

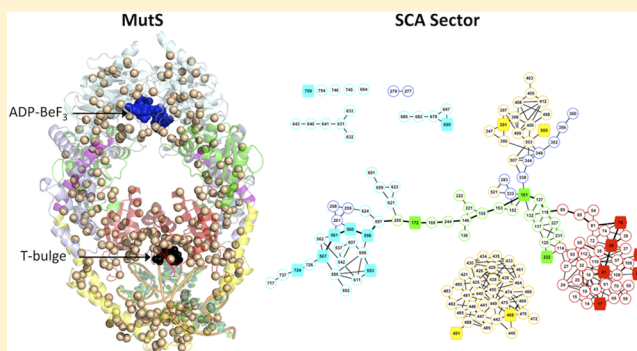
Evolutionary Covariance Combined with Molecular Dynamics Predicts a Framework for Allostery in the MutS DNA Mismatch Repair Protein

Bharat Lakhani,^{†,‡} Kelly M. Thayer,^{§,||} Manju M. Hingorani,^{†,‡} and David L. Beveridge^{*,‡,§,||}

[†]Molecular Biology and Biochemistry Department, [‡]Molecular Biophysics Program, [§]Chemistry Department, and ^{||}Computer Science Department, Wesleyan University, Middletown, Connecticut 06459, United States

S Supporting Information

ABSTRACT: Mismatch repair (MMR) is an essential, evolutionarily conserved pathway that maintains genome stability by correcting base-pairing errors in DNA. Here we examine the sequence and structure of MutS MMR protein to decipher the amino acid framework underlying its two key activities—recognizing mismatches in DNA and using ATP to initiate repair. Statistical coupling analysis (SCA) identified a network (sector) of coevolved amino acids in the MutS protein family. The potential functional significance of this SCA sector was assessed by performing molecular dynamics (MD) simulations for alanine mutants of the top 5% of 160 residues in the distribution, and control nonsector residues. The effects on three independent metrics were monitored: (i) MutS domain conformational dynamics, (ii) hydrogen bonding between MutS and DNA/ATP, and (iii) relative ATP binding free energy. Each measure revealed that sector residues contribute more substantively to MutS structure–function than nonsector residues. Notably, sector mutations disrupted MutS contacts with DNA and/or ATP from a distance via contiguous pathways and correlated motions, supporting the idea that SCA can identify amino acid networks underlying allosteric communication. The combined SCA/MD approach yielded novel, experimentally testable hypotheses for unknown roles of many residues distributed across MutS, including some implicated in Lynch cancer syndrome.



INTRODUCTION

DNA mismatch repair (MMR) initiates with MutS protein recognizing errors made by DNA polymerases, including base–base mismatches and insertion/deletion loops. MutS then activates MutL, which in turn nicks the error-containing strand, setting into motion strand excision and resynthesis by other DNA repair and replication proteins; in *E. coli* and related bacteria that employ methyl-directed MMR, MutL stimulates MutH to nick the strand.^{1–4} MMR is a highly conserved pathway that is essential for suppressing excessive spontaneous mutagenesis that leads to genome instability. Loss of MMR, e.g., due to defective MutS or MutL proteins, results in a high mutator phenotype that is associated with the hereditary Lynch syndrome as well as 10–30% of sporadic tumors in a variety of tissues.^{5,6}

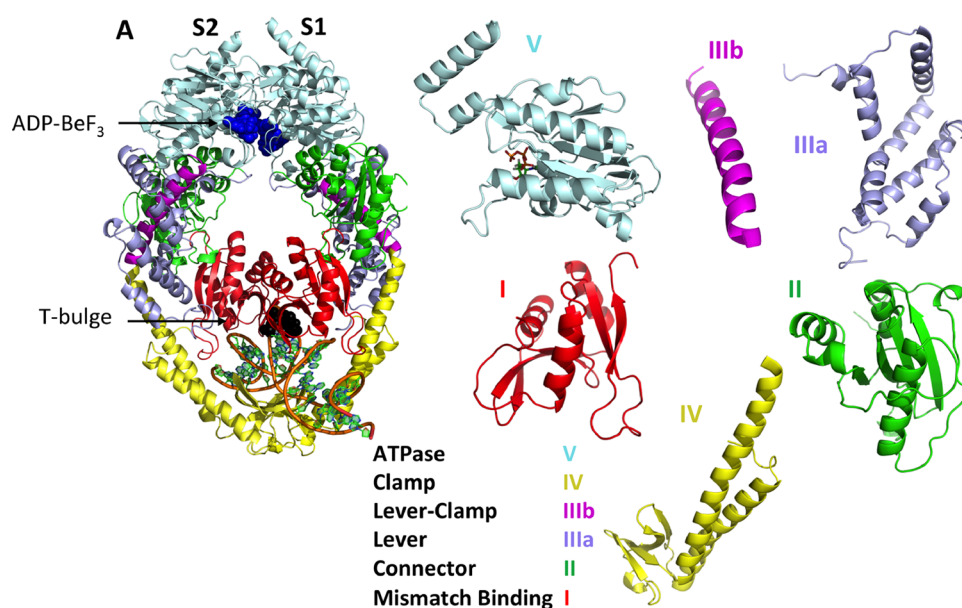
Our study focuses on MutS in order to investigate how this large, multidomain protein employs DNA/mismatch binding and ATPase activities to initiate DNA repair. Crystallographic studies have provided detailed information about the structure–function properties of MutS, especially for the two active sites. However, the functional significance of the vast majority of amino acids beyond the active sites remains unresolved, particularly with respect to residues involved in allosteric signaling between the sites, which are located ~70 Å

apart. One bottleneck is the development of systematic hypotheses to direct mutational analysis of these ~1000-amino acid proteins. We approached this problem by applying statistical coupling analysis (SCA) on the evolutionarily conserved MutS protein family, which identified a network (sector) of 160 covarying residues distributed widely across the protein. We then tested the hypothesis that this coevolved network has functional significance by performing molecular dynamics (MD) simulations on point mutants of the top 5% sector residues and control nonsector residues, to measure their specific contributions to MutS structure–function. The results showed that sector residue mutations predominantly perturb MutS domain structure/dynamics and its interactions with DNA and/or ATP. Moreover, many of these disruptive effects are long-range, indicating that the sector enables allostery between the two active sites on MutS. Thus, the combined SCA/MD approach provided novel, experimentally testable hypotheses about previously unknown functions of MutS residues, including some whose mutations are associated with the Lynch syndrome.

Received: November 29, 2016

Revised: January 24, 2017

Published: January 30, 2017

Domain structure of *T. aquaticus* MutS bound to ADP-BeF₃ and T-bulge DNA

Minimal scheme of MutS actions in MMR

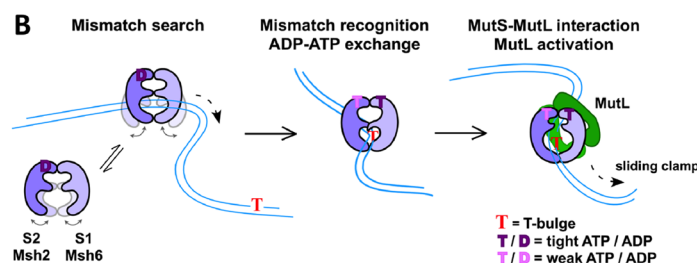


Figure 1. Structure of *T. aquaticus* MutS homodimer. (A) MutS bound to T-bulge DNA (black spheres) and ADP-BeF₃⁻ (blue spheres), with the five domains color coded as follows: mismatch binding domain I (red), connector domain II (green), lever domain IIIa (periwinkle), lever-clamp domain IIIb (magenta), clamp domain IV (yellow), and ATPase domain V (cyan); PDB ID: 1NNE.⁸ (B) Schematic of coupled MutS mismatch-binding and ATPase activities during MMR; briefly, MutS hydrolyzes ATP rapidly and is stabilized in an ADP-bound form that can scan DNA, and mismatch binding suppresses ATP hydrolysis to stabilize the protein in an ATP-bound form that can interact with MutL and initiate DNA repair.

Crystal structures of several MutS proteins bound to mismatched DNA have been solved, including *T. aquaticus*^{7,8} and *E. coli* MutS homodimers^{9–11} as well as human MSH2-MSH6¹² and MSH2-MSH3 heterodimers¹³ (Figure 1A). The interactions of MutS with DNA during mismatch search and recognition, and subsequent interactions with other proteins to initiate repair, are modulated by ATP binding/hydrolysis (Figure 1B).^{14–23} The structures show that the mismatch-binding and ATPase sites on MutS are separated by a distance of about 70–100 Å and intervening protein domains; nonetheless, biochemical studies have revealed that these two active sites are tightly coupled through the repair process, as outlined below (reviewed in ref 14). When not bound to a mismatch, one subunit of the dimer hydrolyzes ATP rapidly and the other slowly (fast: S1/MSH6; slow: S2/MSH2), and MutS with ADP bound to at least one subunit (S2/MSH2) is predominant in steady state.^{15,19,24–27} ADP-bound MutS is capable of enclosing DNA and diffusing along the helical contour probing for mismatched bases.^{16,28–30} Mismatch binding results in ADP-ATP exchange and suppression of ATP hydrolysis, such that ATP-bound MutS becomes

predominant.^{15,24} ATP-bound MutS interacts with and activates MutL, and different studies indicate that the complex can pause at the mismatch or slide away to initiate repair (Figure 1B).^{11,19,21,22,31,32} Thus, ATP binding/hydrolysis and mismatch binding/release are reciprocally linked through allosteric communication between the two active sites on MutS, enabling the protein to transit through different conformations in order to seek, recognize, and initiate repair of a mismatch in DNA.

