

Supporting Information

An Abridged Transition State Model to Derive Structure, Dynamics and Energy Components of DNA Polymerase β Fidelity

Martin Klvaňa¹, Petr Jeřábek¹, Myron F. Goodman² and Jan Florián^{1*}

¹Department of Chemistry, Loyola University Chicago, Chicago, IL 60626

²Departments of Biological Sciences and Chemistry, University of Southern California, Los Angeles, CA 90089-2910

Table 1S: Atom types and partial charges for the triphosphate moiety in GSA and PTS states. Comparison with charges on pentavalent phosphoranes and corresponding ground state compounds used in other studies.

Atom	GSA ^b		PTS		Pentavalent Intermediate ^c	2' Deprotonated Ribose Phosphate ^d	Ribose Phosphorane ^d	Dimethyl Phosphate ^e	Trimethyl Phosphorane ^e	TS Model by Warshel et al. ^f
Name ^a	Type	Charge	Type	Charge	Charge	Charge	Charge	Charge	Charge	Charge
O_{1□}	O3	-0.9100	O3	-0.9100	-0.9100	-	-	-	-	-0.915
O_{2□}	O3	-0.9100	O3	-0.9100	-0.9100	-	-	-	-	-0.915
O_{3□}	O3	-0.9100	O3	-0.9100	-0.9100	-	-	-	-	-0.915
P_□	P	1.2000	P	1.2000	1.2000	-	-	-	-	1.20
O_{1□}	O2	-0.8700	O2	-0.8900	-0.8900	-	-	-	-	-0.905
O_{2□}	O2	-0.8700	O2	-0.8900	-0.8900	-	-	-	-	-0.905
O_{3□}	OS	-0.5000	OS	-0.5000	-0.5000	-	-	-	-	-0.50
P_□	P	1.1900	P	1.1900	1.1900	-	-	-	-	1.195
O_{1□}	O2	-0.8703	ON	-0.8803	-0.8803	-0.84	-0.92	-0.83	-0.98	-0.83
O_{2□}	O2	-0.8703	ON	-0.8803	-0.8803	-0.84	-0.92	-0.83	-0.98	-0.83
O_{3□}	OS	-0.5500	OR	-0.6500	-0.6500	-0.65	-0.75	-0.46	-0.73	-0.785
P_□	P	1.1800	P5	1.1400	1.1400	1.64	1.81	1.14	1.33	1.155
O_{5'}	OS	-0.4954	OT	-0.4954	-0.4554	-0.63	-0.62	-0.46	-0.30	-0.50
O_{3'}	OH	-0.8549	OR	-0.6549	-0.6549	-0.91	-0.68	-	-0.63	-0.585
C_{3'}	CT	0.0713	CT	0.0713	0.0713	0.23	0.10	-	0.15	

^a See also Figure 1S. Atoms that constituted the probe region (in our electrostatic linear-response free energy calculations) are shown in bold font. Additionally, the probe region included the Mg ion (Figure 1S).

^b 3'OH group of the 3'-terminal primer nucleotide is fully deprotonated in GSA state. This state may or may not correspond actual ground state of the nucleotidyl transfer reaction (depending on solution pH and pKa of the 3'OH group).

^c Charges of the pentavalent intermediate used in empirical valence bond (EVB) simulations of chemical reaction catalyzed by T7 DNA polymerase (see Table 1S in Florian et al, JACS 125, 8263, 2003).

^d CHARMM force field charges for the ground- and transition-states of the transphosphorylation reaction, Mayaan et al. (2007) J. Comp. Chem. 28: 495-507.

^e *Ab initio* PCM/B3LYP/TZVP charges for the specific base-catalyzed methanolysis of dimethyl phosphate, Borden et al. (2006) J. Phys. Chem. 110: 14988 - 14999.

^f TS model for pol β used by Warshel and coworkers (2006) Biochemistry 45: 7036 - 7048. Note that charges on Pa and O3' were changed to 0.955 and -0.655 a.u. in the subsequent study of Rucker, Oelschlaeger and Warshel (2009) Proteins 78: 671 - 680.

Table 2S: Van der Waals well-depths (ϵ) and atom radii for atoms belonging to the triphosphate moiety^a.

Atom type	Radius [Å]	ϵ [kcal mol ⁻¹]	Mass [g mol ⁻¹]
O2	1.6612	0.2100	16.00
O3	1.7000	0.2100	16.00
ON	1.6612	0.2100	16.00
OR	1.6837	0.1700	16.00
OS	1.6837	0.1700	16.00
OT	1.6837	0.1700	16.00
P	2.1000	0.2000	30.97
P5	2.1000	0.2000	30.97

^a Standard 6-12 potential function was used.

