Mechanism of mismatch recognition revealed by human MutSβ bound to unpaired DNA loops

Shikha Gupta, Martin Gellert & Wei Yang

DNA mismatch repair corrects replication errors, thus reducing mutation rates and microsatellite instability. Genetic defects in this pathway cause Lynch syndrome and various cancers in humans. Binding of a mispaired or unpaired base by bacterial MutS and eukaryotic MutSα is well characterized. We report here crystal structures of human MutSβ in complex with DNA containing insertion-deletion loops (IDL) of two, three, four or six unpaired nucleotides. In contrast to eukaryotic MutSα and bacterial MutSβ, which bind the base of a mismatched nucleotide, MutSβ binds three phosphates in an IDL. DNA is severely bent at the IDL; unpaired bases are flipped out into the major groove and partially exposed to solvent. A normal downstream base pair can become unpaired; a single unpaired base can thereby be converted to an IDL of two nucleotides and recognized by MutSβ. The C-terminal dimerization domains form an integral part of the MutS structure and coordinate asymmetrical ATP hydrolysis by Msh2 and Msh3 with mismatch binding to signal for repair.

Nucleotide misincorporation or strand slippage at repetitive sequences during replication results in mispaired or unpaired DNA bases. Mismatched bases are recognized by MutS, which in the presence of ATP recruits MutL to initiate the repair process. Eukaryotes have two MutS homologs, α and β. MutSα, a heterodimer of MutS homologs Msh2 and Msh6, recognizes a base mispair or one or two unpaired bases, like homodimeric bacterial MutS. By contrast, MutSβ, a heterodimer of Msh2 and Msh3, recognizes insertion-deletion loops (IDLs) of 1–15 nucleotides, as well as DNA with a 3′ single-stranded overhang. Inactivation of MutS or MutL by mutation, or reduced expression of human MutLα due to promoter hypermethylation, leads to increased mutation rates and microsatellite instability. In humans, such defects are correlated with susceptibility to hereditary nonpolyposis colorectal cancer (HNPCC), also known as Lynch syndrome. Loss of MSH3 in tumor cells is correlated with increased microsatellite instability. In mice, loss of Msh3 knockout out develops cancers only late in life, but a double knockout of Msh3 and Msh6 renders mice far more susceptible to cancer than the single knockout of either gene.

Mechanistically, MutSα functions similarly to bacterial MutS, by inserting the mismatch-binding domain (MBD) of Msh6 into the minor groove at the mismatch site and pulling a mispaired or unpaired base toward it to form π-stacking with a conserved phenylalanine. As a result, normal base stacking is disrupted and the DNA is bent 45–60°. Although MutSβ recognizes one or two unpaired nucleotides, as does MutSα, and Msh3 is homologous to MutS and Msh6, Msh3 lacks this conserved phenylalanine. Mutation of the phenylalanine in Msh6 or replacement of the corresponding residue in Msh3 (lysine) by phenylalanine diminishes the ability of these MutS homologs to bind all forms of mismatched DNA. Although Msh2 is present in both MutSα and MutSβ, the MBD of Msh2 is required for mismatch binding only by MutSβ, not by MutSα. Extensive mutagenic studies have revealed residues critical for IDL recognition and led to the conclusion that MutSβ recognizes mismatches very differently from MutS and MutSα.

MutSβ is involved in more than mismatch repair. In Saccharomyces cerevisiae, MutSβ participates in double-strand break repair and binds a 3′ overhang as well as a 3′ flap formed during double-strand break repair by single-strand annealing. In mammals, MutSβ is required for the mutagenic expansion of trinucleotide repeats and may thus be a culprit in myotonic dystrophy, fragile X syndrome and Huntington’s disease. Trinucleotide repeats can form unique IDLs and be bound by MutSβ. Although short IDLs in these repeats are repaired efficiently by MutSβ, long or clustered IDLs are refractory to repair, and attempted repair may lead to expansion.

To clarify the functions of MutSβ, we report here crystal structures of mismatched DNA in complex with MutSβ that include the previously missing C-terminal dimerization domains present in all MutS functional homologs. The molecular mechanisms of recognition of a variety of IDLs and of the ATP-mediated signaling for later steps of mismatch repair are presented.

RESULTS

Structure of MutSβ–DNA complexes

A form of human MutSβ consisting of full-length MSH2 and trimmed MSH3 (211–1,125 residues, abbreviated as MutSβ hereafter) was generated for structural studies (see Methods). It retains the ATPase activity and binding affinity for IDLs of the full-length species.
MutSβ binds mismatched DNA with sub-nanomolar Kₜₜ values, which are much lower than the values previously measured in the presence of competitor DNAs. MutSβ was cocryrstallized with IDLs of two, three, four or six unpaired bases flanked by duplexes that are 10 to 12 base pairs (bp), in three different crystal lattices (Table 1). These structures, named according to the IDL size as Loop2, Loop3, Loop4 and Loop6, were determined and refined (Methods, Supplementary Fig. 2). Only 1–2 nucleotides at the DNA ends, some residues at the extreme N- and C-termini, and a few solvent-exposed internal loops are disordered. Even at the low resolution of 4.3 Å, the DNA and protein domains are well defined in Loop6. Except for the DNA-binding domains, the four MutSβ structures are superimposable.

