

## Supporting Information

### Supplementary Figure Legends

**Supplementary Figure S1.** DNA changes in the A:C system. (a) DNA in A:C system (colored by atom) superimposed with DNA in the active position (colored purple) from the ternary pol  $\lambda$ /DNA/dTTP complex (PDB entry 1XSN). Black arrows indicate how the DNA shifts from its initial active position. A is the templating adenine at the gap. (b) Time evolution of key distances showing proximity between A and dCTP (yellow) and distance changes between dCTP's base and triphosphate group (blue). Accompanying image shows the positions of the atoms included in the graph.

**Supplementary Figure S2.** Important changes in the A:G system. (a)–(d) Time evolution of key distances involving A5, A6, dGTP, Arg517, and Tyr505 in pol  $\lambda$ 's active site. A5 is the templating adenine at the gap. A6 is the adjacent template adenine that pairs with the primer terminus.

**Supplementary Figure S3.** Changes in Arg517 and dGTP interactions and positions in the T:G system. Time evolution of torsions in dGTP (a) and Arg517 (b). (c) Time evolution of Arg517/dGTP electrostatic interaction energy.

**Supplementary Figure S4.** Key distance changes in the A(*syn*):A system. (a)–(g) Time evolution of key distances involving A(*syn*)5, A6, T6, dATP, and Arg517 in pol  $\lambda$ 's active site. A(*syn*)5 is the templating adenine at the gap. A6 is the adjacent template adenine that pairs with the primer terminus. T6 is the primer terminus thymine.

**Supplementary Figure S5.** Changes occurring during the equilibration phase of the A(*syn*):A system. Snapshots from the equilibration phase are arranged chronologically from left to right. For the A(*syn*):A system, the templating base at the gap (A(*syn*)5), dATP, and Arg517 are colored by atom, while the template adenine base (A6), adjacent to A(*syn*)5, is colored in blue. Green structures represent the corresponding positions of the template adenine (A5) and Arg517 in the ternary pol  $\lambda$  complex with the correct A:dTTP nascent base pair (PDB entry 1XSN). Hydrogen bonding is indicated by dashed yellow lines. Blue arrows indicate movement by Arg517.

**Supplementary Figure S6.** Important active site changes in the A(*syn*):G system. (a) Time

evolution of dGTP torsion defining the change between the minor “straight” and major “sideways” orientations. (b)–(f) Time evolution of distances involving A(*syn*)5, A6, T6, dGTP, and Arg517 in pol  $\lambda$ 's active site. A(*syn*)5 is the templating adenine at the gap. A6 is the adjacent template adenine that pairs with the primer terminus. T6 is the primer terminus thymine.

**Supplementary Figure S7.** Key distance changes in the A:A system. (a)–(f) Time evolution of key distances involving A5, A6, T6, Arg517, Tyr505, and Arg488 in pol  $\lambda$ 's active site. A5 is the templating adenine at the gap. A6 is the adjacent template adenine that pairs with the primer terminus. T6 is the primer terminus thymine.

**Supplementary Figure S8.** Related motions of Phe506 and dATP with key active-site interatomic distances in the A:A system. (a) Time evolution of Phe506 side-chain torsion illustrating changes between active and “sideways” orientations. (b) Time evolution of dATP pseudorotation phase angle. (c)–(d) Time evolution of distances involving T6 and dATP. T6 is the primer terminus thymine.

**Supplementary Figure S9.** Key distance changes in the T:G system. (a)–(h) Time evolution of key distances involving T5, A6, T6, dGTP, Arg517, and Tyr505 in pol  $\lambda$ 's active site. T5 is the templating thymine at the gap. A6 is the adjacent template adenine that pairs with the primer terminus. T6 is the primer terminus thymine.

**Supplementary Figure S10.** Lys273 interactions with template bases A(*syn*)5 and A6 in the A(*syn*):G system. (a) Simulation snapshot showing hydrogen bonds (dashed black lines) formed between Lys273 and both A(*syn*)5 and A6. (b) Time evolution of key distances between Lys273 and A6 or A(*syn*)5 heavy atoms for hydrogen bond formation. A(*syn*)5 is the template adenine at the gap. A6 is the adjacent template adenine that pairs with the primer terminus, T6.

**Supplementary Figure S11.** Lys273 interactions with template base A(*syn*)5 in the A(*syn*):A system. (a) System snapshot showing hydrogen bonds (dashed green line) formed between Lys273 and A(*syn*)5. (b) Time evolution of key distance between Lys273 and A(*syn*)5 heavy atoms for hydrogen bond formation. A(*syn*)5 is the template adenine at the gap.

**Supplementary Figure S12.** Lys273 interactions with template base A6 in the A:A system. (a)–(b) System snapshots showing hydrogen bonds (dashed pink lines) formed between Lys273 and

A6. (c) Time evolution of key distance between Lys273 and A6 heavy atoms for hydrogen bond formation. (d) Time evolution of A6 pseudorotation phase angle. A6 is the template adenine that pairs with the primer terminus, T6. A5 is the templating base at the gap.

**Supplementary Figure S13.** Time evolution of distances between Tyr505 and dGTP in the A(*syn*):G system.

**Supplementary Figure S14.** Diagram showing the hypothetical relationship between the position of the simulated structure and the crystal states. The sides of the triangle are labeled as a, b and c. The shift distance, h, represents the displacement of the simulated structure in the direction perpendicular to the line joining the geometric centers of the crystal structures, c. The variable lengths L1 and L2 describe the RMSD of the simulated structure projected onto c with respect to the crystal binary and ternary states, respectively.

## Supplementary Figures

## A:C Simulation

a)

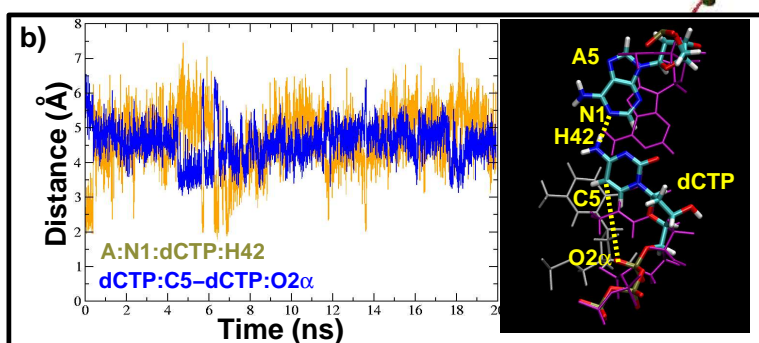
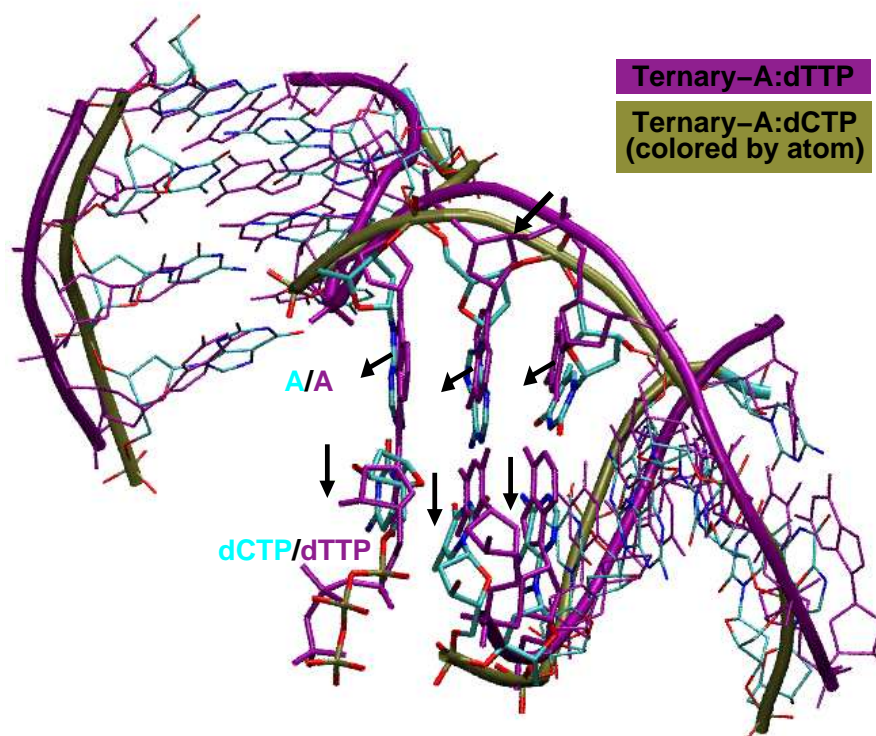


