

Supporting Material

Replica Exchange Molecular Dynamics (REMD)

REMD is a parallel simulation technique used to enhance conformational sampling in simulations (1). It combines the idea of multiple-copy simulation, simulated annealing, and Monte Carlo methods and is one of the generalized ensemble algorithms that can perform a random walk in the energy space using non-Boltzmann weighting factors. The random walk procedure allows a simulation to pass energy barriers and sample more conformational space compared to regular dynamics.

In REMD simulations, independent simulations were conducted on a set of non-interacting replicas of the same molecular system over a range of temperatures. Systems at different temperatures can exchange configurations at fixed time intervals with a transition probability satisfying the detailed balance. The exchange takes place using the Metropolis criterion with an acceptance probability,

$$P = \exp [(\beta_i - \beta_j) (E_i - E_j)],$$

where E_i is the total potential energy of replica i at temperature T_i ; $\beta_i = 1 / (k_B T_i)$ and k_B is Boltzmann's constant.

REMD was designed to enhance sampling efficiency and a single REMD simulation can yield thermodynamic quantities at a range of temperatures. Comparison of REMD with MD has shown that REMD can enhance sampling compared to regular MD within comparable time scales (2, 3).

REMD was used in our loop modeling to scan more possible conformations. We compare the average RMSD of Loop1 in all replicas to that in the regular MD in Fig. S11. Though the loop conformation in each replica varies somewhat, the average RMSD of Loop1 in all replicas does not change much after 3 ns. Full global convergence appears to require more time, but the core loop position and fluctuations appear to contribute little to the global motion of pol μ .

Fig. S11

Principal Components Analysis (PCA)

PCA aims to decompose the motions of a trajectory into independent modes in a way that the first several modes describe most of the positional fluctuations. PCA has been widely used to study the intrinsic motions of various biological systems, including nucleic acids (4, 5) and proteins (5, 6). To describe the collective motions, a covariance matrix C of atomic fluctuations along the dynamics trajectory is constructed. The matrix elements are given by

$$C = \frac{1}{M} \sum_{k=1, M} \langle (X_k - \langle X \rangle)(X_k - \langle X \rangle)^T \rangle$$

where X_k is the coordinate vector at the k th snapshot, and $\langle X \rangle$ is the average structure from the dynamics simulation: $\langle X \rangle = \frac{1}{M} \sum_{k=1, M} X_k$. Diagonalization of covariance matrix produces the

eigenvectors and eigenvalues as entries of Λ from the decomposition:

$$V^T CV = \Lambda, \text{ or } CV_n = \lambda_n V_n, n = 1, 2, \dots, 3N$$

where Λ is the diagonal matrix with eigenvalues λ_i : $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{3N})$. Each eigenvector V_n defines the direction of motion of N atoms as an oscillation about the average structure $\langle X \rangle$.

The normalized magnitude of corresponding eigenvalue ($\lambda / \sum \lambda$) indicates the relative percentage of motion along the eigenvector V_n . If eigenvalues are arranged in a decreasing order, the first few describe the largest positional fluctuations.

We applied PCA to the heavy atoms of the DNA and protein of pol μ , except the flexible Loop2. A total of 3185 atoms were included in this analysis, resulting in 9555 principal components (PCs). Snapshots were sampled from 40 ns of each trajectory at a frequency of $\Delta t = 100$ ps. PCA was performed using the CARMA package (7).

To identify the possible large-scale motions, PCA was performed on two dynamics trajectories from simulation II and IV, both without dTTP/Mg²⁺; or PCA on pol λ (from ternary form, also with dTTP/Mg²⁺ removed) (8). The resulting eigenvalues for the first 20 modes are shown in Fig. S7. Compared to pol λ , there is less (if any) large-scale motion in pol μ , suggested by smaller eigenvalues. Global motions corresponding to the top two principal components of each trajectory are shown in Fig. S12. Both the DNA and protein atoms in the pol μ complex are virtually superimposable, as is evident from distances analyzed (base A5 and A6 around the gap in the DNA template).

Fig. S12

MM-PBSA Calculations

We use the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) method (9-11) to compute the binding free energy between the DNA and protein in pol μ , and compare it to such binding free energy in pol λ (12). To use MM-PBSA to calculate the binding free energy, we remove waters, ions and the incoming nucleotide from the trajectory of our 120 ns MD simulation of pol μ (simulation I). All calculations are performed using the CHARMM program (version c35b2) (13) with the CHARMM force field (14-16).

In the MM-PBSA approach, the binding free energy is calculated from the free energy difference between the polymerase/DNA complex and the two unbound components of the complex. The free energy of the complex, protein, and DNA is calculated separately according to:

$$G = E_{MM} + G_{solv} - TS,$$

where E_{MM} is the molecular mechanical energy, G_{solv} is the solvation energy, and $-TS$ is the solute entropic contribution. E_{MM} is the sum of the internal (e.g., bonds, angles, dihedrals), van der Waals, and electrostatic components. G_{solv} is the sum of the hydrophobic energy (G_{np}) and the electrostatic solvation energy (G_{PB}). The nonpolar contribution to the solvation free energy is calculated from:

$$G_{np} = \gamma SASA + \beta,$$

where $\gamma = 0.00542 \text{ kcal}/\text{\AA}^2$, $\beta = 0.92 \text{ kcal/mol}$, and SASA is the solvent accessible surface area, which is determined using a water probe radius equal to 1.4 \AA . The polar contribution to the solvation free energy (G_{PB}) is determined by solving the finite-difference Poisson-Boltzmann equation using the PBEQ module in CHARMM (17); these calculations were run at 300 K with a solute dielectric constant of 1.0, solvent dielectric constant of 80, reference gas phase dielectric constant of 1.0, water probe radius of 1.4 \AA , monovalent salt concentration of 0.15 M, and ion exclusion radius of 2 \AA . A two-step focusing procedure was used with an initial grid spacing of 0.4 \AA and a final grid spacing of 0.1 \AA . A maximum of 1000 iterations were used for each calculation. The entropy was approximated from quasiharmonic analysis of the MD trajectories using the VIBRAN module in CHARMM (18, 19).

Once the free energy of each species is calculated, we compute the binding free energy (ΔG) from:

$$\Delta G_{\text{bind}} = G(\text{complex}) - G(\text{protein}) - G(\text{DNA})$$

Since these calculations are based on our MD trajectories, averages are obtained for each free energy term. To monitor the convergence of the free energy calculations, we analyzed the data from 10 ~ 40 ns and 10 ~ 120 ns of simulation I, by 1500 frames (every 20 ps) and 2200 frames (every 50 ps) respectively. As shown in Table S2, the computed free energies from the first part and the total of simulation are similar. For the quasiharmonic analysis, a total of 22,000 frames were used, spanning 10 ~ 120 ns with 5 ps frame spacing.

Reference:

1. Mitsutake, A., Y. Sugita, and Y. Okamoto. 2001. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers* 60:96-123.
2. Zhang, W., C. Wu, and Y. Duan. 2005. Convergence of replica exchange molecular dynamics. *J. Chem. Phys.* 123:154105.
3. Sanbonmatsu, K. Y., and A. E. Garcia. 2002. Structure of Met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. *Proteins* 46:225-234.
4. Qian, X., D. Strahs, and T. Schlick. 2001. Dynamic simulations of 13 TATA variants refine kinetic hypotheses of sequence/activity relationships. *J. Mol. Biol.* 308:681-703.
5. Strahs, D., D. Barash, X. Qian, and T. Schlick. 2003. Sequence-dependent solution structure and motions of 13 TATA/TBP (TATA-box binding protein) complexes. *Biopolymers* 69:216-243.
6. Arora, K., and T. Schlick. 2004. In silico evidence for DNA polymerase- β 's substrate-induced conformational change. *Biophys. J.* 87:3088-3099.
7. Glykos, N. M. 2006. Software news and updates. Carma: a molecular dynamics analysis program. *J. Comput. Chem.* 27:1765-1768.
8. Foley, M. C., K. Arora, and T. Schlick. 2006. Sequential side-chain residue motions transform the binary into the ternary state of DNA polymerase λ . *Biophys. J.* 91:3182-3195.
9. Chong, L. T., Y. Duan, L. Wang, I. Massova, and P. A. Kollman. 1999. Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7. *Proc. Natl. Acad. Sci. U. S. A.*

