Structural basis for MutH activation in *E.coli* mismatch repair and relationship of MutH to restriction endonucleases

Changill Ban and Wei Yang¹

Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

¹Corresponding author
e-mail: Wei.Yang@nih.gov

MutS, MutL and MutH are the three essential proteins for initiation of methyl-directed DNA mismatch repair to correct mistakes made during DNA replication in *Escherichia coli*. MutH cleaves a newly synthesized and unmethylated daughter strand 5' to the sequence d(GATC) in a hemi-methylated duplex. Activation of MutH requires the recognition of a DNA mismatch by MutS and MutL. We have crystallized MutH in two space groups and solved the structures at 1.7 and 2.3 Å resolution, respectively. The active site of MutH is located at an interface between two subdomains that pivot relative to one another, as revealed by comparison of the crystal structures, and this presumably regulates the nuclease activity. The relative motion of the two subdomains in MutH correlates with the position of a protruding C-terminal helix. This helix appears to act as a molecular lever through which MutS and MutL may communicate the detection of a DNA mismatch and activate MutH. With sequence homology to *Sau3AI* and structural similarity to *PvuII* endonuclease, MutH is clearly related to these enzymes by divergent evolution, and this suggests that type II restriction endonucleases evolved from a common ancestor.

Keywords: activation/DNA repair/evolution/MutH structures/restriction enzymes

Introduction

DNA repair is an essential process in all living organisms. Mismatches in base pairing occur frequently during DNA replication or homologous recombination and routinely are repaired by a number of different yet overlapping mechanisms (Modrich, 1991; Modrich and Lahue, 1996). One prominent DNA repair pathway in *Escherichia coli* is methylation dependent and is initiated by the MutS, MutL and MutH proteins (Modrich, 1991). In a wild-type cell, the DNA template is methylated at d(GATC) sequences by the DAM methylase, but a newly synthesized strand is not methylated until it is modified by the methylase (Barras and Marinus, 1989). Different methylation states distinguish a parental strand from a daughter strand and direct DNA repair machinery to the latter. MutS is an ATPase that recognizes a mismatched base pair as well as an insertion or deletion of 1–4 nucleotides in one strand. MutH is an endonuclease that is both sequence-specific and methylation-specific; when activated, it

cleaves 5' to the unmethylated d(GATC) sequence in a hemimethylated duplex. MutL mediates the interaction between the site of mismatch bound by MutS and the d(GATC) sequence cleaved by MutH. The two sites can be separated by as much as 1000 bp. Once a nick is introduced in the daughter strand by MutH, DNA exonuclease, helicase, single-strand binding protein and polymerase are recruited to remove nucleotides from the nick to beyond the mismatch and to fill in the resulting gap (Modrich, 1991). Homologs of MutS and MutL have been found in various organisms, including humans, and have also been shown to be involved in DNA repair (see reviews by Kolodner, 1996; Modrich and Lahue, 1996). The phenotype of defects in these DNA repair proteins in eukaryotes is instability of microsatellite repeats. The overwhelming majority of hereditary non-polyposis colorectal cancers (HNPCC) and some sporadic cancers have been attributed to mutations in genes encoding MutS and MutL homologs.

MutH is a 28 kDa endonuclease composed of 229 residues, which is inactive until activated by MutS and MutL associated with a mismatched base pair (Au *et al.*, 1992). The endonuclease activity of MutH is dependent on Mg²⁺. Sequence analysis shows that MutH is homologous to the type II restriction endonuclease *Sau3AI* (Figure 1), but not homologous to *DpnII* or *MboI*, which are isoschizomers of *Sau3AI*. Both MutH and *Sau3AI* recognize the sequence d(GATC) and cleave 5' to the G. However, MutH cleaves only the unmethylated strand of hemimethylated d(GATC) sequence while *Sau3AI* makes a double strand break regardless of its methylation state. An interesting question arises as to how a restriction enzyme-like protein, MutH, has evolved to be a regulated endonuclease that is specific for methyl-directed DNA mismatch repair.

Hundreds of restriction enzymes have been characterized, which display little amino acid sequence similarity and show no apparent evolutionary relationship. Many of these enzymes form dimers in solution and cleave DNA symmetrically at inverted repeats. Structures of *BamHI* (Newman *et al.*, 1994), *EcoRV* (Winkler *et al.*, 1993), *PvuII* (Athanasiadis *et al.*, 1994) and their complexes with DNA have been determined separately (Winkler *et al.*, 1993; Cheng *et al.*, 1994; Newman *et al.*, 1995). Crystal structures of *EcoRI* (Kim *et al.*, 1990) and *FokI* (Wah *et al.*, 1997) in complex with DNA have also been determined. Among these five endonucleases, only three key residues in the active sites are similar in composition and tertiary arrangement. Based on cleavage pattern and overall structure, these enzymes can be divided into at least three classes. *BamHI* and *EcoRI* make a 4 bp staggered cleavage and generate 5' overhangs. Their structural similarity has been noted and an evolutionary relationship has been proposed (Newman *et al.*, 1994).