The structural basis for allosteric communication is difficult to understand at the amino acid level, particularly in a large, multidomain protein such as MutS, although crystal structures and computational studies have offered many insights. The crystal structure of *Thermus aquaticus* (Taq) MutS,^{7,8} the subject of this study, shows a homodimer bound to a 23 basepair DNA containing a T-bulge and ADP-BeF₃⁻ bound in both ATPase sites (Figure 1; PDB ID: 1NNE);⁸ this structure likely reflects an ATP-bound MutS state prior to conformational changes that enable interaction with MutL and movement on DNA.¹¹ Each MutS subunit (S1 and S2) is subdivided into five domains (I–V). The DNA(+T) binds in

the lower channel of the “ θ ” shaped dimer and is kinked by $\sim 60^\circ$ at the T-bulge (all MutS-DNA structures reported thus far have captured this kinked DNA complex).³ The N-terminal mismatch-binding domain (domain I) contains a highly conserved Phe-X-Glu motif that serves as a “reading head” and makes base stacking and hydrogen bonding contacts with the T-bulge. Like MutS ATPase activity, mismatch binding is also asymmetric and the base-specific contacts occur only between one subunit and DNA (S1/MSH6).^{7,12} Together with domains I, the clamp domains (domain IV) from both subunits complete the DNA binding site, enclosing the duplex and making nonspecific contacts with the sugar–phosphate backbone. The C-terminal domain (domain V) contains the highly conserved Walker A (P-loop) and Walker B (Mg^{2+} binding) motifs belonging to the ABC transporter (ATP binding cassette) superfamily that form ATPase active sites at dimer interfaces.³³ The C-terminal domain also contains a conserved helix–turn–helix motif that is involved in MutS dimerization and ATPase activity.³⁴ The remaining connector (domain II), lever (domain IIIa) and lever-clamp (domain IIIb) domains form the core of the protein between the mismatch binding and ATPase active sites; these domains are also part of the interface that binds MutL.^{11,35} The crystal structures of *E. coli* MutS homodimer and human MSH2-MSH6 heterodimer also show the same overall domain organization and interaction with mismatched DNA.^{9,12} Based on these structures, it has been proposed that the long α helix spanning the protein from domain IV to V could transmit allosteric signals between the mismatch binding and ATPase sites (Figure 1A).⁷ Elements in connector domain II, which is surrounded by domains III and V,¹² the junction where domains II, III, and V intersect,^{7,12} the flexible loops lining the upper channel near the ATPase sites,¹² are also proposed to be involved in allostery.^{8,12}

A comparison between the crystal structure of G:T mismatch- and ADP-bound human MSH2-MSH6 and molecular dynamics (MD) simulations of the G:T bound/free and nucleotide-free protein showed subtle reorientation of amino acid residues in the ATPase sites associated with mismatch binding, providing theoretical and experimental evidence of allosteric communication.³⁶ Normal mode calculations of G:T bound/free MSH2-MSH6 and *E. coli* MutS revealed strong motional correlation between the lever (III) and ATPase (V) domains, and between the mismatch binding (I) and ATPase (V) domains, in low frequency modes. These findings suggested that allostery between the two active sites involves coupled motions across MutS domains.³⁷ A subsequent all-atom MD study compared ATP-bound/free forms of Taq MutS to assess nucleotide binding-induced changes in protein structure and dynamics.³⁸ Dynamical cross-correlation maps of the atomic fluctuations indicated inter-residue and inter-domain motional coupling across the protein, and principal component analysis (PCA) of the maps revealed clusters of residues with correlated motions between the active sites, most prominently in the fully liganded ATP- and T-bulge-bound MutS complex. In addition to basepair mismatches, MSH2-MSH6 also binds DNA damage lesions and triggers a cell death response. MD analysis of MSH2-MSH6 bound to G:T mismatch v/s platinum cross-linked DNA revealed distinct changes in disordered loops in the MSH6 lever and MSH2 ATPase domains, which could reflect different allosteric responses to the two DNAs.³⁹ All of these computational studies implicated protein dynamics in allostery and identified the key parts of MutS and the motions that might be involved.

Nonetheless, the underlying amino acid-level architecture in the MutS protein family that enables long-range communication remains elusive. A recent dynamic network analysis based on MD simulations of *E. coli* MutS in all ATP/ADP-bound/free forms (apo, ADP, ATP in each of the two sites), identified sets of physically contiguous residues between the DNA binding and ATPase sites, and postulated that these constitute pathways of communication that change with the nucleotide-liganded state of the protein.⁴⁰ The amino acids making up these pathways (151 of 800) could constitute the underlying architecture for allosteric signaling, although the assumption that allostery occurs through chains of interacting residues in MutS, and the functional significance of the particular residues identified in this study, have not been explicitly tested. Below we describe a complementary, bioinformatics, and molecular modeling-based approach to the problem, using statistical coupling analysis (SCA) to identify network(s) of coevolving residues from multiple sequence alignments of the MutS protein family, followed up by MD analysis to empirically test their involvement in allostery.

The SCA method of identifying cooperative networks of amino acids from evolutionary covariation within a protein family has been applied to a few different proteins in recent years and has yielded new insights into the structural basis of their functions.^{41,42} These groups of coevolved residues are termed statistical or SCA sectors, and they appear to play a significant role in the structure and activity of the proteins analyzed thus far, including PDZ domains,⁴² serine protease (S1A),⁴¹ dihydrofolate reductase (DHFR),⁴³ Hsp70 chaperone,⁴⁴ cathepsin K,⁴⁵ as well as PAS, SH2, and SH3 domains.⁴¹ SCA sectors typically comprise a small fraction ($\sim 20\%$) of the total number of residues in a protein, and present as a distributed, contiguous network that includes the active site(s) and spans distant locations on the protein.⁴² Intra- and inter-domain connections within this network suggest that the residues are involved in long-range allosteric communication, and support amino acid coevolution as a means for establishing a major function within a protein family. For example, sector residues identified in the PDZ domain family connect the ligand-binding site with an allosteric site at the opposite surface of the protein. A recent high-throughput experiment tested the predictive capability of SCA by mutating each of the 83 amino acids in the PSD95^{pdz3} domain to every other amino acid and testing the impact of these changes on ligand binding affinity.⁴² The results showed that sector residues were much more likely to be important for PDZ function compared with nonsector residues. Mutational analysis of some sector residues in Hsp70 also supports the possibility that SCA can identify amino acid networks that allosterically couple distant active/regulatory sites on a protein.⁴⁴ In this study, we applied SCA to the MutS family of mismatch repair proteins to discover the presence of such a network shaped by coevolution, and then subjected several residues in the predicted network to mutational analysis by MD in order to independently assess their contributions to structure–function. The results provide a foundation for new ideas about the roles of evolutionarily correlated amino acids in MutS, especially in allosteric communication.

METHODS

Multiple Sequence Alignment (MSA) of MutS Homologues. A total of 105 orthologous MutS gene sequences were obtained from a previous study on the evolution of archaeal and bacterial MutS, and eukaryotic MSH2, MSH6 genes.⁴⁶ An

additional 59 orthologous sequences were obtained from the Conserved Domain Database: family ID 235444), an NCBI-curated collection of domain models based on 3D structure.⁴⁷ The MutS structure-based alignment from Warren et al.¹² was used as a reference to guide alignment of all 164 collected sequences using Clustal Omega.^{48,49} Sequences with >90% identity were removed, resulting in 142 aligned sequences in the final set subjected to SCA analysis. The LoCo program⁵⁰ was used to postprocess the alignment and detect misaligned positions by local covariance values. The alignment was found to be significant at all positions; therefore, no modifications were required.

Statistical Coupling Analysis (SCA). The SCA 5.0 method as implemented in MATLAB (The MathWorks, Inc., Natick, MA, USA) was used for analysis of evolutionary covariance in the MutS protein family. SCA is described in a series of articles by Ranganathan and co-workers,^{41–44,51–53} and the program was obtained from the Web site: http://systems.swmed.edu/rr_lab/sca.html. SCA is a multivariate analysis in which the calculated weighted positional correlation matrix of the MSA is decomposed into eigenvalues and eigenvectors by spectral decomposition. Co-evolved positions were obtained from the statistically significant eigenmode identified by comparing the original and 100 randomized alignment distributions. The first eigenvector of the weighted positional correlation was analyzed, and a ~20% cutoff yielded a sector of residues that were outside the expectation of randomized alignments (160 of 739 MutS positions).

Contact Mapping of Sector Residues. Crystal structure based native contact maps have been used extensively as geometrical frameworks to understand the structural basis of protein folding and function.^{54–56} Native contacts formed by heavy atoms of MutS sector residues were calculated by the “shadow map” algorithm;⁵⁵ shadow map identifies all heavy atom based direct contacts within 4–6.5 Å cutoff distance in the crystal structure (default 6.0 Å) and discards potential contacts occluded by intervening atoms. The resulting MutS contact map was visualized using the Visone program⁵⁷ in which sector residues were projected as nodes that were connected if their interatomic distances were lower than the cutoff distance. Maps were created at the default 6.0 Å cutoff and at a slightly longer 6.8 Å cutoff. Contact distance distributions for the most distant residue pairs, A146 (C α)–Y244 (C ϵ) and M250 (C ϵ)–A597 (carbonyl O), were calculated from a 15 ns wild type MutS MD trajectory.

Molecular Dynamics (MD) Simulation. The contribution of sector residues to MutS structure and function was investigated by performing all atom MD simulations for alanine mutants of the top 5% sector residues (21 residues that are 2 σ beyond the mean in the distribution) as well as 10 least covarying residues from the bottom of the distribution as control (termed as nonsector residues). The structure of a MutS-(ATP)₂-DNA(+T) complex modified from 1NNE in a previous MD study was used as the starting point for this analysis; changes to 1NNE in that study entailed modification of ADP-BeF₃⁻ to ATP and introduction of residues at noncrystallizing disordered positions.³⁸ Alanine point mutations were introduced by deleting the side chains of selected residues from the PDB input file except the beta carbon. Standard H-building was implemented on all structures to add hydrogens to the modified crystal structure.

