Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA

Galina Obmolova*, Changill Ban†‡, Peggy Hsieh* & Wei Yang†

** Genetics and Biochemistry Branch and † Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA*

DNA mismatch repair is critical for increasing replication fidelity in organisms ranging from bacteria to humans. MutS protein, a member of the ABC ATPase superfamily, recognizes mispaired and unpaired bases in duplex DNA and initiates mismatch repair. Mutations in human MutS genes cause a predisposition to hereditary nonpolyposis colorectal cancer as well as sporadic tumours. Here we report the crystal structures of a MutS protein and a complex of MutS with a heteroduplex DNA containing an unpaired base. The structures reveal the general architecture of members of the MutS family, an induced-fit mechanism of recognition between four domains of a MutS dimer and a heteroduplex kinked at the mismatch, a composite ATPase active site composed of residues from both MutS subunits, and a transmitter region connecting the mismatch-binding and ATPase domains. The crystal structures also provide a molecular framework for understanding hereditary nonpolyposis colorectal cancer mutations and for postulating testable roles of MutS.

DNA mismatch repair (MMR) primarily corrects mispaired or unpaired bases caused by DNA polymerase, thus increasing the overall fidelity of DNA replication by 100- to 1,000-fold[1,2]. Cells harbouring mutations in MMR genes are characterized by greatly elevated rates of spontaneous mutation and instability of microsatellite repeats[1]. Inactivation of MMR genes by mutation or epigenetic processes predisposes humans to hereditary nonpolyposis colorectal cancer (HNPCC) and sporadic tumours[3–5]. In *Escherichia coli*, three MMR proteins, MutS, MutL and MutH, initiate a strand-specific mismatch repair process that takes advantage of the transiently unmethylated state of the newly synthesized daughter strand[1]. MutS recognizes mispaired or 1–4 unpaired bases and, in an ATP-dependent fashion, activates the endonuclease MutH with the aid of another ATPase, MutL[1,6]). Strand-specificity of mismatch repair resides in MutH, which recognizes a hemimethylated GATC sequence and cleaves only the unmethylated daughter strand. The GATC site can be over 1,000 base pairs (bp) away and on either side of the mismatch. Both MutS and MutL are also essential in the subsequent excision and resynthesis of the daughter strand.

Homologues of *E. coli* MutS have been found in nearly all organisms[1,7]. Prokaryotic MutS proteins are encoded by a single gene and form homodimers[8,9]. Eukaryotic genomes contain multiple *mutS* homologue (*msh*) genes. Except for mitochondrial MSH1, eukaryotic MutS proteins are heterodimeric. For instance, MSH2 pairs with MSH6 or MSH3 to form MutSα and MutSβ[7,10]. MutS and its homologues can be divided into two lineages[11]. The majority of prokaryotic and eukaryotic MutS proteins belong to the lineage that participates in MMR, prevents homologous DNA recombination between heterologous sequences, and mediates cell death induced by DNA damaging agents[12,13]. The other lineage, including MSH4–MSH5, does not function in MMR, but promotes DNA pairing and crossing over during meiosis[13].

All members of the MutS family possess a conserved ATPase activity[1,7]. Both mismatch recognition and the ATPase activities of MutS are essential for MMR[14–16] even though each activity is independently detectable[17,18]. However, it is unclear how MutS recognizes a broad range of mismatches embedded in random DNA sequence, how the MutS ATPase activity is involved, and how these two essential functions of MutS are related. We report here the crystal structures of *Thermus aquaticus* (TAQ) MutS and its complex with a heteroduplex DNA at 3.2 Å and 2.2 Å resolution, respectively, and discuss new insights into the function of MutS in MMR.

General features of MutS and a MutS–DNA complex

TAQ MutS, a thermostable homologue of *E. coli* MutS, comprises 811 amino-acid residues[19]. We have co-crystallized a nearly full-length TAQ MutS protein (residues 1–768) and a 21-bp heteroduplex containing a single, unpaired thymidine (T). The truncated TAQ MutS retains the dimerization, mismatch-recognition and ATPase activities of the full-length protein; a similarly truncated *E. coli* MutS is proficient in mismatch repair *in vivo* (P.H. *et al.*, unpublished data). The crystal contains one MutS dimer and one heteroduplex DNA in each asymmetric unit and diffracts X-rays to 2.0 Å resolution (see Table 1 in Supplementary Information). The co-crystal structure was phased by multiwavelength anomalous diffraction with selenylmethionyl substitution[20]. The final refined model contains all nucleotides and residues 1–765 of MutS with the exception of 6 disordered residues, 629–634 (see Table 1 in Supplementary Information). The two subunits of MutS are related by a molecular pseudo-two-fold axis and form an oval disk (~125 × 90 × 70 Å; Fig. 1). Two adjacent large channels with dimensions of ~30 × 20 Å and ~40 × 20 Å penetrate the disk-like protein structure, and the latter is occupied by the heteroduplex DNA. The DNA is kinked sharply towards the major groove by ~60° at the unpaired T. Only one protein subunit interacts with the unpaired T, thereby breaking the molecular two-fold symmetry of the homodimer.

We crystallized a larger TAQ MutS fragment (residues 1–782) in the absence of DNA. The crystal structure was determined by molecular replacement and partially refined to 3.2 Å (see Fig. 1 in Supplementary Information). The protein-only structure is much more mobile than the MutS–DNA complex as indicated by much higher overall *B* values (see Table 1 in Supplementary Information).

‡ Present address: Hauptman-Woodward Medical Institute, Buffalo, New York, USA.

MutS protein

A MutS subunit consists of five structural domains arranged in the shape of a comma (Fig. 2a). Each of the five domains is structurally identical between the two subunits, even though the relative orientations of domains I, IV and V differ (Fig. 2b). Domains I (residues 1–118), II (residues 132–245), IV (residues 406–513) and V (residues 543–765) are of mixed α/β-fold, and domain III (residues 247–385 and 514–540) is entirely α-helical. Each of the five domains resembles at least one previously determined structure (see Fig. 2 in Supplementary Information). For example, domain II resembles RNase H and RNase H-like domains found in Hsc70, HIV integrase and DNA polymerase[21]; domain V, which contains the Walker A motif, is structurally similar to the ATPase domain of ABC transporters[22,23].

The structure of MutS is extended and contains multiple hydrophobic cores. Domain III is central to the MutS structure; it is directly connected to domains II, IV and V by peptide bonds and shares the most extensive interdomain contacts with domain V. Domain IV is elongated, encompassing a pair of antiparallel α-helices over 40 Å in length with a small β-sheet at the distal end. The globular domain I and domain IV are involved in DNA binding. Each of the two domains is independently folded and appended to the core of MutS by flexible peptide linkages and limited interactions (Fig. 2), which indicate the possibility of domain rearrangement. The potential mobility of these domains is confirmed in the crystal structure of MutS alone. Although domains II, III and V retain similar structures in the presence or absence of DNA, domains I and IV are too mobile to be discerned in the absence of DNA (Fig. 3a).

