Base-Flipping Mechanism in Postmismatch Recognition by MutS

Sean M. Law[†] and Michael Feig[†‡]

[†]Department of Biochemistry & Molecular Biology, and [‡]Department of Chemistry, Michigan State University, East Lansing, Michigan

SI Methods

Minimization, Equilibration, and MD Protocol

Initial minimization involved 50 steps with the steepest descent method (SD) followed by 10,000 steps of adopted basis Newton-Raphson minimization. During the minimization, a 10 kcal/mol/Å^2 harmonic restraint was applied to all solute heavy atoms. Following minimization, each structure was gradually heated to 300 K and equilibrated in the NVT ensemble through three consecutive stages: (1) 1.4 ns of MD were carried out during which the solute was restrained as described above but water and ions were allowed to move freely; (2) solute restraints were released gradually over a period of 100 ps; (3) unrestrained MD over 6.4 ns was carried out to further equilibrate the system. All minimization and equilibration steps were carried out using the CHARMM program (1), version c35a1 in conjunction with the MMTSB Tool Set (2). After the initial equilibration phase, each of the nine simulations was then continued for another 200 ns using the NAMD simulation program, version 2.6 (3). The unrestrained NAMD simulations were carried out in the NPT ensemble that was maintained using a Langevin thermostat and barostat with a friction coefficient of 5 ps^{-1} and a 2 fs integration time step was used in conjunction with SETTLE (4) to holonomically constrain bonds involving hydrogen atoms.

F36A Mutant Set Up and Simulations

Fully equilibrated structures from four wild type simulations (ADP:NONE, ADP:ADP, ADP:ATP, and NONE:NONE) were taken and the S1-Phe[36] residue was mutated to an alanine residue using the MMTSB Tool Set (2). These mutated structures were subjected to the solute restraint protocol described above followed by a gradual release of the restraints and then equilibrated for an additional 10 ns (completely free of restraints). Each of the F36A mutant simulations were simulated for a total of 60 ns using an identical production simulation protocol as described above.

Water Residence Time

The residence time of water molecules located within 4 Å of the G10 base using a coordinate correlation function which has been previously used to assess solvent and ion residence times (5-6). Briefly, the water correlation function, $C_\alpha(t)$, is written as:

$$C_\alpha(t) = \frac{1}{N_{water}} \sum_{i=1}^{N_{water}} \frac{1}{N_{\alpha,i}(0, t_{total} - t)} \sum_{t'=0}^{t_{total}-t} p_{\alpha,i}(t', t'+t; t^*)$$

where $p_{\alpha,i}(t', t'+t; t^*)$ is a binary function that is set to 0 unless water molecule i is found within the predefined area α between time t' and $t'+t$. To ignore waters that escape and quickly rebind, the rebinding time t^* was set to 1ps. The binary function is then accumulated across the total simulation time t_{total} and divided by the number of times $N_{\alpha,i}(0, t_{total} - t)$ water molecule i is found within the confines of α. Finally, N_{water} corresponds to the total number of water molecules that participate in the residence time calculation. Depending on the overall sampling, water residence times may be fit to a bi-exponential decay function or, in some cases, a tri-exponential function. Alternatively, by taking the natural logarithm of the water correlation function the residence times can be easily obtained by calculating the inverse slope of portions of $\ln(C_\alpha(t))$ that can be fit to a linear curve.

Solvent-Accessible Surface Area

Analysis of the solvent-accessible surface area for the H1’ atom in the DNA minor groove of bases near the mismatch site was obtained from the COOR SURF command in CHARMM using a 1.4 Å probe radius (which represents a single water molecule). Solvent-accessible surface areas were then calculated with and without MutS. The reported change in accessibility upon MutS binding, referred to as ΔSASA, is the difference between these two values.

References

1. Brooks, B. R., C. L. Brooks, III, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. 2009. CHARMM: The Biomolecular Simulation Program. J. Comput. Chem. 30:1545-1614.
2. Feig, M., J. Karanicolas, and C. L. Brooks, III. 2004. MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology. J. Mol. Graphics Modell. 22:377-395.
3. Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. 2005. Scalable molecular dynamics with NAMD. J. Comput. Chem. 26:1781-1802.
4. Miyamoto, S., and P. A. Kollman. 1992. SETTLE - An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. J. Comput. Chem. 13:952-962.
5. Feig, M., and B. M. Pettitt. 1999. Sodium and chlorine ions as part of the DNA solvation shell. Biophys. J. 77:1769-1781.
6. Garcia, A. E., and G. Hummer. 2000. Water penetration and escape in proteins. Proteins 38:261-272.

SI Figure Legends

Figure S1. MutS-DNA stability and clustering analysis. (A) C_α protein RMSD (red) and heavy atom DNA RMSD (black) for each of the nine simulation models calculated with respect to a common starting structure. Refer to *Methods* for the notation used to describe each simulation model. (B) K-means clustering (implemented in the MMTSB Tool Set (1)) of the nine simulations using a 2.5 Å radius (large gray overlapping circles). Structures were extracted at every 1 ns of production simulation and clustered based on the C_α RMSD. The area of each of cluster (six colored circles) is proportional to the number of structures in that cluster and the individual colored slices show the contribution of structures from the nine different simulations. The colored edges correspond to the sampling of each simulation and the length of the edge is proportional to the C_α RMSD between any two connected cluster centers. The largest C_α RMSD of 3.6 Å was between cluster 2 and cluster 4.

Figure S2. Comparison of the G·T and G/C(-1) major groove widths from the unbiased NONE:NONE simulation. The solid blue line corresponds to a canonical major groove width of 17 Å estimated from the B-DNA crystal structure 1BNA (2). Major groove widths were calculated using the 3DNA program v2.0 (3).

Figure S3. Comparison of C21 ζ and χ torsion angles from the unbiased wild type and F36A mutant simulations. The 3DNA program v2.0 (3) and the MMTSB Tool Set (1) was used for all structural analyses. (A)-(B) Comparison of the C21 backbone ζ torsion angle. (C)-(D) Comparison of T22 glycosyl rotation angle, χ.

Figure S4. Pseudodihedral angle definition and sampling for each replica along the pseudodihedral base-flipping reaction coordinate from the HREM simulation. (A) Pseudodihedral angle definition (see *Methods*). (B) HREM sampling. Thick red bars correspond to the range of equilibrium pseudodihedral angles, θ_i, prescribed by a given harmonic potential while thin black lines correspond to the actual range of pseudodihedral angles, θ, that is sampled by each replica (see *Methods*). (C) Free energy profiles from the same HREM simulation but generated from different length simulations. The PMF from Figure 4*A* is included here for comparison (red).