- 96:14330-14335.
10. Gorfe, A. A., and I. Jelesarov. 2003. Energetics of sequence-specific protein-DNA association: computational analysis of integrase Tn916 binding to its target DNA. *Biochemistry* 42:11568-11576.
 11. Adcock, S. A., and J. A. McCammon. 2006. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem Rev* 106:1589-1615.
 12. Foley, M. C., V. A. Padow, and T. Schlick. 2010. DNA pol λ 's extraordinary ability to stabilize misaligned DNA. *J. Am. Chem. Soc.* 132:13403-13416.
 13. Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* 4:187-217.
 14. MacKerell, A. D., and N. K. Banavali. 2000. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comput. Chem.* 21:105-120.
 15. MacKerell, A. D., D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102:3586-3616.
 16. Mackerell, A. D., Jr., M. Feig, and C. L. Brooks, 3rd. 2004. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25:1400-1415.
 17. Im, W., D. Beglov, and B. Roux. 1998. Continuum Solvation Model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Comput. Phys. Commun.* 111:59-75.
 18. Janezic, D., R. M. Venable, and B. R. Brooks. 1995. Harmonic-Analysis of Large Systems .3. Comparison with Molecular-Dynamics. *J. Comput. Chem.* 16:1554-1566.
 19. Andricioaei, I., and M. Karplus. 2001. On the calculation of entropy from covariance matrices of the atomic fluctuations. *J. Chem. Phys.* 115:6289-6292.

Supplementary Tables

	Pol μ	Pol μ	Pol λ
	(data from 10-40 ns)	(data from 10-120 ns)	(data from 10-20 ns)
G(complex)	-6799.6	-6782.0	-5806.8
G(protein)	-5251.3	-5280.5	-4431.5
G(DNA)	-861.1	-822.3	-750.7
ΔG_{bind}	-687.2	-679.2	-624.6

Table S1. Energetic analysis of the formation of pol μ and pol λ complexes.

All energetic values are in kcal/mol.

Distance (Å)	pol λ	pol β	pol μ	pol μ (crystal structure)	pol X
dTTP(P α) - P6(O3')	6.14	4.2	5.12	3.59	~5
Mg ²⁺ (A) - Mg ²⁺ (B)	4.35	5.66	4.14	4.02	N/A
Mg ²⁺ (A) - Asp-I(OD2)	1.82	1.78	1.81	1.76	Y
Mg ²⁺ (A) - Asp-II(OD1)	1.85	1.84	1.84	1.77	Y
Mg ²⁺ (A) - Asp-III(OD2)	1.92	1.83	1.79	1.77	Y
Mg ²⁺ (A) - Asp-III(OD1)	1.91	N/A	3.95	3.90	
Mg ²⁺ (A) - WAT1(OH2)	2.10	2.11	2.11		Y
Mg ²⁺ (A) - WAT2(OH2)	1.98		1.90		Y
Mg ²⁺ (A) - dNTP(O1A)	3.85	1.84	3.43	3.26	Y
Mg ²⁺ (A) - P6(O3')	4.73	4.73	4.25	1.94	
Mg ²⁺ (B) - Asp-I(OD1)	1.84	1.78	1.87	1.84	Y
Mg ²⁺ (B) - Asp-I(O)	4.03	N/A	4.06	3.82	
Mg ²⁺ (B) - Asp-II(OD2)	1.88	1.80	1.89	1.86	Y
Mg ²⁺ (B) - dNTP(O1B)	1.92	1.83	2.00	1.92	Y
Mg ²⁺ (B) - dNTP(O1A)	1.90	3.98	1.84	1.89	

Mg ²⁺ (B) - dNTP(O3G)	1.85	1.89	1.81	1.82	Y
Mg ²⁺ (B) - WAT3(OH2)	2.07	2.00	2.11	2.04	Y
Mg ²⁺ (B) - WAT4(OH2)		2.02			Y

Table S2. Geometry around Mg²⁺ in the active site.

* WAT, water molecules; dNTP, incoming nucleotide; P6, upstream primer; Mg²⁺(A), catalytic ion; Mg²⁺(B), nucleotide-binding ion. Different atom names may apply (e.g., in pol X and pol β , the residue number of primer terminus is 10 instead of 6). Asp-I/II/III corresponds to the three aspartates at the active site, in the sequence of residue number, namely Asp^{427/429/490} for pol λ , Asp^{190/192/256} for pol β , Asp^{330/332/420} for pol μ , and Asp^{49/51/100} for pol X.

† Distances in red correspond to coordination between the given Mg²⁺ and residue atom. In pol X, specific distances were not mentioned, though the atoms coordinating with Mg²⁺ were indicated (shown as “Y”).

Supplementary Figure Legends

Fig. S1. Open-closed transition of pol β and pol λ revealed in crystal structures.

(a) Crystallographic pol β conformations before chemistry in closed (PDB entry 1BPY, *left*) and open (PDB entry 1BPX, *right*) forms.

(b) Superimposed C α trace of crystallographic pol λ conformations before chemistry in closed (PDB entry 1XSN, *red*) and open (PDB entry 1XSL, *green*) forms. Pol λ displays a shift of DNA template and templating base at the gap (*right*).

Fig. S2. Protein motions (*left*) and DNA motions (*right*) of pol μ in simulations I-X (*red*, simulation I; *green*, simulation II; *blue*, simulation III; *cyan*, simulation IV; *black*, simulation V; *gray*, simulation VI; *pink*, simulation VII; *tan*, simulation VIII; *purple*, simulation IX; *yellow*, simulation X).

Fig. S3. Motions of protein and DNA revealed by PCA. Global DNA and protein motions according to the first two PCs of pol λ , pol μ in simulation II and IV, all without substrate. Ten frames are taken at equally spaced intervals. Bases A5 and A6 in the template DNA are indicated by arrows, with the distance of their movement shown.

Fig. S4. Active site coordination of pol μ in simulations XVI and XVII.

(a) Conformation of His329 changed in simulations XVI and XVII (N δ to N ϵ).

(b) Active site rearrangement of pol μ in simulations XVI (*red*) and XVII (*green*). Dashed lines indicate hydrogen bonds formed between His³²⁹ and dTTP.

Fig. S5. Dihedral angles of (a) His³²⁹:CG - His³²⁹:CB - His³²⁹:CA - His³²⁹:C and (b) His³²⁹:C - Asp³³⁰:N - Asp³³⁰:CA - Asp³³⁰:CB in simulations I-VIII. Simulations without dTTP (simulation II, IV, VI, and VIII) are drawn with dashed lines. In simulation VII (larger damping coefficient), His³²⁹ flips through an intermediate conformation, while Asp³³⁰ stays in its original conformation.

Fig. S6. The flip of His³²⁹ through an intermediate state observed in simulation VII: (1) closed, 1 ns; (2) intermediate, 20 ns; (3) open, 40 ns.

Fig. S7. Conformations of Glu⁴⁴³ and Gln⁴⁴⁰ in simulations I-VIII (*red*, simulation I; *green*, simulation II; *blue*, simulation III; *cyan*, simulation IV; *black*, simulation V; *gray*, simulation VI; *pink*, simulation VII; *tan*, simulation VIII). Simulations without dTTP and Mg²⁺, namely simulation II, IV, VI and VIII, are drawn with a smaller bond radius. All structures are from the last frame of individual 40 ns simulation, superimposed along the backbone of protein with the starting structure used in simulation I. dTTP and Gly⁴³⁵-Trp⁴³⁶ in simulation I are also shown red.

Fig. S8. Time evolution of thumb (pol μ : residues 451, 448, 447 and 444; pol λ : residues 544, 538, 521, 517 and 514)/DNA (T5 - T8) interaction energy divided into van der Waals (VdW) and electrostatics energy components. Energy averages and standard deviations (SD) are provided.

