Supporting Information

Dynamical Allosterism in the Mechanism of Action of DNA Mismatch Repair Protein MutS

Susan N. Pieniazek[†*], Manju M. Hingorani[‡], D. L. Beveridge[†]

[†]Department of Chemistry and [‡]Molecular Biology and Biochemistry Department, Wesleyan University, Middletown, Connecticut

Supporting Materials and Methods

S1. MD Simulations

MD trajectories for MutS•DNA(+T)•ATP and MutS•DNA(+T) complexes were obtained for 150 ns and repeated from a different configuration (described in 'model system preparation', below). As a starting point for the repeat simulations, we used snapshots at time 120 ns and 35 ns from the original MutS•DNA(+T)•ATP and MutS•DNA(+T) simulations, respectively. The solvent molecules and Na+ and Cl- ions were stripped and from each snapshot, solvent boxes rebuilt (to match the size of the original as close as possible), and ions were added. Minimization, heating, and equilibration procedures for these repeated matched those followed in the original simulations (described below). We also carried out a relatively short simulation that included a Mg^{2+} cation coordinated with the ATP in the MutS•DNA(+T)•ATP complex.

Table S1. MD Simulations (1-3) carried out in this study.

MD	Model	Time (ns)
1A	**MutS•DNA(+T)•ATP–** *original*	150.0
1B	**MutS•DNA(+T)•ATP–** *duplicate*	100.0
2A	**MutS•DNA(+T)–** *original*	150.0
2B	**MutS•DNA(+T)–** *duplicate*	100.0
3A	**MutS•DNA(+T)•ATP•Mg^{2+}**	40.0

S2. MD Model System Preparation

MutS•DNA(+T)•BeF_3^- PDB Coordinates. The crystal structure of *T. aquaticus* MutS (PDB code 1nne) serves as the reference structure for all simulations. The MutS coordinate file as utilized for MD consists of 1530 amino acids, 11 of which are not resolved. The two sets of missing loop coordinates are 629-634 (S1, domain V) and 1101-1106 (S2, domain IV). The full model MutS complex was generated from the crystal structure using the Leap module of AMBER 9.0, molecular modeling software Pymol [Delano Scientific, 2006], and manual editing. The backbone and side chains atoms of the ATPase domains in S1 overlap well with S2. In the case of 629-634 coordinates, information about the conformation of its backbone was available since the coordinates for the analogous loop in S2 (residues 1629-1634) are present in the crystal structure. The procedure for building both sets missing coordinate regions was the similar. We first generated a copy of the coordinates in present in opposite subunit. This copy included the coordinates the first set of five residues, plus an additional five residues on each side, thus providing a total of 18 residues. The last three residues were then superimposed and fitted on each side of the anchor residues. Finally, the new loop coordinates were copied into the original crystal structure file.

MD 1A-B

MutS•DNA(+T)•ATP Complex. Initial placement of ATP molecules in the complex was modeled from the coordinates of the crystallographic ligands ADP•BeF_3^- (one per nucleotide binding site). The ADP·BeF_3^-coordinates present in the crystal structure were converted to ATP by replacing the BeF_3^- group with a γ-phosphate ($-PO_3^{2-}$) group. Minimization, heating and heating, and equilibration procedures are described in the main text.

MD 2A-B

MutS•DNA(+T) Complex. The ATP-free complex was generated by removing the ATP coordinates from the complex. Both complexes were then minimized and equilibrated.

MD 3

MutS•DNA(+T)•ATP•Mg^{2+}Complex. The initial structure was based on that from *Taq* MutS 1NNE used for the MutS•DNA(+T)•ATP complex simulation. A best fit of the nucleotide binding site residues in the aforementioned structure and the nucleotide binding site of an available crystal structure (PDB ID: 1w7a) of *E.coli* MutS in complex with ATP and Mg^{2+} provided the relative orientation of an ATP• Mg^{2+}•H_2O complex in the site. ATP molecules in the MutS•DNA(+T)•ATP structure were then replaced by ATP• Mg^{2+}•H_2O. D662 was assigned a standard protonation sate. 1w7a reference: Lamers et al. J. *Biol. Chem.* **2004**, *279*, 43879.

S3. MutS Structure

S.3.1. ATPase Domain V

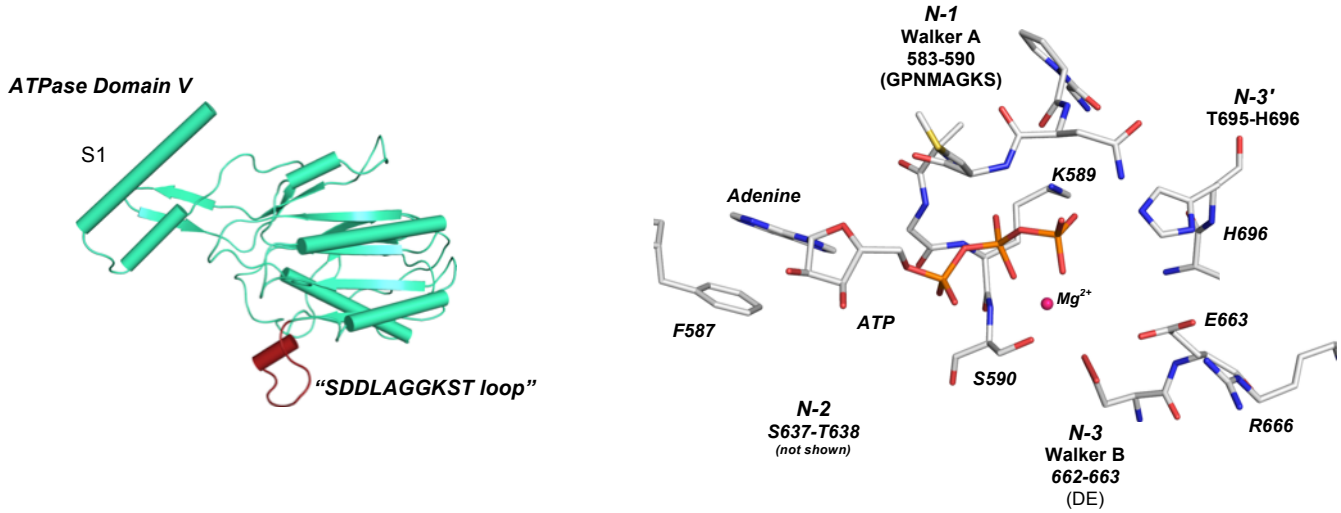


FIGURE S1. (*A*) The "SDDLAGGKST loop" (colored in red) is present in MutS ATPase domains V (subunits S1 and S1). Subunit S2 in omitted in this graphic for clarity. (*B*) MutS ABC (ATP-Binding Cassette) nucleotide-binding site occupied by a ATP• Mg^{2+}•H_2O complex. Water molecules surrounding the Mg^{2+} are omitted to enhance clarity. Highly conserved motifs and

other key residues are labeled according to residue number and name where applicable. The conserved motifs are: N-1 (Walker A motif: 583-590) ATP phosphate binding loop; N-3 (Walker B motif: Glu-Asp 662-663); N-2 (Ser-Thr 637-638); and N-3' (Thr-His 695-696).

S.3.2. DNA-binding Region (I and IV)

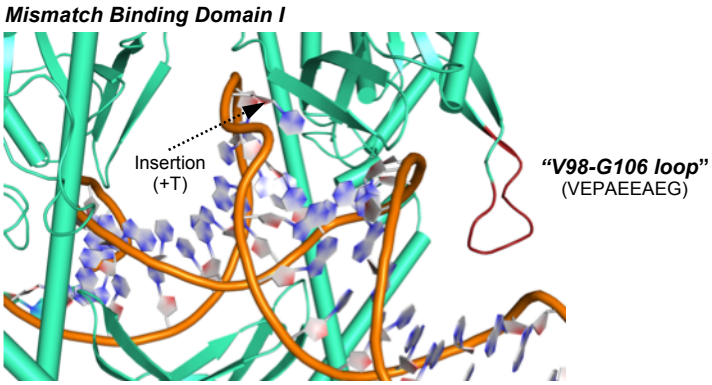


FIGURE S2. The "V98-G106 loop" (colored in red) is present in MutS Mismatch Binding domains (I). The full sequence of this loop is VEPAEEAEG.

S4. MD Sampling and Convergence

S4.1. Sampling of Configurational Space: PCA of Covariance Matrices and 2D-RMSD Analyses

The MutS simulations were carried out for up to 150 ns, We carried out principal component analysis (PCA) of the covariance of atomic fluctuations and two-dimensional route-mean-square deviation analysis (2D-RMSD) to evaluate the configurational space sampled by each trajectory. These analyses yield complimentary information regarding local ensembles of conformations (i.e., substates) and dynamical ranges of snapshots within a single simulation, as well as between snapshots from different simulations.

