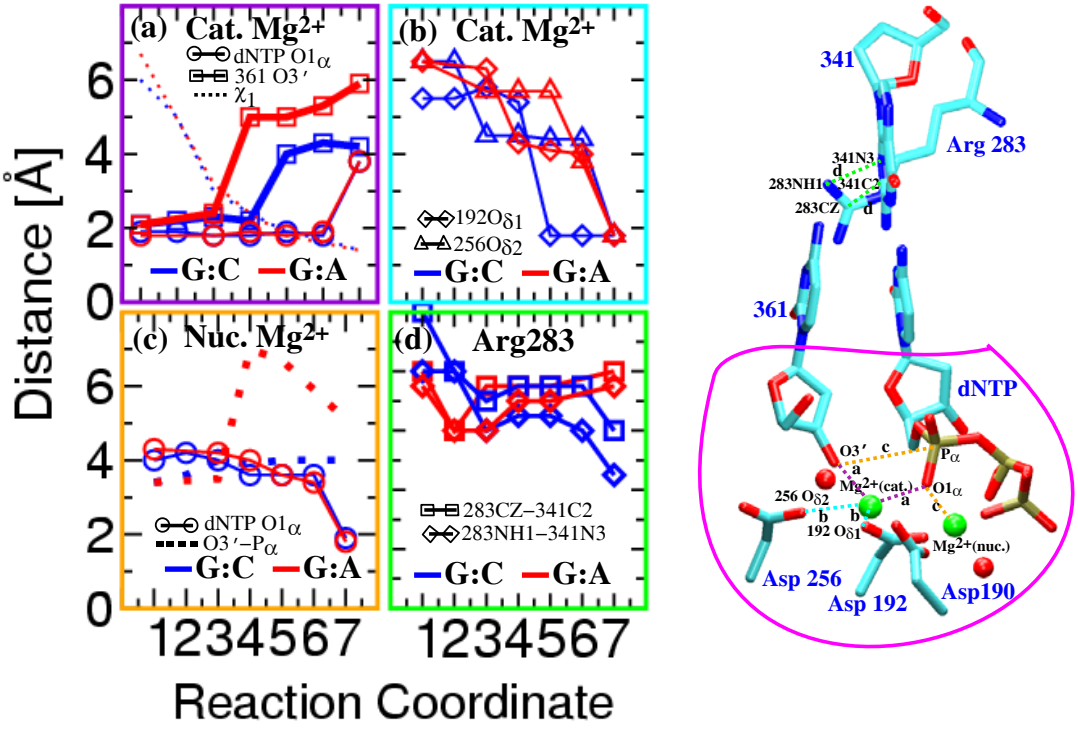
(a) Cat. Mg2+
dNTP O1α
361 O3′
χ1
G:C G:A
(b) Cat. Mg2+
192Oδ1
256Oδ2
G:C G:A
(c) Nuc. Mg2+
dNTP O1α
O3′–Pα
G:C G:A
(d) Arg283
283CZ–341C2
283NH1–341N3
G:C G:A
Distance [Å]
0
2
4
6
1234567
1234567
Reaction Coordinate
341
Arg 283
341N3
283NH1
341C2
283CZ
361
dNTP
O3′
Pα
Mg2+(cat.)
O1α
256 Oδ2
192 Oδ1
Mg2+(nuc.)
Asp 256
Asp 192
Asp190

