Supporting Information

Environmental Effects on Guanine-Thymine Mispair Tautomerization Explored with Quantum Mechanical/Molecular Mechanical Free Energy Simulations

Pengfei Li[1], Atul Rangadurai[2], Hashim M. Al-Hashimi[2], and Sharon Hammes-Schiffer[1]*

[1]Department of Chemistry, Yale University, 225 Prospect Street,
New Haven, CT 06520
[2]Department of Biochemistry, Duke University, Durham, NC, 27710

*Corresponding author email: sharon.hammes-schiffer@yale.edu

TABLE OF CONTENTS

Simulation details of G-T mispair in aqueous solution

A. Structure Preparation

The G-T nucleoside pair was generated based on residues 3 and 14 of PDB entry 1D92.[1] The OL15 force field[2] was used to model the nucleoside pair. Hydrogen atoms were added using the LEaP program in AmberTools.[3] A cuboid water box was employed to solvate the system with the distance from the box boundary to the nucleoside pair at least 25 Å, giving a box size of ~64 Å × 70 Å × 69 Å. The TIP3P water model[4] was used to model the solvent. Na^+ and Cl^- ions were added to the system to provide a salt concentration of ~0.15 M. The van der Waals (VDW) parameters of Na^+ and Cl^- were obtained from the hydration free energy (HFE) parameter set developed by Li et al.[5]

B. Structure Equilibration

The following steps were carried out to equilibrate the system.

(1) 10000 steps of minimization using the steepest descent algorithm, with 200 kcal/mol•Å^{-2} restraints on the heavy atoms of the nucleoside pair.

(2) 500 ps NVT molecular dynamics (MD) at 300 K, with 200 kcal/mol•Å^{-2} restraints on the heavy atoms of the nucleoside pair.

(3) 500 ps NPT MD at 300 K and 1 atm, with 200 kcal/mol•Å^{-2} restraints on the heavy atoms of the nucleoside pair.

(4) 10000 steps of minimization using the steepest descent algorithm followed by 10000 steps of minimization using the conjugate gradient algorithm, with 100 kcal/mol•Å^{-2} restraints on the heavy atoms of the nucleoside pair during these 20000 steps of minimization.

(5) 10000 steps of minimization using the steepest descent algorithm followed by 10000 steps of minimization using the conjugate gradient algorithm, with 50 kcal/mol•Å^{-2} restraints on the heavy atoms of the nucleoside pair during these 20000 steps of minimization.

(6) 10000 steps of minimization using the steepest descent algorithm followed by 10000 steps of minimization using the conjugate gradient algorithm, with 20 kcal/mol•Å^{-2} restraints on the heavy atoms of the nucleoside pair during these 20000 steps of minimization.

(7) 360 ps NVT heating using the following procedure: (a) 50 ps heating from 0 to 50 K followed by 50 ps equilibration at 50 K; (b) 50 ps heating from 50 to 100 K followed by 50 ps equilibration at 100 K; (c) 50 ps heating from 100 to 150 K followed by 50 ps equilibration at 150 K; (d) 50 ps heating from 150 to 200 K followed by 50 ps equilibration at 200 K; (e) 50 ps heating from 200 to 250 K followed by 50 ps equilibration at 250 K; (f) 50 ps heating from 250 to 300 K followed by 50 ps equilibration at 300 K. During this heating procedure 20 kcal/mol•Å^{-2} restraints were applied to the heavy atoms of the nucleoside pair to keep its planarity in the aqueous solution.

(8) 2 ns NPT equilibration at 300 K and 1 atm, with 20 kcal/mol•Å^{-2} restraints on the heavy atoms of the nucleoside pair to keep its planarity in the aqueous solution.

Periodic boundary conditions (PBCs) were used for all of the simulations. The particle mesh Ewald (PME) method[6] was employed to account for the long-range electrostatic interactions with the nonbonded cut-off set as 10 Å. The Langevin thermostat was utilized to control the temperature using a collision frequency of 2 ps^{-1}. The Berendsen barostat was used to control the pressure in the NPT simulations. The SHAKE algorithm[7] was employed to constrain the covalent bonds involving hydrogen atoms, while a "three-point" SHAKE algorithm was used for the water

molecules.[8] The time step was 1 fs for the MD simulations. The restraint strength mentioned in this document is for the equation $U=k(r-r_0)^2$, where k is the restraint strength.

C. Generation of the Initial Strings

For all of the systems investigated in the present work, the following protocol was used for generating the initial strings. The QM/MM minimizations were carried out using the sander program in AMBER[3] and the Q-Chem/AMBER QM/MM interface.[9] The nonbonded cut-off was set as 999 Å in these minimizations so that no cut-off was applied. The electrostatic embedding scheme was used in these QM/MM calculations.

The following procedure was employed to generate the initial strings for all four systems.

(1) The structure after the Structure Equilibration was truncated by keeping the water molecules and ions containing at least one atom that is within 23 Å of the solute molecule.
(2) The structure was minimized using the quantum mechanical/molecular mechanical (QM/MM) method. The system was minimized by allowing only the QM region to move in a frozen aqueous environment. These minimizations were performed to generate the initial strings describing the tautomerization process of the G-T mispair (Figure S2).
(3) The structure was minimized by 50 steps of minimization using the steepest descent algorithm followed by 50 steps of minimization using the conjugate gradient algorithm. During these minimizations, restraints obtained from the optimized wG-T geometry for the G-T base pair in gas phase (Table S1) were applied to the reaction coordinates with restraint strengths of 100 kcal/mol•Å^{-2}. Subsequently, the system was minimized without any restraints by 50 steps of minimization using the steepest descent algorithm and then 50 steps of minimization using the conjugate gradient algorithm. The final configuration was treated as the wG-T structure in the structure interpolation.
(4) The structure was minimized by 50 steps of minimization using the steepest descent algorithm followed by 50 steps of minimization using the conjugate gradient algorithm. During these minimizations, restraints obtained from the optimized transition state geometry, denoted TS(wG-T↔G-T*), for the G-T base pair in gas phase (Table S1) were applied to the reaction coordinates with restraint strengths of 100 kcal/mol•Å^{-2}. The final configuration was treated as the wG-T↔G-T* structure in the structure interpolation.
(5) The structure was minimized by 50 steps of minimization using the steepest descent algorithm followed by 50 steps of minimization using the conjugate gradient algorithm. During these minimizations, restraints obtained from the optimized G-T* geometry for the G-T base pair in gas phase (Table S1) were applied to the reaction coordinates with restraint strengths of 100 kcal/mol•Å^{-2}. Subsequently, the system was minimized without any restraints by 50 steps of minimization using the steepest descent algorithm and then 50 steps of minimization using the conjugate gradient algorithm. The final configuration was treated as the G-T* structure in the structure interpolation.
(6) The structure was minimized by 50 steps of minimization using the steepest descent algorithm followed by 50 steps of minimization using the conjugate gradient algorithm. During these minimizations, restraints obtained from the optimized TS(G-T*↔G*-T) geometry for the G-T base pair in gas phase (Table S1) were applied to the reaction coordinates with restraint strengths of 100 kcal/mol•Å^{-2}. The final configuration was treated as the G-T*↔G*-T structure in the structure interpolation.

(7) The structure was minimized by 50 steps of minimization using the steepest descent algorithm followed by 50 steps of minimization using the conjugate gradient algorithm. During these minimizations, restraints obtained from the optimized G*-T geometry for G-T base pair in gas phase (Table S1) were applied to the reaction coordinates with restraint strengths of 100 kcal/mol•Å^{-2}. Subsequently, the system was minimized without any restraints by 50 steps of minimization using the steepest descent algorithm and then 50 steps of minimization using the conjugate gradient algorithm. The final configuration was treated as the G*-T structure in the structure interpolation.

(8) Finally, two initial strings were generated for the QM/MM finite temperature string simulations. To generate the wG-T→G-T* string, a quadratic interpolation was performed based on the QM/MM minimized wG-T, wG-T↔G-T*, and G-T* structures. A total of 13 reaction coordinates (Figure 3) were used, and 25 images were generated. A set of restraints for the reaction coordinates was generated based on these images. To generate the G-T*→G*-T string, a quadratic interpolation was performed based on the QM/MM minimized G-T*, G-T*↔G*-T, and G*-T structures. The same 13 reaction coordinates were used, and 15 images were generated. Fewer images were used for this string because of the smaller structural changes associated with the G-T*→G*-T tautomerization compared to the wG-T→G-T* tautomerization. A set of restraints for the reaction coordinates were generated based on these images.

D. QM/MM String Simulations

For all of the systems investigated in the present work, the following protocol was used for the finite-temperature string simulations. The CHARMM software package[10] was used to perform the classical MD simulations, and the Q-Chem/CHARMM interface[11] was used for the QM/MM simulations. The Langevin thermostat was used to control the temperature at 300 K, and the nonbonded cut-off was set to 999 Å so that no cut-off was applied. The electrostatic embedding scheme was used in these QM/MM simulations with the point charges of the link host groups set to zero. This scheme was also used to obtain the results given in Tables S6 and S8, i.e., the point charges of the link host groups were also treated as zero in these electrostatic interaction analyses.

The following procedure was used for the QM/MM string simulations for all four systems.

