Supplementary Information for

Uncovering a unique approach for damaged DNA replication: A computational investigation of a mutagenic tobacco-derived thymine lesion

Katie A. Wilson, Carl D. Holland, and Stacey D. Wetmore

Corresponding Author: Stacey D. Wetmore
Email: stacey.wetmore@uleth.ca

This PDF file includes:

- Supplementary Information Text (Full Computational Details)
- Figures S1 to S12
- Tables S1 to S14
- References for SI reference citations
Supplementary Information Text

Full Computational Details

Nucleobase. To parallel a previous study on POB-G (1), we initially investigated the flexibility of the POB-T nucleobase based on a fully extended orientation (i.e., $\alpha'$, $\beta'$, $\gamma'$, $\delta'$, $\epsilon'$, and $\rho'$ = 180°, Figure 1a). The rotational barriers with respect to $\alpha'$, $\beta'$, $\gamma'$, $\delta'$, $\epsilon'$, and $\rho'$ were determined by altering each dihedral angle in 10° increments from 0 to 360° using B3LYP-D3(BJ)/6-31G(d). Due to the absence of steric constraints, this model will provide the lowest energy rotational barriers within the bulky moiety. Subsequently, to gain a more comprehensive picture of the possible orientations of the POB-T nucleobase, a usage directed conformational search with respect to $\alpha'$, $\beta'$, $\gamma'$, $\delta'$, $\epsilon'$, and $\rho'$ as implemented in the HyperChem program (2,3) was performed using AMBER99 (see Table S13 for charges and atom types). A usage directed conformational search identifies all low energy conformations of a molecule by overcoming the multiple-minimum problem through uniformly using saved structures as starting points in the Monte Carlo algorithm. During the search, a maximum of 100,000 iterations or 1000 optimizations were performed. Any retained conformations were required to be within 42 kJ/mol of the lowest energy conformation, have an energy difference from any other conformation of $>0.2$ kJ/mol, and have a heavy atom root-mean-square deviation from any other conformation of < 0.25 Å. All 282 POB-T conformers isolated from this conformational search were subsequently optimized using B3LYP-D3(BJ)/6-31G(d) and the relative energies were determined using B3LYP-D3(BJ)/6-311++G(2df,2p). All duplicate conformations following the B3LYP-D3(BJ) analysis (defined based on an energy difference from any other conformation of < 0.2 kJ/mol, and a difference in $\alpha'$, $\beta'$, $\gamma'$, $\delta'$, $\epsilon'$, and $\rho'$ from any other conformation of < 0.2°) were removed from the data set. The 164 remaining conformations were visually categorized based on the type of interaction between the bulky moiety and the adducted T. The barrier about the glycosidic bond in the POB-T nucleoside was determined by scanning with respect to $\chi$ in 10° increments from 0 to 360° using B3LYP-D3(BJ)/6-31G(d). In the scan, a fully extended nucleoside model was considered that contains the 2'-endo sugar pucker, and the O3' hydroxy group oriented and the O5' hydroxy group fixed to reflect the geometry of a non-terminal nucleoside (i.e., $\angle$(HC3'O3'H) ≈ −60° and $\angle$(C4'C5'O5'H) = 180°). Stationary points from the scan were fully optimized using B3LYP-D3(BJ)/6-31G(d) and the relative energies were calculated with B3LYP-D3(BJ)/6-311++G(2df,2p) and include zero-point vibrational energy corrections scaled by 0.9670.

Isolated Base Pairs. Based on previous work on the POB-G adduct (1), M06-2X/6-31G(d) was used to optimized the hydrogen-bonding pairs between T or POB-T and the canonical nucleobases. The sugar–phosphate backbone of DNA was replaced by a methyl group to reduce computational costs and yet permit analysis of changes in the base pair width (C1′–C1′ distance) and base pair opening $\angle$(N9C1′C1′). Additionally, the base pair distortion was monitored by measuring the angle between the two nucleobase planes (denoted interplanar angle). The interaction energies were calculated with B3LYP-D3(BJ)/6-311++G(2df,2p) and include zero-point vibrational energy corrections scaled by 0.9670. The deformation energy and basis set superposition error corrections were not included in the reported binding strengths due to the significant flexibility of the POB-T nucleobase.

All DFT calculations were performed using Gaussian 09 (revision D.01) (4).

Polymerase. Initial structure for MD simulations on the pol $\kappa$ replication complexes were obtained from chains A, C, and D for a crystal structure corresponding to 2'-deoxy-5'-O-[(R)-hydroxy][(R)-hydroxy(phosphonoxy)phosphoryl]amino]phosphoryl]adenosine (dAMPNP) insertion opposite T for the extension past a lucidin guanine adduct derivative (PDB ID: 5W2C). Glycerol, 1,2-
ethanediol, and di(hydroxyethyl)ether were removed from the system. Missing N-terminal residues (Met–24 – Gly30), a loop (Met225 – Gln281) and C-terminal residues (Glu519 – Gln526) were not included in the model. Indeed, previous studies on pol κ mutants, as well as work on the pol κ bypass of the N2-furfuryl-G and N6-furfuryl-A lesions has shown that the exclusion of these residues does not affect the activity of the enzyme (5,6).

