# Supplementary Information

*Fidelity Discrimination in DNA polymerase $\beta$: differing closing profiles for a mismatched (G:A) versus matched (G:C) base pair by Ravi Radhakrishnan and Tamar Schlick.*

**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 S1. 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.

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 $\beta$ 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.

Table S1
Protonation States of titratable groups in pol $\beta$

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

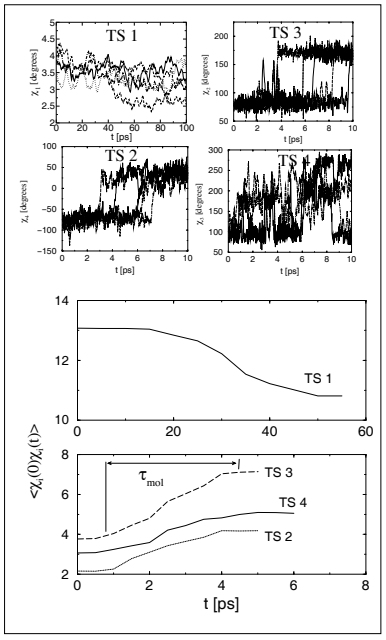


Figure S1: For four transition state (TS) regions, 5 sample trajectories out of 150–200 harvested in path sampling are shown. Order parameter autocorrelation functions $\langle \chi_i(0)\chi_i(t)\rangle$ (in units of Å$^2$ for TS 1, and rad$^2$ for TS 2–4), where $\langle\cdot\rangle$ denotes the average 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 $\tau$; $\chi_4$ was shifted by 180° before computing $\langle \chi_4(0)\chi_4(t)\rangle$ to include in the same plot.

## Appendix B: TPS Supplementary Figures

Depicted in Fig. S1 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. Figures S2 and S3 depict the free energy profiles obtained by calculating the potential of mean force using the BOLAS protocol [6].

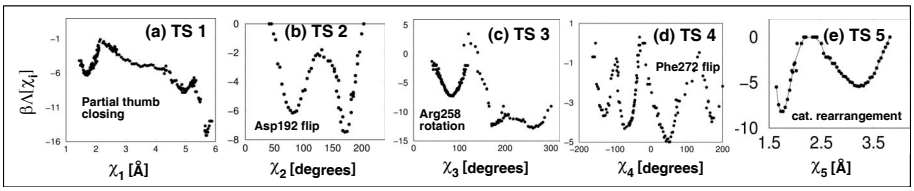


Figure S2: Potential of mean force along the reaction coordinate for different transition state regions for the **G:C** system. ($a$) Partial thumb closing (TS 1). ($b$) Asp-192 flip (TS 2). ($c$) Arg-258 rotation (TS 3). ($d$) Phe-272 flip (TS 4). ($e$) Rearrangement of catalytic region and the stabilization of Arg-258 in the fully rotated state (TS 5); the reaction coordinate $\chi_5$ is the distance between the nucleotide binding $Mg^{2+}$ ion and the oxygen atom $O1_\alpha$ of d**C**TP.

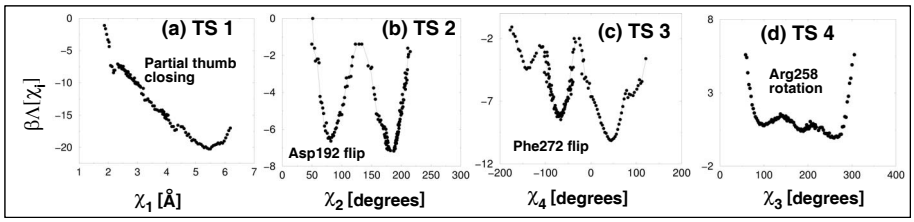


Figure S3: Potential of mean force along the reaction coordinate for different transition state regions for the **G:A** system. ($a$) Partial thumb closing (TS 1). ($b$) Asp-192 flip (TS 2). ($c$) Phe-272 flip (TS 3). ($d$) Arg-258 rotation (TS 4).