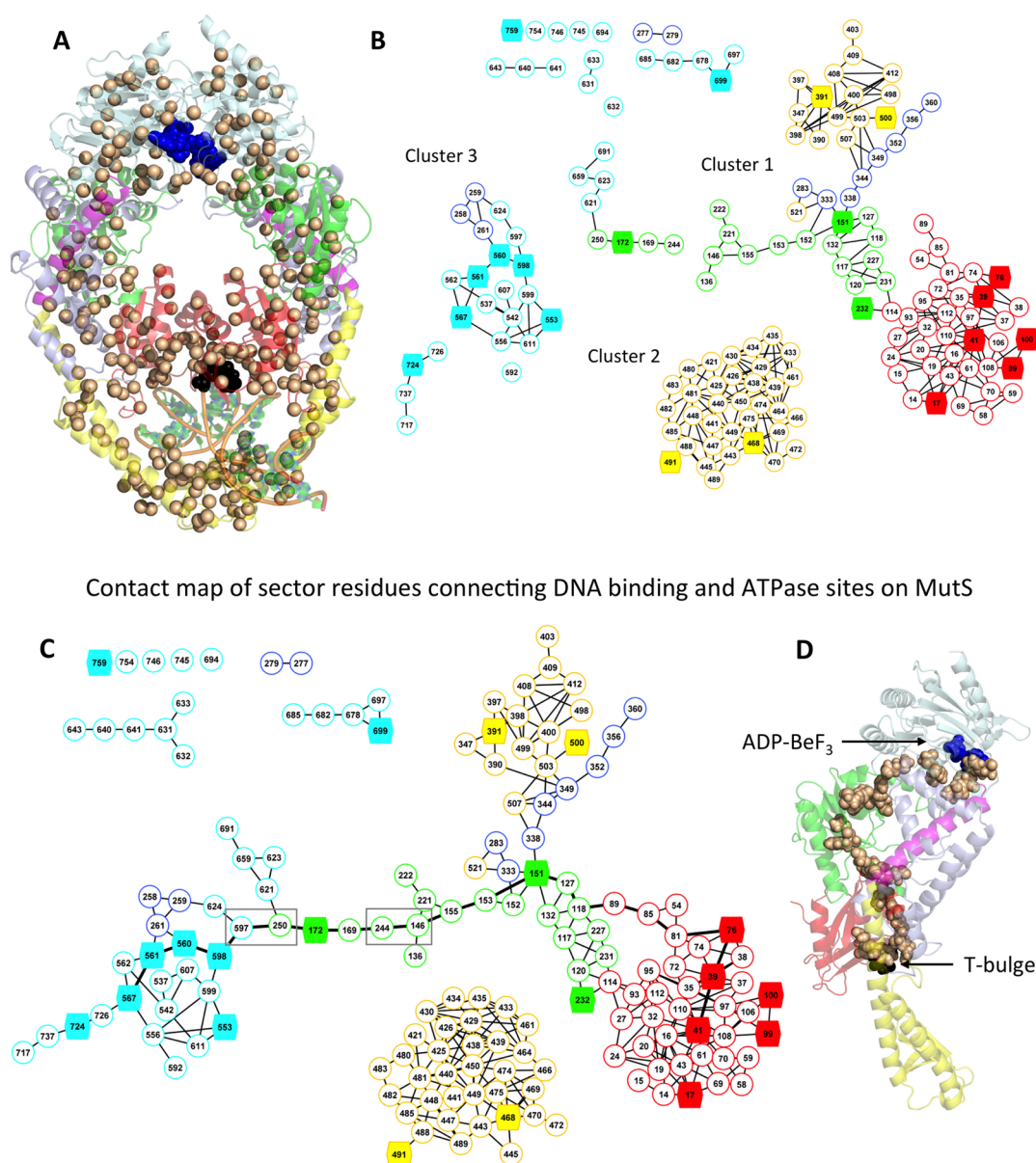
The AMBER12.0⁵⁸ suite was used with the ff14SB force field,^{59–61} which implements ff99bsc0⁶² for DNA and uses

parameters for ions supplied in frcmod.ion08;⁶³ polyphosphate parameters were used for ATP.⁶⁴ Alanine point mutations were created using tleap in the AMBER suite. Each system was solvated in a 12 Å truncated octahedron box of TIP3P⁶⁵ water molecules and electroneutrality was achieved with the addition of Na⁺ counterions. In addition, 150 mM NaCl was added to achieve a biologically relevant ionic strength. The system was treated under periodic boundary conditions. Long-range electrostatic interactions were treated with the particle mesh Ewald (PME) algorithm^{66–68} with a 10 Å Lennard-Jones cutoff. The Berendsen algorithm⁶⁹ maintained the simulations at the target temperature of 300 K. SHAKE⁷⁰ was applied for hydrogen bond motions. The trajectory snapshots were saved at every 2 ps. The equilibrium phase of the simulation, initiated with 1000 steps of steepest descent energy minimization, was followed by 500 steps of conjugate gradient minimization to relax the solvent. This process was iterated four times with successively decreasing harmonic constraints: 100, 100, 10, 0 kcal/mol on solute and 20, 0, 0, 0 kcal/mol on the ions, respectively. Equilibration with harmonic constraints (25, 25, 15, 5 kcal/mol on solute and 20, 20, 10, 0 kcal/mol on the ions, respectively) facilitated heating of the system to 300 K over four 10 ps simulations. An additional 2 ns of dynamics without constraint was considered as equilibration and excluded from the analysis. Production of NPT simulation trajectories proceeded for 15 ns with an integration time of 2 fs. The wild type simulation was continued to 50 ns to check the stability of the trajectories. Equilibration and production level simulations were run on CUDA-enabled NVidia Tesla K20 GPU 2496 cores,^{71–73} which utilized the PMEMD version of sander in the Amber12.0 suite of programs.^{71,72} MD stability verification, dynamic cross-correlation map, and analysis of metrics assessing protein structure–function were computed with the AMBER cpptraj⁷⁴ and MM-GBSA⁷⁵ module on the resultant trajectories. Root-mean-square deviation (RMSD) and hydrogen bond donor–acceptor distances were monitored. Molecular visualization was carried out using VMD⁷⁶ and PyMol.⁷⁷

Sector Residue Contributions to MutS Structure–Function. Metric 1: RMSD Analysis of MutS Domain Structure. The RMSD of MutS domains was chosen as the first metric to assess the effects of alanine point mutations in 21 sector and 10 nonsector residues on protein structure and function. Average domain-wise RMSDs of both subunits in the wild type and mutant MD trajectories were calculated over five 3 ns windows of the total 15 ns simulation by global fitting of heavy backbone atoms onto the initial reference structure. A difference was considered significant if it varied at least 2 σ from the mean of the wild type within the 3 ns window (Table S2). Variations were detected in the mismatch binding (I) and clamp (IV) domains, but not in the connector (II), lever (III), or ATPase domains (V); hence results are shown for I and IV (as well as V for comparison).

Metric 2: H-Bonding in the Mismatch Binding (R76) and ATPase (K589) Active Sites. The second metric to assess the effects of sector and nonsector residue alanine mutations was hydrogen bonding to DNA and ATP. The MD ensemble of the MutS-(ATP)₂-DNA(+T) complex³⁸ reveals stable hydrogen bonds between R76 N η ¹ and R76 N η ² in the S1 subunit mismatch binding domain with cytosine 1545 and guanine 1546 backbone oxygens, respectively (note: R76 is present in the top 5% of sector residues and its mutation to alanine reduces the relative ATP binding free energy by 9.1 kcal/mol as

Spatial distribution of 160 sector residue positions on MutS identified by SCA



Contact map of sector residues connecting DNA binding and ATPase sites on MutS

Figure 2. SCA sector residues in MutS and their contact maps at 6.0 and 6.8 Å cutoff distances. (A) The 160 sector residues identified by SCA are projected onto both S1 and S2 subunits of Taq MutS (wheat spheres depict side chains and backbone). (B) A shadow contact map of the residues with the interatomic distance of heavy atoms at 6.0 Å cutoff⁶⁵ partitioned the sector into three major contiguous clusters (residues: numbered circles or hexagons; contacts: black lines), with cluster 1 linking domains I (red), II (green), IIIa (periwinkle), and IV (yellow); cluster 2 covering domain IV (yellow); and cluster 3 linking domains IIIa (periwinkle) and V (cyan). The top 5% sector residues selected across these clusters for alanine mutation and MD analysis are shown as filled hexagons (21 positions). (C) A shadow contact map at 6.8 Å cutoff extends the network across all five MutS domains, with contiguous pathways connecting the mismatch-binding and ATPase active sites. Gray boxes denote residue pairs at the longest distance (6.5–6.8 Å) that were analyzed further by MD (Figure S2). One possible pathway is highlighted by bold lines in the contact map and (D) projected on the MutS structure (wheat, blue and black spheres depict pathway residues, ATP and T-bulge, respectively).

described below in Metric 3); since the results are similar for both H-bonds, only the one with guanine is shown. K589 is a conserved residue in the Walker A motif required for MutS ATPase activity (K589M mutation increases K_M by 18-fold³³). Two H-bonds are formed between K589 N_ζ and ATP P^β and P^γ oxygens. The alanine mutation was considered disruptive if the H-bond was present in less than 20% of the MD trajectory snapshots.

Metric 3: MM-GBSA Free Energy of Ligand (ATP) Binding. The third metric used to assess the effects of alanine mutations

was the relative ATP binding free energy calculated by the molecular mechanics generalized Born surface area (MM-GBSA) method.⁷⁵ MM-GBSA combines molecular mechanics and continuum solvent-based generalized Born energies to calculate the relative ligand binding free energy.^{78,79} The calculation was performed on 100 snapshots averaged over each 15 ns MD simulation. The relative binding free energy ($\Delta\Delta G_{\text{binding}}$) on ATP binding to both subunits of the MutS-DNA(+T) complex between wild type ($\Delta G_{\text{binding}}(\text{wt})$) and each

of the 21 sector and 10 nonsector mutated residues ($\Delta G_{\text{binding}(\text{mut})}$) was determined by

$$\Delta\Delta G_{\text{binding}} = \Delta G_{\text{binding}(\text{wt})} - \Delta G_{\text{binding}(\text{mut})}$$

where

$$\Delta G_{\text{binding}(\text{wt or mut})} = G_{\text{MutS-ATP-DNA(+T),solvated}} - [G_{\text{MutS-DNA(+T),solvated}} + G_{\text{ATP,solvated}}]$$

The free energy associated with each term on the right side of above equation is estimated by standard MM-GBSA methods:

$$G_{\text{solvated}} = E_{\text{internal}} + E_{\text{electrostatic}} + E_{\text{VDW}} + E_{\text{polar solvation}} + G_{\text{nonpolar solvation}} - TS$$

where E_{internal} (bond, angle and torsion), $G_{\text{electrostatic}}$ (electrostatic), and E_{VDW} (van der Waals) interaction energies are molecular mechanical energies; $G_{\text{nonpolar solvation}}$ is solvation free energy calculated by generalized Born implicit solvent model; $G_{\text{nonpolar solvation}}$ is calculated with a linear dependence to the solvent accessible surface area.⁷⁵ Note that the entropy contribution is neglected in the relative binding free energy calculation. $\Delta\Delta G_{\text{binding}}$ was considered to be significant if the value was >2.5 kcal/mol.