MutS forms a stable dimer due to the extensive interactions between the ATPase domains, which buries more than 2,400 Å^2 of molecular surface (Fig. 3a, b). The presence of DNA brings domains

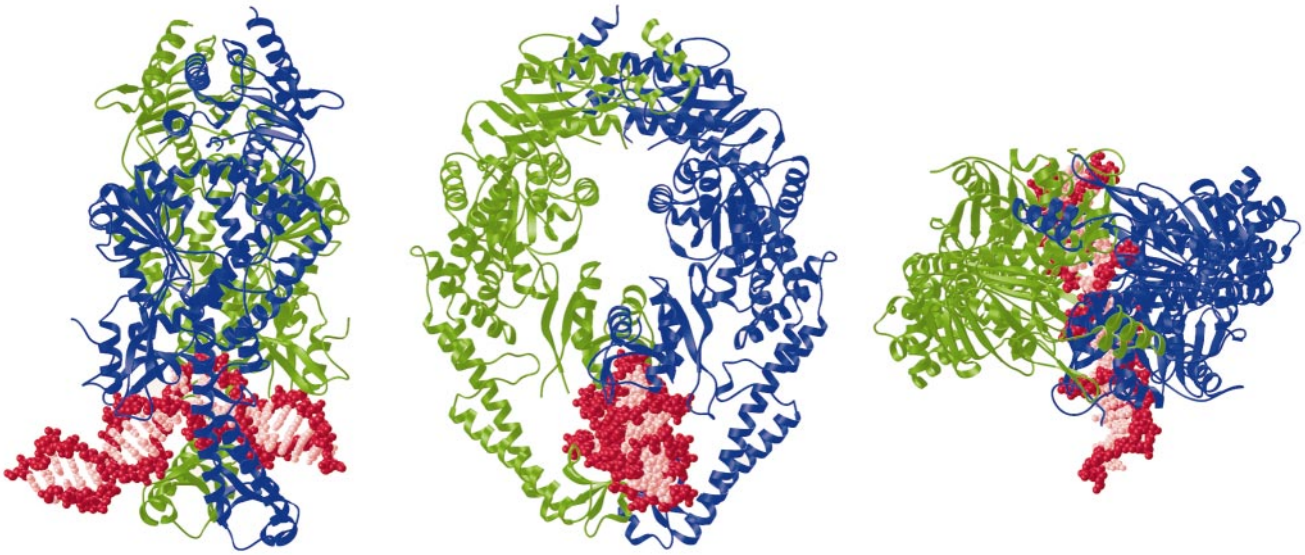

Figure 1 Crystal structure of the TAQ MutS–DNA complex. The two protein subunits are represented by ribbon diagrams in blue (A) and green (B). The DNA is shown in a space-filling model, in which the backbone atoms are red and bases are pink. Three orthogonal views from the side, front and top are shown.

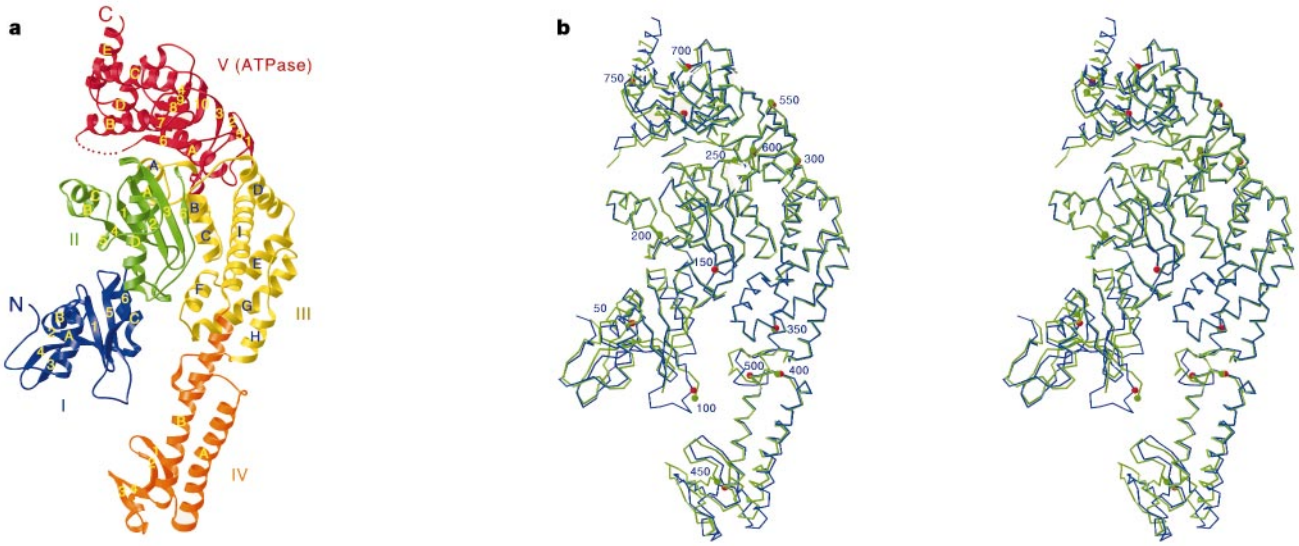

Figure 2 Crystal structure of a TAQ MutS subunit. **a**, Ribbon diagram of subunit A. The five domains are coloured blue, green, yellow, orange and red from the N to the C terminus. Secondary structures are named in alphabetic or numeric order for α-helices and β-strands, respectively, following their order in the primary sequence in each domain. **b**, A stereo view of the superposition of Cα traces of the two subunits. Subunit A is shown in blue and subunit B in green. Every fiftieth Cα of each subunit is marked by a red (subunit A) or a green (subunit B) sphere.

I and IV into the dimer interface, which adds ~600 Å^2 to the buried surface and forms the lower channel that encircles the DNA (Fig. 3b). DNA binding also induces an outward rotation of the two MutS subunits (see Fig. 3 in Supplementary Information). The arrangement of the domains I and IV is reminiscent of the structures of DNA gyrase and topoisomerase II (ref. 24), which also contain globular DNA-binding domains linked by a pair of long anti-parallel helices and form a large channel for DNA passage. However, the two families of proteins share no further structural similarity.

At the dimer interface formed between domains V, a pair of anti-parallel helices (VαD and VαE) extends from the carboxy terminus of one subunit and interacts with a pair of parallel helices (VαB and VαC) from the other subunit (Fig. 3c). The interactions between these helices are mostly hydrophobic except for two hydrogen bonds involving the side chains of D672, S736 and the main chain of V739 (Fig. 3d), and they account for ~70% of the total dimer interface. Although VαD and VαE were thought to be involved in DNA binding by virtue of a conserved helix–turn–helix motif, mutational studies indicated that these two helices are involved in dimerization and ATPase activity[15], a finding confirmed here. Removal of the last five residues of VαE abolishes not only dimerization of TAQ MutS, but also the ATPase and mismatch binding activities (P.H. *et al.*, unpublished data). As VαE is far too distant to participate directly in DNA and ATP binding (70 Å and 20 Å, respectively), dimerization must be required for both activities.