PCA was carried out by fitting the protein and DNA backbone atoms of each configuration to the average structure of the full-length trajectory. A covariance matrix—or, a 3*N*x3*N* matrix, where *N* is the number of protein and DNA atoms in each molecule—was calculated from the new fitted trajectory using the ptraj module of AmberTools 1.2. The elements of the matrix consist of covariances of atomic displacements relative to their average positions. We diagonalized the matrix to obtain a set of eigenvectors and eigenvalues, which represent the principal components and variance along each principal component, respectively. Each trajectory was projected onto its respective principal coordinate; from these projections, two-dimensional plots along the first two principal components were generated to represent the sampled distribution and corresponding populations in configurational space. We assessed these plots (Figure S3) and designated a "final substate" as the final and most populated block of trajectory snapshots for each simulation. The final substates of the ATP-bound and ATP-free simulations are 65-150 ns and 105-150 ns, respectively.

We calculated a matrix of pair-wise fit RMSD (2D-RMSD) values of every structure with every other in the equilibrated portion (i.e the final substate) of the ATP-bound (65-150 ns) and ATP-free (105-150 ns) simulations. These plots are provided in Fig. S4 and show snapshot number on the x and y axes, with each nanosecond comprised of 50 snapshots. Thus, each full-length 150 ns simulation consists of 1500 snapshots 2D-RMSD plot. Here we refine our substate definition as an off-diagonal grouping of structures within ~3 Å RMSD, as well as a members of the final substate established by the PCA.

Our final step was to combine final substate of ATP-bound (65-150 ns) and ATP-free (105-150 ns) trajectory snapshots (*vide supra*) to determine the RMSD differences between the two simulations as a function of time, or the "dynamical range". These plots are shown in Figure S5.

S4.2 Temporal Convergence and **Reproducibility**

To assess convergence, we calculated the RMSD with respect to the average structure for 10 ns blocks of the equilibrated portion of ATP-bound and ATP-free simulations. The RMSD plots and corresponding superposition of average structures from the 10 blocks are shown in Fig. S6 –S7, which show them be in substantial agreement – all are within an rmsd of ~1.5 Å. We also analyzed the corresponding 10 ns windows of correlated atomic fluctuations (Fig. S8 – S9).

As noted above, MD trajectories for MutS•DNA(+T)•ATP and MutS•DNA(+T) complexes were obtained for 150 ns and repeated from a different configuration As a starting point for the repeat simulations, we used snapshots at time 120 ns and 35 ns from the original MutS•DNA(+T)•ATP and MutS•DNA(+T) simulations, respectively. The solvent molecules and Na+ and Cl- ions were stripped and from each snapshot, solvent boxes rebuilt (to match the size of the original as close as possible), and ions were added. Minimization, heating, and equilibration procedures for these repeated matched those followed in the original simulations. RMSD plots and corresponding superposition of structures are shown in Fig. S10 – S11.

We reproduced the original and duplicate ATP-free and ATP-bound MD calculated average structures (ATPase and DNA-binding region domains) (Fig. S10–S11). The protein backbone RMSD for original versus duplicate simulations are shown in parentheses. The overall RMSDs are all relatively small (~1-2 Å), with the highest RMSD in the DNA-binding region (S2)

S5. Atomic Fluctuations Analyses

S5.1. *B*-factors

Atomic positional fluctuations were computed for all residues with the ptraj module of AmberTools 1.2. *B*-factors, as reported in the text, are from multiplication of the squared atomic positional fluctuations by $(8/3)\pi^2$. RMS fittings of the protein backbone were carried out to remove the rotation and translation for each simulation prior to calculating the fluctuations.

We report relative differences in the calculated *B*-factors between ATP-bound and ATP-free simulations. *B*-factors represent the degree of motion in a crystal structure, but a direct

comparison of absolute crystallographic *B*-factor values and MD calculated *B*-factor values could be problematic—crystal packing effects, force fields, and trajectory sampling length are known to cause discrepancies.

S5.2 Dynamical Cross-Correlation Matrices (DCCMs)

Matrices with pair-wise correlations of atomic fluctuations were generated in the ptraj module of AmberTools 1.2. The pair-wise correlations of atomic fluctuations are computed using backbone atoms (N, C_a, C in amino acids, and P, O5', C5', C4', C3', O3' in nucleotides).

The elements of the atomic cross-correlation matrix **C** are defined by:

$$C(i,j) = \frac{\left\langle \Delta r_i \bullet \Delta r_j \right\rangle}{\left(\left\langle \Delta r_i \right\rangle^2 \left\langle \Delta r_j \right\rangle^2 \right)^{\frac{1}{2}}}$$

where $\Delta \boldsymbol{r}_i$ and $\Delta \boldsymbol{r}_i$ are displacement vectors for atoms i and j, respectively. Positive correlated motions are denoted $C(i,j) = 1$ and anti-correlated motions $C(i,j) = -1$. The reported C_{ij} values of the pair-wise correlations of atomic fluctuations averaged by residue, i.e. the average of $N_r(i)$x$N_r(j)$ values involving all pair-wise combinations of the $N_r(i)$ atoms in residue (i) with the $N_r(j)$ atom in residue (j).

S5.3. Clusters of Mutually Correlated Residues Theory

We carry our principal component analysis on the NxN matrix of correlated atomic fluctuations, where N is the average displacement of all atoms per residues. Principal component analysis of correlated atomic fluctuations is similar to PCA in Cartesian coordinate space. The correlation matrix elements, $C(i,j)$, are normalized covariances in the space of atomic fluctuations, $\{D\boldsymbol{r}_i\}$, as opposed to covariances in the Cartesian space, $\{x_i,y_i,z_i\}$.

The matrix of correlated atomic fluctuations is diagonalized by using:

$$\Lambda = \mathbf{T}^{\mathrm{T}}\mathbf{C}\mathbf{T}$$

where the columns of **T** and diagonal elements of Λ are eigenvectors and eigenvalues of **C**, respectively. Each eigenvector represents a cluster of mutually correlated atomic motions and its eigenvalue represents the weight of that cluster

Thus, each eigenvector, v_j, of the correlation matrix can be written as a linear combination of atomic fluctuations:

$$v_i = \sum_{j=1}^{N} \chi_{ij} \Delta r_j$$

Since the eigenvectors (v_j) are normalized, a linear combination in which all the motions contribute equally have all coefficients with an absolute value, $|c_{ij}|$, of $(1/N)^{1/2}$. In our analysis of the correlation matrix eigenvectors, we focused on contributions from motions ($D\boldsymbol{r}_j$) to eigenvector i that are $(1/N)^{1/2} \leq |c_{ij}| < 2*(1/N)^{1/2}$ and $|c_{ij}| \geq 2*(1/N)^{1/2}$. Finally, the overlaps of eigenvectors from ATP-bound and ATP-free simulations were computed to allow for a direct comparison. High eigenvector overlap indicates high similarity between two eigenvectors.

Results: Sampling of Configurational Space: PCA of Covariance Matrices and 2D-RMSD Analyses