Table 1 Data collection and refinement statistics

| Loop2 | Loop3 | Loop4 | Loop6 |
|-------|-------|-------|-------|
| Data collection | 23-ID | In-house | 22-BM | 23-ID |
| Space group | P1 | P1 | P2₁,2₁,2₁ | P2₁,2₁,2₁ |
| Cell dimensions | | | |
| a, b, c (Å) | 66.3, 91.1, 95.6 | 67.1, 91.5, 95.6 | 105.8, 116.1, 180.0 | 103.7, 154.6, 161.6 |
| α, β, γ (°) | 67.8, 87.0, 73.4 | 67.9, 86.5, 72.9 | 90, 90, 90 | 90, 90, 90 |
| Resolution (Å) | 50.0–2.9 (2.95–2.90) | 30.0–2.7 (2.75–2.70) | 32.0–3.09 (3.14–3.09) | 50.0–4.3 (4.37–4.30) |
| Rmerge | 4.6 (24.8) | 4.7 (33.2) | 10.0 (55.4) | 10.2 (68.7) |
| Rmerge | 1/α² | 14.8 (1.4) | 5.0 (1.1) | 12.2 (1.8) |
| Completeness (%) | 94.4 (64.5) | 91.8 (43.3) | 99.0 (98.5) | 96.0 (92.8) |
| Redundancy | 1.9 (1.5) | 1.9 (1.5) | 2.4 (2.4) | 4.0 (4.0) |

Refinement

| | Resolution (Å) | No. reflections | Rwork | Rfree | Reflection number used in Rfree calculation |
|--------------------------|----------------|----------------|-------|-------|------------------------------------------|
| | 44.0–2.9 | 39,845 (1,125) | 19.4 | 19.4, 27.4 | 4.6 (24.8) |
| | 27.0–2.7 | 49,511 (1,315) | 21.0 | 21.0, 27.3 | 4.7 (33.2) |
| | 32.0–3.09 | 38,509 (1,919) | 22.9 | 22.9, 28.7 | 10.0 (55.4) |
| | 49.1–4.3 | 16,879 (999) | 23.9 | 23.9, 28.5 | 10.2 (68.7) |

Redundancy

| | | | |
|--------------------------|----------------|-------|-------|
| Protein, DNA | 13,730, 916 | 13,270, 999 | 13,777, 1,115 | 13,737, 980 |
| Ligand | 27 | 27 | 27 | 27 |
| Water | 15 | 20 | 4 | 0 |
| B-factors | | | |
| Protein, DNA | 84.2, 138.7 | 66.3, 114.8 | 66.8, 136.6 | 275.2, 408.0 |
| Ligand | 89.6 | 78.0 | 86.1 | 268.35 |
| Water | 55.2 | 39.3 | 23.6 | – |
| R.m.s. deviations | | | |
| Bond lengths (Å) | 0.003 | 0.003 | 0.003 | 0.013 |
| Bond angles (°) | 0.68 | 0.67 | 0.563 | 0.994 |

*Data in the highest resolution shell is shown in parenthesis. *Reflection number used in Rfree calculation is shown in parenthesis.

The MutSβ structures reveal two new features. First, the dimerization domains (DMDs) at the C-termini, which were excluded or disordered in all previously known MutS structures, are observed atop the rest of MutS for the first time (Fig. 1a). They strengthen the heterodimer and establish the asymmetry of the nucleotide binding by MSH2 and MSH3 (see later). Second, the degree of DNA bending and the mechanism of mismatch recognition by MutSβ differ appreciably from those of bacterial MutS and MutSβ. DNA is severely bent at an IDL because the extra nucleotides in one strand disrupt normal base stacking. The bending angle increases from 90° in Loop2 to 120° in Loop6 (Fig. 1b). The bases in the IDL are rotated away from MutSβ toward the solvent-exposed major groove. They often stack with one another and roughly maintain their 3.4 Å spacing.

Recognition of IDLs

The IDLs interact chiefly with domain I of MSH3 (MBD) and partially with domain I of MSH2. Although the structure of the MBD is identical among MutS homologs (Fig. 2e), and MSH3 inserts its a r t i c l e s

© 2012 Nature America, Inc. All rights reserved.