Table 3S: Stretching force constants (k) and equilibrium bond distances (R_0) for chemical bonds between atoms of the triphosphate moiety^a.

Type	k [kcal mol ⁻¹ Å ⁻²]	R_0 [Å]
P – O2	1050	1.480
P – O3	950	1.520
P – OR	800	1.600
P – OS	460	1.610
P5 – ON	900	1.550
P5 – OR	90	1.900
P5 – OT	600	1.650

^a Bond energy (R) = $\frac{1}{2} k(R - R_0)^2$

Table 4S: Bending force constants (k) and equilibrium bond angles (α_0) for chemical bonds between atoms of the triphosphate moiety^a.

Type	k [kcal mol ⁻¹ rad ⁻²]	α_0 [deg]
O3 – P – O3	250	109.90
O3 – P – OS	200	108.50
P – OS – P	200	133.50
OS – P – O2	200	108.23
OS – P – OR	90	102.60
O2 – P – O2	280	119.90
OS – P – OS	90	102.60
O2 – P – OR	200	108.23
P – OR – P5	200	133.50
OR – P5 – OR	30	170.00
OR – P5 – ON	140	90.00
OR – P5 – OT	140	90.00
ON – P5 – ON	200	120.00
ON – P5 – OT	200	120.00

^a Bond energy $(\alpha) = \frac{1}{2} k(\alpha - \alpha_0)^2$

Table 5S: Force constants (K), number of minima per full turn (n) and phase-shift (δ) for torsions around chemical bonds between atoms of the triphosphate moiety^a.

Type	K [kcal mol ⁻¹]	n	δ [deg]
O3 – P – OS – P	0.75	3	0.0
P – OS – P – O2	0.75	3	0.0
P – OS – P – OS	0.75	3	0.0
P – OS – P – OR	0.75	3	0.0
O2 – P – OR – P5	0.75	3	0.0
P – OR – P5 – ON	0.75	3	0.0
P – OR – P5 – OT	0.75	3	0.0

^a Bond energy $(\phi) = K(1 + \cos(n\phi - \delta))$

Table 6S: Control simulations for model systems^a in the PTS state.

Solute	Solvent	P _α -O _{lg} shortening
dRib-MDP	water	yes
MI-MDP	water	yes
MI-MDP	two Mg ²⁺ ; water	yes
MI-MDP ^b	two Mg ²⁺ ; water	no
MI-MDP	two Mg ⁰ ; water	yes
MI-MP	two Mg ²⁺ ; water	yes
MI-MP ⁰	two Mg ²⁺ ; water	no
MI ⁰ -MP ⁰	two Mg ⁰ ; water	no
MI-MP	two Mg ²⁺	yes
MI ⁰ -MP ⁰	two Mg ⁰	no
MI-MP	-	no

^a dRib = 2'-deoxyribose with deprotonated O3'; MI = methanol ion; MI⁰ = methanol ion with zero atomic charges; MDP = methyl diphosphate; MP = methyl phosphate; MP⁰ = methyl phosphate with zero atomic charges; Mg⁰ = magnesium with zero atomic charge;

^b force constant for P_α-O_{lg} distance set to 450 (default is 90);

Table 7S: Free-energy changes accompanying the transition from the GSA to PTS state, evaluated from the MD ensemble averages ($\langle \rangle$) of electrostatic interaction energies (U) between the probe region and the rest of the simulated system.

Substrate geometry	$\langle \Delta U \rangle^{\text{a, b}}$ [kcal mol ⁻¹]	$\Delta G_{\text{GSA} \rightarrow \text{PTS}}^{\text{c}}$ [kcal mol ⁻¹]
<i>Aqueous solution</i> ^d		
PTS	-25.9; -25.7	
GSA	-39.2; -40.0	
		-32.7 ± 0.7
<i>Pol β – right dNTP</i> ^e		
PTS	-27.3	
GSA	-48.7	
		-38.0
<i>Pol β - wrong dNTP</i> ^e		
PTS	-27.1	
GSA	-44.4	
		-35.8

$$^{\text{a}} \langle \Delta U \rangle_{\text{PTS}} = \langle U(\text{PTS}) - U(\text{GSA}) \rangle_{\text{PTS}}$$

$$^{\text{b}} \langle \Delta U \rangle_{\text{GSA}} = \langle U(\text{PTS}) - U(\text{GSA}) \rangle_{\text{GSA}}$$

$$^{\text{c}} \Delta G_{\text{GSA} \rightarrow \text{PTS}} = 0.5 * [\langle \Delta U \rangle_{\text{GSA}} + \langle \Delta U \rangle_{\text{PTS}}]$$

^d Results from two independent 2.5 ns simulations.

