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] and Manju M. Hingorani[2,*]
[1]Chemistry Department, Wesleyan University, Middletown, Connecticut and [2]Molecular Biology and Biochemistry Department, Wesleyan University, Middletown, Connecticut

ABSTRACT   The mismatch repair protein MutS safeguards genomic integrity by finding and initiating repair of basepairing errors in DNA. Single-molecule studies show MutS diffusing on DNA, presumably scanning for mispaired/unpaired bases, and crystal structures show a characteristic "mismatch-recognition" complex with DNA enclosed within MutS and kinked at the site of error. But how MutS goes from scanning thousands of Watson-Crick basepairs to recognizing rare mismatches remains unanswered, largely because atomic-resolution data on the search process are lacking. Here, 10 μs all-atom molecular dynamics simulations of *Thermus aquaticus* MutS bound to homoduplex DNA and T-bulge DNA illuminate the structural dynamics underlying the search mechanism. MutS-DNA interactions constitute a multistep mechanism to check DNA over two helical turns for its 1) shape, through contacts with the sugar-phosphate backbone, 2) conformational flexibility, through bending/unbending engineered by large-scale motions of the clamp domain, and 3) local deformability, through basepair destabilizing contacts. Thus, MutS can localize a potential target by indirect readout due to lower energetic costs of bending mismatched DNA and identify a site that distorts easily due to weaker base stacking and pairing as a mismatch. The MutS signature Phe-X-Glu motif can then lock in the mismatch-recognition complex to initiate repair.

SIGNIFICANCE   MutS scans the genome at high speed to find and initiate repair of basepair mismatches in DNA, thereby suppressing mutations and guarding against carcinogenesis. Structures of mismatch-bound MutS show DNA bent sharply and the mismatch site distorted within a "recognition" complex. But how does MutS check long stretches of DNA and form this recognition complex on encountering a mismatch? Microseconds-length molecular dynamics simulations provide the first atomic-resolution view of MutS actions in search mode. Both MutS and DNA undergo concerted conformational changes whereby MutS coarsely checks DNA shape and bendability for mismatch-induced differences and finely checks basepairs for deformability. Through this multipronged mechanism, MutS can quickly detect DNA segments that bend more easily and confirm sites therein that distort more easily as mismatches.

INTRODUCTION

How does a DNA repair protein find and fix infrequent but potentially disastrous basepairing errors in a vast background of canonical basepairs? The mismatch repair (MMR) pathway maintains genome integrity by correcting mistakes made by polymerases during DNA replication, including mismatched basepairs and nucleotide insertions/deletions, suppressing homeologous recombination, which involves erroneous basepairing between partially homologous DNA sequences and signaling cellular responses to damage lesions in DNA (1–4). MMR begins when MutS

finds a basepairing defect in DNA through a process of search and recognition. This evolutionarily conserved protein forms a homodimer, e.g., *Thermus aquaticus* (*Taq*) or *E. coli* MutS (5,6), or heterodimer, e.g., *S. cerevisiae* or human Msh2-Msh6 (MutSα) or Msh2-Msh3 (MutSβ) (7,8), and uses its coupled, asymmetric mismatch binding and ATPase activities to initiate MMR (9). After recognizing a mismatch, MutS undergoes ATP binding-dependent conversion into a sliding clamp state (10–13) that can interact with MutL, an evolutionarily conserved protein that is itself an endonuclease (14) or interacts with an endonuclease, MutH, to nick the error-containing DNA strand (15,16). Subsequent controlled excision of the cut DNA strand past the mismatch site followed by accurate resynthesis completes MMR (17–19). Mutations in MutS and MutL proteins cause the hereditary Lynch cancer syndrome as well as

several sporadic cancers, highlighting the importance of MMR in safeguarding the accuracy of genetic information (20,21).

In the past two decades, numerous crystal structures of MutS have been solved in complex with different mismatches and insertion/deletions—all showing the two subunits arranged in a closed, θ-shaped dimer with DNA enclosed in one chamber (Fig. 1; *Taq* MutS) (5–8,22). In this "mismatch-recognition" complex, DNA is sharply kinked at the mismatch site (45–60°), and a Phe and Glu residue from the mismatch binding domain (I) of MutS subunit A (equivalent to Msh6 in Msh2-Msh6) are stacked and hydrogen-bonded with the mispaired/unpaired base, respectively. Several other residues from the mismatch binding and clamp domains (IV) of both MutS subunits undergo hydrogen bonding and ionic interactions with the DNA backbone over about 5–8 nucleotides on either side of the mismatch. This large MutS footprint, encompassing more than one helical turn of DNA, and the observation that mismatch binding affinity and repair efficiency is influenced by neighboring nucleotide sequence (23,24), implies an important, although unclear, role for these nonspecific interactions during mismatch search and recognition. In addition to the structural data, single-molecule tracking of MutS proteins on DNA has provided a bird's eye view of the search process (11,25,26). MutS encounters DNA via 3D diffusion, and it can encircle and move along the helix over thousands of basepairs at ∼700 bp/s via 1D rotational diffusion. A mismatch halts MutS movement, and it has been proposed that the conformational flexibility of DNA flanking the mismatch and its susceptibility to kinking enable MutS to

stall and switch from scanning to mismatch-recognition state (27–29). But there is little atomic-resolution data to reveal the mechanism whereby MutS interrogates the flexibility of homoduplex DNA and identifies the odd basepairing error from the large excess of normal basepairs. An exciting breakthrough occurred recently when a cryo-EM structure of *E. coli* MutS bound to homoduplex DNA was solved (30). In this "scanning" complex, MutS clamp domains are separated, and the DNA is held in a straight conformation between the mismatch binding and clamp domains of the dimer (DNA-free MutS structures also indicate that clamp domains can adopt various conformations (5,13,31)). The stark differences between the scanning and mismatch-recognition complex structures attest to dynamic interactions and conformational changes in both MutS and DNA as the protein diffuses on the duplex in search of mismatches.

Here, we report another substantive advance in elucidating the mismatch search and recognition mechanism employed by MutS on DNA. We conducted multiple independent 10–15 μs, unbiased molecular dynamics (MD) simulations of *Taq* MutS in complex with correctly basepaired homoduplex DNA and with DNA containing a single thymidine insertion (T-bulge). The data from these nonequilibrium dynamics simulations provide high-resolution temporal views of MutS interactions with the DNA backbone and bases spanning over almost two helical turns. Large conformational changes coupled through conserved contacts enable MutS to probe the helical shape, global bendability, and local basepair deformability of DNA, and transition from scanning mode to mismatch
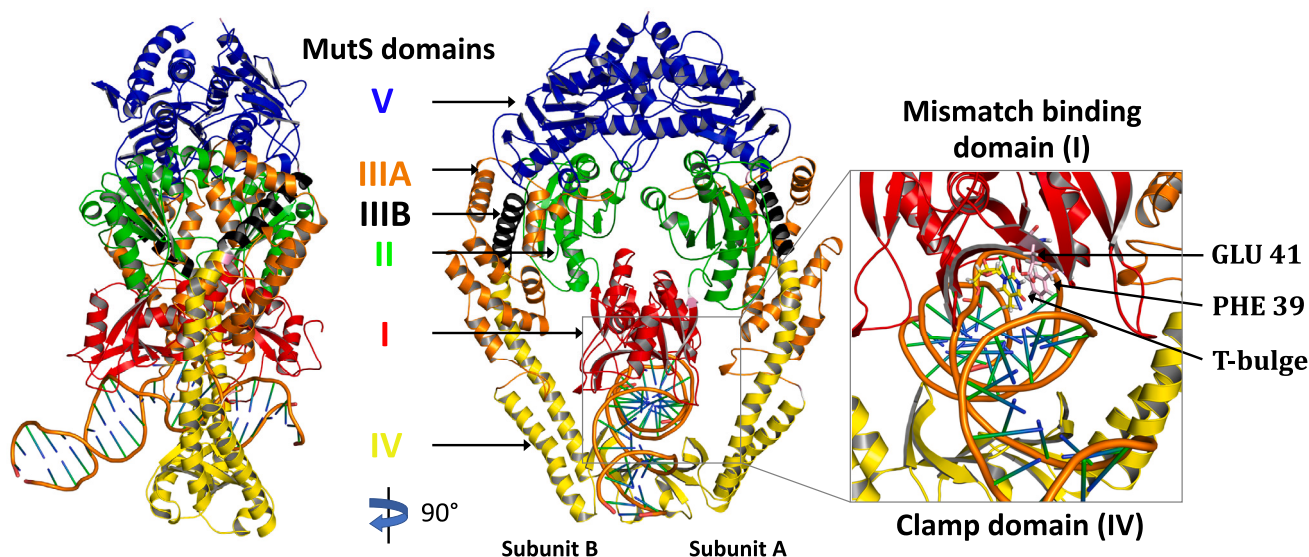

FIGURE 1 Side and front view of the crystal structure of *Taq* MutS bound to mismatched DNA with a T-bulge (PDB: 1NNE). Each subunit in the MutS dimer comprises five domains: I (mismatch binding, *red*), II (connector, *green*), IIIA (lever, *orange*), IV (clamp, *yellow*), IIIB (lever clamp, *black*), and V (ATPase, *blue*). The mismatch detection interface formed by domains I and IV is highlighted in the inset (T-bulge as *yellow sticks*; mismatch binding residues Phe-39 and Glu-41 in subunit A as *pink sticks*).

recognition at a bona fide error, such as a T-bulge, to initiate repair.

MATERIALS AND METHODS

Modeling MutS and DNA

The workflow of modeling MutS and DNA followed by MD simulations and analysis is outlined in Fig. S1. The *Taq* MutS protein sequence (811 residues) was retrieved from UniProt (32) (UniProt ID: Q56215). Of 811 residues, 765 could be accurately modeled using MutS structural data available in the RCSB databank (33). The crystal structure of *Taq* MutS-T-bulge (PDB: 1NNE) (34) was stripped of water molecules, and ADP-BeF$_3$ was removed from both nucleotide binding sites. Residues 630–635 (loop) were missing in both *Taq* MutS subunits and were modeled on their counterparts in *E. coli* MutS (PDB: 2WTU) (35) using Modeller (36). ADP was modeled in subunit A using a related *Taq* MutS structure (PDB: 1FW6) (37). ADP was parameterized with Mg^{2+} coordinated by Ser-590, four water molecules, and an oxygen atom from the β-phosphorous using MCPB.py (38) methodology implemented in AMBER v18 (39) (Fig. S2). Gaussian (g16) (40) was used to estimate partial charges, bond distances, bond angles, torsions, and force constants for the ADP-Mg^{2+}-water-serine complex. The calculations were performed using the B3LYP level of theory and 6-31G++ basis set in implicit water conditions. Mg^{2+}-water bonds were treated using a nonbonded model with harmonic restraints while all other bonds were treated using the bonded model. The protonation state of amino acids was estimated using the H++ server (41). AMBER force field ff14SB (42) was utilized for amino acids and bsc0 (43) for DNA parameterization. Three additional DNA basepairs (G:C) were appended to the DNA to increase its length and modeled using Modeller. *Taq* MutS-homoduplex DNA was generated from MutS-T-bulge by removing the extra thymidine nucleotide in T-bulge DNA, followed by relaxation using Modeller and stepped minimization using AMBER v18 as described above. The t-LEaP module of AMBER v18 was used for hydrogen addition and general parametrization of amino acids and DNA basepairs.