Initial structures for MD simulations on the pol η replication complexes were obtained from a crystal structure corresponding to the insertion of dATP opposite T (PDB ID: 4ECS). Missing residues (Thr155–Glu159) were added by hand using GaussView. Glycerol, Ca²⁺, and pyrophosphate were removed from the system. In cases where multiple orientations of amino acids were present in the experimental data, the orientation that best aligns the residue to interact with the surrounding DNA or protein was measured. Although this approach does not account for rearrangement of the active site to accommodate different lesion orientations, the resulting structures are reasonable starting points for MD simulations. Representative lesion orientations were then used to initiate MD simulation to understand the dynamics of POB-T binding in the pol η and κ active sites, and further explore conformations that can be accommodated by the enzymes. Additionally, directed by DFT calculations on the isolated base pairs, pol κ insertion complexes with a pairing dGTP or dATP opposite anti-POB-T, and pol η insertion complexes with dCTP, dTTP or dATP paired opposite syn-POB-T were modeled.

The resulting complexes were prepared for simulation using the tleap module of AMBER 14 (7). Specifically, missing atoms, including hydrogen atoms, were added to generate the standard protonation states of all DNA and protein residues. Furthermore, NaCl was added to the water box to yield an approximate concentration of 0.150 M for DNA–polymerase complexes. Mg²⁺ ions were included in the DNA–polymerase insertion complexes. All systems were solvated in a TIP3P octahedral water box such that the solute was at least 10.0 Å from the edge of the box. All canonical amino acids, nucleotides, and solvent were modeled with AMBER ff14SB parameters. Parameters for POB-T were assigned according to the GAFF and AMBER ff14SB force fields using ANTECHAMBER 1.4 (8), and partial charges for the lesions were developed using RESP charge fitting from a HF/6-31G(d) calculation by the R.E.D.v.III.4 program (9,10) to be consistent with the AMBER force field (Table S12). The parameters for the dNTPs were adapted from the literature (11-13), Mg²⁺ was modeled using the parameters from Allner et al. (14), and Na⁺ and Cl⁻ were modeled using the monovalent ion parameters from Joung and Cheatham (15).

All systems were minimized stepwise, with the first phase minimizing the water for 2500 steps of steepest decent and 2500 steps of conjugant gradient minimization, while applying a 50 kcal/(mol Å²) restraint on the rest of the system. Next, the hydrogen atoms were minimized for 4000 steps of steepest decent and 4000 steps of conjugant gradient minimization, with a 50 kcal/(mol Å²) restraint on the heavy atoms. Subsequently, the systems were minimized with a 50 kcal/(mol Å²) restraint on the amino acid backbone for 12500 steps of steepest decent and 12500 steps of conjugant gradient minimization. Finally, the entire system was minimized without restraints for 5000 steps of steepest decent and 5000 steps of conjugant gradient minimization. The systems were then heated from 10 to 310 K in 6 steps, each increasing the temperature by 50 K over 10 ps using a 1 fs time step. Heating was performed with a 10 kcal/(mol Å²) restraint on the solute using the
The systems were then equilibrated with a constraint weight of 20, 15, 10, 5 and 1 kcal/(mol Å²) on the solute. Each constraint weight was simulated for 20 ps using a 2 fs time step. Subsequently, 20 ns unconstrained pre-production simulations were performed to identify geometries that were distinctly different from the initial structures and cannot be easily converted during standard MD sampling, with particular attention being paid to the active site and lesion site (including lesion orientation and hydrogen bonding). Structures of the polymerase-DNA complexes with a hydrogen-bonding geometry and lesion orientation most conducive for the reaction were selected as initial structures for 100 ns MD simulations, which were run in triplicate to ensure statistical significance. Since the replicas lead to minimal differences in the active site region (rmsd of 0.7–1.8 Å, Table S11), one replica was extended to yield the final 0.5 μs MD production simulation results discussed in the main text.