Figure S5. Water residence time calculations from the NONE:NONE simulation (see *SI Methods*) and solvent-accessible surface area (SASA) calculations from NONE:NONE and ATP:NONE. (A)-(B) Diagram depicting the water molecules that have entered into the minor groove side as a result of C21 base flipping. The S1 DNA binding domain is shown as a gray surface, the G/C(-1) base pair is colored yellow, the G·T mismatch is colored in magenta, the rest of the DNA is colored in brown (in (B) only), and water molecules located within 4 Å of the G10 base are colored as orange, blue, and green spheres (only waters within a 4 Å radius of the G10 base are used for the residence time calculation). Additional waters within 6 Å of the G10 base are colored in red and were included to illustrate the crowded environment. For clarity, only protein atoms within a 10 Å radius of the G10 base are shown. The white arrow points to the narrow 6 Å wide channel that is created when C21 is flipped out. Orange spheres correspond to fast moving waters with residence times less than 500 ps while green and blue spheres correspond to trapped waters with long residence times in the 1-10 ns range. (C)-(D) Water residence times for water molecules on the surface of the protein (away from the DNA) (C) before and (D) after C21 base flipping (note that the time is in picoseconds). (E)-(F) Water residence times for water molecules located within 4 Å of the G10 base (E) before and (F) after C21 base flipping. The black lines in (C)-(F) correspond to the inverse slope used to calculate the accompanying residence time. (G) Time series for change in solvent-accessible surface area (ΔSASA) of the H1' atom (bound to C1' on the minor groove side) upon binding to MutS. Results from the NONE:NONE simulation (top) where

the base is flipped out after 10 ns are compared with the ATP:NONE simulation (bottom) where the base remains stacked.

Figure S6. Protein domain motions. (A) Side and back view of the starting MutS-DNA complex along with the three orthogonal vectors, X, Y, and Z, used to describe the protein domain motions. The S1 DNA binding domain is red, the S2 DNA binding domain is pink, the DNA is brown, and the rest of the protein is colored white. Vector X is identical to the DNA helical axis, vector Y points upwards towards the phosphorous of cytosine 7, and vector Z is perpendicular to X and Y and points towards the S1 core domain. (B) Side and back view of the S1 and S2 DNA binding domains bound to DNA after ~200 ns of simulation time. The final simulation structure is colored in green and the orientation of the starting structure is identical to (A). (C)-(H) Comparison of the range of motion along the three orthogonal axes X, Y, and Z between trajectories with and without base flipping. The trajectory where base flipping is observed is colored in black and the remaining eight trajectories where no base flipping is observed is collectively colored in red. Movement of the S1 DNA binding domain is shown in (C) – (E) while movement of the S2 DNA binding domain is shown in (F) – (H).

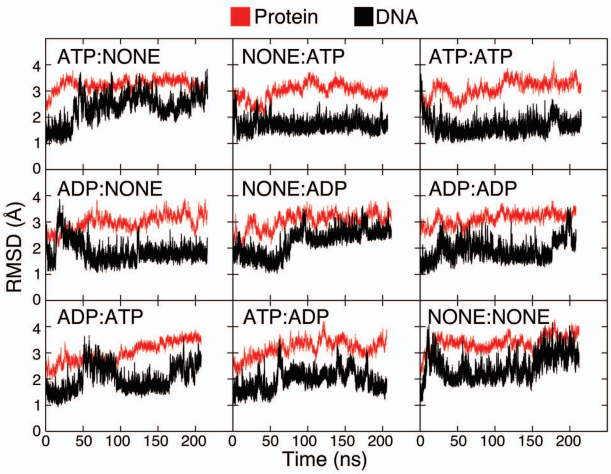
Figure S7. S1 DNA binding domain movement from unbiased and HREM (NONE:NONE) simulations along Y and Z directions. (A) Movement of the S1 DNA binding domain (S1-D1) along Y. (B) Movement of S1-D1 along Z. (C) Free energy profile of the movement of S1-D1 along Y with respect to the base opening angle. (D) Free energy profile of the movement of S1-D1 along Z with respect to the base opening angle.

Figure S8. Allosteric signaling from the DNA binding domain to the ATPase domains. (A) A structure-based sequence alignment was used (4) to map highly conserved residues onto the NONE:NONE model. The protein is shown as white ribbons and is in the same orientation as the full length structure found in Figure 1*A*. The conserved residues are highlighted as red spheres for the S1 monomer only. The same residues are conserved in the S2 monomer as well. DNA, water, and nucleotides have been omitted for clarity. (B)-(C) Visualization of the effects of base flipping on the ATPase domains. The S1 and S2 monomers are colored in blue and white, respectively. Ser^{668} is colored green, Asn^{616} is colored pink, Arg^{667} is colored cyan, Glu^{594} is colored yellow, and ATP is colored magenta. (B) Starting simulation structure modeled from the 1W7A crystal structure. ATP is modeled in for reference but is absent in the base-flipping simulation. (C) A post-base-flipping conformation from the base-flipped trajectory showing the Asn^{616} Ψ angle reorientation, salt bridge formation between Arg^{667} and Glu^{594}, and stabilization of Ser^{668} and the S2 signature loop.

1. Feig, M., J. Karanicolas, and C. L. Brooks, III. 2004. MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology. *J. Mol. Graphics Modell.* 22:377-395.
2. Drew, H. R., R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. E. Dickerson. 1981. Structure of a B-DNA Dodecamer - Conformation and Dynamics .1. *P Natl Acad Sci-Biol* 78:2179-2183.
3. Lu, X. J., and W. K. Olson. 2003. 3dna: A Software Package for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic Acid Structures. Nucleic Acids Res. 31:5108-5121.
4. Warren, J. J., T. J. Pohlhaus, A. Changela, R. R. Iyer, P. L. Modrich, and L. S. Beese. 2007. Structure of the Human MutS Alpha DNA Lesion Recognition Complex. Mol. Cell 26:579-592.