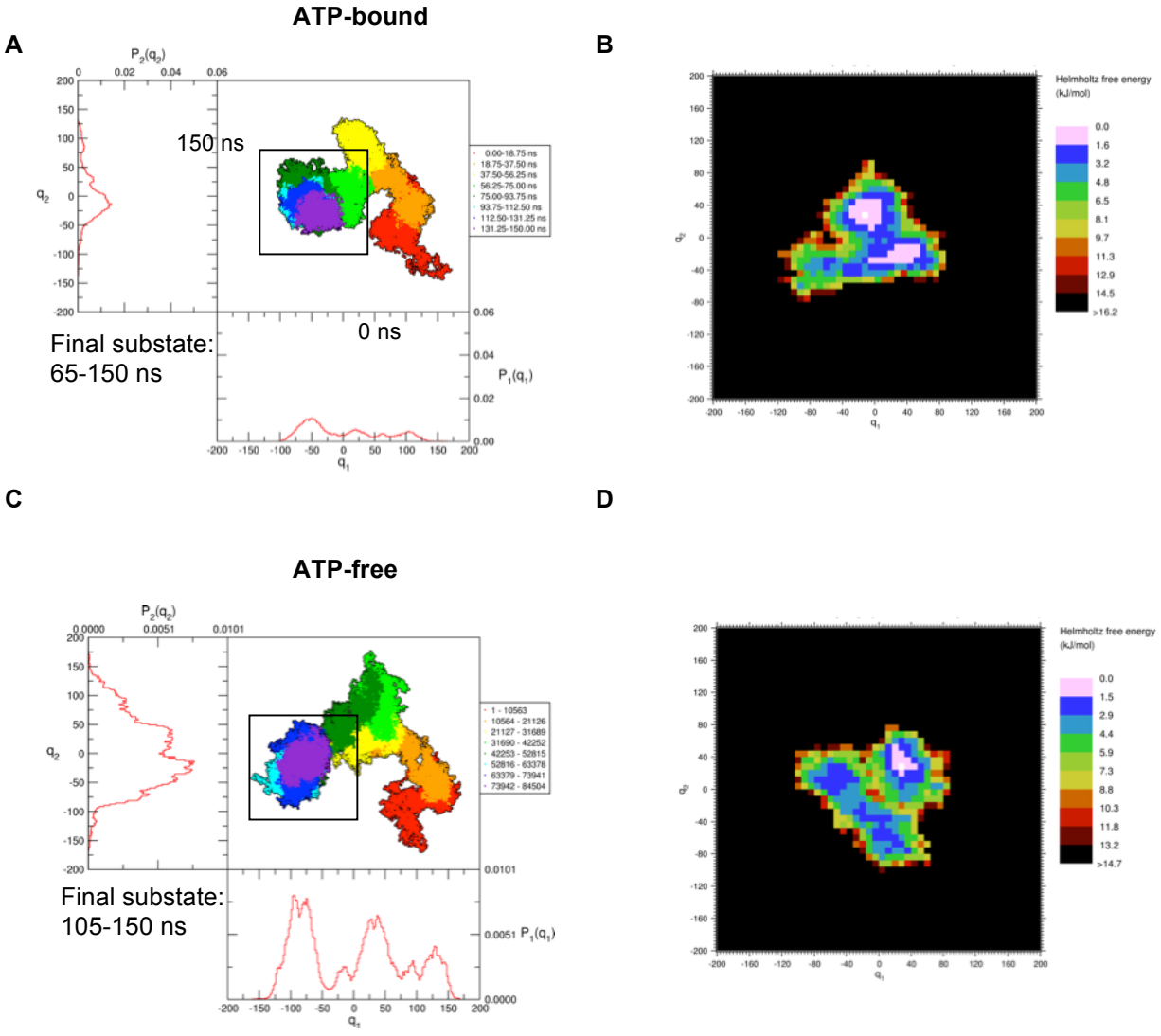


FIGURE S3. Principal component analysis of the covariance of atomic fluctuations for MD on MutS•DNA(+T)•ATP and MutS•DNA(+T): *(A* and *C)* Principal components $q_1(t)$ vs $q_2(t)$; (*B* and D) Free energy vs. PC_1 vs. PC_2 for the equilibrated portion (final substate) of the trajectories. We have analyzed the trajectory snapshots in the equilibrated portion of the simulations (black square: ATP-bound 65-150 ns; ATP-free 105-150 ns).

Below 2D-RMSD plots displayed as a function of snapshot number (1 nanosecond contains 50 snapshots). Relatively low RMSD values are displayed in cool colors (blue and green) and high with warm (yellow and red).

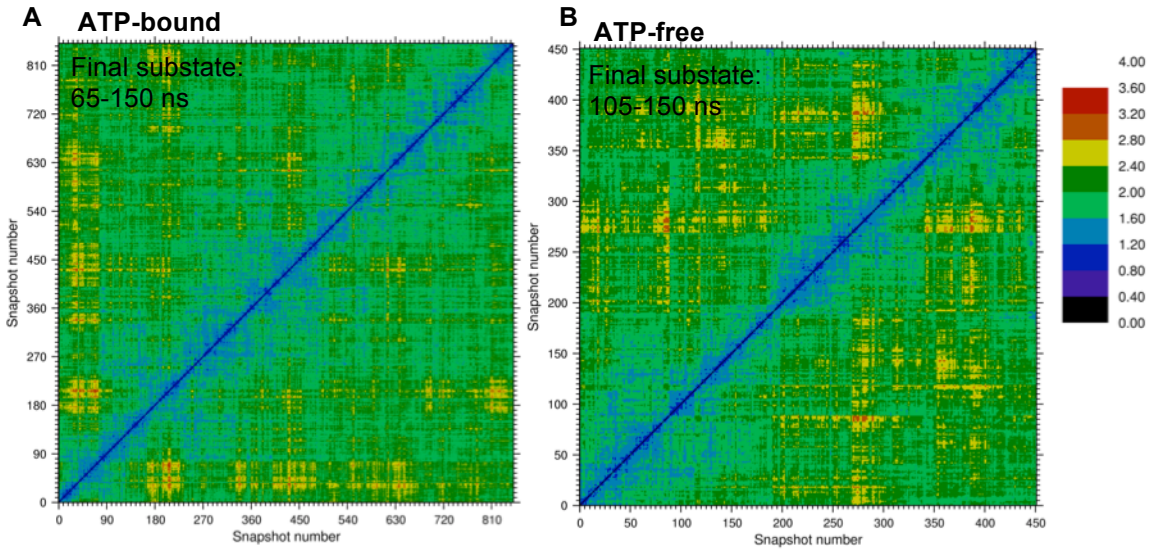


FIGURE S4. 2D-RMSD maps for the equilibrated portion (final substate) of (*A*) MutS•DNA(+T)•ATP and (*B*) MutS•DNA(+T) simulations. Low RMSD values are displayed in cool colors (blue and green) and high with warm (yellow and red).

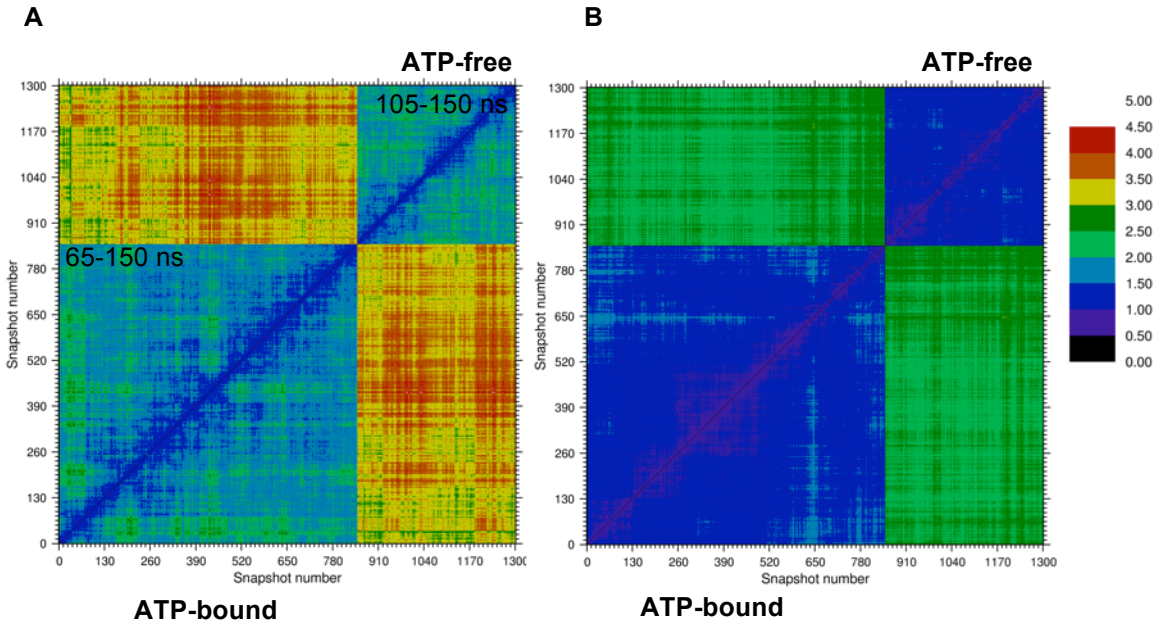
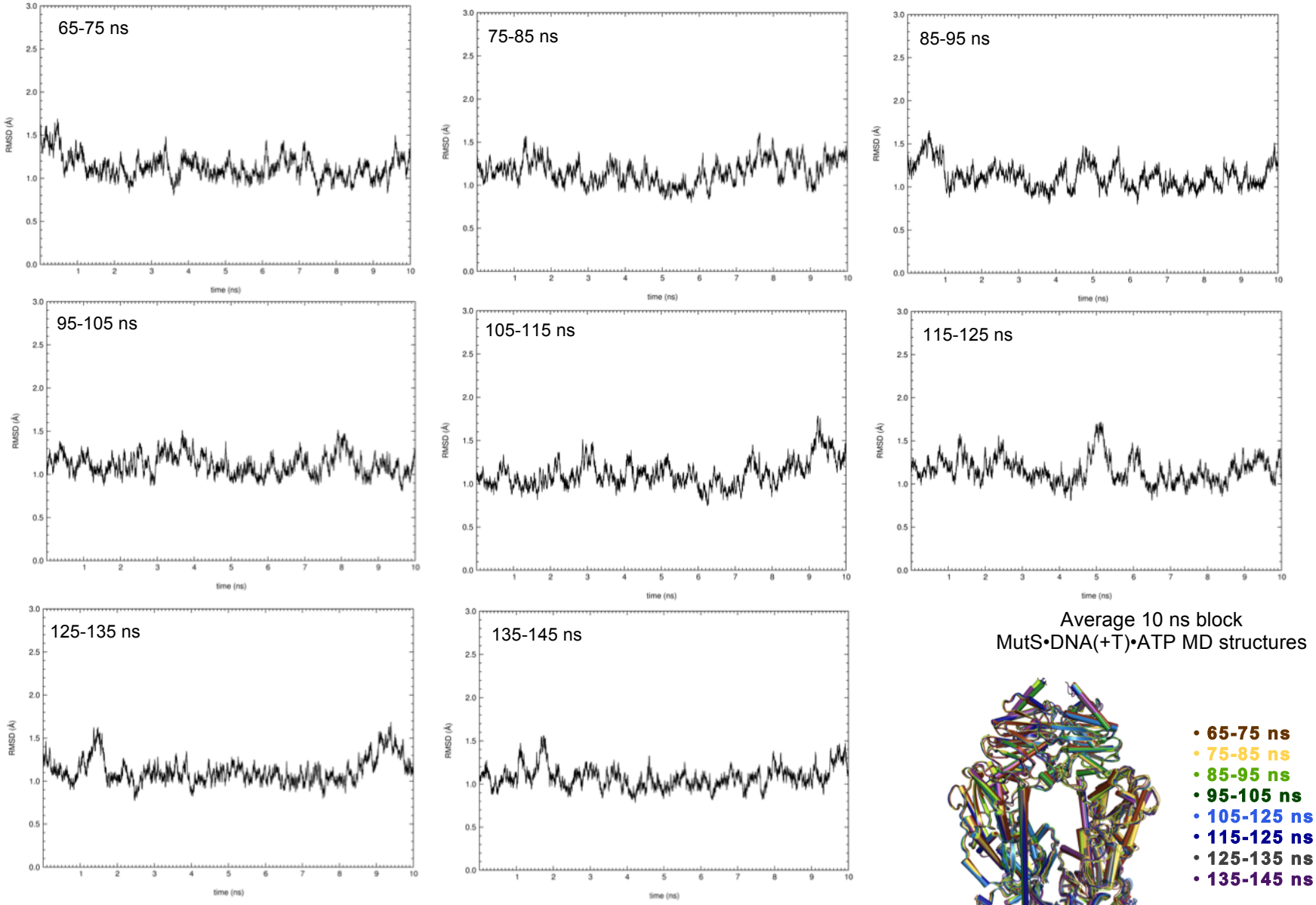