Fig. S9. Comparison of dTTP-A5/A6 distance in simulation I (WT) and simulation XIII (R448A). A5 is originally paired with dTTP, while A6 is originally paired with the upstream primer terminus (T17).

Fig. S10. Distances between the $C\alpha$ atoms of Gln³⁷⁹ in Loop1 and Pro⁴⁶⁶ in the thumb loop in simulations I-VIII and XI-XIII.

Fig. S11. RMSD of Loop1 in REMD and regular MD in loop modeling.

Fig. S12. Cumulative contribution of the top 20 PCs to the total motion.

Supplementary Figures

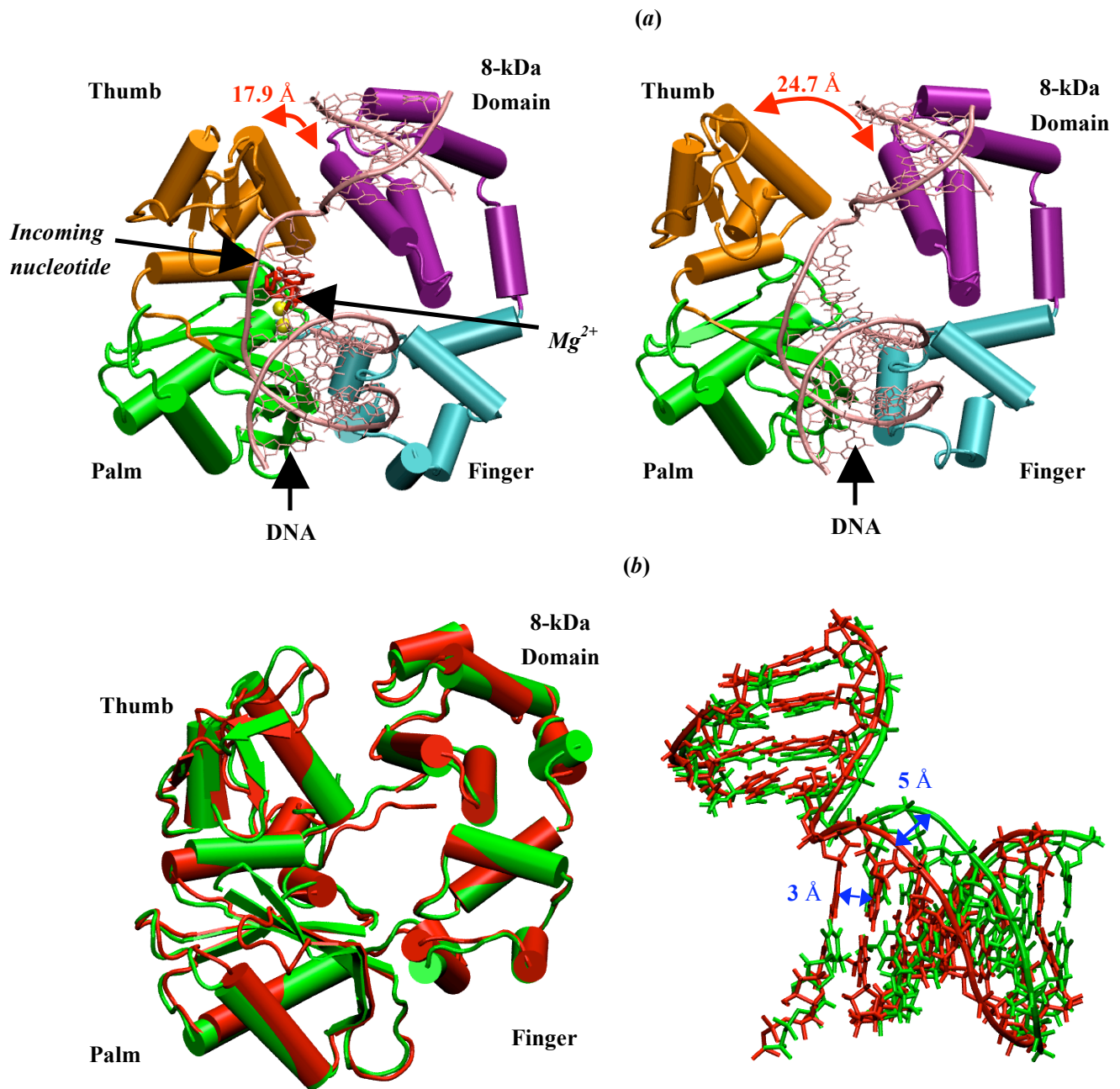
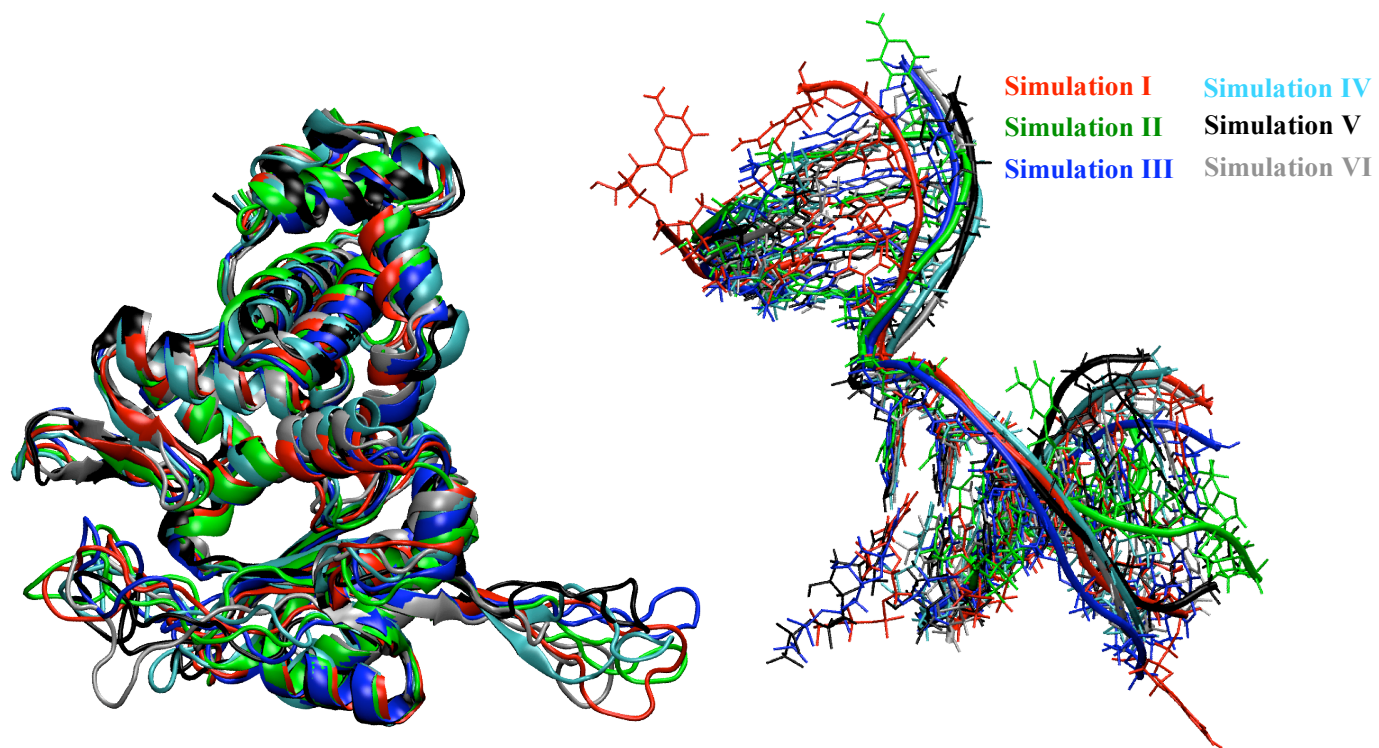


FIGURE S1

(a)



(b)