Figure 3

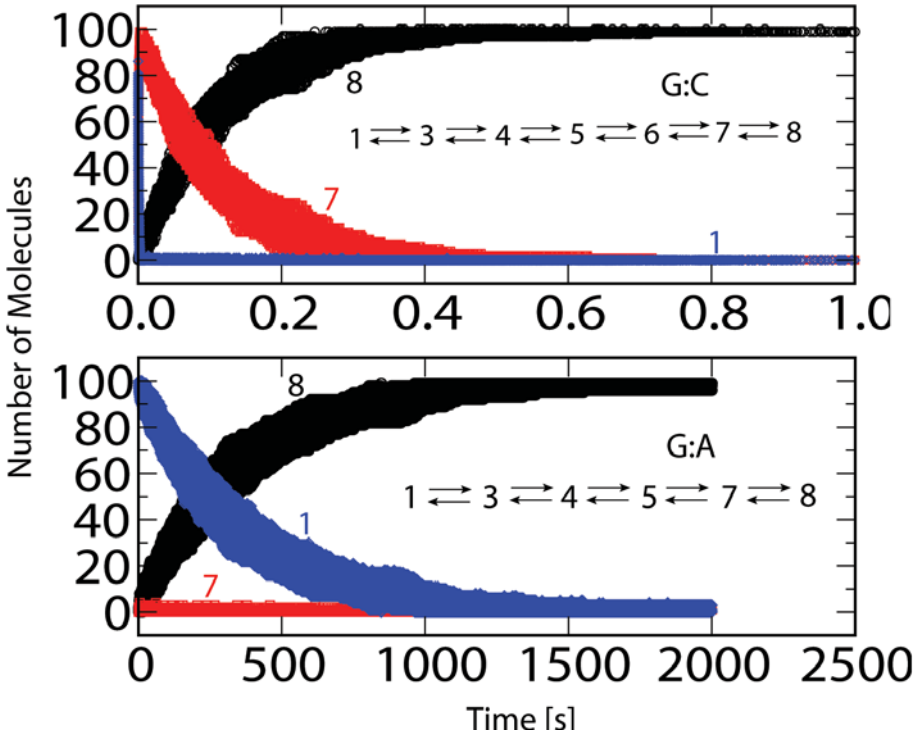
Number of Molecules
100
80
60
40
20
0
8
G:C
1 ⇄ 3 ⇄ 4 ⇄ 5 ⇄ 6 ⇄ 7 ⇄ 8
7
1
0.0
0.2
0.4
0.6
0.8
1.0
100
80
60
40
20
0
8
G:A
1 ⇄ 3 ⇄ 4 ⇄ 5 ⇄ 7 ⇄ 8
1
7
0
500
1000
1500
2000
2500
Time [s]

Figure 4

Table I: Rates k_{TST} estimated by transition state theory

	TS1	TS2	TS3	TS4	TS5	TS6
Matched **G:C** system						
τ_{mol} [ps]	70	4	4	3	0.2[a]	0.2[a]
$\beta\Delta F_{AB}^{\mathrm{barrier}}$	14	4	10	5	5	27
$\beta\Delta F_{BA}^{\mathrm{barrier}}$	3	6	15	6	8	–
$k_{\mathrm{TST}}^{A\to B}$ [s^{-1}]	1.2×10^{4}	4.6×10^{9}	1.1×10^{7}	2.2×10^{9}	3.3×10^{10}	9.4
$k_{\mathrm{TST}}^{B\to A}$ [s^{-1}]	7.1×10^{8}	6.2×10^{8}	7.6×10^{4}	8.3×10^{8}	1.7×10^{9}	0
Mismatched **G:A** system						
τ_{mol} [ps]	70	4	4	3		0.2[a]
$\beta\Delta F_{AB}^{\mathrm{barrier}}$	14	5	6	1		27
$\beta\Delta F_{BA}^{\mathrm{barrier}}$	1	5	8	3		–
$k_{\mathrm{TST}}^{A\to B}$ [s^{-1}]	1.2×10^{4}	1.7×10^{9}	6.2×10^{8}	1.2×10^{11}		9.4
$k_{\mathrm{TST}}^{B\to A}$ [s^{-1}]	5.3×10^{9}	1.7×10^{9}	8.4×10^{7}	2×10^{10}		0

[a]Calculated as $(k_B T/h)^{-1}$

[b] $F_{AB}^{\mathrm{barrier}}$ is the free energy of the transition-state region between basins A and B relative to basin A. For example, considering the adjacent states A and B as metastable states 3 and 4 (separated by TS 2), $\Delta F_{AB}^{\mathrm{barrier}} = F(\mathrm{TS}\ 2) - F(A)$ and $\Delta F_{BA}^{\mathrm{barrier}} = F(\mathrm{TS}\ 2) - F(B)$.