MD simulations

MD simulations were performed on four constructs, two MutS-DNA complexes (10 μs each): 1) MutS-T-bulge in A$_{ADP}$-B$_0$ liganded state and 2) MutS-homoduplex in A$_{ADP}$-B$_0$ liganded state, and the two DNAs (15 μs each): 3) T-bulge and 4) homoduplex. Two additional independent 10 μs replicates of the MutS-homoduplex simulation were performed, starting with different initial velocities and atomic positions. MutS-DNA constructs were solvated in TIP3P (44) water molecules in a truncated octahedron box with 10 Å water clearance on each side. Na$^+$ and Cl$^-$ ions were used to achieve electroneutrality and bring the system to 150 mM ionic strength. Periodic boundary conditions were employed along with the particle mesh Ewald algorithm (45–47) implementing a 10 Å cutoff to handle long-range electrostatic interactions. SHAKE (48) was used to restrict the motion of bonded hydrogen atoms; this also enabled use of 2 fs time steps in the simulations. Temperature and pressure were controlled using a Berendsen thermostat and barostat (49), respectively, with isotropic position scaling for pressure regulation. Random seeds were utilized to initiate the simulations. Minimization was achieved by a 5-step process (each comprising 2000 steps of steepest decent followed by 3000 steps of conjugate gradient), successively decreasing harmonic constraints on the MutS-DNA and DNA constructs in each of the 5 steps (100, 50, 20, 10, and 0 kcal/mol). Each construct was then heated to 300 K over 10 ps of simulation time. Stepped equilibration was performed in steps of 50 ps with successively decreasing harmonic constraints (100, 50, 10, 1, and 0.1 kcal/mol). Additional equilibration of 20 ns was performed without any restraints on MutS or DNA. Root mean-square deviation (RMSD),

potential, kinetic, and total energy was monitored for structural and energetic stability during the equilibration. Production MD runs were carried out under NPT conditions. Trajectory frames were saved at intervals of 20 ps. All MD simulations were performed at Wesleyan University's High Performance Computing Cluster using AMBER v18's pmemd (50,51) implementation on nVIDIA GPUs.

Clustering, data visualization, and analysis

Dynamic cross-correlation matrix (DCCM) calculations, K-means clustering, and principal component analysis (PCA) were performed as implemented in the Cpptraj (52) module of AMBER. Clustering of the MutS-homoduplex trajectory data was performed based on RMSD. A shift analysis was performed by calculating the RMSD of each frame to the centroid of its assigned cluster and the centroids of all other clusters and shown as a frequency plot (53). The DNA substrates were analyzed using Curves+ (54). Incremental bending in the helical axis was determined at each basepair step in each frame of the trajectory and plotted as a function of simulation time. Images and videos were rendered using Gnuplot (55), VMD (56), PyMOL (57), and Chimera (58).

RESULTS AND DISCUSSION

Changes in MutS structural dynamics in the absence and presence of a mismatch

We performed several extended, unbiased MD simulations of two MutS-DNA complexes to compare the structural dynamics and interactions of MutS and DNA before and after mismatch detection. No high-resolution crystal structure of MutS bound to homoduplex DNA is available. Therefore, we remodeled DNA in the *Taq* MutS-T-bulge structure (PDB: 1NNE) by removing the extra unpaired thymidine in one strand to create homoduplex DNA. We also set the nucleotide-liganded state of the MutS dimer to A$_{ADP}$-B$_0$. MutS exhibits asymmetric ATPase activity (only one subunit hydrolyzes ATP rapidly in the absence of a mismatch) (59–61), and the A$_{ADP}$-B$_0$ state has been captured in many crystal structures of *E. coli* MutS bound to different mismatches in DNA (6,22). Both MutS-T-bulge and MutS-homoduplex complexes were simulated for 10 μs in explicit water conditions (note: two additional independent 10 μs simulations were performed for the MutS-homoduplex complex after analysis of the original trajectory revealed novel and dramatic conformational changes, described below). In addition, free T-bulge DNA and homoduplex DNA substrates in the absence of MutS were simulated for 15 μs. The workflow of modeling MutS and DNA followed by MD simulations and analysis is outlined in Fig. S1.

The RMSD values of each MutS subunit plotted over the entire simulation time show that MutS bound to T-bulge exhibits less conformational variability, with an average RMSD of 3.4 Å (0.35 SD) and 2.9 Å (0.18 SD) for subunits A and B in the T-bulge complex compared with 3.8 Å (0.8 SD) and 4 Å (0.86 SD) in the homoduplex complex, respectively (Fig. 2 *A*). This result is in accordance with the large number of crystal structures solved for MutS bound to mismatched DNA and none bound to homoduplex
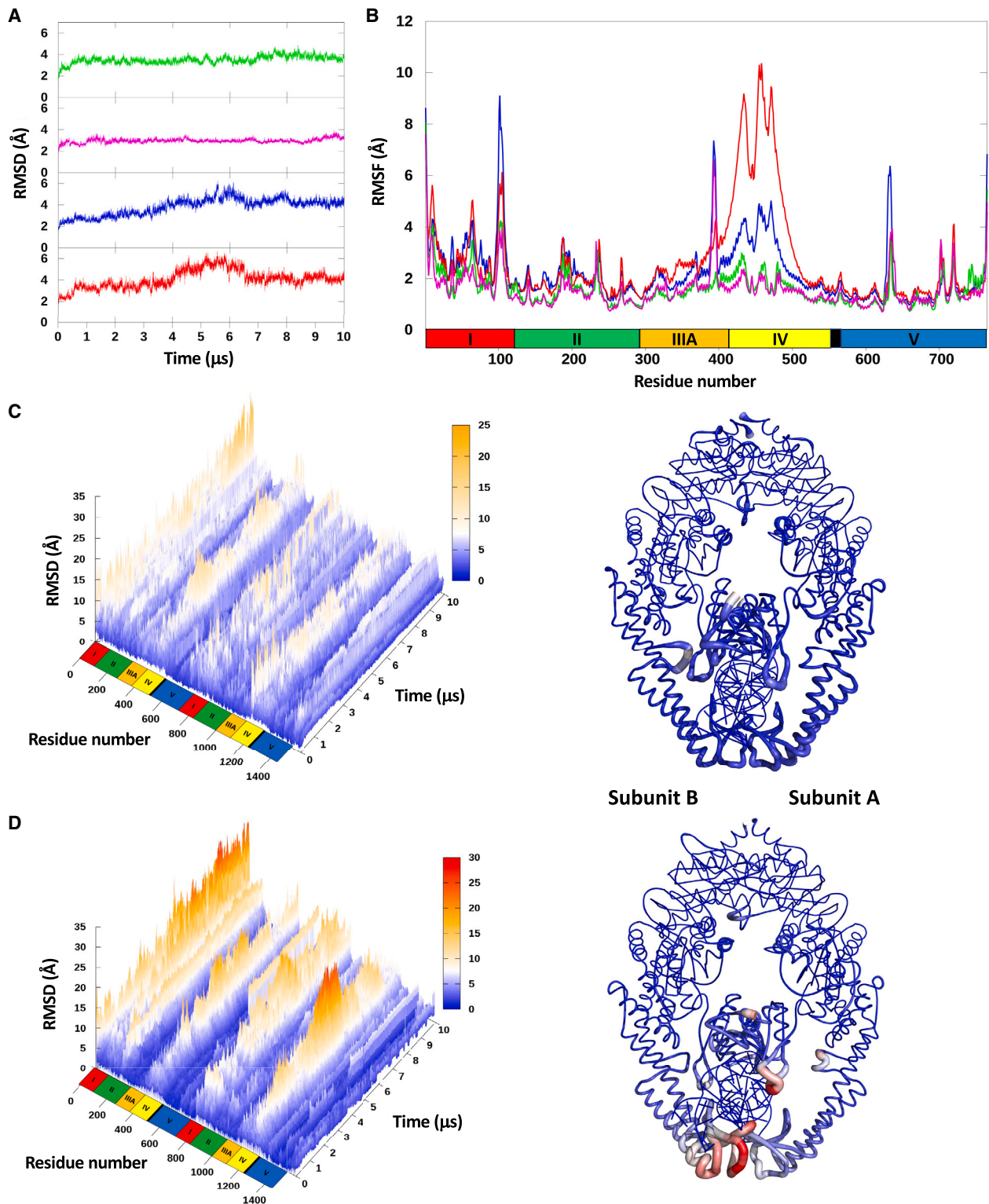
FIGURE 2 RMSD, RMSF, and B-factor values from MD simulations of MutS in complex with T-bulge (MutS-T-bulge) or homoduplex DNA (MutS-homoduplex). (*A*) RMSD as a function of time and (*B*) RMSF for each residue in each subunit of MutS-T-bulge (subunit A, *green*; B, *magenta*) and MutS-homoduplex (subunit A, *blue*; B, *red*); MutS domains I–V and residue numbers are noted below the graph. (*C*) RMSD as a function of both residue number and time for MutS-T-bulge (*left panel*, variation from initial conformation: low [*blue*] to high [*yellow*]). B-factor variation of MutS-T-bulge super-imposed on the initial structure (*right panel*; variation from initial conformation: low [*blue/thinner line*] to high [*red/thicker line*]). (*D*) RMSD as a function of both residue number and time for MutS-homoduplex (*left panel*; variation from initial conformation: low [*blue*] to high [*red*]). B-factor variation of MutS-homoduplex superimposed on the initial structure (*right panel*; variation from initial conformation: low [*blue/thinner line*] to high [*red/thicker line*]).

DNA. To resolve conformational variation at the amino acid level, we plotted the root mean-square fluctuation (RMSF) for each residue in the MutS subunits of both complexes (Fig. 2 B). Subunits A and B in the MutS-T-bulge complex did not diverge significantly from the initial structure, with RMSF values remaining at ~2 Å for most residues (sharp spikes are due to loop regions). In contrast, both subunits in the MutS-homoduplex complex diverged significantly from the initial structure. The deviations were located primarily in clamp domain IV and were more pronounced in subunit B compared with A, revealing an asymmetric response in the MutS dimer to removal of the extra thymidine from DNA. At the other end of MutS, the RMSF values for the ATPase and dimerization domain V remained steady in all the subunits, consistent with the high stability of this domain observed in previous structural studies (5,62).

To assess the structural variation in MutS amino acids as a function of time, we aligned the MD trajectory ensembles with the ATPase domain and determined the RMSD for each residue during the simulation. There is little variation in MutS bound to T-bulge, with per-residue RMSD remaining under 10 Å through most of the simulation (Fig. 2 C, left). A B-factor plot of the trajectory relative to the initial conformation illustrates that the structure remains the same overall (Fig. 2 C, right). However, MutS dynamics are strikingly different when bound to homoduplex DNA, with large variations in per-residue RMSD occurring through most of the 10 μs simulation, often rising to 20–25 Å in parts of the protein (Fig. 2 D, left). The most notable structural variance occurs in the mismatch binding domain I of subunit A and clamp domain IV of subunit B, which make

up the main "mismatch detection interface" and have the most extensive contacts with the unpaired thymidine and flanking nucleotides in the MutS-T-bulge complex (5) (Fig. 1). The B-factor plot illustrates the enhanced dynamics of these two domains when MutS is bound to homoduplex DNA (Fig. 2 D, right), in sharp contrast to when it is bound to T-bulge DNA (Fig. 2 C, right).

