Supplemental information

Molecular dynamics of mismatch detection—How MutS uses indirect readout to find errors in DNA

Abhilash Jayaraj, Kelly M. Thayer, David L. Beveridge, and Manju M. Hingorani

Supplemental Information

Molecular Dynamics of Mismatch Detection – How MutS Uses Indirect Readout to Find Errors in DNA

Abhilash Jayaraj[1,*], Kelly M. Thayer[1], David L. Beveridge[1], Manju M. Hingorani[2,*]

[1]Chemistry Department, Wesleyan University, Middletown, CT 06459, USA.

[2]Molecular Biology and Biochemistry Department, Wesleyan University, Middletown, CT 06459, USA.

*Correspondence: mhingorani@wesleyan.edu; ajayaraj@wesleyan.edu

Supplementary Video Legends

Video 1. MutS-homoduplex DNA motions defining PC1 of the MD simulation (see Figure S4). The observed motions are primarily separation of MutS clamp domains (IV) and lateral movement of DNA.

Video 2. MutS-homoduplex DNA motions defining PC2 of the MD simulation (see Figure S4). The observed motions are primarily unbending of homoduplex DNA from kinked to straight conformation.

Video 3. Series of representative structures from MutS-homoduplex DNA sub-states 1 through 9. Key contact residues are shown (described in Figure 7). The overall DNA shape, bendability, and local base stacking and pairing stability could be assessed through dynamic, concerted interactions of MutS with the sugar-phosphate backbone and bases over 1-2 helical turns.

Video 4. Closer-up view of the MutS-homoduplex DNA complex shown in video 3.
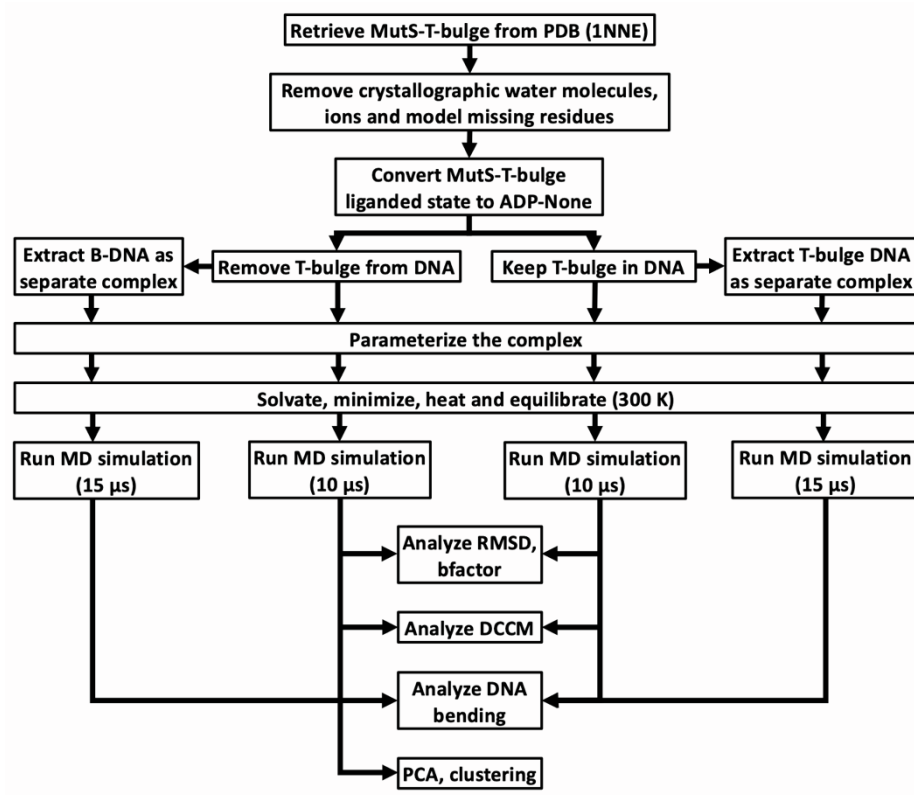
Figure S1