## Appendix C: Mixed QM/MM calculations

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

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

where the vector $\Phi$ is the wave function describing the QM atoms, and $\Phi^t$ is its transpose. The inner product $\Phi^t \hat{H}_{\rm eff}\, \Phi$ represents the expectation value of the effective Hamiltonian in the quantum state characterized by $\Phi$, which upon functional minimization yields the ground state of the system. The Hamiltonian $\hat{H}_{\rm QM/QM}$ 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 $\Phi$, and the density functional theory (DFT) formalism with the B3LYP functional incorporated in GAMESS-UK to compute the terms $\Phi^t \hat{H}_{\rm QM/QM}\, \Phi$ and $\Phi^t \hat{H}_{\rm QM/MM}\, \Phi$. The molecular mechanics Hamiltonian $H_{\rm MM/MM}$ depends solely on the positions of the classical atoms; we will use CHARMM27 for consistency with our prior work. The boundary Hamiltonian $H_{\rm QM/MM}$ 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. [7], essentially consisting of Coulombic terms that are treated using a self-consistent field procedure in DFT [8], 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 [7,9] to satisfy valences of broken bonds in the QM region [7].

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. S4. The evolution of crucial active-site distances (O3′–$P_\alpha$, Cat. $Mg^{2+}$–O3′, and Cat. $Mg^{2+}$–$O1_\alpha$ distances) in the left plot of Fig. S4 through the geometry optimization procedure indicates that these key distances differ from those predicted based on CHARMM 27 force-field alone.

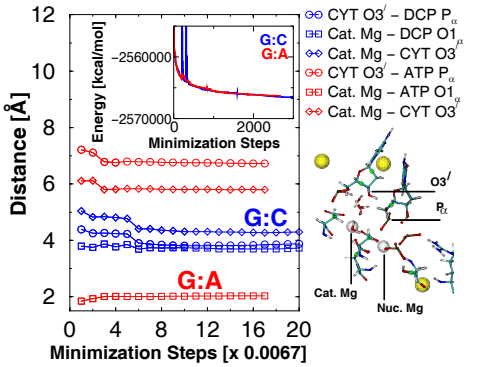


Figure S4: Evolution of key active-site distances (left) and energies (left inset) in the geometry optimization procedure of solvated pol $\beta$ system for **G:C** and **G:A** base pairs using a QM/MM Hamiltonian. Snapshot (bottom right) depicts active site region; green spheres are link atoms, yellow spheres are $Na^+$ ions, and white spheres are $Mg^{2+}$ ions. The protonation states are as described in Table S1

In addition to the two QM/MM models in Fig. S4, we performed geometry optimizations for a model of solvated pol $\beta$ 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 O$\delta_2$ oxygen of Asp256. This model was considered because, according to Warshel et al. [10], 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 S1.

## 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.

Here we outline the procedure to estimate the rates (based on transition state theory [11]) 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. S2 and S3). Using transition state theory [11], the rate of the transition between

adjoining metastable states in is given by

$$k_{\rm TST}^{A\to B} \;=\; \frac{1}{\tau_{\rm mol}}\,\exp(-\beta\Delta F_{AB}^{\rm barrier}), \tag{S-2}$$

where $\tau_{\rm mol}$ is the time to cross the transition-state region and commit to basin $B$, and $\Delta F_{AB}^{\rm barrier}$ 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), $\Delta F_{AB}^{\rm barrier} = F(\text{TS 2}) - F(A)$ and $\Delta F_{BA}^{\rm barrier} = F(\text{TS 2}) - F(B)$; Eq. S-2 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 $1/\tau_{\rm mol} = k_B T/h$, where $h$ is the Plank's constant. In the reactive flux formalism [12], an estimate for $\tau_{\rm mol}$ 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. S1) 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. S-2*.

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:

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

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/$\text{DNA}_{\text{n+1}}$+$\text{PP}_i$). The overall process for the mismatch (**G:A**) system is represented by:

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

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/$\text{DNA}_{\text{n+1}}$+$\text{PP}_i$). 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 $\text{s}^{-1}$ [13–24]. The network of reactions is solved with the stochastic algorithm of Gillespie [25].

**Gillespie algorithm.** Following Gillespie [25], we consider a system composed of $N$ chemical species $S_i\,(i = 1, \cdots, N)$ undergoing $M$ possible chemical reactions $R_\mu\,(\mu = 1, \cdots, M)$ in a given volume $V$. Every reaction $\mu$ is characterized by its stochastic rate constant $c_\mu$, such that $c_\mu dt$ gives the average probability that a particular combination of $R_\mu$ 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 $c_\mu$s and $k_{ij}$s). Given

*A correction to the transition state theory approximation may be obtained by computing the transmission coefficient using the Bennett-Chandler method [12].

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 $R_\mu$ 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 $h_\mu$ is the number of distinct combinations for the reaction $R_\mu$ 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 [25]:

$$P(\tau, \mu) \;=\; \begin{cases} a_\mu \exp(-a_0 \tau) & \text{if } 0 \leq \tau \leq \infty \\ 0 & \text{otherwise.} \end{cases} \tag{S-5}$$

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 $R_\mu$ reaction.

