Supporting Information

Relative Binding Free Energies of Adenine and Guanine to Damaged and Undamaged DNA in Human DNA Polymerase \( \eta \): Clues for Fidelity and Overall Efficiency

Melek N. Ucisik and Sharon Hammes-Schiffer*
Department of Chemistry, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, IL 61801-3364

*Corresponding author: shs3@illinois.edu
Computational Methods

The thermodynamic integration (TI) calculations\textsuperscript{1} used the pmemd implementation of alchemical transformations in the AMBER 14 suite of programs,\textsuperscript{2} which was found to be 2.5 times more computationally efficient than the traditionally used sander implementation.\textsuperscript{3} Additionally, the pmemd version is capable of running softcore simulations at the coupling constant $\lambda$ end points and thus obviates the need to extrapolate the data from the middle states, thereby enhancing the accuracy of the free energy calculations.\textsuperscript{3}

The starting structures for all of the free energy simulations were extracted from a series of long MD trajectories using a protocol described in detail elsewhere.\textsuperscript{4} The initial structures for these long MD trajectories were based on crystal structures, and the charged amino acids were modeled in the protonation states obtained with the H++ protonation state server at neutral pH.\textsuperscript{5} The equilibration process for these long MD trajectories consisted of 63,000 steps of energy-minimization, 200 ps of NVT heating, and 10 ns of NPT equilibration. For each system, three independent 1 µs NPT trajectories were obtained. These simulations were shown to be well-converged and equilibrated in terms of the behavior of the total energy, root-mean-square deviation (RMSD), and other standard metrics. The Mg$^{2+}$ ions remained very close to their original positions throughout the 12 µs sampling. Random configurations from the later portions of the long MD trajectories were used as the starting configurations for the TI simulations.

We followed the protocol described by Kaus \textit{et al}.\textsuperscript{3} for the TI calculations. For the first two thermodynamic cycles shown in Scheme 1, we created topologies including both $\lambda$ end states using the solute coordinates from a configuration from the molecular dynamics (MD) trajectory including Pol $\eta$, a DNA template/primer construct with a T-T dimer (TTD) in the template.
strand, and a dATP opposite the 3’ T of the TTD (“TTD3’-A” from Ref. 4, Table S1). The system was solvated with the TIP3P triangulated water model in a periodically replicated truncated octahedral water box with sides at least 10 Å away from all solute atoms and neutralized by the addition of Na\(^+\) ions. Then additional Na\(^+\) and Cl\(^-\) ions were added to bring the salt concentration to 125 mM. For the states in the thermodynamic cycle that do not include the protein, no buffer ions could be added because the box volume was too small to adjust to the desired buffer concentration. Extra care was taken to ensure that the atomic coordinates of the starting structures of the two \(\lambda\) end states for each transformation differed by only the modified atoms.

For the first thermodynamic cycle (Scheme 1a), the calculations were divided into three steps: first the charges on the mutated residue were turned off, then the softcore van der Waals potential\(^7\) was employed to carry out the mutation from the TTD to two consecutive thymines (TT), and lastly the charges were put back on the TT motif. 25,000 steps of steepest descent energy minimization were performed on each system at \(\lambda=0.5\). Then these minimized systems were equilibrated and simulated at a series of \(\lambda\) values with intervals of 0.1 ranging from 0 to 1. During equilibration for each \(\lambda\) value, the system was heated from 0 to 300 K over 200 ps of MD within the canonical ensemble (NVT) in the presence of a weak harmonic restraint on the solute. Subsequently, the density of the system was allowed to adjust in the isobaric, isothermal ensemble (NPT) over 1 ns without any harmonic restraints employing Langevin dynamics with a collision frequency of 1.0 ps\(^{-1}\) to maintain a pressure of 1.0 bar and temperature of 300 K. Following the NPT equilibration, a 5 ns production trajectory with a time step of 2 fs was propagated at each \(\lambda\) value within the NPT ensemble. For all MD simulations, the SHAKE algorithm was used to constrain covalent bonds involving hydrogen, and the Particle Mesh
Ewald (PME) method\textsuperscript{8} was employed for long-range electrostatic interactions, with a 12 Å non-bonded cutoff to limit the direct space sum. Five independent cycles were completed using the same energy minimized structures at $\lambda=0.5$ but different initial velocities for the equilibration. The $\partial V/\partial \lambda$ data were collected from the production MD trajectories at the discrete $\lambda$ values every 2 ps. The average $\partial V/\partial \lambda$ values were plotted versus $\lambda$, and the area under the resulting curve was calculated with the trapezoidal rule to obtain the free energy change for the reaction under investigation.