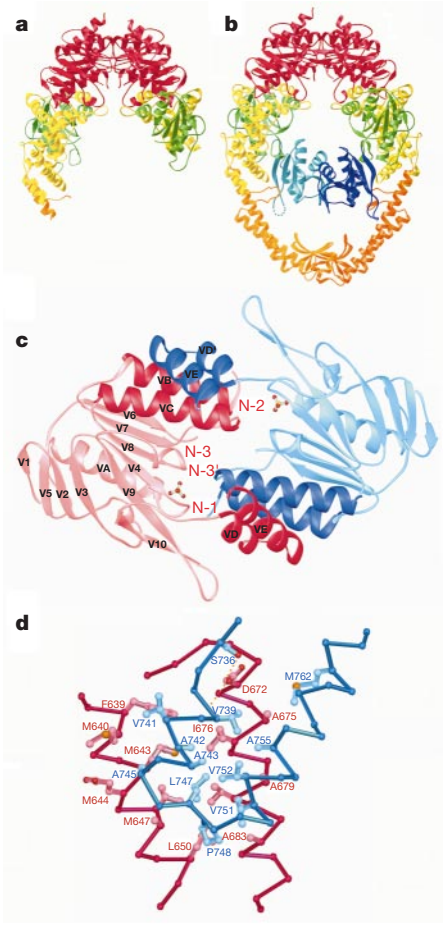

Figure 3 Structures of the MutS dimer. **a**, **b**, Ribbon diagram of the MutS dimer in the absence (**a**) or the presence (**b**) of the heteroduplex DNA. Each domain is shown in the same colours as in Fig. 2a. Domains I and IV of Subunit B are drawn in lighter colour than those of subunit A. **c**, Ribbon diagram of the dimer interface at domain V. Domain V of subunit A and B are coloured in blue and red, respectively. The helices that dominate the dimer interface are in a darker shade. The sulphate ions are shown as ball-and-stick. The four nucleotide-binding motifs in the red subunit are labelled. **d**, One of the four-helix bundles at the dimer interface. The mainchains are represented by the Cα traces. The sidechains that interact between the two subunits are labelled and shown as ball-and-stick. Two hydrogen bonds are represented by yellow dashed lines.

Mismatch DNA recognition by MutS

The unpaired T in the MutS co-crystal structure is partially stacked in the DNA duplex (Fig. 4), in agreement with the observation that a flipped out nucleotide is not recognized by MutS[25]. The base is displaced towards the minor groove by ~3 Å and tilts towards the 5′ base pair so that its O4 oxygen is hydrogen bonded to the N2 of the 5′ G. This conformation of the unpaired T is stabilized by the interactions with MutS, which include an aromatic-ring stack with Phe 39 on the 3′ side and a hydrogen bond with the side chain of Glu 41 (Fig. 4c). Phe 39 was postulated to be adjacent to the unpaired T based on ultraviolet crosslinking experiments[26]. In the crystal structure, Phe 39 approaches the DNA from the minor-groove side, stacks onto the unpaired T, and contacts the base 3′ to the T with the tip of its phenyl ring (Fig. 4c). As a result, the DNA is sharply kinked towards the major groove, and the minor groove is as wide as the adjacent major groove (Fig. 4a, b). All base pairs on either side of the kink assume a canonical B-form conformation except for a slight buckle at the base pair 5′ to the unpaired T.

The kinked heteroduplex DNA is stabilized by extensive interactions with four MutS domains. Domains I and IV of the two subunits crisscross each other and surround 5 and 8 bp on each side of the kink, burying a total of ~3,300 Å^2 of molecular surface (Fig. 4a, b). Domains I and IV share a similar folding topology, two pairs of β-hairpins (β1, β2 and β3, β4) linked by a helical segment form an anti-parallel β-sheet (Fig. 4d). Similar anti-parallel β-sheets have been observed (see Fig. 2 in Supplementary Information), but this is the first example in which these β-sheets are found binding to DNA and interacting with either the major or the minor groove (Fig. 4). The protein–DNA interactions, except at the unpaired T, involve only the phosphate–sugar backbone of the DNA and thus impose no direct sequence specificity on mismatch recognition (Fig. 4e).

The interactions between MutS and DNA are asymmetric. Domain I of subunit A, which donates the intercalating Phe 39, interacts with nine nucleotides by wedging the β-sheet between phosphate backbones in the widened minor groove and anchoring a third β-hairpin (Iβ5 and Iβ6) on the outside (Fig. 4). The amino terminus of IαA also contributes to DNA binding (Fig. 4e). Many of these residues contacting phosphate–sugar backbones are conserved (Figs 4e, 5). Domain IV of subunit B, situated on the opposite side of DNA from domain I of subunit A, uses the two β-turns on the opposite ends of the β-sheet to contact the DNA backbones across the major groove, spanning at least 8 bp (Fig. 4). In contrast, domain IV of subunit A and domain I of subunit B, whose Phe 39 is not engaged in DNA binding and Iβ5–Iβ6 turn is partially disordered (Fig. 4a), contact only two or three adjacent nucleotides on a single strand (Fig. 4e). The asymmetric protein–DNA interactions surrounding the mismatch site and the sharp kink of DNA towards the major groove are in good agreement with previous DNase I and chemical footprinting results[17,27]. The requirement for four domains from two protein subunits to bind DNA explains the dependence of mismatch recognition on protein dimerization[15] (P.H. *et al.*, unpublished data).

The observed flexibility of MutS provides a structural basis for its broad substrate specificity. First, the mobility of the DNA-binding domains evident in the structure of MutS alone (Fig. 3a) provides an entry for DNA to the mismatch-binding channel. Second, the adaptability indicated by the asymmetric use of identical domains in DNA binding enables MutS to accommodate a broad range of heteroduplex DNAs, while maximizing the protein–DNA interface. MutS has been reported to bind DNAs containing mispaired or unpaired bases, 8-oxoG, O^{6}meG or cisplatin adducts[7], which are

likely to have different local structures. In support of this assessment, the replacement of the unpaired T with a T·T mismatch in the MutS–DNA complex results in a new crystal lattice (G.O. *et al.*, unpublished data).

Kinked DNA is likely to be a general feature in MutS–DNA complexes because a straight DNA would clash severely with one domain I and reduce the protein–DNA interface (Fig. 4a). Although mismatch-containing DNAs appear to be straight in the absence of protein, base stacking is less stable at a mismatch site and more susceptible to deformation[28]. DNA modified by cisplatin is intrinsically kinked towards the major groove[29] and recognized by human MutSα[7]). MutS proteins also bind nonspecifically to homoduplex DNA[14,16], which evidently can be bent or sharply kinked upon protein binding[30]. A mutation of Phe 39 to Ala in TAQ MutS, or equivalent mutations in *E. coli* MutS and MSH6 of yeast MutSα, abolishes MutS binding to both hetero- and homoduplex DNA[31,32] (A. Yamamoto and P.H., unpublished data) indicating that they may share a similar binding site in MutS.

The union of MutS and a heteroduplex DNA is thus a result of induced fit of both molecules that involves kinking of DNA and rearrangment of protein domains. The preferred binding of MutS to a heteroduplex DNA is most probably due to a reduction in the energetic cost of kinking DNA in the presence of a mismatch or a modified base[28]. A propensity for stacking mismatched bases within DNA helix rather than flipping them out may further explain the different binding efficiency of MutS for a T·G versus T·T mismatch and dependence on the surrounding base content, that is, AT- versus GC-rich sequence[25].