Figure S1

A


Protein
DNA
ATP:NONE
NONE:ATP
ATP:ATP
ADP:NONE
NONE:ADP
ADP:ADP
ADP:ATP
ATP:ADP
NONE:NONE
RMSD (Å)
Time (ns)

B

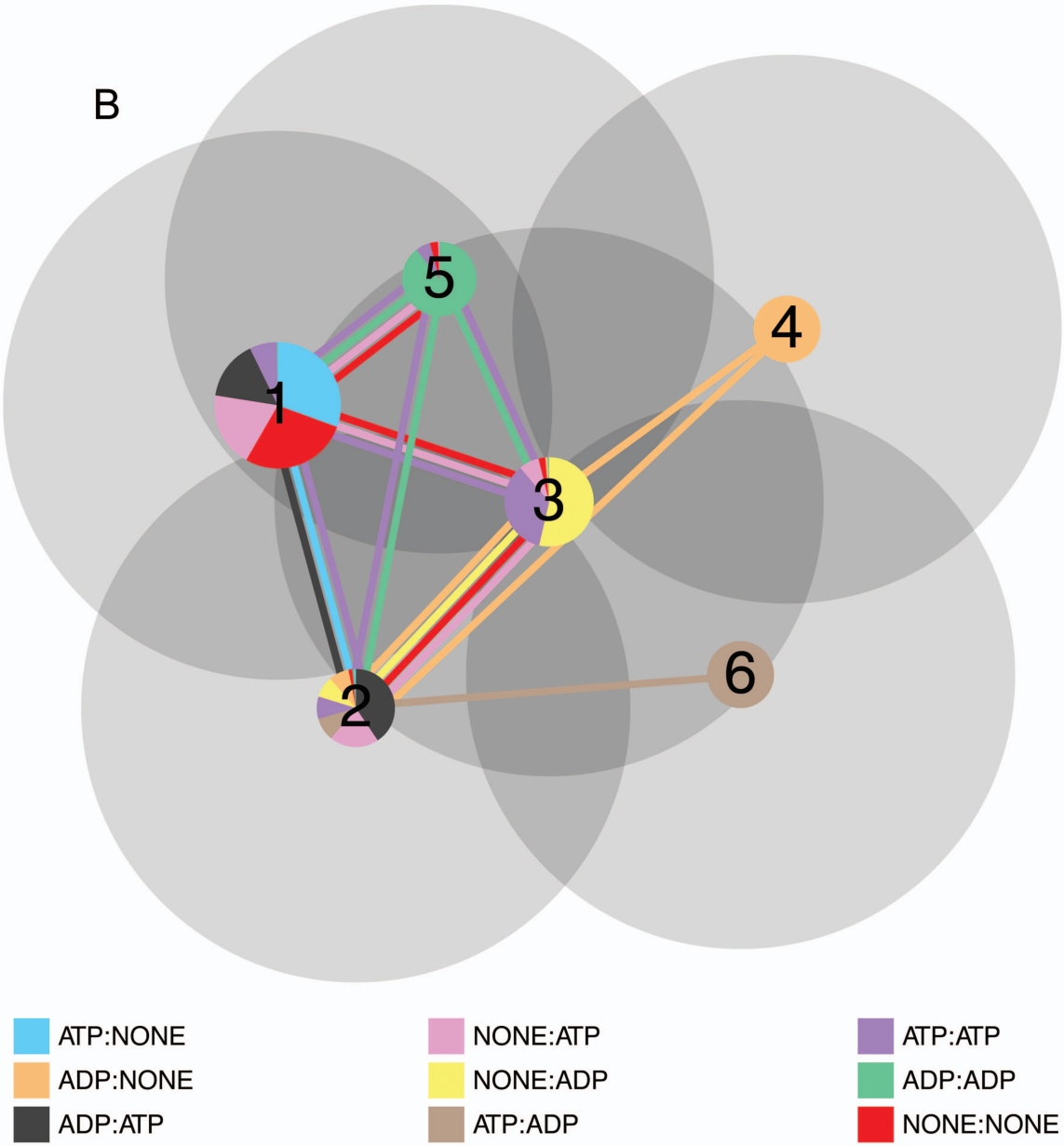

1
2
3
4
5
6
ATP:NONE
NONE:ATP
ATP:ATP
ADP:NONE
NONE:ADP
ADP:ADP
ADP:ATP
ATP:ADP
NONE:NONE

Figure S2