For the second and third thermodynamic cycles (Schemes 1b and 1c), we employed an intermediate dNTP (“INT”) structure (Figure S1) to facilitate sampling and convergence during the transformation of the A base to the G base in the free dNTP molecule. A random configuration from each of the previous MD trajectories comprised of Pol $\eta$, a DNA template/primer construct with a TTD in the template strand, and a dATP opposite the 3’ T of the TTD or a dATP opposite the 5’ T of the TTD (“TTD3’-A” or “TTD5’-A” in Ref. 4)\textsuperscript{4} was chosen as the starting structure for the second and third cycles, respectively. The relevant atoms on the dNTP base were modified to their counterparts in the INT structure (Figure S1). The bond, angle, dihedral, and van der Waals parameters were obtained with the parmchk2 utility of AmberTools\textsuperscript{13}.\textsuperscript{9} The atomic charges were determined as described below. The same preparation, simulation, and data collection steps as in the first thermodynamic cycle were applied in conjunction with softcore van der Waals and electrostatic potentials.\textsuperscript{7,10,11}

\textit{Obtaining the Charges for the INT Base with RESP}

The restrained electrostatic potential (RESP)\textsuperscript{12} charges for the INT base (Figure S1) were obtained after a geometry optimization at the Hartree-Fock level using the 6-31G* basis set with
an in-house code. The atomic charges of the deoxyribose and the phosphate groups in the AMBER libraries for DNA were adopted except for the C1’ atoms of the sugar rings and the base. The H-capped C1’ atoms and the base were treated as a new entity, on which RESP calculations were carried out, where the charge fitting was restricted to all atoms except C4 and C5 of the pyrimidine ring. First the charges of the more electronegative N and O atoms were fit to the electrostatic potential at points chosen in accordance with the Merz-Singh-Kollman scheme, and then the charges of the remaining C and H atoms were optimized in the same way. The rest of the charges were fixed to their values of the G base in the AMBER libraries for DNA. The charges for the T-T dimer, dATP, and dGTP were obtained with a similar protocol in our previous work, the details of which can be found in the associated Supporting Information of reference 4.

**Figure S1.** An intermediate structure was employed in the second and the third TI cycles to enhance the convergence. The above structure comprises the base of that intermediate molecule, denoted “INT”.