NATURE STRUCTURAL & MOLECULAR BIOLOGY VOLUME 19 NUMBER 1 JANUARY 2012 73
When an IDL is longer than three nucleotides, Phe42 of MSH2 (an equivalent to the π-stacking phenylalanine in Msh6 and MutS) forms a π-stack with the fourth unpaired base. Asp41, Lys65, Val79 and Ser81 of MSH2 interact with the plus strand (Fig. 3d and Supplementary Fig. 4). These interactions are absent in the MutSα–DNA complex structures and explain why domain I of MSH2 is required for MutSβ to bind an IDL(22,23). Domains I of MSH2 and MSH3 interact directly (Fig. 2d) and maintain the same interactions in all four MutSβ structures (Figs. 1b and 3d). Different IDLs are fitted into this framework by bending at different angles. Domains IV of MSH2 and MSH3 barely contact one another and move independently (Fig. 2a–c), and this

a plus strand that has additional nucleotides and a complementary minus strand. Tyr245 is inserted between the adjacent nucleotides on the minus strand opposite the IDL and bends the strand sharply (Fig. 3c,d). It interacts with the base pair immediately upstream of the IDL and forms a π-stack with the downstream minus strand in the Loop2 and Loop3 structures. The IDL on the plus strand bends more smoothly. Lys246 along with Ser275, which replaces a conserved glycine in Msh6 and MutS, form numerous interactions with the first phosphates in an IDL inside the widened minor groove. Through the tyrosine-lysine pair, MutSβ binds an IDL with strand and orientation specificity. Mutations of the tyrosine-lysine pair or the Ser275 equivalent in Saccharomyces cerevisiae (yeast) MSH3 cause a mutator phenotype and microsatellite instability(2,22).

The main chain atoms of Ile276 and Pro277 and side chains of His279 and Arg280 (unique among Msh3 homologs) fix the second and third phosphates of an IDL (or the downstream nucleotide in Loop2). Although Pro277 is highly conserved among MutS homologs, it contacts and stabilizes the π-stacking phenylalanine instead in MutS and Msh6. Notably, mutations of the equivalent of Pro277 in yeast MSH3 and MSH6 lead to very different levels of repair deficiency(22). His284 of Msh3, which replaces tyrosine in Msh6 and MutS (Supplementary Fig. 3) and is critical for IDL repair(21), makes favorable interactions with the third phosphate of an IDL and avoids steric clashes (Fig. 3). The distorted IDL is further stabilized through phosphate backbone interactions by Ile263, Tyr264 and His266 on a β strand outside of the minor groove (Supplementary Fig. 4). As a result, the positions of the first three phosphates of each IDL are the same in all four structures.

Figure 1 Overall structures of MutSβ–DNA complexes. (a) Orthogonal views of Loop3 structure in ribbon diagrams, with MSH2 in green and MSH3 in blue. The DNA is shown in a space-filling model with the backbone in red, bases in light pink, and the unpaired nucleotides in yellow and orange. ADP bound to MSH2 is shown in purple sticks. (b) Side views of DNA-binding domains and DNA in Loop2, Loop4 and Loop6 structures. Each unpaired CA dinucleotide repeat is shown in yellow and orange. Domain I of MSH3 is the MBD. MSH2 and MSH3 subunits are indicated by circled numbers 2 and 3, respectively. Domains I–V and dimerization domains are indicated.

Figure 2 Comparison of MutSα and MutSβ proteins. (a) Ribbon diagram of MSH2 from MutSβ. Domains I, II, III, IV, V and the DMDs are shown in blue, green, yellow, orange, pink and red, respectively. MSH2 of MutSα is superimposed and shown in gray. Domain interfaces between I and II and between II, III and V are highlighted in magenta. (b) Ribbon diagram of MSH3 in the same orientation and same color codes. Domains I (MBD) and IV (clamp) interact as indicated by the dotted oval. (c) Superposition of the mismatch-binding subunits in MutSα. MutSβ, E. coli and Taq MutSα in the same orientation as in a and b. Except for domain IV, they superimpose very well. (d) A ribbon diagram of the interface between domains I in MutSβ. Protein residues in all figures are labeled in one-letter code for clarity. (e) On the left is a ribbon diagram of MSH3 MBD decorated by residues conserved among MutS homologs (shown as yellow-blue-red stick-and-ball models) and residues unique among MSH3 homologs (pink-blue-red stick-and-ball models). On the right is the superposition of MSH3 (blue) and MSH6 (gray) MBDs. The r.m.s. deviation between them is 0.7 Å over 87 pairs of Cα atoms.
Figure 3 IDL recognition by MutSβ. (a) A closeup comparison of MSH3–IDL interaction and Taq MutS with a single unpaired base (AT). (b) DNA-binding domains and DNA in Loop4. Domains I and IV of MSH2 are shown in green and yellow, and MBD and clamp domains of MSH3 in blue and orange, respectively. (c) Diagram of the protein-DNA interactions using the same color scheme as in b. (d) Space-filling model of four IDLs and their interaction with domain I of MSH2 (green) and MBD of MSH3 (blue). The base pairs surrounding the IDL are shown in light (upstream) and dark (downstream) pink. For Loop2 and Loop4, a back view looking into the minor groove is also shown.