A composite ATPase active site

Four nucleotide-binding motifs (N-1, -2, -3 and -3′) conserved in the MutS family[22] are located at the dimer-interface region of domain V (Figs 3c, 5). N-1 and N-3 correspond to the Walker A and B motifs present in many ATPases. N-2 and N-3′ are unique to the ABC ATPase superfamily, which includes MutS, UvrA, Rad50, CFTR and ABC transporters[22]. In the crystal structure of the TAQ MutS–DNA complex, a sulphate ion is hydrogen bonded to the P-loop formed by N-1 (GPNMAGKS, residues 583–590) and marks the ATP-binding site (Fig. 3c). N-3 (DE, resisdues 662–663) and N-3′ (TH, residues 695–696) are located on the β-strands adjacent to the P-loop (Fig. 3c). Interestingly N-2 (ST, residues 636–637), which potentially binds the γ-phosphate, is near the ATP-binding site of the neighbouring subunit, but far from that of the same subunit (Fig. 3c). The crystal structure of a MutS–DNA–ADP ternary complex and mutational studies confirm that N-2 acts in *trans* in ATP hydrolysis (M. S. Junop *et al.*, manuscript in preparation). The composite ATPase active site is supported by the highly conserved MutS dimer interface (Fig. 5) and the observation that

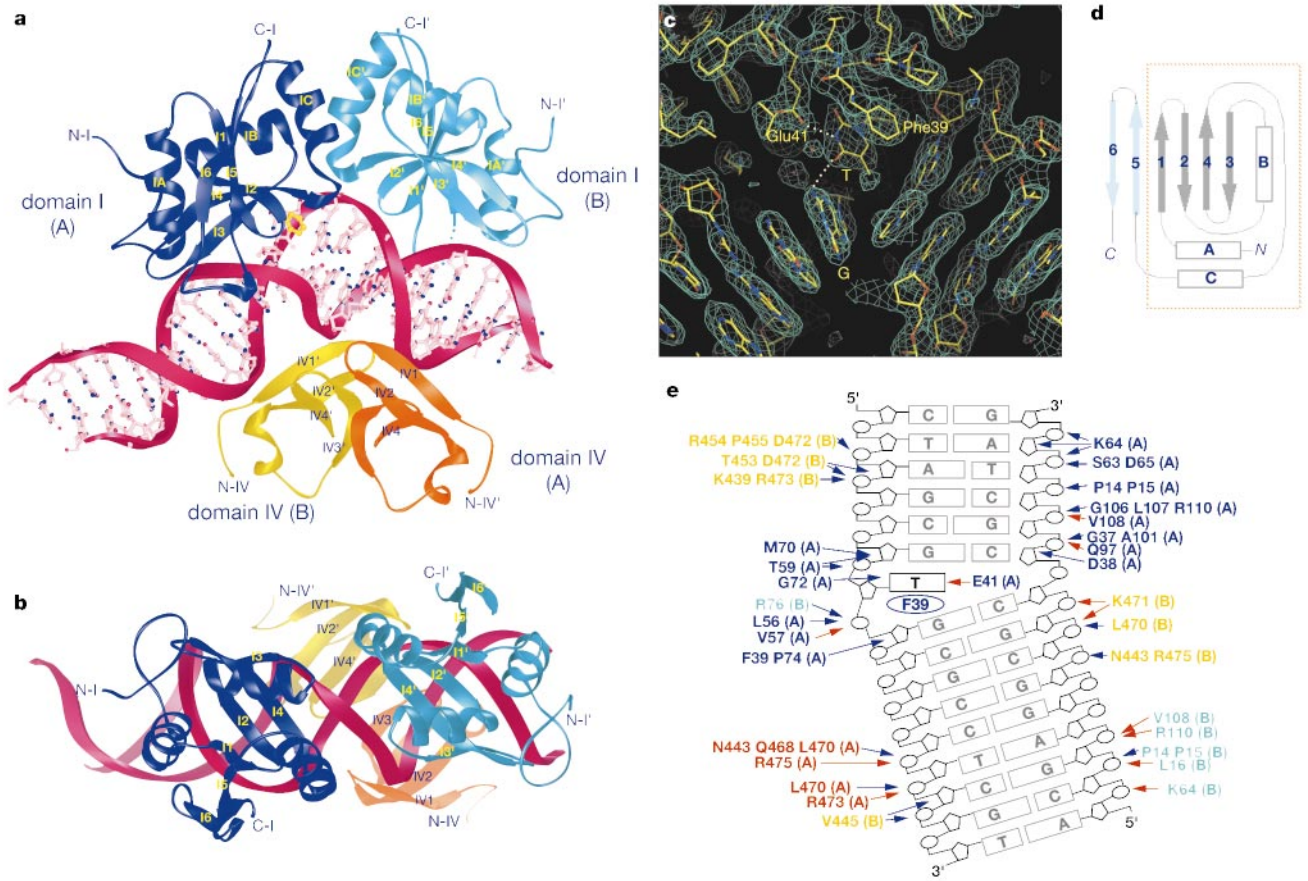

Figure 4 Mismatch recognition by MutS. **a**, The DNA-binding domains represented by ribbon diagrams are shown with the heteroduplex DNA. Domains I and IV of subunit A are shown in dark blue and orange; the corresponding domains of subunit B are in light blue and yellow. The DNA phosphate backbones are represented by red ribbons and sugars and bases are shown as ball-and-stick. Phe 39 of subunit A is shown in yellow. Structural elements of subunit B are denoted with a prime. **b**, A view from the top of **a**. **c**, A $2F_o–F_c$ electron density map contoured at 1.0σ is shown with the refined structure surrounding the unpaired T. The unpaired T and the G 5′ to it are labelled. Both Phe 39 and Glu 41 belong to subunit A. The dashed lines represent hydrogen bonds. **d**, Topology diagram of domains I and IV. Features common between the two domains are boxed and shown in grey. IαB is replaced by a 3_{10} helix in domain IV. **e**, A diagram of protein–DNA interactions. Only those base pairs that contact MutS are shown. Phosphate groups are represented by ovals between the sugar pentagons. Interactions made by each protein domain are labelled in the same colour as the ribbon diagram of this domain in **a** and **b**. Hydrogen bonds are represented by red arrows and van der Waals contacts by blue arrows.