DCCMs were calculated for MutS in both complexes to examine the motions reported by the RMSD, RMSF, and B-factor values above (Fig. 3; summed over 10 μs). MutS bound to T-bulge displayed less motional correlation overall compared with MutS bound to homoduplex DNA (Fig. 3, A and B, respectively). The most prominent difference between the two complexes is in clamp domain IV, where positively correlated motions between subunits A and B in the MutS-T-bulge are negatively correlated in the MutS-homoduplex (black circle). These data indicate that the clamp domains move away from each other in the absence of a mismatch, which is consistent with reports of clamp separation in MD simulations of human MSH2-MSH6 in the absence of DNA (63) and in the recent cryo-EM structure of E. coli MutS bound to homoduplex DNA (30). The DCCMs plotted at 1 μs intervals revealed that the switch from positive to negative correlated motions in the MutS-homoduplex complex occurs at about 3 μs (consistent with RMSD data in Fig. 2 D), and remains negative through the end of the 10 μs simulation (Fig. S3). A difference plot of the DCCMs reveals more subtle details (Fig. 3 C). For instance, correlated motions in the subunit B clamp show greater deviation than in subunit A (black squares), indicating larger conformational changes in subunit B. Correlated motions between the
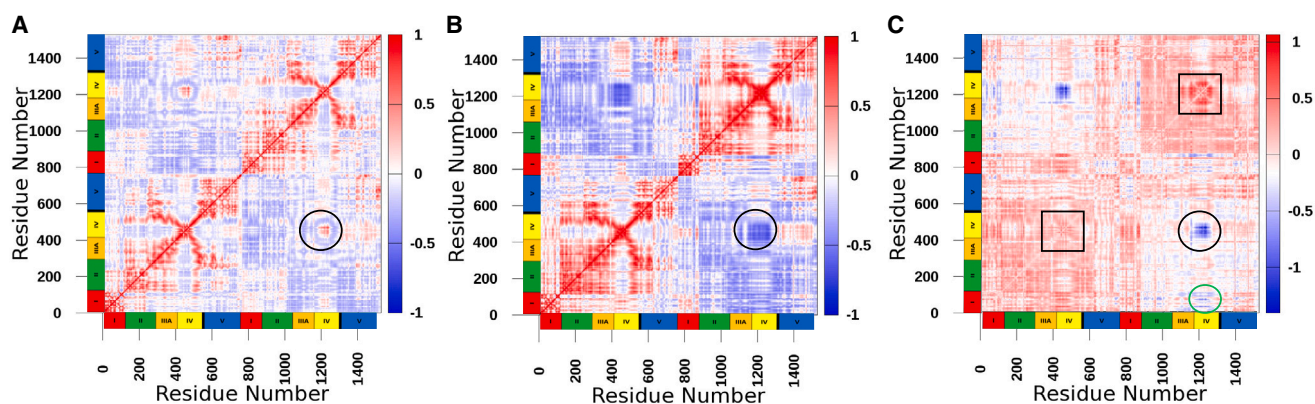

FIGURE 3 Dynamic cross correlation matrices of MutS conformations summed over the MD simulation time (10 μs). The color scale ranges from red (high positive correlation) to blue (high negative correlation) in (A and B), and domains I–V and residue numbers of MutS subunits A and B are noted on the axis (left to right). (A) DCCM of MutS from the complex with T-bulge DNA. The black circle highlights a region of positive structural cross correlation between the clamp domains (IV) of subunits A and B. (B) DCCM of MutS from the complex with homoduplex DNA. The black circle highlights the same region, which now shows negative structural cross correlation between the clamps. (C) Difference plot between DCCMs of MutS from homoduplex and T-bulge complexes. The color scale ranges from red (high positive difference) to blue (high negative difference). Black squares highlight regions of higher deviation in subunit B clamp from the initial conformation compared with subunit A. The black circle highlights correlation between subunit A and B clamps. The green circle highlights correlation between subunit A mismatch binding domain (I) and subunit B clamp domain (IV) that occurs through contacts with DNA (see Fig. 6).

subunit A mismatch binding domain I and subunit B clamp domain also differ between the two MutS complexes (Fig. 3 C, *green circle*), indicating concerted actions by the entire mismatch detection interface during the search and recognition process.

Changes in DNA structural dynamics in the absence and presence of a mismatch

We also examined the role of DNA dynamics in differential interactions of MutS with homoduplex DNA versus mismatched DNA. The DNA substrates from MD simulations of both MutS-bound complexes were analyzed using Curves+ (54) to determine the degree of axial bending at each basepair step in each frame of the trajectory, which was plotted versus the simulation time (Fig. 4, A and B). The T-bulge and homoduplex DNA were also simulated independently of MutS for 15 μs and similarly analyzed to account for the effects of MutS on DNA dynamics (Fig. 4, C and D). MutS-bound T-bulge DNA was sharply kinked at the start and remained bent through the simulation, as expected from crystal structures (Fig. 4 A, *middle red band* in the trajectory; T-bulge at position 13); note: bending is elevated at the free DNA ends and exacerbated by an unpaired nucleotide at position 1 (PDB: 1NNE). In comparison, the free T-bulge DNA in the absence of MutS exhibited modest axial bending to various degrees through

the 15 μs simulation (Fig. 4 C), in accordance with the conformational flexibility reported for DNA containing nucleotide insertions (64). The global bending propensity of this DNA in addition to local deformability at the T-bulge likely favors the kinked state stabilized by MutS in the mismatch-recognition complex (Figs. 1 and 4 A). This could explain the higher affinity of MutS for DNA containing an insertion compared with a mismatched basepair (65), which is not as intrinsically prone to bending (66,67).

MutS-bound homoduplex DNA was bent by 45° in the starting complex formed by removal of the extra thymidine from T-bulge DNA in the crystal structure. This state lasted 0.48 μs, and then the homoduplex exhibited transient periods of unbending and bending between 0.48 and 5.5 μs (Fig. 4 B). After 5.5 μs, the homoduplex remained mostly unbent through to the end of the 10 μs simulation. In comparison, free homoduplex DNA in the absence of MutS did not display any axial bending and remained straight through the 15 μs simulation (Fig. 4 D). These results affirm that sharp kinking of homoduplex DNA is unfavorable to such an extent that it straightens out in the MutS-homoduplex complex despite structural hindrance from the protein in the initial conformation. As described in Figs. 2 and 3, MutS correspondingly converts to a more open state in which the clamp domains separate from each other. MutS-bound homoduplex DNA has been observed in straight and bent conformations by cryo-EM and atomic force
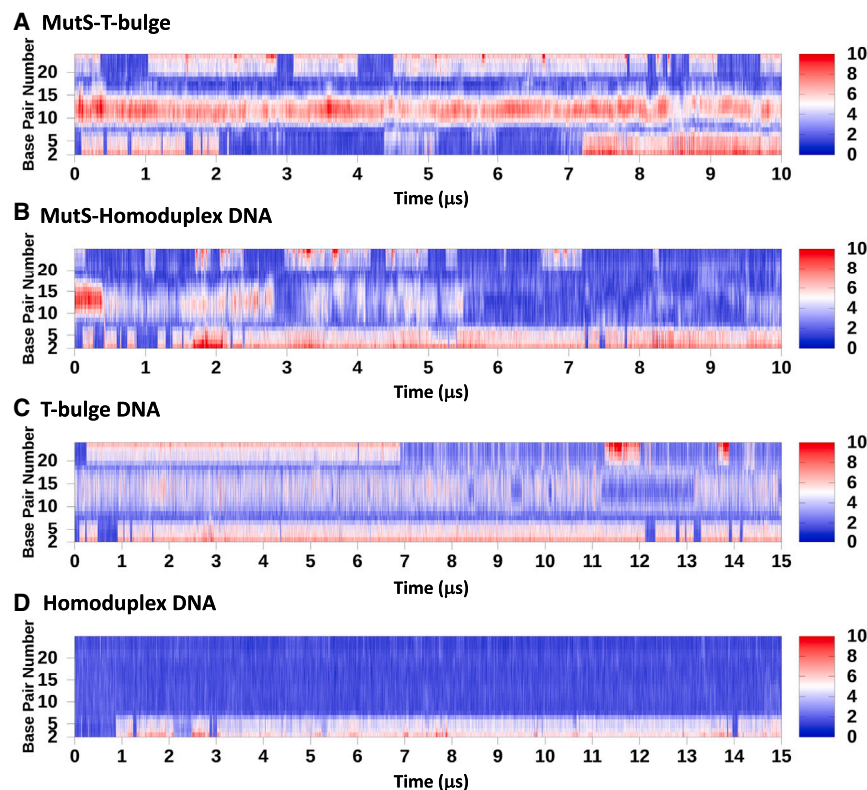

FIGURE 4 Axial DNA bending as a function of simulation time. Dynamic DNA bending at each basepair step was calculated in degrees (0–10°) using Curves+ software. The color scale ranges from red (high bend angle) to blue (low bend angle) and nucleotide numbers 1–23 are noted on the left (T-bulge at position 13). DNA bending in (A) MutS-T-bulge complex and (B) MutS-homoduplex complex. (C) T-bulge DNA and (D) homoduplex DNA bending in the absence of MutS. Note: elevated bending near the DNA end is because nucleotide 1 (G) is unpaired.

microscopy (AFM), respectively (30,68,69). The MD trajectories reveal that these conformations are interconvertible, with DNA transitioning between straight-bent states and MutS between open-closed clamp states. These MutS-DNA dynamics may constitute a scanning mechanism for errors in DNA.

MutS and DNA conformational states on the path to finding a mismatch

We analyzed the MutS-homoduplex trajectory data by principal component analysis and clustering (70) to address our hypothesis that the observed conformational dynamics constitute a mismatch scanning and detection mechanism. The top 10 principal components (PC) are shown in Fig. S4, and the first two components (PC1 and PC2)

explain 70.64% of the conformational variation. PC1 mainly comprises motions of the clamp domains moving apart, with accompanying adjustment of the mismatch binding domains and translation of DNA (Video S1). PC2 mainly comprises DNA unbending motions (Video S2). Together these components account for the dominant changes in MutS-homoduplex during the simulation. Plotting the trajectory frames as a function of PC1 and PC2 parsed the data into nine different conformational groups (Fig. 5 A, black dots are the centroids). A shift analysis was performed to assess the separation between the nine clusters by calculating the RMSD of each frame to the centroid of its assigned cluster and the centroids of all other clusters. The data in Fig. S5 show that the frames in a given cluster lie closer to their own centroid and are distant from the centroids of the other clusters, which supports their classification as distinct
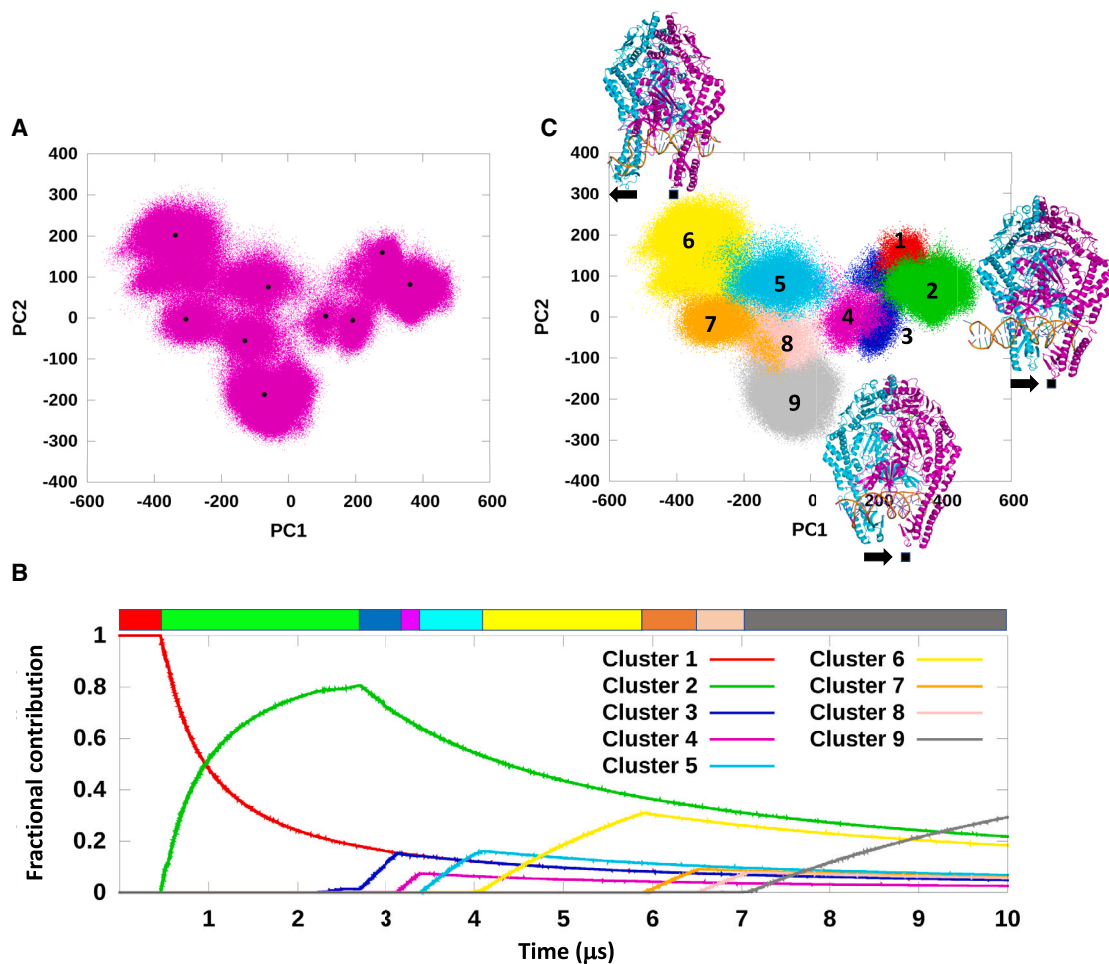