Analysis of the MD simulations was performed on frames spaced by 0.1 ns. A representative structure based on the lesion orientation and hydrogen bonding was chosen using the average-linkage clustering algorithm. All MD simulations were performed using AMBER 14 (6). The strengths of the interactions between the dNTP and lesion were calculated using frames spaced by 1 ns, B3LYP-D3(BJ)/6-311++G(d,p), hydrogen-capped nucleobases (hydrogen atoms optimized with M06-2X/6-31G(d)) and Gaussian 09 (revision D.01) (4). A smaller basis set was used for calculating the interaction energies relative to the basis set used in the nucleobase dimer calculations due to the number of structures considered. Nevertheless, the difference between nucleobase dimer interaction strengths calculated with B3LYP-D3(BJ)/6-311++G(2df,2p) and B3LYP-D3(BJ)/6-311++G(d,p) is ~1–8 kJ/mol (Table S13).
**Figure S1.** B3LYP-D3(BJ)/6-31G(d) minimum energy structures for rotation about the a) $\alpha'$, b) $\beta'$, c) $\gamma'$, d) $\delta'$, e) $\epsilon'$, and f) $\rho'$ dihedral angles initiated from a fully extended POB-T nucleobase model. See Figure 1 for definitions of the dihedral angles.
Figure S2. B3LYP-D3(BJ)/6-31G(d) potential energy surfaces (kJ/mol) for rotation about important POB-T dihedral angles (deg.). See Figure 1 for definitions of dihedral angles.
Figure S3. Overlay of all unique conformers isolated in the conformational search (based on damaged T heavy atoms, left), and lowest energy conformation (right) with the bulky moiety in red and the T nucleobase carbon atoms in grey.
Figure S4. a) Distribution of the POB-T nucleobase conformations resulting from the DFT conformational search and b) distribution of the B3LYP-D3/6-311++G(2df,2p) relative energies (kJ/mol) for each structural category. Energies are reported relative to the overall lowest energy structure.
Figure S5. Energetic effects (B3LYP-D3/6-311++G(2df,2p), kJ/mol) of different bulky moiety orientations in the isolated stacked (red), T-shaped (yellow), hydrogen bonded (blue), or extended (orange) structures from the conformational search.
Figure S6. a) Closest heavy atom distance and b) orientations of the POB-T bulky moiety from the DFT conformational search that fit in the pol κ (PDB ID: 5W2C) active site, with the lesion in the *anti* orientation.
Figure S7. a) Overlay of MD representative structure of the entire enzyme–DNA complex for anti-POB-T bound in the pol κ active site (purple) and the crystal structure (grey) for dATP insertion opposite dT (PDB ID: 5W2C). b) Representative MD structure of the pre-insertion complex and c) conformations of POB-T adopted throughout the MD simulation on the pre-insertion complex when anti-POB-T is bound in the pol κ active site.
Figure S8. a) Closest heavy atom distance and b) orientations of the POB-T bulky moiety from the DFT conformational search that fit in the pol η (PDB ID: 4ECS) active site, with the lesion in the anti orientation (left), or pol η active site (PDB ID: 4ECS), with the lesion in the syn orientation (right).
**Figure S9.** a) Overlay of MD representative structures of the entire enzyme–DNA complex for *anti*-POB-T (left) or *syn*-POB-T (right) bound in the pol η active site (purple) and the crystal structure (grey) for dATP insertion opposite dT (PDB ID: 4ECS). b) Representative MD structure and c) conformations adopted throughout the MD simulation on the pre-insertion complex when *anti*-POB-T (left) or *syn*-POB-T (right) is bound in the pol η active site.
**Figure S10.** B3LYP-D3/6-311++G(2df,2p)//B3LYP-D3(BJ)/6-31G(d) structures and relative energies (kJ/mol) for rotation about the POB-T glycosidic bond.
Figure S11. Overlay (based on damaged T heavy atoms, left) and 2D structural representation (right) of lesion conformations adopted throughout the MD simulation on the pol κ complex for the insertion of a) anti-dATP, b) syn-dATP or c) dGTP opposite POB-T, highlighting the deviation in the bulky moiety (red) and pairing base orientations.
Figure S12. Overlay (based on damaged T heavy atoms, left) and 2D structural representation (right) of lesion conformations adopted throughout the MD simulation on the pol η complex for the insertion of a) anti-dATP, b) syn-dATP, c) dCTP, or d) dTTP opposite syn-POB-T, highlighting the deviation in the bulky moiety (red) and pairing base orientations.
Table S1. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for active site hydrogen bonds in the pol κ complex for insertion of *anti*-dATP opposite POB-T.$^a$

| Hydrogen Bond                        | Occupancy | Average Distance | Average Angle |
|--------------------------------------|-----------|------------------|---------------|
| 3′-dG(O6)···dC(N4H)                  | 97%       | 2.958            | 161.0         |
| 3′-dG(N1H)···dC(N3)                  | 100%      | 2.960            | 162.2         |
| 3′-dG(N2H)···dC(O2)                  | 100%      | 2.861            | 161.1         |
| POB-T(N14)···Ser102(OγH)             | 24%       | 2.965            | 155.3         |
| POB-T(O11)···Ser107(OγH)             | 19%       | 2.895            | 148.8         |
| dATP(O3′)···Tyr82(NH)                | 80%       | 3.170            | 165.1         |
| dATP(Oβ1)···Phe81(NH)                | 93%       | 3.149            | 157.4         |
| dATP(Oβ2)···Thr108(OG1)              | 98%       | 2.689            | 164.4         |
| dATP(Oα)···Wat-Mg$^{2+}$(OH)         | 80%       | 2.705            | 146.6         |
| dATP(Oβγ)···Arg114(NH2H)             | 49%       | 3.096            | 141.6         |
| dATP(Oγ3)···Lys241(NZH)              | 92%       | 2.791            | 149.5         |
| dATP(Oγ1)···Ala80(NH)                | 99%       | 2.944            | 164.7         |
| dATP(Oγ1)···Arg114(NH2H)             | 87%       | 2.904            | 154.8         |
| dATP(Oγ2)···Arg114(NH1H)             | 96%       | 2.914            | 155.8         |