(1) The finite temperature string method with umbrella sampling[12-13] was used to carry out the free energy simulations. Each image of the string was equilibrated for 10 ps using classical MD simulations. These equilibrations were performed allowing only the inner MM region, which was defined as residues (including water molecules and ions) with at least one atom within 18 Å of the N1 atom of the guanine base, to move, while the QM region and outer MM region were kept frozen.

(2) 100 fs QM/MM equilibration was performed for each image while applying the restraints generated from the quadratic interpolation described above. The restraint strength was 50 kcal/mol•Å^{-2} for each reaction coordinate. Subsequently, an updated string was fitted based on the average values of the reaction coordinates for each image during the equilibration. The images were redistributed evenly along this updated string. This updated string was used in the first iteration of the QM/MM string simulations. For each image, the final configuration

of the QM/MM equilibration was used as the initial configuration in the first iteration of the QM/MM string simulations.

(3) The subsequent iterations of the QM/MM string simulations were performed in a similar manner. Specifically, an updated string, which was generated based on the previous iteration, was used in the current iteration, and the final configurations of the previous iteration were used as the initial configurations of the current iteration. To enhance sampling of the high-energy regions, larger restraints (with strengths of 75 or 100 kcal/mol•Å^{-2}) were applied for some reaction coordinates in certain iterations. For the wG-T→G-T* string, 24 iterations were used in the analysis because the convergence criteria were satisfied at this point, and oscillations were observed after this iteration, presumably caused by the flexibility of the nucleoside pair in aqueous solution. For the G-T*→G*-T string, the string simulations were carried out for 25 iterations and also satisfied the convergence criteria. The convergence criteria were: (1) the root-mean square deviation (RMSD) between each reaction coordinate of the final string and the same reaction coordinate averaged over the previous five iterations is within 0.1 Å, and (2) the calculated free energy barrier along the mean free energy path (MFEP) is within 0.5 kcal/mol for the five most recent datasets (i.e. data from iterations 1–21, 1–22, 1–23, 1–24, and 1–25 for the G-T*→G*-T string). The string convergence is illustrated in Figures S3A and S4A. For each data set, the string that was used in the last iteration was considered to be the MFEP. The binless weighted histogram analysis method (WHAM) approach[14] was used to generate the free energy surfaces using the data from all iterations for each string. The statistical uncertainties of the free energies were analyzed using the bootstrapping approach, as shown in Figure S5.

Simulation details of G-T mispair in the A-DNA duplex

A. Structure Preparation

The structure from PDB entry 1D92[1] was used to model the system. It has two G-T mispairs that are composed of residues 3 and 14 and residues 6 and 11. Only the first G-T mispair was retained, and residue 6 was mutated from T to C. The same protocol was used to prepare the DNA duplex system as was used to prepare the nucleoside pair system except that for the DNA duplex system, Mg^{2+} ions were added to neutralize the system before adding the Na^{+} and Cl^{-} ions to a final concentration of 0.15 M. For this system, the solvent box was ~82 Å × 81 Å × 90 Å. The VDW parameters of the Mg^{2+} ions were obtained from Allner et al.[15]

B. Structure Equilibration

The following steps were carried out to equilibrate the system.

(1) 10000 steps of minimization using the steepest descent algorithm, with 200 kcal/mol•Å^{-2} restraints on the backbone heavy atoms (i.e., P, O3', O5', C3', C4', and C5').
(2) 500 ps NVT MD at 300 K, with 200 kcal/mol•Å^{-2} restraints on the backbone heavy atoms.
(3) 500 ps NPT MD at 300 K and 1 atm, with 200 kcal/mol•Å^{-2} restraints on the backbone heavy atoms.
(4) 10000 steps of minimization using the steepest descent algorithm followed by 10000 steps of minimization using the conjugate gradient algorithm, with 100 kcal/mol•Å^{-2} restraints on the backbone heavy atoms during these 20000 steps of minimization.
(5) 10000 steps of minimization using the steepest descent algorithm followed by 10000 steps of minimization using the conjugate gradient algorithm, with 50 kcal/mol•Å^{-2} restraints on the backbone heavy atoms during these 20000 steps of minimization.
(6) 10000 steps of minimization using the steepest descent algorithm followed by 10000 steps of minimization using the conjugate gradient algorithm, with 20 kcal/mol•Å^{-2} restraints on the backbone heavy atoms during these 20000 steps of minimization.
(7) 10000 steps of minimization using the steepest descent algorithm followed by 10000 steps of minimization using the conjugate gradient algorithm, with 10 kcal/mol•Å^{-2} restraints on the backbone heavy atoms during these 20000 steps of minimization.
(8) 10000 steps of minimization using the steepest descent algorithm followed by 10000 steps of minimization using the conjugate gradient algorithm.
(9) 360 ps NVT heating using the following procedure: (a) 50 ps heating from 0 to 50 K followed by 50 ps equilibration at 50 K; (b) 50 ps heating from 50 to 100 K followed by 50 ps equilibration at 100 K; (c) 50 ps heating from 100 to 150 K followed by 50 ps equilibration at 150 K; (d) 50 ps heating from 150 to 200 K followed by 50 ps equilibration at 200 K; (e) 50 ps heating from 200 to 250 K followed by 50 ps equilibration at 250 K; (f) 50 ps heating from 250 to 300 K followed by 50 ps equilibration at 300 K.
(10) 2 ns NPT equilibration at 300 K and 1 atm.

In order to keep the wobble structure of the G-T mispair, a set of restraints was applied during the steps listed above. Two distances were restrained: (G)O6-(T)N3 with an equilibrium distance of 2.76 Å, and (G)N1-(T)O2 with an equilibrium distance of 2.74 Å. These two values were obtained from the crystal structure 1D92.[1] The restraint strength was 50 kcal/mol•Å^{-2} for each of these two restraints.

The same protocol was used for the generation of the initial strings and for the QM/MM string simulations for the G-T mispair in the A-DNA duplex as was used for the G-T mispair in aqueous solution. 25 iterations were propagated for both the wG-T→G-T* and G-T*→G*-T strings. The string convergence is illustrated in Figures S3B and S4B. The statistical uncertainties of the free energies were analyzed using the bootstrapping approach, as shown in Figure S5.

Simulation details of G-T mispair in the B-DNA duplex

The nucleic acid builder (NAB) program in AmberTools[3] was used to generate a B-DNA duplex in the Arnott form with the sequence of the first chain as GGGGCCCC (from 5′-terminal to 3′-terminal). Subsequently, the complementary base of the third G was mutated from C to T, producing the identical DNA sequence as the A-DNA duplex investigated herein. For this system, the solvent box was ~76 Å × 77 Å × 87 Å. The protocol to build and simulate the B-DNA duplex were the same as those used to build and simulate the A-DNA duplex except that in the Structure Equilibration step, an additional 20 ns NVT equilibration was performed to further equilibrate the system after the final 2 ns NPT equilibration while still retaining the same wobble restraints. This additional equilibration was performed because the initial geometry of the DNA duplex was built by a computer program and was not obtained from a crystal structure. 25 iterations were propagated for both the wG-T→G-T* and G-T*→G*-T strings. The string convergence is illustrated in Figures S3C and S4C. The statistical uncertainties of the free energies were analyzed using the bootstrapping approach, as shown in Figure S5.

Simulation details of G-T mispair in the DNA polymerase λ variant

Chains A, C, D, and H from PDB entry 3PML[16] were used to build the protein-DNA-GTP complex, which contains a protein monomer of the DNA polymerase λ variant. This structure has the G-T mispair in the dGMPCPP:T WC-like form, but the free energy simulations were initiated in the wobble form for consistency with the other simulations. dG1 in Chain D was modified to be a 5′ terminal residue. The H++ webserver[17] was used to add hydrogen atoms to the protein-DNA complex. Subsequently, the protonation state of Asp429 (residue number in the original PDB file) in Chain A was corrected by considering its coordination to a metal ion. The GMPCPP residue in Chain A was changed to GTP by renaming it from "1GC" to "GTP" and changing its "C3A" atom to "O3A". The partial charges of GTP were obtained from previous work in our group,[18] and the phosphate parameters of GTP were obtained from Meagher et al.[19] A cuboid TIP3P water box was used to solvate the system with the distance from the protein-DNA-GTP complex to the box boundary at least 14 Å, giving a box size of ~91 Å × 96 Å × 95 Å. Na^+ and Cl^- ions were added to neutralize the system and then provide a salt concentration of ~0.15 M. The VDW parameters for Mg^{2+} from Allnér et al.[15] were used to model the Mg^{2+} ions. The VDW parameters for Na^+ and Cl^- ions were obtained from Li et al.[5]

The same protocol was used for the structure equilibration, generation of the initial strings, and the QM/MM string simulations of the DNA polymerase λ variant as was used for the A-DNA duplex except the following:

(1) For the Structure Equilibration procedure, in addition to the restraints on the backbone P, O3', O5', C3', C4', and C5' atoms of DNA and GTP, as well as the restraints maintaining the G-T mispair wobble form, restraints on the backbone N, CA, and C atoms of the protein were also applied to the system in the first seven steps.

(2) For the Generation of the Initial Strings procedure, the structure obtained after the Structure Equilibration was truncated by keeping the water molecules and ions with at least one atom within 12 Å of the protein-DNA-GTP complex.