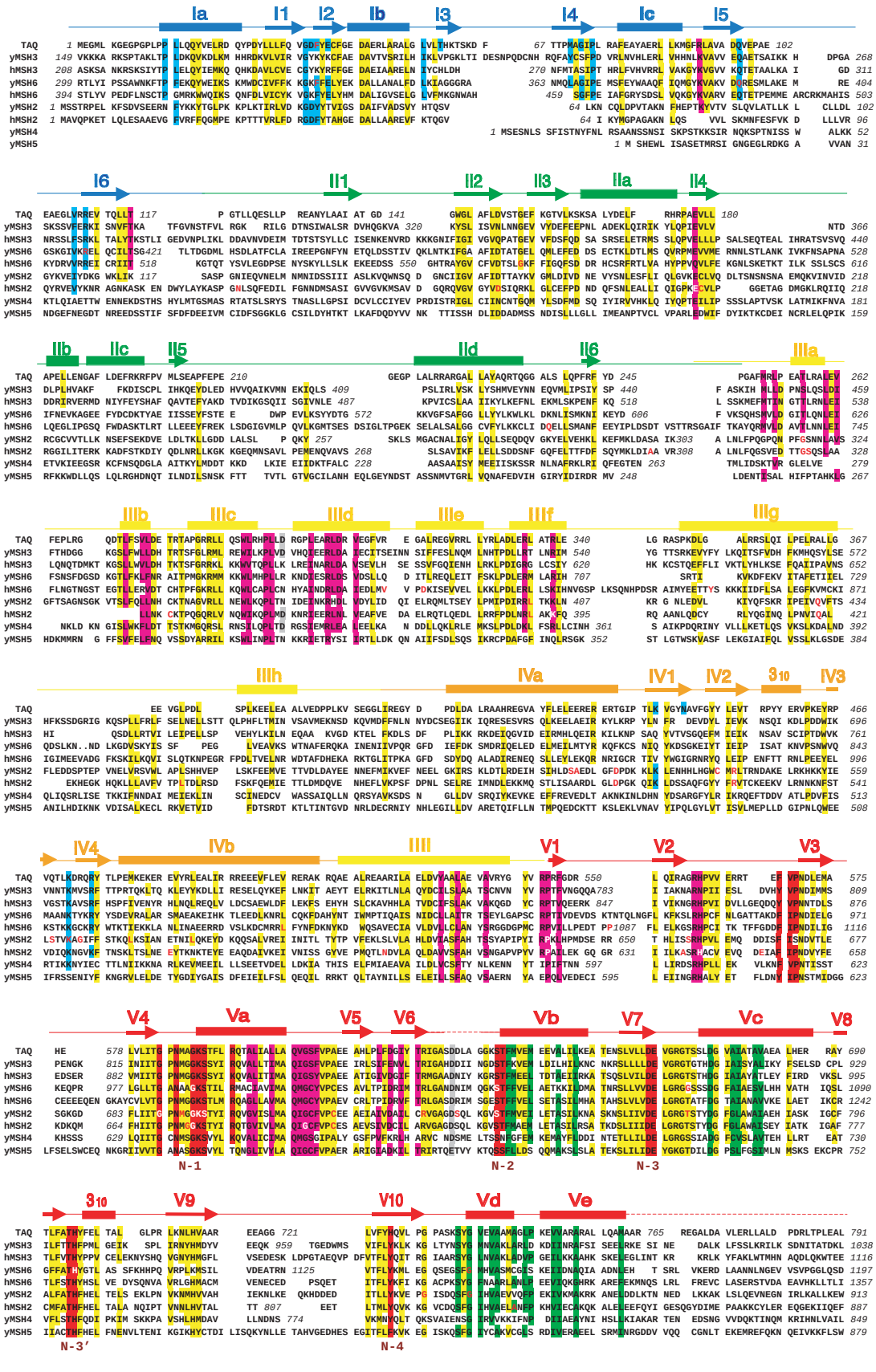

Figure 5 Structure-based sequence alignment. MutS homologues from human (hMSH), *S. cerevisiae* (yMSH) and *T. aquaticus* are shown. The secondary structures observed in TAQ MutS are indicated above the aligned sequence. Conserved residues for structural integrity are highlighted in yellow, for DNA recognition in blue, for protein dimerization in green, for ATPase activity in red and for interdomain interactions in purple. Residues that, when mutated, cause defective mismatch repair in yeast[32,36,48–50], or HNPCC in humans[3,4] are coloured red, or, if they are highlighted in purple and red, white. The five nucleotide-binding motifs are indicated beneath the sequence alignment.

the ATPase activity depends on a native MutS dimer[15] (P.H. *et al.*, unpublished data).

A composite ATPase active site is probably a common feature among members of the ABC superfamily, which share the conserved nucleotide-binding motifs[22] and require dimerization for activity[33]. Like MutS, Rad50 also contains a composite ATPase active site[34]. However, in the crystal structure of HisP, the ATPase subunit of His permease, two protein subunits are packed back-to-back and the conserved nucleotide-binding motif equivalent to N-2 is far away from the bound ATP[23]. Perhaps when HisP forms a functional histidine transporter with HisQM subunits, it will dimerize similarly to the dimerization observed in Rad50 and MutS.

Structure prediction of the MutS family

The entire TAQ MutS sequence can be aligned with every member of the MutS family (Fig. 5; and see Fig. 4 in Supplementary Information). The sequence conservation among prokaryotic members indicates that their three-dimensional structures are homologous to that of TAQ MutS (see Fig. 4 in Supplementary Information). The general architecture of TAQ MutS is also conserved among more divergent eukaryotic MutS homologues (Fig. 5). Deletions and insertions in MutS proteins often correlate with the functional diversity. For example, MSH4, MSH5 and Him14 are involved in meiotic DNA segregation but not mismatch repair[10,35], and sequences in domain I for mismatch recognition are largely absent in these proteins (Fig. 5). In the absence of domain I, which separates the two channels in TAQ MutS (Fig. 3b), the MSH4–MSH5 dimer is likely to contain a single ~30 Å by 70 Å hole, which should be large enough to encompass two DNA duplexes side by side such as might be encountered during Holliday junction migration.

MutSα and MutSβ have different substrate specificity in MMR. MutSα recognizes duplex DNA that contains mispaired or 1–2 unpaired bases, as does TAQ MutS[1,7]. MutSβ has evolved to recognize large unpaired loops, arising with long sequence repeats frequent in eukaryotic genomes. MutSα and MutSβ are composed of MSH2–MSH6 and MSH2–MSH3 heterodimers, respectively. On the basis of the asymmetric usage of the two TAQ MutS subunits in mismatch binding, only one subunit in MutSα and MutSβ is expected to directly contact mismatched bases. The unique subunit, MSH6 or MSH3, is likely to recognize mismatches as is subunit A of TAQ MutS. All residues important for mismatch recognition by TAQ MutS are conserved in MSH6 (Fig. 5). In MSH3, Lys and Arg residues are found to replace Phe 39 and Glu 41 of TAQ MutS (Fig. 5) and probably interact with the phosphate backbone instead of an unpaired base (Fig. 4c). These substitutions may be important in substrate specificity of MutSα versus MutSβ[7].

The MSH2 subunit lacks the equivalent of Q97 and R110 that are essential for TAQ MutS to anchor domain I in the widened minor groove (Figs 4, 5). It also contains a deletion of 12–14 residues in the region encoding the Iβ3–Iβ4 hairpin (Fig. 5), which wedges in the widened DNA minor groove and buttresses the nearby Phe 39 in the TAQ MutS–DNA complex (Fig. 4). Mutating the residue equivalent to Phe 39 of TAQ MutS in MSH2 has no effect on mismatch binding by yeast or human MutSα[31] (J. Jiricny, personal communication). MSH2 probably plays a role equivalent to that of subunit B of TAQ MutS and stabilizes the kinked conformation of mismatch DNA (Fig. 4). Consistent with this prediction, point mutations in the DNA-binding region of domain IV in yMSH2 cause a mutator phenotype in yeast[36] (Fig. 5).

A transmitter connecting ATP- and DNA-binding sites

DNA stimulates the MutS ATPase activity[14–16]. Conversely, the presence of ATP reduces the affinity of MutS for DNA[8,14,15]. The crystal structures show that a bound DNA not only immobilizes domains I and IV, but also alters the structures of the ATPase domains (Fig. 2b; and see Fig. 3 in Supplementary Information).