Fig. 1. The structure and sequence of MutH. (A) A ribbon diagram of MutH. Helices are shown in red and β -strands in blue. Secondary structure elements, mobile loops, active site residues and the sites of the two temperature-sensitive mutants are labeled. (B) Sequence alignment of the MutH family, including MutH from *Haemophilus influenzae* and Sau3AI (Seeber *et al.*, 1990) from *Staphylococcus aureus*. Active site residues are printed in red. Conserved residues in the hydrophobic cores are shaded in yellow, residues important for structural stability in green, residues in the cleft and potentially interacting with DNA in blue, and other conserved residues in pink.

Both *PvuII* and *EcoRV* cleave DNA to generate blunt ends. Their structures are related but clearly differ from those of *BamHI* and *EcoRI* (Athanasidis *et al.*, 1994). These two classes of endonucleases also differ in their mode of DNA binding. *BamHI* and *EcoRI* approach DNA in the major groove, while *PvuII* and *EcoRV* approach the minor groove. *FokI* represents a type of bipartite restriction enzymes. It is monomeric and has two separated domains, a DNA-binding domain and a catalytic domain. The catalytic domain of *FokI* is very similar to a *BamHI* protomer but cleaves any sequence of DNA 9 and 13 bp away from the recognition site (Wah *et al.*, 1997). In the

crystal structure of the *FokI*-DNA complex, the catalytic domain is packed next to the DNA-binding domain rather than being close to the scissile bonds. It has been a puzzle whether these different classes of restriction endonucleases are derived from a common ancestor.

We have crystallized native and selenomethionine-substituted MutH from *E. coli* in two space groups and have determined the structure of MutH in each of the crystal forms to resolutions of 2.3 and 1.7 Å, respectively. One of the crystal forms has two MutH molecules related by an imperfect dyad axis in one asymmetric unit. The three distinct structures thus obtained reveal a number of

interesting aspects of MutH. First, MutH is composed of two subdomains that adopts multiple conformations, with the two subdomains pivoting relative to one another. Secondly, the active site of MutH is similar to those of restriction endonucleases in primary, secondary and tertiary structure. Thirdly, a large cleft is formed between the two subdomains within a single molecule of MutH, with the active site located at its bottom, thus identifying a site for DNA binding. Fourthly, the active site of MutH is located at the interface between the two pivoting subdomains and undergoes conformational changes, which conforms with its regulated activity. The conformational change of the subdomains is correlated with the movement of a protruding C-terminal helix and the following hydrophobic tail, which appear to act as a molecular lever for MutS and MutL to activate MutH. Finally, the fact that monomeric MutH is structurally similar to a *PvuII* protomer prompts comparison of the structures of restriction enzymes at the tertiary rather than quaternary level. The structural similarity noted here between two subsets of restriction endonucleases, represented by *Bam*HI and *Eco*RV, supports a shared ancestor for these enzymes.

Results and discussion

Overall structure

The structure of MutH resembles a clamp, with a large cleft dividing the molecule into two halves (Figure 1A). Each half of MutH forms a subdomain that contains similar structural elements. The subdomain N-arm contains residues 1–83 and 120–145 that form helices α A, α B and α C, and a mixed β -sheet (strands 1, 2, 4 and 5). The subdomain C-arm contains residues 90–117 and 148–229 that fold into three helices (α D, α E and α F), an anti-parallel β -sheet (strands 3, 6 and 9) and a β -hairpin (strands 7 and 8) appended to its tip (Figure 1A). The two subdomains share a hydrophobic interface and are connected by three polypeptide linkers. While each subdomain as a structural entity is conserved among the three MutH structures, the interface between them is apparently flexible and allows the two subdomains to pivot relative to each other. All secondary structures are well defined in MutH. Three loops BC, C1 and 67, which are named after the helices and strands that they connect, appear rather flexible and are ordered in some cases due to lattice contacts in the crystals.

Unexpectedly, MutH is structurally very similar to *PvuII* endonuclease (Figure 2) despite a lack of detectable amino acid sequence similarity and functional differences including recognition of different DNA sites and a blunt versus staggered cleavage pattern. Unlike *PvuII*, MutH is a monomer in solution as well as when crystallized. Monomeric MutH is probably the active species of this enzyme because it cleaves only one DNA strand in mismatch repair. When MutH makes a double strand break in an unmethylated d(GATC) sequence, cleavage of each strand occurs independently (Welsh *et al.*, 1987; Au *et al.*, 1992).

Putative active site

The active site of MutH is apparent based on its structural similarity to those of the restriction enzymes and the conservation of the active site residues among MutH

Fig. 2. A ribbon diagram of the *PvuII* dimer. *PvuII* is viewed in exactly the same orientation as MutH in Figure 1A after superimposition of one of its subunits onto MutH. The superimposed subunit is colored in the same fashion as MutH in Figure 1A. The second subunit is colored pink. Secondary structure elements are labeled as in MutH. The additional β -strand at the C-terminus is labeled strand 10, which has no counterpart in MutH. The three conserved active site residues are also labeled. The coordinates of *PvuII* are taken from the complex with DNA (Cheng *et al.*, 1994). If shown, the DNA would be in the cleft formed between the two subunits, with the helical axis roughly perpendicular to the plane of the page.