For the DNA polymerase λ variant, 25 iterations were propagated for both the wG-T→G-T* and G-T*→G*-T strings. The string convergence is illustrated in Figures S3D and S4D. The statistical uncertainties of the free energies were analyzed using the bootstrapping approach, as shown in Figure S5.

Optimized geometries of G-T base pair in gas phase

To obtain reasonable structures for the procedure used to generate the initial strings, we reproduced the results from a previous publication.[20] The structures of the G-T mispair in the wG-T, TS(wG-T↔G-T*), G-T*, TS(G-T*↔G*-T), and G*-T states were optimized at the M06/6-311++G** level of theory, as in the previous work.[20] A frequency analysis was performed to ensure that a minimum was found for the wG-T, G-T*, and G*-T states, and only one imaginary frequency was found for the TS(wG-T↔G-T*) and TS(G-T*↔G*-T) states. The relative energy values were reproduced for these states. These calculations were performed using the Gaussian09 program.[21] The Cartesian coordinates of the optimized structures are given below, and the corresponding values of the 13 reaction coordinates for these structures are given in Table S1. These values were used as restraints in the geometry optimizations performed prior to the quadratic interpolation procedure implemented to generate the initial strings. These gas phase optimized structures were not used in the production level QM/MM string simulations.

wG-T:

7	3.439095	1.666637	-0.051170
6	2.181181	1.114384	-0.042129
8	1.183852	1.825342	-0.062447
7	2.161358	-0.246778	-0.008713
6	3.278311	-1.101675	0.018218
8	3.140286	-2.299648	0.047940
6	4.576564	-0.421815	0.007179
6	5.797966	-1.270601	0.033877
6	4.595713	0.918332	-0.025979
1	1.225832	-0.696295	-0.004489
1	6.705244	-0.660930	0.019879
1	5.811846	-1.903079	0.927277
1	5.816497	-1.950774	-0.823611
1	5.519709	1.487833	-0.035629
7	-5.070807	-0.674095	-0.016580
6	-4.757243	-2.019476	-0.035653
7	-3.479957	-2.240669	-0.030637
6	-2.912035	-0.988572	-0.006121
6	-1.542194	-0.586946	0.006349
8	-0.524364	-1.261040	-0.006019
7	-1.442616	0.818041	0.036465
6	-2.481019	1.697188	0.041195
7	-2.147308	3.013587	0.102321
7	-3.739557	1.341765	0.020596
6	-3.886185	0.002088	0.002803
1	-5.530281	-2.776281	-0.052497
1	-0.480044	1.181713	0.037283
1	-2.891944	3.669887	-0.070443
1	-1.203649	3.298011	-0.113153
1	3.467112	2.675355	-0.070590

1 -5.986336 -0.249537 -0.014219

TS(wG-T↔G-T*):

7 3.492819 1.607996 -0.000096
6 2.149857 1.278830 -0.000231
8 1.307975 2.201749 -0.000394
7 1.811657 -0.010846 -0.000189
6 2.743089 -1.021115 0.000006
8 2.365013 -2.189606 0.000083
6 4.165892 -0.662112 0.000109
6 5.177549 -1.752549 0.000284
6 4.473432 0.644901 0.000053
1 0.404416 -1.298526 -0.000171
1 6.196100 -1.352524 0.000332
1 5.052917 -2.398876 0.875433
1 5.053058 -2.399025 -0.874775
1 5.497007 1.008514 0.000128
7 -4.884441 -0.300490 0.000078
6 -4.766198 -1.680913 -0.000042
7 -3.539798 -2.091321 -0.000118
6 -2.797191 -0.932258 -0.000046
6 -1.421455 -0.700755 -0.000084
8 -0.541616 -1.633147 -0.000199
7 -1.063699 0.609632 0.000007
6 -1.956872 1.655808 0.000156
7 -1.411714 2.873832 0.000297
7 -3.264001 1.484718 0.000186
6 -3.618599 0.200570 0.000077
1 -5.641789 -2.316490 -0.000065
1 -0.054394 0.848695 -0.000108
1 -2.030900 3.666799 0.000266
1 -0.394906 2.970931 0.000183
1 3.715933 2.592059 -0.000156
1 -5.731210 0.249397 0.000155

G-T*:

7 1.012048 0.427790 -0.000386
6 1.747998 1.578756 -0.000260
7 1.051148 2.728695 -0.000499
7 3.062864 1.592931 -0.000002
6 3.585521 0.356537 0.000095
6 2.938187 -0.875717 -0.000017
6 1.516772 -0.875256 -0.000248
8 0.739585 -1.830626 -0.000321

7 3.843457 -1.913085 0.000169
6 5.001727 -1.332706 0.000382
7 4.914892 0.047079 0.000352
1 1.579067 3.585159 0.000062
1 0.030413 2.747064 -0.000172
1 -0.019583 0.479396 -0.000433
1 5.960544 -1.834380 0.000575
7 -3.877760 1.433094 0.000294
6 -2.484813 1.473462 0.000148
8 -1.906032 2.549674 0.000207
7 -1.851362 0.270895 -0.000057
6 -2.520027 -0.861013 -0.000072
8 -1.883423 -1.995146 -0.000233
6 -3.956051 -0.929909 0.000085
6 -4.662789 -2.241107 0.000045
6 -4.581449 0.267629 0.000262
1 -4.391151 -2.836504 -0.877334
1 -4.391060 -2.836616 0.877320
1 -5.746869 -2.099348 0.000112
1 -5.663357 0.356417 0.000390
1 -0.880402 -1.875359 -0.000295
1 5.669162 0.717291 0.000487
1 -4.345277 2.329825 0.000444

TS(G-T*↔G*-T):
7 -0.853675 0.332231 -0.000175
6 -1.546858 1.519618 0.000202
7 -0.823470 2.643687 0.000475
7 -2.869900 1.620774 0.000351
6 -3.480266 0.442154 0.000128
6 -2.896638 -0.823669 -0.000143
6 -1.499509 -0.859853 -0.000261
8 -0.856855 -1.974588 -0.000505
7 -3.852662 -1.817889 -0.000242
6 -4.976064 -1.175860 -0.000051
7 -4.823046 0.201092 0.000200
1 -1.343337 3.505828 0.000483
1 0.203889 2.660673 0.000184
1 0.358707 0.355314 -0.000133
1 -5.959303 -1.627992 -0.000053
7 3.855192 1.374674 -0.000239
6 2.474369 1.489207 -0.000290
8 1.969658 2.608008 -0.000425
7 1.767671 0.332034 -0.000159
6 2.362136 -0.894778 0.000127
8 1.672295 -1.928457 0.000366

6 3.815316 -0.987943 0.000146
6 4.455071 -2.331777 0.000396
6 4.500310 0.168729 -0.000022
1 4.149548 -2.912604 0.876631
1 4.149585 -2.912912 -0.875648
1 5.545512 -2.247566 0.000400
1 5.585434 0.206235 0.000005
1 0.163931 -1.901100 -0.000174
1 -5.545690 0.905444 0.000265
1 4.362972 2.247804 -0.000312

G*-T:
7 0.948023 0.316085 -0.024590
6 1.643642 1.483983 -0.033347
7 0.915619 2.621588 -0.068804
7 2.969777 1.618379 -0.018417
6 3.592183 0.447244 0.000686
6 3.014862 -0.821442 0.007890
6 1.615201 -0.837881 -0.005191
8 0.975630 -1.982760 0.002230
7 3.977397 -1.810467 0.027691
6 5.098340 -1.161671 0.032532
7 4.939288 0.212919 0.017322
1 1.420813 3.490699 -0.030876
1 -0.098732 2.627053 -0.022165
1 6.083721 -1.609294 0.047032
7 -3.981709 1.386396 0.034649
6 -2.609007 1.527187 0.027875
8 -2.066859 2.612971 0.041477
7 -1.935352 0.335015 0.004603
6 -2.488318 -0.933355 -0.011337
8 -1.766569 -1.919685 -0.030817
6 -3.942381 -0.995715 -0.002884
6 -4.589772 -2.335462 -0.019864
6 -4.615663 0.168033 0.019484
1 -0.889881 0.379381 -0.005894
1 -4.289440 -2.902665 -0.906614
1 -4.280166 -2.929477 0.845891
1 -5.679148 -2.247489 -0.012641
1 -5.700425 0.205471 0.026976
1 -0.003739 -1.873014 -0.014001
1 -4.505093 2.250242 0.051041
1 5.658667 0.919958 0.015901

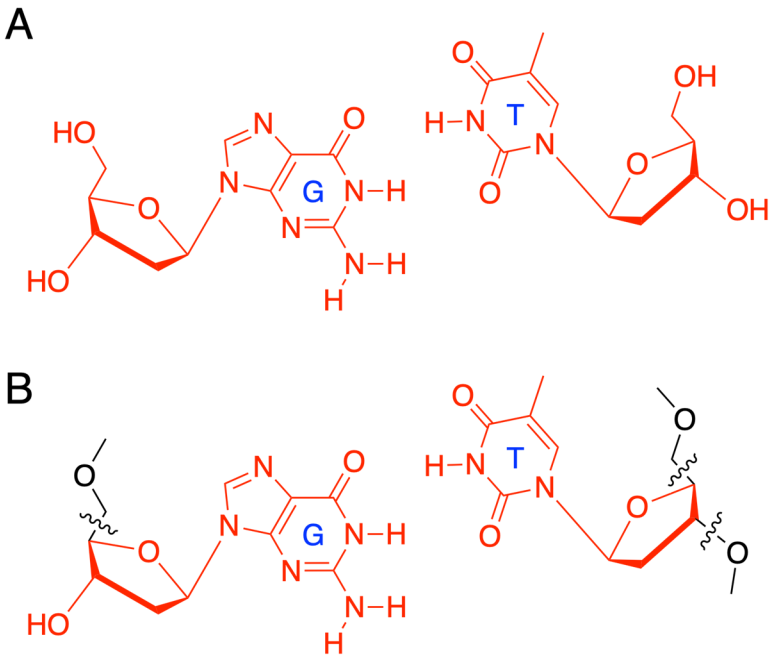

Figure S1. Illustration of the QM regions in the QM/MM simulations of the G-T mispair in (A) aqueous solution, and (B) the DNA polymerase λ variant. The atoms depicted in red are in the QM region, and the cut bonds are indicated by wavy curves. Illustration of the QM region for the G-T mispair in the DNA duplexes is shown in Figure 2 of the main paper.