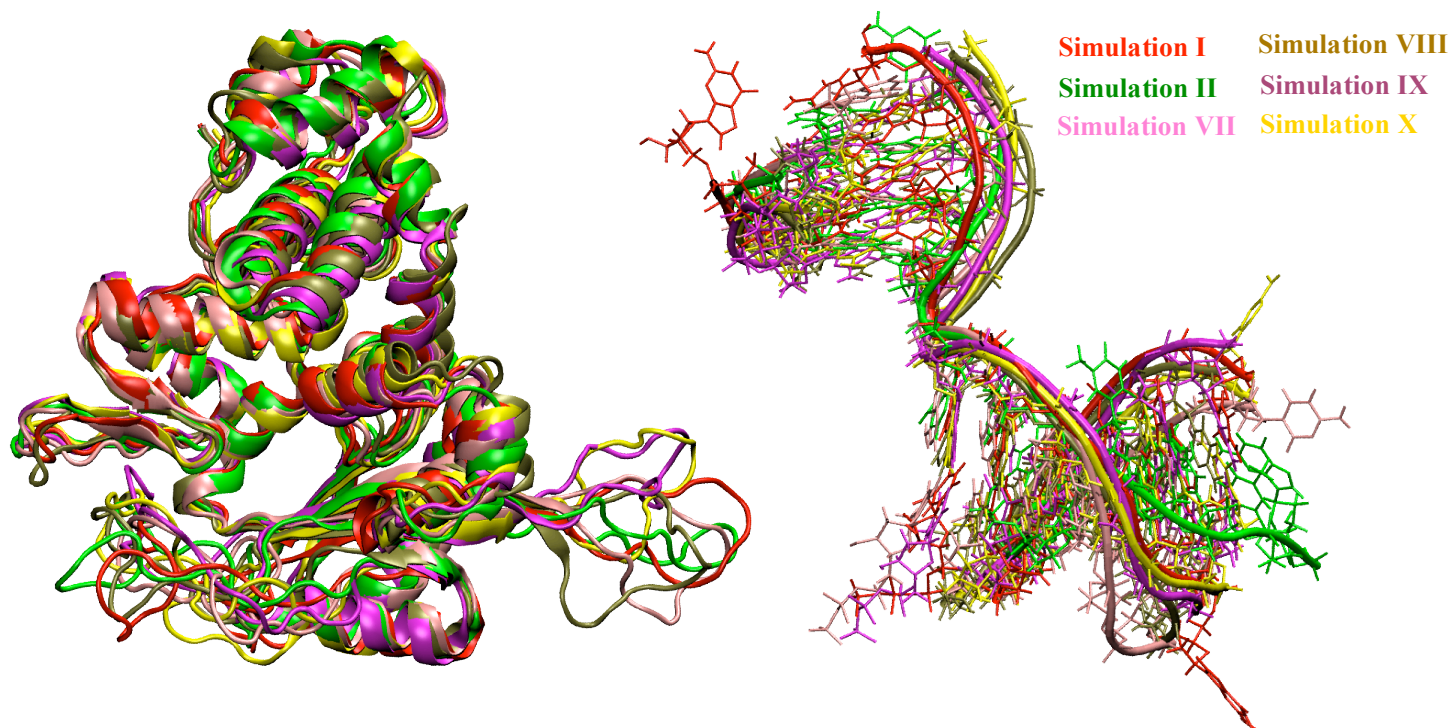


FIGURE S2

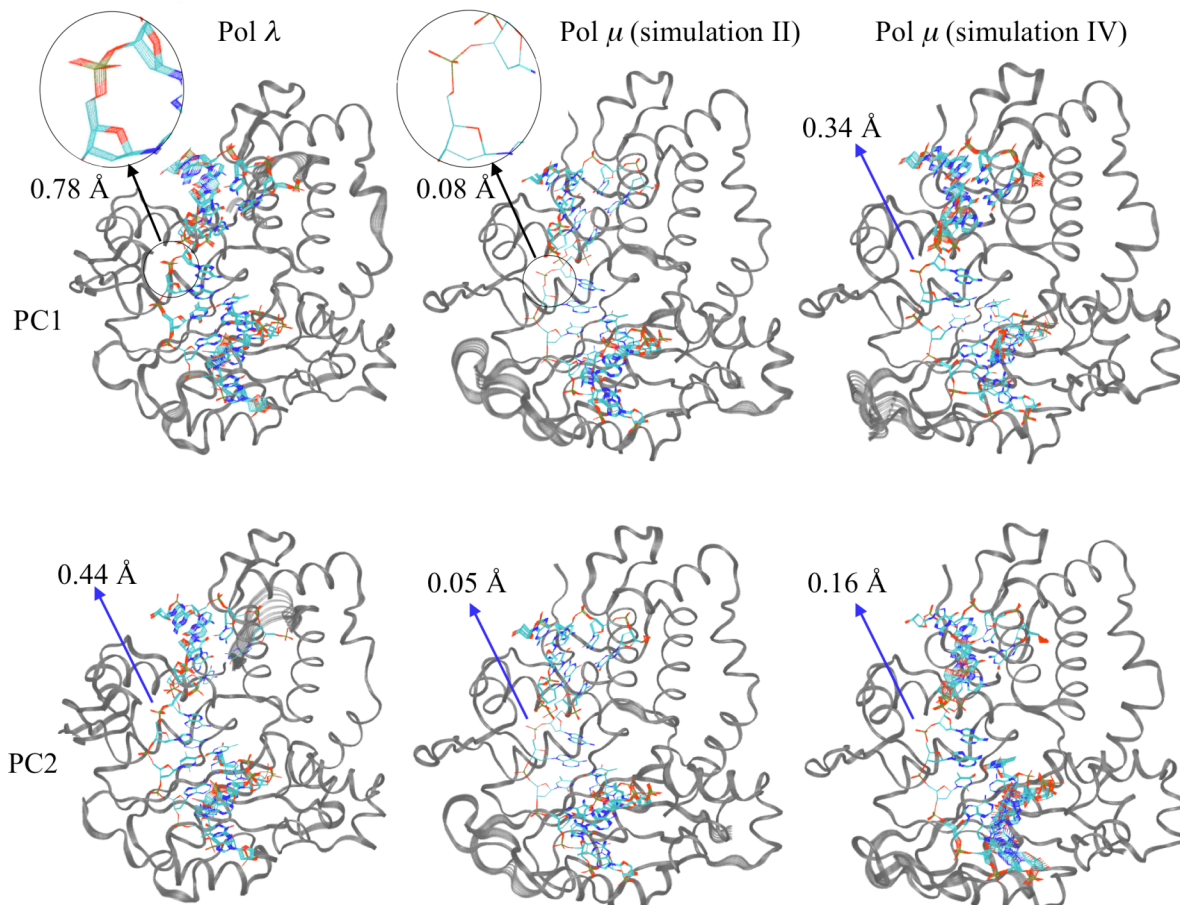


FIGURE S3