FIGURE 5 Principal component and clustering analysis of the MutS-homoduplex DNA complex MD simulation. (*A*) MutS-homoduplex conformations projected as a function of the top 2 principal components of PCA (*pink dots*). The data sort into nine structurally distinct clusters (centered at the *black dots*). (*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 (substate) is numbered in the order of appearance and has a different color (substate 1, *red*; 2, *green*; 3, *blue*; 4, *magenta*; 5, *cyan*; 6, *yellow*; 7, *orange*; 8, *pink*; and 9, *gray*). (*C*) Individual substates from the clustering analysis projected as a function of PC1 and PC2. Conformations of prominent substates 2, 6, and 9 are shown (see Fig. 6 and Table S1). Similar results were obtained from two replicate MutS-homoduplex simulations, shown in Figs. S6 and S7.

conformational populations of the MutS-homoduplex complex. Further analysis of the full trajectory data by K-means clustering based on RMSD between conformations provided additional support that the nine clusters are structurally distinct from each other. Moreover, the fractional composition of each cluster plotted over time showed that they form and decay sequentially through the simulation (Fig. 5 B; clusters numbered 1–9 in order of appearance). Fig. 5 C shows the clustering data projected onto the PC1 and PC2 components, colored and numbered 1 through 9 by cluster as identified in Fig. 5 B. Together, the results indicate a kinetic pathway in which the MutS-homoduplex complex transitions between bent and straight DNA conformations through a series of different substates (Videos S3 and S4; detailed description provided below).

A concern that these striking results are from a single, albeit long, simulation was addressed by performing two additional 10 μs MD simulations of the MutS-homoduplex complex, starting from different initial velocities and structures. The data were analyzed as described above and the results are shown in Figs. S6 and S7. These new simulations again reveal clustering of MutS-homoduplex into nine distinct conformational substates that transition in series along the trajectory. We also see evidence that substates 8 and 9 can transition to substates 3–4, suggesting that the complex can cycle through parts of the pathway. We also tested the significance of the transitions between substates individually by selecting representative structures from each one 1–8 and simulating them independently for at least 2 μs each. The trajectories were projected onto to the PC1 and PC2 components obtained from the original full 10 μs simulation (Fig. S8). The data confirm that each substate can transition to the next one in the pathway, and indicate that some steps maybe reversible (e.g., between substates 5 and 6).

The nine MutS-homoduplex substates have different lifetimes, reflecting varying degrees of stability (Fig. 5 B and Table S1). Three substates have longer lifetimes than the others: 2 (2.18 μs), 6 (1.83 μs), and 9 (2.93 μs). Fig. 6 illustrates representative structures of these three prominent substates. Similar structures were obtained for substates 2, 6, and 9 from the two replicate MutS-homoduplex trajectories as well, further highlighting the significance of these conformational populations in MutS-homoduplex dynamics (Figs. S6 and S7). In substate 2, the DNA has almost straightened out from the initial bent state; however, there is little change in the overall MutS conformation and the clamp domains remain closed, reflecting the high stability of the dimeric θ structure (Fig. 6 A; clamps separated by 18 Å). In substate 6, the subunit B clamp has moved forward on DNA relative to subunit A, resulting in a separation of 36 Å between the two clamp domains (Fig. 6 B); Fig. S6 shows that this separation can be as wide as 50 Å. Notably, the DNA is bent, held out of plane through newly formed contacts with the subunit B clamp. In substate 9, the subunit B clamp has moved back closer to subunit A, although the two clamp

domains remain partially open (separated by 27 Å) and the DNA is fully straight (Fig. 6 C). A coarse-grained simulation of human MSH2-MSH6 has shown that the protein does not dissociate from DNA even when the clamp domains are modeled in a relatively open position (71). This finding provides strong support for our proposition, discussed in detail below, that the open-to-closed conformational dynamics of clamp domains underlie the mismatch search mechanism employed by MutS as it diffuses on homoduplex DNA (E. coli MutS diffuses on DNA for ~1 s per binding event (11)). In addition to clamp domains, the mismatch binding domains also undergo rearrangement in concert with the changes in DNA. An overlay of the initial and final complex conformations highlights the changes in the mismatch binding and clamp domains of subunit A (Fig. 6 D) and subunit B (Fig. 6 E) and the corresponding changes in homoduplex DNA.

Representative structures of all substates 1 through 9 can be viewed in series in Videos S3 and S4. Homoduplex DNA adopts a range of straight-to-bent conformations and MutS adopts open-to-closed clamp conformations in these substates. After DNA first straightens out in substate 2, subsequent movements evoke a cantilever, with MutS anchoring the DNA polymer through relatively stable contacts at one location with the mismatch binding and clamp domains of subunit A (*magenta*) and enabling it to bend (substate 6) and unbend (substate 9) through dynamic contacts at other locations with the mobile clamp domain of subunit B (*cyan*). The videos also provide a preview of interactions between MutS residues and DNA, whose role in interrogating DNA to detect a basepairing error is discussed in detail below.

MutS-DNA interactions assess DNA shape, global bending, and local basepair deformability in the search for mismatches

Since the first crystal structures of *E. coli* and *Taq* MutS encircling mismatched DNA were solved (5,6), it has been proposed that the protein employs an indirect readout mechanism, testing the conformational flexibility of DNA to find sites prone to bending and deformation. The conformational dynamics of *Taq* MutS-homoduplex DNA complex reported here provide, for the first time, a molecular underpinning for this hypothesis. Fig. 7 illustrates several interactions between MutS residues and homoduplex DNA in substate 9 (straight DNA) and substate 6 (bent DNA), selected on the basis of contact frequency (>20% of all frames in the substate trajectory have at least one hydrogen bond; Table S2). The interactions can be categorized into three functional classes. The first set of residues is distributed across the mismatch binding domains of both subunits and contacts the sugar-phosphate backbone in both substates 9 and 6 (Fig. 7, A and B, *blue labels*). Ser-63$^A_{mm}$ and Lys-64$^A_{mm}$ from subunit A and Lys-64$^B_{mm}$ and Arg-110$^B_{mm}$ from
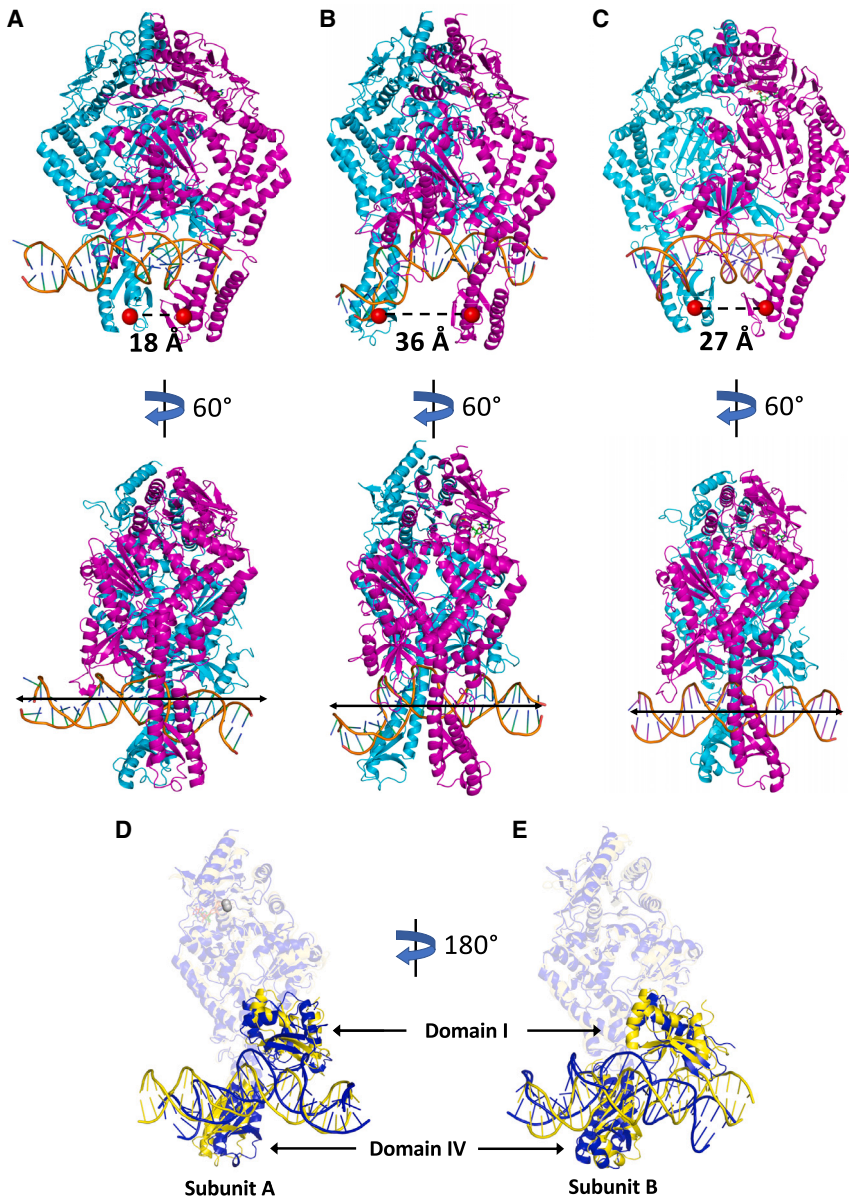
FIGURE 6 Prominent MutS-homoduplex DNA complex substates observed during MD simulation. Subunit A, magenta; subunit B, cyan (*A–C*); distance between clamp domains IV (*dashed line*). (*A*) Substate 2 (straightened DNA, clamps very close), (*B*) substate 6 (bent DNA, clamps farthest apart), and (*C*) substate 9 (straight DNA, clamps close). Corresponding 60° rotated views show DNA conformation. Similar substate 2, 6, and 9 conformations were obtained from the two replicate MutS-homoduplex simulations as well (Figs. S6 and S7). (*D*) Overlay of initial (*blue*) and final (*yellow*) subunit A conformations, showing rearrangement of the mismatch binding (I) and clamp (IV) domains and coupled DNA unbending (subunit B is hidden and the other subunit A domains are translucent). (*E*) 180° rotation shows the analogous view of subunit B.