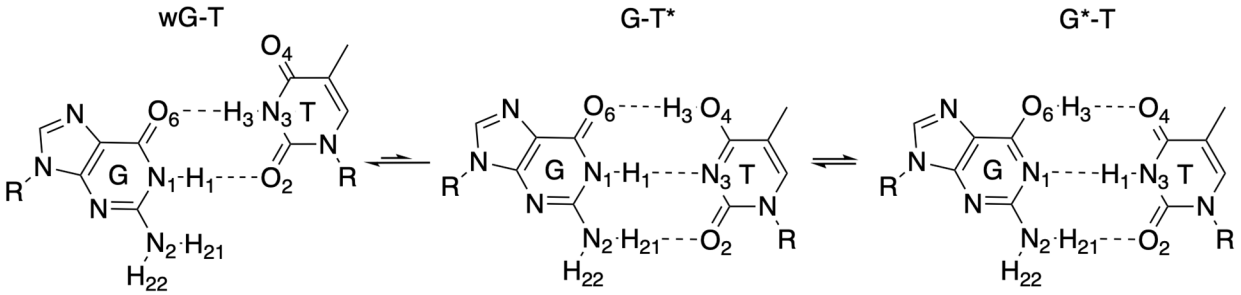

Figure S2. Schematic depiction of G-T mispair tautomerization. Here wG-T denotes the wobble structure, and G* and T* represent G and T, respectively, in their enol forms. The N, O, and H atoms used to describe the reaction coordinates in Table S1 are marked with numbers.

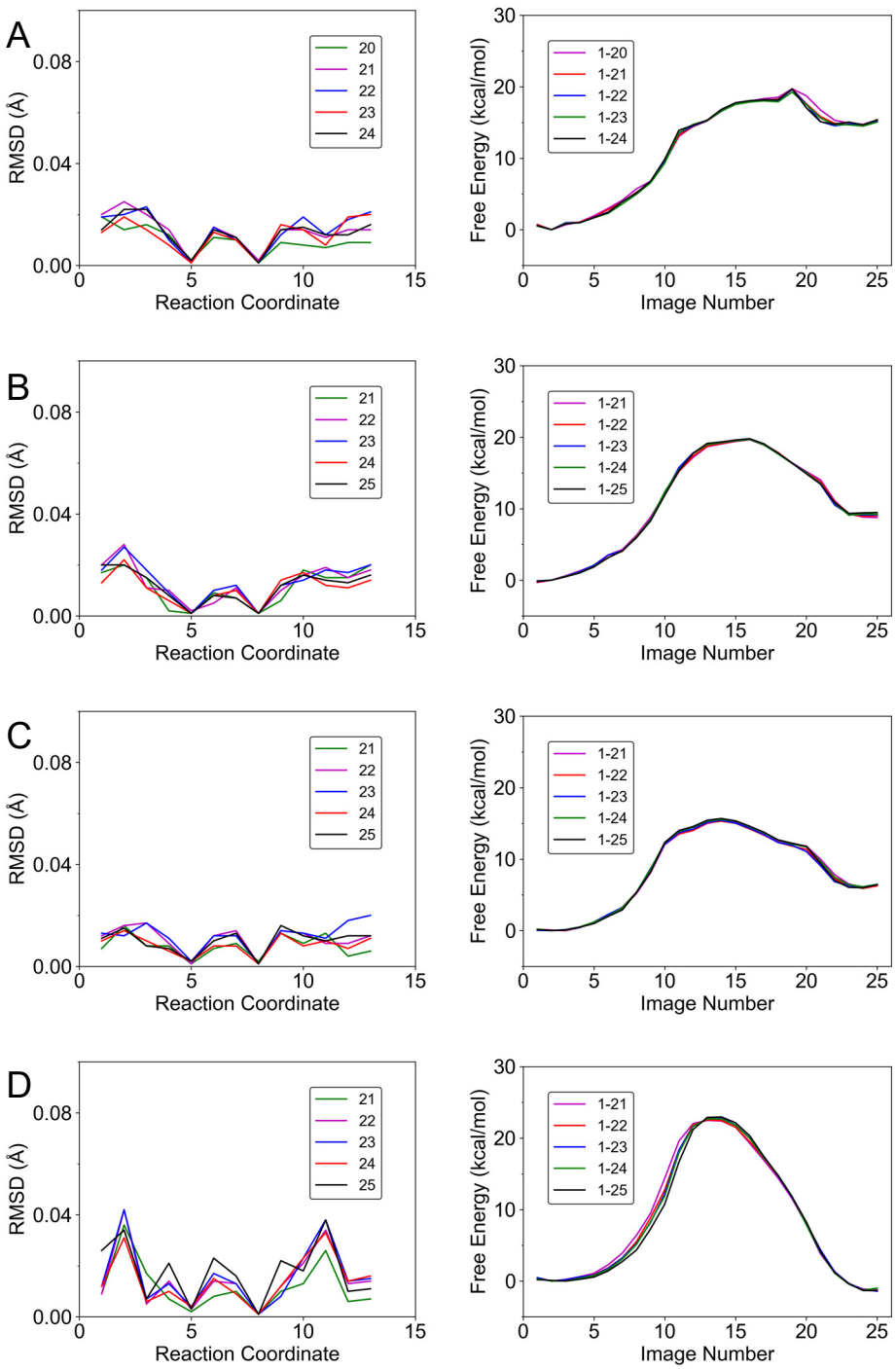

Figure S3. Illustration of the convergence of the wG-T→G-T* tautomerization string in (A) aqueous solution, (B) the A-DNA duplex, (C) the B-DNA duplex, and (D) the DNA polymerase λ variant. The left panels depict the RMSD values for each reaction coordinate in each of the five most recent iterations when compared to the average values obtained from the previous five iterations. The right panels depict the free energy profiles along the MFEPs as obtained from the five most recent iterations.