^e Results from a single 10 ns trajectory (for each system).

Table 8S: Average RMSD (Å) of the thumb subdomain in MD simulations of GSA and PTS complexes that were evaluated with respect to the crystallographic coordinates of the open pol β - DNA complex (1BPX), and closed ternary complexes (GS) containing the right (2FMP) and wrong (3C2M) dNTP substrates.

Pol β complex	1BPX			2FMP			3C2M		
	GSA ^a	PTS ^{a,b}	GS ^c	GSA ^a	PTS ^{a,b}	GS ^c	GSA ^a	PTS ^{a,b}	GS ^c
Right	5.1	5.0	5.1	0.6	0.7	0	1.3	1.2	1.1
Wrong	4.8	4.5	3.8	0.8	0.8	1.1	1.1	1.0	0

^a The presented averages are based on trajectories shown in Figures 5S and 6S.

^b Average of eight independent 10 ns MD simulations (standard deviation of the average is 0.2 Å).

^c RMSD of the thumb subdomain for the pair of crystal structures.

Table 9S: Percentage of configurations of a water molecule near O_{lg} (O_{wat}-O_{lg} distance < 3.5 Å) that form H-bonds with oxygen atoms of dNTP substrate^a. Configurations with H_{wat}...O distance < 2.0 Å were considered as H-bonded.

Atoms ^b	GSA		PTS	
	dCTP (1293 ps)	dATP (3142 ps)	dCTP (5415 ps)	dATP (3395 ps)
O _{2α}	24.9	41.2	14.0	11.7
O _{3α}	4.7	0	0.7	1.1
O _{1β}	0.1	46.3	0	0.3
O _{1γ}	2.3	0	17.4	16.2
O _{2α} & O _{γ}	0	0	60.9	62.8
O _{3α} & O _{γ}	0	0	4.4	4.5

^a The identity of water molecules near O_{lg} changes during the simulation. The data in this Table reflect behavior of a single water molecule that showed the longest residence time during our 10 ns simulation. This residence time is shown in parenthesis.

^b See Figure 1S for atom numbering. O_{lg} = O_{3 α} . O _{γ} includes H-bonding to O_{1 γ} , O_{2 γ} or O_{3 γ} atoms.

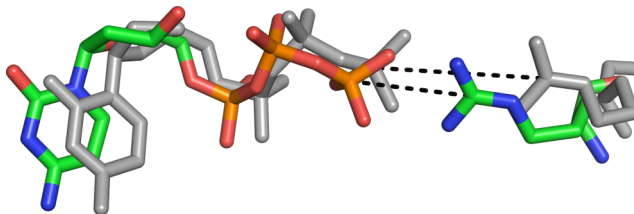
Table 10S: Percentage of configurations of a water molecule near O_{lg} ($O_{wat}-O_{lg}$ distance < 3.5 Å) that form H-bonds with oxygen atoms of dNTP substrate^a. Configurations with $H_{wat...O}$ distance < 2.5 Å were considered as H-bonded.

Atoms ^b	GSA		PTS	
	dCTP (1293 ps)	dATP (3142 ps)	dCTP (5415 ps)	dATP (3395 ps)
$O_{2\alpha}$	64.3	45.8	0.5	1.4
$O_{3\alpha}$	22.4	0.3	0.1	8.5
$O_{1\beta}$	0.2	53.1	0	0.3
$O_{1\gamma}$	0.2	0	0.5	0.6
$O_{1\alpha}$ & O_{γ}	0	0	0.5	0
$O_{2\alpha}$ & $O_{2\alpha}$	0.2	0	0	0
$O_{2\alpha}$ & $O_{3\alpha}$	2.2	0.1	0	0.6
$O_{2\alpha}$ & O_{γ}	0	0	88.4	85.7
$O_{3\alpha}$ & O_{γ}	0	0	9.8	10.3

^a The identity of water molecules near O_{lg} changes during the simulation. The data in this Table reflect behavior of a single water molecule that showed the longest residence time during our 10 ns simulation. This residence time is shown in parenthesis.

^b See Figure 1S for atom numbering. $O_{lg} = O_{3\alpha}$. O_{γ} includes H-bonding to $O_{1\gamma}$, $O_{2\gamma}$ or $O_{3\gamma}$ atoms.

Table 11S: Distance between dNTP substrate and arginine 149 in crystal structures of pol β .