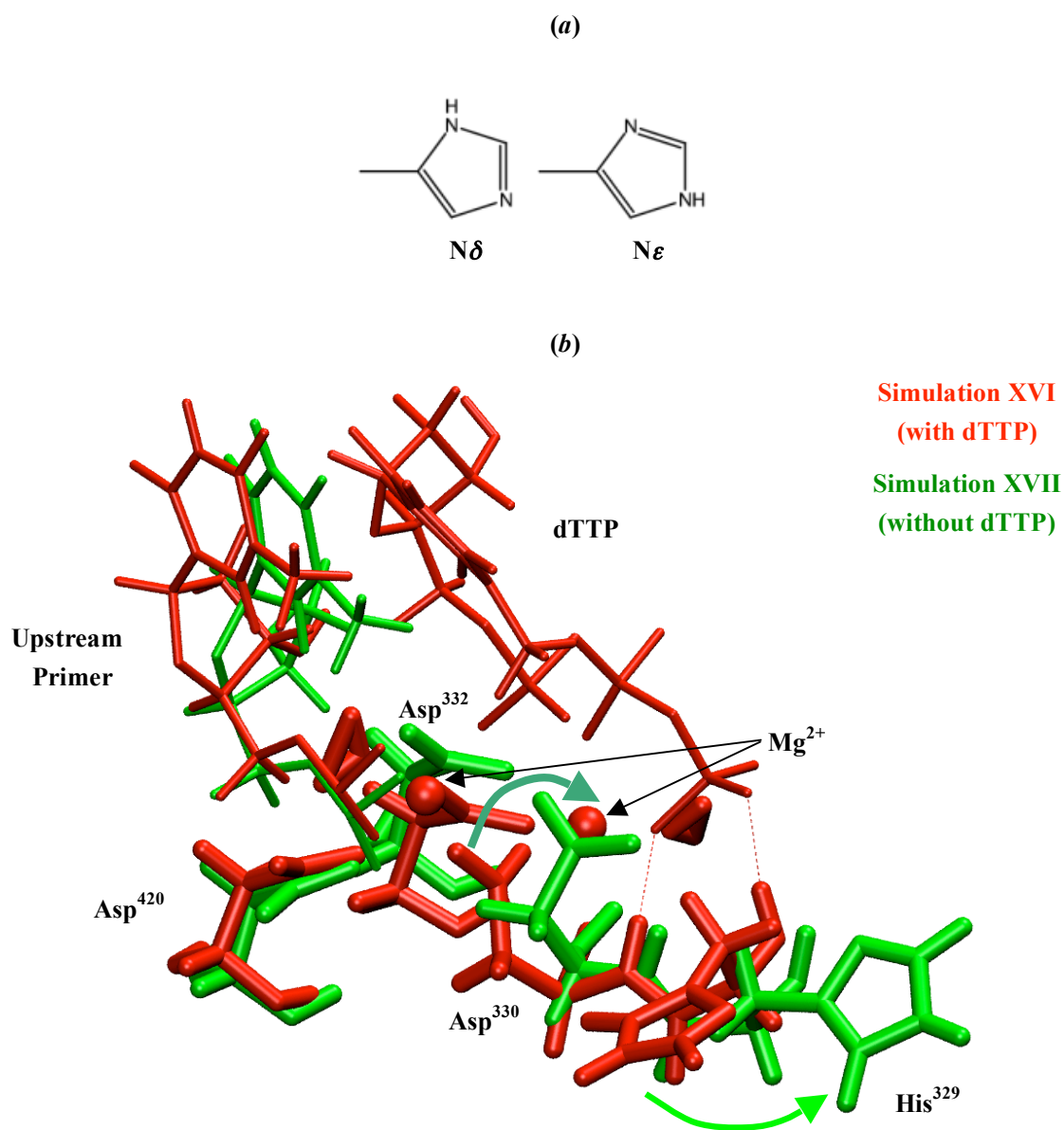


FIGURE S4

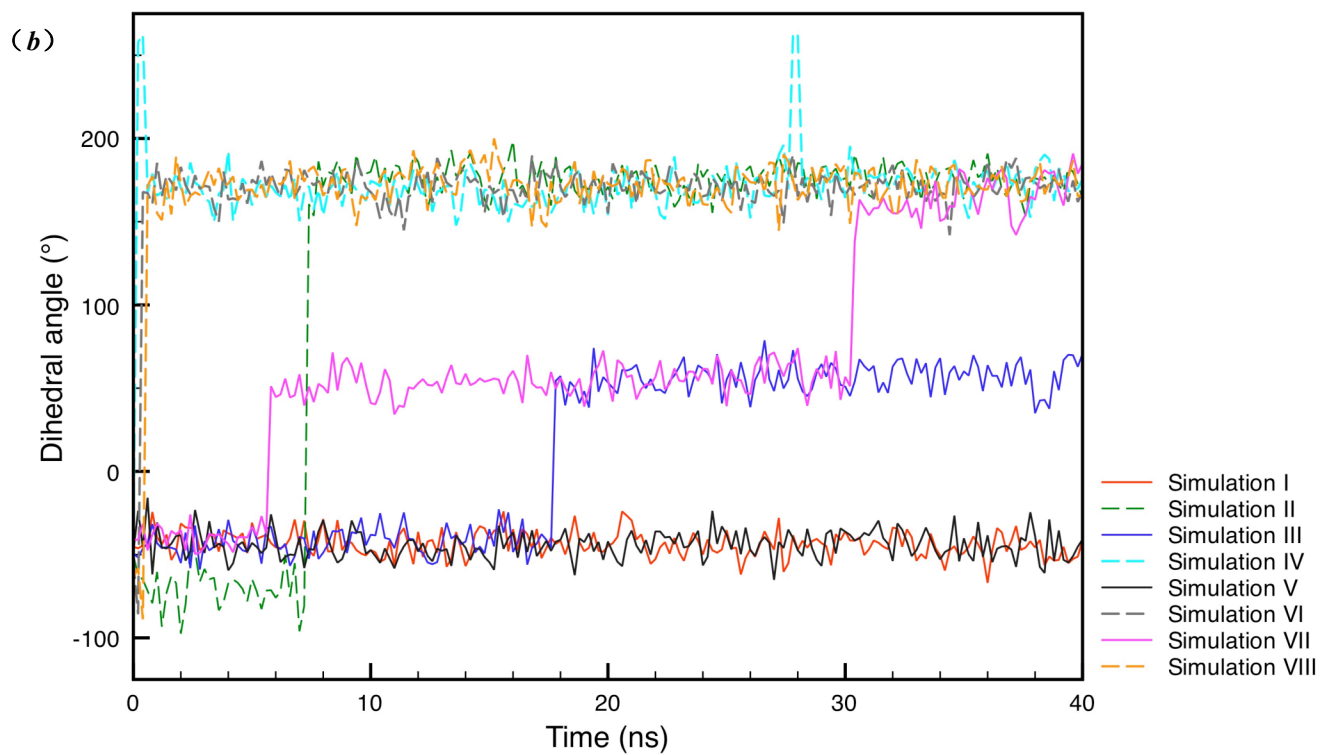
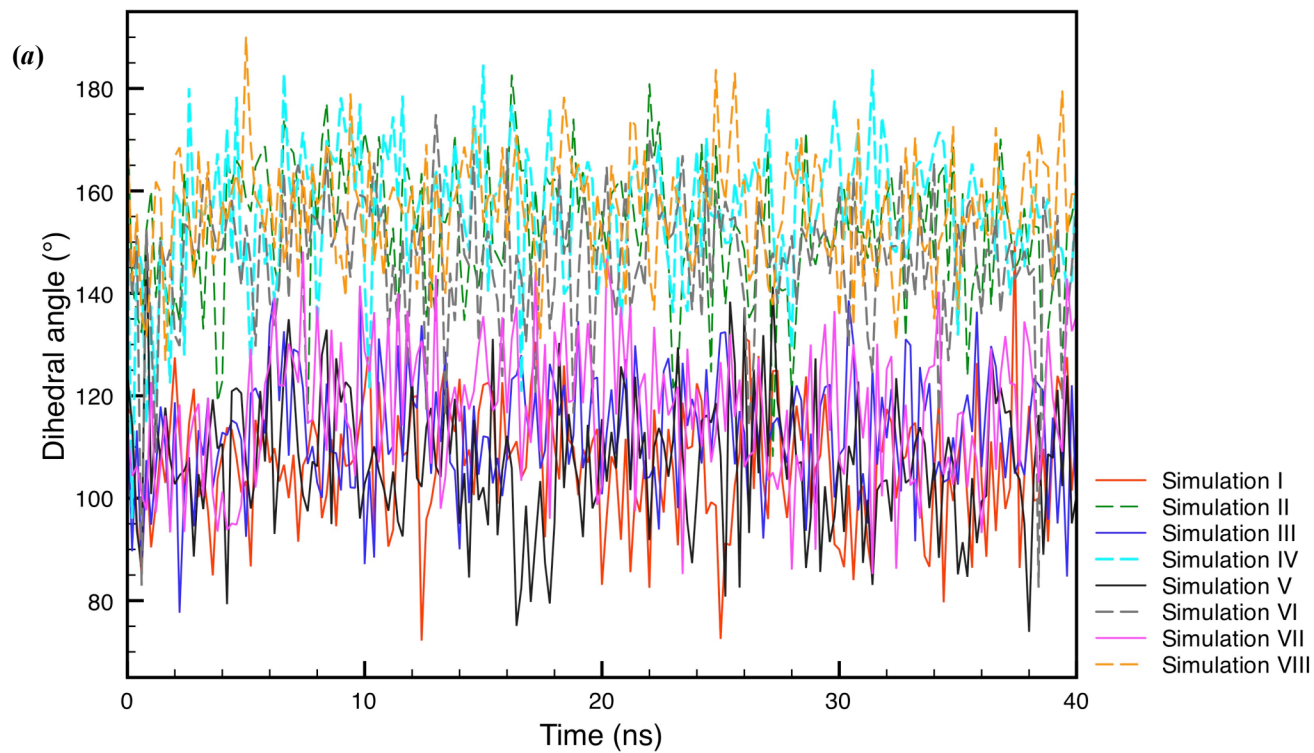