$^a$Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°.
Table S2. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for active site hydrogen bonds in the polκ complex for insertion of syn-dATP opposite POB-T.\(^a\)

| Bond Description                  | Occupancy | Average Distance | Average Angle |
|-----------------------------------|-----------|------------------|---------------|
| 3′-dG(O6)···dC(N4H)               | 98%       | 2.951            | 158.0         |
| 3′-dG(N1H)···dC(N3)               | 100%      | 2.964            | 164.6         |
| 3′-dG(N2H)···dC(O2)               | 100%      | 2.875            | 162.0         |
| POB-T(N14)···Ser102(OγH)          | 16%       | 2.946            | 156.9         |
| dATP(O3′)···Tyr82(NH)             | 29%       | 3.050            | 161.1         |
| dATP(Oβ1)···Phe81(NH)             | 95%       | 3.141            | 157.0         |
| dATP(Oβ2)···Thr108(OG1)           | 100%      | 2.669            | 165.6         |
| dATP(Oα)···Wat-Mg\(^{2+}\)(OH)   | 82%       | 2.705            | 147.7         |
| dATP(Oβγ)···Arg114(NH2H)          | 73%       | 3.016            | 152.9         |
| dATP(Oγ3)···Tyr111(OH)            | 61%       | 2.631            | 166.8         |
| dATP(Oγ3)···Lys241(NZH)           | 87%       | 2.800            | 147.4         |
| dATP(Oγ1)···Ala80(NH)             | 100%      | 2.904            | 165.6         |
| dATP(Oγ1)···Arg114(NH2H)          | 99%       | 2.927            | 157.4         |
| dATP(Oγ2)···Arg114(NH1H)          | 72%       | 2.998            | 149.0         |

\(^a\)Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°.
Table S3. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for active site hydrogen bonds in the pol κ complex for insertion of dGTP opposite POB-T.\(^a\)

| Bond                        | Occupancy | Average Distance | Average Angle |
|-----------------------------|-----------|------------------|---------------|
| 3′-dG(O6)···dC(N4H)         | 99%       | 2.931            | 163.0         |
| 3′-dG(N1H)···dC(N3)         | 100%      | 2.976            | 163.8         |
| 3′-dG(N2H)···dC(O2)         | 99%       | 2.941            | 163.3         |
| POB-T(O4)···GTP(N1H)        | 99%       | 2.924            | 160.0         |
| POB-T(N3)···GTP(N2H)        | 96%       | 3.074            | 154.6         |
| dCTP(O3′)···Tyr82(NH)       | 64%       | 3.194            | 165.8         |
| dCTP(Oβ1)···Phe81(NH)       | 100%      | 2.995            | 158.1         |
| dCTP(Oβ2)···Thr108(OG1)     | 99%       | 2.767            | 160.5         |
| dTTP(Oα)···Wat-Mg\(^{2+}\)(OH) | 73%       | 2.849            | 149.4         |
| dCTP(Oβγ)···Arg114(NH2H)    | 18%       | 2.920            | 160.3         |
| dCTP(Oγ1)···Tyr111(OH)      | 19%       | 2.608            | 165.7         |
| dCTP(Oγ1)···Lys241(NZH)     | 91%       | 2.788            | 149.1         |
| dCTP(Oγ3)···Lys241(NZH)     | 76%       | 3.093            | 144.1         |
| dCTP(Oγ1)···Ala80(NH)       | 77%       | 3.015            | 135.5         |
| dCTP(Oγ2)···Arg114(NH1H)    | 21%       | 3.040            | 142.7         |

\(^a\)Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°.
Table S4. Coordination of the active site Mg\(^{2+}\) ions across the simulation for the pol κ and η insertion complexes corresponding to various lesion replication outcomes.\(^a\)

| Enzyme | Base Pair | Mg\(^{2+}\) (binding) | Mg\(^{2+}\) (catalytic) |
|--------|-----------|------------------------|------------------------|
| pol κ  | T:dATP    | 6/6                    | 6/6                    |
|        | POB-T:anti-dATP | 6/6              | 6/6                    |
|        | POB-T:syn-dATP | 6/6              | 6/6                    |
|        | POB-T:dGTP  | 5/6                    | 5/6                    |
| pol η  | T:dATP    | 6/6                    | 6/6                    |
|        | POB-T:anti-dATP | 6/6              | 6/6                    |
|        | POB-T:syn-dATP | 6/6              | 6/6                    |
|        | POB-T:dCTP  | 6/6                    | 5/6                    |
|        | POB-T:dTTP  | 6/6                    | 5/6                    |

\(^a\)Atom is considered coordinated if the distance is < 2.5 Å for > 90% of the simulation.
Table S5. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for active site hydrogen bonds in the pol κ complex for insertion of anti-dATP opposite T.

| Bond Description | Occupancy | Average Distance | Average Angle |
|------------------|-----------|------------------|---------------|
| 3′-dG(O6)···dC(N4H) | 99%       | 2.944            | 163.3         |
| 3′-dG(N1H)···dC(N3)  | 100%      | 2.966            | 161.9         |
| 3′-dG(N2H)···dC(O2)  | 100%      | 2.864            | 160.9         |
| dATP(O3′)···Tyr82(NH) | 93%       | 3.128            | 166.6         |
| dATP(Oβ1)···Phe81(NH) | 94%       | 3.154            | 155.7         |
| dATP(Oβ2)···Thr108(OG1) | 100%    | 2.670            | 166.1         |
| dATP(Oα)···Wat·Mg^{2+}O(H) | 86% | 2.705            | 147.0         |
| dATP(Oβγ)···Arg114(NH2H) | 61%      | 2.971            | 158.5         |
| dATP(Oγ3)···Thr111(OH) | 63%       | 2.686            | 164.1         |
| dATP(Oγ3)···Lys241(NZH) | 65%      | 2.880            | 145.6         |
| dATP(Oγ1)···Ala80(NH) | 99%       | 2.909            | 166.5         |
| dATP(Oγ1)···Arg114(NH2H) | 77%       | 3.014            | 148.4         |
| dATP(Oγ2)···Arg114(NH1H) | 99%       | 2.926            | 160.0         |

*Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°.
### Table S6. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for active site hydrogen bonds in the pol η complex for insertion of *anti*-dATP opposite T.\(^a\)

| Bond                                | Occupancy | Average Distance | Average Angle |
|-------------------------------------|-----------|------------------|---------------|
| 3’-dT(O4)···dA(N6H)                 | 96%       | 2.989            | 159.8         |
| 3’-dT(N3H)···dA(N1)                 | 99%       | 2.964            | 163.3         |
| dATP(N7)···Arg61(N1H)               | 38%       | 2.958            | 159.1         |
| dATP(O3’)···Phe18(NH)               | 87%       | 3.144            | 167.5         |
| dATP(Oβ2)···Phe17(NH)               | 90%       | 3.182            | 161.3         |
| dATP(Oα1)···Wat-Mg\(^{2+}\)(OH)     | 46%       | 2.800            | 142.5         |
| dATP(Oγ3)···Tyr52(OH)               | 99%       | 2.607            | 164.7         |
| dATP(Oγ1)···Lys231(NZH)             | 66%       | 2.924            | 141.0         |
| dATP(Oγ3)···Lys231(NZH)             | 73%       | 3.001            | 157.2         |
| dATP(Oγ3)···Cys16(NH)               | 99%       | 2.928            | 164.6         |
| dATP(Oβγ)···Arg55(NH2H)             | 98%       | 2.917            | 152.7         |
| dATP(Oγ3)···Arg55(NH1H)             | 92%       | 2.920            | 158.9         |

\(^a\)Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°.
**Table S7.** Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for active site hydrogen bonds in the pol η complex for insertion of *anti*-dATP opposite *syn*-POB-T.\(^a\)

| Hydrogen Bond | Occupancy | Average Distance | Average Angle |
|---------------|-----------|------------------|---------------|
| 3′-T(O4)···A(N6H) | 90% | 2.976 | 150.9 |
| 3′-T(N3H)···A(N1) | 100% | 2.927 | 164.0 |
| Ser116(Oγ)···3′-T(O3′H) | 85% | 2.953 | 156.0 |
| dATP(N7)···Arg61(N1H) | 67% | 2.948 | 158.6 |
| dATP(O3′)···Phe18(NH) | 84% | 3.146 | 167.4 |
| dATP(Oβ2)···Phe17(NH) | 94% | 3.134 | 160.4 |
| dATP(Oα1)···Wat-Mg\(^{2+}\)(OH) | 53% | 2.792 | 143.8 |
| dATP(Oγ3)···Tyr52(OH) | 77% | 2.609 | 166.3 |
| dATP(Oγ1)···Lys231(NZH) | 62% | 2.914 | 141.3 |
| dATP(Oγ3)···Lys231(NZH) | 62% | 2.914 | 141.3 |
| dATP(Oγ3)···Cys16(NH) | 99% | 2.948 | 165.0 |
| dATP(Oβγ)···Arg55(NH2H) | 96% | 2.949 | 149.6 |
| dATP(Oγ3)···Arg55(NH1H) | 88% | 2.921 | 157.3 |

\(^a\)Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°.
| Hydrogen Bond                  | Occupancy | Average Distance | Average Angle |
|-------------------------------|-----------|------------------|---------------|
| 3'-T(O4)···A(N6H)             | 79%       | 3.064            | 149.0         |
| 3'-T(N3H)···A(N1)             | 100%      | 2.928            | 163.7         |
| Ser116(Oγ)···3'-T(O3'H)       | 32%       | 2.956            | 155.5         |
| dATP(Oα)···Arg61(N1H)         | 23%       | 2.968            | 159.1         |
| dATP(O3')···Phe18(NH)         | 74%       | 3.160            | 166.9         |
| dATP(Oβ2)···Phe17(NH)         | 92%       | 3.146            | 157.9         |
| dATP(Oα1)···Wat·Mg²⁺(OH)      | 32%       | 2.791            | 144.1         |
| dATP(Oγ3)···Tyr52(OH)         | 56%       | 2.609            | 165.9         |
| dATP(Oγ1)···Lys231(NZH)       | 50%       | 2.914            | 142.9         |
| dATP(Oγ3)···Lys231(NZH)       | 40%       | 2.992            | 154.4         |
| dATP(Oγ3)···Cys16(NH)         | 100%      | 2.914            | 165.6         |
| dATP(Oβγ)···Arg55(NH2H)       | 91%       | 2.992            | 144.3         |
| dATP(Oγ2)···Arg55(NH1H)       | 72%       | 2.962            | 150.3         |
| dATP(Oγ3)···Arg55(NH1H)       | 66%       | 2.974            | 154.8         |

*Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°.
Table S9. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for active site hydrogen bonds in the pol η complex for insertion of dCTP opposite syn-POB-T.\textsuperscript{a}

| Bond Description                        | Occupancy | Average Distance | Average Angle |
|-----------------------------------------|------------|------------------|---------------|
| 3′-T(O4)···A(N6H)                        | 89%        | 3.020            | 158.3         |
| 3′-T(N3H)···A(N1)                        | 99%        | 2.952            | 160.2         |
| dCTP(O3')···Phe18(NH)                   | 96%        | 3.062            | 161.6         |
| Ser113(Oγ)···3′-dT(O3'H)                | 90%        | 2.880            | 158.4         |
| dCTP(Oα1)···Arg61(NH1H)                 | 65%        | 2.854            | 154.3         |
| dCTP(Oβ2)···Phe17(NH)                   | 78%        | 3.206            | 159.1         |
| 3′-T(OP)···Wat-Mg\textsuperscript{2+}(OH)| 26%        | 2.809            | 142.4         |
| dCTP(Oβγ)···Arg55(NH2H)                 | 97%        | 2.936            | 149.1         |
| dCTP(Oγ3)···Tyr52(OH)                   | 99%        | 2.581            | 166.3         |
| dCTP(Oγ3)···Lys231(NZH)                 | 45%        | 3.038            | 155.8         |
| dCTP(Oγ1)···Lys231(NZH)                 | 69%        | 2.915            | 144.5         |
| dCTP(Oγ3)···Cys16(NH)                   | 97%        | 2.948            | 165.3         |
| dCTP(Oγ3)···Arg55(NH2H)                 | 71%        | 2.977            | 147.8         |
| dCTP(Oγ3)···Arg55(NH1H)                 | 70%        | 2.972            | 158.2         |
| dCTP(Oγ2)···Arg55(NH1H)                 | 53%        | 3.054            | 142.4         |

\textsuperscript{a}Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°.
Table S10. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for active site hydrogen bonds in the pol η complex for insertion of dTTP opposite syn-POB-T.\(^a\)

| Hydrogen Bond | Occupancy | Average Distance | Average Angle |
|---------------|-----------|------------------|---------------|
| 3′-T(O4)···A(N6H) | 97%       | 2.933            | 155.7         |
| 3′-T(N3H)···A(N1)  | 98%       | 3.031            | 161.8         |
| POB-T(O4)···dATP(N3H) | 95%       | 2.835            | 151.8         |
| dTTP(O3′)···Phe18(NH) | 95%       | 3.040            | 161.5         |
| Ser113(Oγ)···3′-dT(O3′H) | 43%       | 2.980            | 154.0         |
| dTTP(Oα1)···Arg61(NH1H) | 69%       | 2.879            | 154.3         |
| dTTP(Oβ2)···Phe17(NH) | 90%       | 3.166            | 159.9         |
| dTTP(Oα)···Wat-Mg\(^{2+}\)(OH) | 27%       | 2.832            | 139.6         |
| dTTP(Oβγ)···Arg55(NH2H) | 96%       | 2.971            | 146.4         |
| dTTP(Oγ3)···Tyr52(OH) | 85%       | 2.577            | 165.2         |
| dTTP(Oγ3)···Lys231(NZH) | 26%       | 3.031            | 136.2         |
| dTTP(Oγ1)···Lys231(NZH) | 37%       | 2.986            | 150.7         |
| dTTP(Oγ3)···Cys16(NH) | 96%       | 2.976            | 166.3         |
| dTTP(Oγ3)···Arg55(NH2H) | 84%       | 2.947            | 148.9         |
| dTTP(Oγ3)···Arg55(NH1H) | 71%       | 3.029            | 145.0         |
| dTTP(Oγ2)···Arg55(NH1H) | 51%       | 3.066            | 152.8         |

\(^a\)Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°.
Table S11. Heavy atom rmsd (Å) with respect to the representative structure for Trial 1.