PDB Code	Resolution [Å]	Source	Template:NTP	NTP[PG]...R149[CZ] Distance [Å]
1ZJN	2.61	Human	dC·dGTP	4.65
2BPF	2.90	Rat	dG·ddCTP	4.67
2BPG	3.60	Rat	dG·ddCTP	5.24
3ISD	2.60	Human	THF·dAMPCPP	5.54
1MQ3	2.80	Human	8-oxodG·dCTP	5.56
3LK9	2.50	Human	dA·dTMP(CXY)P(CXY)P [X=C-Y; X=Y=F]	5.61
3JPP	2.10	Human	dC·dGMP(CX)PP [X=Met]	5.92
3JPN	2.15	Human	dC·dGMP(CXY)PP [X=C-Y; X=Y=Cl]	5.98
3JPS	2.00	Human	dC·dGMP(CXY)PP [X=C-Y; X=F; Y=Met]	5.98
2FMP	1.65	Human	dG·ddCTP	5.99 (4.35, 4.33) ^a
3C2M	2.15	Human	dG·dAMPCPP	5.99 (4.40, 4.34) ^a
3JPT	2.15	Human	dC·dGMP(CXY)PP [X=C-Y; X=F; Y=Cl]	5.99
3MBY	2.00	Human	dA·8-oxodGTP	6.04
3JPO	2.00	Human	dC·dGMP(CX)PP [X=Cl]	6.07
2FMS	2.00	Human	dA·dUMPNPP	6.10
1BPY	2.20	Human	dG·ddCTP	6.14
3JPR	2.10	Human	dC·dGMP(CXY)PP [X=C-Y; X=Y=Met]	6.15
3JPQ	1.90	Human	dC·dGMP(CX)PP [X=Br]	6.19

^a Linear response model simulation of GSA and PTS model, respectively (this study).

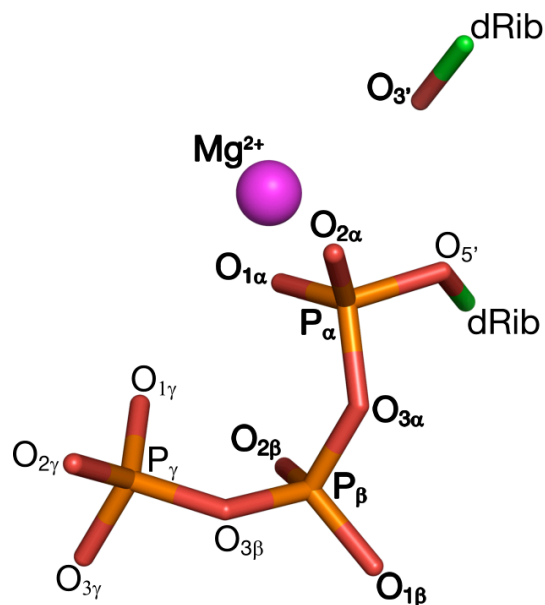


Figure 1S: Atom numbering for the triphosphate group that was used in its force field specifications (Tables 1S-5S). Names of atoms that encompass the probe region of the substrate are printed in bold font.

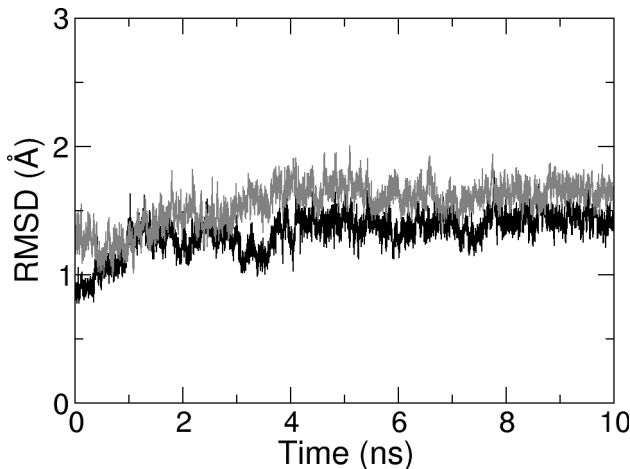


Figure 2S: Time evolution of the RMSD for the activated ground state (GSA) pol β complex containing the dG•dCTP (black) and dG•dATP (grey) nascent base pair. The black and grey graph lines reflect deviations of protein backbone atoms residing inside the simulation sphere (see Methods) from their coordinates in the crystal structures of pol β ternary complexes containing the dG•dCTP (PDB code 2FMP) and dG•dATP (PDB code 3C2M), respectively.