S5
Figure S2. $\frac{\partial V}{\partial \lambda}$ values plotted against the coupling parameter $\lambda$ for the reactions 1 and 2 of the thermodynamic cycle in Scheme 1a. Every reaction was divided into three parts, where first the charges on the TTD were removed (a, d), then the transformation from the TTD to two regular, consecutive T’s occurred (b, e), and lastly the charges on the two consecutive, healthy T’s were restored (c, f). The panels a, b, and c display data from transformations in the presence of the dATP free nucleotide opposite the 3’ T of the TTD or the two healthy T’s replacing it. The panels d, e, and f display data from transformations in the absence of dATP. The different colors represent the results from the five parallel TI simulations. Error bars correspond to the standard deviation of the $\frac{\partial V}{\partial \lambda}$ values obtained for each TI simulation.
Figure S3. RMSD profiles of the TTD/TT motif obtained from three independent trajectories, depicted in three different colors, associated with the systems TTD3'-A (A), TTD3'-G (B), N/A-A (C), and TTD5'-A (D). TTD3'-A is comprised of the catalytic core of Pol η, a DNA primer/template with a TTD in the template strand, and an incoming dATP opposite the 3’ T of the TTD lesion. TTD3’-G is comprised of the same components except dATP is replaced by dGTP. N/A-A differs from TTD3’-G in that the TTD is replaced by two normal thymines. Finally, TTD5’-A is comprised of the same components as TTD3’-A except the incoming dATP pairs with the 5’ T of the TTD. The N/A-A system exhibits the least stabilization for this motif.
Figure S4. ∂V/∂λ values plotted against the coupling parameter λ for the reactions numbered as 1, 2, 5, and 6 according to the thermodynamic cycle of Scheme 1b. The corresponding reaction numbers are also shown on the bottom left corner of each plot. (1) Transformation of the A base in dNTP to INT in the field of Pol η and the damaged TTD-containing DNA template/primer, where dNTP is opposite the 3’ T of the TTD lesion, (2) transformation of the INT base in dNTP to G in the field of Pol η and the damaged TTD-containing DNA template/primer, where dNTP is opposite the 3’ T of the TTD lesion, (5) transformation of the A base in dNTP to INT in bulk water, (6) transformation of the INT base in dNTP to G in bulk water. The different colors represent the results from the five parallel TI simulations. Error bars correspond to the standard deviation of the ∂V/∂λ values obtained for each TI simulation.
Figure S5. $\frac{\partial V}{\partial \lambda}$ values plotted against the coupling parameter $\lambda$ for the reactions numbered as 1, 2, 5, and 6 in the thermodynamic cycle of Scheme 1c. The corresponding reaction numbers are also shown on the bottom left corner of each plot. (1) Transformation of the A base in dNTP to INT in the field of Pol $\eta$ and the damaged TTD-containing DNA template/primer, where dNTP is opposite the 5' T of the TTD lesion, (2) transformation of the INT base in dNTP to G in the field of Pol $\eta$ and the damaged TTD-containing DNA template/primer, where dNTP is opposite the 5' T of the TTD lesion, (5) transformation of the A base in dNTP to INT in bulk water, (6) transformation of the INT base in dNTP to G in bulk water. The numbers for the transformations in bulk water (5 and 6) were the same as the previous thermodynamic cycle for the insertion against the 3' thymine. The different colors represent the results from the five parallel TI simulations. Error bars correspond to the standard deviation of the $\frac{\partial V}{\partial \lambda}$ values obtained for each TI simulation.
Figure S6. Normalized population distributions of the $\beta V/\beta \lambda$ values for the reactions 1 (upper half) and 2 (lower half) of the thermodynamic cycle in Scheme 1a. The data from the five replicate TI simulations are shown in each column. Every leg was divided into three parts, where first the charges on the TTD were removed (panels 1-5 and 16-20), then the transformation from the TTD to two regular, consecutive T’s occurred (panels 6-10 and 21-25), and lastly the charges on the two consecutive, healthy T’s were restored (panels 11-15 and 26-30). The panels 1-15 display data from transformations in the presence of the dATP free nucleotide opposite the 3’-T of the TTD or the two healthy T’s replacing it. The panels 16-30 display data from transformations in the absence of dATP. The different colors represent the different $\lambda$ values for the transformations. These histograms were generated to show the phase space overlap in terms of the $\beta V/\beta \lambda$ values at the distinct $\lambda$ values constituting a transformation.
Figure S7. Normalized population distributions of $\partial V/\partial \lambda$ values for the reactions numbered as 1, 2, 5, and 6 in the thermodynamic cycle of Scheme 1b. The data from the five replicate TI simulations are shown in each column. The panels display the following: transformation of the A base in dNTP to INT in the field of Pol $\eta$ and the damaged TTD-containing DNA template/primer, where dNTP is opposite the 3’ T of the TTD lesion (panels 1-5), transformation of the INT base in dNTP to G in the field of Pol $\eta$ and the damaged TTD-containing DNA template/primer, where dNTP is opposite the 3’ T of the TTD lesion (panels 6-10), transformation of the A base in dNTP to INT in bulk water (11-15), and transformation of the INT base in dNTP to G in bulk water (panels 16-20). These histograms were generated to show the phase space overlap in terms of the $\partial V/\partial \lambda$ values at the distinct $\lambda$ values constituting a transformation.