FIGURE S5

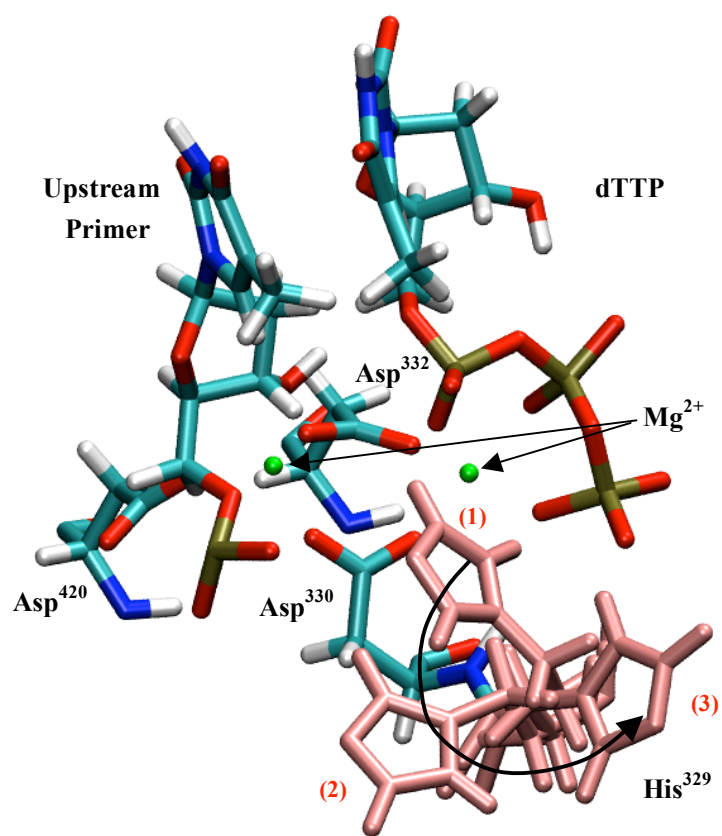


FIGURE S6

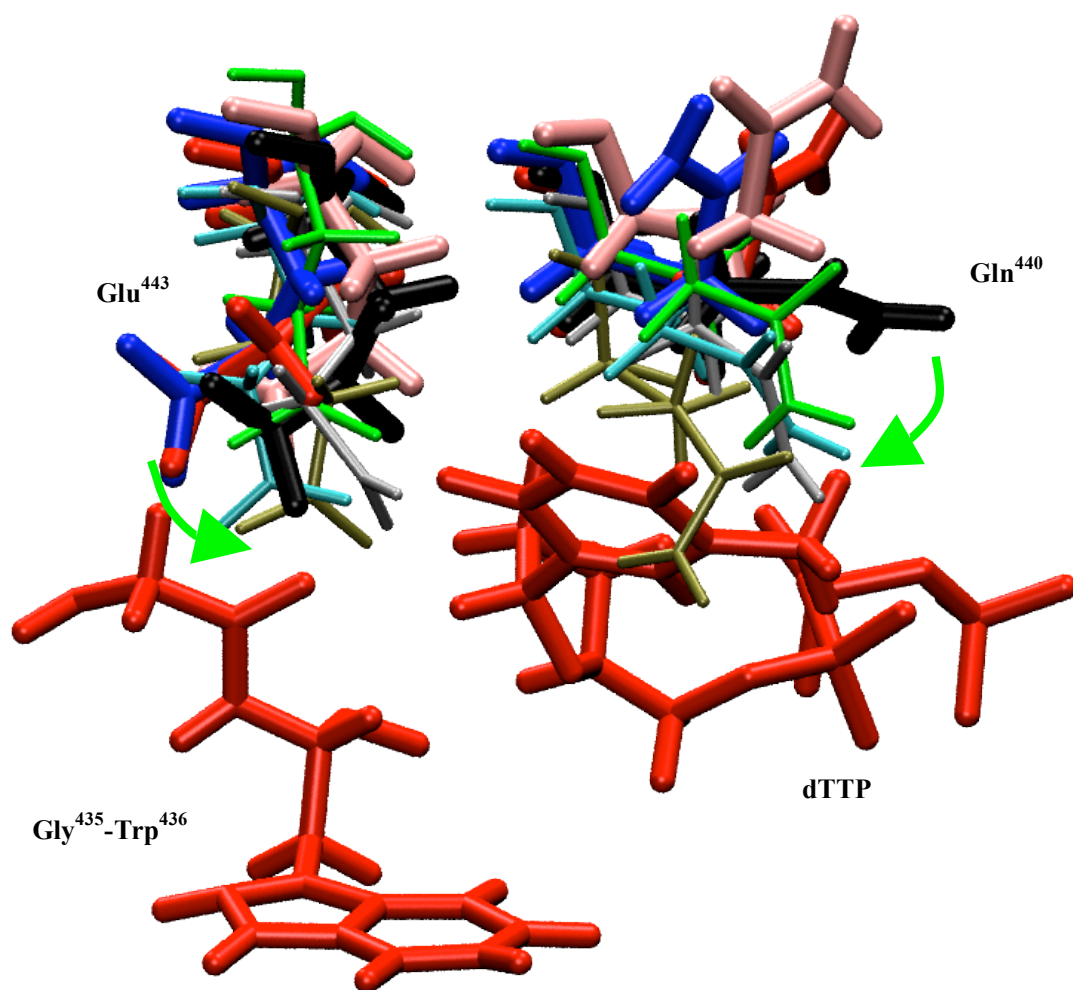


FIGURE S7