The C-terminal dimerization domain

The C-terminal end of MSH3 forms two α helices linked by a tight turn, and MSH2 has a third short helix in addition (Fig. 4a,b). They dimerize predominantly by hydrophobic interactions and bury 1.195 Å² of surface area. The five-helix bundle is further strengthened by salt bridges within each subunit and N and C capping of the MSH3 helices by MSH2. Preceding the helical appendage is a helix-turn-helix (HTH) motif, which extends from one subunit and embraces the ATPase domain of the other in all MutS homolog structures (Figs. 4 and 5a). In MSH3 the second helix of HTH and the first helix in the dimerization domain are merged into one 55 Å long helix, resulting in a three-helix unit (Fig. 4c). Based on secondary structure prediction, a similar three-helix unit also exists in MSH6. We therefore redefine the dimerization domain to include the preceding HTH. Each DMD is thus bivalent and interacts with the ATPase domain and the DMD of another subunit. In MSH2 the two halves of the DMD are linked by a disordered 15-aa loop and are juxtaposed rather than opposite to each other (Fig. 4c). The two DMDs are thus structurally different and do not follow the dyad axis that relates domains I–V of MSH2 and MSH3. The DMDs lean toward MSH3 and predominantly shield the MSH2 ATPase site (Figs. 1a and 5b).

The structure of the last 34 residues of E. coli MutS, which were deleted to facilitate crystallization of MutS–DNA complexes17, was determined as a fusion to maltose-binding protein38. It contains two α-helices and dimerizes by forming a four-helix bundle. These last two helices of E. coli MutS can be roughly superimposed on those of MSH3, but its dimerization partner is ~10Å away from MSH2’s DMD (Supplementary Fig. 6). As in MSH2, the two halves of the E. coli MutS-DMD are linked by a 20-aa flexible linker38. Without the second half of the DMD, E. coli MutS is predominantly dimeric, but with it, MutS becomes a tetramer or even an octamer (R. Ghirlando, S. Ramon-Maiques and W.Y., unpublished data). By contrast, Thermus aquaticus (Taq) MutS, which is predicted to have a short linker (4 aa) in its DMD, is a stable dimer in solution39. The short linker may prohibit the cross-linking effect of the two halves of the DMD and restrict its interacting partners within a single dimer.

In addition to stabilizing a MutS dimer, the DMD is the conduit for communication of nucleotide (ADP or ATP) binding between the two

Isomerization of mismatched DNA

MSH3 has little contact with the IDL bases, and these unpaired bases adopt different conformations in the four structures. Loop2, Loop4 and Loop6 have one, two and three CA dinucleotide repeats, respectively. In these structures the first C is flipped out and solvent exposed, and the second, third and fourth bases, if present, are stacked with one another (Fig. 3d). But in Loop3, where the IDL sequence is ACA, the second A becomes stacked with the other two unpaired bases and contacts the side chain of Lys246. In Loop6, the fifth base slips out and the sixth is stacked with the downstream duplex (Fig. 3d). Additional unpaired nucleotides may slip out between the fourth and sixth base of an IDL as observed in Loop6.

In Loop3 a normal G-C base pair immediately downstream of the IDL becomes unpaired ~50% of the time, and the G is flipped out and stacks with Phe42 of MSH2 like the fourth unpaired IDL base in Loop4 and Loop6 (Supplementary Figs. 4 and 5). Conversely, the fourth IDL base in Loop4 has a second minor conformation of stacking with the downstream duplex, like the sixth IDL base in Loop6 (Supplementary Figs. 2b, 4 and 5). Conversion of an unpaired base to a mismatched base pair was also observed with E. coli MutS37. The consistent interactions with the first three phosphates of the IDL in all four structures indicate that MutSβ prefers to bind an IDL of at least two unpaired bases. Interconversion of paired and unpaired bases suggest that an unpaired base (a substrate of MutSβ) may be converted to a two-base IDL and recognized by MutSβ. Furthermore, one unpaired base in mononucleotide repeats may be isomerized to a larger loop by creating a bulge at a distance on the opposite strand (Supplementary Fig. 5).

gives rise to an ‘accordion’ effect and accommodates different sizes of IDL with different DNA bending angles (Fig. 3d).

Recognition of IDLs is aided by substantial contacts between DNA duplexes and MSH3. A loop from residues 300 to 319 in the MSH3 MBD interacts with the upstream duplex in the major groove, covering 8 bp (Fig. 3a–c). The short helix on this loop interacts with the helical arm of domain IV (Fig. 2b) and coordinates the binding of the downstream duplex by the globular head of domain IV, which is hydrogen bonded with three consecutive phosphates on the plus strand. No interaction with the minus strand is observed in the downstream duplex. Our structures fully agree with the DNA footprinting analyses8. The interaction of MSH3 with both strands upstream of the downstream duplex by the globular head of domain IV, which accommodates different sizes of IDL with different DNA bending angles (Fig. 3d).