homologs (Figure 1B). The catalytic sequence motif D(X)_{6–30}(E/D)XK, found in many type II restriction enzymes including several with known structures (Anderson, 1993), is present in MutH (Asp70, Glu77 and Lys79) and resides on the consecutive strands 1 and 2 at the bottom of the cleft (Figure 1A). On one side of the cleft, Glu56 forms a water-mediated hydrogen bond to Glu77. Glu56 is structurally equivalent to Glu45 in *EcoRV*, which has been shown to coordinate a Mg²⁺ ion and is critical for catalysis in *EcoRV* (Kostrewa and Winkler, 1995). On the other side of the cleft, Lys116 from the C-arm, which is conserved among the MutH homologs, makes a salt bridge to Glu77. The sites of two reported temperature-sensitive mutants in MutH, T28L and D157N (Grafstrom and Hoess, 1987), are both >20 Å away from the active site (Figure 1A). In the crystal structures, Thr28 provides an N-capping to helix α B, and Asp157 stabilizes a β -turn; therefore, they are probably important for the structural integrity of MutH. Immediately adjacent to the active site, loop C1 (between α C and β 1) adopts two different conformations or is completely disordered in the three crystal structures of MutH arising from the two space groups. Interestingly, counterparts of this loop in other restriction enzymes are often observed to be mobile in the absence of cognate DNA (Winkler *et al.*, 1993; Athanasiadis *et al.*, 1994; Newman *et al.*, 1994).

Fig. 3. DNA-binding cleft. Orthogonal views of the MutH molecular surface mapped with positive (blue) and negative (red) electrostatic potentials. The molecular surface is generated from the molecule with the most 'open' conformation (the red one in Figure 4A). A bent DNA borrowed from the *EcoRV*-DNA complex is modeled into the DNA-binding cleft. Loops BC and 67 are partially removed in order to produce an unobstructed view of the cleft. The subdomains, the molecular lever and the conserved residues in the cleft are labeled.

DNA-binding cleft

The location of the MutH active site indicates that DNA must bind in the cleft between the N-arm and the C-arm. This cleft, which is 15–18 Å wide and 12–14 Å deep, is of a size comparable with the substrate-binding cleft found in *PvuII* and *EcoRV* complexed with DNA (Winkler *et al.*, 1993; Cheng *et al.*, 1994). The active site is off-center at one end of the cleft, suggesting that the major groove of DNA approaches the protein, because only in such an arrangement would d(GATC) be in contact with the cleft when the scissile bond 5' to the G is positioned in the active site (Figure 3). The length of the cleft (~25 Å) allows it to contact at least 7 bp if the DNA is in B-form. Positively charged loop BC (between α B and α C) and loop 67 (between β 6 and β 7) form a gateway to the cleft and are quite flexible. The importance of these two loops is suggested by the number of strictly conserved residues which they include (Figure 1B). In addition, the DNA-binding cleft is lined with Asp91 and Phe94, which are conserved in the MutH family. Phe94 is exposed entirely to solvent and perhaps plays a role in DNA recognition or may even intercalate between base pairs (Figure 3).

The β -hairpin composed of strands 7 and 8 is tethered to the C-arm by two loops, one of which is the flexible loop 67 (Figure 1A), and it interacts with the C-arm by only a few van der Waals contacts. With three conserved hydrophobic residues, Leu193, Ile197 and Ile204 (Figure 1B), exposed to solvent, this structural appendage is stabilized at the tip of the C-arm by lattice contacts in the crystals. This β -hairpin may be dislodged from the C-arm and brought into contact with DNA by rotation around the connecting loops when DNA is present.

MutH activation mechanism

The most remarkable feature of the MutH structures is the pivoting of one subdomain relative to the other, which results in the closing and opening of the DNA-binding cleft (Figure 4A). Comparison of the three structures of MutH shows that the individual subdomains can be superimposed with an r.m.s. deviation of <1.0 Å over ~100 C α atoms. However, a small shift at the hydrophobic interface between the two subdomains results in displacements as large as 11 Å at the tip of one subdomain after superposition of the other (Figure 4A). Although there are substantial conformational changes in *BamHI*, *EcoRV* and *PvuII* upon association with DNA, their active sites remain the same with or without DNA (Winkler *et al.*, 1993; Athanasiadis *et al.*, 1994; Cheng *et al.*, 1994; Newman *et al.*, 1994, 1995). In contrast, the active site of MutH, located at the subdomain interface, undergoes conformational changes even in the absence of DNA. Pivoting of the two subdomains in MutH results in changes of C α distances between active site residues of ~1.0 Å. Although small, such a change may be sufficient to act as a switch that regulates MutH catalytic activity.

Rotation of the subdomains in the MutH crystal structures appears to correlate with the position of the C-terminal helix α F. The N-terminal half of α F is buried in the hydrophobic core and forms an integral part of the C-arm. The C-terminal half of α F makes hydrophobic contacts to the N-arm and serves as a pivoting point. When α F is shifted toward the N-arm, the entire C-arm rotates and widens the cleft, as observed in the most open structure of MutH (Figure 4A). A C-terminal tail following α F provides a handle to maneuver this molecular lever. In addition, it can be lengthened by unwinding of the C-terminus of α F, thus extending further away from the main body of the protein as observed in one of three MutH structures (Figure 4A). In both crystal forms, this rather hydrophobic tail is involved in intermolecular contacts. This C-terminal tail in MutH, therefore, is probably a site of interaction with MutS and MutL that directs the opening and closing of MutH subdomains, thereby activating MutH (Figure 4B). This proposed interface of MutH with MutS and MutL and the activation mechanism need to be verified in future experimental tests such as protein cross-linking and mutagenic studies.