Figure S8. Normalized population distributions of $\frac{\partial V}{\partial \lambda}$ values for the reactions numbered as 1, 2, 5, and 6 in the thermodynamic cycle of Scheme 1c. The data from the five replicate TI simulations are shown in each column. The panels display the following: transformation of the A base in dNTP to INT in the field of Pol $\eta$ and the damaged TTD-containing DNA template/primer, where dNTP is opposite the 5’ T of the TTD lesion (panels 1-5), transformation of the INT base in dNTP to G in the field of Pol $\eta$ and the damaged TTD-containing DNA template/primer, where dNTP is opposite the 5’ T of the TTD lesion (panels 6-10), transformation of the A base in dNTP to INT in bulk water (11-15), and transformation of the INT base in dNTP to G in bulk water (panels 16-20). These histograms were generated to show the phase space overlap in terms of the $\frac{\partial V}{\partial \lambda}$ values at the distinct $\lambda$ values constituting a transformation.
Figure S9. 10-point running averages of the $\partial V/\partial \lambda$ values at each $\lambda$ for the reactions 1 (upper half) and 2 (lower half) of the thermodynamic cycle in Scheme 1a. The data from the five replicate TI simulations are shown in each column. Every reaction was divided into three parts, where first the charges on the TTD were removed (panels 1-5 and 16-20), then the transformation from the TTD to two regular, consecutive T’s occurred (panels 6-10 and 21-25), and lastly the charges on the two consecutive, healthy T’s were restored (panels 11-15 and 26-30). The panels 1-15 display data from transformations in the presence of the dATP free nucleotide opposite the 3’-T of the TTD or the two healthy T’s replacing it. The panels 16-30 display data from transformations in the absence of dATP. The different colors represent the different $\lambda$ values for the transformations.
Figure S10. 10-point running averages of the $\partial V / \partial \lambda$ values at each $\lambda$ for the reactions 1 (upper half) and 2 (lower half) of the thermodynamic cycle in Scheme 1a. This figure encompasses the same set of graphs in Figure S9 with smaller ranges of the y-axes for clarity.
Figure S11. 10-point running averages of $\partial V/\partial \lambda$ values at each $\lambda$ for the reactions numbered as 1, 2, 5, and 6 in the thermodynamic cycle of Scheme 1b. The data from the five replicate TI simulations are shown in each column. The panels display the following: transformation of the A base in dNTP to INT in the field of Pol $\eta$ and the damaged TTD-containing DNA template/primer, where dNTP is opposite the 3’ T of the TTD lesion (panels 1-5, reaction 1 of Scheme 1b), transformation of the INT base in dNTP to G in the field of Pol $\eta$ and the damaged TTD-containing DNA template/primer, where dNTP is opposite the 3’ T of the TTD lesion (panels 6-10, reaction 2 of Scheme 1b), transformation of the A base in dNTP to INT in bulk water (11-15, reaction 5 of Scheme 1b), and transformation of the INT base in dNTP to G in bulk water (panels 16-20, reaction 6 of Scheme 1b). The different colors represent the different $\lambda$ values for the transformations. The curves for $\lambda=1$ (in orange) in panels 1-5 and 18-20 show elevated fluctuations. We checked these simulations visually for possible multiple binding modes and detected no problems. The simulations at the endpoints of TI calculations are known to be more difficult to converge. The same feature is exhibited with the larger error bars obtained at $\lambda=1$ in Figure S4 and the broader distributions of $\lambda=1$ values in Figure S7. Quantitatively, Table S3 displays larger standard deviations for the reaction 1, where the greatest fluctuations are observed, but the simulations still yield meaningful relative binding free energy values with the reported standard deviations.
Figure S12. 10-point running averages of $\partial V/\partial \lambda$ values at each $\lambda$ for the reactions numbered as 1, 2, 5, and 6 in the thermodynamic cycle of Scheme 1c. The data from the five replicate TI simulations are shown in each column. The panels display the following: transformation of the A base in dNTP to INT in the field of Pol $\eta$ and the damaged TTD-containing DNA template/primer, where dNTP is opposite the 5’ T of the TTD lesion (panels 1-5, reaction 1 of Scheme 1c), transformation of the INT base in dNTP to G in the field of Pol $\eta$ and the damaged TTD-containing DNA template/primer, where dNTP is opposite the 5’ T of the TTD lesion (panels 6-10, reaction 2 of Scheme 1c), transformation of the A base in dNTP to INT in bulk water (11-15, reaction 5 of Scheme 1c), and transformation of the INT base in dNTP to G in bulk water (panels 16-20, reaction 6 of Scheme 1c). The different colors represent the different $\lambda$ values for the transformations. We checked these simulations visually for possible multiple binding modes and detected no problems. The simulations at the endpoints of TI calculations are known to be more difficult to converge. The same feature is exhibited with the larger error bars obtained at $\lambda=1$ in Figure S5 and the broader distributions of $\lambda=1$ values in Figure S8. Quantitatively, Table S4 displays larger standard deviations for the reaction 1, where the greatest fluctuations are observed, but the simulations still yield meaningful relative binding free energy values with the reported standard deviations.
Table S1. Information about the systems used in the previous molecular dynamics studies.