E. coli MutS is predominantly dimeric, but with it, MutS becomes a tetramer or even an octamer (R. Ghirlando, S. Ramon-Maiques and W.Y., unpublished data). By contrast, Thermus aquaticus (Taq) MutS, which is predicted to have a short linker (4 aa) in its DMD, is a stable dimer in solution39. The short linker may prohibit the cross-linking effect of the two halves of the DMD and restrict its interacting partners within a single dimer.

In addition to stabilizing a MutS dimer, the DMD is the conduit for communication of nucleotide (ADP or ATP) binding between the two
subunits. Each ATPase site of MutS is composite and consists of the N1, N3 and N4 nucleotide-binding motifs from one subunit, and the N2 from the partner subunit (Fig. 5a). Within each subunit, the N-terminal HTH of the DMD contacts the nucleotide-binding site directly through its conserved SYG (serine, tyrosine or phenylalanine, glycine) motif and also indirectly by interacting with the trans-acting N2 region of the other subunit (Fig. 4c,d and Supplementary Fig. 7). The MSH6 DMD alters its conformation upon ADP binding, and the movement of the MSH6 DMD is coupled with the ATPase domain of MSH2 (ref. 18) (Fig. 5a,b and Supplementary Fig. 8). We find it interesting but not surprising that the ATPase domain and DMD of MSH3, which is devoid of ADP, are superimposable with ADP-free MSH6 (Fig. 4c,d). ATP binding can alter the structure of the DMD in the same subunit and possibly cause a sliding of the DMD relative to the N2 of the other subunit (Fig. 4d). Because movement of the DMDs of the two subunits is coupled (Fig. 5b), binding of ATP by one subunit can thus alter the ATP-binding site of the other.

The communication of the two ATPase domains through the DMD is substantiated by the G1142D mutation in the SYG motif of yeast MSH6, which not only reduces ATP binding by MSH6 but also inhibits ATP binding by MSH2 (ref. 40). The roles of the DMD in dimerization and bridging of two ATP-binding sites also explain why a complete DMD is required for the cellular functions of MutS in E. coli (ref. 41). Furthermore, mutations in the MSH2 DMD have been found among HNPCC families, which highlights its importance in mismatch repair (http://www.insight-group.org/mutations/) (Supplementary Fig. 3).

An occluded MSH3 ATP-binding site
Both MutSβ and MutSα show asymmetric ATP binding and hydrolysis in the two subunits. Msh3 and Msh6 appear to have higher ATPase activity than Msh2 without DNA, but they have low affinity for ATP and greatly inhibited ATPase activity after binding to a mismatched DNA (ref. 42). Structural studies of MutS–DNA complexes confirmed that Msh6 has lower affinity for ADP than Msh2 but without revealing the reason (ref. 43). In all four MutSβ structures the ATPase site of MSH3 is devoid of nucleotide (Fig. 5a). Addition of ADP or ATP in the crystal-soaking buffer reduces the X-ray diffraction quality of MutSβ–DNA complexes. When bound to a MutS homolog, the adenine of ADP is sandwiched between a pair of conserved aromatic residues (underlined) in the YUP (tyrosine or phenylalanine, isoleucine or valine, and proline) and FLY (Phe-Leu-Tyr) motifs (Phe650 and Tyr815 of MSH2) (Fig. 3c). In MSH3 the nonconserved residues surrounding YUP and FLY change the peptide backbone conformation and close the nucleotide-binding site.

Figure 4 Dimerization domains (DMD) of MutSβ. (a,b) Orthogonal views of the C-terminal halves of DMDs. The hydrophobic side chains at the interface, and polar residues forming salt bridges that stabilize intrasubunit interactions, are shown as sticks with carbon in light gray, nitrogen in blue and oxygen in red. Glu901 and Lys912 of MSH2 form N- and C-caps of the MSH3 helices. (c) The ATPase domain (light green) and DMD (green) of MSH2 are shown with the trans-acting N2 (red and cyan) and DMD (blue) of MSH3. The ADP bound to MSH2 is shown as purple sticks. The two shaded ovals indicate the enlarged areas shown in a, b and d. (d) A closeup view of the interactions between MSH3 DMD(N) and the ATPase domain of MSH2. Interactions between hydrophobic residues dominate, and two pairs of salt bridges (red dashes) may have alternative interacting partners (black dashes) if the two subunits slide relative to each other.