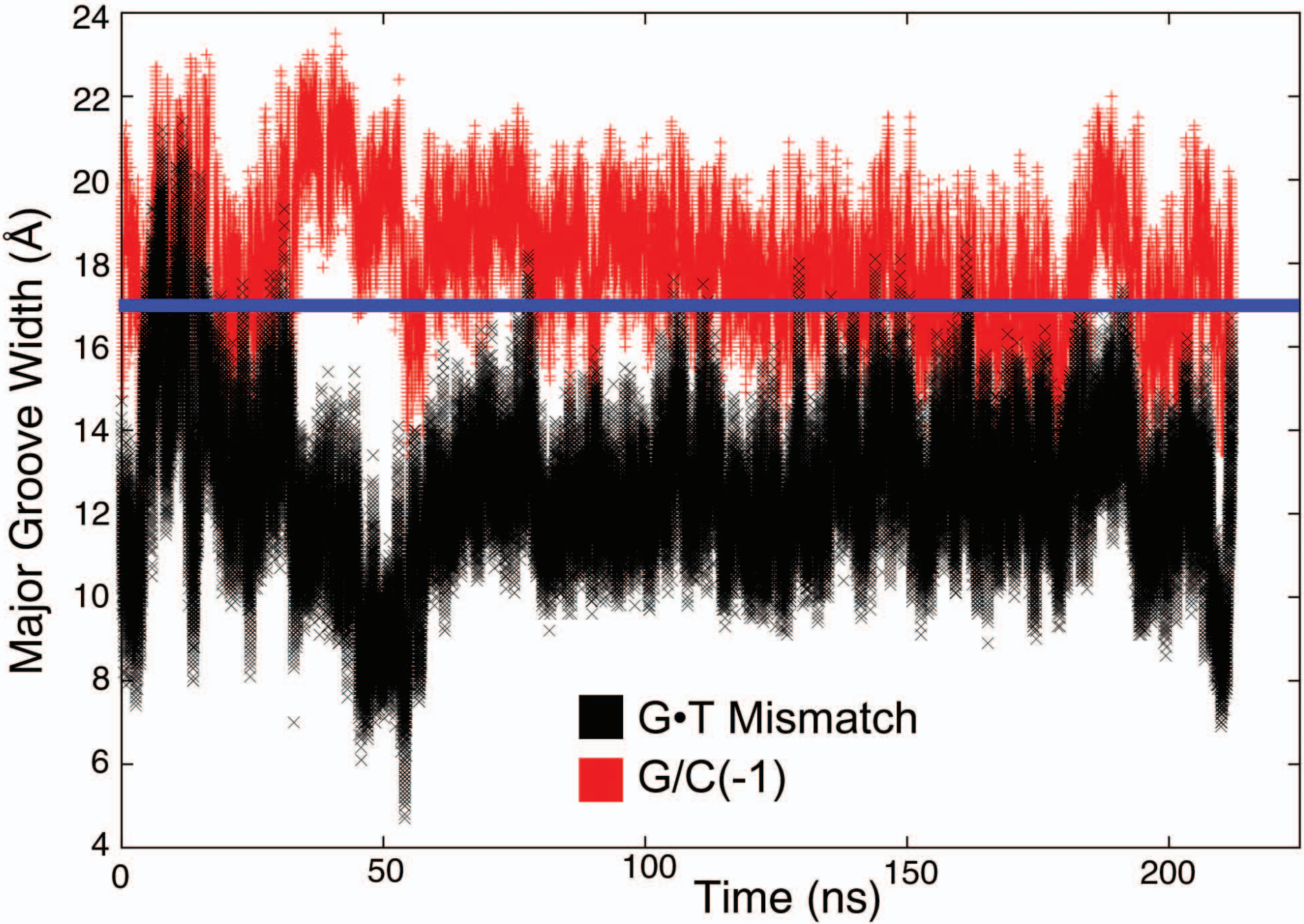

Major Groove Width (Å)
24
22
20
18
16
14
12
10
8
6
4
0
50
100
150
200
Time (ns)
G•T Mismatch
G/C(-1)

Figure S3

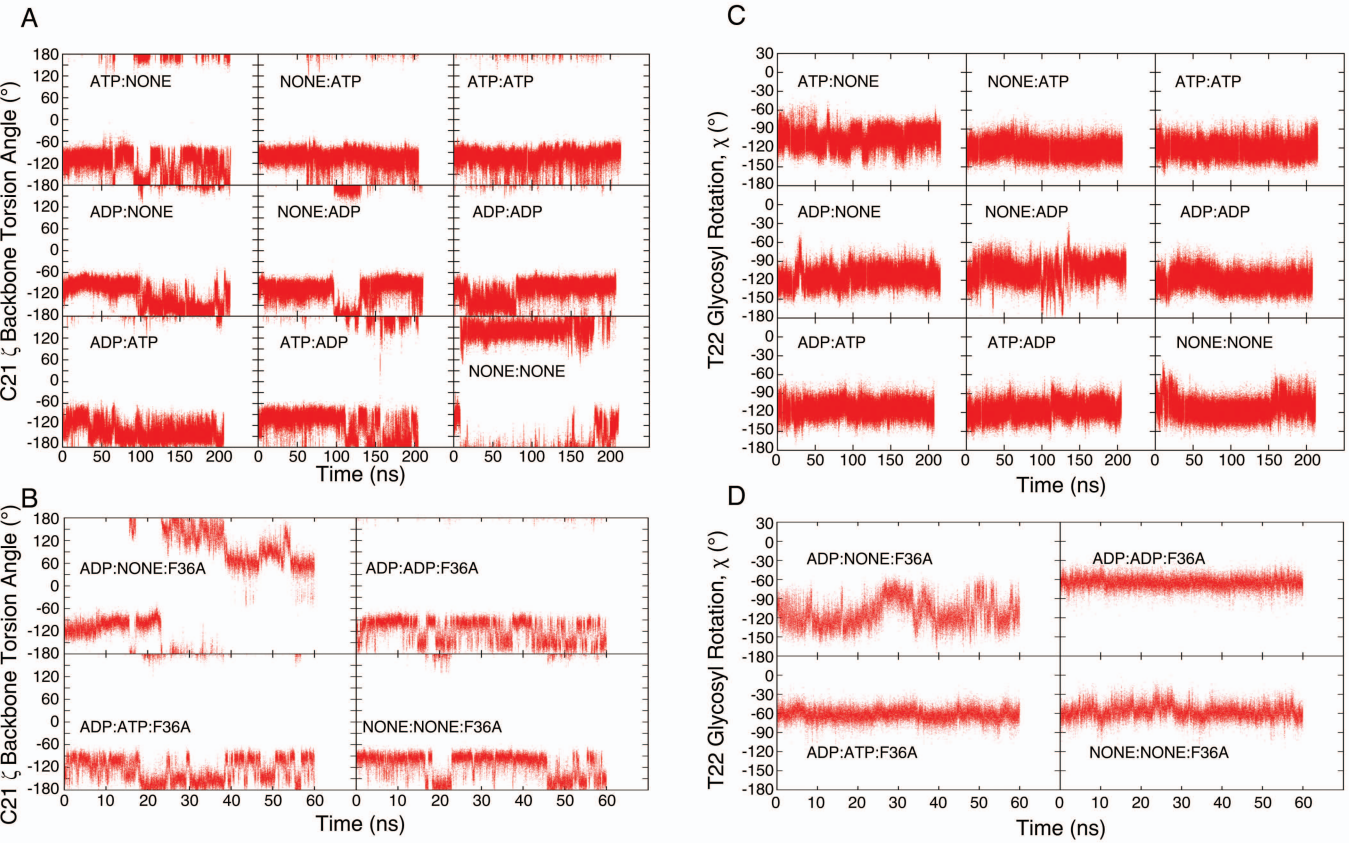
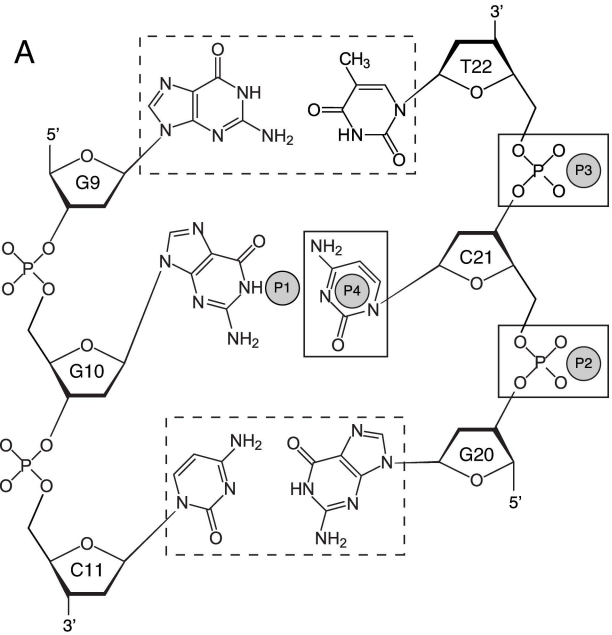
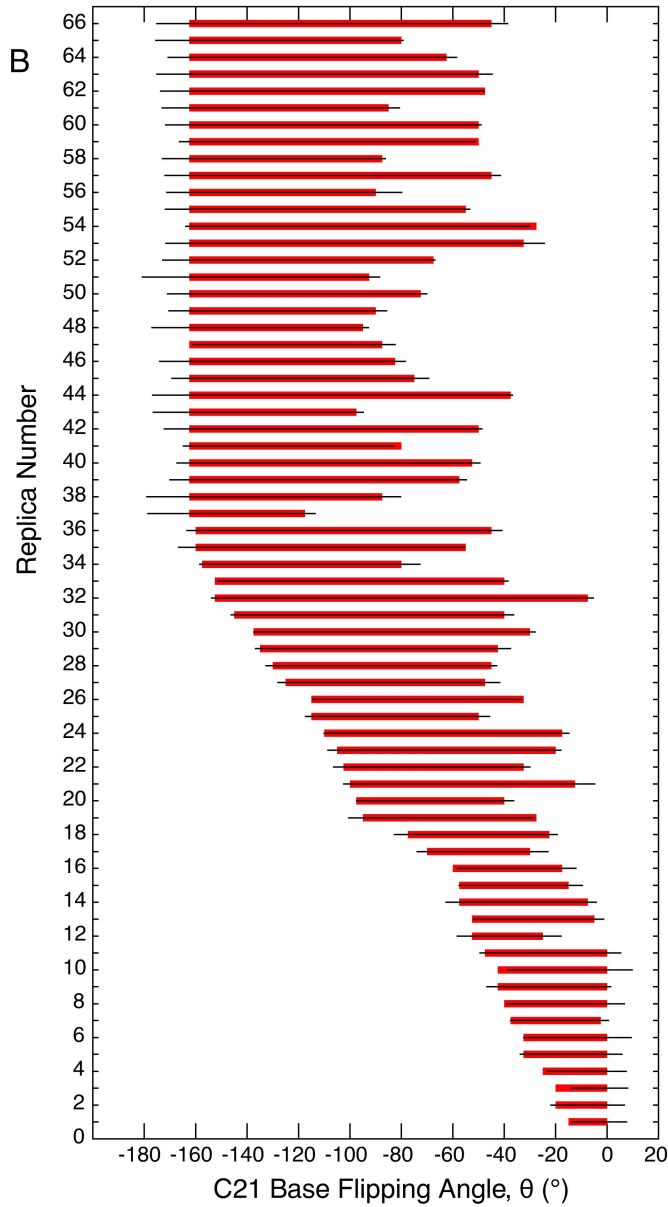