Supplementary Information

Appendix A: Protonation States

We choose protonation states of the titratable side chain groups in the enzyme based on individual pKa values consistent with a solution pH of 7.0 as reported in Table A1. In the open crystal complex the three conserved Asp groups are well separated from each other and not closely interacting with the d**C**TP, and therefore this choice of the protonation state based on pKa of the amino acid group and an overall pH of 7.0 is reasonable.

Table A1: Protonation states of amino acids in pol β

Residue	Charge	pKa
Asp	-1	3.9
Glu	-1	4.3
His	0	6.5
Lys	$+1$	10.8
Arg	$+1$	12.5

Still, a body of recent simulation data suggests that the protonation states are unclear. In Ref. [1], the authors show on the basis of a truncated model of the active site in *ab-initio* calculations, density functional theory (DFT) functionals, and specific basis-set used that the geometry could only be optimized if the assumption that Asp192 was protonated was made. A report by a different group [2] on the same system, truncated pol β active site claims that geometries can be optimized using high-level DFT without assuming that Asp192 is protonated. These contrasting observations may reflect artifacts of truncating the active site and ignoring the rest of the protein/DNA/solvent environment.

Note also that the protonation state may change as the conformational change occurs. In classical simulations, it is not possible to allow this change in a physically consistent manner, and that is part of the inherent limitations of classical force fields. These are discussed under quantum mechanics/ molecular mechanics (QM/MM) simulations described in Appendix C.

Appendix B: TPS Supplementary Figures

Depicted in Fig. B1 are results for for the mismatched **G:A** system (the results for **G:C** system are provided in [3]): trajectories in each transition state region are harvested using the shooting algorithm [4,5] to connect two metastable states via a Monte Carlo protocol in trajectory space.

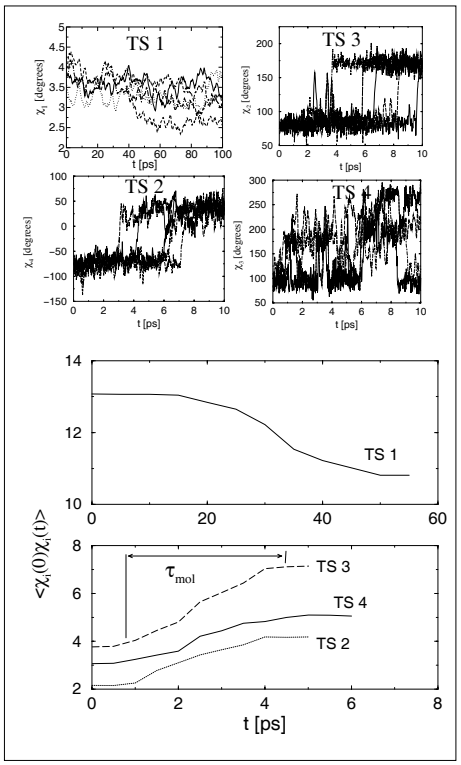

Figure B1: For four transition state (TS) regions, 5 sample trajectories out of 150–200 harvested in path sampling are shown. Order parameter autocorrelation functions $\langle \chi_i(0)\chi_i(t)\rangle$ (in units of Å^2 for TS 1, and rad^2 for TS 2–4), where $\langle\cdot\rangle$ denotes the average over the ensemble of generated trajectories. Autocorrelation functions are plotted with initial point $\langle \chi_i(0)\chi_i(0)\rangle \approx \langle \chi_A\rangle^2$ and end point $\langle \chi_i(0)\chi_i(\tau)\rangle \approx \langle \chi_A\rangle\langle \chi_B\rangle$, to indicate crossing the barrier region between A and B over time τ; χ_4 was shifted by 180° before computing $\langle \chi_4(0)\chi_4(t)\rangle$ to include in the same plot.