subunit B bookend about 1.5 helical turn and could check DNA shape as MutS undergoes 1D rotational diffusion along the axis in scanning mode (11). Arg-76$_{mm}^{A}$ is an additional point of contact with the DNA backbone midway between the flanking Lys-64 residues. The second set of residues is concentrated in the clamp domains of both subunits and is involved in large conformational changes that could check the global bendability of DNA (Fig. 7, *A* and *B*, *green labels*; Videos S3 and S4). Arg-473$_{clamp}^{A}$ and Arg-475$_{clamp}^{A}$ from subunit A maintain contact with the same site on DNA through the entire MD trajectory (along with Arg-76$_{mm}^{A}$ from the mismatch binding domain). In stark contrast, Asn-443$_{clamp}^{B}$, Lys-471$_{clamp}^{B}$, Arg-473$_{clamp}^{B}$, Arg-475$_{clamp}^{B}$ from subunit B contact different sites on DNA when the clamp extends to a fully open conformation as

the DNA bends (substate 6) or retracts to a partially closed conformation as the DNA straightens (substate 9); note: the highly conserved Lys-439$_{clamp}^{B}$ residue is also likely part of this set, but does not present as a high-frequency contact because the DNA is short. Extension and retraction of the subunit B clamp coupled with DNA bending and unbending, respectively, can be seen clearly in Videos S3 and S4 (substates 1–9 linked in series). Finally, at the center of the MutS footprint on DNA, Asp-38$_{mm}^{A}$ remains inserted in the minor groove during the entire trajectory while Lys-471$_{clamp}^{B}$ is inserted in the corresponding major groove when the DNA is straight, as in substate 9 (Fig. 7, *A* and *B*, *orange labels*). These two highly conserved residues evoke a pincer that could check local deformability by destabilizing stacking and hydrogen bonding between bases (5–7).
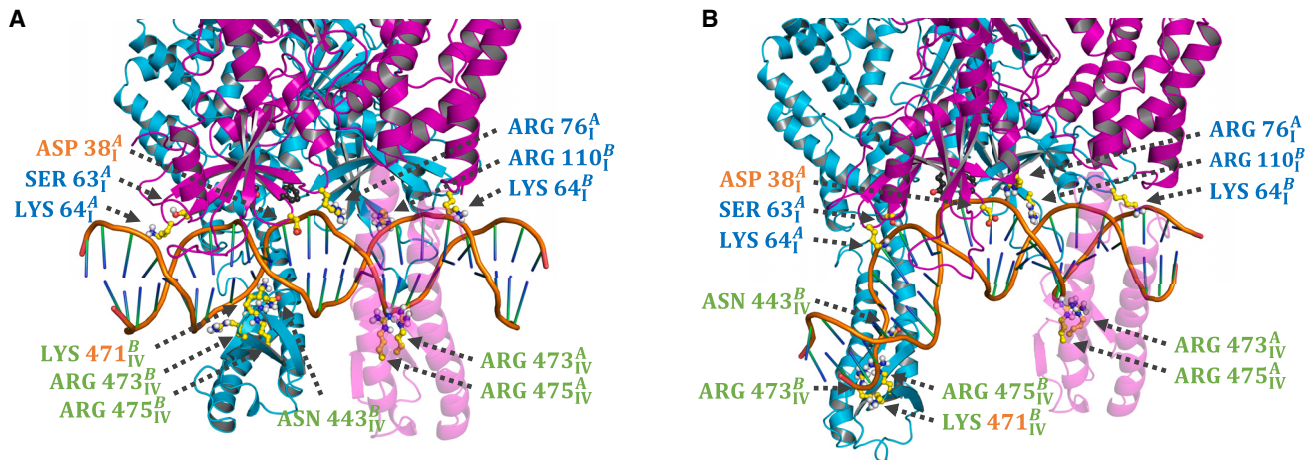
FIGURE 7 MutS-DNA interactions constituting the mismatch search mechanism. Three classes of interactions are shown for (*A*) substate 9 and (*B*) substate 6 from the MD trajectory of MutS-homoduplex DNA complex. Subunit A, magenta; subunit B, cyan; residues shown as ball-and-stick, yellow; subunit A Phe-39 and Glu-41 that lock in the mismatch-recognition complex at a bona fide error, black. Contacts from the mismatch binding (I) and clamp (IV) domains are shown that could enable MutS to check 1) the shape of DNA over ~1.5 helical turn (*blue labels*), 2) the bendability of DNA (*green labels*) as subunit B clamp extends (substate 6) and retracts (substate 9) along the axis, and 3) the local deformability of basepairs at the center of the MutS binding site on DNA (*orange labels*).

Figure360▷ For a Figure360 author presentation of this figure, see https://doi.org/10.1016/j.bpj.2023.06.006.

We noted that Phe-39$^A_{mm}$, which undergoes the signature π-π stacking with the mispaired/unpaired base in the mismatch-recognition complex, is twice as flexible in the MutS-homoduplex versus the MutS-T-bulge trajectory. This finding is consistent with previous AFM studies that showed similar bending of homoduplex DNA by *Taq* MutS wild-type and Phe-39 (Ala) mutant (69), suggesting that Phe-39 does not make significant contact with the homoduplex during the mismatch search process. On the other hand, almost all of the residues listed above that undergo nonspecific electrostatic and hydrogen bonding interactions with the homoduplex also interact with kinked T-bulge DNA in the mismatch-recognition complex, reflecting their comprehensive role in both search and recognition mechanisms. Moreover, these residues are also conserved in diverse organisms, including *E. coli* MutS and human, mouse, and/or *S. cerevisiae* Msh2-Msh6, affirming their importance in this fundamental DNA repair pathway (5–7). Structural alignment of the *Taq* MutS (PDB: 1NNE) and human MSH2-MSH6 (PDB: 2O8F) complexes with T-bulge DNA shows that the conserved residues are positioned to perform analogous functions in both proteins (Fig. S9 and Table S3). Several are also reported in the InSiGHT database of variants for Lynch syndrome, although the effects of the point mutations are unknown (72).

Together the dynamic interactions described above constitute an indirect readout mechanism that combines a coarse search for regions of unusual conformational flexibility in DNA and a fine search that discriminates between a canonical basepair and a mismatch (Videos S3 and S4). At a coarse level, MutS switches between conformations that enable DNA bending and unbending as it diffuses along the helix,

and the lower energetic cost of bending a mismatch-containing segment could help localize the protein to a potential target site. These actions alone are insufficient as MD studies of DNA bending and deformability have shown that bending propensity contributes to mismatch detection but does not confer high enough selectivity (29,67), which is consistent with reports from AFM analysis of MutS-homoduplex DNA complexes as well (69). At a fine level, MutS interactions could preferentially promote destabilization and opening of a mismatched basepair and enable kinking of DNA. This mechanism is consistent with the report that basepair opening energetics account for the higher affinity of MutS for mismatched over homoduplex DNA (73). At this stage, the signature Phe-39$^A_{mm}$ and Glu-41$^A_{mm}$ residues from subunit A can stack and hydrogen bond with the displaced base and lock in the mismatch-recognition complex (Fig. 1).

Implications of asymmetry in MutS structure and function for mismatch repair

Despite being a homodimer, bacterial MutS is a functional heterodimer whose subunits have asymmetric mismatch recognition and ATPase activities (9). Our data reveal striking DNA binding asymmetry even in the mismatch search mechanism, with distinct roles for the two clamp domains in bending DNA. It is not known if the subunits originally designated as A and B in the homodimer (from the MutS-T-bulge complex) retain their roles while scanning homoduplex DNA, or if they can be switched. A recent study of mixed wild-type/mutant *E. coli* MutS dimers found that subunit functions are exchangeable for initiation of repair after mismatch recognition (30), which supports

the possibility that switching occurs during the mismatch search phase as well. In eukaryotic heterodimers, DNA binding asymmetry is codified, with Msh6 and Msh2 serving as subunits A and B, respectively. The division of labor in the mismatch search mechanism employed by bacterial homodimers may have allowed Msh2 to retain the common function of checking global DNA bendability through evolution, while Msh6 and Msh3 specialized to check local basepair stability for mismatches or small insertion/deletions and large insertion/deletion loops, respectively (7,8). Part of this specialization appears to be the ability of Msh2-Msh6 to stay in tighter contact and diffuse slower on DNA as it scans for small discrepancies compared with Msh2-Msh3, which exhibits faster diffusion and more hopping (dissociation) events as it scans for large discrepancies in the double helix (71).

Due to its asymmetric ATPase activity, the MutS dimer can adopt nine different nucleotide-liganded states. Some of these states are coupled to specific MutS actions on DNA, e.g., after mismatch recognition, MutS converts to an A_{ATP}-B_{ATP} state that can interact with MutL (74,75). While the computationally intensive full 10 μs simulations were performed with MutS-T-bulge and MutS-homoduplex complexes in A_{ADP}-B_0 state, we also simulated substate 6 of MutS-homoduplex for 1 μs each in the remaining eight nucleotide-liganded states and found several capable of transitioning between substates 6 and 9. It is possible that the DNA binding and ATPase activities are uncoupled in the mismatch search phase such that the properties of DNA rather than the ATPase reaction govern MutS actions as it diffuses along DNA to find an error. Once a mismatch is recognized, the two activities may be tightly coupled again to trigger ADP-ATP exchange and stabilize MutS in the A_{ATP}-B_{ATP} state. Finally, while the precise mechanism of allosteric communication between the distant mismatch binding and ATPase sites is unclear, many of the MutS residues that contact homoduplex DNA identified in this study are also part of a coevolved network of contiguous residues that connect the two sites and could modulate their activities (Asp-38, Arg-76, Arg-110, Asn-443, and Arg-475) (76). This network, identified by statistical coupling analysis, also contains residues that contact kinked DNA in the mismatch-recognition complex, including Phe-39 and Glu-41. It is an intriguing possibility that distinct but overlapping subnetworks could convey information specific to the mismatch search, recognition, and repair initiation phases of mismatch repair between the DNA binding and ATPase sites. As more dynamic views of MutS actions on DNA become available from structural and computational studies, we can look forward to higher-resolution understanding of the DNA mismatch repair mechanism.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2023.06.006.

REFERENCES

1. Kunkel, T. A., and D. A. Erie. 2015. Eukaryotic mismatch repair in relation to DNA replication. *Annu. Rev. Genet.* 49:291–313.

2. Jiricny, J. 2013. Postreplicative mismatch repair. *Cold Spring Harb. Perspect. Biol.* 5:a012633.

3. Spies, M., and R. Fishel. 2015. Mismatch repair during homologous and homeologous recombination. *Cold Spring Harb. Perspect. Biol.* 7:a022657.

4. Li, Z., X. Li, …, D. Zhang. 2016. DNA mismatch repair and the DNA damage response. *Food Chem.* 202:94–98.

5. Obmolova, G., C. Ban, …, W. Yang. 2000. Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature.* 407:703–710.

6. Lamers, M. H., A. Perrakis, …, T. K. Sixma. 2000. The crystal structure of DNA mismatch repair protein MutS binding to a G· T mismatch. *Nature.* 407:711–717.

7. Warren, J. J., T. J. Pohlhaus, …, L. S. Beese. 2007. Structure of the human MutSα DNA lesion recognition complex. *Mol. Cell.* 26:579–592.

8. Gupta, S., M. Gellert, and W. Yang. 2011. Mechanism of mismatch recognition revealed by human MutSβ bound to unpaired DNA loops. *Nat. Struct. Mol. Biol.* 19:72–78.

9. Hingorani, M. M. 2016. Mismatch binding, ADP–ATP exchange and intramolecular signaling during mismatch repair. *DNA Repair.* 38:24–31.

10. Gradia, S., D. Subramanian, …, R. Fishel. 1999. hMSH2–hMSH6 forms a hydrolysis-independent sliding clamp on mismatched DNA. *Mol. Cell.* 3:255–261.

11. Lee, J.-B., W.-K. Cho, …, R. Fishel. 2014. Single-molecule views of MutS on mismatched DNA. *DNA Repair.* 20:82–93.

12. Sharma, A., C. Doucette, …, M. M. Hingorani. 2013. Slow conformational changes in MutS and DNA direct ordered transitions between mismatch search, recognition and signaling of DNA repair. *J. Mol. Biol.* 425:4192–4205.

13. Borsellini, A., V. Kunetsky, …, M. H. Lamers. 2022. Cryogenic electron microscopy structures reveal how ATP and DNA binding in MutS coordinates sequential steps of DNA mismatch repair. *Nat. Struct. Mol. Biol.* 29:59–66.