| System        | Based on PDB structure | dNTP  | TTD? | To which base of the TTD is the dNTP opposite? |
|---------------|------------------------|-------|------|-----------------------------------------------|
| TTD3'-A       | 3MR3                   | dATP  | Yes  | 3'                                            |
| TTD3'-G       | 3MR3                   | dGTP  | Yes  | 3'                                            |
| N/A-A         | 3MR2                   | dATP  | No   | N/A                                           |
| TTD5'-A       | 3SI8                   | dATP  | Yes  | 5'                                            |

Table S2. Free energies calculated for the reactions 1 and 2 of Scheme 1a. The $\Delta \Delta G_{binding}$ is the relative free energy of binding dATP to the enzyme-DNA complex with the TTD at the 3' T versus the enzyme-DNA complex with two healthy thymines at the same position. On average, dATP binding to the TTD-containing DNA is thermodynamically favored by 4.29 kcal/mol over binding to undamaged DNA.

|                      | Run 1 | Run 2 | Run 3 | Run 4 | Run 5 | Average | Standard Deviation |
|----------------------|-------|-------|-------|-------|-------|---------|--------------------|
| Reaction 1           | 1.68  | 5.43  | 8.50  | 5.56  | 6.36  | 5.51    | 2.21               |
| Reaction 2           | 5.78  | 9.63  | 11.44 | 9.91  | 12.20 | 9.79    | 2.22               |
| $\Delta \Delta G_{binding}$ | -4.10 | -4.20 | -2.94 | -4.35 | -5.84 | -4.29   | 1.03               |

Table S3. Free energies calculated for the reactions 1, 2, 5, and 6 of Scheme 1b. The $\Delta \Delta G_{binding}$ is the relative free energy of binding dATP to the enzyme-DNA complex with the TTD at the 3' T versus binding dGTP at the same position in this complex. On average, dATP binding is thermodynamically favored over dGTP binding by 2.37 kcal/mol.

|                      | Run 1 | Run 2 | Run 3 | Run 4 | Run 5 | Average | Standard deviation |
|----------------------|-------|-------|-------|-------|-------|---------|--------------------|
| Reaction 1           | -7.92 | -6.54 | -5.76 | -8.79 | -8.31 | -7.47   | 1.27               |
| Reaction 2           | 79.01 | 78.79 | 79.13 | 79.62 | 79.76 | 79.26   | 0.41               |
| Reaction 5           | -9.16 | -9.20 | -9.25 | -9.18 | -9.31 | -9.22   | 0.06               |
| Reaction 6           | 78.52 | 78.96 | 78.40 | 78.57 | 78.75 | 78.64   | 0.22               |
| $\Delta \Delta G_{binding}$ | -1.74 | -2.48 | -4.22 | -1.43 | -2.00 | -2.37   | 1.10               |

Table S4. Free energies calculated for the reactions 1, 2, 5, and 6 of Scheme 1c. The $\Delta \Delta G_{binding}$ is the relative free energy of binding dATP to the enzyme-DNA complex with the TTD at the 5' T versus binding dGTP at the same position in this complex. On average, dATP binding is thermodynamically favored over dGTP binding by 2.94 kcal/mol.

|                      | Run 1 | Run 2 | Run 3 | Run 4 | Run 5 | Average | Standard deviation |
|----------------------|-------|-------|-------|-------|-------|---------|--------------------|
| Reaction 1           | -6.79 | -7.04 | -6.11 | -6.93 | -6.73 | -6.72   | 0.36               |
| Reaction 2           | 79.02 | 78.89 | 79.28 | 79.19 | 79.04 | 79.08   | 0.15               |
| Reaction 5           | -9.16 | -9.20 | -9.25 | -9.18 | -9.31 | -9.22   | 0.06               |
| Reaction 6           | 78.52 | 78.96 | 78.40 | 78.57 | 78.75 | 78.64   | 0.22               |
| $\Delta \Delta G_{binding}$ | -2.87 | -2.08 | -4.02 | -2.87 | -2.87 | -2.94   | 0.69               |
Table S5. The coordinates, AMBER atom types, and charges for the dATP molecule used in the simulations.