Molecular Snapshots of Allosteric Effects as Reported by R76 and K589. In the mismatch binding site, all atoms of R76 and guanine 1546 were aligned between the wild type (green) and each mutant structure (magenta) in Pymol.⁷⁷ K589 and ATP were similarly aligned in the ATPase site. The final MD snapshots of the wild type and mutant proteins are shown to illustrate changes in both sites; movies of the MD trajectories for wild type MutS and two mutants, R172A and I553A, are included in [Supporting Information](#).

RESULTS AND DISCUSSION

Identification of a Coevolved Amino Acid Network in MutS. Statistical Coupling Analysis. The promise of SCA is the discovery of patterns of correlation among amino acids arising from functional constraints on a protein through evolution. Unlike multiple sequence alignment-based approaches that provide information on evolutionarily conserved residues, SCA resolves residues that are coevolving, not just pairwise but as a network. The hypothesis currently being tested in the field is that this coevolved network can provide a basis for understanding the structure/dynamics and interactions underlying protein function. We applied SCA to the MutS family of proteins because little is known about how the architecture of this critical DNA mismatch repair protein enables it to find and initiate repair of base pairing errors in DNA with the help of ATP. The SCA matrix calculated for a curated multiple sequence alignment of 142 MutS homologues yielded a top eigenvector that was well separated from statistical noise compared to a randomized sequence alignment ([Figure S1](#)). This eigenvector revealed a protein sector comprising 160 residues of 739 total in the aligned sequences ($\sim 22\%$; [Table S1](#)). The sector residue positions did not correlate strongly with highly conserved residues ([Figure S1](#)), indicating that SCA captured information beyond conservation in the MutS protein family.⁸⁰

Projection of the sector residues on the *T. aquaticus* MutS dimer structure shows that they are widely distributed over all five domains of the protein ([Figure 2A](#)), with a large majority (77.5%) located in the DNA binding (mismatch binding, I;

clamp, IV) and ATPase (V) domains. Only about 30% of the residues are less than 8 Å from the mismatch binding and ATPase sites; thus, the SCA sector includes key active site residues as well as distant residues that may have indirect effects on DNA binding and ATP binding/hydrolysis. Some of these 160 residues have been previously identified or postulated as important for MutS structure–function, providing support for the significance of the SCA sector. These include (i) F39 and E41 of the F-X-E motif in domain I that stack and hydrogen bond with the mismatched base, respectively;^{7,81} (ii) E99, P100, G106 in the glutamate-rich VEPAAEEAEG loop in domain I, which is part of a β hairpin that contacts DNA⁷ and moves away when MutS binds ATP, possibly facilitating sliding of the protein on DNA;³⁸ (iii) 18 positions that are implicated in key hydrogen bonding and salt bridge interactions for carboplatin- and cisplatin-lesion recognition by human MSH2-MSH6;⁸² (iv) F567 in domain V that stabilizes the adenosine base in the ATPase active site;³³ (v) L631, A632 and G633 in the SDDLGGKST loop in domain V, containing the highly conserved N-2 motif (ST) that binds the ATP γ -phosphate and contacts lever domain III in the other subunit;^{7,8,38} (vi) A745, G746, R754 and L759 in the helix-turn-helix motif in domain V, which is involved in MutS dimerization and ATPase activity;^{7,34} (vii) A537, F724, and H726 (A562, F758, and H760 in *E. coli* MutS, respectively) that lie at the interface between MutS and MutL;¹¹ (viii) 35 residues in the region proposed as a “transmitter” of signals between the DNA binding and ATPase sites, formed by the junction of domains II, III, and V and an α helix in domain IV ([Table S1](#));⁷ (ix) 48 positions that overlap with 151 pathway residues identified by a recent dynamic network analysis of *E. coli* MutS;⁴⁰ and finally, (x) 48 positions that overlap with mutations in human MSH2 and MSH6 subunits associated with Lynch cancer syndrome ([Table S1](#); <http://insight-group.org>).⁵

Mapping of SCA Sector Residues. The literature survey above shows that the SCA sector includes amino acids previously implicated in MutS function, providing initial support for the utility of the method. However, given that most of the sector residues have no previously assigned function, a systematic analysis was initiated to examine whether and how the sector might enable MutS actions in MMR. First, the spatial relationship between sector residues was determined by a shadow contact map, in which interactions between amino acids are defined by an interatomic distance cutoff between heavy atoms and those potentially occluded by intervening atoms are removed.⁵⁵ [Figure 2B](#) shows all 160 sector residues colored by MutS domain, with lines connecting those in contact with each other at a default 6.0 Å cutoff. The residues group into three major structurally contiguous clusters, with cluster 1 linking domains I (red), II (green), IIIa (periwinkle) and IV (yellow); cluster 2 covering domain IV (yellow); and cluster 3 linking domains IIIa (periwinkle) and V (cyan). In the case of smaller proteins examined previously by SCA, such as PDZ and multidomain Hsp70, most of the sector residues form a single contiguous network spanning the protein.^{42,44,53} In case of MutS, the sector manifests as discrete clusters of connected residues, which might reflect local architecture that supports distinct MutS activities, such as DNA/mismatch binding, ATPase, interaction with MutL, and/or allostery through mechanisms other than physically connected pathways between distant sites.^{83,84} Notably, a small increase in the cutoff distance to 6.8 Å results in a large network that includes 67% of the sector residues and connects all five MutS domains; cluster 2,

Table 1. Metrics for Assessing the Effects of Sector and Nonsector Residue Mutations on MutS^a

	position of mutation in MutS wild type MutS	H-bond to DNA 73%	H-bond to ATP 87%	minimum distance: mutation to DNA NA	minimum distance: mutation to ATP NA	relative ATP binding free energy v/s wild type NA	domain destabilization in S1, S2 subunits NA
sector residues							
1.	L17A	11%	80%	3.9 Å	62 Å	3.332 (0.528)	S2–IV
2.	F39A	60%	0%	1.6 Å	57 Å	12.611 (2.291)	S1–IV, S2–I, IV
3.	E41A	82%	90%	3.5 Å	55.2 Å	10.409 (2.285)	S2–I, IV
4.	R76A	NA	68%	4.4 Å	49.8 Å	9.154 (2.432)	S2–I, IV
5.	E99A	13%	56%	7.5 Å	63.9 Å	9.159 (1.986)	
6.	P100A	19%	0%	6.3 Å	65.1 Å	−0.311 (2.148)	S2–IV
7.	S151A	18%	59%	28.8 Å	38.7 Å	7.891 (1.864)	S1–I
8.	R172A	11%	0%	41.8 Å	25.0 Å	24.626 (10.134)	S1–IV, S2–IV
9.	Q232A	54%	0%	20.9 Å	48.5 Å	8.684 (2.174)	S2–IV
10.	P391A	62%	63%	15 Å	53.8 Å	−17.926 (1.833)	S1–I
11.	Q468A	69%	0%	2.7 Å	75.7 Å	1.082 (2.476)	S1–I, IV, S2–IV
12.	L491A	56%	85%	19.3 Å	72.7 Å	4.662 (1.765)	S1–IV, S2–IV
13.	E500A	74%	0%	15 Å	61.7 Å	−15.905 (1.365)	
14.	I553A	15%	0%	60 Å	18 Å	13.804 (1.976)	S1–I, S2–I
15.	V560A	17%	0%	40 Å	11 Å	0.717 (2.321)	S1–IV
16.	V561A	67%	56%	45 Å	8.5 Å	0.612 (0.130)	S2–I, IV
17.	F567A	61%	0%	49 Å	3 Å	2.614 (3.050)	S2–I, IV
18.	L598A	2%	0%	45 Å	13 Å	15.052 (1.550)	S2–I, IV
19.	E699A	9%	0%	70 Å	20 Å	7.562 (2.254)	S2–I, IV
20.	F724A	0%	81%	57 Å	10 Å	11.504 (0.335)	S1–IV
21.	L759A	9%	0%	70 Å	16 Å	−14.838 (0.936)	S1–I, IV, S2–IV
	position of mutation in MutS wild type MutS	H-bond to DNA 73%	H-bond to ATP 87%	minimum distance: mutation to DNA NA	minimum distance: mutation to ATP NA	relative ATP binding free energy v/s wild type NA	domain destabilization in S1, S2 subunits NA
nonsector residues							
1.	F78A	67%	0%	10 Å	45 Å	7.955 (2.123)	
2.	K161A	19%	83%	41 Å	30 Å	3.047 (2.191)	
3.	Y167A	55%	65%	43 Å	23 Å	0.650 (1.947)	
4.	L191A	62%	61%	39 Å	21 Å	0.734 (1.478)	
5.	E204A	67%	54%	36 Å	34 Å	0.069 (0.946)	
6.	P206A	58%	69%	39 Å	39 Å	0.685 (1.222)	
7.	P209A	75%	55%	45 Å	47 Å	1.857 (2.864)	
8.	L341A	58%	67%	52 Å	28 Å	0.304 (1.335)	
9.	I358A	61%	64%	36 Å	59 Å	1.156 (2.868)	S1–IV, S2–IV
10.	R550A	74%	70%	41 Å	65 Å	0.916 (0.324)	S1–IV, S2–IV

^aPercent hydrogen bond retention in MD trajectory snapshots between reporter R76 N η^2 and guanine 1546 backbone oxygen in DNA, and reporter K589 N ϵ and ATP P $^{\beta}$, in the S1 subunit, and the minimum distances between the mutation and reporter sites are listed. Also listed are the MM-GBSA calculated ATP binding free energies of mutants relative to wild type MutS (std. dev. in parentheses), and the domains destabilized by mutations in S1 and S2 subunits. Label color is keyed to that of MutS domains as shown in Figure 1.

which covers domain IV, remains distinct (Figure 2C). Two residue pairs make critical contacts in this larger network (highlighted by gray rectangles in Figure 2C), A146 (C α)–Y244 (C ϵ) and M250 (C ϵ)–A597 (carbonyl O), at average interatomic distances of 6.8 and 6.5 Å, respectively, as calculated from a wild type MutS MD trajectory (Figure S2); the minimum distances between these pairs are 5.7 and 5.0 Å, respectively, within the range for short-range interactions. This network does include contiguous pathways linking the DNA binding and ATPase sites, one of which is projected on a subunit of Taq MutS (Figure 2D).