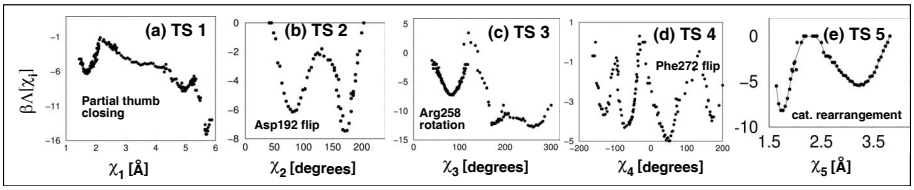

Figure B2: Potential of mean force along the reaction coordinate for different transition state regions for the **G:C** system. (*a*) Partial thumb closing (TS 1). (*b*) Asp-192 flip (TS 2). (*c*) Arg-258 rotation (TS 3). (*d*) Phe-272 flip (TS 4). (*e*) Rearrangement of catalytic region and the stabilization of Arg-258 in the fully rotated state (TS 5); the reaction coordinate χ_5 is the distance between the nucleotide binding Mg^{2+} ion and the oxygen atom $O1_\alpha$ of d**C**TP.

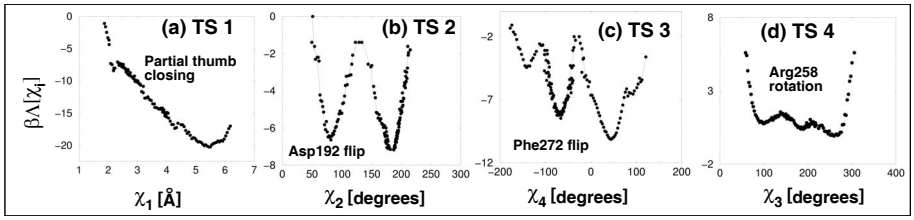

Figure B3: Potential of mean force along the reaction coordinate for different transition state regions for the **G:A** system. (*a*) Partial thumb closing (TS 1). (*b*) Asp-192 flip (TS 2). (*c*) Phe-272 flip (TS 3). (*d*) Arg-258 rotation (TS 4).

Appendix C: Mixed QM/MM calculations

In the QM/MM formalism, the effective Hamiltonian ($\hat{H}_{\text{eff}}$) is the sum of terms representing the QM (quantum mechanical) region, the MM (molecular mechanical) region, and the interaction between them, i.e., $\hat{H}_{\text{eff}} = \hat{H}_{\text{QM/QM}} + \hat{H}_{\text{MM/MM}} + \hat{H}_{\text{QM/MM}}$. Given $\hat{H}_{\text{eff}}$, the energy of the system has the form

$$E = \Phi^t \hat{H}_{\text{eff}} \Phi = \Phi^t \hat{H}_{\text{QM/QM}} \Phi + \Phi^t \hat{H}_{\text{QM/MM}} \Phi + E_{\text{MM/MM}} , \qquad \text{(C-1)}$$

where the vector Φ is the wave function describing the QM atoms, and Φ^t is its transpose. The inner product $\Phi^t \hat{H}_{\text{eff}} \Phi$ represents the expectation value of the effective Hamiltonian in the quantum state characterized by Φ, which upon functional minimization yields the ground state of the system. The Hamiltonian $\hat{H}_{\text{QM/QM}}$ describes the nuclei and electrons (within the Born-Oppenheimer approximation) in the QM region. We use the 6-311G basis set to describe the wave function Φ, and the density functional theory (DFT) formalism with the B3LYP functional incorporated in GAMESS-UK to compute the terms $\Phi^t \hat{H}_{\text{QM/QM}} \Phi$ and $\Phi^t \hat{H}_{\text{QM/MM}} \Phi$. The molecular mechanics Hamiltonian $H_{\text{MM/MM}}$ depends solely on the positions of the classical atoms; we will use CHARMM27 for consistency with our prior work. The boundary Hamiltonian $H_{\text{QM/MM}}$ describes the interactions between the atoms (nuclei and electrons) in the QM region with those in MM region and has the form as described in ref. [6], essentially consisting of Coulombic terms that are treated using a self-consistent field procedure in DFT [7], nonbonded van der Waals terms, and certain bonded terms. Since the boundary between the MM and QM regions cuts through covalent bonds, we employ the "single link atom" procedure [6,8] to satisfy valences of broken bonds in the QM region [6].