Figure 5 Asymmetric ATPase sites of MutSβ. (a) Ribbon diagram of the ATPase and dimerization domains. MSH2 is shown in light and dark green, and MSH3 in light and dark blue. The trans-acting N2 regions of MSH2 and MSH3 are highlighted in red. The aromatic side chains connecting the ATPase site to the DMD are shown as blue (MSH3) and green sticks (MSH2). (b) A view 180° from a showing the asymmetric DMDs, biased toward the MSH2 ATPase site. (c) Comparison of the connection between the ATP binding site and DMD in MSH3 (blue), MSH6 with ADP (pink) and without (yellow) after superposition. The critical aromatic side chains are shown as sticks.
In the absence of a nucleotide, the P-loops of all MutS homologs assume an α-helical conformation by repositioning a glycine in the Walker A motif (Gly-Gly-Lys-Ser, Gly891 in human MSH3) and are closed for phosphate binding. In bacterial MutS (PDB 1EWQ16 and 1E3M17) and human MSH6 (PDB 2O8E18), the YUP and FLY motifs stay put, leaving the adenine-binding site open. In MSH3, Tyr868 of YUP occupies the adenine-binding site because of an insertion of Ghu-Gln-Asp-Gln before YUP and an altered peptide conformation (Fig. 5c). This is due to the replacement of a small side chain (alanine or cysteine) in bacterial MutS, eukaryotic Msh2 and Msh6 by a phenylalanine conserved among Msh3 homologs (Phe1023 in human MSH3, Supplementary Fig. 3). Phe1023 occupies the space normally occupied by the phenylalanine of the FLY motif (Phe1046 of human MSH3), causing FLY to alter its course and close the ATP-binding site. Phe1023 also links the adenine stacking FLY (Tyr1048) and the SYG motif (Tyr1059) in the DNA-binding domains and ATP-binding by Msh2 is promoted. With Msh3 or Msh6 bound, the ATPase activity of these subunits is linked to the moderate change of its ATPase activity upon DNA binding.

First, MSH3 clearly dominates DNA binding, using both its MBD and clamp domains (Fig. 3c). The DNA is off-center and biased toward MSH3. Similarly in the crystal structure of MutSβ, MSH6 also dominates the DNA binding of the short DNA used43. Second, MSH2 and MSH3 are structurally different. In MSH3, domains I and II are extended toward a mismatched DNA and associated intimately with domains III, IV and V (Fig. 2a, b). The tight domain association, which is conserved among Msh3, Msh6 and the mismatch-binding subunit in bacterial MutS (Fig. 2c and Supplementary Fig. 3), is coupled with the strong inhibition of ATPase activity of these subunits upon binding to a mismatched DNA43-45. Meanwhile the loose domain association of MSH2 in MutSα and MutSβ is linked to a conformational change in its ATPase activity upon DNA binding.

Mutational studies of yeast MutSα indicate that ATP binding by Msh2 is essential for recruiting MutLα, whereas ATP binding by both subunits releases a bound DNA40,43. Combining the existing data, we propose the following model of mismatch-repair initiation in humans (Fig. 6). In the absence of DNA, the four DNA-binding domains of MutSα or MutSβ move around freely46. When bound to a normal DNA, which is resistant to deformation, domains VI close and form a clamp, but domains I remain flexible. Both MutS subunits can bind ATP, and this results in the dissociation or sliding of the protein along DNA48. In the presence of a mismatch, the MBD of Msh3 or Msh6 lodges into the lesion site and induces a conformational change in its ATPase domain and DMDs. Consequently, nucleotide is occluded from Msh3 or Msh6, and ATP binding by Msh2 is promoted. With Msh3 or Msh6 bound to a mismatch and Msh2 to ATP, either MutSα or MutSβ can recruit MutLα and activate mismatch repair. This model provides a detailed molecular picture that can help guide future experimentation.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

Accession codes. Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under accession codes 3THY (Loop2), 3THX (Loop3), 3THW (Loop4) and 3THZ (Loop6).