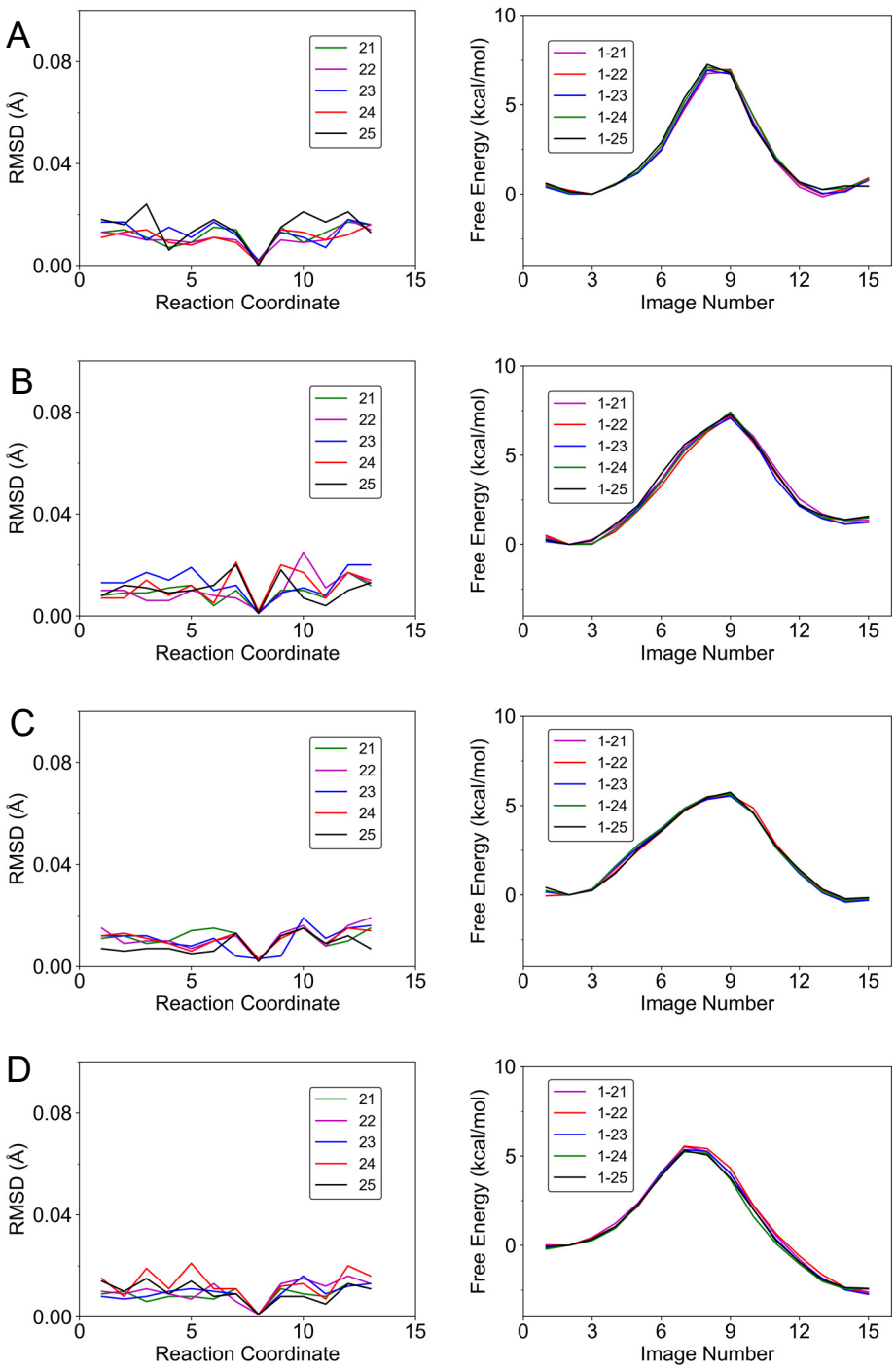

Figure S4. Illustration of the convergence of the G-T*→G*-T tautomerization string in (A) aqueous solution, (B) the A-DNA duplex, (C) the B-DNA duplex, and (D) the DNA polymerase λ variant. The left panels depict the RMSD values for each reaction coordinate in each of the five most recent iterations when compared to the average values obtained from the previous five iterations. The right panels depict the free energy profiles along the MFEPs as obtained from the five most recent iterations.

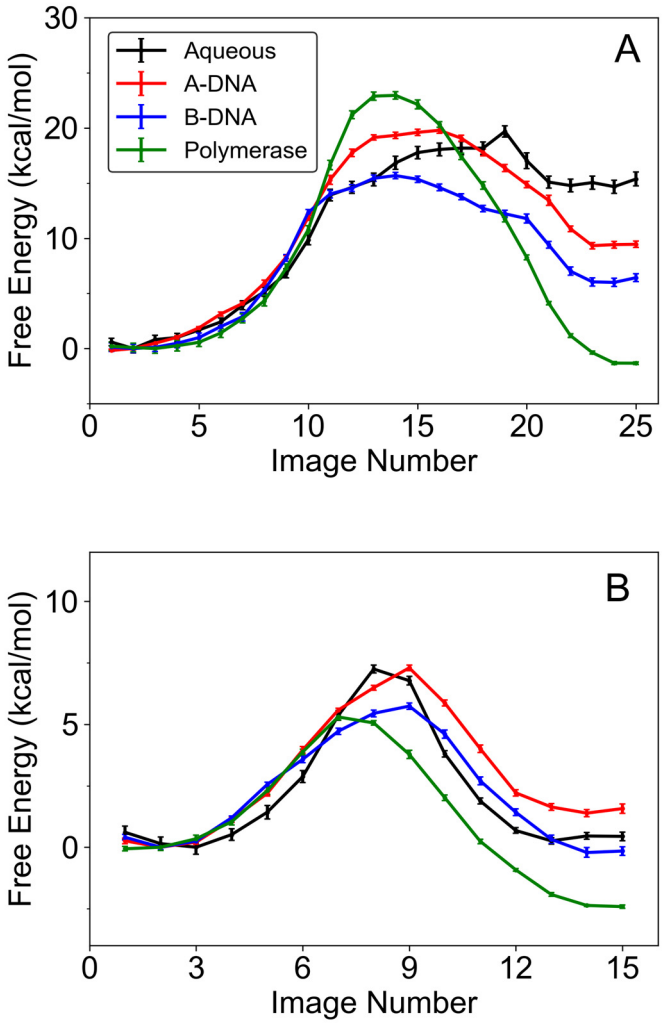

Figure S5. Free energies along the MFEPs with statistical error bars for the (A) wG-T→G-T* and (B) G*-T→G-T* tautomerization in aqueous solution, the A-DNA duplex, the B-DNA duplex, and the DNA polymerase λ variant. These error bars correspond to the statistical errors and were obtained using the bootstrapping error analysis[22] with ten "fake" data sets. Note that these statistical error bars do not account for errors arising from the level of theory used to generate the potential energy surface and the limited conformational sampling, which introduce greater uncertainties.

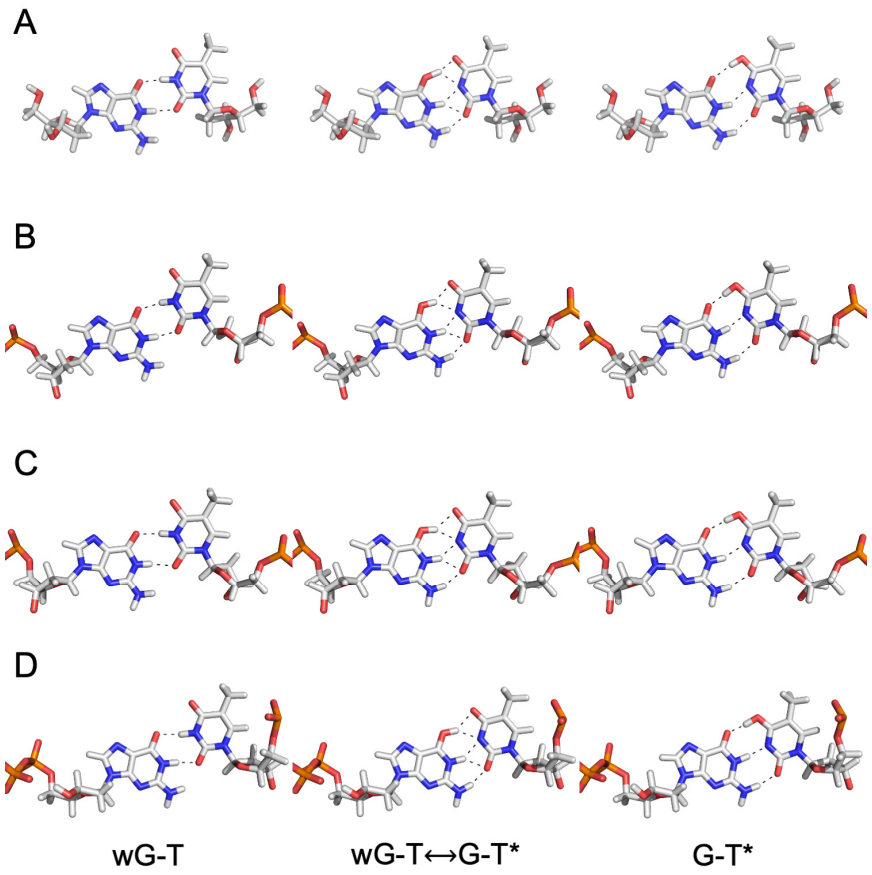

Figure S6. Representative configurations along the wG-T→G-T* tautomerization process in (A) aqueous solution, (B) the A-DNA duplex, (C) the B-DNA duplex, and (D) the DNA polymearse λ variant. These structures are the last configurations of the last iteration of each string simulation for the images corresponding to the reactant state, the "transition state" (i.e., top of the free energy barrier), and the product state in Figure 4. Specifically, these configurations were obtained from images 2, 19, and 24 of the string in aqueous solution, images 1, 16, and 23 of the string in the A-DNA duplex, images 2, 14, and 24 of the string in the B-DNA duplex, and images 3, 14, and 25 of the string in the DNA polymerase λ variant.