A
C21 ζ Backbone Torsion Angle (°)
ATP:NONE
NONE:ATP
ATP:ATP
ADP:NONE
NONE:ADP
ADP:ADP
ADP:ATP
ATP:ADP
NONE:NONE
Time (ns)
B
C21 ζ Backbone Torsion Angle (°)
ADP:NONE:F36A
ADP:ADP:F36A
ADP:ATP:F36A
NONE:NONE:F36A
Time (ns)
C
T22 Glycosyl Rotation, χ (°)
ATP:NONE
NONE:ATP
ATP:ATP
ADP:NONE
NONE:ADP
ADP:ADP
ADP:ATP
ATP:ADP
NONE:NONE
Time (ns)
D
T22 Glycosyl Rotation, χ (°)
ADP:NONE:F36A
ADP:ADP:F36A
ADP:ATP:F36A
NONE:NONE:F36A
Time (ns)

Figure S4


A
3'
T22
CH3
NH
N
NH2
O
HN
5'
G9
P3
C21
NH2
P1
P4
P2
G10
G20
5'
C11
3'


B
Replica Number
C21 Base Flipping Angle, θ (°)

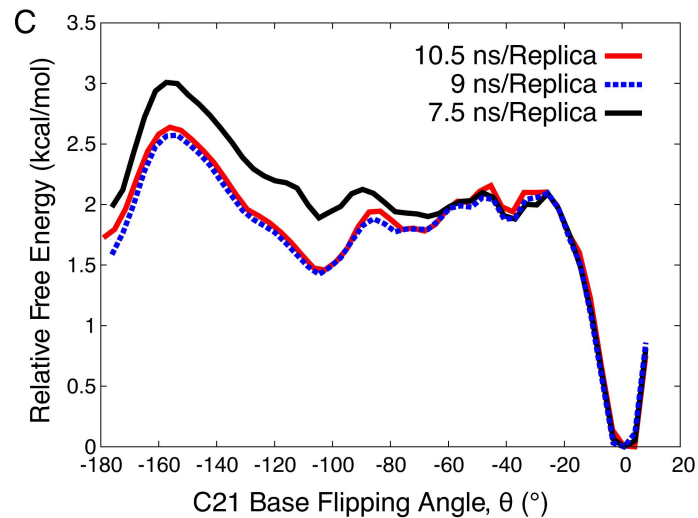
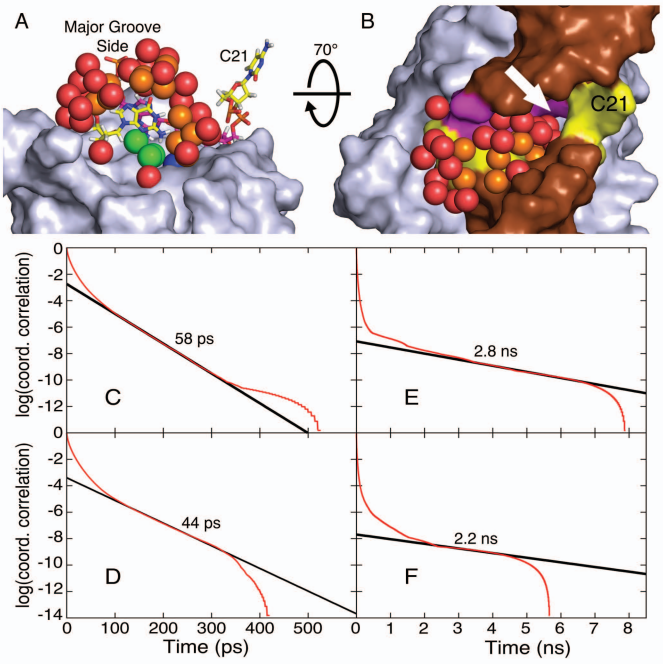