| Atom # | Atom Name | X     | Y     | Z     | Amber Atom Type | Charge  |
|--------|-----------|-------|-------|-------|-----------------|---------|
| 1      | N1        | -24.18 | 49.62 | 6.46  | NC              | -0.7624 |
| 2      | C2        | -23.02 | 49.83 | 5.73  | CQ              | 0.5716  |
| 3      | N3        | -22.24 | 48.91 | 5.40  | NC              | -0.7417 |
| 4      | C4        | -22.55 | 47.57 | 5.62  | CB              | 0.38    |
| 5      | C5        | -23.69 | 47.22 | 6.45  | CB              | 0.0725  |
| 6      | C6        | -24.61 | 48.37 | 6.79  | CA              | 0.6897  |
| 7      | N6        | -25.64 | 48.12 | 7.44  | N2              | -0.9123 |
| 8      | N7        | -23.72 | 45.94 | 6.52  | NB              | -0.6175 |
| 9      | C8        | -22.66 | 45.58 | 5.93  | CK              | 0.1607  |
| 10     | N9        | -21.95 | 46.52 | 5.44  | N*              | -0.0268 |
| 11     | PA        | -21.58 | 41.37 | 5.75  | P               | 1.12763 |
| 12     | PB        | -18.91 | 41.12 | 7.14  | P               | 1.44258 |
| 13     | PG        | -19.19 | 38.50 | 8.09  | P               | 1.34166 |
| 14     | C1'       | -20.69 | 46.39 | 4.64  | CT              | 0.0431  |
| 15     | O1A       | -21.29 | 40.39 | 5.10  | O2              | -0.871294 |
| 16     | O1B       | -18.55 | 40.74 | 5.98  | O2              | -0.888027 |
| 17     | O1G       | -19.77 | 38.29 | 6.99  | O3              | -0.968624 |
| 18     | C2'       | -19.46 | 46.25 | 5.59  | CT              | -0.0854 |
| 19     | O2A       | -22.19 | 41.53 | 6.30  | O2              | -0.871294 |
| 20     | O2B       | -18.34 | 42.18 | 7.54  | O2              | -0.888027 |
| 21     | O2G       | -18.11 | 37.70 | 8.36  | O3              | -0.968624 |
| 22     | C3'       | -19.17 | 44.77 | 5.63  | CT              | 0.0713  |
| 23     | O3'       | -17.93 | 44.37 | 5.88  | OH              | -0.6549 |
| 24     | O3B       | -18.69 | 40.01 | 8.21  | OS              | -0.647602 |
| 25     | O3G       | -20.03 | 38.37 | 9.15  | O3              | -0.968624 |
| 26     | C4'       | -19.69 | 44.46 | 4.20  | CT              | 0.1629  |
| 27     | O4'       | -20.69 | 45.29 | 3.88  | OS              | -0.3691 |
| 28     | C5'       | -20.18 | 42.94 | 4.21  | CT              | -0.0069 |
| 29     | O5'       | -21.38 | 42.65 | 4.93  | OS              | -0.623459 |
| 30     | H2        | -22.75 | 50.83 | 5.43  | H5              | 0.0598  |
| 31     | HN1       | -25.89 | 47.15 | 7.63  | H               | 0.4167  |
| 32     | HN2       | -26.22 | 48.87 | 7.77  | H               | 0.4167  |
| 33     | H8        | -22.41 | 44.53 | 5.87  | H5              | 0.1877  |
| 34     | H1'       | -20.54 | 47.31 | 4.07  | H2              | 0.1838  |
| 35     | H2'1      | -18.63 | 46.82 | 5.18  | HC              | 0.0718  |
| 36     | H2'2      | -19.70 | 46.63 | 6.58  | HC              | 0.0718  |
| 37     | H3'       | -19.83 | 44.27 | 6.32  | H1              | 0.0985  |
| 38     | HO        | -18.01 | 43.56 | 6.40  | HO              | 0.4396  |
| 39     | H4'       | -18.88 | 44.55 | 3.48  | H1              | 0.1176  |
### Table S6. The coordinates, AMBER atom types, and charges for the dGTP molecule used in the simulations.