Interestingly, the catalytic domain of *Sau3AI* shows extensive sequence homology to MutH; 16% of the sequences are identical, including all the catalytic residues, and >40% are homologous when the conservative changes are included (Figure 1B). Based on the sequence homology and similarity in catalysis of the two endonucleases, we predict that the structure of the catalytic domain of *Sau3AI* is similar to that of MutH. In contrast to the conditional

catalytic activity of MutH, *Sau3AI* is independently active and possesses an additional 270 residues at the C-terminus of the catalytic domain. The C-terminus in MutH is distant from the active site and is located on the opposite surface relative to the DNA-binding cleft. Therefore, the additional residues in *Sau3AI* are unlikely to be contributing directly to either DNA binding or catalysis but may function to support the catalytic domain structurally. The structural role of these residues in *Sau3AI* may be substituted by MutL and MutS in the case of MutH.

Relationship between MutH and PvuII

The structures of MutH and *PvuII* are strikingly similar in overall fold except that *PvuII* with only 157 residues has a few deletions relative to MutH. The most prominent one is the deletion of strands 6, 7 and 8, which results in a direct connection between helix E and strand 9 in *PvuII* (Figure 2). MutH and *PvuII* differ in conformation at the N-terminus. In *PvuII*, helices A and B are fused into one long and bent helix, which extends away from one subunit and interacts with helix C of the other subunit to form a dimer (Figure 2). Excluding the first 33 amino acid residues that form the dimer interface in *PvuII*, MutH and *PvuII* can be superimposed with an r.m.s. deviation of 2.3 over 83 pairs of C α atoms.

The DNA-binding cleft in MutH is formed between two subdomains within one molecule, while in the case of *PvuII* it is formed between two subunits (Figures 1 and 2). *PvuII* possesses an additional β -strand after helix α F at its C-terminus, which hydrogen-bonds to strand 5 of the central β -sheet in the N-arm (Figure 2) and intimately joins the two subdomains in *PvuII*. In addition, the linkers

between the two subdomains, particularly between β 5 and α E, are shorter in *PvuII* than those in MutH (Figure 2); thus, the region in *PvuII* that is equivalent to the C-arm appears to be much closer to the N-arm and the cleft found in MutH is closed off. The regions of *PvuII* homologous to the two subdomains in MutH are, therefore, merged into a single structural unit.

If the dimeric structure of *PvuII* is compared with MutH, the regions equivalent to the N-arm in the two proteins become more similar, since helix A from the neighboring subunits in *PvuII* occupies the same position as helix A in MutH (Figures 1 and 2). Moreover, the DNA-binding clefts in MutH and *PvuII*, although formed very differently, actually overlap with one another after superposition of one subunit of *PvuII* with MutH (Figures 1 and 2). The structural conservation between MutH and *PvuII* strongly suggest that they arose by divergent evolution.

Comparison of MutH and restriction endonucleases

The active site of MutH consisting of Asp70, Glu77 and Lys79 (the DEK triad) is similar to that of *BamHI*, *EcoRI*, *EcoRV*, *FokI* and *PvuII* (Figure 5A). This DEK triad is the core of the active site motif D(X)₆₋₃₀(E/D)XK found in many restriction endonucleases (Anderson, 1993). The first Asp residue in the DEK triad is absolutely conserved in all of the above six endonucleases. The second carboxylate is more often a glutamate than an aspartate. The lysine is always separated from the second carboxylate by one residue that tends to be hydrophobic. *BamHI* is exceptional, with a glutamate replacing the Lys residue in

Fig. 4. Pivoting of the subdomains. (A) A stereo view of the overlay of the three MutH C α traces after superimposing the first 60 C α atoms at the N-terminus. Every tenth C α atom is labeled with a ball. The blue and the red structures are derived from the monoclinic crystal and the yellow one from the orthorhombic crystal. The catalytic residues of the blue molecule, which has the most complete trace, are shown. Three linkers between the two pivoting subdomains are labeled 1, 2 and 3. (B) A simplified diagram of how the molecular lever pivots the C-arm relative to the N-arm. The correlated motion of the lever and the C-arm is denoted by arrows of the same color. Two nearly orthogonal rotations are observed and indicated by blue and green arrows.

Fig. 5. Structural comparison of endonucleases. (A) Ribbon diagrams of MutH, the catalytic domain of *FokI*, and single protomers of *BamHI*, *EcoRI*, *EcoRV* and *PvuII*. They are viewed in exactly the same orientation after superimposing the DEK triad and the central β -sheet. Active sites are labeled with the catalytic residues in ball-and-stick, red for the acidic and blue for the basic residues. The β -sheet common among these six structures is colored blue, the pair of helices N-terminal to the active site in green (in *EcoRI*, one of the two helices is actually from the C-terminus). The second helix in *PvuII* is provided by a dimeric partner not shown here. The structural elements in *EcoRV* and *PvuII* which are similar to those in the C-arm of MutH are colored magenta. The pair of helices that are part of the Rossmann fold in *BamHI*, *EcoRI* and *FokI* are shown in lime green. (B) Structures of *BamHI* and *EcoRV* in complex with DNA. *BamHI* and *EcoRV* are viewed in similar orientations after superposition of the DEK triad and the $\alpha\beta$ barrel in one protomer. After the view of the *BamHI*-DNA complex was chosen, the *EcoRV*-DNA complex was rotated slightly to bring its dyad axes into the viewing plane. Proteins are shown as ribbon diagrams using the same color scheme as in (A). The active site of the second protomer is omitted for clarity. DNA molecules are represented as red sticks; the scissile bonds are highlighted in dark green.