C
10.5 ns/Replica
9 ns/Replica
7.5 ns/Replica
Relative Free Energy (kcal/mol)
C21 Base Flipping Angle, θ (°)

Figure S5


A
Major Groove Side
C21
70°
B
C21
log(coord. correlation)
58 ps
C
2.8 ps
E
44 ps
D
2.2 ps
F
Time (ps)
Time (ns)

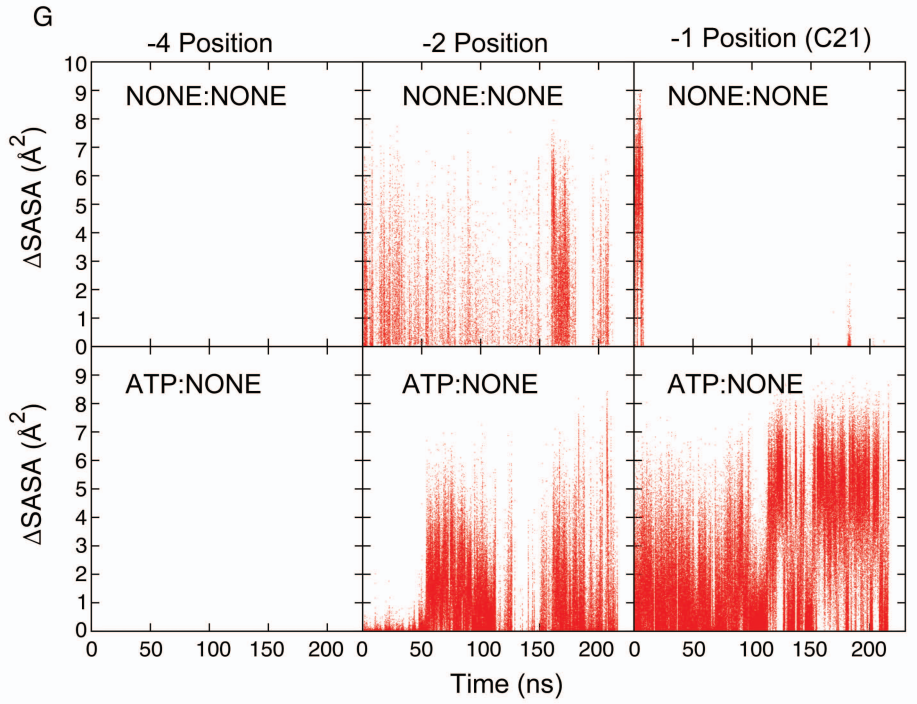
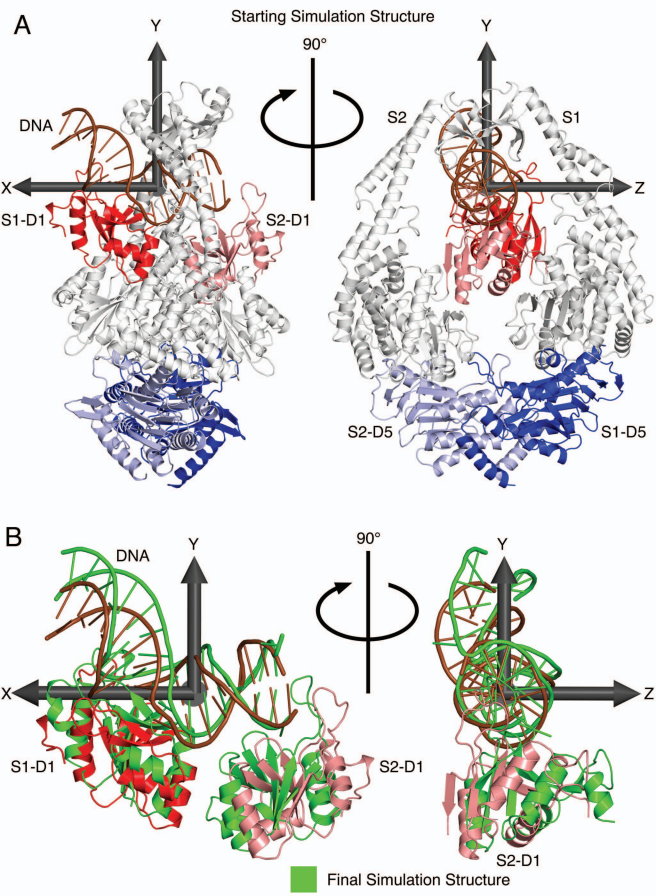

G
-4 Position
-2 Position
-1 Position (C21)
NONE:NONE
ATP:NONE
ΔSASA (Å²)
Time (ns)

Figure S6


A
Starting Simulation Structure
90°
Y
X
Z
DNA
S1-D1
S2-D1
S2
S1
S2-D5
S1-D5
B
DNA
S1-D1
S2-D1
Final Simulation Structure