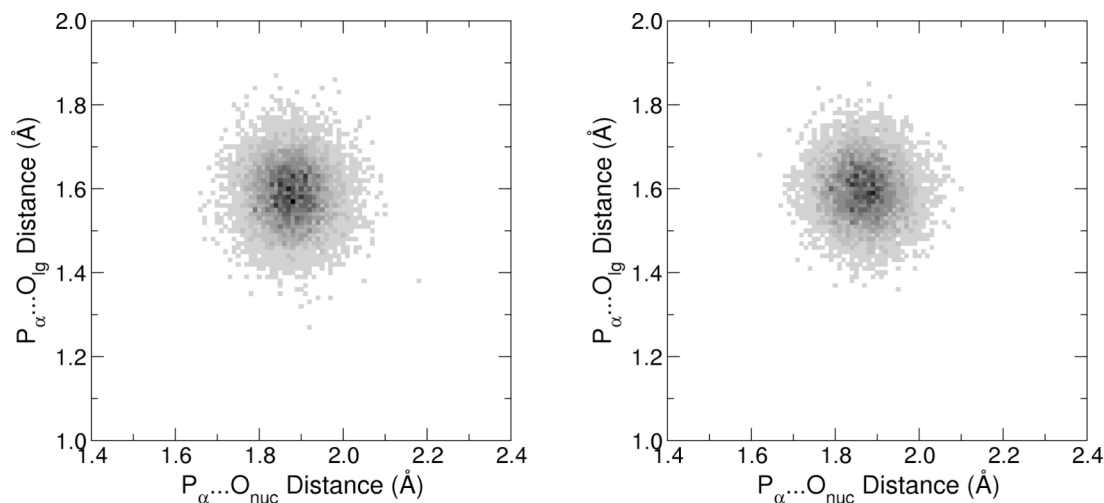


Figure 3S: Distribution of the sampled O_{nuc} - P_{α} and P_{α} - O_{lg} distances in the simulations of PTS complex with the correct (left) and incorrect (right) dNTP substrate.

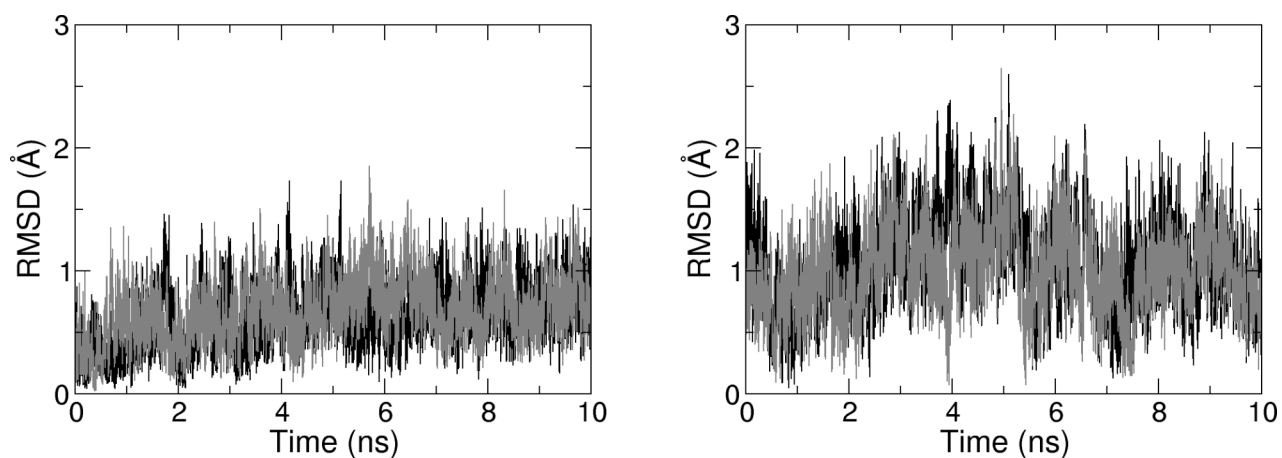


Figure 4S: Time evolution of the RMSD of the protein backbone atoms in the thumb subdomain during MD simulations of the GSA (black) and PTS model (grey) complexes containing the correct (left) and incorrect (right) dNTP. The RMSD displayed in the left and right figures refer to atomic coordinates in the crystal of the pol β ternary complex containing the correct (PDB code 2FMP) and incorrect (PDB code 3C2M) dNTP substrate, respectively.