the DEK triad. The spatial arrangement of the DEK triad on two consecutive and structurally adjacent β -strands is conserved regardless of the separation between the two carboxylates, which varies from 6 to 30 amino acid

residues (Figure 5A). The first Asp residue always appears at the N-terminus of the first β -strand and adjacent to the second carboxylate on the next strand. Sequence variation between the two carboxylates only affects the length of

Fig. 6. Topology diagrams of MutH and BamHI. Helices are shown as open bars and β -strands as arrows. Conserved structural elements in the $\alpha\beta$ barrel are colored green (helices) and blue (strands). The locations of the active site residues are marked by red stars. Regions that are proposed to be involved in domain substitution are boxed with dashed lines.

the strands, which are shortest in MutH and longest in *EcoRI*. These two strands are conserved in a central β -sheet that appears in all six endonuclease structures. Depending on the sequence inserted between the two carboxylates, the DEK triad occurs either at one end of the central β -sheet as in MutH and *PvuII*, in the middle as in *EcoRV* or towards the other end as in *BamHI*, *EcoRI* and *FokI* (Figure 5A). Despite functional differences among these six endonucleases, conservation of the active site in primary, secondary and tertiary structure provides compelling evidence for divergent evolution of these enzymes.

Structural similarities among these six endonucleases extend to the region equivalent to the N-arm of MutH, which consists of most of the N-terminal half of each of the other five proteins. After superposition of the DEK triad and the central β -sheet, a similar pair of helices, equivalent to αA and αC in MutH, emerges on the same side of the β -sheet as in *BamHI* and *EcoRV* (Figure 5A). These two helices and the central β -sheet can best be characterized as an $\alpha\beta$ barrel, which is quite prominent in *BamHI* and *EcoRV* (Figure 5B). Between *BamHI* and *EcoRV*, 74 pairs of C α atoms in this $\alpha\beta$ barrel can be superimposed with an r.m.s. deviation of 2.6 Å. Superposition of the same regions between *BamHI* and *EcoRI* results in an r.m.s. deviation of 2.7 Å. Structural elements in this $\alpha\beta$ barrel do have some variations. The β -strand corresponding to $\beta 5$ in MutH orients in opposite directions in *BamHI*, *EcoRI* and *FokI* compared with *EcoRV*, *PvuII* and MutH (Figure 6). In *PvuII*, the N-terminal helices are exchanged between the dimeric partners so that one of the helices in the $\alpha\beta$ barrel comes from another protomer (Figure 2). Curiously, this helix in *EcoRI* is from the C- instead of the N-terminus and runs in the opposite direction to that of *BamHI* or *EcoRV*, which may seem to be a result of convergent rather than divergent evolution. However, after carefully comparing the structures of *BamHI* and *EcoRI*, Newman *et al.* (1994) proposed a divergent evolutionary scheme to account for the differences between these two endonucleases.

The structures of these six endonucleases differ most in the C-terminal half of the molecule, which is involved in DNA recognition, and can be divided into two structural subsets as illustrated in Figure 5A. *BamHI*, *EcoRI* and

FokI form one subset; MutH, *PvuII* and *EcoRV* form the other. Members in the subset represented by *BamHI* contain a Rossmann fold, a β - α - β motif frequently found in nucleotide-binding proteins (Rossmann *et al.*, 1975). Members of the other subset, including MutH, possess a rather extended structure instead of a Rossmann fold. This difference in the C-terminal half of the molecule appears to be a result of domain substitution (Figure 6). The β -strand that runs in the opposite direction in the $\alpha\beta$ barrel can be explained by the domain substitution as well since it is one of the strands in the Rossmann fold (Figure 6). Domain substitution clearly took place in the bipartite *FokI* endonuclease; its catalytic domain is tethered to a helix-turn-helix-containing DNA recognition domain (Wah *et al.*, 1997).