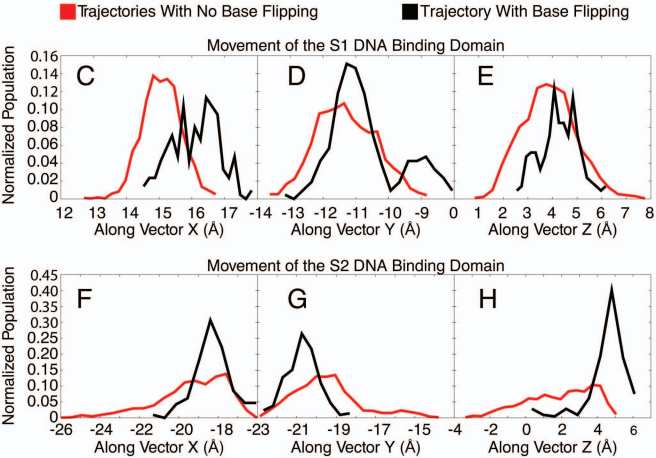
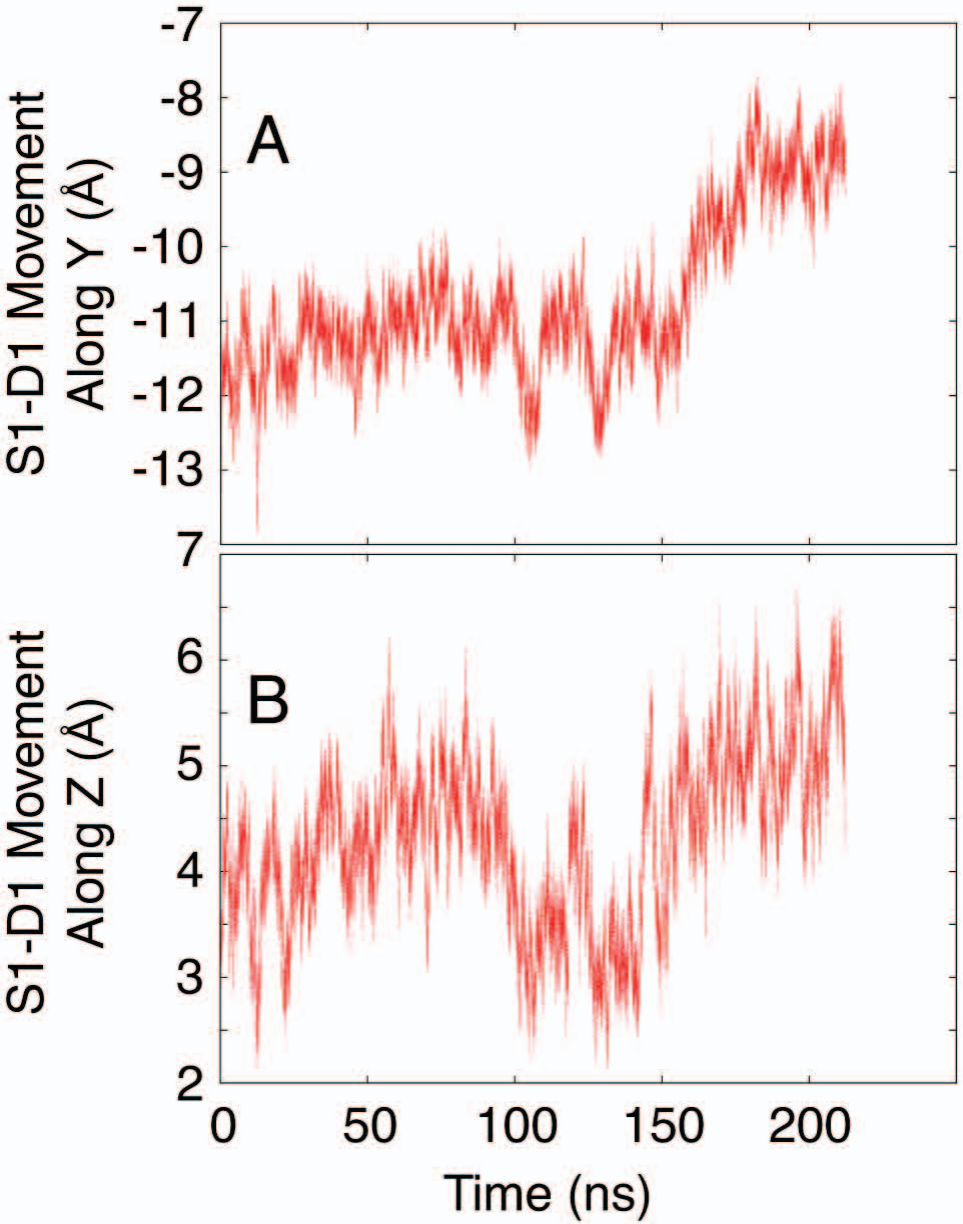

Trajectories With No Base Flipping
Trajectory With Base Flipping
Movement of the S1 DNA Binding Domain
C
D
E
Normalized Population
Along Vector X (Å)
Along Vector Y (Å)
Along Vector Z (Å)
Movement of the S2 DNA Binding Domain
F
G
H
Normalized Population
Along Vector X (Å)
Along Vector Y (Å)
Along Vector Z (Å)

Figure S7


A
B
S1-D1 Movement Along Y (Å)
S1-D1 Movement Along Z (Å)
-7
-8
-9
-10
-11
-12
-13
7
6
5
4
3
2
0
50
100
150
200
Time (ns)