| Atom # | Atom Name | X     | Y     | Z     | Amber Atom Type | Charge   |
|--------|-----------|-------|-------|-------|----------------|----------|
| 1      | C2        | -23.05| 49.897| 5.738 | CA             | 0.7432   |
| 2      | N1        | -24.102| 49.641| 6.631 | NA             | -0.5053  |
| 3      | N3        | -22.298| 48.872| 5.211 | NC             | -0.6636  |
| 4      | C4        | -22.685| 47.598| 5.519 | CB             | 0.1814   |
| 5      | C5        | -23.711| 47.28 | 6.421 | CB             | 0.1991   |
| 6      | C6        | -24.425| 48.391| 6.927 | C              | 0.4918   |
| 7      | N7        | -23.837| 45.917| 6.583 | NB             | -0.5725  |
| 8      | C8        | -22.836| 45.5  | 5.759 | CK             | 0.0736   |
| 9      | N9        | -22.118| 46.466| 5.106 | N*             | 0.0577   |
| 10     | PA        | -21.927| 41.319| 5.363 | P              | 1.127635 |
| 11     | PB        | -19.232| 40.901| 6.272 | P              | 1.442587 |
| 12     | PG        | -19.397| 38.295| 7.181 | P              | 1.341652 |
| 13     | C1'       | -21.002| 46.379| 4.21  | CT             | 0.0358   |
| 14     | O1A       | -21.891| 40.183| 4.395 | O2             | -0.871294|
| 15     | O1B       | -18.971| 40.448| 4.888 | O2             | -0.888027|
| 16     | O1G       | -20.151| 38.219| 5.919 | O3             | -0.968624|
| 17     | C2'       | -19.681| 46.136| 4.944 | CT             | -0.0854  |
| 18     | O2A       | -23.238| 41.443| 6.009 | O2             | -0.871294|
| 19     | O2B       | -18.424| 42.098| 6.554 | O2             | -0.888027|
| 20     | O2G       | -18.221| 37.427| 7.198 | O3             | -0.968624|
| 21     | C3'       | -19.504| 44.633| 4.874 | CT             | 0.0713   |
| 22     | O3'       | -18.164| 44.177| 4.925 | OH             | -0.6549  |
| 23     | O3B       | -18.856| 39.766| 7.295 | OS             | -0.647602|
| 24     | O3G       | -20.105| 38.194| 8.442 | O3             | -0.968624|
| 25     | C4'       | -20.184| 44.315| 3.547 | CT             | 0.1629   |
| 26     | O4'       | -21.228| 45.263| 3.364 | OS             | -0.3691  |
| 27     | C5'       | -20.735| 42.884| 3.561 | CT             | -0.0069  |
| 28     | O5'       | -21.806| 42.664| 4.496 | OS             | -0.623459|
| 29     | O         | -20.753| 41.342| 6.432 | OS             | -0.405899|
| 30     | H8        | -22.618| 44.46 | 5.628 | H5             | 0.1997   |
| 31     | H1'       | -20.925| 47.305| 3.68  | H2             | 0.1746   |
| 32     | H2'       | -19.818| 46.406| 5.97  | HC             | 0.0718   |
| 33     | H2"       | -18.897| 46.588| 4.372 | HC             | 0.0718   |
| 34     | H3'       | -19.919| 44.135| 5.726 | H1             | 0.0985   |
Table S6. The coordinates, AMBER atom types, and charges for the TTD molecule used in the simulations.