In order to generate $\tau$ and $\mu$ according to the distribution specified in Eq. (S-5), 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

$$\tau = (1/a_0) \ln(1/r_1), \tag{S-6}$$

and $\mu$ to be that integer for which

$$\sum_{\nu=1}^{\mu-1} a_\nu \;<\; r_2 a_0 \;\leq \sum_{\nu=1}^{\mu} a_\nu. \tag{S-7}$$

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 $\tau$ and $\mu$

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

This procedure is implemented in the STOCKS simulation software [26], 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 [25] using the STOCKS simulation software [26]. 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 $k_{\mathrm{pol}}$ of 10 $\mathrm{s}^{-1}$ [13–24], 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. *J. Phys. Chem. B*, **2001**, *105*, 287–292.
2. Rittenhouse, R. C.; Apostoluk, W. K.; Miller, J. H.; Straatsma, T. P. *Phys. Rev. Lett.*, **2003**, *53*, 667–682.
3. Radhakrishnan, R.; Schlick, T. *Proc. Nat. Acad. Sci.*, **2004**, *101*, 5970–5975.
4. Bolhuis, P. G.; Dellago, C.; Chandler, D. *Faraday Discuss.*, **1998**, *110*, 421–436.
5. Dellago, C.; Bolhuis, P. G.; Geissler, P. L. *Adv. Chem. Phys*, **2002**, *123*, 1–81.
6. Radhakrishnan, R.; Schlick, T. *J. Chem. Phys.*, **2004**, *121*, 2436–2444.
7. Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A*, **2000**, *104*, 1720–1735.
8. Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry*. Dover Publications, Mineola, New York, 1996.
9. Das, D.; Eurenius, K. P.; Billings, E. M.; Sherwood, P.; Chatfield, D. C.; Hodoscek, M.; Brooks, B. R. *J. Chem. Phys.*, **2002**, *117*, 10534–10547.
10. Florián, J.; Goodman, M. F.; Warshel, A. *J. Am. Chem. Soc.*, **2003**, *125*, 8163–8177.
11. Frost, A. A.; Pearson, R. G. *Kinetics and mechanism*. John Wiley and Sons, New York, NY, 1961.
12. Chandler, D. *J. Chem. Phys.*, **1978**, *68*, 2959–2970.
13. Shah, A. M.; Li, S.-X.; Anderson, K. S.; Sweasy, J. B. *J. Biol. Chem.*, **2001**, *276*, 10824–10831.
14. Vande Berg, B. J.; Beard, W. A.; Wilson, S. H. *J. Biol. Chem.*, **2001**, *276*, 3408–3416.
15. Ahn, J.; Kraynov, V. S.; Zhong, X.; Werneburg, B. G.; Tsai, M.-D. *Biochem. J.*, **1998**, *331*, 79–87.
16. Suo, Z.; Johnson, K. A. *J. Biol. Chem.*, **1998**, *273*, 27250–27258.
17. Kraynov, V. S.; Werneburg, B. G.; Zhong, X.; Lee, H.; Ahn, J.; Tsai, M.-D. *Biochem. J.*, **1997**, *323*, 103–111.
18. Zhong, X.; Patel, S. S.; Werneburg, B. G.; Tsai, M.-D. *Biochemistry*, **1997**, *36*, 11891–11900.
19. Dahlberg, M. E.; Benkovic, S. J. *Biochemistry*, **1991**, *30*, 4835–4843.
20. Kuchta, R. D.; Mizrahi, V.; Benkovic, P. A.; Johnson, K. A.; Benkovic, S. J. *Biochemistry*, **1987**, *26*, 8410–8417.
21. Wong, I.; Patel, S. S.; Johnson, K. A. *Biochemistry*, **1991**, *30*, 526–537.
22. Patel, S. S.; Wong, I.; Johnson, K. A. *Biochemistry*, **1991**, *30*, 511–525.
23. Frey, M. W.; Sowers, L. C.; Millar, D. P.; Benkovic, S. J. *Biochemistry*, **1995**, *34*, 9185–9192.
24. Capson, T. L.; Peliska, J. A.; Kaboord, B. F.; Frey, M. W.; Lively, C.; Dahlberg, M.; Benkovic, S. J. *Biochemistry*, **1992**, *31*, 10984–10994.
25. Gillespie, D. T. *J. Phys. Chem.*, **1977**, *81*, 2340–2361.
26. Kierzek, A. M. *Bioinformatics*, **2002**, *18*, 470–481.