The structural diversity among *PvuII*, *EcoRV*, *BamHI* and *EcoRI* has been noted as striking, and differences in their subunit interfaces as fundamental (Cheng *et al.*, 1994). MutH, as a monomeric endonuclease, shows that dimerization is not a prerequisite for DNA binding or cleavage and offers a new perspective from which to compare the rather diverse structures of the four dimeric restriction enzymes. After superposition of the DEK triad and the $\alpha\beta$ barrel in one protomer, as illustrated in Figure 5B, the differences among the endonucleases in quaternary structure, that is in the mode of DNA binding and in the dimer formation, do not appear as fundamental as previously thought (Cheng *et al.*, 1994). Regarding DNA-protein interactions, the DNA duplex adopts a similar orientation relative to the DEK triad in both *BamHI* and *EcoRV* (Figure 5B), each of which represents one of the two structural subsets defined earlier. In both cases, recognition of a specific DNA sequence occurs in the major groove despite the fact that this is accomplished by two helices in *BamHI* and by two loops in *EcoRV*. With regard to dimerization, the second protomer of *BamHI* joins at the 'C-arm' and the two $\alpha\beta$ barrels lie side by side; the second protomer of *EcoRV*, however, approaches toward the 'N-arm' and the two $\alpha\beta$ barrels are stacked head to head (Figure 5B). The orientation of the second protomer in *BamHI* and *EcoRV* depends on the position of the second scissile bond (Figure 5B). When the cleavages are staggered with 5' overhangs, as carried out by *BamHI* and *EcoRI*, two active sites in the nuclease dimer approach the major groove because the distance between the two scissile bonds is shortest across the major groove. When cleavages are blunt as in the case of *PvuII* and *EcoRV* or staggered with 3' overhangs, the minor groove is approached instead, but for the same reason (Figure 5B). The overall orientation of protein relative to DNA thus reflects the nature of the double strand cleavage as suggested previously (Anderson, 1993). The differences in quaternary structures that Cheng *et al.* (1994) observed can be viewed as a result of rearrangement of protein subunits in order to adapt to various DNA cleavage patterns. These differences, therefore, provide evidence in support of divergent evolution.

In conclusion, MutH and the five other endonucleases compared here share a catalytic core of an $\alpha\beta$ barrel which presents the conserved DEK triad, but fall into two subsets because of structural differences in surrounding regions. We propose an evolutionary relationship between these two subsets by domain substitution. The conserved

Table I. Summary of crystallographic data

Space group	P2 ₁	P2 ₁ 2 ₁ 2 ₁	
Unit cell (<i>a</i> , <i>b</i> , <i>c</i>)	39.9, 111.7, 55.1 (Å), β = 96.5°	41.6, 72.6, 72.9 (Å)	
Crystal content	Native	EMP	SeMet
ASU content	2 molecules	2 molecules	1 molecule
Diffraction data	20.0–2.28 Å	20.0–2.48 Å	20.0–1.7 Å
Last resolution bin	2.37–2.28 Å	2.61–2.48 Å	1.76–1.7 Å
Completeness (%) ^a	95.2 (83.8)	93.1 (69.2)	86.7 (43.8)
<i>R</i> _{merge} (%) ^{a,b}	7.9 (36.9)	6.3 (20.6)	3.0 (12.4)
<i>R</i> _{iso} (F) ^c		18.7% (20.0–2.5 Å)	
Phasing power ^d		1.56 (10–2.8 Å)	
Figure-of-merit		0.457 (10–2.8 Å)	
Refinement			
Protein atoms	3502		1689
Solvent atoms	147		258
Unique reflections	19 944 (20–2.3 Å)		20 976 (8–1.7 Å)
<i>R</i> and (<i>R</i> -free) ^e	22.5% (28.9%)		20.8% (26.3%)
R.m.s. bond length	0.006 Å		0.004 Å
R.m.s. bond angle	1.399°		1.156°
Average <i>B</i> of MC ^f	23.5 and 33.7 Å ²		16.3 (1.1) Å ²
Average <i>B</i> of SC ^f	26.9 and 36.7 Å ²		19.8 (2.3) Å ²

EMP, ethyl mercury phosphate; SeMet, selenomethionine-substituted MutH.

^aThe data of the last resolution bin is shown in parentheses.

^b $R_{\text{merge}} = \sum_i |I_{hi} - \langle I_h \rangle| / \sum_i I_{hi}$.

^c $R_{\text{iso}} = \sum_i |F_p| - |F_{ph}| / \sum_i |F_p|$.

^dPhasing power = $\langle |F_h| \rangle / \text{r.m.s.}(\epsilon)$, i.e. the mean amplitudes calculated from the heavy-atom model divided by the lack of closure error.

^e $R = \sum_i |F_{\text{obs}}| - |F_{\text{calc}}| / \sum_i |F_{\text{obs}}|$; *R*-free is monitored with 10% of the reflection data excluded from refinement.

^fMC, main chain atoms; SC, side chain atoms. The *B*-value of each molecule in the P2₁ space group is listed separately; r.m.s. deviations are shown in parentheses.

catalytic core and variations in its quaternary arrangement that correlate with functional differences provide a strong argument for a divergent evolutionary relationship among a larger number of restriction enzymes than previously suspected. Restriction enzymes, which are highly variable in sequence, still pose a challenge for structure prediction even after several structures have been determined and the catalytic core identified. Frequent alteration in the sequence of the catalytic core and the structure elements surrounding it seem to have enabled this family of endonucleases to acquire distinct biological functions. MutH can be viewed as a restriction enzyme whose activity is regulated by MutS and MutL. The revelation of two subdomains in MutH and their pivoting motion suggest a mechanism for MutH activation. As the first structure available in the MutSLH trio, MutH also provides a cornerstone for studying the mechanism of mismatch repair.