FIGURE S5. (*A*) Combined 2D-RMSD maps for MutS•DNA(+T)•ATP and MutS•DNA(+T) protein backbone atoms and (*B*) MutS•DNA(+T)•ATP and MutS•DNA(+T) ATPase (domain V) atoms (85 ns in the ATP-bound, 45 ns in the ATP-free). The maps show how the RMSD of the two equilibrated portions of the simulations vary with time (the length of the equilibrated position of each simulation is different, hence the blocks are different sizes. The RMSD of the two simulations is ~2.0-4.5 Å in the case of (*A*), and 2-3 Å in the case of (*B*). Each nanosecond contains 50 snapshots.

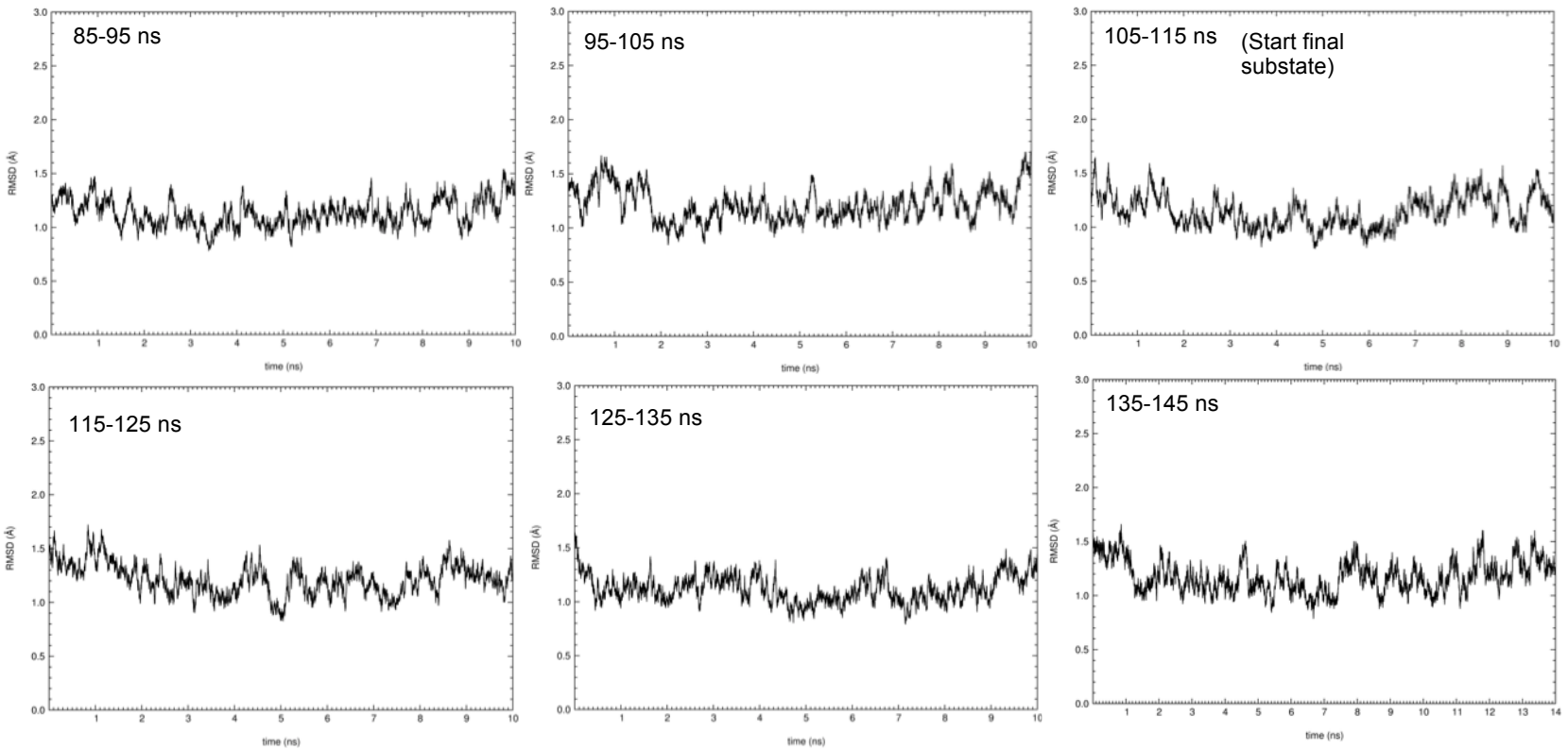
Results: **Temporal Convergence** and **Reproducibility**



Average 10 ns block
MutS•DNA(+T)•ATP MD structures

- 65-75 ns
- 75-85 ns
- 85-95 ns
- 95-105 ns
- 105-125 ns
- 115-125 ns
- 125-135 ns
- 135-145 ns

FIGURE S6. MutS•DNA(+T)•ATP protein (N, C_a, C) and DNA (P, O5', C5', C4', C3', O3') backbone RMSD of with respect to the MD average structures from each 10 ns block.



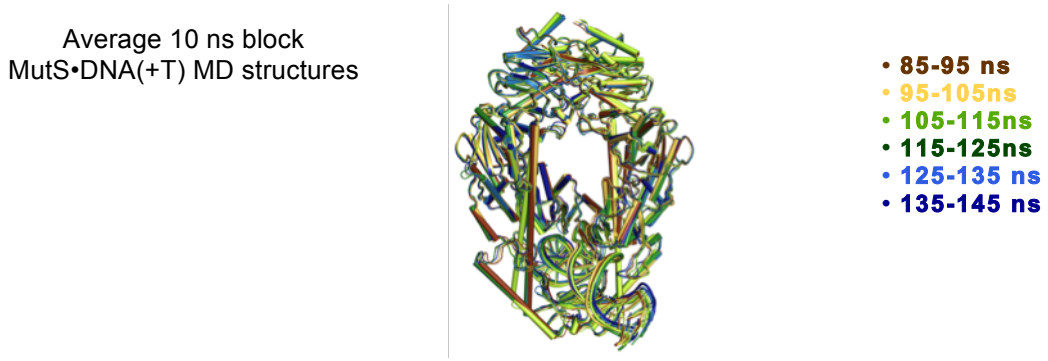


FIGURE S7. MutS•DNA(+T) protein (N, C_a, C) and DNA (P, O5', C5', C4', C3', O3') backbone RMSD with respect to the average MD structures from each 10 ns block.