Figure S1. Flowchart of modeling, molecular dynamics (MD) simulations and analysis performed on MutS-DNA complexes and DNA in this study.
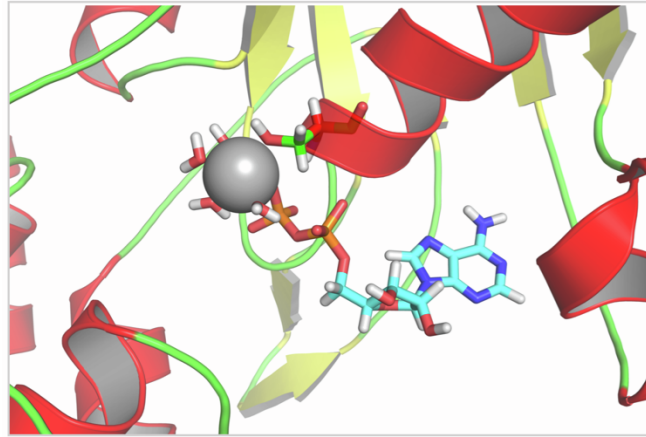
Figure S2

Figure S2. Parametrization of ADP in the *Taq* MutS nucleotide binding site. ADP was parameterized with Mg^{2+}, water molecules, and a serine residue in subunit A. Mg^{2+} was bonded to eight oxygen atoms in an octahedral formation (four from four water molecules, one from the ADP β phosphate and one from SER 590).

Figure S3

Figure S3. Dynamic Cross Correlation Matrices (DCCM) of MutS protein conformations shown in sequential or summed 1 μs timeframes over the MD simulation time (10 μs). The color scale ranges from red (high positive correlation) to blue (high negative correlation). The black circle highlights a region of different structural cross correlation between subunit A and B clamp domains (IV) in the two MutS-DNA complexes. (*A*) DCCM of MutS from the complex with T-bulge DNA for each 1 μs interval and (*B*) summed over 1 μs intervals (note: panel B(J) is also shown as Figure 3A in the main manuscript). (*C*) DCCM of MutS from the complex with homoduplex DNA for each 1 μs interval and (*D*) summed over 1 μs intervals (note: panel D(J) is also shown as Figure 3B in the main manuscript).
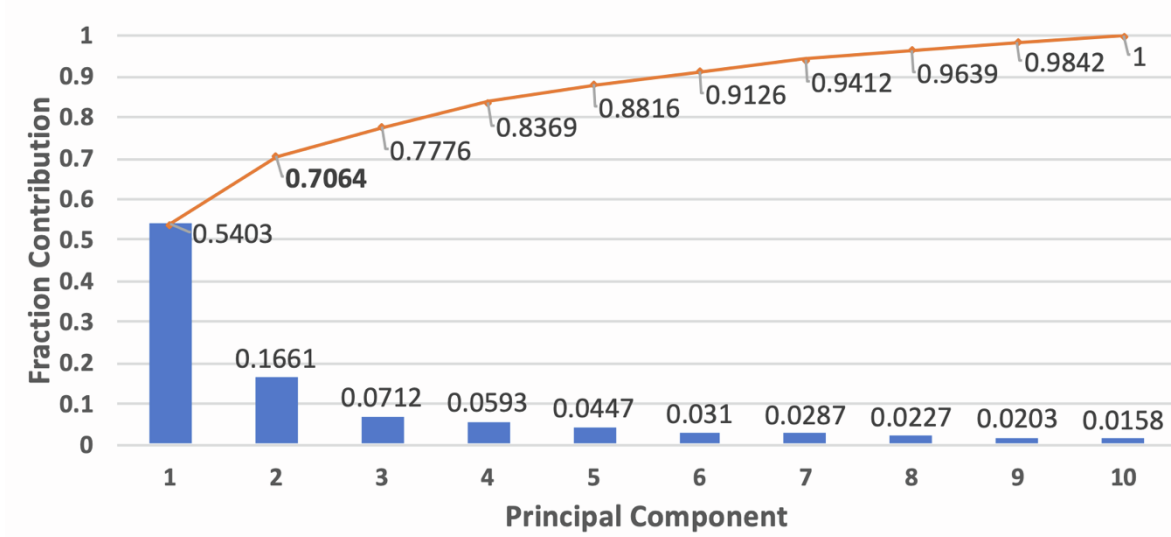
Figure S4