Minimizations are performed using the QM/MM Hamiltonian to allow for the relaxation of the active site geometry under the more realistic forcefield in which the active-site atoms are treated ab-initio. Subtle differences in the key distances between the geometries resulting from CHARMM27 and the QM/MM are observed, depicted in Fig. C1. The evolution of crucial active-site distances (O3′–P_α, Cat. Mg^{2+}–O3′, and Cat. Mg^{2+}–$O1_\alpha$ distances) in the left plot of Fig. C1 through the geometry optimization procedure indicates that these key distances differ from those predicted based on CHARMM 27 force-field alone.

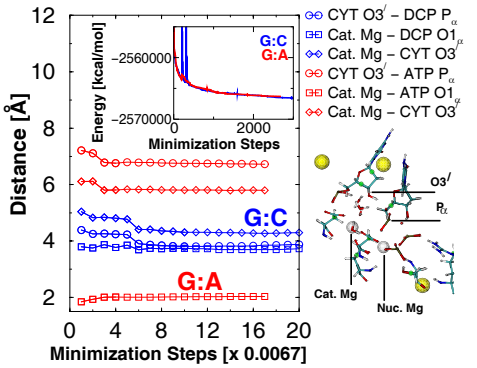

Fig. C1

Figure C1: Evolution of key active-site distances (left) and energies (left inset) in the geometry optimization procedure of solvated pol β system for **G:C** and **G:A** base pairs using a QM/MM Hamiltonian. Snapshot (bottom right) depicts active site region; green spheres are link atoms, yellow spheres are Na^+ ions, and white spheres are Mg^{2+} ions. The protonation states are as described in Table A1

In addition to the two QM/MM models in Fig. C1, we performed geometry optimizations for a model of solvated pol β system with a different protonation state (Asp256 was protonated in this case). The proton was restrained by an external harmonic potential with a force constant of 2000 kcal/mol/Å^2 at a distance of 1 Å from the Oδ_2 oxygen of Asp256. This model was considered because, according to Warshel et al. [9], the first step in the nucleotide incorporation reaction in a T7 polymerase is protonation of one the acidic residues holding the catalytic Mg^{2+} ion, as a result of which the O3′ hydroxyl group of the terminal DNA primer is converted to an oxyanion. It was also evident from the relative energies that the protonated state of Asp256 was significantly less stable than the unprotonated state suggesting that the protonation

does not happen during the closing conformational change prior to chemistry. The barrier to deprotonate Asp256 is close to zero as the depotonation occurred during a geometry optimization with the restraint on the proton removed, and the proton was transferred back to the O3′ oxyanion of the terminal base of the primer DNA. This further validates the protonation states assumed in Table A1.

Appendix D: Calculating Reaction Rates

We further demonstrate the significance of the cascade of subtle events orchestrating the active-site assembly for the correct and mismatch systems prior to the chemical incorporation and subsequent catalysis by solving a network model (inset in Figure 4) to produce the overall rates for the combined process. Fig. 4

Here we outline the procedure to estimate the rates (based on transition state theory [10]) associated with the transitions between adjacent metastable states in our overall free energy profile, as well as the overall rate for the closing transition.

The free energies of the different metastable states and transition-state regions relative to the open and closed states are obtained from the potential of mean force calculations (see Figs. B2 and B3). Using transition state theory [10], the rate of the transition between adjoining metastable states in is given by

$$k_{\mathrm{TST}}^{A\to B} = \frac{1}{\tau_{\mathrm{mol}}} \exp(-\beta \Delta F_{AB}^{\mathrm{barrier}}), \tag{D-1}$$

where τ_{mol} is the time to cross the transition-state region and commit to basin B, and $\Delta F_{AB}^{\mathrm{barrier}}$ is the free energy of the transition-state region between basins A and B relative to basin A. For example, considering the adjacent states A and B as metastable states 3 and 4 (separated by TS 2), $\Delta F_{AB}^{\mathrm{barrier}} = F(\mathrm{TS}\ 2) - F(A)$ and $\Delta F_{BA}^{\mathrm{barrier}} = F(\mathrm{TS}\ 2) - F(B)$; Eq. D-1 is then used to compute $k^{A\to B}$ and $k^{B\to A}$ associated with TS 2.