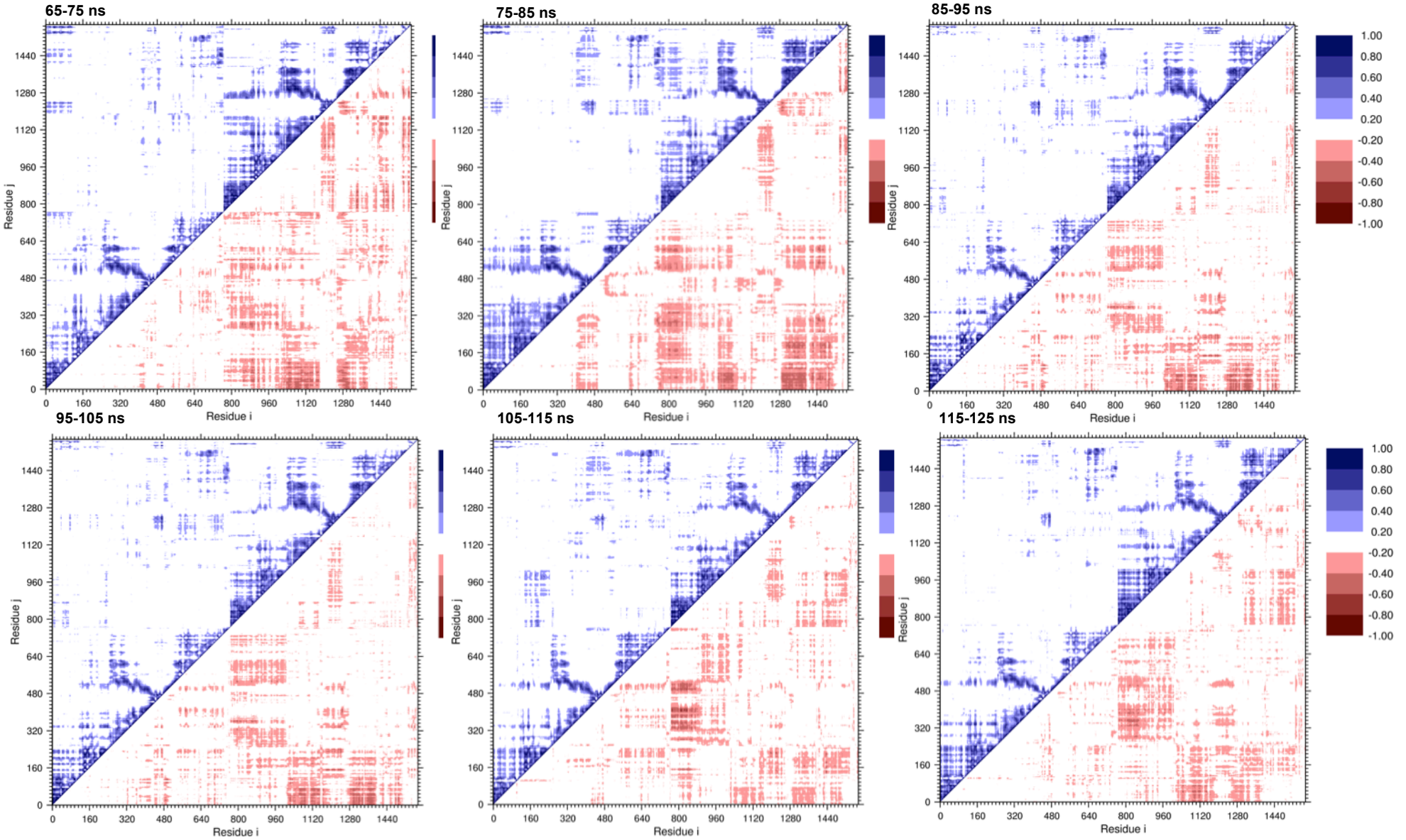


FIGURE S8. DCCMS of 10 ns blocks of ATP-bound trajectory snapshots from the equilibrated portion of the simulation. Protein and DNA backbone RMSD fitting of the protein and DNA backbones was carried out with respect to the MD average structure from each 10 ns block.

Temporal Convergence of Correlated Atomic Fluctuations

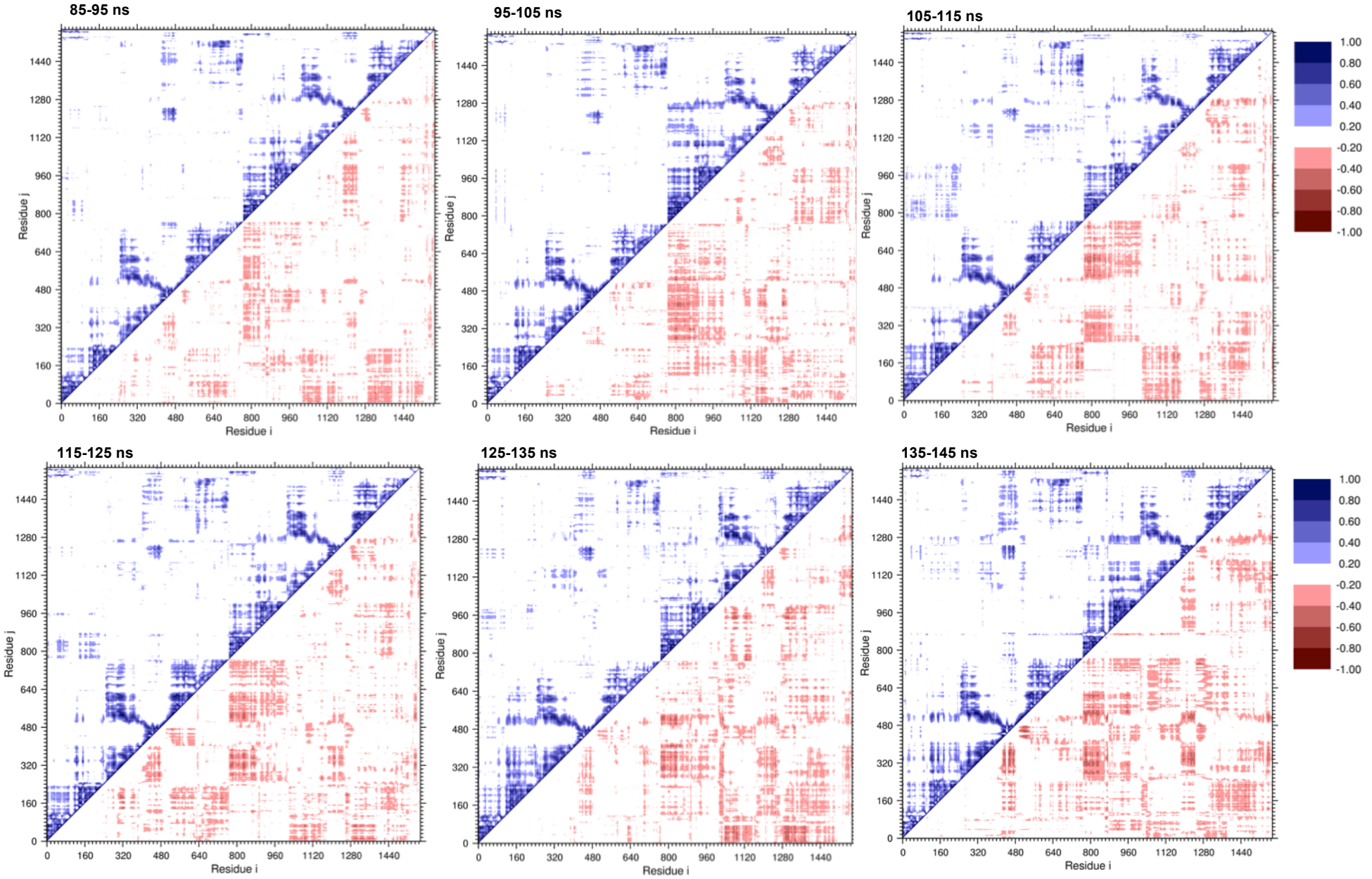
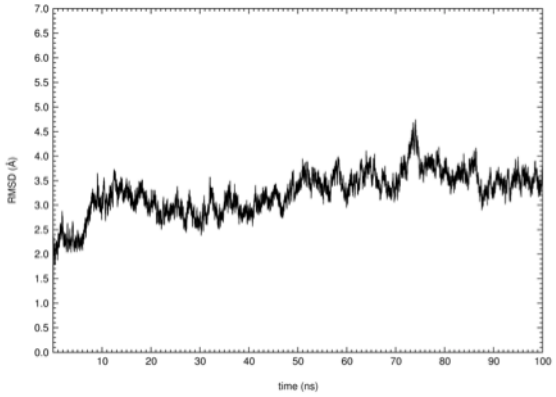


FIGURE S9. DCCMS of 10 ns blocks of ATP-free trajectory snapshots from the equilibrated portion of the simulation. Protein and DNA backbone RMSD fitting of the protein and DNA backbones was carried out with respect to the MD average structure from each 10 ns block.