Figure S4. Principal components from analysis (PCA) of the MutS-homoduplex DNA complex MD simulation. Fractional contributions of the eigenvalue of each principal component (blue histograms) and the cumulative contribution (orange line) are shown.
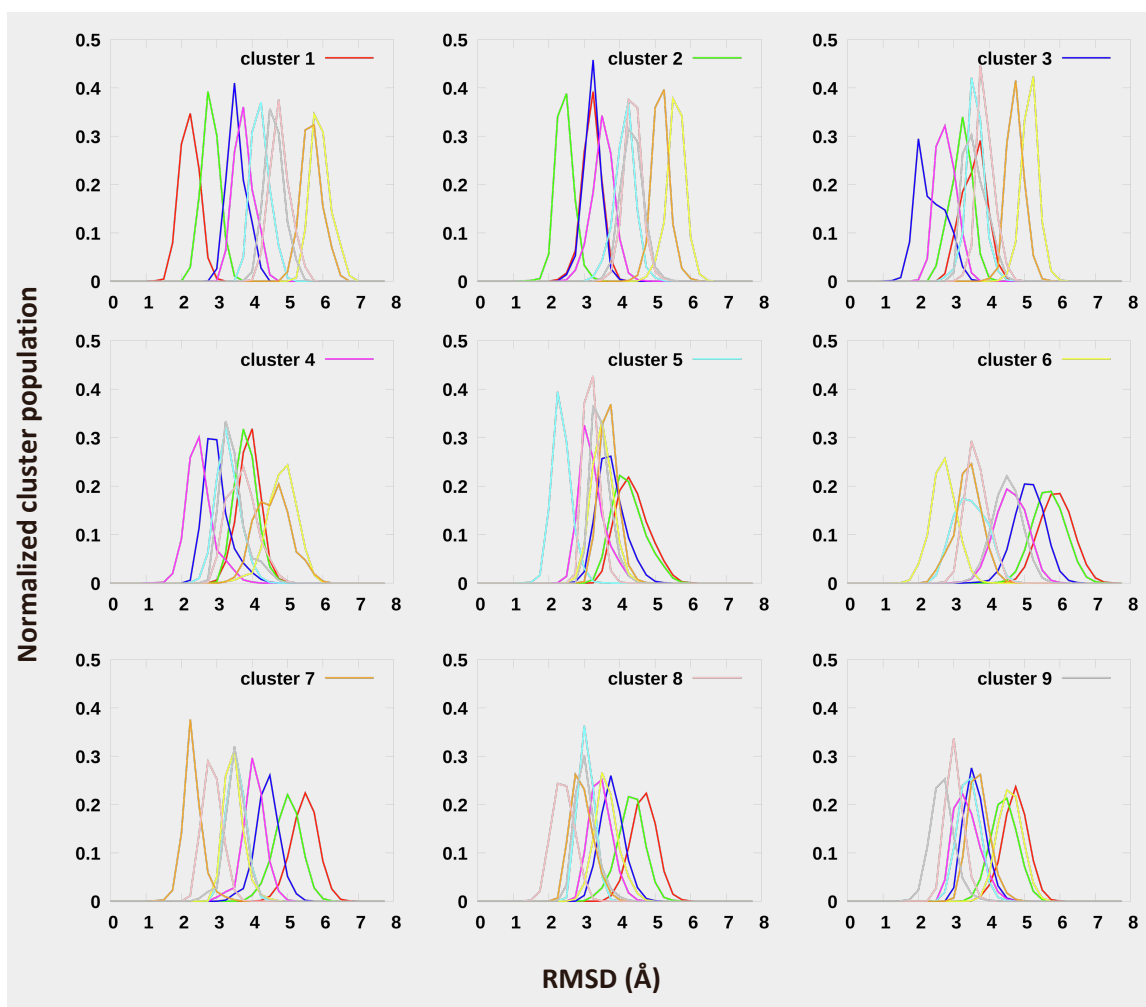
Figure S5