14. Kadyrov, F. A., L. Dzantiev, …, P. Modrich. 2006. Endonucleolytic function of MutLα in human mismatch repair. *Cell.* 126:297–308.

15. Acharya, S., P. L. Foster, …, R. Fishel. 2003. The coordinated functions of the E. coli MutS and MutL proteins in mismatch repair. *Mol. Cell.* 12:233–246.

16. Schofield, M. J., S. Nayak, …, P. Hsieh. 2001. Interaction of Escherichia coli MutS and MutL at a DNA mismatch. *J. Biol. Chem.* 276:28291–28299.

17. Grilley, M., J. Griffith, and P. Modrich. 1993. Bidirectional excision in methyl-directed mismatch repair. *J. Biol. Chem.* 268:11830–11837.

18. Constantin, N., L. Dzantiev, …, P. Modrich. 2005. Human mismatch repair: reconstitution of a nick-directed bidirectional reaction. *J. Biol. Chem.* 280:39752–39761.

19. Jeon, Y., D. Kim, …, J.-B. Lee. 2016. Dynamic control of strand excision during human DNA mismatch repair. *Proc. Natl. Acad. Sci. USA.* 113:3281–3286.

20. Lynch, H. T., C. L. Snyder, …, M. P. Hitchins. 2015. Milestones of Lynch syndrome: 1895–2015. *Nat. Rev. Cancer.* 15:181–194.

21. Peltomäki, P. 2003. Role of DNA mismatch repair defects in the pathogenesis of human cancer. *J. Clin. Oncol.* 21:1174–1179.

22. Natrajan, G., M. H. Lamers, …, T. K. Sixma. 2003. Structures of Escherichia coli DNA mismatch repair enzyme MutS in complex with different mismatches: a common recognition mode for diverse substrates. *Nucleic Acids Res.* 31:4814–4821.

23. Jones, M., R. Wagner, and M. Radman. 1987. Repair of a mismatch is influenced by the base composition of the surrounding nucleotide sequence. *Genetics.* 115:605–610.

24. Mazurek, A., C. N. Johnson, …, R. Fishel. 2009. Sequence context effect for hMSH2-hMSH6 mismatch-dependent activation. *Proc. Natl. Acad. Sci. USA.* 106:4177–4182.

25. Gorman, J., A. Chowdhury, …, E. C. Greene. 2007. Dynamic basis for one-dimensional DNA scanning by the mismatch repair complex Msh2-Msh6. *Mol. Cell.* 28:359–370.

26. Qiu, R., V. C. DeRocco, …, K. R. Weninger. 2012. Large conformational changes in MutS during DNA scanning, mismatch recognition and repair signalling. *EMBO J.* 31:2528–2540.

27. Yang, W. 2008. Structure and mechanism for DNA lesion recognition. *Cell Res.* 18:184–197.

28. Isaacs, R. J., and H. P. Spielmann. 2004. A model for initial DNA lesion recognition by NER and MMR based on local conformational flexibility. *DNA Repair.* 3:455–464.

29. Bouchal, T., I. Durník, and P. Kulhánek. 2021. Bending of Canonical and G/T Mismatched DNAs. *J. Chem. Inf. Model.* 61:6000–6011.

30. Fernandez-Leiro, R., D. Bhairosing-Kok, …, M. H. Lamers. 2021. The selection process of licensing a DNA mismatch for repair. *Nat. Struct. Mol. Biol.* 28:373–381.

31. Bhairosing-Kok, D., F. S. Groothuizen, …, T. K. Sixma. 2019. Sharp kinking of a coiled-coil in MutS allows DNA binding and release. *Nucleic Acids Res.* 47:8888–8898.

32. The UniProt Consortium. 2017. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 45:D158–D169.

33. Bank, P. D. 1971. Protein data bank. *Nature New Biol.* 233:223.

34. Alani, E., J. Y. Lee, …, W. Yang. 2003. Crystal structure and biochemical analysis of the MutS· ADP· beryllium fluoride complex suggests a conserved mechanism for ATP interactions in mismatch repair. *J. Biol. Chem.* 278:16088–16094.

35. Lebbink, J. H. G., A. Fish, …, T. K. Sixma. 2010. Magnesium coordination controls the molecular switch function of DNA mismatch repair protein MutS. *J. Biol. Chem.* 285:13131–13141.

36. Webb, B., and A. Sali. 2016. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* 86:5–6.

37. Junop, M. S., G. Obmolova, …, W. Yang. 2001. Composite active site of an ABC ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair. *Mol. Cell.* 7:1–12.

38. Li, P., and K. M. Merz, Jr. 2016. MCPB. Py: A python Based Metal Center Parameter Builder.

39. Case, D. A., I. Y. Ben-Shalom, …, A. W. Goetz. 2018. AMBER 2018. University of California, San Francisco.

40. Frisch, M.J., G.W. Trucks, … Hratch. Gaussian 16, Revision C.01. .

41. Anandakrishnan, R., B. Aguilar, and A. V. Onufriev. 2012. H++ 3.0: automating p K prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* 40:W537–W541.

42. Maier, J. A., C. Martinez, …, C. Simmerling. 2015. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11:3696–3713.

43. Galindo-Murillo, R., J. C. Robertson, …, T. E. Cheatham, III. 2016. Assessing the current state of amber force field modifications for DNA. *J. Chem. Theory Comput.* 12:4114–4127.

44. Jorgensen, W. L., J. Chandrasekhar, …, M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

45. Essmann, U., L. Perera, …, L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.

46. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.

47. Crowley, M., T. Darden, …, D. Deerfield Ii. 1997. Adventures in improving the scaling and accuracy of a parallel molecular dynamics program. *J. Supercomput.* 11:255–278.

48. Ryckaert, J.-P., G. Ciccotti, and H. J. Berendsen. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23:327–341.

49. Berendsen, H. J. C., J. P. M. Postma, …, J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.

50. Salomon-Ferrer, R., A. W. Götz, …, R. C. Walker. 2013. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J. Chem. Theory Comput.* 9:3878–3888.

51. Götz, A. W., M. J. Williamson, …, R. C. Walker. 2012. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. *J. Chem. Theory Comput.* 8:1542–1555.

52. Roe, D. R., and T. E. Cheatham, III. 2013. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* 9:3084–3095.

53. Thayer, K. M., B. Lakhani, and D. L. Beveridge. 2017. Molecular dynamics–Markov state model of protein ligand binding and Allostery in CRIB-PDZ: conformational selection and induced fit. *J. Phys. Chem. B.* 121:5509–5514.

54. Blanchet, C., M. Pasi, …, R. Lavery. 2011. CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. *Nucleic Acids Res.* 39:W68–W73.

55. Janert, P. K. 2016. Gnuplot in Action: Understanding Data with Graphs. Simon and Schuster.

56. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38.

57. DeLano, W. L. 2002. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter on protein crystallography.* 40:82–92.

58. Pettersen, E. F., T. D. Goddard, …, T. E. Ferrin. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25:1605–1612.

59. Antony, E., S. Khubchandani, …, M. M. Hingorani. 2006. Contribution of Msh2 and Msh6 subunits to the asymmetric ATPase and DNA mismatch binding activities of Saccharomyces cerevisiae Msh2–Msh6 mismatch repair protein. *DNA Repair.* 5:153–162.

60. Antony, E., and M. M. Hingorani. 2004. Asymmetric ATP binding and hydrolysis activity of the Thermus aquaticus MutS dimer is key to modulation of its interactions with mismatched DNA. *Biochemistry.* 43:13115–13128.

61. Antony, E., and M. M. Hingorani. 2003. Mismatch recognition-coupled stabilization of Msh2-Msh6 in an ATP-bound state at the initiation of DNA repair. *Biochemistry.* 42:7682–7693.

62. Mukherjee, S., S. M. Law, and M. Feig. 2009. Deciphering the mismatch recognition cycle in MutS and MSH2-MSH6 using normal-mode analysis. *Biophys. J.* 96:1707–1720.

63. Mukherjee, S., and M. Feig. 2009. Conformational change in MSH2-MSH6 upon binding DNA coupled to ATPase activity. *Biophys. J.* 96:L63–L65.

64. Kalnik, M. W., D. G. Norman, …, D. J. Patel. 1990. Conformational transitions in thymidine bulge-containing deoxytridecanucleotide duplexes. Role of flanking sequence and temperature in modulating the equilibrium between looped out and stacked thymidine bulge states. *J. Biol. Chem.* 265:636–647.

65. Yang, Y., L. E. Sass, …, D. A. Erie. 2005. Determination of protein–DNA binding constants and specificities from statistical analyses of single molecules: MutS–DNA interactions. *Nucleic Acids Res.* 33:4322–4334.

66. Rossetti, G., P. D. Dans, …, M. Orozco. 2015. The structural impact of DNA mismatches. *Nucleic Acids Res.* 43:4309–4321.

67. Sharma, M., A. V. Predeus, …, M. Feig. 2013. DNA bending propensity in the presence of base mismatches: implications for DNA repair. *J. Phys. Chem. B.* 117:6194–6205.

68. Wang, H., Y. Yang, …, D. A. Erie. 2003. DNA bending and unbending by MutS govern mismatch recognition and specificity. *Proc. Natl. Acad. Sci. USA.* 100:14822–14827.

69. Tessmer, I., Y. Yang, …, D. A. Erie. 2008. Mechanism of MutS searching for DNA mismatches and signaling repair. *J. Biol. Chem.* 283:36646–36654.

70. Wolf, A., and K. N. Kirschner. 2013. Principal component and clustering analysis on molecular dynamics data of the ribosomal L11·23S subdomain. *J. Mol. Model.* 19:539–549.

71. Pal, A., H. M. Greenblatt, and Y. Levy. 2020. Prerecognition diffusion mechanism of human DNA mismatch repair proteins along DNA: Msh2-Msh3 versus Msh2-Msh6. *Biochemistry.* 59:4822–4832.

72. Plazzer, J.-P., R. H. Sijmons, …, F. Macrae. 2013. The InSiGHT database: utilizing 100 years of insights into Lynch syndrome. *Fam. Cancer.* 12:175–180.

73. Bouchal, T., I. Durník, …, P. Kulhánek. 2020. Importance of base-pair opening for mismatch recognition. *Nucleic Acids Res.* 48:11322–11334.

74. Mazur, D. J., M. L. Mendillo, and R. D. Kolodner. 2006. Inhibition of Msh6 ATPase activity by mispaired DNA induces a Msh2 (ATP)-Msh6 (ATP) state capable of hydrolysis-independent movement along DNA. *Mol. Cell.* 22:39–49.

75. Hao, P., S. J. LeBlanc, …, K. R. Weninger. 2020. Recurrent mismatch binding by MutS mobile clamps on DNA localizes repair complexes nearby. *Proc. Natl. Acad. Sci. USA.* 117:17775–17784.

76. Lakhani, B., K. M. Thayer, …, D. L. Beveridge. 2017. Evolutionary covariance combined with molecular dynamics predicts a framework for allostery in the MutS DNA mismatch repair protein. *J. Phys. Chem. B.* 121:2049–2061.