| pol | anti-POB-T | Trial 1<sup>a</sup> | Trial 1<sup>b</sup> | Trial 2<sup>a</sup> | Trial 3<sup>a</sup> |
|-----|------------|-------------------|-------------------|-------------------|-------------------|
|     | Entire System | Active Site | Entire System | Active Site | Entire System | Active Site |
| κ   | 1.936±0.332  | 1.346±0.295  | 2.305±0.333  | 1.568±0.255  | 2.289±0.194  | 1.770±0.214  | 2.373±0.239  | 1.578±0.132  |
|     | anti-POB-T:anti-dATP | 2.304±0.359 | 1.390±0.538 | 2.305±0.359 | 1.390±0.538 | 2.178±0.147 | 0.863±0.123 | 2.398±0.307 | 1.001±0.177 |
|     | anti-POB-T: syn-dATP | 1.864±0.372 | 0.670±0.161 | 1.984±0.293 | 1.006±0.254 | 2.404±0.206 | 0.863±0.140 | 2.816±0.365 | 0.832±0.104 |
|     | anti-POB-T: anti-dGTP | 1.834±0.186 | 1.050±0.452 | 2.397±0.346 | 1.785±0.451 | 2.164±0.142 | 1.119±0.157 | 2.119±0.234 | 0.853±0.123 |
| η   | anti-POB-T | 2.056±0.272  | 1.966±0.329  | 2.189±0.273  | 1.231±0.319  | 2.472±0.312  | 1.432±0.209  | 2.157±0.132  | 1.526±0.164  |
|     | syn-POB-T | 2.075±0.244  | 1.488±0.262  | 2.906±0.342  | 1.941±0.158  | 2.751±0.436  | 1.647±0.140  | 2.401±0.126  | 1.657±0.193  |
|     | syn-POB-T: anti-dATP | 1.948±0.285 | 0.700±0.201 | 2.830±0.364 | 0.988±0.438 | 2.187±0.160 | 1.230±0.129 | 2.712±0.194 | 1.542±0.106 |
|     | syn-POB-T: syn-dATP | 2.109±0.356 | 0.959±0.263 | 2.104±0.308 | 1.226±0.326 | 2.424±0.317 | 1.439±0.354 | 2.562±0.158 | 1.513±0.118 |
|     | syn-POB-T: anti-dCTP | 2.169±0.244 | 1.015±0.140 | 2.936±0.484 | 0.975±0.137 | 2.480±0.103 | 0.961±0.121 | 2.632±0.201 | 1.262±0.254 |
|     | syn-POB-T: anti-dTTP | 1.865±0.204 | 1.020±0.340 | 2.179±0.284 | 1.299±0.245 | 2.401±0.198 | 0.910±0.139 | 2.920±0.220 | 1.345±0.328 |

<sup>a</sup>100 ns simulations. <sup>b</sup>500 ns simulations.
Table S12. Atom types and charges for POB-T.

| Atom Name | Atom Type | Charge  |
|-----------|-----------|---------|
| P         | P         | 1.1623  |
| O5'       | OS        | -0.4914 |
| OP1       | O2        | -0.7616 |
| OP2       | O2        | -0.7706 |
| O3'       | OS        | -0.5532 |
| C5'       | CT        | 0.0105  |
| H5'1      | H1        | 0.0681  |
| H5'2      | H1        | 0.0681  |
| C4'       | CT        | 0.1734  |
| H4'       | H1        | 0.0965  |
| O4'       | OS        | -0.3982 |
| C1'       | CT        | 0.2113  |
| H1'       | H2        | 0.0585  |
| C3'       | CT        | 0.1826  |
| H3'       | H1        | 0.0609  |
| C2'       | CT        | -0.0839 |
| H2'1      | HC        | 0.0486  |
| H2'2      | HC        | 0.0486  |
| N1        | N*        | -0.1376 |
| C6        | CM        | -0.2167 |
| H6        | H4        | 0.2227  |
| C5        | CM        | -0.0481 |
| C7        | CT        | -0.2565 |
| H71       | HC        | 0.0828  |
| H72       | HC        | 0.0828  |
| H73       | HC        | 0.0828  |
| C4        | C         | 0.7946  |
| O4        | O         | -0.609  |
| N3        | NC        | -0.7289 |
| C2        | CA        | 0.6597  |
| O2        | OS        | -0.2901 |
| C8        | CT        | 0.0419  |
| H81       | H1        | 0.0654  |
| H82       | H1        | 0.0654  |
| C9        | CT        | 0.0684  |
| H91       | HC        | 0.0007  |
| H92       | HC        | 0.0007  |
| C10       | CT        | 0.0492  |
| H101      | HC        | -0.0082 |
| H102      | HC        | -0.0082 |
| C11       | C         | 0.4364  |
| O11       | O         | -0.4909 |
| C12       | CA        | -0.1437 |
| C13       | CA        | 0.2872  |
| H13       | H4        | 0.0733  |
| C17       | CA        | -0.0146 |
| H17       | HA        | 0.1505  |
| C16       | CA        | -0.301  |
| H11       | HA        | 0.1509  |
| C15       | CA        | 0.3295  |
| H15       | H4        | 0.0643  |
| N14       | NC        | -0.5864 |
Table S13. Comparison of B3LYP-D3(BJ) binding energies (kJ/mol) calculated with 6-311++G(2df,2p) and 6-311++G(d,p) for base pairs between the Watson-Crick (WC) face or Hoogsteen (H) face of POB-T and the canonical nucleobases.\(^a\)

|                  | B3LYP-D3(BJ)/6-311++G(2df,2p) | B3LYP-D3(BJ)/6-311++G(d,p) | Deviation |
|------------------|-------------------------------|----------------------------|-----------|
| WC-POB-T:WC-A    | –71.3                         | –78.9                      | 7.6       |
| WC-POB-T:H-A     | –68.8                         | –75.8                      | 7.0       |
| WC-POB-T:WC-T    | –70.6                         | –75.5                      | 4.9       |
| WC-POB-T:WC-C    | –80.7                         | –86.5                      | 5.8       |
| WC-POB-T:WC-G    | –93.6                         | –96.8                      | 3.2       |
| H-POB-T:WC-A     | –36.5                         | –37.8                      | 1.3       |
| H-POB-T:H-A      | –35.3                         | –36.5                      | 1.2       |
| H-POB-T:WC-T     | –35.9                         | –36.9                      | 1.0       |
| H-POB-T:WC-C     | –47.7                         | –49.7                      | 2.0       |

\(^a\)Single-point calculations were performed on M06-2X/6-31G(d) optimized geometries.
References

1. Wilson, K.A., Szemethy, K.G. and Wetmore, S.D. (2017) Conformational flexibility and base-pairing tendency of the tobacco carcinogen O6-[4-oxo-4-(3-pyridyl)butyl]guanine. *Biophys. Chem.*, **228**, 25-37.