Figure S5. RMSD shift analysis of the nine MutS-homoduplex conformational clusters. In each panel, the RMSD of every frame in the MD trajectory to the centroid of its assigned cluster (1 – 9) and the centroids of all other clusters is shown as a frequency plot. The independent peaks observed for each cluster confirm that the frames assigned to that cluster lie closer to their own centroid relative to all the other centroids, which supports their classification as distinct conformational populations.
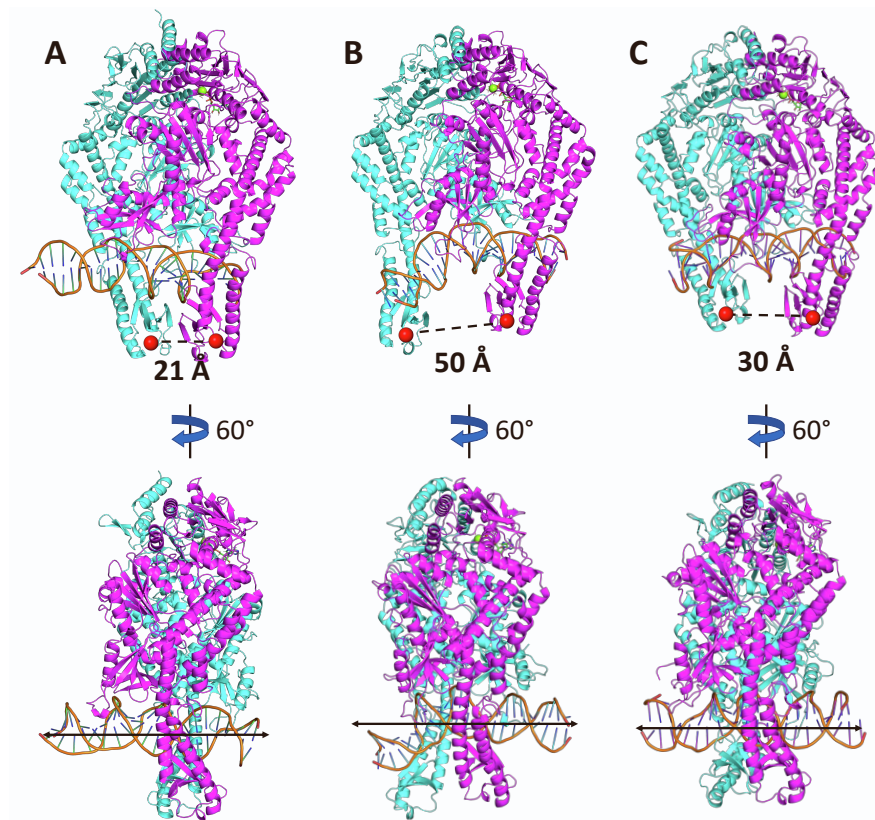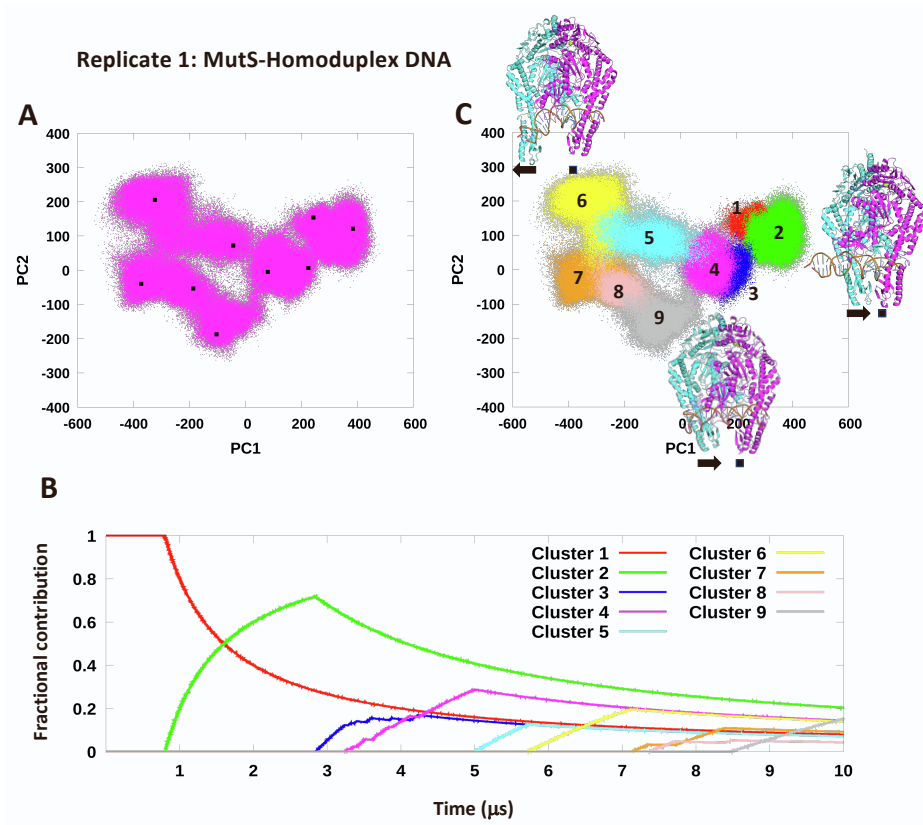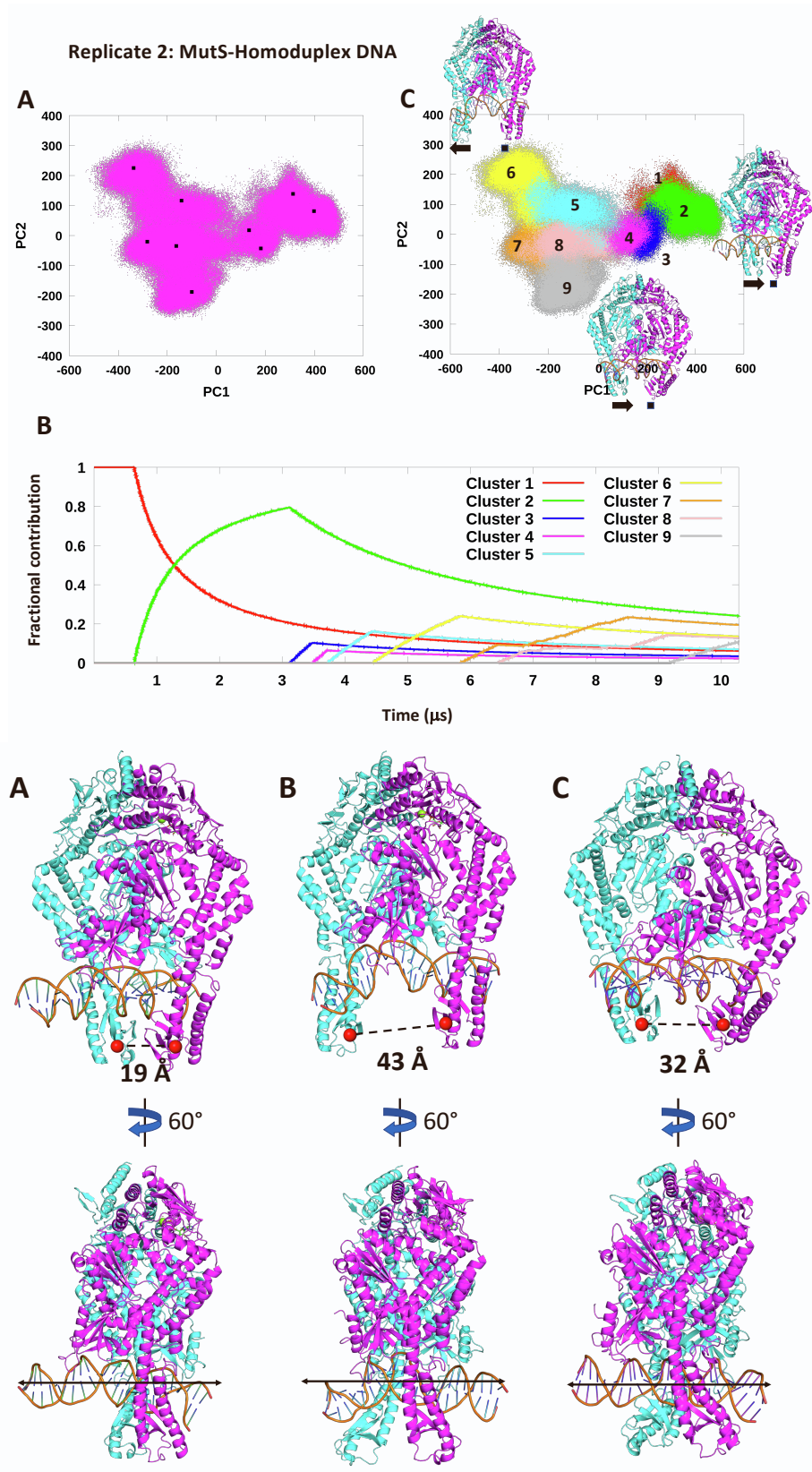
Figure S6