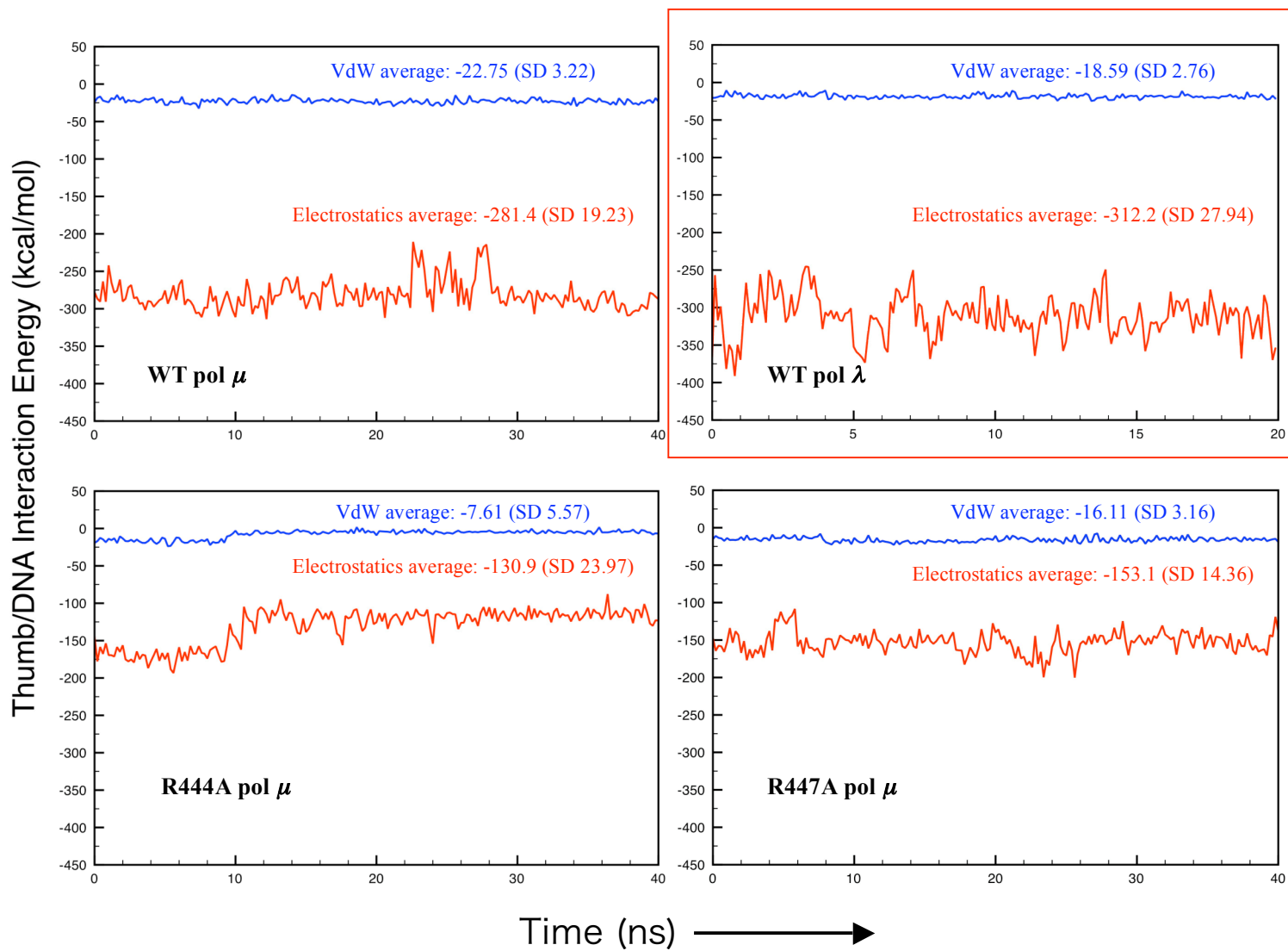


FIGURE S8

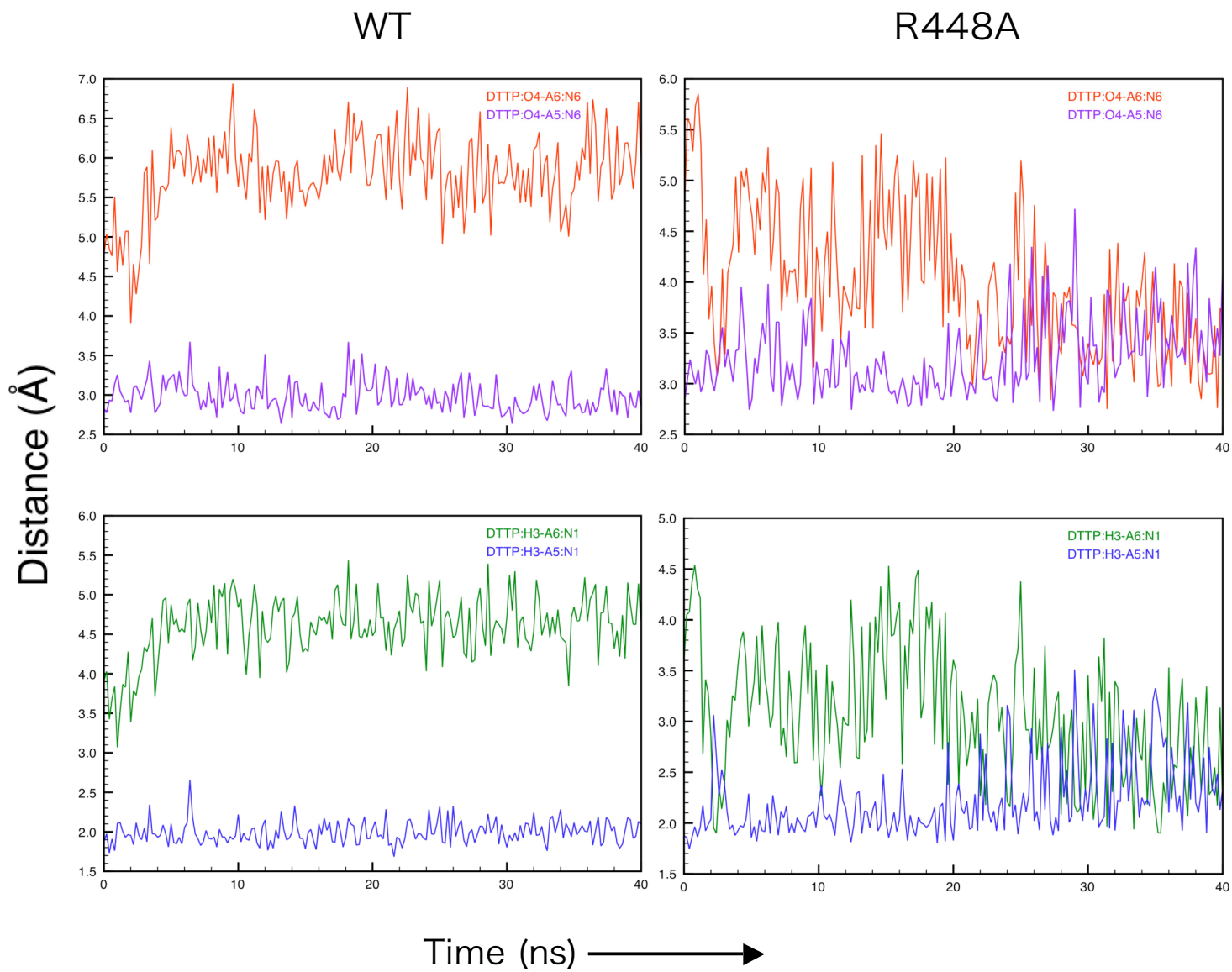


FIGURE S9

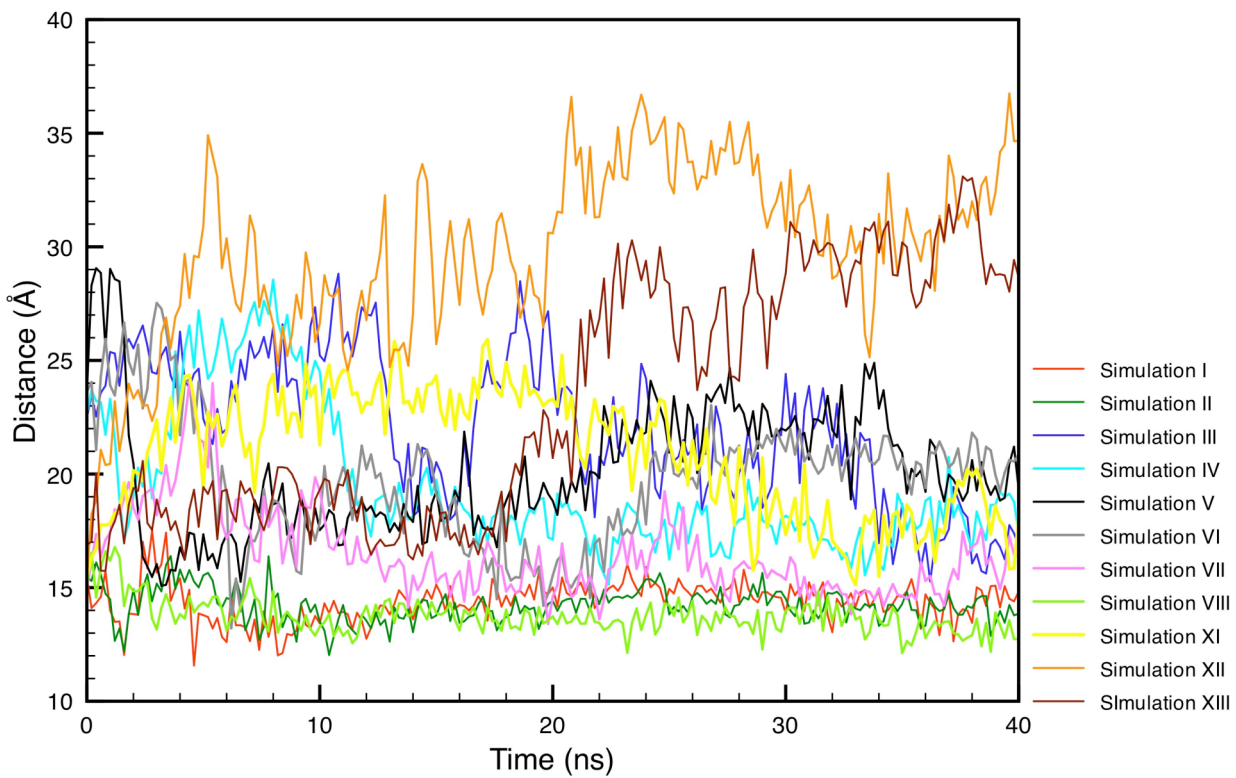