Functional Significance of the SCA Sector. *MD Simulations of Alanine Mutants.* As noted in the Introduction Section, the structural network defined by an SCA sector has potential functional significance, assuming that the amino acids have coevolved to retain key structure–function properties of the protein. A few studies have tested this hypothesis

experimentally by mutating a subset of sector residues and assessing the impact on function, e.g., Hsp70.⁴⁴ Such analysis is generally limited in scale, especially for large proteins like MutS, given the need to generate a vast number of mutants and monitor multiple interactions and activities for systematic testing. The difficulty is compounded by the need for combinatorial mutagenesis to investigate any coordination among residues in the network. We tackled the problem by combining SCA with all-atom molecular dynamics (MD) simulations of sector residues mutated to alanine. The goal was to monitor any changes in local or global dynamics associated with the mutations, and determine whether these changes could perturb MutS structure and related functions.⁸⁵ We selected 21 positions constituting the top 5% of the sector residue distribution (Figure 2B, C, filled hexagons), and 10 least covarying positions as control “nonsector” residues for a proof-of-principle mutational study. Of the 21 sector residues, ten

Sector residue mutations disrupt distant MutS domains

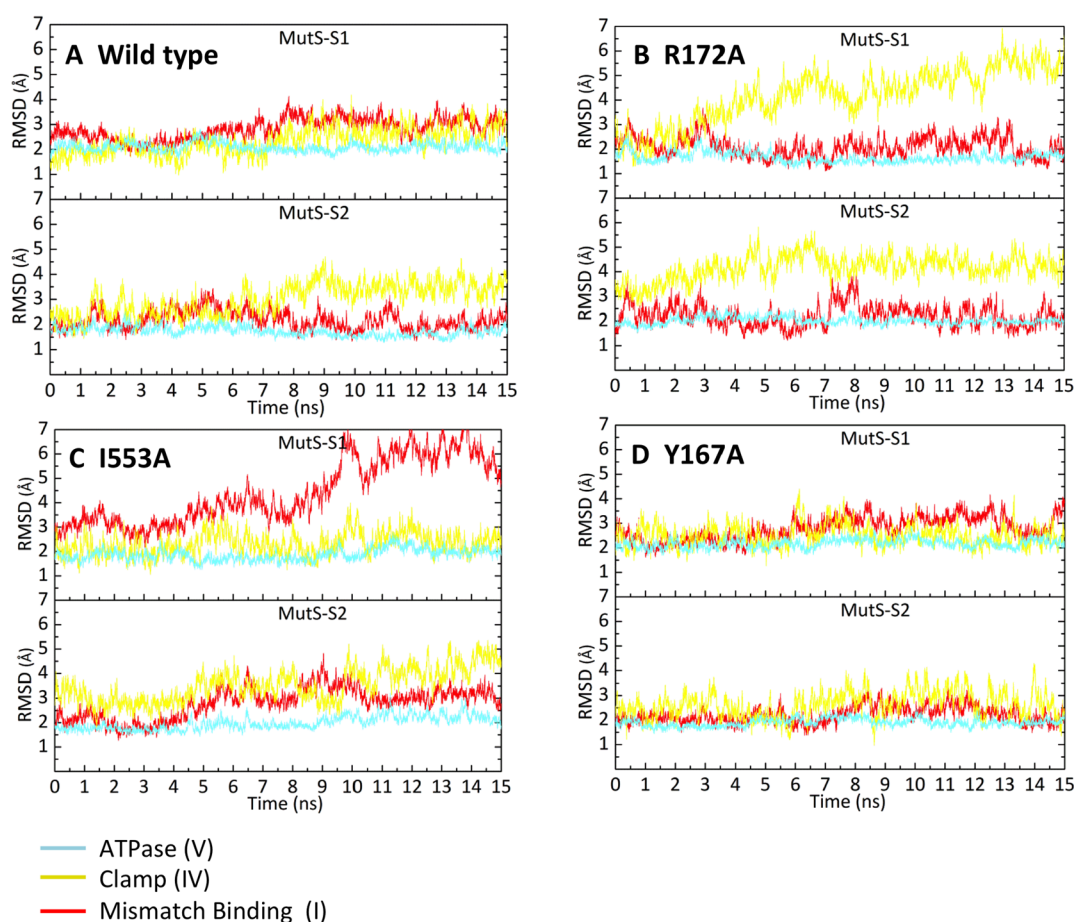


Figure 3. Domain-wise 1D-RMSD comparison between wild type and sector/nonsector residue alanine mutants. The 1D-RMSD was calculated as a function of MD time for (A) wild type MutS and alanine mutants of 21 sector and 10 nonsector residues. Three of these are highlighted here (see Figure S3 and S4 for the remainder; Table 1). (B) R172A (located in connector domain II, Figure 2C) and (C) I553A (located in ATPase domain V, Figure 2C), as well as (D) Y167A, a nonsector control residue in the connector domain. Data are shown for the mismatch binding (I) and clamp (IV) domains of both S1 and S2 subunits, color-coded as in Figure 1; the ATPase (V) domain, which does not undergo significant changes, is shown for comparison (no significant changes were detected in domains II and III either).

were from various MutS domains in cluster 1, two were from cluster 2, five from cluster 3, and the remaining four were distributed across individual or small groups of residues that did not map to these networks (Figure 2B, C). MD simulations were performed with the 31 total mutants to compare the effects of alanine mutations in sector versus nonsector residues.

The simulations were performed with a MutS-(ATP)₂-DNA(+T) complex³⁸ derived from the MutS-(ADP-BeF₃)₂-DNA(+T) crystal structure,⁸ which represents a critical MutS intermediate in the MMR pathway. Additional simulations were performed with ADP-bound and mixed ATP/ADP-bound intermediates for select mutants as well. Each mutant was subjected to 15 ns MD simulations, and the data were analyzed with respect to three key measures of MutS structure/dynamics and function, namely domain conformation and interactions with DNA and ATP. The first metric was domain-wise RMSD values calculated for both S1 (mismatch binding) and S2 subunits from the MD trajectories of wild type and mutant protein dimers (Table 1, Figures 3, S3, and S4). The second metric was the stability of hydrogen bonds in the active sites—between the mismatch binding domain I and DNA, and between the Walker A motif and ATP (Table 1, Figures 4, S5,

and S6). R76, a top 5% sector residue in the mismatch binding subunit S1, makes two stable hydrogen bonds with DNA flanking the T-bulge (between R76 N η ¹ and N η ² hydrogens and cytosine 1545 and guanine 1546 backbone oxygens, respectively; only the bond with guanine is shown for clarity).³⁸ These contacts may help distort the DNA to facilitate F-X-E interaction with the mismatched base (widening of the minor groove and kinking toward the major groove).⁷ The second set of hydrogen bonds is between K589, a highly conserved lysine in the Walker A motif/P-loop that is essential for ATPase activity, and ATP (between K589 N ζ and ATP P $^{\beta}$ and P $^{\gamma}$ oxygens).^{7,33} The third metric was the relative ATP binding free energy calculated by the molecular mechanics combined with generalized Born surface area method (MM-GBSA; Table 1), which is used often to estimate ligand binding affinities based on MD simulations of ligand-macromolecule complexes,^{85,86} including recently to illustrate changes in *E. coli* MutS affinity for DNA when bound to ATP⁸⁷ (MM-PBSA yielded similar results; data not shown).

Sector Residue Mutations Perturb MutS Active Site Structure/Dynamics. Results for all three measures described above are presented in Table 1 for the 21 sector and 10

Sector residue mutations exert long range effects on DNA binding and ATPase sites

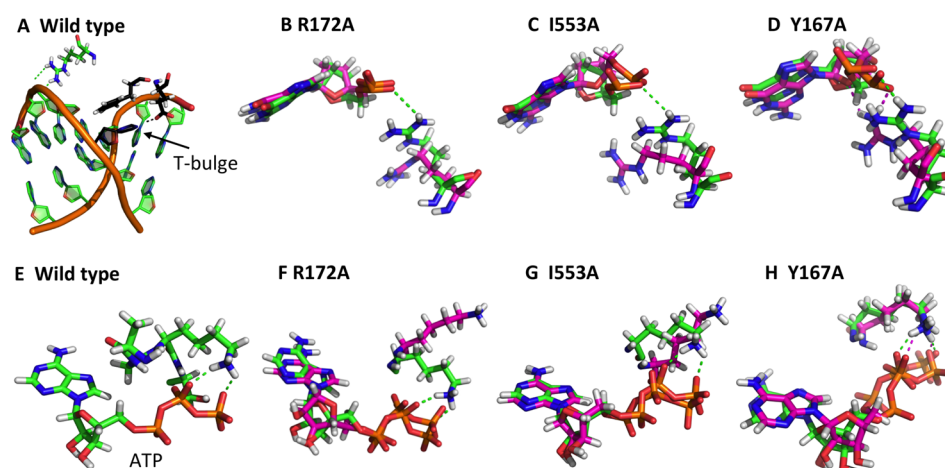


Figure 4. MD snapshots of the long-range effects of alanine mutations on active site contacts between MutS and DNA/ATP. (A) In wild type MutS, R76 (green) makes two H-bonds with the DNA backbone near the T-bulge in DNA stabilized by F-X-E motif (black sticks). Three mutants are highlighted here (see Figure S5 and S6 for the remainder; Table 1). (B–D) All atom local alignment of R76 and guanine 1546 between wild type (green) and mutant (magenta) proteins shows the H-bonds disrupted in (B) R172A and in (C) I553A sector mutants (Table 1), but not in (D) Y167A nonsector mutant. (E) In wild type MutS, K589 (green) in the Walker A motif makes two H-bonds with the ATP β and γ phosphate oxygens (red). (F–H) All atom local alignment of K589 and ATP between wild type (green) and mutant (magenta) proteins shows disruption of the H-bonds in (F) R172A and (G) I553A sector mutants, but not in (H) Y167A nonsector mutant.