| Atom # | Atom Name | X     | Y     | Z     | Amber Atom Type | Charge |
|--------|-----------|-------|-------|-------|-----------------|--------|
| 1      | P         | -29.085 | 57.994 | 16.211 | P               | 1.1659 |
| 2      | OP1       | -28.371 | 57.018 | 17.093 | O2              | -0.7761|
| 3      | OP2       | -30.405 | 58.515 | 16.625 | O2              | -0.7761|
| 4      | O5'       | -29.38  | 57.496 | 14.708 | OS              | -0.4954|
| 5      | C5'       | -28.421 | 56.621 | 14.068 | C               | -0.0069|
| 6      | C4R       | -27.375 | 57.35  | 13.375 | CT              | 0.1629 |
| 7      | O4'       | -26.491 | 56.272 | 12.998 | OS              | -0.3691|
| 8      | C3R       | -27.665 | 58.029 | 11.99  | CT              | 0.0713 |
| 9      | O3R       | -26.591 | 58.955 | 11.783 | OS              | -0.5232|
| 10     | C2'       | -27.492 | 56.843 | 11.004 | CT              | -0.0854|
| 11     | C1'       | -26.235 | 56.262 | 11.611 | CT              | 0.068  |
| 12     | N1        | -25.947 | 54.867 | 11.193 | N*              | -0.020321|
| 13     | C2        | -24.798 | 54.241 | 11.611 | C               | 0.551271|
| 14     | O2        | -23.855 | 54.931 | 12.057 | O               | -0.553208|
| 15     | N3        | -24.627 | 52.881 | 11.39  | NA              | -0.551009|
| 16     | C4        | -25.609 | 52.022 | 10.859 | C               | 0.52168 |
| 17     | O4        | -25.404 | 50.796 | 10.839 | O               | -0.502676|
| 18     | C5        | -26.871 | 52.518 | 10.333 | CT              | 0.088656|
| 19     | C5A       | -28.049 | 52.032 | 11.274 | CT              | -0.169457|
| 20     | C6        | -26.833 | 54.041 | 10.372 | CT              | -0.255178|
| 21     | PB        | -26.396 | 59.706 | 10.415 | P               | 1.1659 |
| 22     | O5P       | -25.434 | 60.691 | 10.804 | O2              | -0.7761|
| 23     | O4P       | -27.736 | 60.083 | 9.844  | O2              | -0.7761|
| 24     | O5R       | -25.76  | 58.648 | 9.452  | OS              | -0.4954|
| 25     | C5R       | -24.377 | 58.515 | 9.425  | CT              | -0.0069|
| 26     | O4R       | -24.593 | 56.152 | 8.829  | OS              | -0.3691|
| 27     | C2R       | -25.274 | 56.313 | 6.644  | CT              | -0.0854|
| 28     | C1R       | -24.725 | 55.359 | 7.704  | CT              | 0.068  |
Complete References for the Article

(27) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham III, T. E.; 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.;
References for SI

(1) Straatsma, T. P.; McCammon, J. A. J. Chem. Phys. 1991, 95, 1175.
(2) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham III, T. E.; 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.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A. AMBER 14; University of California, San Francisco, 2014.
(3) Kaus, J. W.; Pierce, L. T.; Walker, R. C.; McCammon, J. A. J. Chem. Theory Comput. 2013, 9, 4131.
(4) Ucisik, M. N.; Hammes-Schiffer, S. Biochemistry 2015, Submitted.
(5) Gordon, J. C.; Myers, J. B.; Foltz, T.; Shoja, V.; Heath, L. S.; Onufriev, A. Nucleic Acids Res. 2005, 33, W368.
(6) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. J. Chem. Phys. 1983, 79, 926.
(7) Steinbrecher, T.; Joung, J.; Case, D. A. J. Comput. Chem. 2011, 32, 3253.
(8) York, D. M.; Darden, T. A.; Pedersen, L. G. J. Chem. Phys. 1993, 99, 8345.
(9) D.A. Case; T.A. Darden; T.E. Cheatham, I.; C.L. Simmerling; J. Wang; R.E. Duke; R. Luo; R.C. Walker; W. Zhang; K.M. Merz; B. Roberts; S. Hayik; A. Roitberg; G. Seabra; J. Swails; A.W. Götz; I. Kolossváry; K.F. Wong; F. Paesani; J. Vanicek; R.M. Wolf; J. Liu; X. Wu; S.R. Brozell; T. Steinbrecher; H. Gohlke; Q. Cai; X. Ye; J. Wang; M.-J. Hsieh; G. Cui; D.R. Roe; D.H. Mathews; M.G. Seetin; R. Salomon-Ferrer; C. Sagui; V. Babin; T. Luchko; S. Gusarov; A. Kovalenko; Kollman, P. A.; University of California, San Francisco: 2012.
(10) Shirts, M. R.; Pande, V. S. J. Chem. Phys. 2005, 122.
(11) Steinbrecher, T.; Mobley, D. L.; Case, D. A. J. Chem. Phys. 2007, 127.
(12) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. J. Phys. Chem. 1993, 97, 10269.
(13) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. J. Am. Chem. Soc. 1995, 117, 5179.
(14) Singh, U. C.; Kollman, P. A. J. Comput. Chem. 1984, 5, 129.
(15) Besler, B. H.; Merz, K. M.; Kollman, P. A. J. Comput. Chem. 1990, 11, 431.