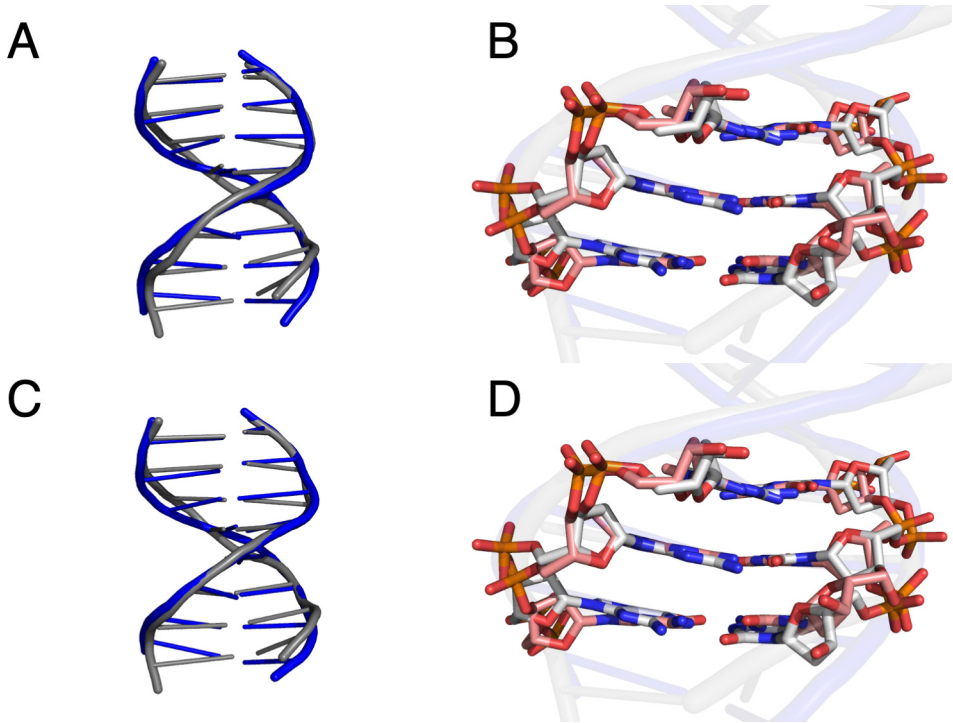

Figure S7. (A) Comparison of the helix structures of A-DNA (blue) and B-DNA (gray) duplexes when the G-T mispair is in the wG-T state. (B) Comparison of the base stacking in the A-DNA (white/gray colors) and B-DNA (salmon color) duplexes when the G-T mispair is in the wG-T state, with the G-T mispair flanked by the neighboring base pairs. (C) Comparison of the helix structures of A-DNA (blue) and B-DNA (gray) duplexes when the G-T mispair is in the G-T* state. (D) Comparison of the base packing in the A-DNA (white/gray colors) and B-DNA (salmon color) duplexes when the G-T mispair is in the G-T* state, with the G-T mispair flanked by the neighbouring base pairs. These structures correspond to the structures shown in Figure S6. The A-DNA and B-DNA were superimposed by aligning the backbone heavy atoms (P, O3, O5, C3', C4', and C5').

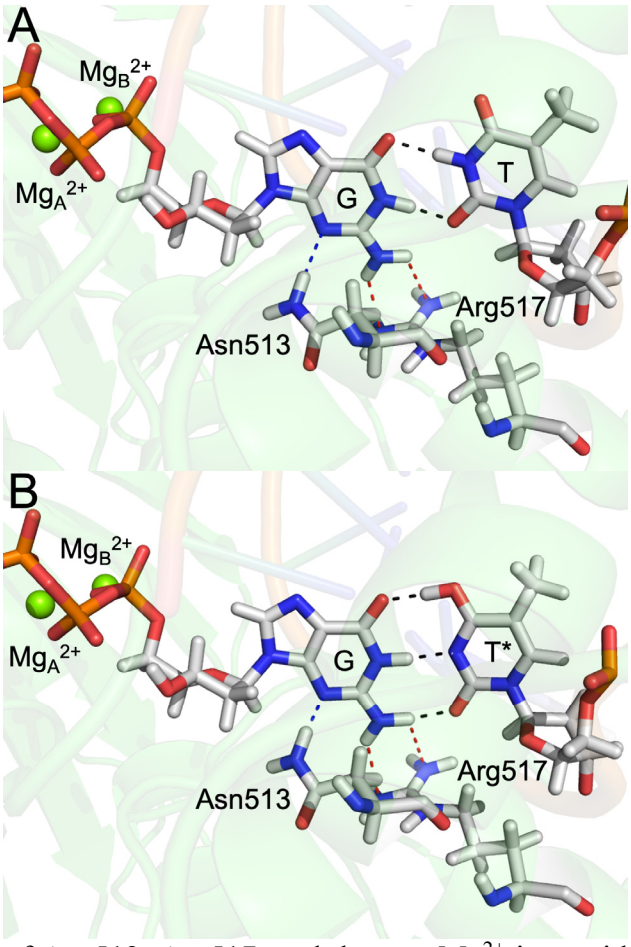

Figure S8. Interactions of Asn513, Arg517, and the two Mg^{2+} ions with the G-T mispair in the DNA polymerase λ variant in (A) the wG-T state, where the structure shown is the last configuration of the last iteration for image 3, and (B) the G-T* state, where the structure shown is the last configuration of the last iteration for image 25 of the string for wG-T→G-T* tautomerization. The hydrogen bonds within the G-T mispair are indicated with black dashed lines, the hydrogen bond between Asn513 and G is indicated with a blue dashed line, and the hydrogen bonds between Arg517 and G are indicated with red dashed lines.

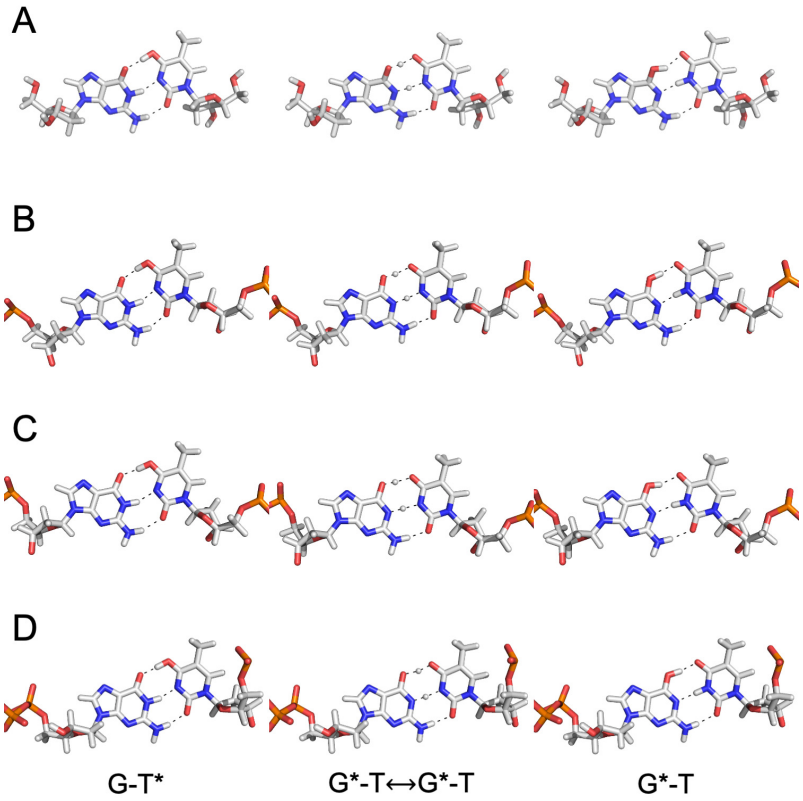

Figure S9. Representative configurations along the G-T*→G*-T tautomerization process in (A) aqueous solution, (B) the A-DNA duplex, (C) the B-DNA duplex, and (D) the DNA polymearse λ variant. These structures are the last configurations of the last iteration of each string simulation for the images corresponding to the reactant state, the "transition state" (i.e., top of the free energy barrier), and the product state in Figure 7A. Specifically, these configurations were obtained from images 3, 8, and 13 of the string in aqueous solution, images 2, 9, and 14 of the string in the A-DNA duplex, images 2, 9, and 14 of the string in the B-DNA duplex, and images 1, 7, and 15 of the string in the DNA polymerase λ variant.

Table S1. Reaction Coordinates of the QM Optimized Geometries in the Gas Phase.[a]

RC	Atom pairs	wG-T	TS (wG-T↔G-T*)	G-T*	TS (G-T*↔G*-T)	G*-T
R1	G(O6)-T(O4)	3.81	2.96	2.63	2.53	2.74
R2	G(O6)-T(H3)	1.84	1.00	1.62	1.02	0.99
R3	T(H3)-T(O4)	2.50	2.15	1.01	1.51	1.76
R4	G(N1)-T(N3)	3.76	2.94	2.87	2.62	2.88
R5	G(N1)-G(H1)	1.03	1.04	1.03	1.21	1.84
R6	G(H1)-T(N3)	3.00	2.05	1.84	1.41	1.05
R7	G(N2)-T(O2)	3.54	2.80	2.96	2.79	2.98
R8	G(N2)-G(H21)	1.01	1.02	1.02	1.03	1.02
R9	G(H21)-T(O2)	2.81	1.87	1.95	1.77	1.97
R10	G(O6)-T(N3)	2.87	2.86	3.34	3.49	3.72
R11	T(N3)-T(H3)	1.04	1.91	2.36	2.75	2.93
R12	G(N1)-T(O2)	2.81	2.86	3.61	3.63	3.79
R13	G(H1)-T(O2)	1.79	1.92	2.80	2.77	2.53

[a]Distances given in Å.

Table S2. Examples of G-T Mispairs in Different Polymerase Environments.[a]

Systems	DNA polymerase λ variant (closed state)	DNA polymerase BF1 variant (ajar state)	DNA polymerase RB69 variant
PDB ID	3PML[16]	3HP6[23]	4M3R[24]
Chain/Residue	ChA/Res701: ChC/Res5:	ChA/Res201: ChC/Res3:	ChP/Res115: ChT/Res5
G-T mispair	GMPCPP:T	TTP:G	T:G
Geometry	WC-like	wobble	wobble
R1	2.66	3.65	3.51
R4	2.97	3.74	3.45
R7	3.06	3.78	3.32
R10	3.83	2.79	2.83
R12	4.01	2.96	2.75
Systems	DNA polymerase Dpo4	DNA polymerase η	DNA polymerase β (closed state)
PDB ID	2AGO[25]	4J9L[26]	4PGX[27]
G-T mispair	ChB/Res1814: ChC/Res1905	ChA/Res503: ChTRes4	ChA/Res404: ChT/Res6
G-T mispair	G:T	GMPNPP:T	TMPNPP:G
Geometry	staggered	wobble	WC-like
R1	4.63	3.50	2.97
R4	3.72	3.47	2.91
R7	3.04	3.55	2.85
R10	5.62	2.62	3.15
R12	5.03	2.55	3.17

[a]Reaction coordinates R1, R4, R7, R10, and R12 are shown in Figure 3. The distances are given in Å.