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.
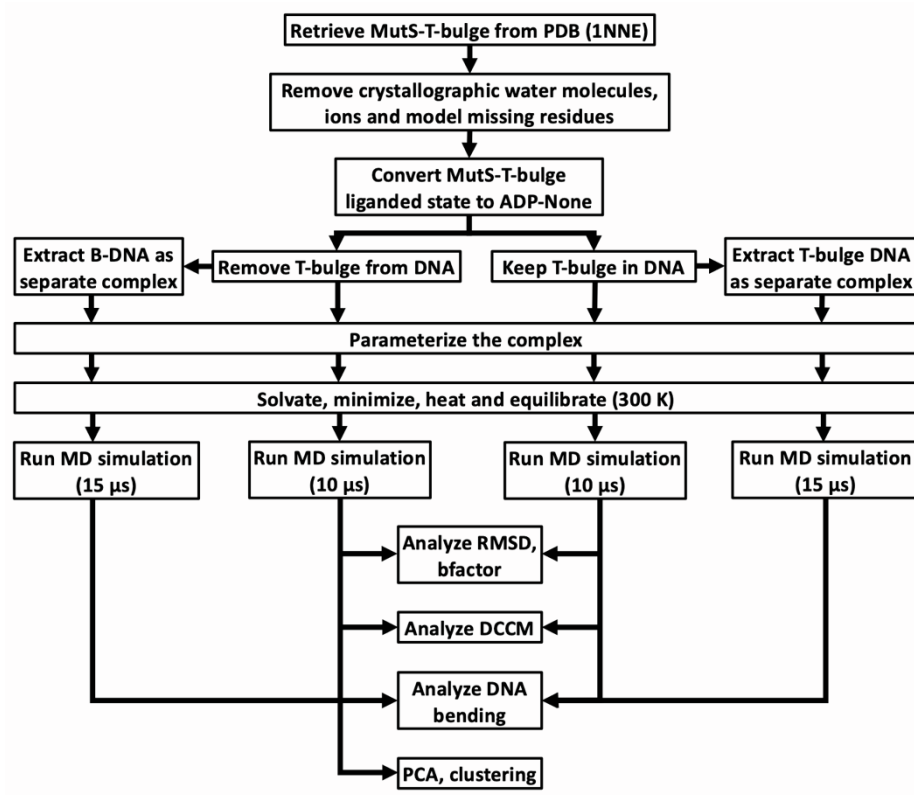
Figure S1

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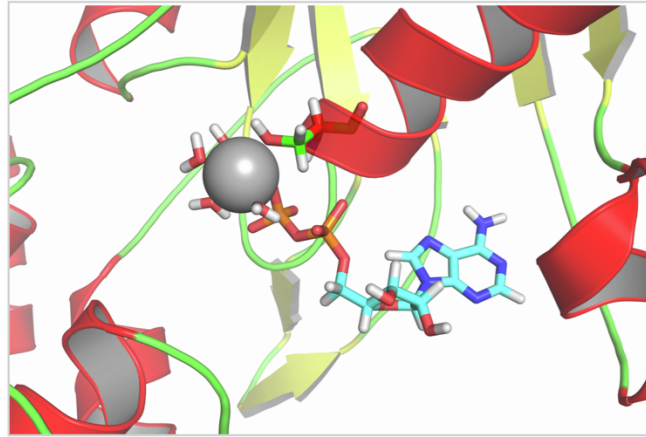
Figure S2

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).
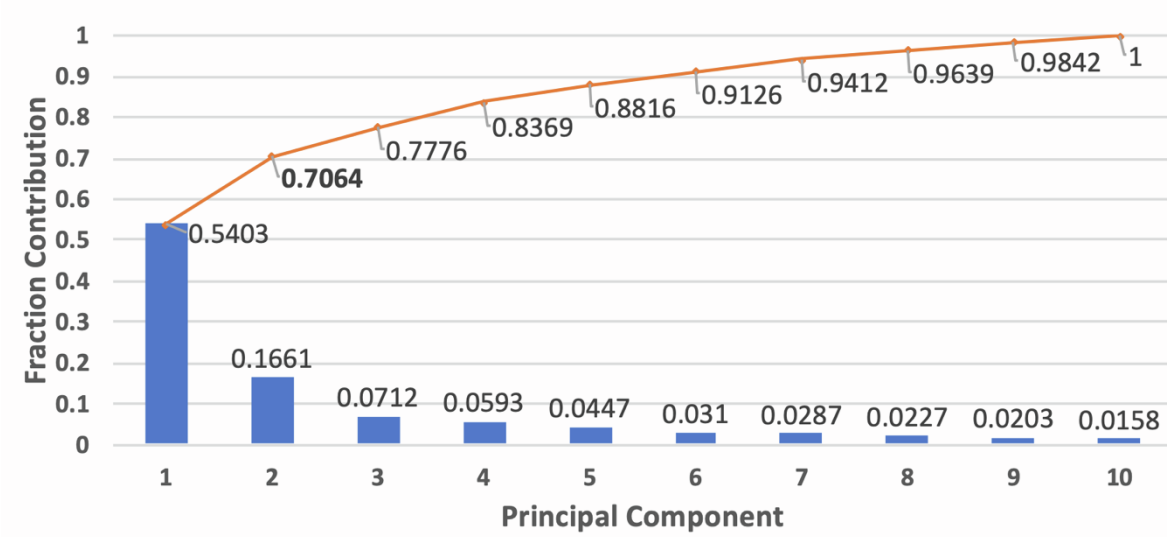
Figure S4

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.
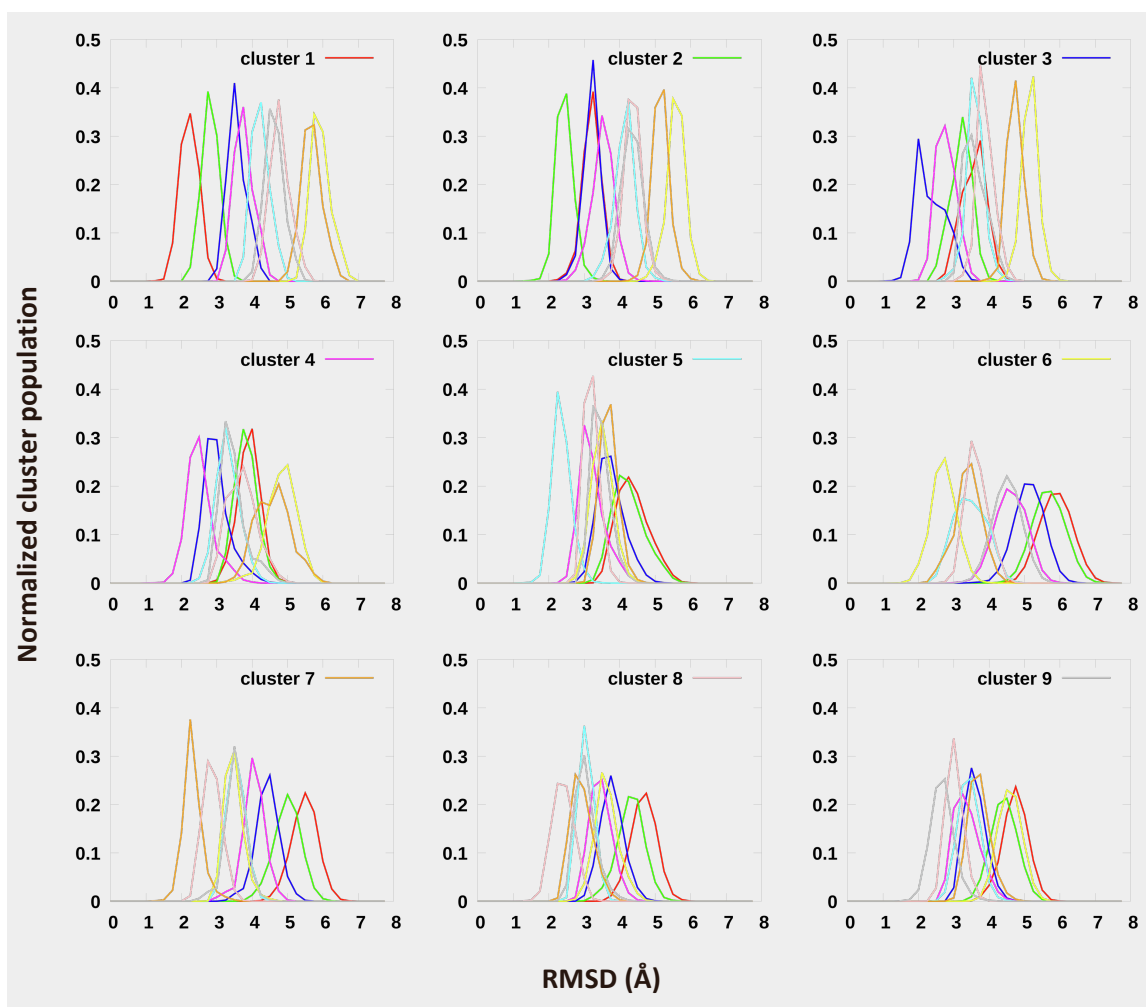
Figure S5