In the ideal gas approximation, the pre-factor $1/\tau_{\mathrm{mol}} = k_B T/h$, where h is the Plank's constant. In the reactive flux formalism [11], an estimate for τ_{mol} is given by $w/\langle|\dot{q}|\rangle^*$, where w is the characteristic width to be crossed along the reaction coordinate q, and $\langle|\dot{q}|\rangle^*$ is the rate of change of the reaction coordinate at the transition state surface.

Where available, we use the characteristic time for the relaxation of the order parameter autocorrelation function (see Fig. B1) as an estimate for $\tau_{\rm mol}$ [3]. We approximate $\tau_{\rm mol} \approx k_B T/h$ if the estimate from the autocorrelation function is not available.

The rates of transitions between the adjacent metastable states are calculated using Eq. D-1[‡].

Using the individual rates of transitions between adjoining metastable basins, the overall rate can be determined by modeling the overall process as a network of reactions. For the matched (**G:C**) system, the overall process can be represented by:

$$\underset{\text{(Open)}}{\text{MS 1}} \underset{k_{31}}{\overset{k_{13}}{\rightleftarrows}} \text{MS 3} \underset{k_{43}}{\overset{k_{34}}{\rightleftarrows}} \text{MS 4} \underset{k_{54}}{\overset{k_{45}}{\rightleftarrows}} \text{MS 5} \underset{k_{65}}{\overset{k_{56}}{\rightleftarrows}} \text{MS 6} \underset{k_{76}}{\overset{k_{67}}{\rightleftarrows}} \underset{\text{(Closed)}}{\text{MS 7}} \underset{k_{87}}{\overset{k_{78}}{\rightleftarrows}} \underset{\text{(Product)}}{\text{MS 8}}, \tag{D-2}$$

where MS 1–7 correspond to the different metastable states (see Fig. 1 of main text), and MS 8 corresponds to the product state after the chemical reaction (Enzyme/$\text{DNA}_{\text{n+1}}$+PP_i). The overall process for the mismatch (**G:A**) system is represented by:

$$\underset{\text{(Open)}}{\text{MS 1}} \underset{k_{31}}{\overset{k_{13}}{\rightleftarrows}} \text{MS 3} \underset{k_{43}}{\overset{k_{34}}{\rightleftarrows}} \text{MS 4} \underset{k_{54}}{\overset{k_{45}}{\rightleftarrows}} \text{MS 5} \underset{k_{75}}{\overset{k_{57}}{\rightleftarrows}} \underset{\text{(Closed)}}{\text{MS 7}} \underset{k_{87}}{\overset{k_{78}}{\rightleftarrows}} \underset{\text{(Product)}}{\text{MS 8}}, \tag{D-3}$$

where MS 1–5 and 7 correspond to the different metastable states (see Fig. 1 of main text), and MS 8 corresponds to the product state (Enzyme/$\text{DNA}_{\text{n+1}}$+PP_i). The individual rate constants derived from the free energy calculations are provided in Table I, main text. The rate-constants for the final step of product formation were derived from the experimentally measured $k_{\rm pol}$ values in the range 3–90 s^{-1} [12–23]. The network of reactions is solved with the stochastic algorithm of Gillespie [24].