Materials and methods

Protein expression, purification and crystallization

His-tagged *E. coli* MutH was overexpressed as described (Feng and Winkler, 1995). Selenomethionine-labeled MutH was produced by transforming the methionine auxotrophic strain B834(DE3) with the MutH expression vector pTX417 and growing these cells in a defined medium containing selenomethionine (LeMaster and Richards, 1985). His-tagged MutH was purified over a Ni column. The His tag was removed by thrombin cleavage at room temperature. MutH was then purified over a mono-Q column and concentrated to 10 mg/ml in 20 mM Tris (pH 8.0), 60 mM KCl, 1 mM dithiothreitol (DTT), 0.5 mM EDTA and 5% glycerol. The monomeric nature of MutH was established by size exclusion chromatography of concentrated protein over a Superdex-75 column. N-terminal sequencing confirmed the removal of the His tag. The final purified MutH consists of its original 229 amino acids and three additional residues (Gly–Ser–His) as the remainder of the His tag.

Crystals were grown using the hanging drop method at 20°C and reached the maximal size within a week. The precipitant buffer contained 100 mM ammonium acetate (pH 5.8), 50 mM magnesium acetate, 1 mM DTT and 12–16% PEG 6000. Both monoclinic and orthorhombic crystals (Table I) were grown under similar conditions but the orthorhombic crystals grew only with fresh preparations of MutH. The ethylmercury phosphate (EMP) derivative was prepared by soaking crystals in the mother liquor with 20% PEG 6000 and 0.6 mM EMP for 24 h.

Structure determination

Diffraction data that contributed to the structure determination were collected at –160°C using an R-axis IPII detector mounted on a Rigaku RU200 generator. Ethylene glycol (20%) was added to the mother liquor for flash freezing. Diffraction data were indexed, integrated and scaled using the programs DENZO and SCALEPACK (Otwinowski and Minor, 1997). CCP4 programs (CCP4, 1994) were used for data reduction, phasing and electron density map calculation. The monoclinic crystal form was obtained first and its structure was determined by SIRAS using EMP as a single isomorphous derivative. There were two Cys residues per asymmetric unit. Two Hg sites were found using Patterson maps and refined using MLPHARE (CCP4, 1994). The initial Fourier map calculated at 3 Å after solvent flattening by DM in OMIT mode (CCP4, 1994) showed α helices. Hg sites were refined further with the external phases generated by DM. Molecular masks for non-crystallographic averaging were generated using RAVE (Kleywegt and Jones, 1994). An electron density map calculated to 2.8 Å after solvent flipping and non-crystallographic averaging (DM) was used to trace the peptide chain. The initial model had ~80% of the residues traced and was refined using standard protocols in X-PLOR (Brünger, 1992). The rest of the molecules were traced after iterations of refinement using X-PLOR and model rebuilding using O (Jones *et al.*, 1991). The experimental map after non-crystallographic averaging and solvent flattening was not as good as expected because the molecules related by a non-crystallographic 2-fold axis had very different average *B*-values (Table I) and different conformations due to the subdomain rotation. Non-crystallographic restraints were only used at early stages of refinement. Molecules related by the non-crystallographic 2-fold axis were packed with the DNA-binding cleft facing opposite directions. The final refined model contains all residues except five at the N-terminus, of which three were from the His tag. Loops BC, C1 and 67 are poorly defined in one of the two molecules.

The structure in the orthorhombic crystal form was solved by molecular replacement. The same solutions were found using either X-PLOR or AMoRe (Navaza, 1994). Because of the high quality of the diffraction data, refinement converged rapidly. Parts of loops BC, C1 and 67 were again disordered. The final refined model at 1.7 Å includes 11 ethylene glycol as well as 214 water molecules. Although Mg^{2+} ions were included in the crystallization buffer, no Mg^{2+} was found in the refined structure even at 2.3 or 1.7 Å resolution. This may be due to the flexibility of the active site and a lack of substrate. In both refined models, the majority of residues are in the most favorable conformation and the rest are favorable as judged by PROCHECK (CCP4, 1994).

Sequence analysis and structure comparison

Sequence similarity among MutH of *E.coli* and *Haemophilus influenzae* and *Sau3AI* was detected by BLITZ (Collins and Coulson, 1990) and the sequences were aligned using MULTALIN (Corpet, 1988). MutH structures were superimposed using the least-square function in O, and their r.m.s. deviation calculated using ALIGN (Cohen, 1998). The structural similarity among MutH, PvuII and EcoRV was first found using DALI (Holm and Sander, 1993). Coordinates of BamHI (1BHM), EcoRI (1ERI), EcoRV (1RVB), FokI (1FOK) and PvuII (1PVI) from their DNA co-crystals were used in the structure comparison with MutH. These structures first were aligned manually using O. ALIGN was then used to improved the superposition and to calculate r.m.s. deviations.

Structure figures were generated using RIBBONS (Carson, 1987), except Figure 3 which was generated using GRASP (Nicholls *et al.*, 1991).

The coordinates of both crystal structures of MutH have been deposited with the Brookhaven Protein Data Bank as 1AZO (orthorhombic form) and 2AZO (monoclinic form) and will be released in May 1998.

Acknowledgements

We thank J.Tormo and J.Wang for help with the structure determination; F.Jerva and S.Rhee for help with data collection at synchrotron; C.Ogata for the access to the beamline X4A at BNL; A.Agarwal and D.Wah for the coordinates of the FokI catalytic domain before scheduled release; and D.Leahy, B.Craigie and colleagues at LMB, NIDDK for comments on the manuscript.