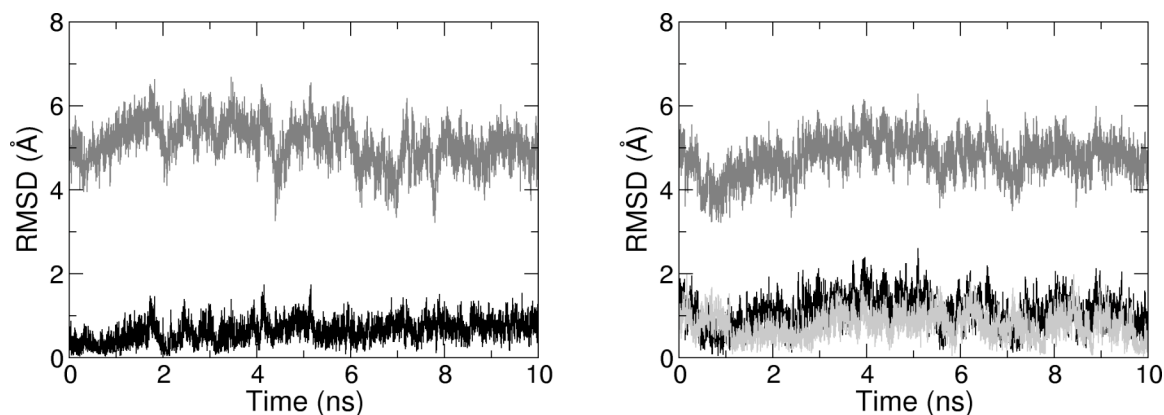


Figure 5S: RMSD of the thumb subdomain determined by MD simulations of the GSA complex containing the dG•dCTP (left) and dG•dATP (right) nascent base pair. The deviations were measured with respect to crystallographic coordinates of the binary pol β -DNA complex (PDB code 1BPX, dark grey), dG•dCTP (PDB code 2FMP, black line in the left figure and light grey line in the right figure) or dG•dATP (3C2M, black line in the right figure) ternary complexes. RMSD values obtained by averaging over the whole MD trajectory are presented in Table 5S.

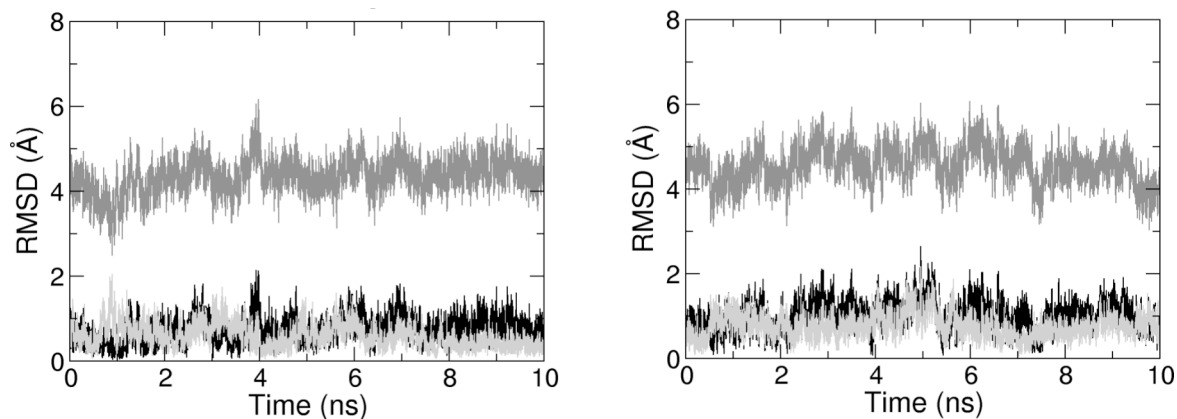


Figure 6S: Time evolution of RMSD of the protein backbone atoms in the thumb subdomain during MD simulations of the PTS complex containing the dG•dATP nascent base pair. The deviations are plotted with respect to the 1BPX (dark grey), 3C2M (black) and 2FMP (light grey) crystallographic coordinates. The left and right figures show two independent MD trajectories for the same simulated system. RMSD values obtained by averaging over the whole MD trajectory are presented in Table 5S.