Table S3. Benchmark Calculations of Different Levels of Theory for Calculating the Relative Energies between the wG-T and the G-T* or G*-T States in Gas Phase.[a]

Method	E(G-T*) − E(wG-T) (kcal/mol)	E(G*-T) − E(wG-T) (kcal/mol)	Average Unsigned Error (kcal/mol)
MP2/CBS[b]	-1.1	-2.3	0.0
CCSD(T)/cc-pVDZ	-0.3	-2.9	0.7
MP2/aug-cc-pVDZ	-1.3	-2.6	0.3
MP2/cc-pVDZ	-0.6	-2.9	0.6
MP2/cc-pVTZ	-1.3	-2.7	0.3
MP2/cc-pVQZ	-1.1	-2.4	0.1
B3LYP/6-31G*	-1.2	-2.5	0.2
B3LYP/6-31+G**	-0.5	-1.3	0.8
B3LYP/6-311G**	0.0	-1.3	1.1
B3LYP/6-311+G**	0.4	-0.6	1.6
M06-2X/6-31G**	-2.9	-4.2	1.9
M06-2X/6-31+G**	-2.4	-3.2	1.1
M06-2X/6-311G**	-1.8	-3.1	0.8
M06-2X/6-311+G**	-1.5	-2.5	0.3
B3LYP-D3/6-31G*	-2.0	-3.4	1.0
B3LYP-D3/6-31+G**	-1.3	-2.2	0.2
B3LYP-D3/6-311G**	-0.9	-2.2	0.2
B3LYP-D3/6-311+G**	-0.5	-1.5	0.7
M06-2X-D3/6-31G**	-3.0	-4.3	2.0
M06-2X-D3/6-31+G**	-2.5	-3.3	1.2
M06-2X-D3/6-311G**	-1.9	-3.2	0.9
M06-2X-D3/6-311+G**	-1.6	-2.6	0.4
ωB97X-D/6-31G*	-1.6	-3.1	0.7
ωB97X-D/6-31+G**	-1.2	-2.2	0.1
ωB97X-D/6-311G**	-0.6	-2.1	0.4
ωB97X-D/6-311+G**	-0.3	-1.5	0.8
ωB97X-D3/6-31G*	-1.7	-3.2	0.8
ωB97X-D3/6-31+G**	-1.2	-2.2	0.1
ωB97X-D3/6-311G**	-0.7	-2.1	0.3
ωB97X-D3/6-311+G**	-0.4	-1.5	0.8

[a]All benchmarking calculations were performed on the geometries of the G-T mispairs optimized at the M06/6-311++G** level of theory given above. These calculations were performed using the Gaussian 09 program[21] except the ωB97X-D3 calculations were performed using the Q-Chem program[28] because this density functional is unavailable in the Gaussian 09 program. [b]The complete basis set (CBS) energies were used as the reference and were extrapolated based on the results of the MP2/cc-pVTZ and MP2/cc-pVQZ calculations using the extrapolation formula: $E_{\mathrm{CBS}} = E_{\mathrm{SCF}}^{(\infty)} + E_{\mathrm{Corr}}^{(\infty)}$, $E_{\mathrm{SCF}}^{(X)} = E_{\mathrm{SCF}}^{(\infty)} + A\exp\left(-\alpha\sqrt{X}\right)$, and $E_{\mathrm{corr}}^{(\infty)} = \frac{X^{\beta}E_{\mathrm{corr}}^{(X)} - Y^{\beta}E_{\mathrm{corr}}^{(Y)}}{X^{\beta} - Y^{\beta}}$, where $X = 3$ and $Y = 4$. Here $\alpha = 5.46$ and $\beta = 3.05$, as given in Section 6.1.3.4 of the ORCA 2.9 manual.[29]

Table S4. Sampling Times for the QM/MM String Simulations.

String	# string iterations	# images per string	Sampling time per image	Sampling time
wG-T→G-T* in aqueous solution	24	25	100 fs	60.0 ps
wG-T→G-T* in A-DNA	25	25	100 fs	62.5 ps
wG-T→G-T* in B-DNA	25	25	100 fs	62.5 ps
wG-T→G-T* in polymerase	25	25	100 fs	62.5 ps
G-T*→G*-T in aqueous solution	25	15	100 fs	37.5 ps
G-T*→G*-T in A-DNA	25	15	100 fs	37.5 ps
G-T*→G*-T in B-DNA	25	15	100 fs	37.5 ps
G-T*→G*-T in polymerase	25	15	100 fs	37.5 ps
			Total	397.5 ps

Table S5. Average Hydrogen Bond Distances and Angles for Key States in G-T Mispair Tautomerization.[a]

wG-T→G-T* tautomerization						
wG-T[b]	O6-N3	O6-H3-N3	N1-O2	N1-H1-O2		
Aqueous	2.86	167.5	2.88	167.7		
A-DNA	2.86	168.1	2.82	167.6		
B-DNA	2.87	169.9	2.86	173.0		
Polymerase	3.01	164.3	2.75	161.5		
G-T*[c]	O6-O4	O6-H3-O4	N1-N3	N1-H1-N3	N2-O2	N2-H21-O2
Aqueous	2.90	164.0	3.02	151.3	3.01	148.8
A-DNA	2.49	164.7	2.75	155.7	2.94	162.0
B-DNA	2.66	170.7	2.87	158.7	2.91	167.5
Polymerase	2.73	169.0	2.81	169.6	2.79	166.0
G-T*→G*-T tautomerization						
G-T*[d]	O6-O4	O6-H3-O4	N1-N3	N1-H1-N3	N2-O2	N2-H21-O2
Aqueous	2.66	172.0	2.85	160.7	2.91	157.6
A-DNA	2.53	169.2	2.85	156.4	3.04	157.6
B-DNA	2.57	166.3	2.79	163.5	2.89	167.4
Polymerase	2.68	161.6	2.76	172.0	2.82	165.8
G*-T[e]	O6-O4	O6-H3-O4	N1-N3	N1-H1-N3	N2-O2	N2-H21-O2
Aqueous	2.70	158.4	2.80	171.8	2.90	164.2
A-DNA	2.58	166.2	2.80	165.8	3.00	166.1
B-DNA	2.55	161.3	2.77	167.4	2.90	167.0
Polymerase	2.59	162.1	2.83	170.0	2.88	165.8

[a]These values were obtained by averaging the distance or angle for the last iteration of each string simulation for the images corresponding to the specified state. Distances are given in Å, and angles are given in degrees. These configurations were weighted according to the umbrella sampling restraints. [b]The images are 2, 1, 2, and 3 of the string for wG-T→G-T* tautomerization in aqueous solution, the A-DNA duplex, the B-DNA duplex, and the DNA polymerase λ variant, respectively. [c]The images are 24, 23, 24, and 25 of the string for wG-T→G-T* tautomerization in aqueous solution, the A-DNA duplex, the B-DNA duplex, and the DNA polymerase λ variant, respectively. [d]The images are 3, 2, 2, and 1 of the string for G-T*→G*-T tautomerization in aqueous solution, the A-DNA duplex, the B-DNA duplex, and the DNA polymerase λ variant, respectively. [e]The images are 13, 14, 14, and 15 of the string for G-T*→G*-T tautomerization in aqueous solution, the A-DNA duplex, the B-DNA duplex, and the DNA polymerase λ variant, respectively.

Table S6. Average Energy of G-T Mispair ($E_{G\text{-}T}$) and Average Electrostatic Interaction Energy (E_{elec}) between G-T Mispair and Environment for Key States of wG-T→G-T* Tautomerization in the A-DNA, B-DNA, and Polymerase Systems.[a]

wG-T→G-T* Tautomerization			
$E_{G\text{-}T}$			
Systems	wG-T	wG-T↔G-T*	G-T*
A-DNA	0.0	17.6 (16.6)	6.2 (5.7)
B-DNA	0.0	5.8 (4.5)	-17.1 (-11.2)
Polymerase	0.0	21.1 (25.4)	2.1 (1.5)
E_{elec}			
Systems	wG-T	wG-T↔G-T*	G-T*
A-DNA	0.0	19.1 (17.1)	12.1 (13.8)
B-DNA	0.0	14.9 (20.9)	23.6 (25.6)
Polymerase	0.0	22.6 (16.5)	-5.7 (-5.0)
$E_{G\text{-}T} + E_{elec}$			
Systems	wG-T	wG-T↔G-T*	G-T*
A-DNA	0.0	36.7 (33.7)	18.4 (19.5)
B-DNA	0.0	20.6 (25.4)	6.5 (14.3)
Polymerase	0.0	43.6 (41.9)	-3.6 (-3.5)

[a]The average energy of the G-T mispair, $E_{G\text{-}T}$, is defined as the average energy of the QM region. The average electrostatic interaction energy between the G-T mispair and the environment, E_{elec}, is defined as the QM/MM electrostatic interaction energy. These average energies were obtained based on the configurations of the last iteration in the string simulations for the images corresponding to the wG-T state (image 1 for A-DNA, image 2 for B-DNA, and image 3 for polymerase), the wG-T↔G-T* state (image 16 for A-DNA, image 14 for B-DNA, and image 14 for polymerase), and the G-T* state (image 23 for A-DNA, image 24 for B-DNA, and image 25 for polymerase). These configurations were weighted according to the umbrella sampling restraints. The values in brackets were obtained from the second-to-last iteration following the same protocol to test the reproducibility. The results of this analysis for the aqueous system are not given because the trends were not found to be reproducible according to this metric, most likely due to the greater flexibility of the base pair in aqueous solution. Energies are given in kcal/mol.