A. MutS•DNA(+T)•ATP duplicate simulation (~100 ns); 1D-RMSD of the protein and DNA backbone atoms with respect to the first snapshot.



B. Overlay of Original and Duplicate MutS•DNA(+T)•ATP Average Structures

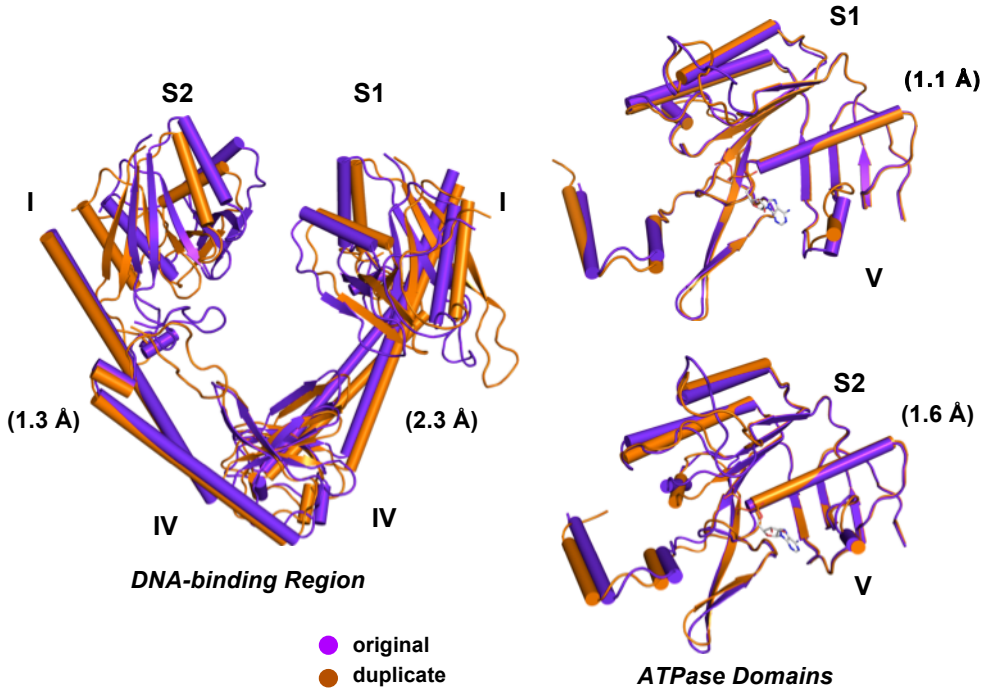
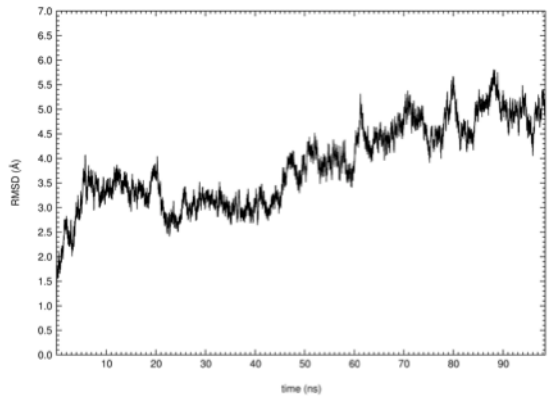


FIGURE S10. (*A*) RMSD of protein (N, C_a, C) and DNA (P, O5', C5', C4', C3', O3') backbone atoms with respect to the first simulation snapshot of the ~100 ns duplicate MutS•DNA(+T)•ATP simulation (~3.0 Å). (*B*) Duplicate MD calculated average DNA binding and ATP-structures (orange) aligned with the protein backbone atoms of the original MD average structures (purple). Average structures calculated from the last 50 ns (final substate). RMSD values are given in parentheses (~1.1 to 2.3 Å).

A. MutS•DNA(+T) duplicate simulation (~100 ns); 1D-RMSD of protein and DNA backbone atoms with respect to the first snapshot.



B. Overlays of Original and Duplicate MutS•DNA(+T) Average Structures

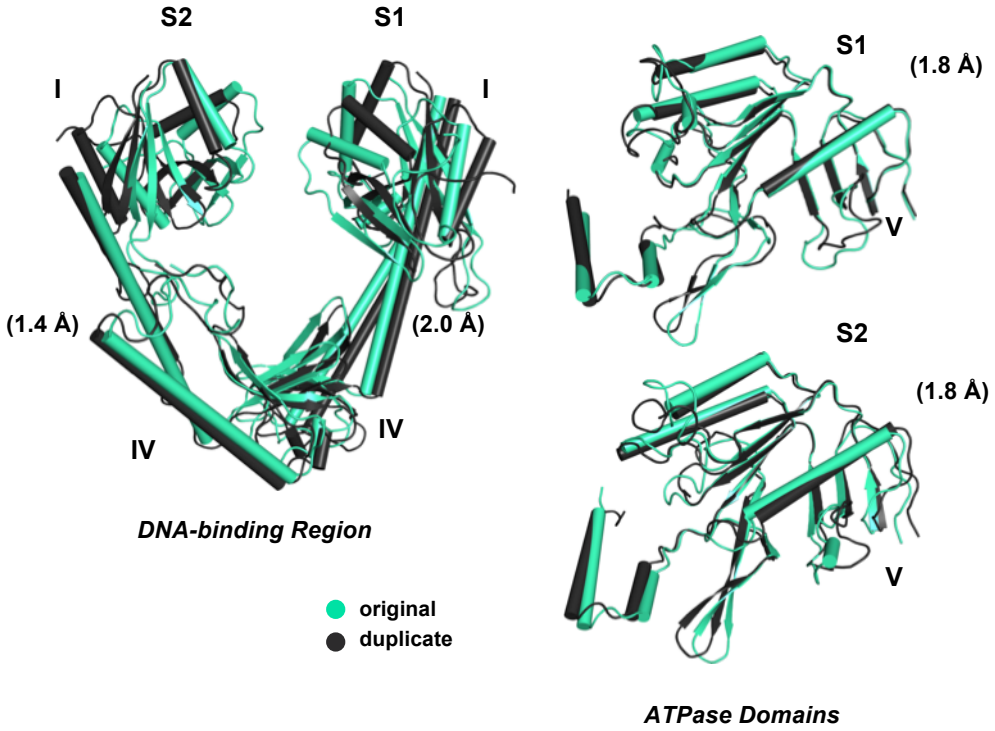
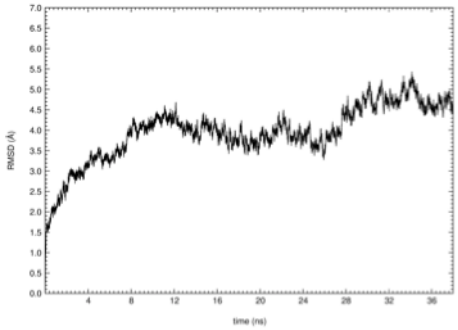


FIGURE S11. (*A*) RMSD of protein (N, C_a, C) and DNA (P, O5', C5', C4', C3', O3') backbone atoms with respect to the first simulation snapshot of the ~100 ns duplicate MutS•DNA(+T) simulation (~4.5 Å). (*B*) Duplicate MD calculated average DNA binding and ATPase structures (black) aligned with the protein backbone atoms of the original MD average structures (green). Average structures calculated from the last 25 ns (final substate). RMSD values are given in parentheses (~1.4 to 2.0 Å).

A. MutS•DNA(+T)-Mg-ATP duplicate simulation (40 ns); 1D-RMSD of protein and DNA backbone atoms with respect to the first snapshot.