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank J. Jiang for collecting X-ray diffraction data of Loop4; A. Howard for help with data processing; T. Kunkel for the term ‘isomerization’; R. Kolodner, E. Alani and J. Surtees for vital discussions; and D. Leahy for critical reading of
1. Iyer, R.R., Pluciennik, A., Burdett, V. & Modrich, P.L. DNA mismatch repair: functions and mechanisms. Chem. Rev. 106, 302–323 (2006).
2. Jurany, M.S., Ombrová, 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 7, 1–12 (2001).
3. Habraken, Y., Sung, P., Prakash, L. & Prakash, S. Binding of insertion/deletion DNA mismatches by the heterodimer of yeast mismatch repair proteins MSH2 and MSH3. J. Mol. Biol. 281, 1185–1196 (1996).
4. Palombo, F. & Kolodner, R.D. Inhibition of Msh6 ATPase activity in mismatch recognition. J. Biol. Chem. 274, 9837–9841 (1999).
5. Mendillo, M.L. et al. A conserved MutS homolog connector domain interface interacts with MutL homologs. Proc. Natl. Acad. Sci. USA 106, 22223–22228 (2009).
6. Habraken, Y., Sung, P., Prakash, L. & Prakash, S. ATP-dependent assembly of a ternary complex consisting of a DNA mismatch and the yeast MSH2–MSH6 and MLH1–PM1 protein complexes. J. Biol. Chem. 273, 9837–9841 (1998).
7. Harfe, B.D. & Jinks-Robertson, S. Sequence composition and context effects on the generation and repair of frameshift intermediates in mononucleotide runs in Saccharomyces cerevisiae. Genetics 156, 571–578 (2000).
8. Surtees, J.A. & Alani, E. Mismatch repair factor MSH2–MSH3 binds and alters the conformation of branched DNA structures predicted to form during genetic recombination. J. Mol. Biol. 360, 523–536 (2006).
9. Marsischky, G.T., Filosi, N., Kane, M.F. & Kolodner, R. Redundancy of Saccharomyces cerevisiae MSH3 and MSH6 in MSH2-dependent mismatch repair. Genes Dev. 10, 407–420 (1996).
10. Herman, J.G. et al. Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. Proc. Natl. Acad. Sci. USA 95, 359–362 (1998).
11. de Wind, N. et al. HNPPC-like cancer predisposition in mice through simultaneous loss of Msh3 and Msh6 mismatch-repair protein functions. Nat. Genet. 23, 153–162 (2000).
12. Sia, E.A., Dominska, M., Stefanovic, L. & Petes, T.D. Isolation and characterization of a human Msh2(ATP)–Msh6(ATP) state capable of hydrolysis-independent movement along DNA. Nucleic Acids Res. 37, 3145–3154 (2009).
13. Hohne, D. et al. Human DNA mismatches are lesion dependent. Nat. Struct. Mol. Biol. 16, 1057–1064 (1999).
14. Skyba, D.A. et al. Genetic studies of DNA mismatch repair in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch-repair proteins. Hum. Mol. Genet. 28, 7248–7265 (2019).
15. Droms, J., Wang, Y. & Kolodner, R. The crystal structure of DNA mismatch repair protein MutSbeta does not hijack (CAG)n hairpin repair in vitro. J. Biol. Chem. 284, 20452–20456 (2009).
16. Li, Y. et al. The DNA mismatch repair genes Msh3 and Msh6 cooperate in mismatch recognition and repair. Nat. Struct. Mol. Biol. 16, 1057–1064 (1999).
17. Lyapin, E., Kebir, B. & Hingorani, M.M. Contribution of Msh2 to DNA repair in human colorectal cancer. Proc. Natl. Acad. Sci. USA 107, 12593–12598 (2010).
18. Wang, W., Guerrette, S. & Fishel, R. Dissociation of mismatch recognition and ATPase activity by MSH2–MSH3. J. Biol. Chem. 274, 21659–21664 (1999).
19. Liu, Y., Li, G.M. & Fishel, R. Distinct nucleotide binding/hydrolysis properties and molar ratio of MutSalpha and MutSbeta determine their differential mismatch binding activities. J. Biol. Chem. 284, 11567–11562 (2009).
20. Owen, B.A.L., Lang, W.H. & McMurray, C.T. The nucleotide binding dynamics of human MSH2–MSH3 are lesion dependent. Nat. Struct. Mol. Biol. 16, 550–557 (2009).
21. Ye, T. et al. Mechanistic characterization of mismatch repair system in vivo. J. Biol. Chem. 282, 16345–16354 (2007).
22. Natrajan, G. et al. Structures of Escherichia coli DNA mismatch repair enzyme MSH5 in complex with different mismatches: a common recognition mode for diverse substrates. Nucleic Acids Res. 31, 4814–4821 (2003).
23. Dominska, M., Petes, T.D. & Alani, E. Escherichia coli MutS tetramerization domain structure reveals that stable dimers but not tetramers are essential for DNA mismatch repair in vivo. J. Biol. Chem. 282, 16345–16354 (2007).
24. Biswas, I. et al. Oligomerization of a MutS mismatch repair protein from Thermus aquaticus. J. Biol. Chem. 274, 23673–23678 (1999).
25. Hess, M.T., Mendillo, M.L., Mazur, D.J. & Kolodner, R.D. Biochemical basis for dominant mutations in the Saccharomyces cerevisiae MSH6 gene. Proc. Natl. Acad. Sci. USA 103, 558–563 (2006).
26. Attwood, J. et al. Crystal structure of mismatch repair protein Msh2 and its complex with a substrate DNA. Nature 407, 703–710 (2000).
27. Lammers, M.H. et al. The crystal structure of DNA mismatch repair protein MutS binding to a G × T mismatch. Nature 407, 711–717 (2000).
28. Warren, J.E. et al. Structure of the human MutSalpha DNA lesion recognition complex. Mol. Cell 26, 579–592 (2007).
29. Bowers, J., Sokolsky, T., Quach, T. & Alani, E. A mutation in the MSH6 subunit of the Saccharomyces cerevisiae MSH2–MSH6 complex disrupts mismatch recognition. J. Biol. Chem. 274, 116125–116125 (1999).
30. Drothmann, K., Yang, W., Brownlowe, F.E., Koo, E.T. & Kolak, T.A. Asymmetric recognition of DNA local distortion. Structure-based functional studies of eukaryotic Msh2–Msh4. J. Biol. Chem. 276, 46225–46229 (2001).
31. Droms, J.M., Putnam, C.D. & Kolodner, R.D. Functional studies and homology modeling of Msh2–Msh3 predict that mismatch recognition involves DNA bending and strand separation. Mol. Cell 30, 3321–3328 (2010).
ONLINE METHODS