nonsector alanine mutants of MutS. Overall, 20 of the 21 sector mutations disrupt at least two metrics and 10 disrupt all three metrics, whereas only 2 of the 10 nonsector mutations disrupt two metrics (F78A, K161A) and none disrupt all three (Table 1). Two sector mutants (R172A, I553A) and one control nonsector mutant (Y167A) are highlighted in Figures 3 and 4 and the remainder are presented in Supporting Information (Figure S3–S6). With respect to the first metric, domain-wise RMSD across the MD trajectories of several mutants showed significant changes in the mismatch binding (I) and DNA binding clamp (IV) domains (defined as 2σ from the mean of the wild type within 3 ns windows; Table S2). No significant changes were observed in domains II, III and V; therefore, trajectories are shown only for domains I and IV, plus the stable ATPase domain (V) for comparison (Figure 3). Overall, 19 of 21 sector mutants and only 2 of 10 nonsector mutants showed domain destabilization (Table 1). For example, the clamp domains of both S1 and S2 subunits are destabilized in R172A (Figure 3B), and the mismatch binding domains of both S1 and S2 subunits are destabilized in I553A (Figure 3C). In contrast, all domains in both S1 and S2 subunits remain stable in the Y167A nonsector mutant (Figure 3D). MD simulations of both R172A and I553A were extended further to 50 ns and yielded the same results (Figure S3). Four independent 50 ns MD simulations were performed for the I553A mutant, and all of them yielded the same results as well (data not shown). With respect to the second metric, hydrogen bonding between MutS and either DNA or ATP is disrupted in 17 of 21 sector mutants and only 2 of 10 nonsector mutants (Table 1). For example, hydrogen bonding between R76 and the DNA backbone flanking the T-bulge is disrupted in both R172A and I553A mutants but not in Y167A (defined as less than 20% H-bond retention over the MD trajectory; Figure 4B–D, respectively). Hydrogen bonding between K589 and ATP is also disrupted in both R172A and I553A mutants but not in Y167A (Figure 4F–H, respectively); movies of the changes in hydrogen bonding

over the MD trajectories in both active sites are included in Supporting Information. Finally, with respect to the third metric, the relative ATP binding free energy also changed significantly for 17 of 21 sector mutants and only 2 of 10 nonsector mutants ($\Delta\Delta G_{\text{binding}} > 2.5$ kcal/mol); e.g., for both R172A and I553A mutants but not for Y167A (Table 1). Thus, an overwhelming majority of the sector residues tested in this proof-of-principle MD analysis contribute to structural properties of MutS that are critical for its function.

As noted earlier, the MutS dimer has asymmetric ATPase activity; accordingly, it can adopt nine different ATP/ADP-bound/free forms during the ATPase reaction cycle. We chose to perform MD simulations first with the MutS-(ATP)₂-DNA(+T) complex, since ATP-bound MutS is a key intermediate that is stabilized after mismatch recognition and undergoes conformational changes resulting in interaction with MutL to license nicking of the error-containing strand. Previous MD studies have shown that other nucleotide-bound forms of *E. coli* MutS maintain the same overall structure,^{37,40} which implies that the structural network revealed by SCA could contribute to function in these forms as well. We tested this hypothesis by performing MD simulations on some of the ADP-bound forms of MutS that occur prior to mismatch recognition: MutS-ADP_{S1}ATP_{S2}-DNA(+T), MutS-ATP_{S1}-ADP_{S2}-DNA(+T), and MutS-(ADP)₂-DNA(+T).¹⁴ Table S3 shows the effects of mutating sector residues R172 and I553 as well as nonsector residue Y167 to alanine on ADP-bound MutS. All ADP-bound forms of R172A and I553A show domain destabilization, but not Y167A (except ATP_{S1}-ADP_{S2}-bound MutS). Also, hydrogen bonding between R76 and the DNA backbone is disrupted in R172A and I553A but not in Y167A for all three ADP-bound MutS forms. The same is true for hydrogen bonding between K589 and the nucleotide (except all mutants retain this contact in ADP₂-bound MutS). Thus, the results are the same overall as for ATP-bound MutS,

where sector mutants are substantively more disruptive than nonsector mutants.

Involvement of the SCA Sector in Allostery Across MutS. Remarkably, 20 of the 21 sector residue mutants tested in this study exert their disruptive effects on MutS domain conformation or interactions with DNA and/or ATP from a distance. This property is indicative of their involvement in allosteric communication across the protein. For example, R172A in connector domain II perturbs DNA binding clamp domain IV conformation as well as interaction with ATP from a distance of ~ 25 Å, and I553A in ATPase domain V perturbs mismatch binding domain I conformation as well as interaction with mismatched DNA from a distance of ~ 60 Å (Table 1, Figures 3 and 4). As noted earlier, physical contiguity is a characteristic of SCA sector residue networks identified in proteins thus far, providing support for the view that allostery occurs through direct interactions between amino acids spanning distant locations.^{51,53,88,89} Such allosteric pathways might operate in MutS as well, since about 2/3 of the sector residues are part of a contiguous network that spans all five domains between the mismatch binding and ATPase sites.

However, we note that sector residues that lie outside of the contiguous network in MutS also show effects at a distance (Figure 2C). For example, the Q468A mutation in domain IV perturbs domain I conformation as well as interaction with ATP from a distance of ~ 76 Å, E699A in domain V perturbs domain I conformation as well as interaction with DNA from a distance of ~ 70 Å, and L759A in domain V perturbs domain IV conformation as well as interaction with DNA from a distance of ~ 70 Å (Table 1, Figures S3 and S5). This finding implies that SCA can also reveal functionally relevant residues that are not in contact with other members of the network (this could be especially relevant for large multidomain proteins like MutS). It also supports the view that allostery need not rely on pathways of physically interacting residues and could emerge from multiple determinants of changes in the protein free energy landscape that stabilize one set of conformations over another.^{84,90} As noted earlier, MD studies of *E. coli*, *T. aquaticus* MutS, and human MSH2-MSH6 have revealed long-range correlated motions across the protein, implicating dynamics-based mechanisms of allostery.^{36–39} We therefore asked whether sector residues that are not part of contiguous pathways are involved in these conformational dynamics. Figure 5 shows difference maps calculated from the motional correlation matrices of Q468A, E699A, and L759A versus wild type MutS-(ATP)₂-DNA(+T) complexes (Figure 2). All three mutants exhibit changes in correlated motions relative to the wild type protein. The changes most closely related to MutS activities, between the mismatch binding domain of S1 subunit and the ATPase domains of S1 and S2 subunits, are highlighted by black rectangles. The differences in motional coupling could reflect contributions of sector residues outside of a physically connected network to shifts in the MutS conformational ensemble that mediate allosteric communication between the DNA binding and ATPase sites.

CONCLUSION

To the best of our knowledge, MutS is the largest protein to be examined by SCA (~ 180 kDa dimer), a statistical method that identifies evolutionarily conserved amino acid networks from sequence alignments of protein families. The analysis yielded a set of residues bearing the hallmarks of previously identified SCA sectors in smaller proteins, such as serine protease and

Sector mutations alter correlated motions between mismatch binding and ATPase sites

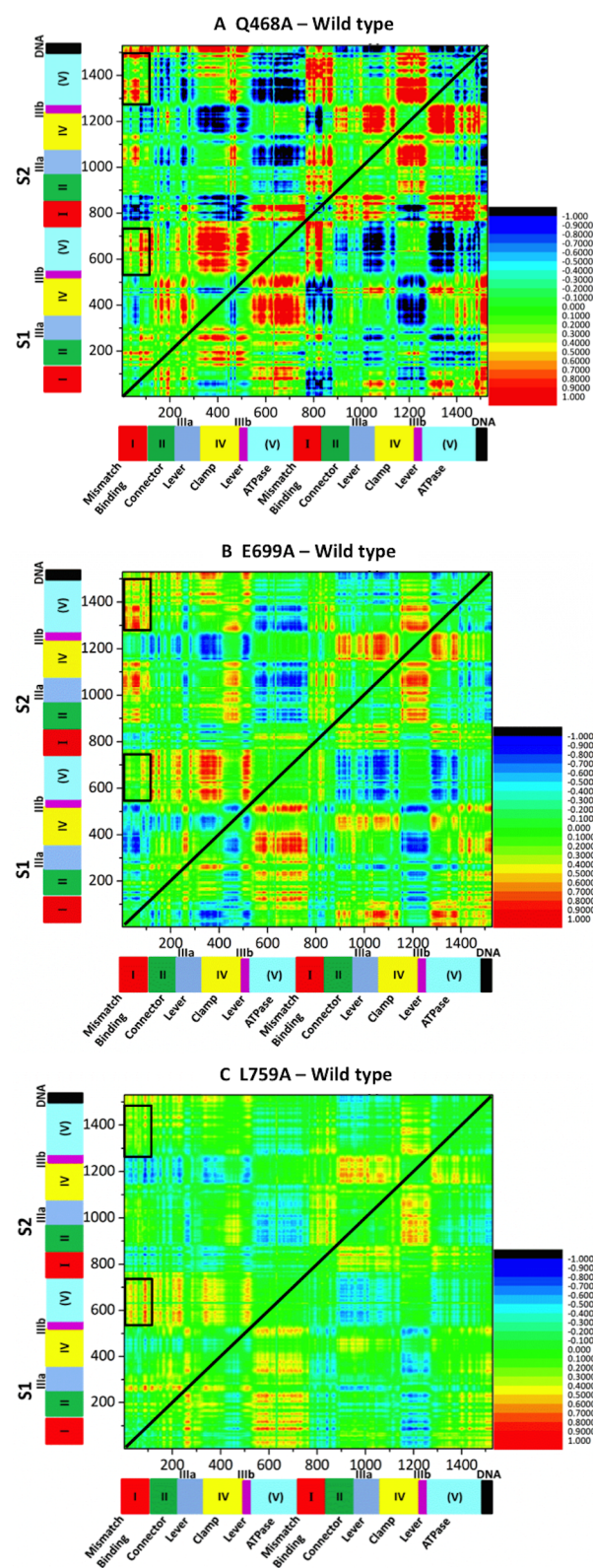


Figure 5. Difference plots of correlated motions of sector residues Q468A, E699A, and L759A relative to wild type. These SCA sector residues, among others, lie outside the contiguous network defined by a 6.8 Å cutoff (Figure 2C). Changes in correlated atomic fluctuations associated with alanine mutations of these residues are shown in a symmetric matrix for (A) Q468A, (B) E699A, and (C) Y759A. Black

Figure 5. continued

rectangles highlight the differences in coupling between the mismatch-binding domain (I) in the S1 subunit and the ATPase domains (V) in both S1 and S2 subunits.

PDZ domains.^{41,42} These residues are sparse (~20% of total), spatially distributed, and the majority are part of a contiguous network defined by van der Waals contacts that spans all MutS domains and connects the two active sites. How to assess the potential evolutionary, structural, and/or functional significance of this coevolved network is an open question in the field. In this study we demonstrate that combining SCA with MD analysis of alanine mutants can provide a predictive model for the structure–function roles of individual amino acids and networks located beyond the active sites in MutS protein. This finding is especially significant in case of large proteins where open-ended experimental mutagenesis is challenging, or for activities that are difficult to resolve experimentally, such as allosteric communication. The MD trajectory data enabled empirical testing of specific sector residue contributions to MutS structure–function, especially allostery, including conformational dynamics, hydrogen bonding with mismatched DNA and ATP ligands in the active sites, and free energy of ligand binding. Most notable among the findings was the disruptive impact of mutating distant sector residues on critical contacts between MutS and DNA/ATP, and that residues both within and outside of contiguous pathways contribute to allostery. Thus, this study provides testable hypotheses for previously unknown functions of widely distributed amino acids in MutS, including their role in allosteric communication. We suggest that it can serve as a model for other large systems to explore the functional relevance of the evolutionarily conserved protein architecture revealed by SCA.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcc.6b11976.