B. Overlays of MutS•DNA(+T)-Mg-ATP and MutS•DNA(+T)-ATP Average Structures

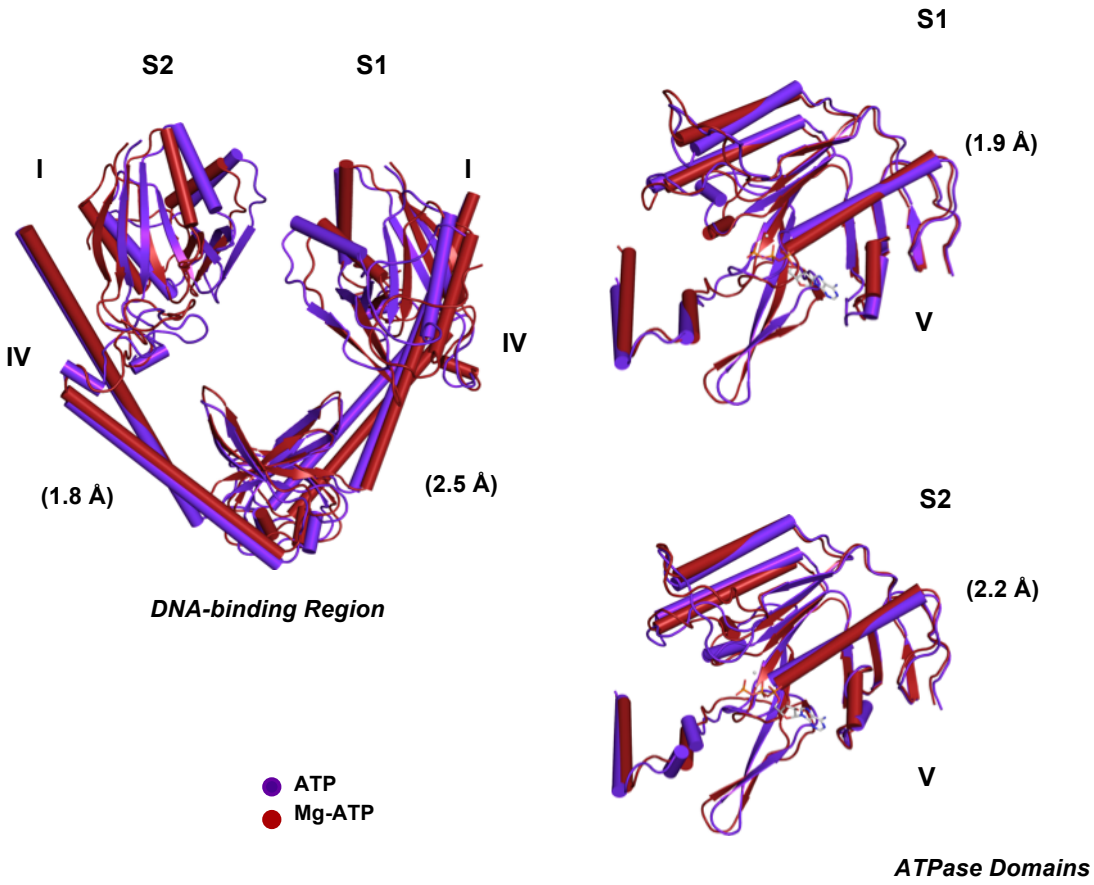


FIGURE S12. (*A*) RMSD of MutS•DNA(+T)•Mg-ATP and MutS•DNA(+T)•Mg-ATP protein (N, C_a, C) and DNA (P, O5', C5', C4', C3', O3') backbone atoms with respect to the first simulation snapshot . (*B*) MD calculated average Mg-ATP-bound (red) and ATP-bound (purple). Average structures calculated from the last 15 ns (final substate). RMSD values are given in parentheses (~1.8 to 2.5 Å). DNA binding (I and V) and ATPase structures (V) aligned with the protein backbone atoms.

Results: Atomic Fluctuations

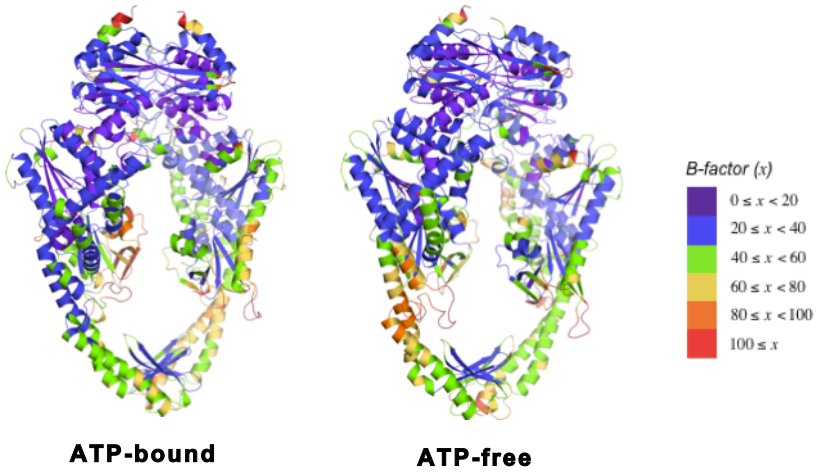


FIGURE S13. Structure-based representations of calculated B factors color-coded from blue to red in the order of increasing values. B factors are squared atomic positional fluctuations multiplied the by $(8/3)\pi^2$. The DNA is not shown.

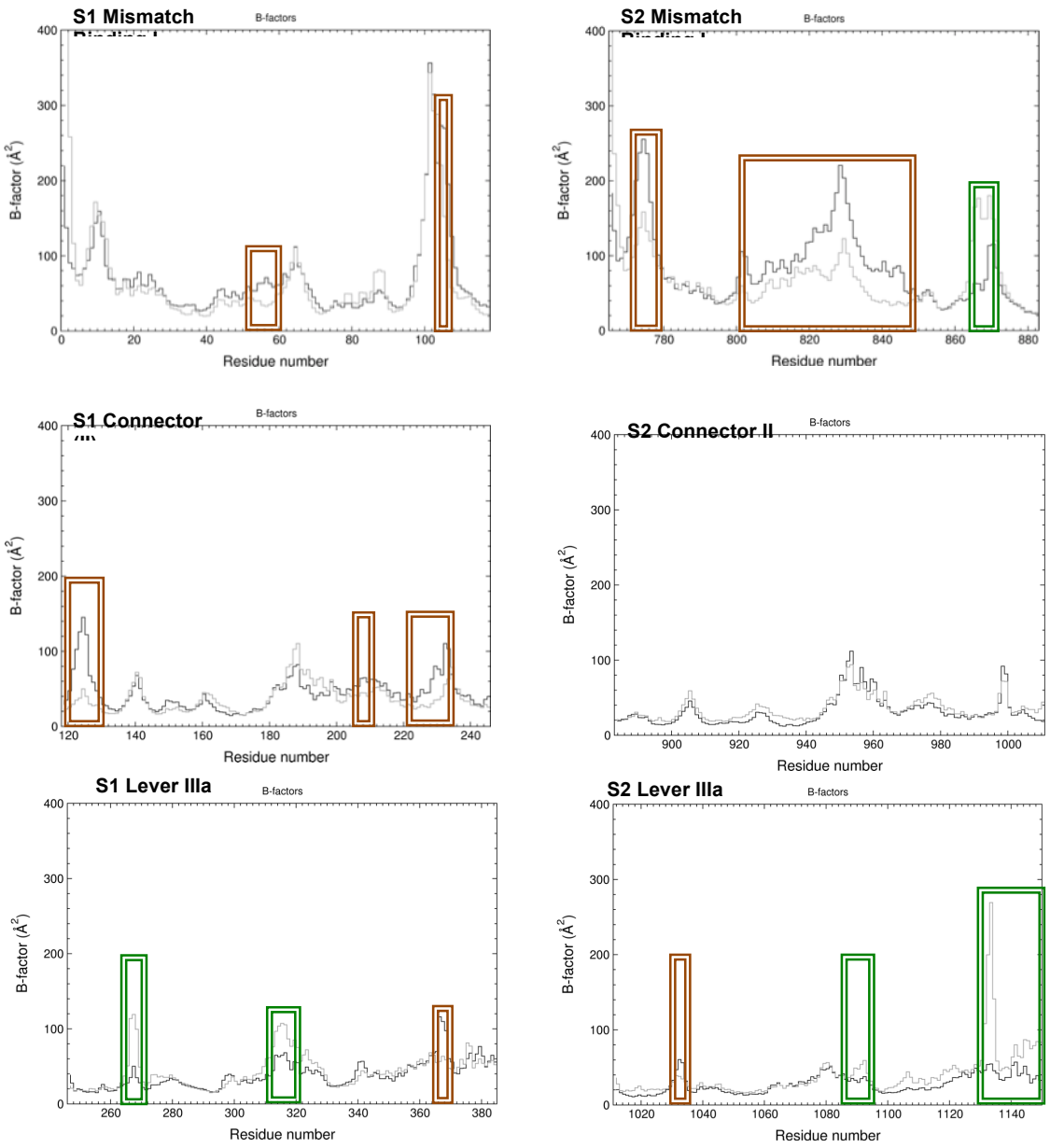


FIGURE S14. Plots of calculated B-factors per residue decomposed by domain. Relative changes in B-factors upon ATP binding: an increase of > 30% highlighted by a brown box, a decrease in > 30% is highlighted by a green box. Only residues with a B-factor > 30 Å^2 are shown.