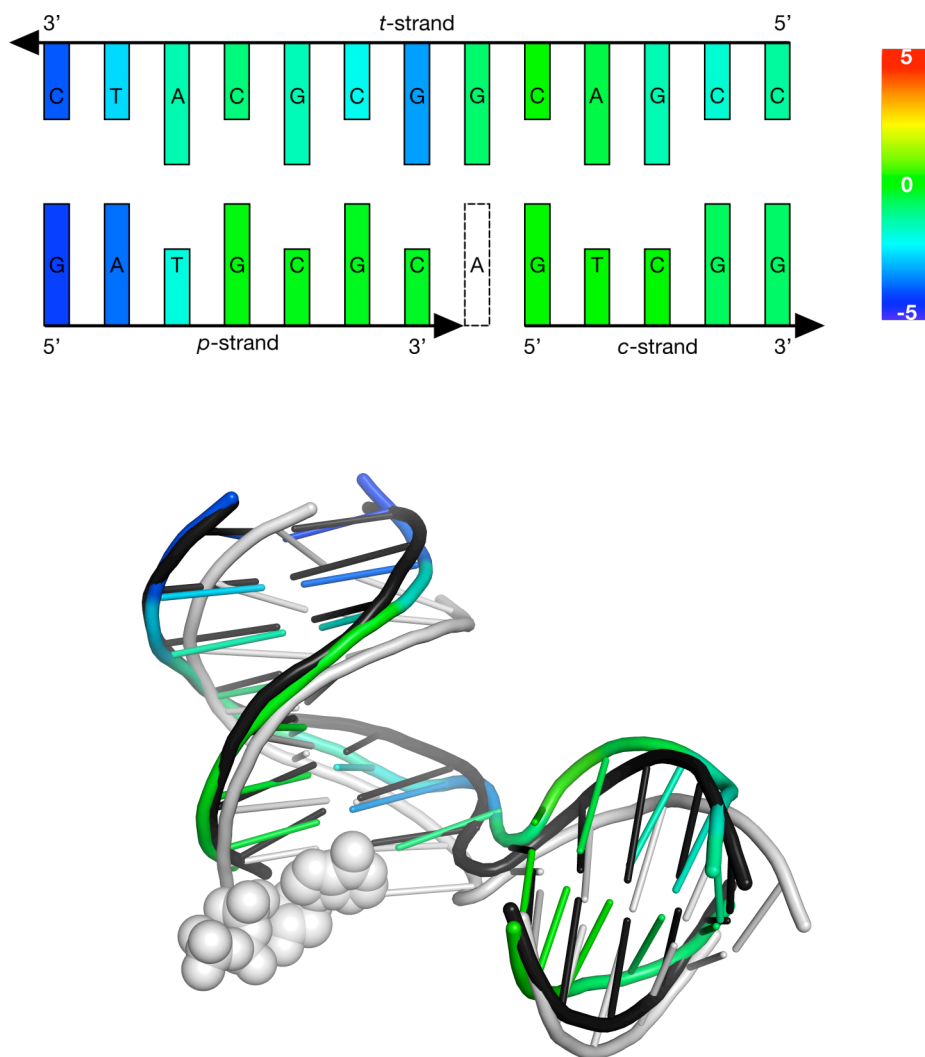


Figure 7S: Relative positions of individual nucleotides (top) and DNA (bottom) from MD simulations and crystal structures of the pol β ternary complexes. *Top:* Schematic representation of the template (t), primer (p) and complementary (c) DNA strands with the individual nucleotide colors reflecting the magnitude of the average nucleotide displacement in the MD simulation of the wrong GSA complex from the corresponding crystal structure (pdb code 3C2M). More specifically, the plotted $\Delta\text{RMSD} = \text{RMSD}_{\text{MD}} - \text{RMSD}_{\text{XC}}$, where RMSD_{MD} refers to average RMSD the nucleotide in the MD simulation of wrong GSA pol β complex (initiated from the 3C2M structure) with respect to the crystal structure of the right complex (pdb code 2FMP), and RMSD_{XC} refers to the average RMSD of the same nucleotide between the 3C2M and 2FMP crystal structures. The green color represents 0 Å displacement from the 3C2M crystal structure, whereas the colors between green and dark-blue represent simulated positions between 3C2M and 2FMP crystal structures. *Bottom:* Aligned average structure of DNA in 2FMP (white), 3C2M (dark grey), and MD simulation of the GSA state of the pol β complex containing the wrong dNTP (simulations was initiated from the 3C2M structure, nucleotides are colored by ΔRMSD , dATP is shown in white spheres).