Further details on determination of SCA sector; contact distances between critical residue pairs in network, domain-wise RMSD; hydrogen bonding for sector and nonsector mutants (other than R172A, I553A, and Y167A shown in Figures 3 and 4); and domain-wise RMSD and hydrogen bonding for R172A, I553A, and Y167A mutants in ADP-bound forms of MutS (PDF)

Comparison of key Taq MutS sector positions with human MutS α residues (XLSX)

MD-calculated domain-wise RMSD values for all 31 mutants (XLSX)

Movie of wildtype MutS R76 H bond with DNA (MPG)

Movie of wildtype MutS K589 H bond with ATP (MPG)

Movie of R172A MutS R76 H bond with DNA (MPG)

Movie of R172A MutS K589 H bond with ATP (MPG)

Movie of I553A MutS R76 H bond with DNA (MPG)

Movie of I553A MutS K589 H bond with ATP (MPG)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: dbeveridge@wesleyan.edu; Phone: (860) 685-2575.

ORCID 

Manju M. Hingorani: 0000-0003-3133-4495

David L. Beveridge: 0000-0003-2150-899X

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank Ishita Mukerji and Sudipta Lahiri for insightful discussions and Henk Meij for technical support. This work was supported by NIH T32 GM008271 (molecular biophysics training grant), NSF grants CNS-0619508 and CNS-0959856 (Wesleyan University computational resources), and NIH grant R15 GM114743 (M.M.H.).

■ REFERENCES

- (1) Modrich, P. Mechanisms in *E. Coli* and Human Mismatch Repair (Nobel Lecture). *Angew. Chem., Int. Ed.* **2016**, *55*, 8490–8501.
- (2) Jiricny, J. Postreplicative Mismatch Repair. *Cold Spring Harbor Perspect. Biol.* **2013**, *5*, a012633.
- (3) Groothuizen, F. S.; Sixma, T. K. The Conserved Molecular Machinery in DNA Mismatch Repair Enzyme Structures. *DNA Repair* **2016**, *38*, 14–23.
- (4) Friedhoff, P.; Li, P.; Gotthardt, J. Protein-Protein Interactions in DNA Mismatch Repair. *DNA Repair* **2016**, *38*, 50–57.
- (5) Peltomäki, P. Update on Lynch Syndrome Genomics. *Fam. Cancer* **2016**, *15*, 385–393.
- (6) Heinen, C. D. Mismatch Repair Defects and Lynch Syndrome: The Role of the Basic Scientist in the Battle against Cancer. *DNA Repair* **2016**, *38*, 127–134.
- (7) Obmolova, G.; Ban, C.; Hsieh, P.; Yang, W. Crystal Structures of Mismatch Repair Protein MutS and Its Complex with a Substrate DNA. *Nature* **2000**, *407*, 703–710.
- (8) Alani, E.; Lee, J. Y.; Schofield, M. J.; Kijas, A. W.; Hsieh, P.; Yang, W. 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.* **2003**, *278*, 16088–16094.
- (9) Lamers, M. H.; Perrakis, A.; Enzlin, J. H.; Winterwerp, H. H.; de Wind, N.; Sixma, T. K. The Crystal Structure of DNA Mismatch Repair Protein MutS Binding to a G X T Mismatch. *Nature* **2000**, *407*, 711–717.
- (10) Natrajan, G.; Lamers, M. H.; Enzlin, J. H.; Winterwerp, H. H. K.; Perrakis, A.; Sixma, T. K. Structures of *Escherichia Coli* DNA Mismatch Repair Enzyme MutS in Complex with Different Mismatches: A Common Recognition Mode for Diverse Substrates. *Nucleic Acids Res.* **2003**, *31*, 4814–4821.
- (11) Groothuizen, F. S.; Winkler, I.; Cristóvão, M.; Fish, A.; Winterwerp, H. H. K.; Reumer, A.; Marx, A. D.; Hermans, N.; Nicholls, R. A.; Murshudov, G. N.; et al. MutS/MutL Crystal Structure Reveals That the MutS Sliding Clamp Loads MutL onto DNA. *eLife* **2015**, *4*, e06744.
- (12) Warren, J. J.; Pohlhaus, T. J.; Changela, A.; Iyer, R. R.; Modrich, P. L.; Beese, L. S. Structure of the Human MutS α DNA Lesion Recognition Complex. *Mol. Cell* **2007**, *26*, 579–592.
- (13) Gupta, S.; Gellert, M.; Yang, W. Mechanism of Mismatch Recognition Revealed by Human MutS β Bound to Unpaired DNA Loops. *Nat. Struct. Mol. Biol.* **2011**, *19*, 72–78.
- (14) Hingorani, M. M. Mismatch Binding, ADP-ATP Exchange and Intramolecular Signaling during Mismatch Repair. *DNA Repair* **2016**, *38*, 24–31.
- (15) Antony, E.; Hingorani, M. M. Asymmetric ATP Binding and Hydrolysis Activity of the *Thermus Aquaticus* MutS Dimer Is Key to Modulation of Its Interactions with Mismatched DNA. *Biochemistry* **2004**, *43*, 13115–13128.
- (16) Qiu, R.; DeRocco, V. C.; Harris, C.; Sharma, A.; Hingorani, M. M.; Erie, D. A.; Weninger, K. R. Large Conformational Changes in MutS during DNA Scanning, Mismatch Recognition and Repair Signalling. *EMBO J.* **2012**, *31*, 2528–2540.