Figure S1:

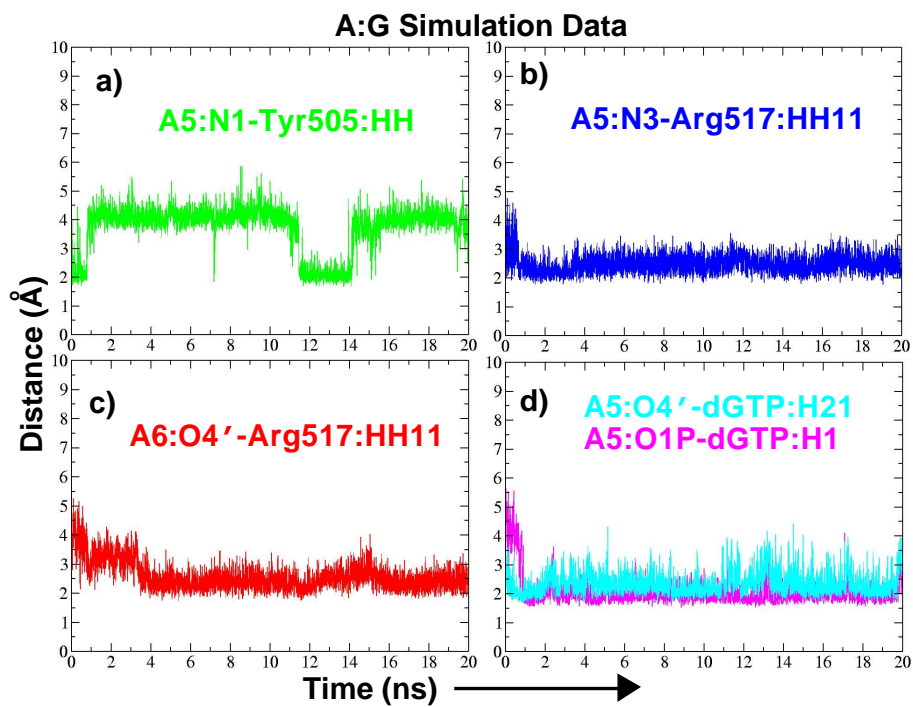


Figure S2:

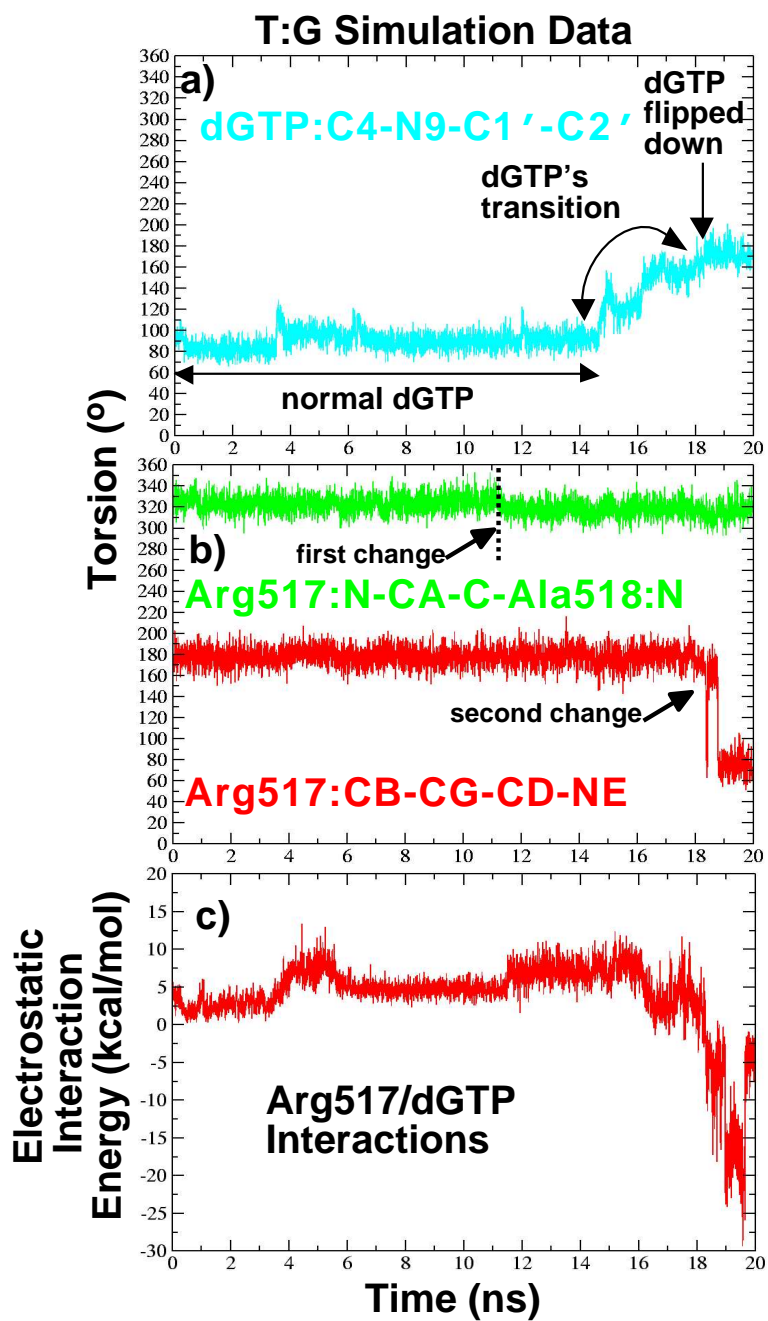


Figure S3:

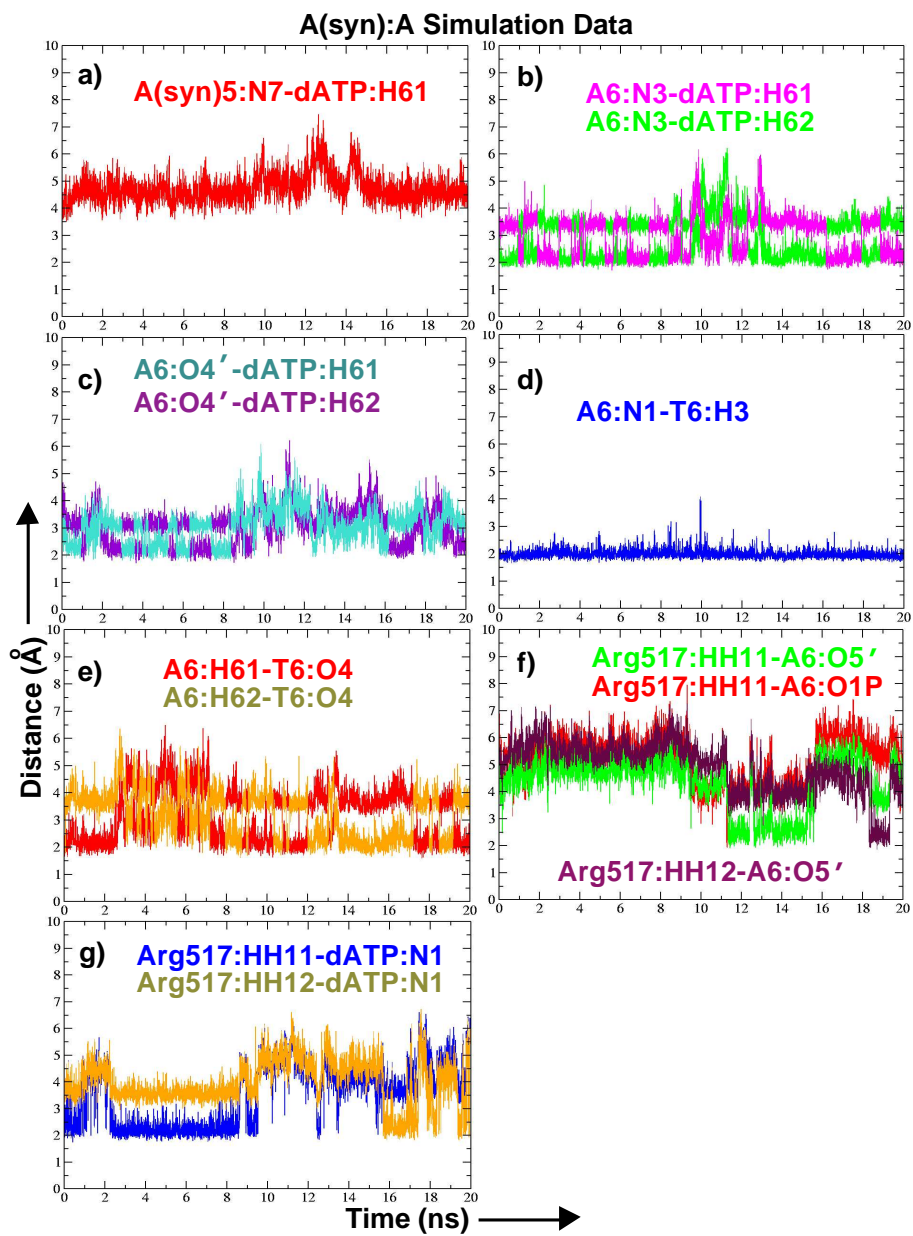


Figure S4:

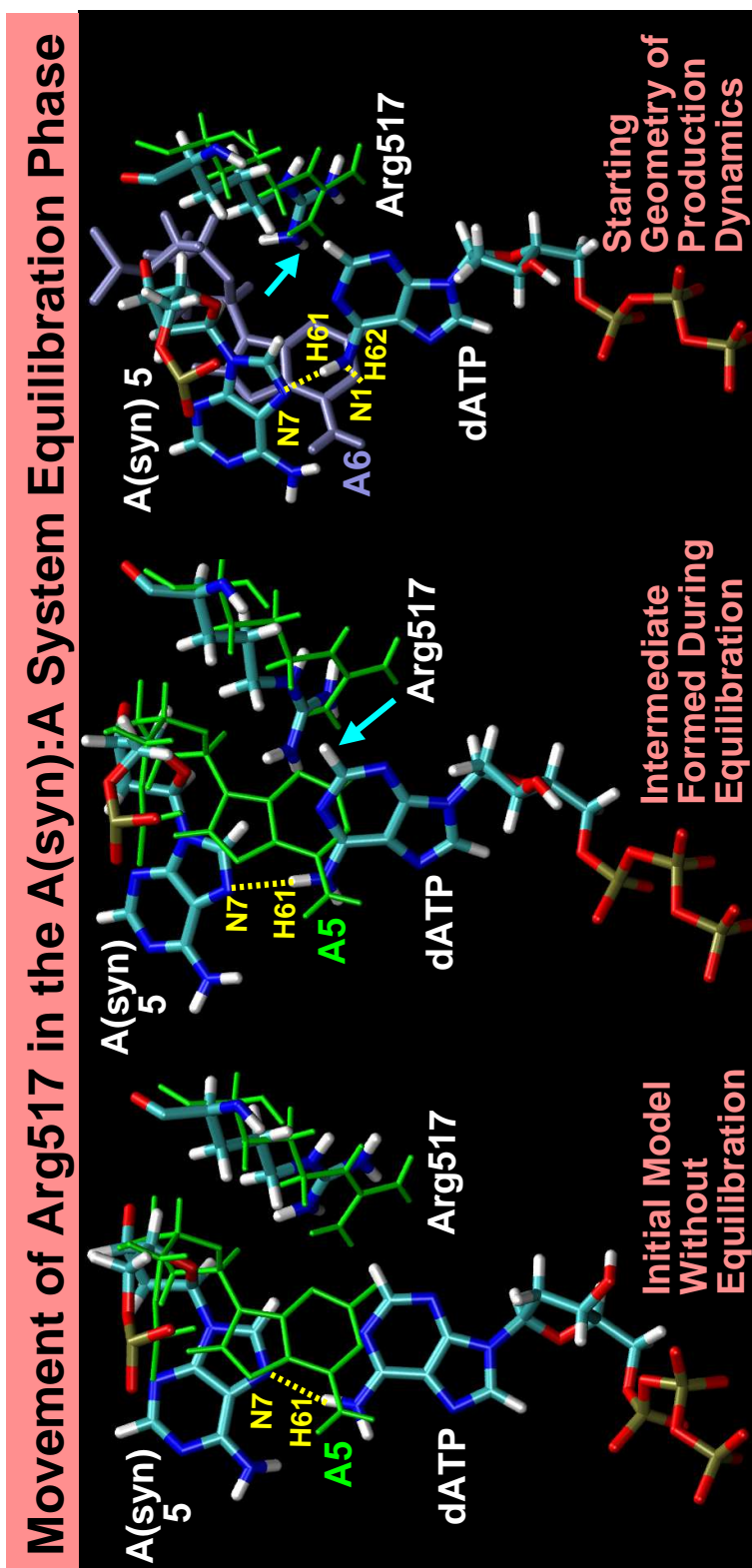


Figure S5:



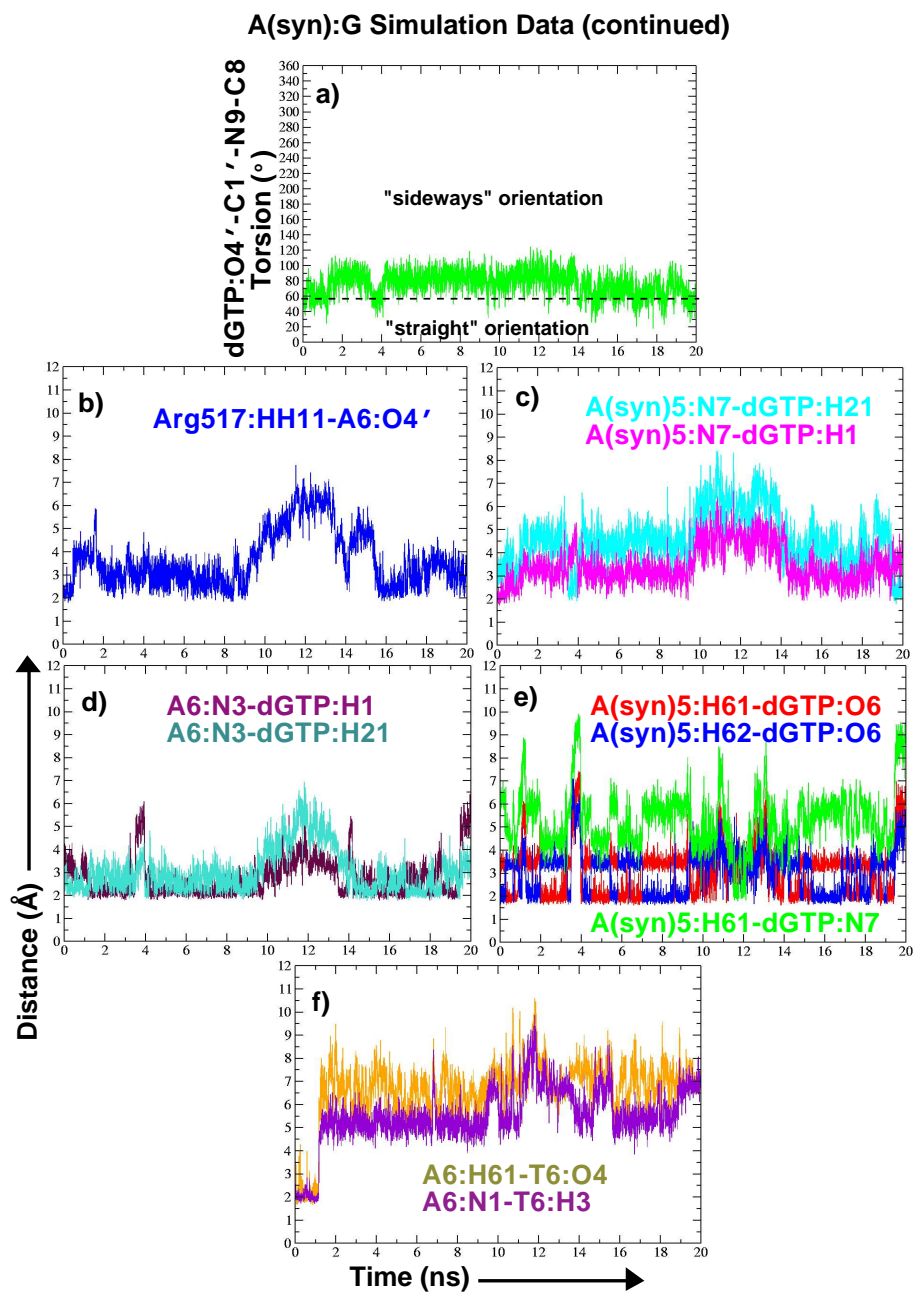


Figure S6:

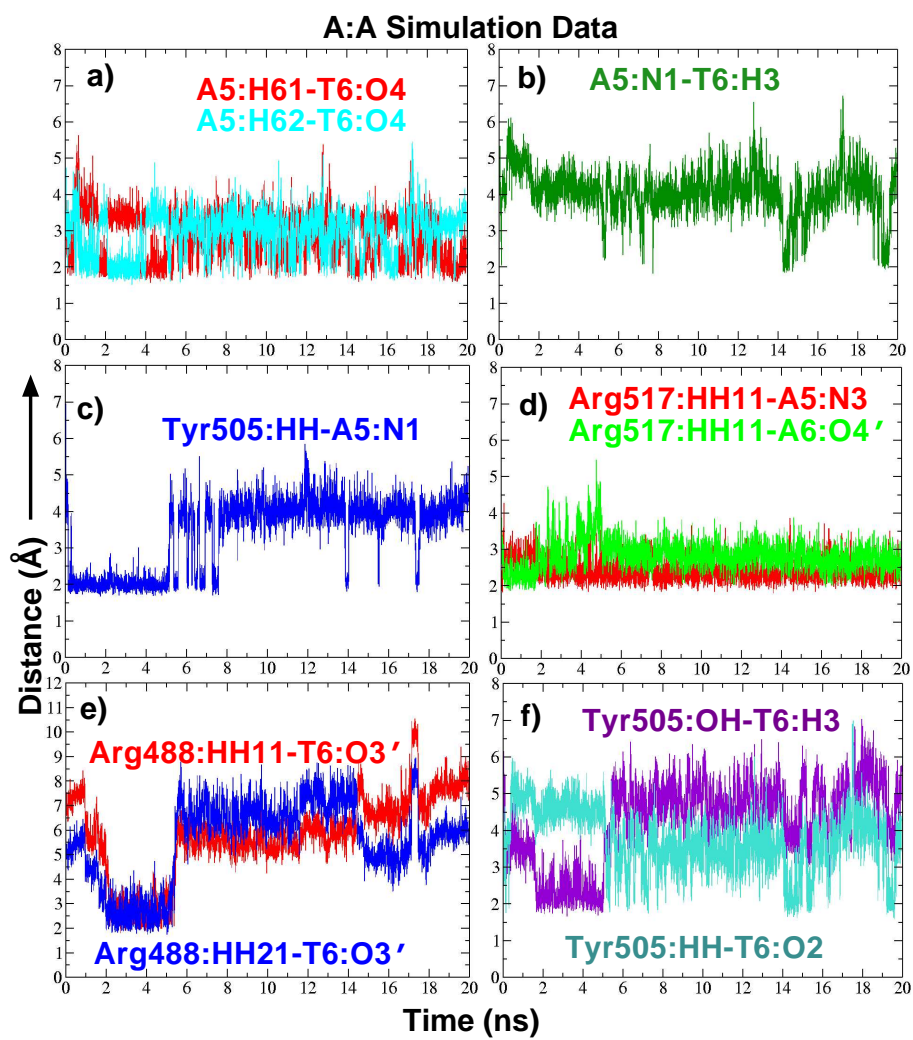


Figure S7:

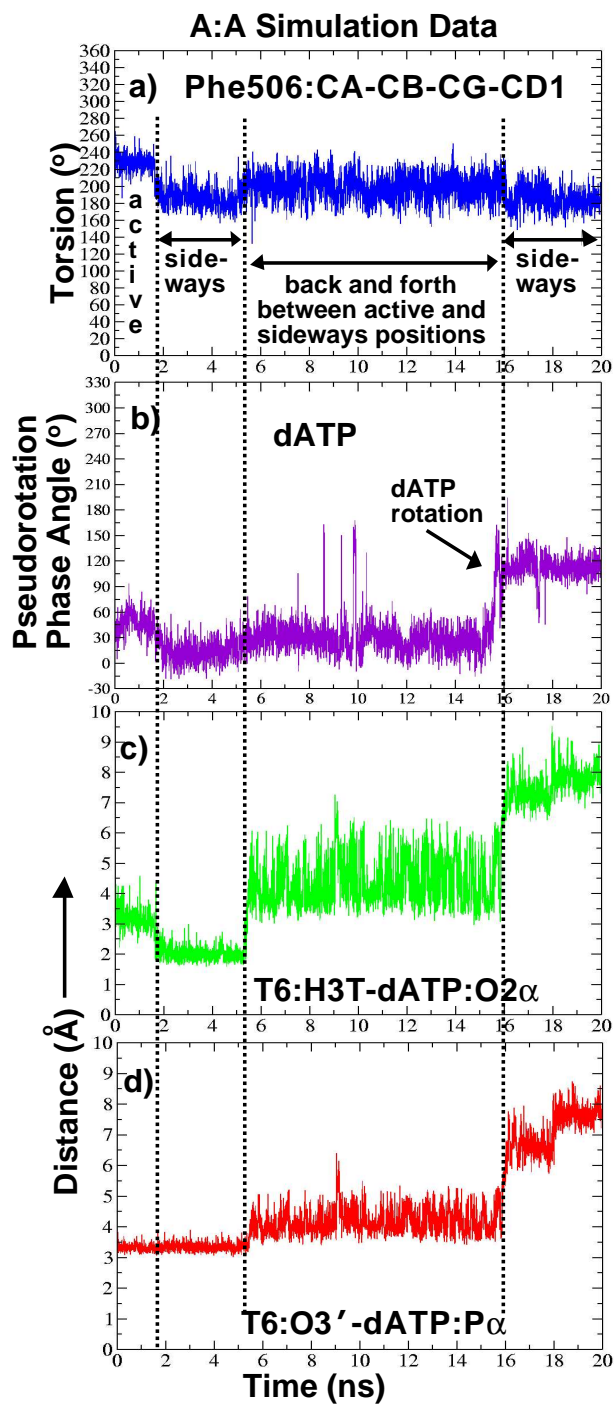


Figure S8:

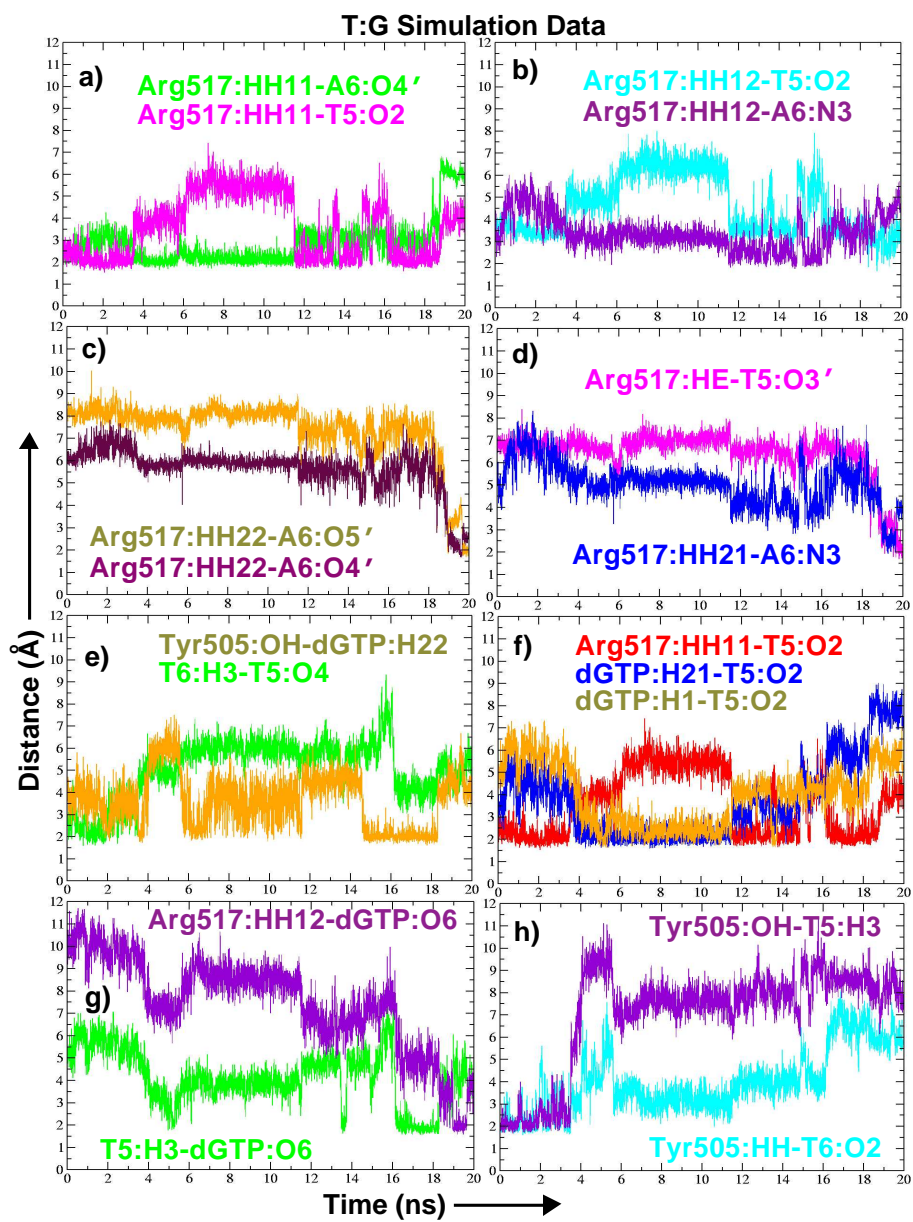


Figure S9:

### A(syn):G Simulation Data

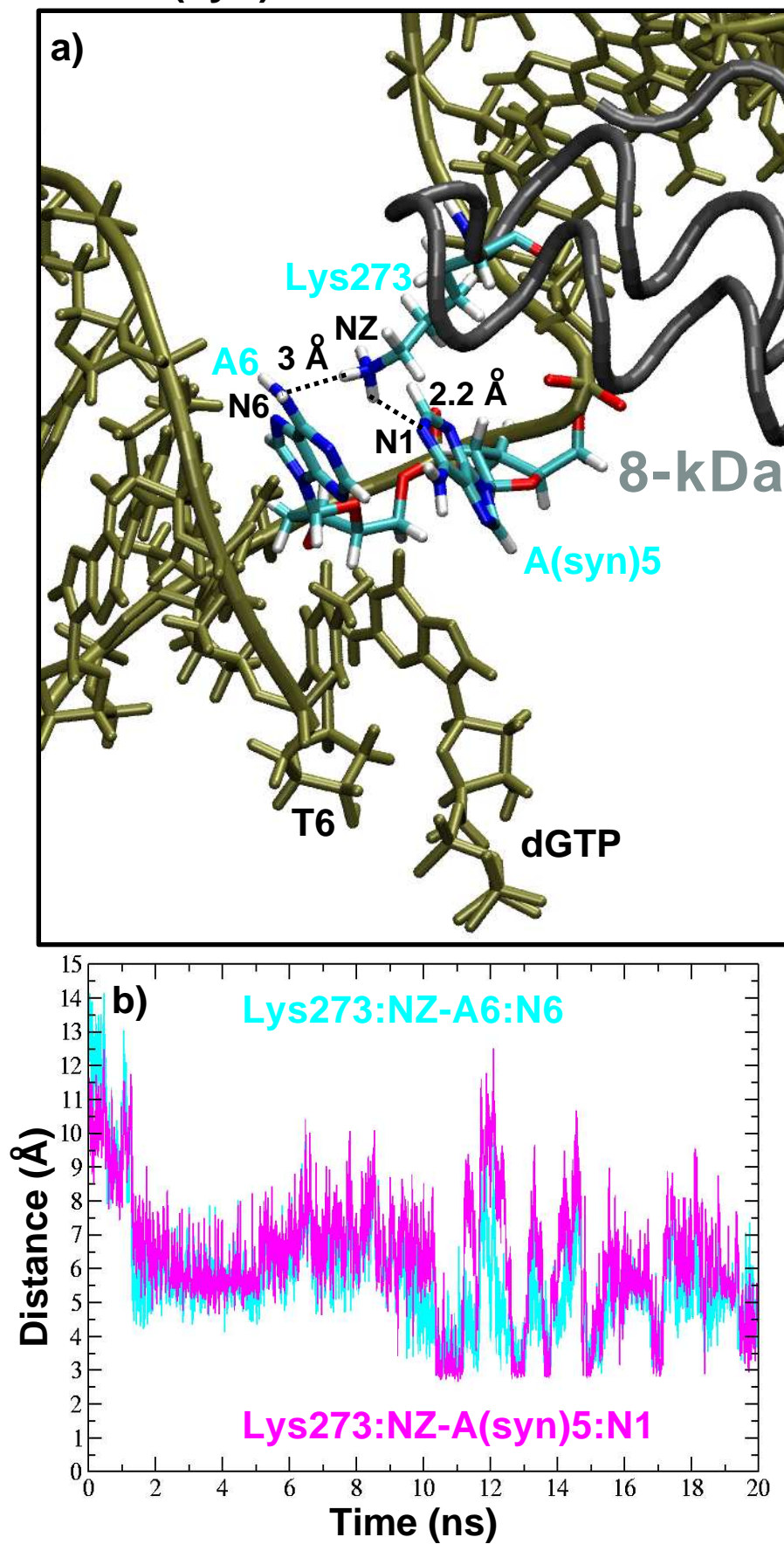


Figure S10:



## A(syn):A Simulation Data

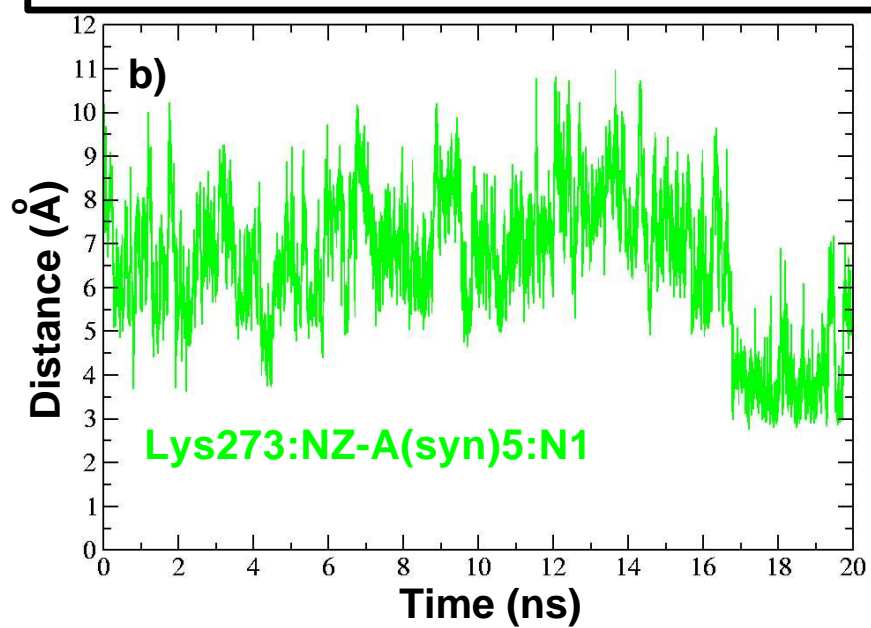
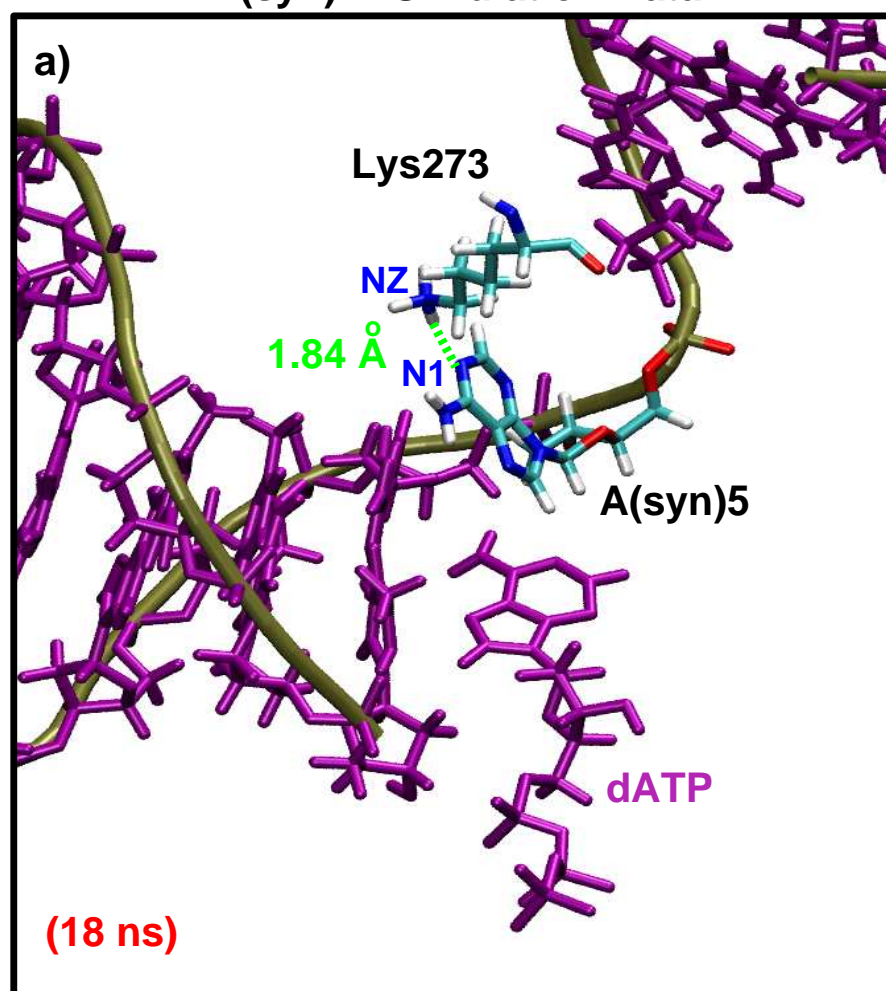


Figure S11:

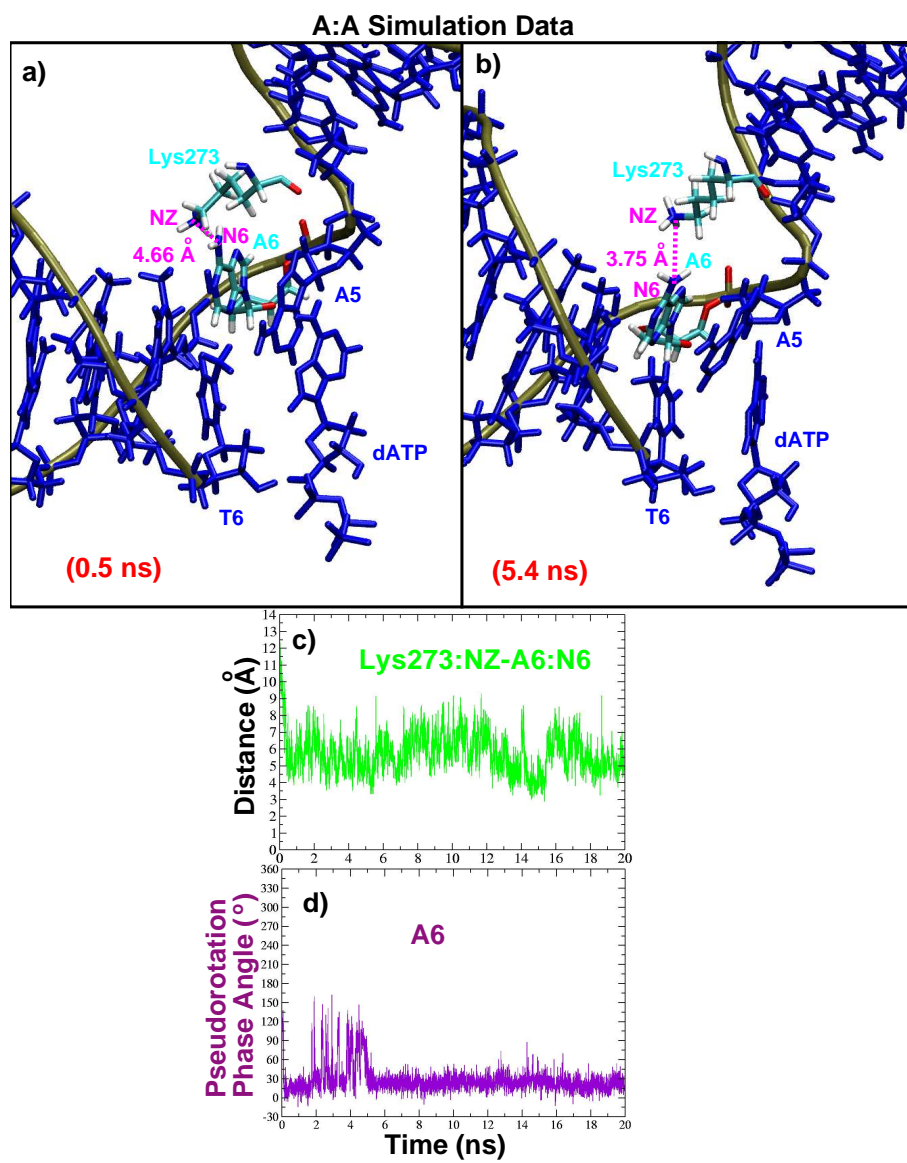


Figure S12:

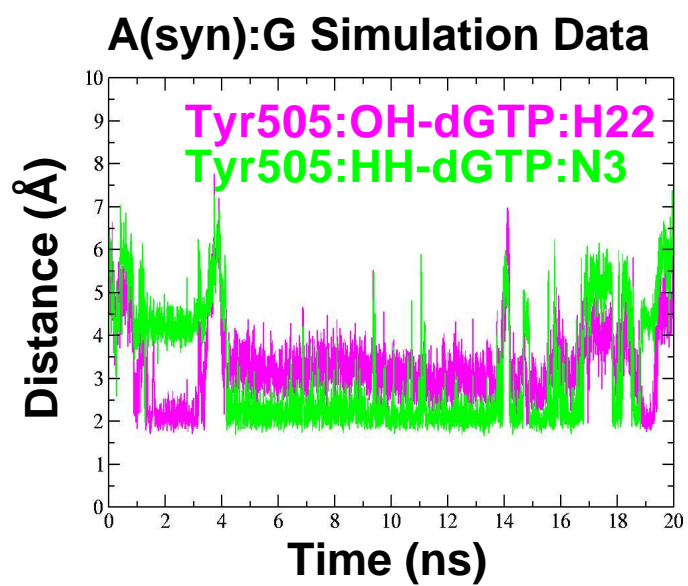


Figure S13:



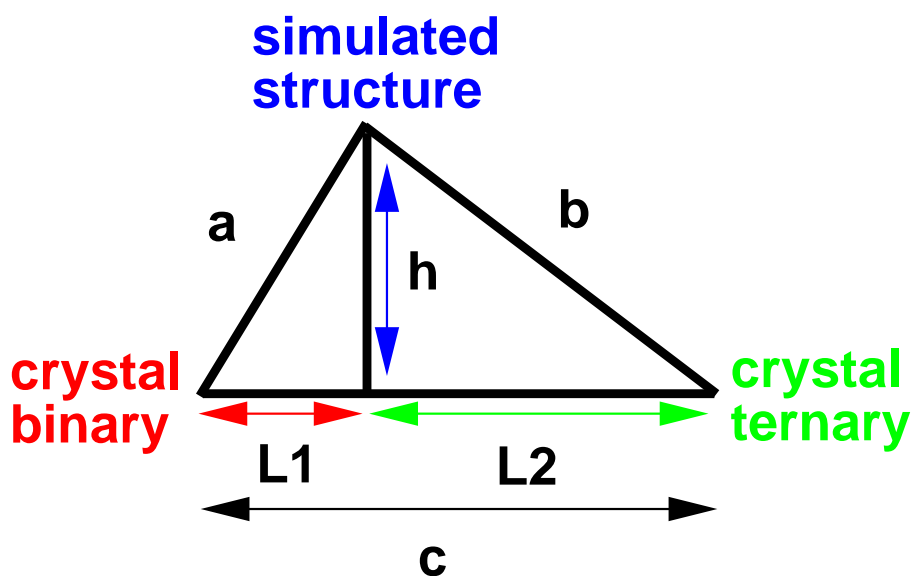


Figure S14:

## Hydrogen Bonding Changes in Mismatch Simulations

### A:G:

In the A:G stacking arrangement, the dGTP forms two hydrogen bonds to the DNA template strand backbone near the templating adenine, A5 (refer to Sup. Fig. S2d for distance data).

### T:G:

The 5 ns simulation snapshot of Figure 3c shows that the T:G base pair is formed by a bifurcated hydrogen bond between T5:O2 and dGTP:H21 and H1 atoms and a second, less stable, hydrogen bond formed between the T5:H3 and dGTP:O6 atoms (refer to Sup. Fig. S9f,g for distance data). Water molecules also frequently bridge interactions between the bases (e.g., between dGTP:O6 and both T5:H3 and T5:O4) as indicated in Figure 3c in the simulation snapshot at 9.5 ns. Before T5 and dGTP wobble base pair, Tyr505 occasionally forms hydrogen bonds to T5 and T6 (Sup. Fig. S9h) and T5 and T6 interact (Sup. Fig. S9e).

**A(*syn*):A:**

In the initial model, a single hydrogen bond forms between A(*syn*)5 and dATP, but it is highly unstable and disappears at the start of production dynamics. As shown in Sup. Fig. S5, Arg517 facilitates the breaking of the hydrogen bond in the equilibration phase by moving partially into A(*syn*)5's position. Arg517's movement pushes A(*syn*)5 temporarily upward (Sup. Fig. S5, center). The dATP then reorients in the active site to form hydrogen bonds simultaneously to both A(*syn*)5 and A6, the adjacent template base (Sup. Fig. S5, right). During production dynamics, dATP loses its hydrogen bond to A(*syn*)5, but continues to interact with A6 (refer to Sup. Fig. S4a,b,c for distance data). Most unexpectedly, the primer terminus base pair remains intact (Sup. Fig. S4d,e) despite the extra hydrogen bond formed between A6 and dATP.

**A(*syn*):G:**

In the minor straight orientation, dGTP (H1 or H21 atoms) can form a hydrogen bond to A(*syn*)5:N7 (refer to Sup. Fig. S6c and other parts for distance data). The major sideways orientation of dGTP tilts the base into the upstream DNA and facilitates hydrogen bonding to A6:N3 (Figure 5a, 8 ns snapshot, and Sup. Fig. S6d). Also, while in the sideways orientation, a third hydrogen bond forms between the dGTP:N7 and A(*syn*)5:H61 atoms briefly around 12 ns (Sup. Fig. S6e). Regardless of the orientation, dGTP:O6 forms a hydrogen bond to the A(*syn*)5:H61 or H62 atoms (Sup. Fig. S6e). The primer terminus base pair breaks after  $\sim 1$  ns (Sup. Fig. S6f).

**A:A:**

The A:dATP bases stack as shown in Figure 5b. The primer terminus base pair does not hydrogen bond because the template base (A6) lifts partially out of the helix (Figure 5b, 0.08 ns snapshot). As Tyr505 interacts with A5, it wedges between A5 and dATP, making it harder for dATP to form a

hydrogen bond to A5. Tyr505 also interacts with the primer terminus (T6) (Sup. Fig. S7f).

## Comparison of Mismatch Hydrogen Bonding in Pol $\lambda$ with Experimental Data:

To further test our conclusions from the pol  $\lambda$  mismatch studies, we compared the simulated mismatch geometries to available structural information regarding mismatches in free DNA and within the context of a DNA polymerase active site. In free DNA under physiological conditions, A:C mismatches occur in both protonated ( $A^+ : C$ ) and neutral ( $A : C$ ) forms with both bases in the *anti* orientation (1, 2, 3, 4, 5, 6, 7). The protonated  $A^+ : C$  favors a wobble base pair with at least two hydrogen bonds between the bases, while neutral  $A : C$  may alternate between wobble and reverse wobble base pairs with only one hydrogen bond between the bases (6, 7). Our studies of pol  $\lambda$  bound to A:dCTP, modeled in the neutral form, reveal occasional reverse wobble base pairing between A and C. This agrees with experimental data which indicate that the neutral base pair is less stable than the protonated form (8, 9). Interestingly, X-ray crystal structures of pol  $\beta$  with A:C mismatches in the active site before and after incorporation do not reveal any hydrogen bonding between the bases (10, 11).

For adenine opposite guanine, most structural studies reveal either a  $G(anti) : A(anti)$  or  $G(anti) : A(syn)$  pairing around pH 7 (12, 13, 14, 15, 16, 17, 5, 18, 19, 20, 21). We modeled both these base pair forms. Interestingly, as in our  $A(syn) : G$  system, A:G base pairs are sometimes stabilized by hydrogen bonds to adjacent DNA bases in free DNA (21, 19). In pol  $\beta$ , the G:A mismatch is not coplanar and one hydrogen bond forms between the bases (11). Simulations of pol X bound to A:G reveal stacking interactions between the bases (22) like those observed in pol  $\lambda$ . An analysis of the impact of the nearest-neighbors of a G:A mismatch on DNA duplex stability suggests that G and C bases on the 5' side of the mismatch increase the DNA stability more than A or T bases (23); our modeled DNA contained an adjacent A:T base pair, which possibly contributes to the amount of dGTP rearrangement that occurs in the  $A(syn) : G$  system.

In X-ray crystal structures of a T:G mismatch in B-DNA (24), two hydrogen bonds form between the bases. A variation of this wobble base pair is captured in the pol  $\lambda$  T:G system. As in our study, water molecules also bridge interactions between the bases in the crystal structures (24). Within the Dpo4 active site, X-ray crystal structures reveal the T:G mismatch either in a wobble base pair or staggered geometry in which the bases are not coplanar (25); this suggests that, like pol  $\lambda$ , the wobble base pair is not very stable in Dpo4's active site. A reverse wobble base pair is also captured in Dpo4 when the T:G base pair is at the primer terminus (26), however, at this position in BF, a wobble base pair is captured (27). The nature of the primer terminus base pair is important because the stacking interactions between the T:G mismatch and adjacent base pairs have been shown to influence DNA duplex stability (28, 29). In the pol  $\lambda$  simulation, the A:T primer terminus base pair provides little stability to the T:G mismatch since the bases do not stack well. Thus, the collapse of the T:G base pair in the pol  $\lambda$  simulation may be related to the poor interactions between the mismatch and the primer terminus base pair.

The experimental data suggest that the presence of an A:A mismatch in DNA is always destabilizing regardless of the sequence context (30). Within B-form DNA, NMR data indicate that both bases of an A:A mismatch are in the *anti* orientation and one hydrogen bond may form between them (31, 15, 32). We do not observe any hydrogen bonds formed between the bases in our A:A or A(*syn*):A systems. This is similar to frayed A:A mismatch in BF's active site (27). Like the pol  $\lambda$  A:A system, the A:A bases stack within the active site of binary pol  $\beta$  complexes (33); however, after the next incoming nucleotide binds, one of the adenine bases rotates to the *syn* orientation (33). An X-ray crystal structure of pol  $\lambda$  also shows a G(*syn*):G geometry at the primer terminus (34), indicating purine:purine mispairs may favor this conformation outside of the incoming nucleotide binding position.