Replicate 1: MutS-Homoduplex DNA

Figure S7

Replicate 2: MutS-Homoduplex DNA

Figures S6 - S7. PCA and clustering analysis of replicate trajectories and the resulting sub-states of MutS-homoduplex complex. The data for replicate 1 (S6) and replicate 2 (S7) were analyzed as described for the original trajectory (Figures 5 and 6). ***Top panel:*** (*A*) MutS-homoduplex conformations projected as a function of the top two principal components of PCA (pink dots). The data sort into nine structurally distinct clusters (centered at the black dots). (*B*) Fractional composition of MutS-homoduplex clusters as a function of time. Clustering analysis was performed using K means clustering based on RMSD. The kinetic trace for each cluster (sub-state) is numbered in the order of appearance and has a different color (sub-state 1 - red, 2 - green, 3 - blue, 4 - magenta, 5 - cyan, 6 - yellow, 7 - orange, 8 - pink and 9 - grey). (*C*) Individual sub-states from the clustering analysis projected as a function of PCA components PC1 and PC2. ***Bottom panel:*** Prominent sub-states observed in all three independent MutS-homoduplex trajectories: (*A*) Sub-state 2 (straightened DNA, clamps very close), (*B*) sub-state 6 (bent DNA, clamps farthest apart), and (*C*) sub-state 9 (straight DNA, clamps close). Corresponding 60° rotated views show the DNA conformation (subunit A - magenta, subunit B - cyan; distance between clamp domains IV shown as dashed line).
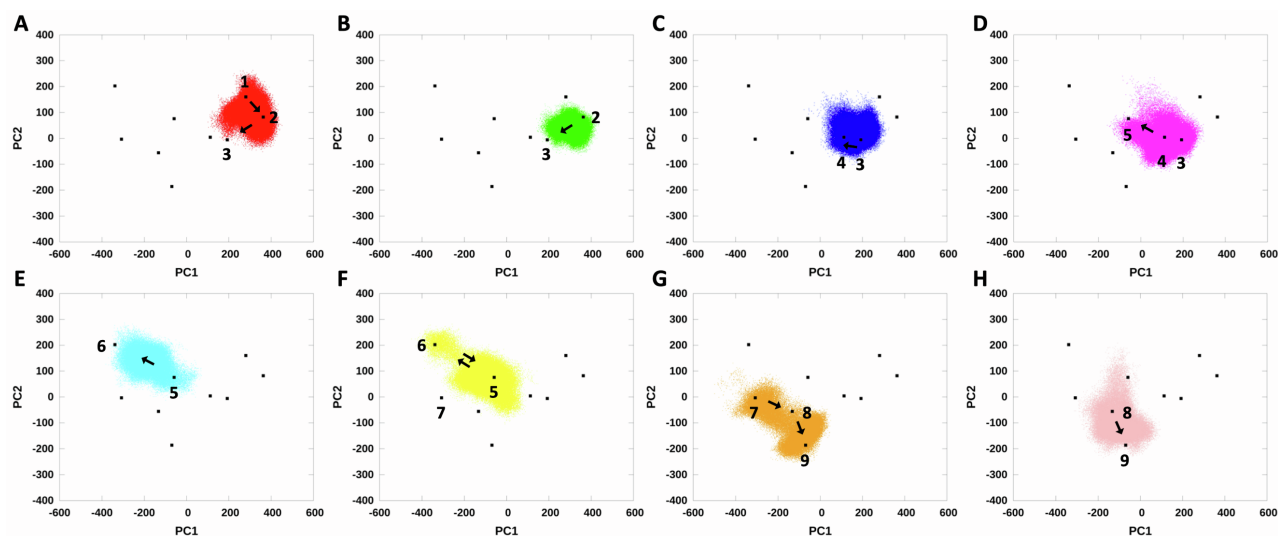
Figure S8