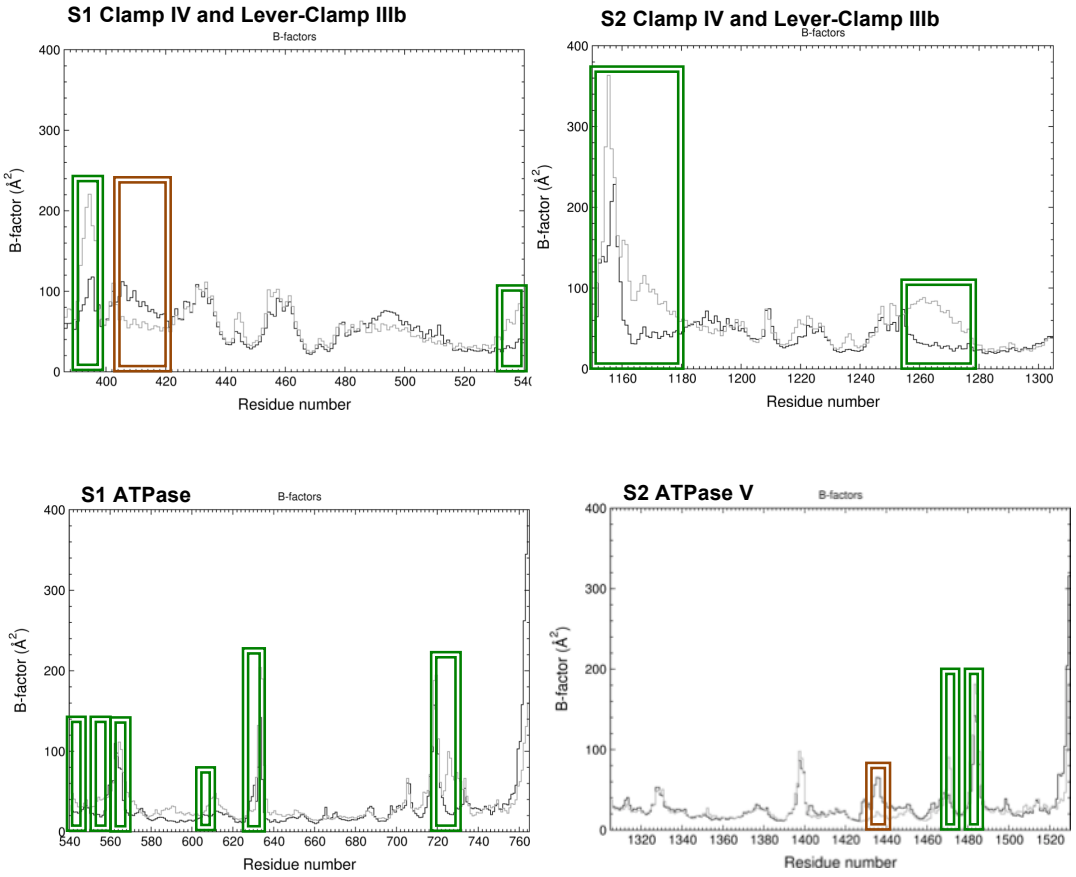


FIGURE S14 (cont). Plots of calculated B-factors per residue decomposed by domain. Relative changes in B-factors upon ATP binding: an increase of > 30% highlighted by a brown box, a decrease in > 30% is highlighted by a green box. Only residues with a B-factor > 30 Å^2 are shown.

TABLE S2A. Eigenvectors of the correlation matrix from the ATP-free complex with the largest overlap to each one of the ten eigenvectors with largest eigenvalues of the correlation matrix from the ATP-bound complex.

ATP-bound Complex	ATP-free Complex	Eigenvector Overlap
1	1	0.78
2	4	–0.74
3	2	–0.85
4	3	–0.55
5	8	0.39
6	10	0.39
7	6	0.66
8	9	–0.42
9	5	0.62
10	8	0.56

TABLE S2B. Eigenvectors of the correlation matrix from the ATP-bound complex with the largest overlap to each one of the ten eigenvectors with largest eigenvalues of the correlation matrix from the ATP-free complex.

ATP-bound Complex	ATP-free Complex	Eigenvector Overlap
1	**1**	0.78
3	**2**	–0.85
4	**3**	–0.55
2	**4**	–0.74
9	**5**	0.62
7	**6**	0.66
6	**7**	0.36
10	**8**	0.56
12	**9**	0.49
6	**10**	0.39

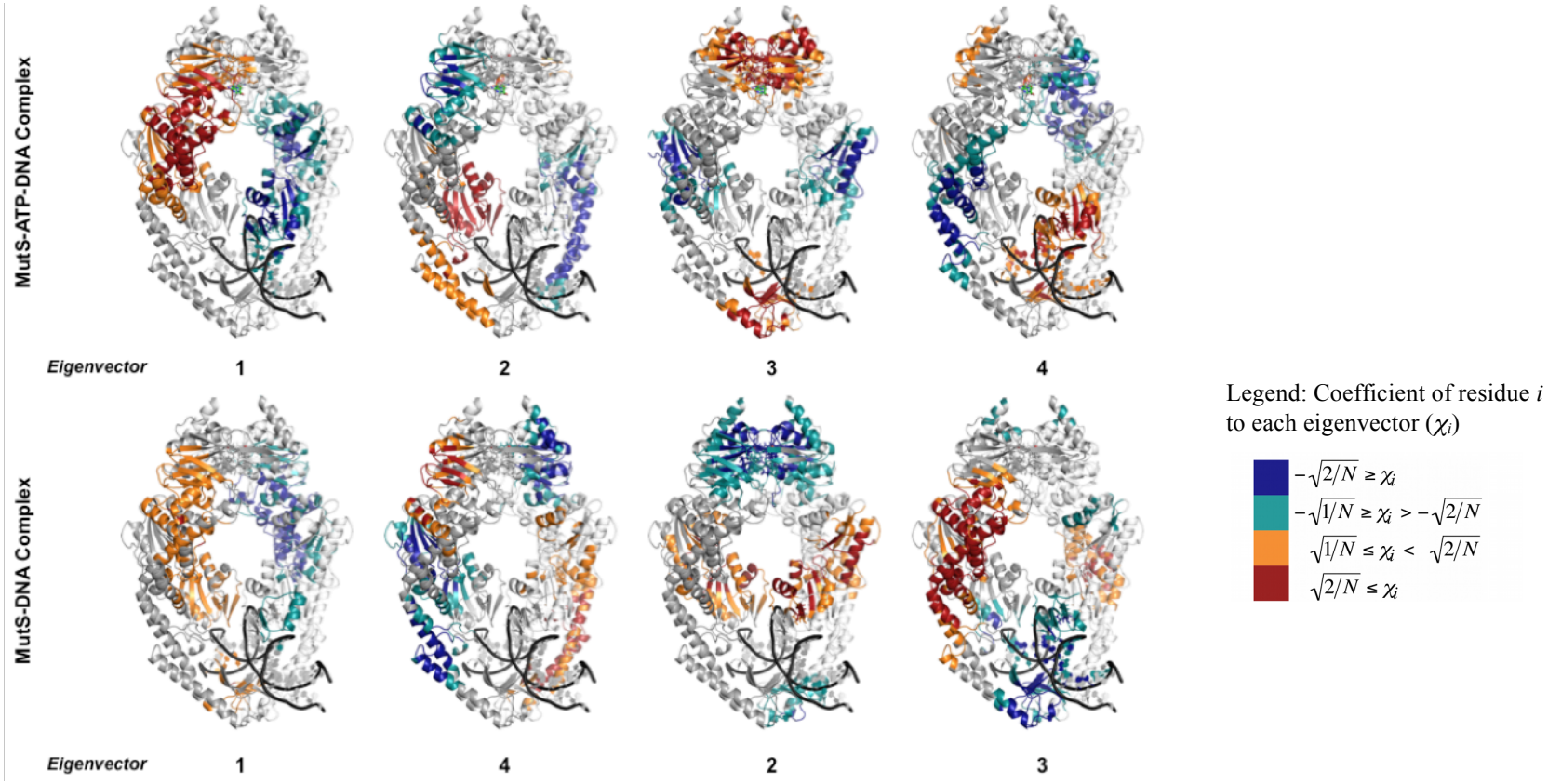


FIGURE S15. Eigenvectors with the four largest eigenvalues of the correlation matrices from MutS-DNA complex (no ATP). Eigenvectors with the four largest eigenvalues of the correlation matrices from ATP-bound and ATP-free mapped onto the corresponding structure (Tables S1A and B). The intensity of the color represents the coefficient (χ_i) of residue motion *i* to each eigenvector. Blue/green vs. red/orange colors correspond to combinations with opposite phase (sign is arbitrary).