Gillespie algorithm. Following Gillespie [24], we consider a system composed of N chemical species $S_i\,(i = 1, \cdots, N)$ undergoing M possible chemical reactions $R_\mu\,(\mu = 1, \cdots, M)$

[‡]A correction to the transition state theory approximation may be obtained by computing the transmission coefficient using the Bennett-Chandler method [11].

in a given volume V. Every reaction μ is characterized by its stochastic rate constant c_μ, such that $c_\mu dt$ gives the average probability that a particular combination of R_μ reactant molecules will react accordingly in the next infinitesimal time interval dt. (For the set of transitions we consider, there is a one-to-one correspondence between c_μs and k_{ij}s). Given that at time t, the system is in state $(X_1, \cdots, X_N)$, where X_i denotes the number of molecules of species i, the probability that an R_μ reaction will occur within the given volume V in the interval $(t, t + \Delta t)$ is $a_\mu dt = h_\mu c_\mu dt$, where h_μ is the number of distinct combinations for the reaction R_μ to occur. For a reaction $R_1 : S_1 + S_2 \rightarrow S_3$, $h_1 = X_1 X_2$. We consider reactions of type $R_1 : S_1 \rightarrow S_2$, for which $h_1 = X_1$. (For other types of reactions, e.g., $R_2 : 2S_1 \rightarrow S_3$, $h_2 = X_1(X_1 - 1)/2$, and $R_3 : S_1 + S_2 \rightarrow S_3$, $h_3 = X_1 X_2$).

With the above definitions, the reaction probability density function $P(\tau, \mu)$ is given by [24]:

$$P(\tau, \mu) = \begin{cases} a_\mu \exp(-a_0 \tau) & \text{if } 0 \leq \tau \leq \infty \\ 0 & \text{otherwise.} \end{cases} \tag{D-4}$$

Here $a_0 = \sum_i a_i$ and $P(\tau, \mu) d\tau$ is the probability that, given the state $(X_l, \cdots, X_N)$ at time t, the next reaction in V will occur in the infinitesimal time interval $(t + \tau, t + \tau + d\tau)$, and will be an R_μ reaction.

In order to generate τ and μ according to the distribution specified in Eq. (D-4), we generate two random numbers r_1 and r_2 between 0 and 1 (end points excluded) from a unit-interval uniform distribution $(0, 1)$ and set

$$\tau = (1/a_0) \ln(1/r_1), \tag{D-5}$$

and μ to be that integer for which

$$\sum_{\nu=1}^{\mu-1} a_\nu \; < \; r_2 a_0 \; \leq \sum_{\nu=1}^{\mu} a_\nu. \tag{D-6}$$

The Gillespie algorithm for simulating the stochastic time evolution of a chemically reacting system is therefore:

Step 0 (Initialization). Input the desired values for the M reaction constants $c_l, \cdots, c_M$ and the N initial molecular population numbers $X_1, \cdots, X_N$. Set the time variable t and the reaction counter n both to zero. Initialize the unit-interval uniform random number generator (URN).

Step 1. Calculate and store the M quantities $a_1 = h_l c_l, \cdots, a_M = h_M c_M$ for the current molecular population numbers, and also a_0.

Step 2. Generate two random numbers r_1 and r_2 using the unit-interval uniform random number generator, and calculate τ and μ

Step 3. Using the τ and μ values obtained in step 2, increase t by τ, and adjust the molecular population levels $X_1, \cdots, X_N$ to reflect the occurrence of one R_μ reaction. Then increase the reaction counter n by 1 and return to step 1.

This procedure is implemented in the STOCKS simulation software [25], which we have applied to determine the temporal evolution of number of reactant, product, and intermediate species. One hundred different evolution trajectories are harvested to account for the stochasticity inherent in the system; this produces the bands in Fig. 4.

The temporal evolution of number of reactant, product, and intermediate species (Figure 4) are solved with the stochastic algorithm of Gillespie [24] using the STOCKS simulation software [25]. The spread in the kinetics (thickness of bands shown) represents the inherent stochasticity of the system, which is based on a copy number of 100 binary (open) complexes

(MS 1), and 100 different evolution trajectories.