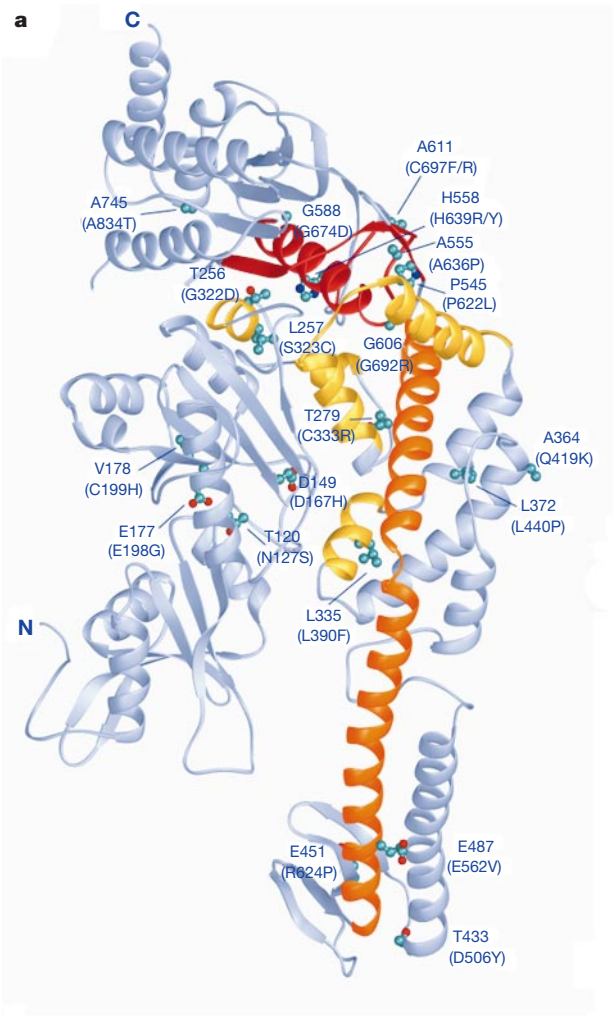

Figure 6 The 'transmitter', HNPCC and MMR. **a**, The ribbon diagram of MutS subunit A. The secondary structures that contain conserved residues to form the 'transmitter' at the three-domain junction are highlighted in red (domain V), yellow (domain III) and orange (IIIα). IVαB is also highlighted in orange for its role in the allosteric connection. Residues corresponding to the missense mutations found in the HNPCC databases[3,4] are also shown. Owing to limited clinical and functional data, it is uncertain whether all of these mutations are pathogenic and cause MMR defect. For instance, G322D, L390F and Q419K appear in normal individuals at an allele frequency of 1–2% and might be polymorphisms. However, putative equivalents of these three human MSH2 polymorphisms cause elevated mutation rates in yeast[48]. **b**. A model of the MutS, MutL and MutH complex. The MutS dimer is shown in a blue and green Cα trace with the bound heteroduplex DNA, the LN40 dimer in red with two ADPnP molecules, and MutH in yellow. Subunit B of MutS is modelled to contact LN40. The LN40 and MutH interface is based on a previous proposal[6].

These observations indicate that an allosteric connection exists between the distant ATP- and DNA-binding sites.

An examination of conserved residues in the MutS family reveals the location of a potential allosteric connection. In addition to the ATPase active site and the dimer interface, conserved residues are concentrated at the junction where domains II, III and V converge (Figs 5, 6a). This three-domain junction includes α-helix A and the surrounding β-strands 1, 2, 5 and 6 of domain V, and α-helices A, B, C, D and I of domain III. The functional importance of this region is also indicated by the HNPCC mutations localized here, such as, G322D, S323C, C333R, P622L and H639R/Y (Fig. 6a). Helices VαA, IIIαI and IVαB form the backbone that directly connects the DNA-binding and the ATPase domains (Figs 2a, 6a). The connection between domains I and III is intervened by the least conserved domain II (Figs 2a, 5) and is probably established only after binding of a heteroduplex DNA. Two HNPCC mutations, N127S and D167H, reside in the I–II domain interface (Fig. 6a), indicating that this interface is critical for MMR. We propose that this three-domain junction serves as a 'transmitter' for information exchange between the ATP- and DNA-binding sites.

A model of MutS in mismatch repair

Two previous models propose that MutS leaves the mismatch either by ATP-hydrolysis dependent translocation or ATP-binding dependent sliding before initiating MMR. The bi-directional translocation model based on an α-like DNA-loop structure observed by electron microscopy implied that a heteroduplex DNA passes through a MutS dimer twice[8]. MutS does contain a second channel that by size and electrostatic potential is suitable for DNA binding (Fig. 1). However, a putative MutS dimer of two spheres observed by electron microscopy[8] actually corresponds to two MutS dimers on the basis of their shape and dimensions. In solution, MutS exists predominantly as a dimer, and formation of tetramers is dependent on protein concentration[1,9]. Both biochemical studies and the crystal structure indicate that it is a MutS dimer that recognizes a mismatch[9].

The sliding-clamp model was supported by a clamp-like structure of human MutS observed by electron microscopy in the absence of DNA[37]. Without a bound DNA, domains I and IV are mobile (Fig. 3a). This mobility complicates the interpretation of structural differences shown in the electron microscopy images, which had been attributed to the presence of ATP versus ADP[37]. The most puzzling problem raised by these two models is how the repair system knows whether and where a mismatch has occurred if MutS leaves a mismatch site.

We propose that MutS remains bound to a mismatch site to faithfully invoke MMR. Our crystal structures suggest that MutS is unable to retain the mismatch-bound conformation once it leaves a mismatch (Fig. 3). Interactions between MutL and MutS have been shown to stabilize the MutS–DNA complex and prevent MutS from leaving a mismatch even in the presence of ATP[38,39]. Examination of the MutS structure reveals an intriguing depression adjacent to the 'transmitter' (see Fig. 5 in Supplementary Information). The two depressions of a MutS dimer differ in shape and electrostatic potential as a result of asymmetric DNA binding (see Fig. 5 in Supplementary Information). We propose that one of the depressions, presumably that of subunit B, the equivalent of MSH2, is the site where MutL interacts with MutS (Fig. 6b) and potentially stabilizes the MutS–DNA interaction. On the basis of the sequence alignment, this depressed surface exists in all members of the MutS family even though the sequence is not conserved. The shape and electrostatic potential of the depression vary, perhaps mirroring a non-conserved surface of MutL proteins. Two HNPCC mutants of human MSH2 (Q419K and L440P) that are exposed to solvent at the outer rim of this depression, mark a potential site for MutSα and MutSβ to interact with MutL homologues in humans.

In *E. coli*, MutS that remains at a mismatch site should be able to activate MutH bound to a GATC sequence either upstream or downstream by protein–protein interactions mediated by MutL, much like enhancer-binding proteins in transcriptional regulation. Our preliminary results also indicate that MutS has to bind both ATP and a mismatch simultaneously to activate MutH and initiate MMR (M. S. Junop *et al.*, manuscript in preparation).