DNA constructs and protein preparation. Human MSH2 and MSH3 were amplified by Superscript III one step RT-PCR (Invitrogen) using DNA isolated from HeLa cells (RNasey mini kit; Qiagen). The two cDNAs were cloned into pCR-Blunt (Invitrogen) and sequenced. Our MSH3 clone encodes the same protein as EAW95862.1 in GenBank and does not have a 9-aa (AAAAAAPA) duplication and two amino acid substitutions as does the second version of human MSH3 (AAAB47281.1 in GenBank). His6-MBP-PreScission was added downstream of the polyhedrin promoter using BamHI and SpeI restriction sites in a FastBac dual. The nonconserved N-terminal 210 and C-terminal three residues of human MSH3, also predicted to be unstructured, were deleted for structural studies. The full-length and trimmed human MSH3 (encoding 211–1,125 residues) were cloned downstream of a His6-MBP-PreScission cassette using a SpeI restriction site. Full-length human MSH2 (encoding 1–934 residues) was cloned downstream of the p10 promoter using KpnI and XhoI sites. The vector was transformed in DH10Bac cells (Invitrogen). Bacmids were prepared from white colonies and transfected into SF-9 cells in Grace’s Insect Cell Medium (not supplemented) using Cellfectin II (Invitrogen) according to the manufacturer’s instructions. Human MSH2 and human MSH3 were coexpressed by infecting Hi5 insect cells with amplified viruses were amplified according to the manufacturer’s protocol. Human MSH2 and MSH3 (encoding 1–934 residues) were removed by amylose resin, and the proteins were further purified by Heparin and Mono Q columns (GE Healthcare). Supernatant containing virus was harvested after 48 h after infection.

Insect cells were lysed in Buffer A (25 mM HEPES pH 8.0, 1 M NaCl, 30 mM imidazole, 10% (v/v) glycerol and 1 mM Tris-(2-carboxyethyl)phosphine (TCEP). Full-length and trimmed versions of human MutSβ were affinity purified using a Ni2+ affinity column (GE Healthcare). The proteins were eluted with a step gradient of 30% (v/v) Buffer B (Buffer A + 300 mM imidazole). The protein peaks were collected and the His6-MBP tags were cut off using PreScission protease. The tag and uncleaved human MutSβ were removed by amylose resin, and the proteins were further purified by Heparin and Mono Q columns (GE Healthcare). Supernatant was used as the final purification step. Concentrations of MutSβ were measured optically using an extinction coefficient of 134,565 M−1 cm−1 at 280 nm (0.64 at 1 mg ml−1). Proteins were stored in the final buffer (25 mM HEPES pH 8.0, 100 mM KCl, 5 mM MgCl2, 0.1 mM EDTA, 20% (v/v) glycerol, 2 mM TCEP and 0.2 mg ml−1 BSA) and incubated for 10 min at 25 °C. The complexes were resolved on a 6% (w/v) DNA-Retardation gel (Invitrogen) in 0.5X TBE at 4 °C, 80V for 1 h. The gels were developed using phosphor-imaging plates and read on a Typhoon Trio (GE Healthcare). ImageQuant TL was used to quantify the complex and Prism was used to calculate Kd values from three independent measurements using nonlinear regression and the ‘one site binding’ equation. Addition of nonspecific DNA in the binding assay competes MutSβ away from IDLs and increases the apparent Kd for IDLs by 20- to 40-fold (data not shown).

ATPase assay. ATPase activity was assayed in 15 µl of reaction buffer containing 25 mM HEPES (pH 8.0), 100 mM KCl, 2 mM TCEP, 5% (v/v) glycerol, 5 mM MgCl2, and 100 nM human MutSβ protein and 200 nM DNA. Kinetic data were obtained by varying the ATP concentration from 10 to 1,000 µM, with 50 fmol of [%32P]-ATP (PerkinElmer) in each reaction. The reaction was incubated at 37 °C for 30–45 min and terminated by addition of an equal volume of 50 mM EDTA. One microliter of each reaction mix was spotted on a PEL-Cellulose TLC plate (Grace Discovery Sciences, formerly Alltech). Labeled ATP and ADP were separated by developing the TLC plate in 0.75 M KH2PO4 and visualized by a phosphorimaging plate and Typhoon TRIO (GE Healthcare). ImageQuant TL was used to quantify the products and Prism was used to calculate Kcat and Vmax values from three independent measurements using nonlinear regression and the one site binding equation.

49. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326 (1997).
50. Vagin, A. & Teplyakov, A. Molecule replacement with MOLREP. Acta Crystallogr. D Biol. Crystallogr. 66, 22–25 (2010).
51. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. 66, 486–501 (2010).
52. Brünger, A.T. et al. Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. 54, 905–921 (1998).
53. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 66, 213–221 (2010).