We note a striking difference in the reaction evolution between the **G:C** and **G:A** systems from Fig. 4. The curve corresponding to the open enzyme state for the matched system (blue) rapidly disappears, with the closed state (red band, MS 7) quickly emerging and transitioning into product (black band, MS 8) — where d**C**TP has been incorporated into the primer strand opposite the template guanine residue. At around 0.1 s, which corresponds to k_{pol} of 10 s^{-1} [12–23], the product curve sharply rises, until all species are product at around 1 s.

Literature cited

1. Abashkin, Y. G, Erickson, J. W, & Burt, S. K. (2001) *J. Phys. Chem. B* **105**, 287–292.
2. Rittenhouse, R. C, Apostoluk, W. K, Miller, J. H, & Straatsma, T. P. (2003) *Phys. Rev. Lett.* **53**, 667–682.
3. Radhakrishnan, R & Schlick, T. (2004) *Proc. Nat. Acad. Sci.* **101**, 5970–5975.
4. Bolhuis, P. G, Dellago, C, & Chandler, D. (1998) *Faraday Discuss.* **110**, 421–436.
5. Dellago, C, Bolhuis, P. G, & Geissler, P. L. (2002) *Adv. Chem. Phys* **123**, 1–81.
6. Reuter, N, Dejaegere, A, Maigret, B, & Karplus, M. (2000) *J. Phys. Chem. A* **104**, 1720–1735.
7. Szabo, A & Ostlund, N. S. (1996) *Modern Quantum Chemistry.* (Dover Publications, Mineola, New York).
8. Das, D, Eurenius, K. P, Billings, E. M, Sherwood, P, Chatfield, D. C, Hodoscek, M, & Brooks, B. R. (2002) *J. Chem. Phys.* **117**, 10534–10547.
9. Florián, J, Goodman, M. F, & Warshel, A. (2003) *J. Am. Chem. Soc.* **125**, 8163–8177.
10. Frost, A. A & Pearson, R. G. (1961) *Kinetics and mechanism.* (John Wiley and Sons, New York, NY).
11. Chandler, D. (1978) *J. Chem. Phys.* **68**, 2959–2970.
12. Shah, A. M, Li, S.-X, Anderson, K. S, & Sweasy, J. B. (2001) *J. Biol. Chem.* **276**, 10824–10831.
13. Vande Berg, B. J, Beard, W. A, & Wilson, S. H. (2001) *J. Biol. Chem.* **276**, 3408–3416.
14. Ahn, J, Kraynov, V. S, Zhong, X, Werneburg, B. G, & Tsai, M.-D. (1998) *Biochem. J.* **331**, 79–87.
15. Suo, Z & Johnson, K. A. (1998) *J. Biol. Chem.* **273**, 27250–27258.
16. Kraynov, V. S, Werneburg, B. G, Zhong, X, Lee, H, Ahn, J, & Tsai, M.-D. (1997) *Biochem. J.* **323**, 103–111.
17. Zhong, X, Patel, S. S, Werneburg, B. G, & Tsai, M.-D. (1997) *Biochemistry* **36**, 11891–11900.
18. Dahlberg, M. E & Benkovic, S. J. (1991) *Biochemistry* **30**, 4835–4843.
19. Kuchta, R. D, Mizrahi, V, Benkovic, P. A, Johnson, K. A, & Benkovic, S. J. (1987) *Biochemistry* **26**, 8410–8417.
20. Wong, I, Patel, S. S, & Johnson, K. A. (1991) *Biochemistry* **30**, 526–537.
21. Patel, S. S, Wong, I, & Johnson, K. A. (1991) *Biochemistry* **30**, 511–525.
22. Frey, M. W, Sowers, L. C, Millar, D. P, & Benkovic, S. J. (1995) *Biochemistry* **34**, 9185–9192.
23. Capson, T. L, Peliska, J. A, Kaboord, B. F, Frey, M. W, Lively, C, Dahlberg, M, & Benkovic, S. J. (1992) *Biochemistry* **31**, 10984–10994.
24. Gillespie, D. T. (1977) *J. Phys. Chem.* **81**, 2340–2361.
25. Kierzek, A. M. (2002) *Bioinformatics* **18**, 470–481.