Conclusions

The crystal structures of TAQ MutS and a TAQ MutS–DNA complex capture two states of a flexible and dynamic molecule. They reveal the general architecture of this conserved family of proteins, the basis for recognition of a broad range of mismatches and the existence of a previously unappreciated composite ATPase active site. An allosteric connection between the mismatch- and ATP-binding sites of MutS has emerged from our crystal structures and the biochemical data accumulated over the past decade. The identification of a 'transmitter' region together with previous structures of MutL and MutH[6,40] provide a testable model of the MutS, MutL and MutH complex that initiates MMR in *E. coli*. Finally, the structures and structure-based sequence alignment illuminate the defects of mutations found in various MutS proteins, in particular those of HNPCC kindreds. More experimental data are clearly needed to address whether ATP hydrolysis by MutS is required for activation of MutH through MutL, and to investigate the use of the second channel in the MutS structure that is probably involved in DNA recombination and may also be involved in the search for a mismatch. □

Methods

Crystallization

Full-length TAQ MutS fails to yield diffraction quality crystals of either protein alone or protein–DNA complexes. The truncated TAQ MutS defined by partial proteolysis, amino-acid residues 1–768 and 1–782, were cloned, overexpressed and purifed[9]. Selenomethionlyl-substituted proteins were expressed in B834 (DE3) cells with defined medium[20] and purified over an additional Phenyl-Sepharose column[9]. The oligonucleotides, 5′-GCGACGCTAGCGTGCGGCTCGTC-3′ and 5′-GGACGAGCCGCCGCTAG CGTCG-3′, were synthesized in-house or purchased from Yale University. The MutS–DNA complex was made by incubating MutS (residues 1–768) with the heteroduplex DNA at a protein to DNA molar ratio of 1:1.2 in 20 mM HEPES (pH 7.0), 50 mM KCl, 5 mM $MgCl_2$, 1.0 mM dithiothreitol and 7% glycerol at 37 °C for 15 min and concentrated in a Centricon-30 from an initial protein concentration of 0.5 mg ml^{-1} to 15 mg ml^{-1}. Crystals of the complex in $P2_12_12_1$ space group (see Table 1 in Supplementary Information) were obtained at 20 °C using the hanging-drop vapour diffusion method. The reservoir solution contained 100 mM cacodylate (pH 6.2), 200 mM $(NH_4)_2SO_4$, 1 mM dithiothreitol and 18% PEG4000. Crystals appeared within a day and grew to a maximal dimension of ~0.2 × 0.3 × 0.6 mm within a week. Crystals of the TAQ MutS (residues 1–782) in space group $P3_121$ were obtained also using the hanging-drop vapour diffusion method. The protein solution was at a concentration of 40 mg ml^{-1} in 20 mM Tris (pH 8.0), 200 mM KCl, 1 mM dithiothreitol, 0.1 mM EDTA, 5 mM $MgSO_4$ and 5% glycerol, and the reservoir solution contained 90 mM HEPES (pH 7.2), 7.2% PEH400, 1.8 M $(NH_4)_2SO_4$ and 1.5 mM $CuCl_2$. Crystals of MutS alone grew at 27 °C and were pyramid-like with the longest dimension up to 0.6 mm.

Data collection

Both types of crystals were flash-frozen in liquid propane for storage and data collection. The cryo-solvent for the MutS–DNA complex contained the mother liquor with 25% PEG4000 and 10% ethylene glycol. The MutS alone crystal was cryo-protected by addition of 15% glucose to the mother liquor. The multiwavelength anomalous diffraction dataset of the MutS–DNA complex at 2.5 Å resolution was collected at four wavelengths with inverse beam geometry at Brookhaven National Laboratory (BNL) beamline X9B using a Quantum4 CCD detector at 100 K (see Table 1 in Supplementary Information). For structure refinement, a dataset of the MutS–DNA complex from a larger crystal of the same selenomethionyl-substituted batch was collected to 2.2 Å resolution in-house at 95 K using a RAXIS-II detector mounted on a Rigaku RU-200 generator. The diffraction data of the MutS alone crystal was collected at BNL beamline X4A at 100 K using an R-AXIS IV detector. All diffraction data were processed by HKL (ref. 41) (see Table 1 in Supplementary Information).

Structure determination

SOLVE (ref. 42) found 24 Se sites in the MutS–DNA complex and generated an interpretable electron density map at 2.5 Å (see Table 1 in Supplementary Information). The map was further improved using DM (ref. 43). A complete protein and DNA model were built into the electron density map using O (ref. 44). The resulting model was refined against the 2.2 Å dataset using CNS (ref. 45). The crystal lattice contacts are formed mainly

between MutS molecules with one DNA end packed against a symmetry-related MutS. The final model consists of MutS subunit A (residues 1–628 and 635–765) and B (residues 1–100, 108–628 and 635–762), the DNA heteroduplex, 4 sulphate ions, 7 ethylene glycol and 814 water molecules (see Table 1 in Supplementary Information). In the Ramachandran plot, 91.8% of residues are in the 'most favourable' regions and none in 'disallowed' regions.

The structure of MutS alone was determined by molecular replacement using CNS. The search model with DNA-binding domains removed gave better solutions than that with the entire MutS. The crystal space group was determined to be $P3_121$ with one MutS dimer in each asymmetric unit. The structure was refined to the R and R_{free} of 33.1% and 36.1%, respectively, with tight NCS restraints (see Table 1 in Supplementary Information). With the B value of the diffraction data near 90 Å^2 and the low-resolution diffraction, further refinement was judged to be unlikely to provide more information. The final model contains residues 131–197, 203–327, 340–352, 518–628 and 635–765 of one subunit, and residues 117–196, 202–391, 498–628 and 635–762 of the other subunit. In the Ramachandran plot, 98.9% of residues are in the 'most favoured' and 'additionally allowed' regions and none is in the 'disallowed' regions.

Sequence alignment and structure comparison were carried out as described[40]. Figures 4c and 6b were generated using O (ref. 44), Fig. 5 in Supplementary Information was made using GRASP (ref. 46), and the rest of the structure figures were made using RIBBONS (ref. 47).

Received 14 April; accepted 4 August 2000.