FIGURE S10

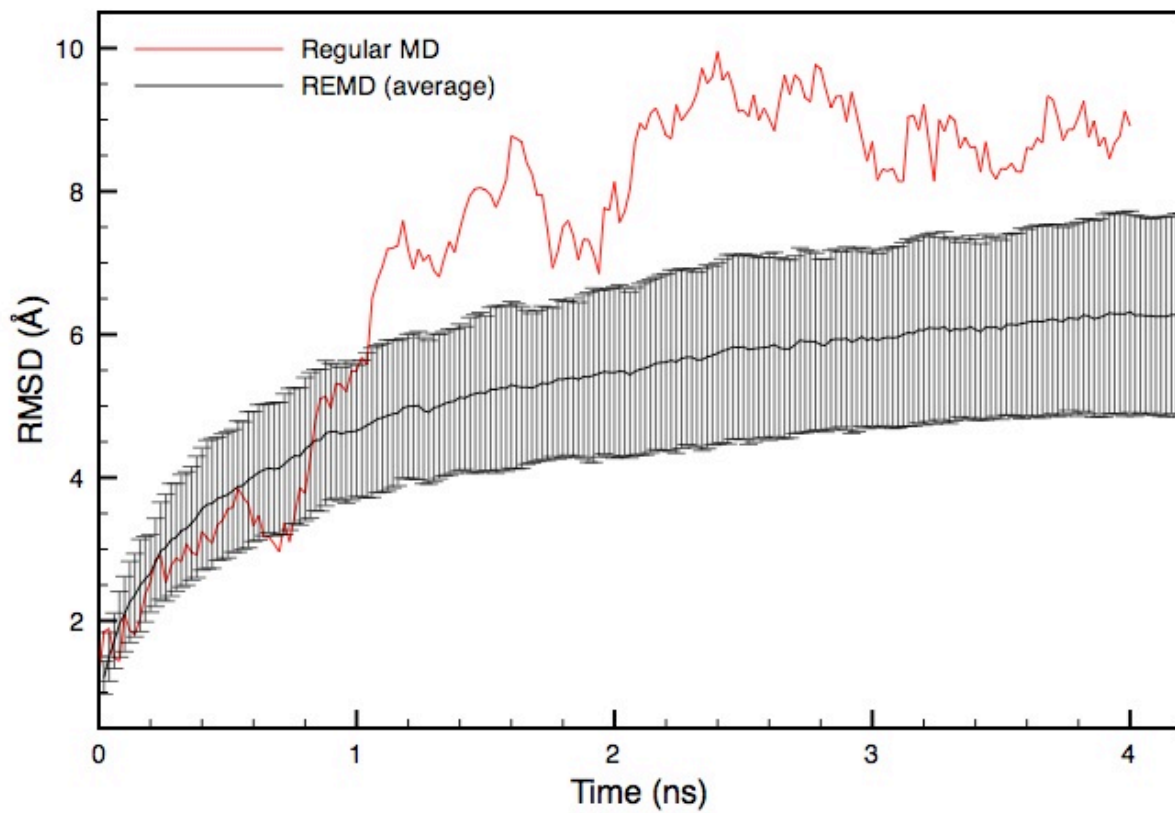


FIGURE S11

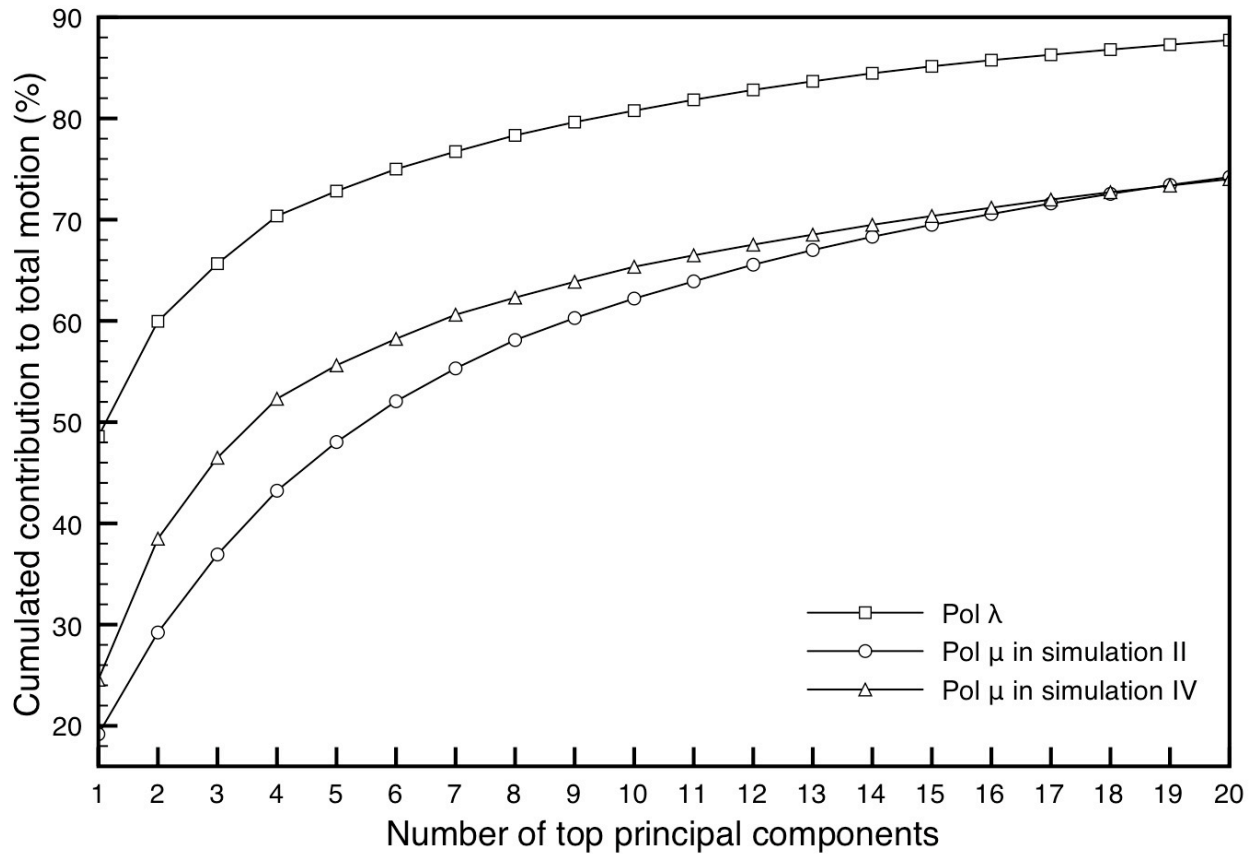


FIGURE S12