In sum, our pol  $\lambda$  studies agree with available structural data. Hydrogen bond formation between the mismatched bases in pol  $\lambda$  appears to play an important role in improving the geometry of the

mismatch and stabilizing the dNTP. However, it is not essential for dNTP incorporation as exemplified by our A:C system where the bases rarely hydrogen bond, but pol  $\lambda$  inserts this mismatch very frequently. Pol  $\beta$  and pol X also appear similar in this respect (35, 22). Thus, these polymerases agree with nonpolar nucleoside isostere data that suggest steric effects in the polymerase active site rather than Watson-Crick hydrogen bonds are more important in determining dNTP incorporation (36). Our results suggest that electrostatics interactions between the polymerase and the incorrect dNTP are much more important for polymerase fidelity than electrostatics interactions between the mismatched bases, which affect base-base hydrogen bonding.

## Geometric Characteristics

The movement of enzymes such as pol  $\lambda$  is not always accurately represented by a RMSD measure due to the magnitude and direction of domain motions. A previously developed scheme to represent the RMSD data more accurately in such situations was used to capture the motion of the DNA and subdomains in our pol  $\lambda$  simulations (37). In this alternative approach, we project the RMSD of the simulated structure on the line joining the geometric centers of the structure in the crystal open and closed conformations. As shown in Figure S14, this can be represented by a triangle where the crystal open and closed conformations form two vertices of the triangle and the simulated structure forms the third. The lengths of side a and b of the triangle are given by the RMSD of the simulated structure relative to the crystal open conformation and crystal closed conformation, respectively, when the simulated structure and the corresponding crystal structure are superimposed with respect to all protein  $C_\alpha$  atoms. The RMSD between the crystal open and closed conformations forms side c of the triangle and is held fixed. The shift distance, h, describes the displacement of the simulated structure in the direction perpendicular to the line joining the geometric centers of the structure in the crystal open and closed conformations. When h is constant, RMSD alone is a good measure of domain motions. When h is not constant, this is an appropriate method to use to represent domain motion toward the crystal states. The variable lengths L1 and L2 correspond to the projected RMSD of the simulated structure with respect to the crystal open and closed conformations, respectively.

## References

- [1] Patel, D. J.; Kozlowski, S. A.; Ikuta, S.; Itakura, K. *Biochemistry* **1984**, *23*, 3218–3226.
- [2] Hunter, W. N.; Brown, T.; Anand, N. N.; Kennard, O. *Nature* **1986**, *320*, 552–555.
- [3] Hunter, W. N.; Brown, T.; Kennard, O. *Nucleic Acids Research* **1987**, *15*, 6589–6606.
- [4] Gao, X.; Patel, D. J. *J. Biol. Chem.* **1987**, *262*, 16973–16984.
- [5] Wang, C.; Gao, H.; Gaffney, B. L.; Jones, R. A. *J. Am. Chem. Soc.* **1991**, *113*, 5486–5488.
- [6] Boulard, Y.; Cognet, J. A. H.; Gabarro-Arpa, J.; LeBret, M.; Sowers, L. C.; Fazakerley, G. V. *Nucleic Acids Research* **1992**, *20*, 1933–1941.
- [7] Boulard, Y.; Cognet, J. A. H.; Gabarro-Arpa, J.; LeBret, M.; Carbonnaux, C.; Fazakerley, G. V. *J. Mol. Biol.* **1995**, *246*, 194–208.
- [8] Brown, T.; Leonard, G. A.; Booth, E. D.; Kneale, G. *J. Mol. Biol.* **1990**, *212*, 437–440.
- [9] Allawi, H. T.; SantaLucia, Jr., J. *Biochemistry* **1998**, *37*, 9435–9444.
- [10] Krahn, J. M.; Beard, W. A.; Wilson, S. H. *Structure* **2004**, *12*, 1823–1832.
- [11] Batra, V. K.; Beard, W. A.; Shock, D. D.; Pedersen, L. C.; Wilson, S. H. *Mol. Cell* **2008**, *30*, 315–324.
- [12] Patel, D. J.; Kozlowski, S. A.; Ikuta, S.; Itakura, K. *Biochemistry* **1984**, *23*, 3207–3217.
- [13] Kan, L.-S.; Chandrasegaran, S.; Pulford, S. M.; Miller, P. S. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 4263–4265.
- [14] Gao, X.; Patel, D. J. *J. Am. Chem. Soc.* **1988**, *110*, 5178–5182.
- [15] Maskos, K.; Gunn, B. M.; LeBlanc, D. A.; Morden, K. M. *Biochemistry* **1993**, *32*, 3583–3595.
- [16] Carbonnaux, C.; van der Marel, G. A.; van Boom, J. H.; Guschlbauer, W.; Fazakerley, G. V. *Biochemistry* **1991**, *30*, 5449–5458.

- [17] Lane, A. N.; Jenkins, T. C.; Brown, D. J. S.; Brown, T. *Biochem. J.* **1991**, *279*, 269–281.
- [18] Brown, T.; Hunter, W. N.; Kneale, G.; Kennard, O. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 2402–2406.
- [19] Prive, G. G.; Heinemann, U.; Chandrasegaran, S.; Kan, L.-S.; Kopka, M. L.; Dickerson, R. E. *Science* **1987**, *238*, 498–504.
- [20] Nikonowicz, E. P.; Gorenstein, D. G. *Biochemistry* **1990**, *29*, 8845–8858.
- [21] Webster, G. D.; Sanderson, M. R.; Skelly, J. V.; Neidle, S.; Swann, P. F.; Li, B. F.; Tickle, I. J. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 6693–6697.
- [22] Sampoli Benitez, B. A.; Arora, K.; Balistreri, L.; Schlick, T. *J. Mol. Biol.* **2008**, *384*, 1086–1097.
- [23] Allawi, H. T.; SantaLucia, Jr., J. *Biochemistry* **1998**, *37*, 2170–2179.
- [24] Hunter, W. N.; Brown, T.; Kneale, G.; Anand, N. N.; Rabinovich, D.; Kennard, O. *J. Biol. Chem.* **1987**, *262*, 9962–9970.
- [25] Vaisman, A.; Ling, H.; Woodgate, R.; Yang, W. *EMBO J.* **2005**, *24*, 2957–2967.
- [26] Trincao, J.; Johnson, R. E.; Wolffe, W. T.; Escalante, C. R.; Prakash, S.; Prakash, L.; Aggarwal, A. K. *Nat. Struct. Mol. Biol.* **2004**, *11*, 457–462.
- [27] Johnson, S. J.; Beese, L. S. *Cell* **2004**, *116*, 803–816.
- [28] Allawi, H. T.; SantaLucia, Jr., J. *Biochemistry* **1997**, *36*, 10581–10594.
- [29] Allawi, H. T.; SantaLucia, Jr., J. *Nucleic Acids Research* **1998**, *26*, 4925–4934.
- [30] Peyret, N.; Ananda Seneviratne, P.; Allawi, H. T.; SantaLucia, Jr., J. *Biochemistry* **1999**, *38*, 3468–3477.
- [31] Arnold, F. H.; Wolk, S.; Cruz, P.; Tinoco, Jr., J. *Biochemistry* **1987**, *26*, 4068–4075.

- [32] Gervais, V.; Gognet, J. A. H.; LeBret, M.; Sowers, L. C.; Fazakerley, G. V. *Eur. J. Biochem.* **1995**, *228*, 279–290.
- [33] Batra, V. K.; Beard, W. A.; Shock, D. D.; Pedersen, L. C.; Wilson, S. H. *Structure* **2005**, *13*, 1225–1233.
- [34] Picher, A. J.; Garcia-Diaz, M.; Bebenek, K.; Pedersen, L. C.; Kunkel, T. A. *Nucleic Acids Res.* **2006**, *34*, 3259–3266.
- [35] Arora, K.; Schlick, T. *J. Phys. Chem. B* **2005**, *109*, 5358–5367.
- [36] Kool, E. T.; Sintim, H. O. *Chem. Commun.* **2006**, *35*, 3665–3675.
- [37] Arora, K.; Schlick, T. *Biophys. J.* **2004**, *87*, 3088–3099.