2. HyperChem. Professional 7.5 ed. Hypercube, Inc., Gainesville, Florida 32601, USA.

3. Chang, G., Guida, W.C. and Still, W.C. (1989) An internal-coordinate monte carlo method for searching conformational space. *J. Am. Chem. Soc.*, **111**, 4379-4386.

4. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G.A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H.P., Izmaylov, A.F., Bloino, J., Zheng, G., Sonnenberg, J.L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Jr., J.A.M., Peralta, J.E., Ogliaro, F., Bearpark, M., Heyd, J.J., Brothers, E., Kudin, K.N., Staroverov, V.N., Keith, T., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J.C., Iyengar, S.S., Tomasi, J., Cossi, M., Rega, N., Millam, J.M., Klene, M., Knox, J.E., Cross, J.B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev, O., Austin, A.J., Cammi, R., Pomelli, C., Ochterski, J.W., Martin, R.L., Morokuma, K., Zakrzewski, V.G., Voth, G.A., Salvador, P., Dannenberg, J.J., Dapprich, S., Daniels, A.D., Farkas, O., Foresman, J.B., Ortiz, J.V., Cioslowski, J. and Fox, D.J. (2013). Revision D.01 ed. Gaussian, Inc., Wallingford CT.

5. Antczak, N.M., Walker, A.R., Stern, H.R., Leddin, E.M., Palad, C., Coulther, T.A., Swett, R.J., Cisneros, G.A. and Beuning, P.J. (2018) Characterization of nine cancer-associated variants in human DNA polymerase κ. *Chem. Res. Toxicol.*

6. Antczak, N.M., Packer, M.R., Lu, X., Zhang, K. and Beuning, P.J. (2017) Human y-family DNA polymerase κ is more tolerant to changes in its active site loop than its ortholog Escherichia coli dinb. *Chem. Res. Toxicol.*, **30**, 2002-2012.

7. Case, D.A., Babin, V., Berryman, J.T., Betz, R.M., Cai, Q., Cerutti, D.S., T.E. Cheatham, I., Darden, T.A., Duke, R.E., Gohlke, H., Goetz, A.W., Gusarov, S., Homeyer, N., Janowski, P., Kaus, J., Kolossváry, I., Kovalenko, A., Lee, T.S., LeGrand, S., Luchko, T., Luo, R., Madej, B., Merz, K.M., Paesani, F., Roe, D.R., Roitberg, A., Saguí, C., Salomon-Ferrer, R., Seabra, G., Simmerling, C.L., Smith, W., Swails, J., Walker, R.C., Wang, J., Wolf, R.M., Wu, X. and Kollman, P.A. (2014). University of California, San Francisco.

8. Wang, J., Wang, W., Kollman, P.A. and Case, D.A. (2006) Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.*, **25**, 247–260.

9. Dupradeau, F.-Y., Pigache, A., Zaffran, T., Savineau, C., Lelong, R., Lelong, D., Rosanski, W. and Cieplak, P. (2010) The r.E.D. Tools: Advances in resp and esp charge derivation and force field library building. *Phys. Chem. Chem. Phys.*, **12**, 7821-7839.

10. Vanquellef, E., Simon, S., Marquant, G., Garcia, E., Klimerak, G., Delepine, J.C., Cieplak, P. and Dupradeau, F.-Y. (2011) R.E.D. Server: A web service for deriving resp and esp charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.*, **39**, W511-W517.
11. Zhang, L., Rechkoblit, O., Wang, L., Patel, D.J., Shapiro, R. and Broyde, S. (2006) Mutagenic nucleotide incorporation and hindered translocation by a food carcinogen C8-dG adduct in sulfolobus solfataricus p2 DNA polymerase IV (Dpo4): Modeling and dynamics studies. *Nucleic Acids Res.*, **34**, 3326-3337.

12. Perlow, R.A. and Broyde, S. (2002) Toward understanding the mutagenicity of an environmental carcinogen: Structural insights into nucleotide incorporation preferences. *J. Mol. Biol.*, **322**, 291-309.

13. Zhang, L., Shapiro, R. and Broyde, S. (2005) Molecular dynamics of a food carcinogen–DNA adduct in a replicative DNA polymerase suggest hindered nucleotide incorporation and extension. *Chem. Res. Toxicol.*, **18**, 1347-1363.

14. Allnér, O., Nilsson, L. and Villa, A. (2012) Magnesium ion–water coordination and exchange in biomolecular simulations. *J. Chem. Theory Comput.*, **8**, 1493-1502.

15. Joung, I.S. and Cheatham, T.E. (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*, **112**, 9020-9041.