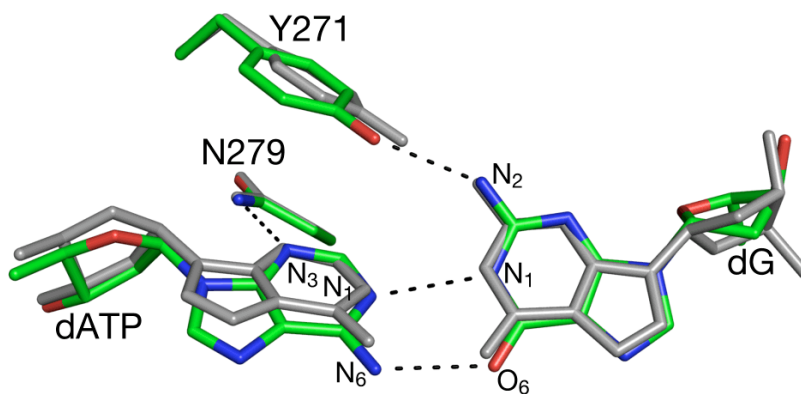


Figure 8S: Overlap of the average calculated structures of the dG•dATP mispair from the simulations of the PTS (atom-type colors) and GSA (grey) complexes. Hydrogen bonding interactions are shown with dashed lines. The calculated average $N_6H_6\cdots O_6$ and $N_1\cdots H_1N_1$ distances are 2.05 and 2.11 Å in the PTS, and 1.90 and 2.02 Å in the GSA state, respectively. These distances are longer than for the dG•dCTP base pair (not shown in this Figure), which equal to 1.94 (1.98), 1.92 (1.93) and 1.85 (1.87) Å for $O_6\cdots H_4N_4$, $N_1H_1\cdots N_3$, and $N_2H_2\cdots O_2$ H-bonds in the GSA (PTS) complexes, respectively.

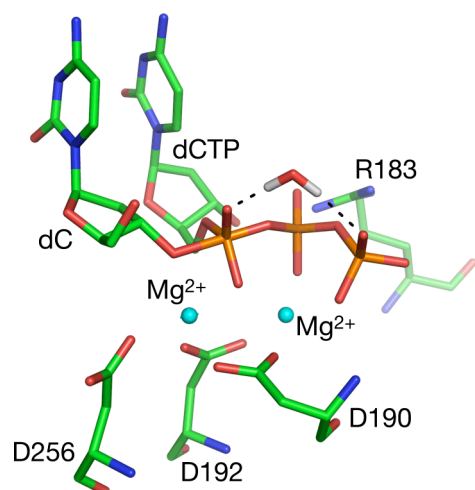


Figure 9S: Representative snapshot from an MD trajectory of the PTS complex. Representative hydrogen bonding interactions of a water molecule in the close distance from O_{lg} are shown with the dashed lines (detailed analysis of prevalent H-bonding patterns is presented in Tables 8S and 9S).

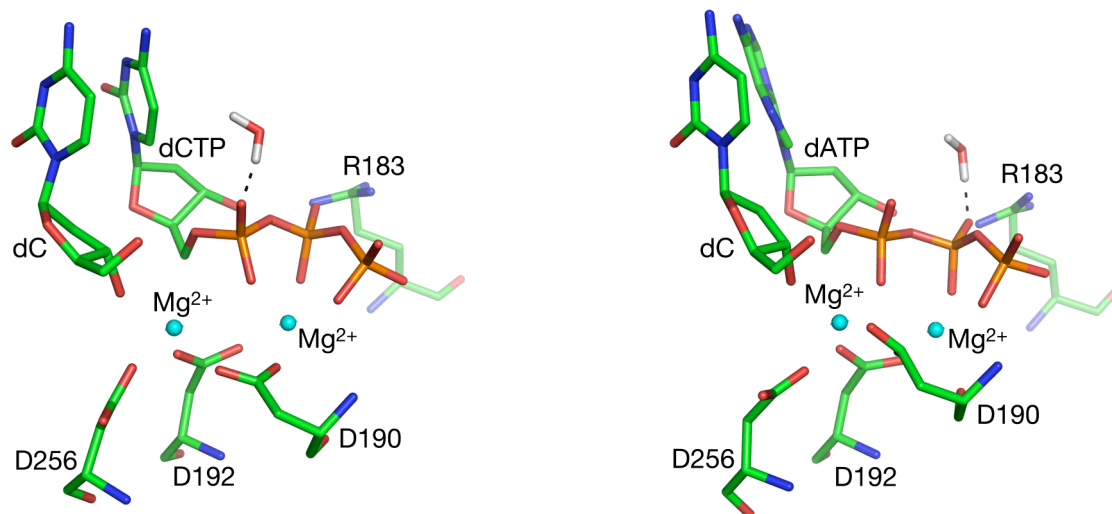


Figure 10S: Representative snapshots from MD simulations of the GSA complex containing correct (left) and incorrect (right) dNTP. Hydrogen bonding interactions of a water molecule near the leaving group oxygen with $O_{2\alpha}$ (left) and $O_{1\beta}$ atoms are shown with the dashed lines. These H-bonds are present in 64.3 and 53.1 percent of snapshots that have $O_{\text{wat}} \dots O_{\text{lg}}$ distance $< 3.5 \text{ \AA}$ (Table 9S).