- (17) Jacobs-Palmer, E.; Hingorani, M. M. The Effects of Nucleotides on MutS-DNA Binding Kinetics Clarify the Role of MutS ATPase Activity in Mismatch Repair. *J. Mol. Biol.* **2007**, *366*, 1087–1098.
- (18) Sharma, A.; Doucette, C.; Biro, F. N.; Hingorani, M. M. Slow Conformational Changes in MutS and DNA Direct Ordered Transitions between Mismatch Search, Recognition and Signaling of DNA Repair. *J. Mol. Biol.* **2013**, *425*, 4192–4205.
- (19) Mazur, D. J.; Mendillo, M. L.; Kolodner, R. D. Inhibition of Msh6 ATPase Activity by Mismatched DNA Induces a Msh2(ATP)-Msh6(ATP) State Capable of Hydrolysis-Independent Movement along DNA. *Mol. Cell* **2006**, *22*, 39–49.
- (20) Acharya, S.; Foster, P. L.; Brooks, P.; Fishel, R. The Coordinated Functions of the E. Coli MutS and MutL Proteins in Mismatch Repair. *Mol. Cell* **2003**, *12*, 233–246.
- (21) Selmane, T.; Schofield, M. J.; Nayak, S.; Du, C.; Hsieh, P. Formation of a DNA Mismatch Repair Complex Mediated by ATP. *J. Mol. Biol.* **2003**, *334*, 949–965.
- (22) Mendillo, M. L.; Mazur, D. J.; Kolodner, R. D. Analysis of the Interaction between the Saccharomyces Cerevisiae MSH2-MSH6 and MLH1-PMS1 Complexes with DNA Using a Reversible DNA End-Blocking System. *J. Biol. Chem.* **2005**, *280*, 22245–22257.
- (23) Winkler, I.; Marx, A. D.; Lariviere, D.; Heinze, R. J.; Cristovao, M.; Reumer, A.; Curth, U.; Sixma, T. K.; Friedhoff, P. Chemical Trapping of the Dynamic MutS-MutL Complex Formed in DNA Mismatch Repair in Escherichia Coli. *J. Biol. Chem.* **2011**, *286*, 17326–17337.
- (24) Antony, E.; Hingorani, M. M. Mismatch Recognition-Coupled Stabilization of Msh2-Msh6 in an ATP-Bound State at the Initiation of DNA Repair. *Biochemistry* **2003**, *42*, 7682–7693.
- (25) Antony, E.; Khubchandani, S.; Chen, S.; Hingorani, M. M. 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* **2006**, *5*, 153–162.
- (26) Gradia, S.; Acharya, S.; Fishel, R. The Human Mismatch Recognition Complex hMSH2-hMSH6 Functions as a Novel Molecular Switch. *Cell* **1997**, *91*, 995–1005.
- (27) Heinen, C. D.; Cyr, J. L.; Cook, C.; Punja, N.; Sakato, M.; Forties, R. A.; Lopez, J. M.; Hingorani, M. M.; Fishel, R. Human MSH2 (hMSH2) Protein Controls ATP Processing by hMSH2-hMSH6. *J. Biol. Chem.* **2011**, *286*, 40287–40295.
- (28) Jeong, C.; Cho, W.-K.; Song, K.-M.; Cook, C.; Yoon, T.-Y.; Ban, C.; Fishel, R.; Lee, J.-B. MutS Switches between Two Fundamentally Distinct Clamps during Mismatch Repair. *Nat. Struct. Mol. Biol.* **2011**, *18*, 379–385.
- (29) Gorman, J.; Chowdhury, A.; Surtees, J. A.; Shimada, J.; Reichman, D. R.; Alani, E.; Greene, E. C. Dynamic Basis for One-Dimensional DNA Scanning by the Mismatch Repair Complex Msh2-Msh6. *Mol. Cell* **2007**, *28*, 359–370.
- (30) Gorman, J.; Wang, F.; Redding, S.; Plys, A. J.; Fazio, T.; Wind, S.; Alani, E. E.; Greene, E. C. Single-Molecule Imaging Reveals Target-Search Mechanisms during DNA Mismatch Repair. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, E3074–E3083.
- (31) Qiu, R.; Sakato, M.; Sacho, E. J.; Wilkins, H.; Zhang, X.; Modrich, P.; Hingorani, M. M.; Erie, D. A.; Wenginger, K. R. MutL Traps MutS at a DNA Mismatch. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 10914–10919.
- (32) Cho, W. K.; Jeong, C.; Kim, D.; Chang, M.; Song, K. M.; Hanne, J.; Ban, C.; Fishel, R.; Lee, J. B. ATP Alters the Diffusion Mechanics of MutS on Mismatched DNA. *Structure* **2012**, *20*, 1264–1274.
- (33) Junop, M. S.; Obmolova, G.; Rausch, K.; Hsieh, P.; Yang, W. Composite Active Site of an ABC ATPase: MutS Uses ATP to Verify Mismatch Recognition and Authorize DNA Repair. *Mol. Cell* **2001**, *7*, 1–12.
- (34) Biswas, I.; Obmolova, G.; Takahashi, M.; Herr, A.; Newman, M. A.; Yang, W.; Hsieh, P. Disruption of the Helix-U-Turn-Helix Motif of MutS Protein: Loss of Subunit Dimerization, Mismatch Binding and ATP Hydrolysis. *J. Mol. Biol.* **2001**, *305*, 805–816.
- (35) Mendillo, M. L.; Hargreaves, V. V.; Jamison, J. W.; Mo, A. O.; Li, S.; Putnam, C. D.; Woods, V. L.; Kolodner, R. D. A Conserved MutS Homolog Connector Domain Interface Interacts with MutL Homologs. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 22223–22228.
- (36) Mukherjee, S.; Feig, M. Conformational Change in MSH2-MSH6 upon Binding DNA Coupled to ATPase Activity. *Biophys. J.* **2009**, *96*, L63–L65.
- (37) Mukherjee, S.; Law, S. M.; Feig, M. Deciphering the Mismatch Recognition Cycle in MutS and MSH2-MSH6 Using Normal-Mode Analysis. *Biophys. J.* **2009**, *96*, 1707–1720.
- (38) Pieniasek, S. N.; Hingorani, M. M.; Beveridge, D. L. Dynamical Allosterism in the Mechanism of Action of DNA Mismatch Repair Protein MutS. *Biophys. J.* **2011**, *101*, 1730–1739.
- (39) Negureanu, L.; Salsbury, F. R. Insights into Protein–DNA Interactions, Stability and Allosteric Communications: A Computational Study of MutS-DNA Recognition Complexes. *J. Biomol. Struct. Dyn.* **2012**, *29*, 757–776.
- (40) Wang, B.; Francis, J.; Sharma, M.; Law, S. M.; Predeus, A. V.; Feig, M. Long-Range Signaling in MutS and MSH Homologs via Switching of Dynamic Communication Pathways. *PLoS Comput. Biol.* **2016**, *12*, e1005159.
- (41) Halabi, N.; Rivoire, O.; Leibler, S.; Ranganathan, R. Protein Sectors: Evolutionary Units of Three-Dimensional Structure. *Cell* **2009**, *138*, 774–786.
- (42) McLaughlin, R. N., Jr; Poelwijk, F. J.; Raman, A.; Gosal, W. S.; Ranganathan, R. The Spatial Architecture of Protein Function and Adaptation. *Nature* **2012**, *491*, 138–142.
- (43) Reynolds, K. A.; McLaughlin, R. N.; Ranganathan, R. Hot Spots for Allosteric Regulation on Protein Surfaces. *Cell* **2011**, *147*, 1564–1575.
- (44) Smock, R. G.; Rivoire, O.; Russ, W. P.; Swain, J. F.; Leibler, S.; Ranganathan, R.; Gierasch, L. M. An Interdomain Sector Mediating Allosteric Regulation in Hsp70 Molecular Chaperones. *Mol. Syst. Biol.* **2010**, *6*, 414.
- (45) Novinec, M.; Korenč, M.; Cafilisch, A.; Ranganathan, R.; Lenarčič, B.; Baici, A. A Novel Allosteric Mechanism in the Cysteine Peptidase Cathepsin K Discovered by Computational Methods. *Nat. Commun.* **2014**, *5*, 3287.
- (46) Lin, Z.; Nei, M.; Ma, H. The Origins and Early Evolution of DNA Mismatch Repair Genes-Multiple Horizontal Gene Transfers and Co-Evolution. *Nucleic Acids Res.* **2007**, *35*, 7591–7603.
- (47) Marchler-Bauer, A.; Lu, S.; Anderson, J. B.; Chitsaz, F.; Derbyshire, M. K.; DeWeese-Scott, C.; Fong, J. H.; Geer, L. Y.; Geer, R. C.; Gonzales, N. R.; et al. CDD: A Conserved Domain Database for the Functional Annotation of Proteins. *Nucleic Acids Res.* **2011**, *39*, D225–D229.
- (48) Sievers, F.; Higgins, D. G. Clustal Omega, Accurate Alignment of Very Large Numbers of Sequences. In *Methods Mol. Biol.*; SpringerLink, 2013; Vol. 1079, pp 105–116.
- (49) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539.
- (50) Dickson, R. J.; Gloor, G. B. Protein Sequence Alignment Analysis by Local Covariation: Coevolution Statistics Detect Benchmark Alignment Errors. *PLoS One* **2012**, *7*, e37645.
- (51) Süel, G. M.; Lockless, S. W.; Wall, M. A.; Ranganathan, R. Evolutionarily Conserved Networks of Residues Mediate Allosteric Communication in Proteins. *Nat. Struct. Biol.* **2003**, *10*, 59–69.
- (52) Reynolds, K. A.; Russ, W. P.; Socolich, M.; Ranganathan, R. Evolution-Based Design of Proteins. In *Methods in Enzymology*; Elsevier Inc., 2013; Vol. 523, pp 213–235.
- (53) Lockless, S. W.; Ranganathan, R. Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families. *Science* **1999**, *286*, 295–299.
- (54) Shea, J. E.; Onuchic, J. N.; Brooks, C. L., III Probing the Folding Free Energy Landscape of the Src-SH3 Protein Domain. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 16064–16068.
- (55) Noel, J. K.; Whitford, P. C.; Onuchic, J. N. The Shadow Map: A General Contact Definition for Capturing the Dynamics of

Biomolecular Folding and Function. *J. Phys. Chem. B* **2012**, *116*, 8692–8702.

(56) Koga, N.; Takada, S. Roles of Native Topology and Chain-Length Scaling in Protein Folding: A Simulation Study with a Go-like Model. *J. Mol. Biol.* **2001**, *313*, 171–180.

(57) Baur, M. et al. *Visone Software for Visual Social Network Analysis; In Graph Drawing. GD 2001. Lecture Notes in Computer Science*; Mutzel, P., Jünger, M., Leipert, S., Eds.; Springer Link, 2002; Vol. 2265, pp 463–464.

(58) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.

(59) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(60) Wang, J.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.* **2000**, *21*, 1049–1074.

(61) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R. W.; R, C.; Zhang, W.; Merz, K. M.; et al. *AMBER 2014*; University of California: San Francisco, CA, 2014.

(62) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of Alpha/gamma Conformers. *Biophys. J.* **2007**, *92*, 3817–3829.

(63) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–9041.

(64) Meagher, K. L.; Redman, L. T.; Carlson, H. A. Development of Polyphosphate Parameters for Use with the AMBER Force Field. *J. Comput. Chem.* **2003**, *24*, 1016–1025.

(65) Jorgensen, W. L.; Jenson, C. Temperature Dependence of TIP3P, SPC, and TIP4P Water from NPT Monte Carlo Simulations: Seeking Temperatures of Maximum Density. *J. Comput. Chem.* **1998**, *19*, 1179–1186.

(66) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(67) Crowley, M. F.; Darden, T. A.; Cheatham, T. E., III; Deerfield, D. W., II Adventures in Improving the Scaling and Accuracy of a Parallel Molecular Dynamics Program. *J. Supercomput.* **1997**, *11*, 255–278.

(68) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(69) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(70) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(71) Götz, W. A.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with Amber - Part I: Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542–1555.

(72) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. I. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542–1555.

(73) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without Compromise - A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2013**, *184*, 374–380.

(74) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.

(75) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314–3321.

(76) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.

(77) *The PyMOL Molecular Graphics System*, Version 1.7; Schrödinger, LLC.: New York, NY, 2010.

(78) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

(79) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; et al. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33*, 889–897.

(80) Teşileanu, T.; Colwell, L. J.; Leibler, S. Protein Sectors: Statistical Coupling Analysis versus Conservation. *PLoS Comput. Biol.* **2015**, *11*, 1–20.

(81) Schofield, M. J.; Brownwell, F. E.; Nayak, S.; Du, C.; Kool, E. T.; Hsieh, P. The Phe-X-Glu DNA Binding Motif of MutS: The Role of Hydrogen Bonding in Mismatch Recognition. *J. Biol. Chem.* **2001**, *276*, 45505–45508.

(82) Negureanu, L.; Salisbury, F. R., Jr. Non-Specificity and Synergy at the Binding Site of the Carboplatin-Induced DNA Adduct via Molecular Dynamics Simulations of the MutSalph-DNA Recognition Complex. *J. Biomol. Struct. Dyn.* **2014**, *32*, 969–992.

(83) Kumar, S.; Ma, B.; Tsai, C. J.; Sinha, N.; Nussinov, R. Folding and Binding Cascades: Dynamic Landscapes and Population Shifts. *Protein Sci.* **2000**, *9*, 10–19.

(84) Tsai, C. J.; Nussinov, R. A Unified View of “How Allostery Works. *PLoS Comput. Biol.* **2014**, *10*, e1003394.

(85) Ramos, R. M.; Moreira, I. S. Computational Alanine Scanning Mutagenesis-An Improved Methodological Approach for Protein-DNA Complexes. *J. Chem. Theory Comput.* **2013**, *9*, 4243–4256.

(86) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discovery* **2015**, *10*, 449–461.

(87) Ishida, H.; Matsumoto, A. Mechanism for Verification of Mismatched and Homoduplex DNAs by Nucleotides-Bound MutS Analyzed by Molecular Dynamics Simulations. *Proteins: Struct., Funct., Genet.* **2016**, *84*, 1287–1303.

(88) Hatley, M. E.; Lockless, S. W.; Gibson, S. K.; Gilman, A. G.; Ranganathan, R. Allosteric Determinants in Guanine Nucleotide-Binding Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 14445–14450.

(89) Shulman, A. I.; Larson, C.; Mangelsdorf, D. J.; Ranganathan, R. Structural Determinants of Allosteric Ligand Activation in RXR Heterodimers. *Cell* **2004**, *116*, 417–429.

(90) Hilser, V. J.; Wrabl, J. O.; Motlagh, H. N. Structural and Energetic Basis of Allostery. *Annu. Rev. Biophys.* **2012**, *41*, 585–609.