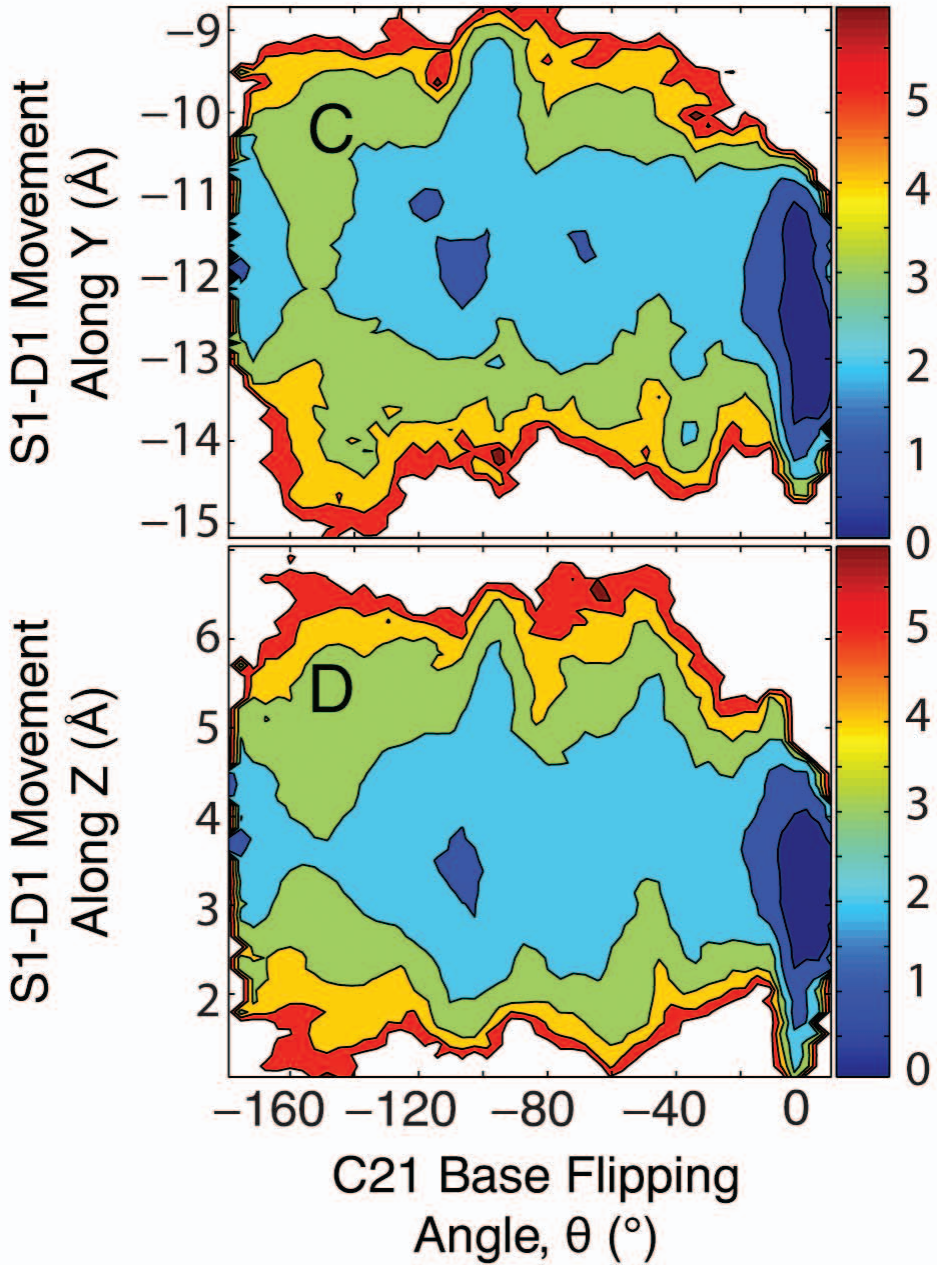

C
D
S1-D1 Movement Along Y (Å)
S1-D1 Movement Along Z (Å)
−9
−10
−11
−12
−13
−14
−15
6
5
4
3
2
5
4
3
2
1
0
−160
−120
−80
−40
0
C21 Base Flipping Angle, θ (°)

Figure S8

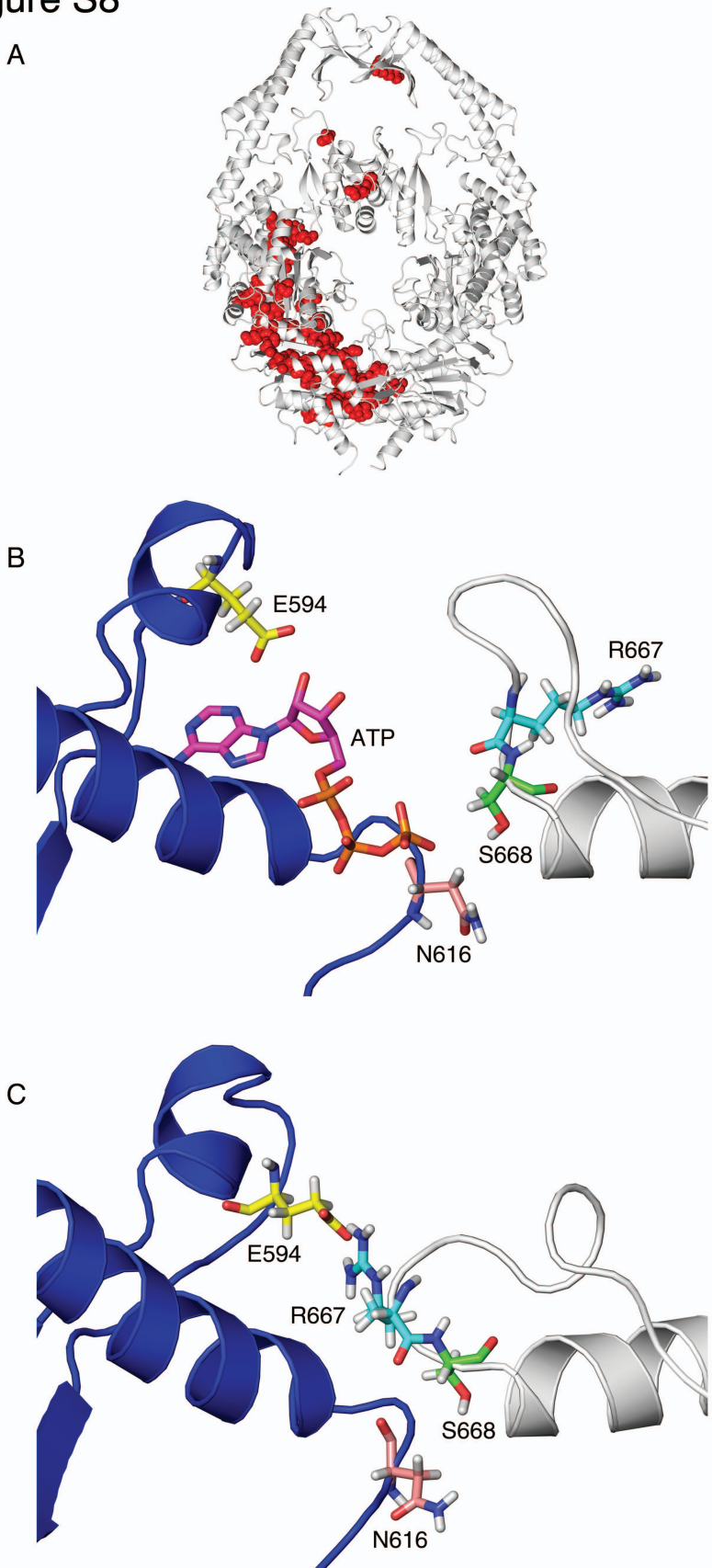


Table S1. DNA Sequence Used in All Simulations*

	G/-(-9)	T/A(-8)	G/C(-7)	A/T(-6)	C/G(-5)	C/G(-4)	A/T(-3)	C/G(-2)	G/C(-1)		A/T(+1)	C/G(+2)	C/G(+3)	G/C(+4)	T/A(+5)	C/G(+6)	G/C(+7)	A/T(+8)	
3'…	G18	T17	G16	A15	C14	C13	A12	C11	G10	**G9**	A8	C7	C6	G5	T4	C3	G2	A1	…5'
5'…	---	A14	C15	T16	G17	G18	T19	G20	C21	**T22**	T23	G24	G25	C26	A27	G28	C29	T30	…3'

*The DNA sequence is identical to the crystal structure found in Ref. 18 and the G·T mismatch is shown in bold.