Figure S8. MD simulations of individual sub-states of the MutS-homoduplex DNA complex. Simulations were initiated independently from the original 10 μs simulation for each sub-state 1 through 8 and run for at least 2 μs. (*A-H*) Each trajectory is projected as a function of the top two principal components obtained from the full simulation (sub-state 1 - red, 2 - green, 3 - blue, 4 - magenta, 5 - cyan, 6 - yellow, 7 - orange, and 8 - pink). Each sub-state shows signs of transition to the next one during the simulation (arrows depict the observed direction). The data also indicate that sub-state 6 is capable of sampling both sub-states 5 (panel *F*) and 7 (Figure 5).
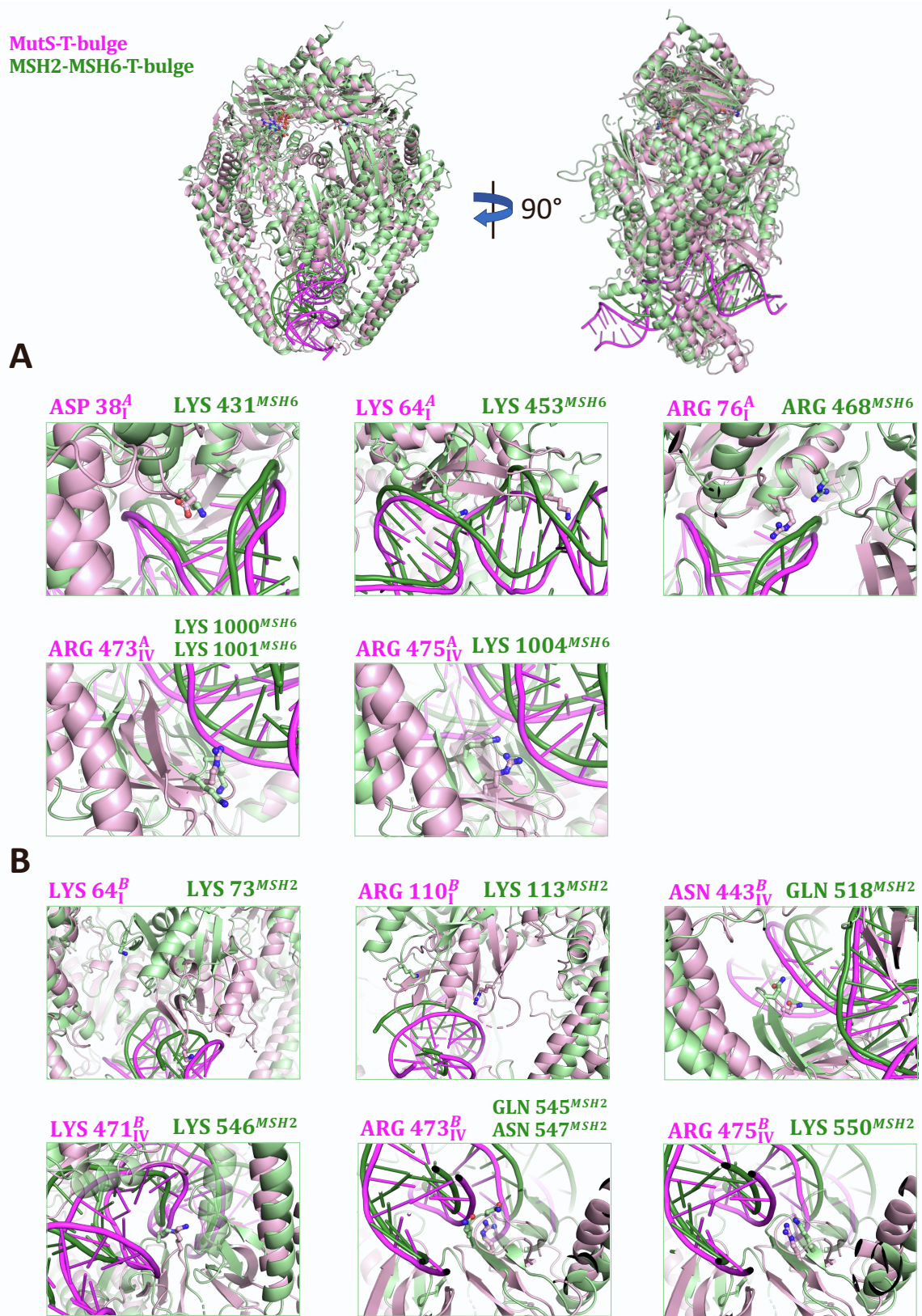
Figure S9