References

- Anderson,J.E. (1993) Restriction endonucleases and modification methylases. *Curr. Opin. Struct. Biol.*, **3**, 24–30.
- Athanasiadis,A., Kotsifaki,D., Tucker,P.A., Wilson,K.S. and Kokkinidis,M. (1994) Crystal structure of PvuII endonuclease reveals extensive structural homologies to EcoRV. *Nature Struct. Biol.*, **1**, 469–475.
- Au,K.G., Welsh,K. and Modrich,P. (1992) Initiation of methyl-directed mismatch repair. *J. Biol. Chem.*, **267**, 12142–12148.
- Barras,F. and Marinus,M.G. (1989) The great GATC: DNA methylation in *E.coli*. *Trends Genet.*, **5**, 139–143.
- Brünger,A.T. (1992) *X-PLOR, Version 3.1, A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT.
- Carson,M. (1987) Ribbon models of macromolecules. *J. Mol. Graphics*, **5**, 103–106.
- CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.
- Cheng,X., Balendiran,K., Schildkraut,I. and Anderson,J.E. (1994) Structure of PvuII endonuclease with cognate DNA. *EMBO J.*, **13**, 3927–3935.
- Cohen,G.H. (1998) ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. *J. Appl. Crystallogr.*, in press.
- Collins,J.F. and Coulson,A.F.W. (1990) Significance of protein sequence similarities. *Methods Enzymol.*, **183**, 474–486.
- Corpet,F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.*, **16**, 10881–10890.
- Feng,G. and Winkler,M.E. (1995) Single-step purifications of His6-MutH, His6-MutL and His6-MutS repair proteins of *Escherichia coli* K-12. *BioTechniques*, **19**, 956–965.
- Grafstrom,R.H. and Hoess,R.H. (1987) Nucleotide sequence of the *Escherichia coli mutH* gene. *Nucleic Acids Res.*, **15**, 3073–3084.
- Holm,L. and Sander,C. (1993) Dali: version 2.0. *J. Mol. Biol.*, **233**, 123–138.
- Jones,T.A., Zou,J.-Y. and Cowan,S.W. (1991) Improved methods for

building models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.

- Kim,Y., Grable,J.C., Love,R., Greene,P.J. and Rosenberg,J.M. (1990) Refinement of EcoRI endonuclease crystal structure: a revised protein chain tracing. *Science*, **249**, 1307–1309.
- Kleywegt,G.J. and Jones,T.A. (1994) Masks and bones. In Wolf,W., Evan,P.R. and Leslie,A.G.W. (eds), *From First Map to Final Model. Proceedings of the CCP4 study weekend*. SERC Daresbury Laboratory, Warrington, UK, pp. 59–66.
- Kolodner,R. (1996) Biochemistry and genetics of eukaryotic mismatch repair. *Genes Dev.*, **10**, 1433–1442.
- Kostrewa,D. and Winkler,F.K. (1995) Mg^{2+} binding to the active site of EcoRV endonuclease: a crystallographic study of complexes with substrate and product DNA at 2 Å resolution. *Biochemistry*, **34**, 683–696.
- LeMaster,D.M. and Richards,F.M. (1985) 1H–15N heteronuclear NMR studies of *Escherichia coli* thioredoxin in samples isotopically labeled by residue type. *Biochemistry*, **24**, 7263–7268.
- Modrich,P. (1991) Mechanisms and biological effects of mismatch repair. *Annu. Rev. Genet.*, **25**, 229–253.
- Modrich,P. and Lahue,R. (1996) Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu. Rev. Biochem.*, **65**, 101–133.
- Navaza,J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr.*, **A50**, 157–163.
- Newman,M., Strzelecka,T., Dorner,L.F., Schildkraut,I. and Aggarwal,A.K. (1994) Structure of restriction endonuclease BamHI and its relationship to EcoRI. *Nature*, **368**, 660–664.
- Newman,M., Strzelecka,T., Dorner,L.F., Schildkraut,I. and Aggarwal,A.K. (1995) Structure of BamHI endonuclease bound to DNA: partial folding and unfolding on DNA binding. *Science*, **269**, 656–663.
- Nicholls,A., Sharp,K.A. and Honig,B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.*, **11**, 281–296.
- Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
- Rossmann,M.G., Liljas,A., Brandén,C.-I. and Banaszak,L.J. (1975) Evolutionary and structural relationships among dehydrogenases. In Boyer,P.D. (ed.), *The Enzymes*. Academic Press, New York, Vol. XI, pp. 61–102.
- Seeber,S., Kessler,C. and Goetz,F. (1990) Cloning, expression and characterization of the *Sau3AI* restriction and modification genes in *Staphylococcus carnosus* TM 300. *Gene*, **94**, 37–43.
- Wah,D.A., Hirsch,J.A., Dorner,L.F., Schildkraut,I. and Aggarwal,A.K. (1997) Structure of the multimodular endonuclease FokI bound to DNA. *Nature*, **388**, 97–100.
- Welsh,K.M., Lu,A.-L., Clark,S. and Modrich,P. (1987) Isolation and characterization of the *Escherichia coli mutH* gene product. *J. Biol. Chem.*, **262**, 15624–15629.
- Winkler,F. *et al.* (1993) The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.*, **12**, 1781–1795.

Received November 21, 1997; revised December 22, 1997;
accepted December 23, 1997