1. Modrich, P. & Lahue, R. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu. Rev. Biochem.* **65,** 101–133 (1996).
2. Umar, A. & Kunkel, T. A. DNA-replication fidelity, mismatch repair and genome instability in cancer cells. *Eur. J. Biochem.* **238,** 297–307 (1996).
3. Human Genome Mutation Database [online] (cited 25 March 2000) ⟨http://www.uwcm.ac.uk/uwcm/mg/ns/1/203983.html⟩ (1997).
4. HNPCC Mutation Database [online] (cited 25 March 2000) ⟨http://www.nfdht.nl/database/msh2.htm⟩ (1997).
5. Herman, J. G. *et al.* Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc. Natl Acad. Sci. USA* **95,** 6870–6875 (1998).
6. Ban, C., Junop, M. & Yang, W. Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. *Cell* **97,** 85–97 (1999).
7. Jiricny, J. Eukaryotic mismatch repair: an update. *Mutat. Res.* **409,** 107–121 (1998).
8. Allen, D. J. *et al.* MutS mediates heteroduplex loop formation by a translocation mechanism. *EMBO J.* **14,** 4467–4476 (1997).
9. Biswas, I. *et al.* Oligomerization of a MutS mismatch repair protein from *Thermus aquaticus*. *JBC* **274,** 23673–23678 (1999).
10. Buermeyer, A. B., Deschênes, S. M., Baker, S. M. & Liskay, R. M. Mammalian DNA mismatch repair. *Annu. Rev. Genet.* **33,** 533–564 (1999).
11. Eisen, J. A. A phylogenomic study of the MutS family of proteins. *Nucleic Acids Res.* **26,** 4291–4300 (1998).
12. Matic, I., Taddei, F. & Radman, M. Genetic barriers among bacteria. *Trends Microbiol.* **4,** 69–72 (1996).
13. Nakagawa, T., Datta, A. & Kolodner, R. Multiple functions of MutS- and MutL-related heterocomplexes. *Proc. Natl Acad. Sci. USA* **96,** 14186–14188 (1999).
14. Gradia, S., Acharya, S. & Fishel, R. The human mismatch recognition complex hMSH2-hMSH6 functions as a novel molecular switch. *Cell* **91,** 995–1005 (1997).
15. Alani, E., Sokolsky, T., Studamire, B., Miret, J. J. & Lahue, R. S. Genetic and biochemical analysis of Msh2p-Msh6p: role of ATP hydrolysis and Msh2p-Msh6p subunit interactions in mismatch base pair recognition. *Mol. Cell. Biol.* **17,** 2436–2447 (1997).
16. Bjornson, K. P., Allen, D. J. & Modrich, P. Modulation of MutS ATP hydrolysis by DNA cofactors. *Biochemistry* **39,** 3176–3183 (2000).
17. Su, S. S. & Modrich, P. *Escherichia coli* mutS-encoded protein binds to mismatched DNA base pairs. *Proc. Natl Acad. Sci. USA* **83,** 5057–5061 (1986).
18. Haber, L. T. & Walker, G. C. Altering the conserved nucleotide binding motif in the *Salmonella typhimurium* MutS mismatch repair protein affects both its ATPase and mismatch binding activities. *EMBO J.* **10,** 2707–2715 (1991).
19. Biswas, I. & Hsieh, P. Identification and characterization of a thermostable MutS homolog from *Thermus aquaticus*. *J. Biol. Chem.* **271,** 5040–5048 (1996).
20. Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9,** 1665–1672 (1990).
21. Yang, W. & Steitz, T. A. Recombining the structures of HIV integrase, RuvC and RNase H. *Structure* **3,** 131–134 (1995).
22. Gorbalenya, A. E. & Koonin, E. V. Superfamily of UvrA-related NTP-binding proteins implication for rational classfication of recombination /repair systems. *J. Mol. Biol.* **213,** 583–591 (1990).
23. Hung, L. -W. *et al.* Crystal structure of the ATP-binding subunit of an ABC transporter. *Nature* **396,** 703–707 (1998).
24. Berger, J. M. & Wang, J. C. Recent developments in DNA topoisomerase II structure and mechanism. *Curr. Opin. Struct. Biol.* **6,** 84–90 (1996).
25. Jones, M., Wagner, R. & Radman, M. Repair of a mismatch is influenced by the base composition of the surrounding nucleotide sequence. *Genetics* **115,** 605–610 (1987).
26. Malkov, V. A., Biswas, I., Camerini-Otero, R. D. & Hsieh, P. Photocross-linking of the NH2-terminal region of Taq MutS protein to the major groove of a heteroduplex DNA. *J. Biol. Chem.* **272,** 23811–23817 (1997).
27. Biswas, I. & Hsieh, P. Interaction of MutS protein with the major and minor grooves of a heteroduplex DNA. *J. Biol. Chem.* **272,** 23811–23817 (1997).
28. Bajski, S. R., Jackson, B. A. & Barton, J. K. DNA repair: models for damage and mismatch recognition. *Mutat. Res.* **447,** 49–72 (2000).
29. Ohndorf, U. M., Rould, M. A., He, Q., Pabo, C. O. & Lippard, S. J. Basis for recognition of cisplatin-modified DNA by high-mobility-group proteins. *Nature* **399,** 708–712 (1999).
30. Dickerson, R. E. DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucleic Acids Res.* **26,** 1906–1926 (1998).
31. 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,** 16115–16125 (1999).
32. Das Gupta, R. & Kolodner, R. D. Novel dominant mutations in *Saccharomyces cerevisiae* MSH6. *Nature Genet.* **24,** 53–56 (2000).
33. Holland, I. B. & Blight, M. A. ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans. *J. Mol. Biol.* **293,** 381–399 (1999).
34. Hopfner, K. P. *et al.* Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell* **101,** 789–800 (2000).
35. Zalevsky, J., MacQueen, A. J., Duffy, J. B., Kemphues, K. J. & Villeneuve, A. M. Crossing over during *Caenorhabditis elegans* meiosis requires a conserved MutS-based pathway that is partially dispensable in budding yeast. *Genetics* **153,** 1271–1283 (1999).
36. Studamire, B., Price, G., Sugawara, N., Haber, J. E. & Alani, E. Separation-of-function mutations in *Saccharomyces cerevisiae* MSH2 that confer mismatch repair defects but do not affect nonhomologous-tail removal during recombination. *Mol. Cell. Biol.* **19,** 7558–7567 (1999).
37. Gradia, S. *et al.* hMSH2-hMSH6 forms a hydrolysis-independent sliding clamp on mismatch DNA. *Mol. Cell* **3,** 255–261 (1999).
38. Grilley, M., Welsh, K. M., Su, S. -S. & Modrich, P. Isolation and characterization of the *Escherichia coli mutL* gene product. *J. Biol. Chem.* **264,** 1000–1004 (1989).
39. 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-PMS1 protein complexes. *J. Biol. Chem.* **273,** 9837–9841 (1998).
40. Ban, C. & Yang, W. Structural basis for MutH activation in *E. coli* mismatch repair and relationship of MutH to restriction endonucleases. *EMBO J.* **17,** 1526–1534 (1998).
41. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
42. Terwilliger, T. C. SOLVE: An automated structure solution for MAD and MIR. Edition 1.16 [online] ⟨http://www.solve.lanl.gov⟩ (1997).
43. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
44. Jones, T. A., Zou, J. -Y. & Cowan, S. W. Improved methods for building models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
45. Brünger, A. T. *et al.* Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
46. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and themodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11,** 281–296 (1991).
47. Carson, M. Ribbon models of macromolecules. *J. Mol. Graphics* **5,** 103–106 (1987).
48. Drotschmann, K., Clark, A. B. & Kunkel, T. A. Mutator phenotypes of common polymorphisms and missense mutations in MSH2. *Curr. Biol.* **9,** 907–910 (1999).
49. Drotschmann, K. *et al.* Mutator phenotypes of yeast strains heterozygous for mutations in the MSH2 gene. *Proc. Natl Acad. Sci. USA* **96,** 2970–2975 (1999).
50. Wu, T. -H. & Marinus, M. G. Dominant negative mutator mutations in the *mutS* gene of *Escherichia coli*. *J. Bacteriol.* **176,** 5393–5400 (1994).

Supplementary information is available on *Nature*'s World-Wide Web site (http://www.nature.com) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

We thank I. Biswas for TAQ MutS expression vectors; Z. Dauter and C. Ogata for synchrotron beamline support; G. Poy for oligonucleotide synthesis; Q. Zhao for assistance in data collection; E. Alani, D. Camerini-Otero, R. Craigie, M. Gellert, M. Junop, T. Kunkel, D. Leahy, K. Mizuuchi, H. Nash and M. Radman for discussions and comments on the manuscript.

Correspondence and requests for materials should be addressed to W.Y. (e-mail: Wei.Yang@nih.gov). Coordinates have been deposited in the Protein Data Bank under accession codes 1EWR (MutS) and 1EWQ (MutS–DNA complex).