Figure S9. Structural alignment of *Taq* MutS-T-bulge and human MSH2-MSH6-T-bulge complexes. The two structures (PDB ID: 1NNE and 2O8F, respectively) were aligned by Pymol (mismatch-binding MutS subunit A with MSH6) and are shown in pink (MutS) and green (MSH2-MSH6). (*A*) MutS subunit A residues that have high contact frequency in MD trajectories and probe the conformational flexibility of homoduplex DNA (Table S2, Figure 7) and corresponding conserved MSH6 residues. (*B*) Analogous residues in MutS subunit B and MSH2. Note: MSH2 domain I is rotated away from DNA in the structure, hence LYS 64^B and ARG 110^B of MutS subunit B do not appear superimposed with corresponding MSH2 residues.

Cluster	Dwell time (µs)	Number of frames
1	0.48	24202
2	2.18	108895
3	0.48	24235
4	0.25	12664
5	0.68	33878
6	1.83	91657
7	0.6	29829
8	0.56	28040
9	2.93	146600

Supplementary Table S1. Fraction of time and number of frames MutS-homoduplex DNA complex dwells in each sub-state during the MD simulation. Sub-states 9, 2 and 6 are the top three most long-lived and populous sub-states.

Residue	Contact Frequency (%)			
	Sub-state 6		Sub-state 9	
	Subunit A	Subunit B	Subunit A	Subunit B
S63	20	-	49	-
K64	-	-	47	28
R76	21	-	43	-
R110	-	41	-	49
N443	-	20	-	30
K471	-	26	-	33
R473	79	78	85	75
R475	79	66	75	75

Supplementary Table S2. Contact frequency of residues from MutS subunits A and B with DNA in sub-states 6 and 9 of the MutS-homoduplex DNA complex, defined as the ratio of frames with at least one hydrogen bond between the residue and DNA/total frames of the sub-state in the MD trajectory.

Taq MutS[A, B]	Human MSH6	Human MSH2
D38[A]	K431	
S63[A]		
K64[A,B]	K453	K73*
R76[A]	R468*	
R110[B]		K113*
N443[B]		Q518*
K471[B]		K546*
K473[A,B]	K1000*, K1001	Q545, N547*
R475[A,B]	K1004	K550

Supplementary Table S3. Conservation of high contact frequency residues from *Taq* MutS-homoduplex DNA sub-states 6 and 9 (Table S2, Figure 7) in human MSH2-MSH6. The residues were identified from sequence and/or structural alignment between the two proteins (structural alignment shown in Figure S9).

Note: * denotes that mutations of these residues are reported in the InSiGHT database of variants for Lynch Syndrome (https://www.insight-group.org/variants/databases/); most are assigned unknown effects.