Table S7. P–P Distance of the G-T Mispair in the A-DNA and B-DNA Duplexes and DNA Polymerase λ Variant Systems for Key States in wG-T→G-T* Tautomerization.[a]

	wG-T	wG-T↔G-T*	G-T*
A-DNA	20.3	20.3	19.9
B-DNA	19.7	20.0	19.5
DNA Polymerase λ variant[b]	17.8	17.7	17.3

[a]These distances are averaged over the last iteration of the QM/MM string simulations for images 1 (wG-T), 16 (wG-T↔G-T*), and 23 (G-T*) for the G-T mispair in the A-DNA duplex, image numbers 2 (wG-T), 14 (wG-T↔G-T*), and 24 (G-T*) for the G-T mispair in the B-DNA duplex, and image numbers 3 (wG-T), 14 (wG-T↔G-T*), and 25 (G-T*) for the G-T mispair in the DNA polymerase λ variant. These configurations were weighted according to the umbrella sampling restraints. The distances are given in Å. [b]For the DNA polymerase λ variant, the P–P distance is the distance between the P atom of the T residue and the PA atom of the GTP ligand.

Table S8. Contributions from Selected Groups in the Polymerase to the Relative Average Electrostatic Interaction Energies between G-T Mispair and Environment for Key States of wG-T→G-T* Tautomerization[a]

ΔE_{elec}(G-T*) – ΔE_{elec}(wG-T)		ΔE_{elec}(wG-T↔G-T*) – ΔE_{elec}(wG-T)	
Asn513	-2.5 (-1.5)	Asn513	0.6 (2.2)
Arg517	-10.0 (-10.4)	Arg517	-3.3 (-5.0)
Mg_A^{2+}	-3.9 (-2.6)	Mg_A^{2+}	7.5 (6.1)
Mg_B^{2+}	-2.9 (-2.6)	Mg_B^{2+}	9.0 (7.2)

[a]The average electrostatic interaction energy between the G-T mispair and the environment, E_{elec}, is defined as the QM/MM electrostatic interaction energy. Each term in the table was calculated based on the following equation: $\Delta E_{elec}(A) - \Delta E_{elec}(B) = [E_{elec}(A) - E_{elec}'(A)] - [E_{elec}(B) - E_{elec}'(B)]$, where A and B represent two states, and the prime indicates the average electrostatic interaction energy computed without inclusion of the charges from the specified residue or ion. A negative value of $\Delta E_{elec}(A) - \Delta E_{elec}(B)$ signifies that the state A is stabilized relative to the state B by the electrostatic environment. These energies were computed for the configurations of iteration 25 in the string simulation for wG-T→G-T* tautomerization in the polymerase for the images corresponding to the wG-T state (image 3), the wG-T↔G-T* state (image 14), and the G-T* state (image 25). These configurations were weighted according to the umbrella sampling restraints. The values in brackets were obtained from iteration 24 following the same protocol to test the reproducibility.

References

(1) Hunter, W. N.; Kneale, G.; Brown, T.; Rabinovich, D.; Kennard, O., Refined crystal structure of an octanucleotide duplex with G· T mismatched base-pairs. *J. Mol. Biol.* **1986,** *190*, 605-618.

(2) Zgarbová, M.; Sponer, J.; Otyepka, M.; Cheatham III, T. E.; Galindo-Murillo, R.; Jurecka, P., Refinement of the sugar–phosphate backbone torsion beta for AMBER force fields improves the description of Z-and B-DNA. *J. Chem. Theory Comput.* **2015,** *11*, 5723-5736.

(3) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005,** *26*, 1668-1688.

(4) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79*, 926-935.

(5) Li, P.; Song, L. F.; Merz Jr, K. M., Systematic Parameterization of Monovalent Ions Employing the Nonbonded Model. *J. Chem. Theory Comput.* **2015,** *11*, 1645-1657.

(6) Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98*, 10089-10092.

(7) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977,** *23*, 327-341.

(8) Miyamoto, S.; Kollman, P. A., SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992,** *13*, 952-962.

(9) Götz, A. W.; Clark, M. A.; Walker, R. C., An extensible interface for QM/MM molecular dynamics simulations with AMBER. *J. Comput. Chem.* **2014,** *35*, 95-108.

(10) Brooks, B. R.; Brooks, C. L.; MacKerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S., CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009,** *30*, 1545-1614.

(11) Woodcock, H. L.; Hodošček, M.; Gilbert, A. T.; Gill, P. M.; Schaefer, H. F.; Brooks, B. R., Interfacing Q‐Chem and CHARMM to perform QM/MM reaction path calculations. *J. Comput. Chem.* **2007,** *28*, 1485-1502.

(12) Rosta, E.; Nowotny, M.; Yang, W.; Hummer, G., Catalytic mechanism of RNA backbone cleavage by ribonuclease H from quantum mechanics/molecular mechanics simulations. *J. Am. Chem. Soc.* **2011,** *133*, 8934-8941.

(13) Ganguly, A.; Thaplyal, P.; Rosta, E.; Bevilacqua, P. C.; Hammes-Schiffer, S., Quantum mechanical/molecular mechanical free energy simulations of the self-cleavage reaction in the hepatitis delta virus ribozyme. *J. Am. Chem. Soc.* **2014,** *136*, 1483-1496.

(14) Souaille, M.; Roux, B., Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput. Phys. Commun.* **2001,** *135*, 40-57.

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

(16) Bebenek, K.; Pedersen, L. C.; Kunkel, T. A., Replication infidelity via a mismatch with Watson–Crick geometry. *Proc. Natl. Acad. Sci. U.S.A.* **2011,** *108*, 1862-1867.

(17) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A., H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* **2005,** *33*, W368-W371.

(18) Ucisik, M. N.; Hammes-Schiffer, S., Relative binding free energies of adenine and guanine to damaged and undamaged DNA in human DNA polymerase η: Clues for fidelity and overall efficiency. *J. Am. Chem. Soc.* **2015,** *137*, 13240-13243.
(19) Meagher, K. L.; Redman, L. T.; Carlson, H. A., Development of polyphosphate parameters for use with the AMBER force field. *J. Comput. Chem.* **2003,** *24*, 1016-1025.
(20) Nomura, K.; Hoshino, R.; Shimizu, E.; Hoshiba, Y.; Danilov, V. I.; Kurita, N., DFT calculations on the effect of solvation on the tautomeric reactions for wobble Gua-Thy and canonical Gua-Cyt base-pairs. *J. Mod. Phys.* **2013,** *4*, 422-431.
(21) 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.; Fox, D. J. *Gaussian 09, Revision D. 01*, Gaussian Inc.: Wallingford, CT, 2013.
(22) Grossfield, A., WHAM: the weighted histogram analysis method. *WHAM version 2.0.9*.
(23) Wu, E. Y.; Beese, L. S., The structure of a high fidelity DNA polymerase bound to a mismatched nucleotide reveals an "ajar" intermediate conformation in the nucleotide selection mechanism. *J. Biol. Chem.* **2011,** *286*, 19758-19767.
(24) Xia, S.; Konigsberg, W. H., Mispairs with Watson-Crick base-pair geometry observed in ternary complexes of an RB69 DNA polymerase variant. *Protein Sci.* **2014,** *23*, 508-513.
(25) Vaisman, A.; Ling, H.; Woodgate, R.; Yang, W., Fidelity of Dpo4: effect of metal ions, nucleotide selection and pyrophosphorolysis. *EMBO J.* **2005,** *24*, 2957-2967.
(26) Zhao, Y.; Gregory, M. T.; Biertümpfel, C.; Hua, Y.-J.; Hanaoka, F.; Yang, W., Mechanism of somatic hypermutation at the WA motif by human DNA polymerase η. *Proc. Natl. Acad. Sci. U.S.A.* **2013,** *110*, 8146-8151.
(27) Koag, M.-C.; Nam, K.; Lee, S., The spontaneous replication error and the mismatch discrimination mechanisms of human DNA polymerase β. *Nucleic Acids Res.* **2014,** *42*, 11233-11245.
(28) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X., Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **2015,** *113*, 184-215.
(29) Neese, F., ORCA–An ab initio DFT and Semi-empirical Electronic Structure Package, version 2.9. *Max Planck Institute for Bioinorganic Chemistry* **2012**.