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.
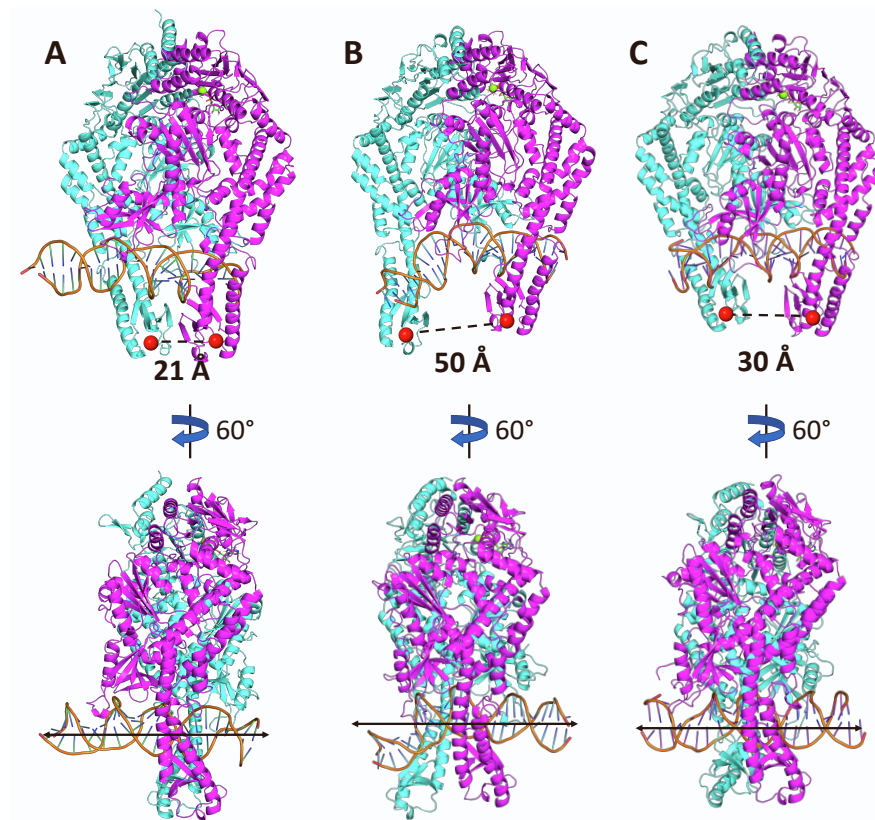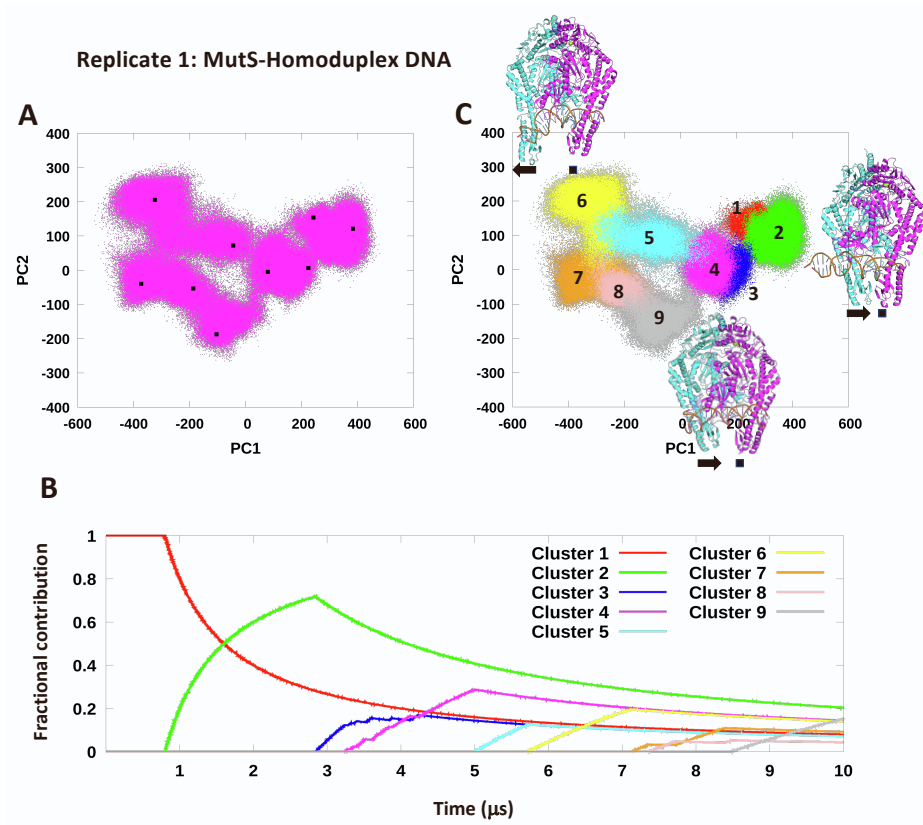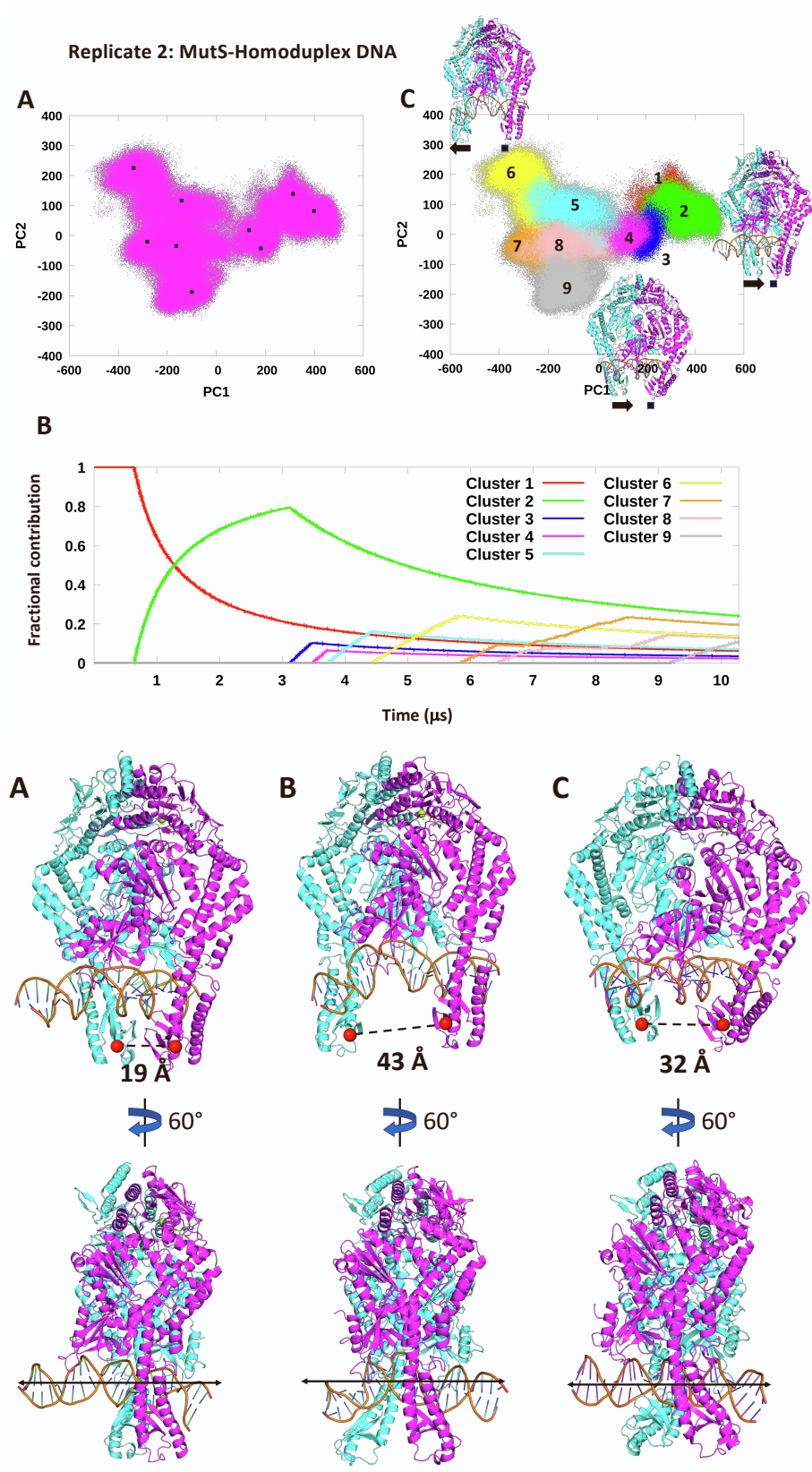
Figure S6

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).
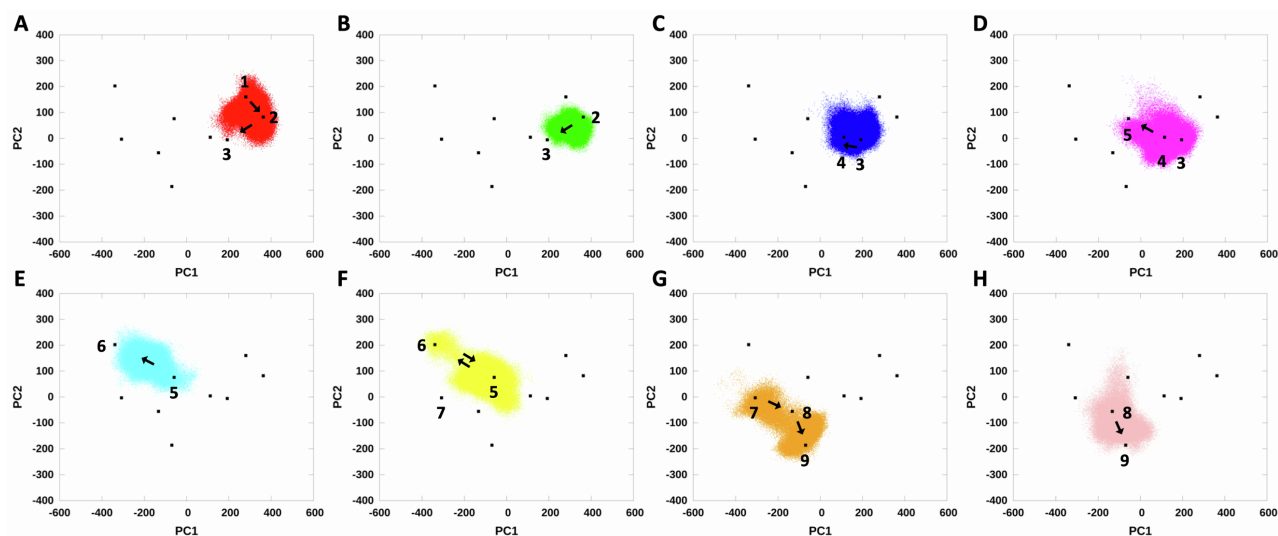
Figure S8

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

MutS-T-bulge
MSH2-MSH6-T-bulge

90°

A

ASP 38$_I^A$ LYS 431^{MSH6}

LYS 64$_I^A$ LYS 453^{MSH6}

ARG 76$_I^A$ ARG 468^{MSH6}

LYS 1000^{MSH6}
ARG 473$_{IV}^A$ LYS 1001^{MSH6}

ARG 475$_{IV}^A$ LYS 1004^{MSH6}

B

LYS 64$_I^B$ LYS 73^{MSH2}

ARG 110$_I^B$ LYS 113^{MSH2}

ASN 443$_{IV}^B$ GLN 518^{MSH2}

LYS 471$_{IV}^B$ LYS 546^{MSH2}

GLN 545^{MSH2}
ARG 473$_{IV}^B$ ASN 547^{MSH2}
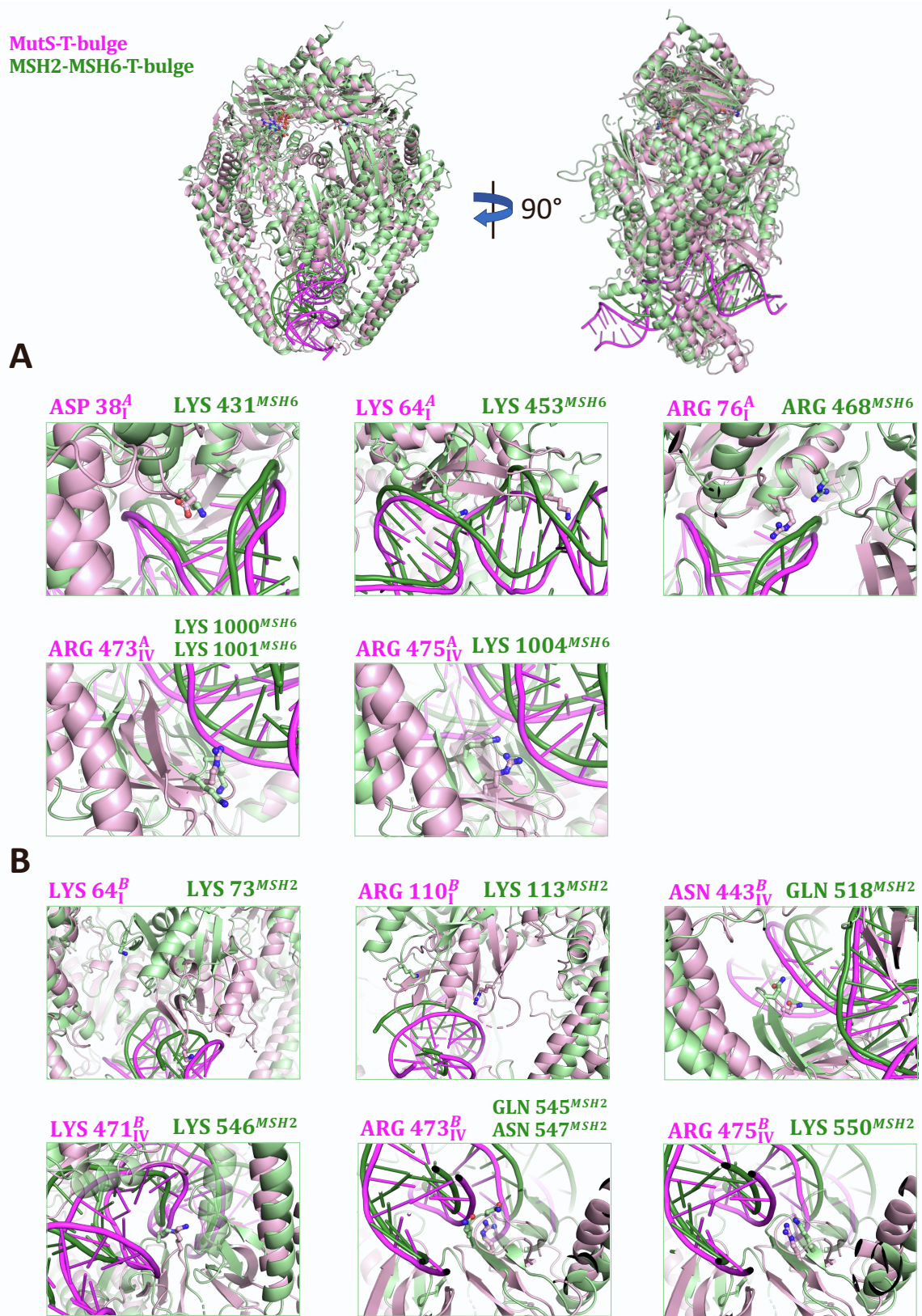
ARG 475$_{IV}^B$ LYS 550^{MSH2}

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.